

**ANALISA PERBANDINGAN MODEL PREDICTION  
DALAM PREDIKSI HARGA SAHAM PADA TIGA  
SEKTOR INDUSTRI DI INDONESIA**



**TESIS**

Diajukan sebagai salah satu syarat untuk memperoleh gelar  
Magister Ilmu Komputer (M.Kom)

**EVITA FITRI**

**14002313**

Program Studi Ilmu Komputer (S2)  
Sekolah Tinggi Manajemen Informatika Dan Komputer  
Nusa Mandiri  
Jakarta  
2020

## PERSEMBAHAN

*“Bacalah dengan (menyebut) nama Tuhanmu  
yang menciptakan” (QS. Al-Alaq)*

*Dengan menyebut nama Allah Yang Maha Pengasih lagi Maha Penyayang,*  
Segala puja dan puji hanya milik Allah, yang tidak pernah berhenti memberika rahmat dan karunia-Nya kepada para hamba-Nya. Allah Maha Kuasa untuk menjadikan segala sesuatu yang sulit menjadi mudah. Shalawat dan salam semoga terlimpahkan kepada Nabi Muhammad *Shallallahu ‘alaihi wassalam*, keluarga dan para sahabatnya.

Atas karunia Allah *Subhanahu wa Ta’ala*, penulisan tesis ini telah dapat ku selesaikan dan dengan sepenuh hati ku persembahkan untuk:

1. Kedua orang tua ku yang sangat amat ku cintai, Mamah dan Ayah yang selalu membimbingku, mendukungku, memotivasi, serta yang selalu memberi apa yang terbaik bagiku, dan Ridhonya yang selalu aku butuhkan didalam hidupku.
2. Kakak ku Dedy Pratomo yang selalu menjadi panutan ku untuk terus bersemangat dalam mencapai pendidikan yang lebih tinggi, serta adik-adikku yang senantiasa menghiburku dengan senyuman-senyuman hangat mereka.
3. Fathurrohman yang selalu setia dalam memberikanku semangat, motivasi serta bantuan yang sangat membantu dalam penyelesaian pendidikan terakhirku saat ini.

Terimakasih untuk berbagai kebaikan yang terlalu panjang untuk ku jabarkan, semoga kita semua senantiasa dalam penjagaan Allah *Azza Wa Jalla*, Aamiin..

*“Ketika aku tahu, aku semakin tahu bahwa aku tidak tahu apa-apa”  
(Imam Safe’i)*

## SURAT PERNYATAAN ORISINAL DAN BEBAS PLAGIARISME

Yang bertanda tangan di bawah ini :

Nama : Evita Fitri  
NIM : 14002313  
Program Studi : Ilmu Komputer  
Jenjang : Program Magister (S2)  
Konsentrasi : *Data Mining*

Dengan ini menyatakan bahwa tesis yang telah saya buat dengan judul : “*Analisa Perbandingan Model Prediction Dalam Prediksi Harga Saham Pada Tiga Sektor Industri Di Indonesia*”, adalah hasil karya sendiri, dan semua sumber yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar dan tesis belum pernah diterbitkan atau dipublikasikan dimanapun dan dalam bentuk apapun.

Demikianlah surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila dikemudian hari ternyata saya memberikan keterangan palsu dan atau ada pihak lain yang mengklaim bahwa tesis yang telah saya buat adalah hasil karya milik seseorang atau badan tertentu, saya bersedia diproses baik secara pidana maupun perdata dan kelulusan saya dari Program Studi Ilmu Komputer (S2) Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri dicabut/dibatalkan.

Jakarta, 12 Januari 2021

Yang Menyatakan,



Evita Fitri

## PERSETUJUAN DAN PENGESAHAN TESIS

Tesis ini diajukan oleh:

Nama : Evita Fitri  
NIM : 14002313  
Program Studi : Ilmu Komputer  
Jenjang : Program Magister (S2)  
Konsentrasi : *Data Mining*  
Judul Tesis : “Analisa Perbandingan Model *Prediction* Dalam Prediksi  
Harga Saham Pada Tiga Sektor Industri Di Indonesia”

Telah dipertahankan pada periode 2020-2 dihadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Magister Ilmu Komputer (M.Kom) pada Program Studi Ilmu Komputer (S2) Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIK Nusa Mandiri).

Jakarta , 21 Januari 2021

### PEMBIMBING TESIS

Dosen Pembimbing : Dr. Dwiza Riana, S.Si, MM, M.Kom



### DEWAN PENGUJI

Penguji I : Dr. Hilman Ferdinandus Pardede,  
S.T, M.EICT



Penguji II : Dr. Agus Subekti, M.T



Penguji III/Dosen Pembimbing : Dr. Dwiza Riana, S.Si, MM, M.Kom



	<b>LEMBAR BIMBINGAN TESIS</b>
	<b>STMIK NUSA MANDIRI</b>

NIM : 14002313  
 Nama Lengkap : Evita Fitri  
 Dosen Pembimbing : Dr. Dwiza Riana, S.Si, MM, M.Kom  
 Judul Tesis : “Analisa Perbandingan Model *Prediction* Dalam  
 Prediksi Harga Saham Pada Tiga Sektor Industri  
 Di Indonesia”

NO	Tanggal Bimbingan	Pokok Bahasan	Paraf Dosen Pembimbing
1	20 September 2020	Pengajuan Judul Tesis	
2	21 September 2020	ACC Pengajuan Judul Tesis	
3	24 Oktober 2020	Konsultasi Jurnal Referensi dan Dataset	
4	14 November 2020	Pengajuan Metode dan Konsultasi Progres Uji Coba Penelitian	
5	06 Desember 2020	Pengajuan Bab 1 – Bab 2	
6	13 Desember 2020	Revisi Bab 1 – Bab 2 dan Pengajuan Bab 3 – Bab 5	
7	19 Desember 2020	Revisi Bab 3 – Bab 5	
8	02 Januari 2021	Acc Keseluruhan Bab 1 – Bab 5	

Catatan Untuk Dosen Pembimbing.

Bimbingan Tesis

- Dimulai pada tanggal : 20 September 2020
- Diakhiri pada tanggal : 02 Januari 2021
- Jumlah pertemuan bimbingan : 8 (Delapan) Kali Bimbingan

Disetujui oleh,

**Dosen Pembimbing**

(Dr. Dwiza Riana, S.Si, MM, M.Kom)



## KATA PENGANTAR

Puji syukur alhamdulillah, penulis panjatkan kehadiran Allah SWT yang telah melimpahkan rahmat serta karunia-NYA sehingga penulis pada akhirnya dapat menyelesaikan tesis ini dengan baik, dan dimana tesis ini penulis sajikan dalam bentuk sebuah buku yang sederhana, adapun judul yang penulis ambil untuk tesis ini adalah “*Analisa Perbandingan Model Prediction Dalam Prediksi Harga Saham Pada Tiga Sektor Industri Di Indonesia*”.

Tujuan dari dibuatnya penulisan tesis ini adalah untuk memenuhi salah satu syarat kelulusan dalam mendapatkan gelar Magister Ilmu Komputer pada Program Studi Ilmu Komputer (S2) STMIK Nusa Mandiri. Sebagai bahan penulisan tesis ini diambil berdasarkan hasil penelitian, observasi dan beberapa sumber literatur yang mendukung penulisan ini. Penulis menyadari bahwa tanpa adanya bimbingan serta dorongan dari semua pihak, maka penulisan tesis ini tidak dapat berjalan dengan baik. Oleh karena itu, perkenankan penulis mengucapkan beberapa kata terima kasih yang sebesar-besarnya kepada:

1. Ibu Dr. Dwiza Riana, S.Si, MM, M.Kom selaku Ketua STMIK Nusa Mandiri sekaligus pembimbing tesis bagi penulis yang telah membantu dalam menyelesaikan penulisan tesis ini.
2. Wakil Ketua STMIK Nusa Mandiri
3. Ketua Program Studi Ilmu Komputer STMIK Nusa Mandiri
4. Bapak/Ibu dosen Sistem Informasi STMIK Nusa Mandiri Jakarta
5. Seluruh staf di lingkungan STMIK Nusa Mandiri yang telah membantu penulis selama perkuliahan.
6. Kedua orang tua tercinta serta keluarga baik kakak ataupun adik yang selalu memberikan dukungan moril maupun spiritual.
7. Seluruh rekan teman-teman penulis yang telah banyak memberikan support positif terhadap penulis
8. Serta semua pihak yang terlalu banyak untuk disebutkan satu persatu sehingga terwujudnya penulisan ini.

Penulis menyadari bahwa penulisan laporan tesis ini masih jauh sekali dari sempurna, untuk itu penulis mohon kritik dan saran yang bersifat membangun demi kesempurnaan penulisan dimasa yang akan datang. Akhir kata semoga penulisan tesis ini dapat berguna khususnya bagi penulis dan bagi para pembaca yang berminat pada umumnya.

Bekasi, 12 Januari 2021

A handwritten signature in black ink, appearing to read 'Evita Fitri', with a stylized flourish at the end.

Evita Fitri

## SURAT PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan di bawah ini :

Nama : Evita Fitri  
NIM : 14002313  
Program Studi : Ilmu Komputer  
Jenjang : Program Magister (S2)  
Konsentrasi : *Data Mining*  
Jenis Karya : Tesis

Demi pengembangan ilmu pengetahuan, dengan ini menyetujui untuk memberikan izin kepada pihak Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIK Nusa Mandiri) **Hak Bebas Royalti Non-Eksklusif (*Non-Exclusive Royalty-Free Right*)** atas karya ilmiah kami yang berjudul: *Analisa Perbandingan Model Prediction Dalam Prediksi Harga Saham Pada Tiga Sektor Industri Di Indonesia beserta perangkat yang diperlukan (apabila ada).*

Dengan **Hak Bebas Royalti Non-Eksklusif** ini pihak STMIK Nusa Mandiri berhak menyimpan, mengalih-media atau *bentuk*-kan, mengelolaannya dalam pangkalan data (*database*), mendistribusikannya dan menampilkan atau mempublikasikannya di internet atau media lain untuk kepentingan akademis tanpa perlu meminta izin dari kami selama tetap mencantumkan nama kami sebagai penulis/pencipta karya ilmiah tersebut.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak STMIK Nusa Mandiri, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 12 Januari 2021

Yang Menyatakan,



Evita Fitri

## ABSTRAK

Nama : Evita Fitri  
NIM : 14002313  
Program Studi : Ilmu Komputer  
Jenjang : Program Magister (S2)  
Konsentrasi : *Data Mining*  
Judul : “Analisa Perbandingan Model *Prediction* Dalam Prediksi Harga Saham Pada Tiga Sektor Industri Di Indonesia”

Saham merupakan salah satu jenis investasi jangka panjang yang cukup banyak diminati oleh kalangan masyarakat karena investasi ini mendatangkan keuntungan yang terbilang cukup besar bagi para investornya. Namun berkaitan dengan hal tersebut, pergerakan harga saham pada umumnya cenderung bersifat non linear dan non stasioner, hal ini dikarenakan harga saham dapat dipengaruhi oleh beberapa faktor yang hasil dapat merubah pola nilai harga saham baik naik atau turun, sehingga hal ini dapat mempersulit dalam memprediksi harga saham. Dalam penelitian ini, dilakukan analisa perbandingan model *prediction* dalam memprediksi harga saham menggunakan pendekatan teknikal berdasarkan data masa lampau, adapun data yang digunakan adalah historis harga saham dengan mengambil sampel data pada tiga emiten dari pasar modal Indonesia. Terdapat tiga metode yang diuji coba pada penelitian ini, diantaranya *Linear Regression* (LR), *Random Forest Regression* (RFR) dan *Multilayer Perceptron* (MLP). Pengujian dilakukan dengan dua pemodelan data yaitu *partitioning* yang divalidasi dengan *Cross Validation* dan pemodelan data dengan *Cross Validation* tanpa *partitioning*. Dalam penelitian ini, model *prediction* dengan LR mampu menghasilkan nilai prediksi *error* yang cukup rendah dengan *score* RMSE sebesar 0.010 paling rendah dan RMSE tertinggi sebesar 0.012, MAPE terendah sebesar 1,2% dan tertinggi sebesar 1,9%, MAE terendah sebesar 0.006 dan tertinggi sebesar 0.009, serta nilai  $R^2$  tertinggi sebesar 99,8% dan terendah sebesar 99,6%. Hal ini dapat disimpulkan bahwa pada penelitian ini, model prediction *Linear Regression* mampu memprediksi data historis harga saham lebih baik dibandingkan dengan model RFR dan MLP.

Kata Kunci: Prediksi Harga Saham, *Linear Regression*, *Random Forest Regression*, *Multilayer Perceptron*.

## ABSTRACT

Name : Evita Fitri  
NIM : 14002313  
Study of Program : Ilmu Komputer  
Levels : Strata Dua (S2)  
Concentration : *Data Mining*  
Title : “Analisa Perbandingan Model *Prediction* Dalam Prediksi Harga Saham Pada Tiga Sektor Industri Di Indonesia”

*Stock are one type of long-term investment that is quite in demand by the public because this investment brings quite a large profit for its investors. However, in relation to this, stock price movements in general tend to be non-linear and non-stationary, this is because stock prices can be influenced by several factors whose results can change the pattern of stock price values either up or down, so this can make it difficult to stock prices prediction. In this study, a comparative analysis of prediction models was carried out in predicting stock prices using a technical approach based on past data, while the data used were historical stock prices by taking data samples from three issuers from the Indonesian capital market. There are three methods that were tested in this study, including Linear Regression (LR), Random Forest Regression (RFR) and Multilayer Perceptron (MLP). The test was carried out with two data modeling, namely partitioning which was validated with Cross Validation and data modeling with Cross Validation without partitioning. In this study, the prediction model with LR is able to produce a fairly low error prediction value with the lowest RMSE score of 0.010 and the highest RMSE of 0.012, the lowest MAPE of 1.2% and the highest of 1.9%, the lowest MAE of 0.006 and the highest. of 0.009, and the highest R2 value was 99.8% and the lowest was 99.6%. It can be concluded that in this study, the Linear Regression prediction model is able to predict historical data on stock prices better than the RFR and MLP models.*

*Keywords: Stock Price Prediction, Linear Regression, Random Forest Regression, Multilayer Perceptron*

## DAFTAR ISI

	Halaman
HALAMAN SAMPUL TESIS .....	i
HALAMAN JUDUL TESIS .....	ii
SURAT PERNYATAAN ORISINALITAS DAN BEBAS PLAGIARISME .....	iii
PERSETUJUAN DAN PENGESAHAN .....	iv
LEMBAR BIMBINGAN TESIS .....	v
KATA PENGANTAR .....	vi
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIK .....	viii
ABSTRAK .....	ix
ABSTRACT .....	x
DAFTAR ISI .....	xi
DAFTAR TABEL .....	xiii
DAFTAR GAMBAR .....	xiv
DAFTAR LAMPIRAN .....	xviii
<b>BAB I PENDAHULUAN</b> .....	1
1.1. Latar Belakang Penulisan .....	1
1.2. Identifikasi Masalah.....	4
1.3. Tujuan Penelitian .....	5
1.4. Ruang Lingkup Penelitian.....	5
1.5. Sistematika Penulisan .....	6
<b>BAB II LANDASAN TEORI</b> .....	8
2.1. Tinjauan Pustaka .....	8
2.2. Tinjauan Studi .....	31
2.3. Tinjauan Organisasi/Objek Penelitian .....	39
2.4. Kerangka Pemikiran .....	40
<b>BAB III METODOLOGI PENELITIAN</b> .....	43
3.1. Analisis Kebutuhan.....	43
3.2. Metode Penelitian .....	43
3.3. Sumber Data Penelitian.....	47
3.4. Pengolahan Data Awal.....	48
3.5. Metode Yang Diusulkan .....	48
3.6. Eksperimen Dan Hasil Pengujian .....	49
3.7. Evaluasi Dan Hasil Validasi .....	49
<b>BAB IV HASIL PENELITIAN DAN PEMBAHASAN</b> .....	50
4.1. <i>Business Understanding</i> .....	50
4.2. <i>Data Understanding</i> .....	51
4.2.1. Pengumpulan Data Awal ( <i>Collect The Initial Data</i> ) .....	51
4.2.2. Mendeskripsikan Data ( <i>Describe The Data</i> ) .....	55

4.3. Data <i>Preparation</i> .....	56
4.3.1. Data <i>Cleaning</i> .....	57
4.3.2. Data <i>Transformation</i> .....	60
4.3.3. Pembagian Data ( <i>Partitioning</i> ).....	62
4.4. <i>Modelling</i> .....	64
4.4.1. <i>Design Modelling</i> .....	64
4.4.2. Desain Model <i>Partitioning</i> Dengan <i>Cross Validation</i> .....	66
4.4.3. Desain Model <i>Cross Validation</i> (Tanpa <i>Partitioning</i> ).....	87
4.5. <i>Evaluation</i> .....	101
4.5.1. Hasil Evaluasi Menggunakan Model Prediksi <i>Linear</i> <i>Regression</i> .....	102
4.5.2. Hasil Evaluasi Menggunakan Model Prediksi RFR .....	104
4.5.3. Hasil Evaluasi Menggunakan Model Prediksi MLP .....	107
4.5.4. Rangkuman Hasil Evaluasi .....	110
<b>BAB V PENUTUP</b> .....	113
5.1. Kesimpulan .....	113
5.2. Saran .....	114
<b>DAFTAR PUSTAKA</b> .....	115
<b>LAMPIRAN</b> .....	121
<b>DAFTAR RIWAYAT HIDUP</b> .....	157

## DAFTAR TABEL

	Halaman
1. Tabel 1.1 <i>Source Dataset</i> .....	6
2. Tabel 2.1 Rangkuman Perbandingan Jurnal Referensi .....	35
3. Tabel 3.1 Spesifikasi Kebutuhan .....	43
4. Tabel 4.1 Daftar Historis Harga Saham PT. Indofood, Tbk .....	51
5. Tabel 4.2 Daftar Historis Harga Saham PT. Kimia Farma, Tbk.....	53
6. Tabel 4.3 Daftar Historis Harga Saham PT. Telkom Indonesia, Tbk	54
7. Tabel 4.4 Deskripsi Data .....	55
8. Tabel 4.5 Jumlah Pembagian Data ( <i>Partitioning</i> ) .....	62
9. Tabel 4.6 Hasil Validasi Data <i>Training</i> Dengan <i>Cross Validation</i> Model LR .....	67
10. Tabel 4.7 <i>Coefficients and Statistic</i> Dari Model <i>Learner</i> LR (INDF)	68
11. Tabel 4.8 <i>Coefficients and Statistic</i> Dari Model <i>Learner</i> LR (KAEF)	69
12. Tabel 4.9 <i>Coefficients and Statistic</i> Dari Model <i>Learner</i> LR (TLKM)	69
13. Tabel 4.10 Daftar Percobaan Pada Model Prediksi MLP .....	83
14. Tabel 4.11 <i>Output</i> Proses <i>Cross Validation</i> Dengan <i>X-Partitioner</i> ...	88
15. Tabel 4.12 <i>Coefficients and Statistic</i> Model <i>Learner</i> LR Dengan Validasi (INDF) .....	88
16. Tabel 4.13 <i>Coefficients and Statistic</i> Model <i>Learner</i> LR Dengan Validasi (KAEF) .....	89
17. Tabel 4.14 <i>Coefficients and Statistic</i> Model <i>Learner</i> LR Dengan Validasi (TLKM) .....	89
19. Tabel 4.15 <i>Score Error</i> Dengan Metode LR Validasi <i>Partitioning</i> ...	102
20. Tabel 4.16 <i>Score Error</i> Dengan Metode LR Dengan <i>Cross Validation</i>	103
21. Tabel 4.17 <i>Score Error</i> Dengan Metode RFR Validasi <i>Partitioning</i> ...	104
22. Tabel 4.18 <i>Score Error</i> Dengan Metode RFR Validasi <i>Cross</i> <i>Validation</i> .....	106
23. Tabel 4.19 <i>Score Error</i> Dengan Metode MLP Validasi <i>Cross</i> <i>Validation</i> .....	108
24. Tabel 4.20 Perbandingan RMSE, MAPE, MAE dan $R^2$ .....	111

## DAFTAR GAMBAR

	Halaman
1. Gambar 2.1 Bidang Ilmu <i>Data mining</i> .....	13
2. Gambar 2.2 Tahapan CIRSP-DM .....	15
3. Gambar 2.3 Struktur Random Forest .....	20
4. Gambar 2.4 Arsitektur Lapisan Tunggal.....	21
5. Gambar 2.5 Arsitektur <i>Multilayer Network</i> .....	21
6. Gambar 2.6 Arsitektur Lapisan Kompetitif .....	22
7. Gambar 2.7 Arsitektur <i>Multilayer Perceptron</i> .....	23
8. Gambar 2.8 Diagram Model MLP Yang Diusulkan .....	26
9. Gambar 2.9 Kriteria Nilai MAPE .....	29
10. Gambar 2.10 Tampilan Halaman Utama KNIME .....	30
11. Gambar 2.9 Kerangka Pemikiran.....	41
12. Gambar 3.1 Metode Penelitian.....	44
13. Gambar 3.2 Detail Sumber Dataset.....	47
14. Gambar 4.1 Proses <i>Preprocessing</i> Dataset .....	56
15. Gambar 4.2 Konfigurasi <i>CSV Reader</i> .....	57
16. Gambar 4.3 Hasil Proses <i>CSV Reader</i> (Dataset INDF).....	58
17. Gambar 4.4 Konfigurasi <i>Column Filter</i> .....	58
18. Gambar 4.5 Hasil Proses <i>Column Filter</i> (Dataset INDF) .....	59
19. Gambar 4.6 Konfigurasi <i>Missing Value</i> .....	59
20. Gambar 4.7 Hasil Proses Pada Node <i>Missing Value</i> (Dataset INDF)	60
21. Gambar 4.8 Konfigurasi Normalisasi Data.....	61
22. Gambar 4.9 Hasil Proses Normalisasi (Dataset INDF).....	61
23. Gambar 4.10 Konfigurasi Pembagian Data ( <i>Partitioning</i> ) .....	62
24. Gambar 4.11 Hasil <i>First Partitioning</i> ( <i>Training</i> INDF) .....	63
25. Gambar 4.12 Hasil <i>Second Partitioning</i> ( <i>Testing</i> INDF) .....	63
26. Gambar 4.13 Model Pengolahan Data Menggunakan <i>Linear Regression</i> (INDF) .....	64
27. Gambar 4.14 Model Pengolahan Data Menggunakan RFR (INDF)..	64
28. Gambar 4.15 Model Pengolahan Data Menggunakan MLP (INDF) .	65
29. Gambar 4.16 Tahap Pembacaan Data dan <i>Preprocessing</i> Pada <i>Design Modelling</i> .....	65
30. Gambar 4.17 <i>Partitioning</i> disertai <i>Cross Validation</i> LR (INDF, KAEF, TLKM) .....	66
31. Gambar 4.18 Konfigurasi <i>Cross Validation</i> .....	67
32. Gambar 4.19 Konfigurasi <i>Linear Regression Learner</i> .....	68
33. Gambar 4.20 Konfigurasi Node <i>Predictor Regression</i> .....	70
34. Gambar 4.21 Hasil <i>Predicted Data</i> pada Node <i>Predictor Regression</i> (INDF).....	70
35. Gambar 4.22 Penyelesaian Proses <i>Cross Validation</i> Menggunakan <i>X-Aggregator</i> .....	71
36. Gambar 4.23 <i>Prediction Table</i> Dari Proses <i>X-Aggregator</i> LR (INDF)	72
37. Gambar 4.24 <i>Error Rate 10 K-Fold Cross Validation</i> Model LR(INDF)	72
38. Gambar 4.25 <i>Error Rate 10 K-Fold Cross Validation</i> Model LR	

(KAEF).....	73
39. Gambar 4.26 <i>Error Rate 10 K-Fold Cross Validation</i> Model LR (TLKM).....	73
40. Gambar 4.27 <i>Prediksi Data Testing</i> Dari <i>Partitioning</i> Dengan Model LR .....	74
41. Gambar 4.28 <i>Predicted Data Testing</i> Hasil <i>Partitioning</i> Model LR (INDF).....	74
42. Gambar 4.29 <i>Partitioning</i> dengan <i>Cross Validation</i> RFR (INDF, KAEF, TLKM).....	75
43. Gambar 4.30 Konfigurasi <i>Random Forest Regression Learner</i> .....	76
44. Gambar 4.31 <i>Attribute Statistic</i> Hasil Pemodelan RFR <i>Partitioning</i> (INDF).....	76
45. Gambar 4.32 Hasil <i>Out of Bag Prediction</i> RFR <i>Partitioning</i> (INDF)	77
46. Gambar 4.33 Hasil <i>Prediksi Data</i> Dengan RFR <i>Predictor</i> (INDF) ...	78
47. Gambar 4.34 Proses Terakhir <i>Cross Validation (X-Aggregator)</i> RFR .....	79
48. Gambar 4.35 <i>Prediction Table</i> Dari Proses <i>X-Aggregator</i> RFR (INDF).....	79
49. Gambar 4.36 <i>Error Rate 10 K-Fold Cross Validation</i> Model RFR (INDF).....	80
50. Gambar 4.37 <i>Prediksi Data Testing</i> Hasil <i>Partitioning</i> Dengan Model RFR.....	80
51. Gambar 4.38 <i>Predicted Data Testing Partitioning</i> Dengan Model RFR (INDF) .....	81
52. Gambar 4.39 <i>Partitioning</i> Disertai <i>Validasi Cross Validation</i> Metode MLP (INDF, KAEF, TLKM).....	81
53. Gambar 4.40 Konfigurasi MLP <i>Learner</i> Data <i>Partitioning</i> Dengan <i>Validasi</i> (INDF) .....	82
54. Gambar 4.41 Hasil MLP <i>Predictor</i> Data <i>Testing Partitioning</i> Dengan <i>Cross Validation</i> (INDF).....	84
55. Gambar 4.42 Hasil MLP <i>Predictor</i> Data <i>Testing Partitioning</i> (INDF)	85
56. Gambar 4.43 Proses <i>Cross Validation</i> Menggunakan <i>X-Aggregator</i> MLP.....	85
57. Gambar 4.44 <i>Prediction Table</i> Dari Proses <i>X-Aggregator</i> MLP (INDF).....	86
58. Gambar 4.45 <i>Error Rate Cross Validation</i> Model MLP <i>Partitioning</i> (INDF).....	86
59. Gambar 4.46 <i>Desain Model</i> Dengan <i>Cross Validation</i> LR (INDF, KAEF, TLKM).....	87
60. Gambar 4.47 Hasil <i>Node Predictor Regression</i> Melalui <i>Cross Validation</i> (INDF).....	90
61. Gambar 4.48 Hasil <i>X-Aggregator</i> (Tanpa <i>Partitioning</i> ) LR (INDF)..	91
62. Gambar 4.49 <i>Error Rate Cross Validation</i> (Tanpa <i>Partitioning</i> )	

Model LR (INDF) .....	92
63. Gambar 4.50 Desain Model Dengan Cross Validation RFR (INDF, KAEF, TLKM).....	92
64. Gambar 4.51 <i>Attribute Statistic</i> Hasil <i>Modelling</i> RFR <i>Cross Validation</i> (INDF).....	93
65. Gambar 4.52 Output Out of Bag Prediction RFR Cross Validation (INDF).....	94
66. Gambar 4.53 <i>Prediction Output</i> Dari Prediksi RFR <i>Predictor Cross Validation</i> (INDF).....	95
67. Gambar 4.54 <i>Prediction Table</i> Hasil Node X-Aggregator RFR <i>Cross Validation</i> (INDF).....	96
68. Gambar 4.55 <i>Error Rates Cross Validation</i> Model RFR (Tanpa <i>Partitioning</i> ) (INDF).....	96
69. Gambar 4.56 Desain Model Dengan <i>Cross Validation</i> MLP (INDF, KAEF, TLKM).....	97
70. Gambar 4.57 Konfigurasi MLP <i>Learner</i> Data Hasil <i>Cross Validation</i>	98
71. Gambar 4.58 Hasil Node MLP <i>Predictor</i> Data <i>Testing Cross Validation</i> (INDF).....	99
72. Gambar 4.59 Proses X-Aggregator Tanpa <i>Partitioning</i> Model MLP..	99
73. Gambar 4.60 <i>Prediction Table</i> Hasil X-Aggregator MLP Model <i>Cross Validation</i> (INDF).....	100
74. Gambar 4.61 <i>Error Rates X-Aggregator</i> Model MLP (INDF).....	100
75. Gambar 4.62 Konfigurasi <i>Numeric Score</i> .....	101
76. Gambar 4.63 <i>Line Plot</i> Prediksi Data <i>Training</i> Metode LR <i>Partitioning</i> (INDF) .....	102
77. Gambar 4.64 <i>Line Plot</i> Prediksi Data <i>Testing</i> Metode LR <i>Partitioning</i> (INDF) .....	103
78. Gambar 4.65 <i>Line Plot</i> Prediksi Data <i>Testing</i> Metode LR <i>Cross Validation</i> (INDF).....	104
79. Gambar 4.66 <i>Line Plot</i> Prediksi Data <i>Training</i> Metode RFR <i>Partitioning</i> (INDF).....	105
80. Gambar 4.67 <i>Line Plot</i> Prediksi Data <i>Testing</i> Metode RFR <i>Partitioning</i> (INDF) .....	105
81. Gambar 4.68 <i>Line Plot</i> Prediksi Data <i>Testing</i> Metode RFR <i>Cross Validation</i> (INDF).....	106
82. Gambar 4.69 <i>Score Error</i> Dengan Metode MLP Validasi <i>Partitioning</i>	107
83. Gambar 4.70 <i>Line Plot</i> Prediksi Data <i>Training</i> Metode <i>Partitioning</i> MLP Dengan <i>Hidden Layer</i> 5 (INDF).....	108
84. Gambar 4.71 <i>Line Plot</i> Prediksi Data <i>Testing</i> Metode <i>Partitioning</i> MLP Dengan <i>Hidden Layer</i> 5 (INDF).....	108
85. Gambar 4.72 <i>Line Plot</i> Prediksi Data <i>Testing Cross Validation</i> (Tanpa <i>Partitioning</i> ) MLP Dengan 7 <i>Hidden Layer</i> (KAEF).....	109
86. Gambar 4.73 Perbandingan RMSE, MAPE, MAE dan R <sup>2</sup> .....	110

## DAFTAR LAMPIRAN

	Halaman
Lampiran 1. Hasil Proses <i>CSV Reader</i> (Dataset KAEF, TLKM).....	121
Lampiran 2. Hasil Proses <i>Column Filter</i> (Dataset KAEF, TLKM) .....	122
Lampiran 3. Hasil Proses <i>Node Missing Value</i> (Dataset KAEF, TLKM)	123
Lampiran 4. Hasil Proses <i>Normalisasi</i> (Dataset KAEF, TLKM).....	124
Lampiran 5. Hasil <i>Partitioning (Training dan Testing KAEF)</i> .....	125
Lampiran 6. Hasil <i>Partitioning (Training Dan Testing TLKM)</i> .....	126
Lampiran 7. Model <i>Linear Regression</i> (KAEF, TLKM).....	127
Lampiran 8. Model <i>RFR</i> (KAEF, TLKM).....	128
Lampiran 9. Model <i>MLP</i> (KAEF, TLKM) .....	129
Lampiran 10. Hasil <i>Prediksi Node Predictor Regression</i> (KAEF,TLKM)	130
Lampiran 11. Hasil <i>Prediksi Proses X-Aggregator LR</i> (KAEF,TLKM) ..	131
Lampiran 12. <i>Data Testing Hasil Partitioning LR</i> (KAEF,TLKM) .....	132
Lampiran 13. Hasil <i>Attribute Statistic RFR Partitioning</i> (KAEF, TLKM) .....	133
Lampiran 14. <i>Output Out of Bag Prediction RFR Partitioning</i> (KAEF) .	133
Lampiran 15. Hasil <i>Output Tree Views RFR Partitioning</i> (INDF, KAEF TLKM) .....	134
Lampiran 16. Hasil <i>Prediksi RFR Predictor Partitioning</i> (KAEF, TLKM) .....	136
Lampiran 17. Hasil <i>Proses X-Aggregator RFR</i> (KAEF, TLKM) .....	137
Lampiran 18. Hasil <i>Error Rate Dari X-Aggregator RFR</i> (KAEF,TLKM)	138
Lampiran 19. Hasil <i>Data Testing Partitioning RFR</i> (KAEF, TLKM).....	139
Lampiran 20. Hasil <i>MLP Predictor Data Testing Partitioning Cross Validation</i> (KAEF,TLKM) .....	140
Lampiran 21. Hasil <i>MLP Predictor Data Testing Partitioning</i> (KAEF, TLKM) .....	141
Lampiran 22. Hasil <i>Dari Proses X-Aggregator MLP</i> (KAEF,TLKM) .....	142
Lampiran 23. Hasil <i>Error Rate X-Aggregator MLP Partitioning</i> (KAEF, TLKM) .....	143
Lampiran 24. Hasil <i>Node Predictor Regression Cross Validation</i> (KAEF, TLKM) .....	144
Lampiran 25. Hasil <i>Prediksi Node X-Aggregator (Tanpa Partitioning) LR</i> (KAEF,TLKM) .....	145
Lampiran 26. Hasil <i>Error Rate Cross Validation (Tanpa Partitioning) LR</i> (KAEF, TLKM) .....	146
Lampiran 27. Hasil <i>Attribute Statistic RFR Cross Validation</i> (KAEF, TLKM) .....	147

Lampiran 28. Hasil <i>Out of Bag Prediction</i> RFR <i>Cross Validation</i> (KAEF, TLKM) .....	147
Lampiran 29. Hasil <i>Tree Views</i> RFR <i>Cross Validation</i> (INDF, KAEF, TLKM) .....	149
Lampiran 30. Hasil Prediksi Node RFR <i>Predictor Cross Validation</i> (KAEF, TLKM) .....	151
Lampiran 31. Hasil Prediksi Node <i>X-Aggregator</i> RFR <i>Cross Validation</i> (KAEF, TLKM) .....	152
Lampiran 32. <i>Error Rate Cross Validation</i> Model RFR Tanpa <i>Partitioning</i> (KAEF, TLKM) .....	153
Lampiran 33. Hasil Node MLP <i>Predictor Data Testing Cross Validation</i> (KAEF, TLKM) .....	154
Lampiran 34. Hasil Prediksi Node <i>X-Aggregator</i> Model MLP <i>Cross Validation</i> (KAEF, TLKM) .....	155
Lampiran 35. <i>Error Rate Cross Validation</i> Model MLP Tanpa <i>Partitioning</i> (KAEF, TLKM).....	156

# **BAB I**

## **PENDAHULUAN**

### **1.1. Latar Belakang Penulisan**

Terkait kondisi keuangan pada masa mendatang, salah satu hal yang dapat mengamankan kondisi tersebut adalah dengan memiliki investasi jangka panjang. Pada pemanfaatannya, investasi dapat digunakan untuk memenuhi kebutuhan jangka panjang diantaranya persiapan pensiun, sebagai biaya pendidikan maupun pernikahan yang cukup tinggi di masa depan ataupun sebagai salah satu alternatif dari terhindarnya masa inflasi. Pada implementasinya seiring berjalannya waktu, investasi membuat dana tersebut semakin banyak tanpa kita sadari, bukan hanya itu melainkan sudah banyak layanan finansial berbasis teknologi canggih yang membantu penggunanya dalam berinvestasi.

Adapun beberapa jenis investasi untuk jangka panjang yang dapat diterapkan diantaranya investasi berupa tanah, emas, mesin maupun aset finansial seperti deposit, saham, properti atau obligasi. Dari berbagai jenis aset finansial yang diperjualbelikan di pasar modal salah satunya adalah saham. Adapun salah satu aset yang sangat penting bagi setiap perusahaan di dunia sendiri adalah investasi saham pada pasar modal. Hal tersebut disebabkan adanya dampak terhadap ekonomi suatu negara dari para investor di negara tersebut.

Saham merupakan surat berharga yang dikeluarkan oleh sebuah perusahaan yang berbentuk Perseroan Terbatas (PT) atau yang biasa disebut emiten. Pada implementasinya, dengan memiliki saham maka menyatakan bahwa pemilik saham tersebut merupakan juga pemilik sebagian dari perusahaan tersebut [1]. Adapun banyak emiten atau perusahaan yang bergerak di pasar modal, para investor pun bebas dalam menentukan di perusahaan mana modal sahamnya akan ditanamkan, terdapat beberapa sektor industri pada pasar modal Indonesia diantaranya sektor pertanian, pertambangan, industri dasar dan kimia, aneka industri, industri barang konsumsi, properti dan konstruksi bangunan, infrastruktur dan transportasi, finansial dan perdagangan, jasa dan investasi.

Berbicara mengenai saham, investasi saham menjadi pilihan investasi yang cukup banyak diminati oleh kalangan dewasa, hal ini dikarenakan dalam berinvestasi saham dapat memberikan keuntungan yang cukup besar dan juga cepat bagi para investornya. Berkaitan dengan itu, terdapat pula beberapa faktor yang dapat mempengaruhi pola naik atau turunnya harga saham serta ketidakpastian sehingga dapat menyebabkan kerugian bagi para para investornya, diantaranya adalah terdiri dari faktor internal serta faktor eksternal, hal ini dikarenakan sifat harga saham yang fluktuatif. Hal ini dibuktikan dalam penelitian [2] yang membahas prediksi harga saham pada PT. Garuda Indonesia, Tbk pada masa pandemi Covid-19. Pada kasus tersebut dinyatakan bahwa pada harga saham PT. Garuda Indonesia cukup mengalami penurunan drastis pada saat pandemi Covid-19. Namun tidak semua emiten cenderung mengalami penurunan, sebagian emiten atau perusahaan pun ada yang mengalami perubahan berupa peningkatan penghasilan, sebagai contoh yakni industri obat-obatan yang cenderung lebih dibutuhkan daripada sebelumnya terkait faktor yang mempengaruhi harga saham seperti adanya sebuah wabah.

Terkait hal tersebut, pada penelitian ini penulis ingin membahas mengenai analisis perbandingan model prediksi harga saham pada beberapa sektor industri di Indonesia berdasarkan data-data terdahulu, adapun data-data yang akan digunakan oleh penulis merupakan data-data historis harga saham harian yang bersifat data *time series* dengan periode tahun 2015 sampai dengan 2020, sehingga dari data-data tersebut dapat memberikan hasil yang cukup akurat pada prediksi harga saham kedepannya. Beberapa metode yang dapat digunakan untuk mendukung analisa prediksi sangatlah beragam, adapun metode yang bisa digunakan dalam prediksi harga saham yang bersifat data *time series* sendiri terdapat beberapa. Terkait hal itu, pada penelitian ini akan dilakukan peramalan data historis harga saham bersifat *time series* dalam prediksi harga saham dengan menggunakan model prediksi *Linear Regression (LR)*, *Random Forest Regression (RFR)* dan *Multilayer Perceptron (MLP)*.

Pada penelitian [3] yang membahas mengenai prediksi harga saham berdasarkan pendekatan *machine learning* dengan menggunakan metode *Logistic*

*Regression* dan *Support Vector Machine*, kedua metode tersebut dinyatakan efektif untuk digunakan dalam memprediksi harga saham yang mana pada penelitian tersebut diambil sampel pada pasar modal China, pada penelitian ini pun ditemukan bahwa model SVM sudah optimal dengan nilai 17,13% serta nilai maksimal *drawdown* 0.32 sehingga model ini dapat membantu dalam strategi memprediksi pasar modal di China. Selanjutnya penelitian peramalan dalam prediksi harga saham pun dilakukan oleh [4] dengan menerapkan algoritma *Artificial Neural Network* dan menggunakan empat atribut yaitu nilai *open*, *high*, *low* sebagai prediktor dan *close* sebagai *class* yang berfokus pada penentuan *Root Mean Squared Error* (RMSE). Adapun hasil dengan mengoptimalkan nilai-nilai parameter dan pemilihan ukuran *hidden layer* memberikan hasil yang cukup baik yakni dengan nilai akurasi dan RMSE masing-masing sebesar 0.266 +/- 0.000. Dan pada penelitian [5] melakukan analisa perbandingan model algoritma *data mining* pada prediksi harga saham GGRM menggunakan model *Linear Regression*, *Neural Network*, *SVM*, *Gaussian Process*, dan *Polynomial Regression* dengan lima atribut yaitu *date*, *open*, *high*, *low* dan *close* dengan hasil penelitian menyimpulkan bahwa data harga saham GGRM dapat diprediksi dengan menggunakan model algoritma *Neural Network* dengan hasil akurasi dan nilai RMSE sebesar 612.474 +/- 89.402 (mikro: 618.916 +/- 0.000) paling kecil dibandingkan dengan model algoritma lainnya, sehingga dengan prediksi hal ini dapat membantu dalam memprediksi harga saham GGRM di pasar modal.

*Multilayer Perceptron* atau dikenal dengan MLP merupakan salah satu metode dari *Neural Network* yang disebutkan cocok digunakan pada penyelesaian masalah yang bersifat tidak linear dan *non deterministik*, serta pada metode ini telah diterapkan dengan sukses pada penyelesaian masalah-masalah yang cukup sulit dengan uji coba atau melatih data menggunakan algoritma propagasi balik dari kesalahan atau *error backpropagation* [6].

*Random Forest Regression* merupakan metode yang digunakan pada pemodelan regresi, adapun pada metode ini banyak digunakan dalam pemodelan prediktif, model ini pun dapat mengatasi masalah prediktor yang berkorelasi tinggi, hal ini karena dibuat berdasarkan *regression tree* dan juga seleksi teknik

peubah, model ini juga menunjukkan bahwa *Random Forest Regression* baik digunakan dalam hal berhubungan dengan nonlinear beserta respon [7]. Adapun regresi adalah ukuran statistik yang digunakan untuk menentukan kekuatan hubungan antara variabel tak bebas dengan variabel bebas, *Linear Regression* memodelkan hubungan antara variabel skalar dan satu atau lebih variabel penjelas. Metode *Linear Regression* sendiri tersusun atas dasar pola hubungan data yang relevan di masa lalu [8].

Berdasarkan hal-hal di atas, maka peneliti bertujuan melakukan peramalan dalam prediksi harga saham pada tiga sektor industri di Indonesia dengan tujuan berfokus pada pencapaian minimnya tingkat *error* diantaranya *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE) dan tingkat nilai  $R^2$  menggunakan pendekatan *Machine Learning* dengan model algoritma *Linear Regression*, *Random Forest Regression* dan *Multilayer Perceptron*.

## **1.2. Identifikasi Masalah**

Adapun berdasarkan latar belakang yang telah dipaparkan di atas, identifikasi masalah yang penulis buat pada penelitian ini adalah diantaranya dilakukannya analisis dalam memprediksi harga saham pada tiga sektor industri di Indonesia berdasarkan data-data harga saham terdahulu, adapun data-data harga saham yang dimiliki merupakan data harian bersifat *time series* dengan *range* waktu 2015 sampai dengan 2020. Dalam peramalan atau memprediksi harga saham, salah satu metode yang banyak digunakan pada analisa prediksi harga saham diantaranya adalah pendekatan regresi. Adapun terdapat beberapa algoritma yang dapat digunakan dalam menganalisa data. Berdasarkan hal tersebut, dalam penelitian ini akan dilakukannya uji coba dalam prediksi harga saham pada tiga sektor industri di Indonesia dengan menggunakan model algoritma *Linear Regression*, *Random Forest Regression* dan *Multilayer Perceptron*. Adapun secara rinci identifikasi masalah yang akan dibahas pada penelitian ini adalah:

1. Apakah model algoritma *Linear Regression*, *Random Forest Regression* dan *Multilayer Perceptron* dapat digunakan dalam memprediksi harga saham pada data historis harga saham tiga sektor industri yang didapat dari pasar modal Indonesia?
2. Selanjutnya dari ketiga model prediksi tersebut, model prediksi apakah yang menghasilkan tingkat *error* terendah dalam memprediksi historis data harga saham yang telah diuji coba?
3. Apakah dengan menerapkan model prediksi tersebut dapat membantu kalangan baik dari segi investor atau perusahaan dalam menganalisa model apa yang lebih baik yang dapat digunakan sebagai langkah dalam memprediksi harga saham dan yang nantinya akan dapat dijadikan sumber informasi?

### **1.3. Tujuan Penelitian**

Adapun tujuan pada penelitian yang penulis lakukan ini diantaranya adalah sebagai berikut:

1. Bertujuan membantu dan memudahkan para investor dalam menganalisa dan memprediksi historis harga saham pada pasar modal Indonesia yang nantinya untuk menanamkan modal sahamnya di perusahaan yang lebih tepat menggunakan model prediksi *Linear Regression*, *Random Forest Regression* dan *Multilayer Perceptron*.
2. Menganalisa tingkat kesalahan atau *error* diantaranya *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE) dan tingkat nilai  $R^2$  dalam memprediksi harga saham pada tiga sektor industri di pasar modal Indonesia dengan menggunakan tiga perbandingan model algoritma tersebut.
3. Membandingkan model-model prediksi yang dapat dilakukan pada prediksi historis harga saham di tiga sektor industri di Indonesia.

Selanjutnya bertujuan sebagai salah satu syarat kelulusan program Strata Dua (S2) pada program studi Magister Komputer di Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Nusa Mandiri.

#### 1.4. Ruang Lingkup Penelitian

Pada penelitian ini, penulis membatasi ruang lingkup hanya pada pengolahan dataset menggunakan tiga model *machine learning* yakni model prediksi *Linear Regression*, *Random Forest Regression* dan *Multilayer Perceptron* sebagai perbandingan dalam mengevaluasi untuk menghasilkan nilai *error* terendah diantaranya *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE) dan tingkat nilai  $R^2$  pada uji coba dataset. Adapun dataset yang digunakan adalah data yang diambil dari pasar modal Indonesia IDX pada laman <https://finance.yahoo.com> yang terdiri dari historis harga pada tiga sektor industri serta dataset diambil pada periode 01 September 2015 sampai dengan 01 September 2020, pada periode tersebut dibagi menjadi data *training* dan data *testing*, adapun data ke tiga sektor tersebut dapat dilihat pada Tabel 1.1 *Source Dataset* :

Tabel 1.1 *Source Dataset*

No	Sektor Industri	Nama Perusahaan	Kode Saham	Source Dataset
1	Industri Barang Konsumsi	PT Indofood Sukses Makmur Tbk	INDF.JK	<a href="https://finance.yahoo.com/quote/INDF.JK/?p=INDF.JK">https://finance.yahoo.com/quote/INDF.JK/?p=INDF.JK</a>
2	Transportasi & Infrastruktur	PT Telekomunikasi Indonesia Tbk	TLKM.JK	<a href="https://finance.yahoo.com/quote/TLKM.JK/">https://finance.yahoo.com/quote/TLKM.JK/</a>
3	Industri Bahan Dasar & Kimia	PT Kimia Farma Tbk	KAEF.JK	<a href="https://finance.yahoo.com/quote/KAEF.JK/history?p=KAEF.JK">https://finance.yahoo.com/quote/KAEF.JK/history?p=KAEF.JK</a>

## **1.5 Sistematika Penulisan**

Sistematika penulisan pada tesis ini terdiri dari lima bab, dimana tiap bab terdiri dari beberapa sub bab yaitu:

### **BAB 1 Pendahuluan**

Bab ini menjelaskan latar belakang masalah mengenai prediksi harga saham serta perumusan masalah, batasan masalah, tujuan penelitian dan manfaat penelitian dan sistematika penulisan yang berkaitan dengan pembahasan *modelling prediction* pada prediksi harga saham, adapun data prediksi diambil dari historis harga saham pada tiga sektor industri di Indonesia yaitu sektor industri barang konsumsi, transportasi dan infrastruktur serta industri bahan dasar dan kimia, adapun data didapatkan dari laman *website yahoo finance*.

### **BAB 2 Kerangka Pemikiran**

Berisi landasan-landasan teori yang digunakan diantaranya adalah teori pasar modal, saham, harga saham, *data mining*, teori algoritma *Linear Regression*, *Random Forest Regression* dan *Multilayer Perceptron*, teori-teori tersebut untuk mendukung pelaksanaan penelitian dan menyertakan uraian sistematis dari penelitian- penelitian terkait.

### **BAB 3 Metode Penelitian**

Berisi metode penelitian yang digunakan pada penelitian dengan pembahasan menguji beberapa model algoritma dalam memprediksi harga saham yang dilakukan dalam penelitian, adapun secara garis besar terbagi menjadi tiga, yaitu *preprocessing* yang dilakukan setelah penarikan dataset sehingga data dapat diolah tanpa adanya *missing value* serta ditetapkannya pembobotan dengan normalisasi, selanjutnya *modeling* dengan menguji coba tiga model prediksi yakni LR, RFR dan MLP, dan terakhir evaluasi hasil penelitian yang menjadi tolak ukur pada penelitian ini yakni dalam melihat nilai terendah *errornya*.

### **BAB 4 Hasil dan Pembahasan**

Pada Bab ini berisi pemaparan hasil penelitian dari uji coba beberapa model algoritma dalam memprediksi historis harga saham yang digambarkan dalam bentuk uraian, tabel, grafik beserta pembahasannya.

## **BAB 5 Penutup**

Pada bab ini berisi tentang kesimpulan dari hasil penelitian yang telah dilakukan dalam menguji coba beberapa model algoritma pada prediksi harga saham dan saran dari penulis mengenai penelitian ini yang selanjutnya dapat dikembangkan.

## **BAB II**

### **LANDASAN TEORI**

#### **2.1. Tinjauan Pustaka**

##### **2.1.1. Pasar Modal**

Pasar modal atau sering disebut dengan *capital market* saat ini menjadi salah satu sumber pembiayaan yang cukup diminati serta dibutuhkan oleh beberapa pihak terkait, diantaranya adalah pelaku usaha yang membutuhkan sumber modal untuk kegiatannya. Tidak sekedar itu, pihak terkait lain pun seperti masyarakat yang berperan sebagai investor menjadikan pasar modal ini sebagai salah satu alternatif sumber pembiayaan terhadap kegiatan yang menghasilkan keuntungan bagi para investor tersebut. Adapun dalam buku Hukum Pasar Modal, mendefinisikan bahwa *capital markets are places where those who require additional funds seek out others who wish to invest their excess*. Definisi yang sama dikemukakan oleh [9] yang menyatakan bahwa pasar modal merupakan tempat bertemunya pihak-pihak yang memiliki kelebihan dalam hal kapasitas modal untuk para investor dengan pihak yang membutuhkan tambahan modal baik modal jangka panjang maupun jangka pendek.

Di Indonesia sendiri, pengertian pasar modal dapat dilihat pada Pasal 1 Angka 13 Undang-Undang Nomor 8 Tahun 1995 tentang pasar modal yang selanjutnya disebut UUPM yang menegaskan bahwa “Pasar modal adalah kegiatan yang bersangkutan dengan penawaran umum dan perdagangan Efek, perusahaan publik yang berkaitan dengan efek yang diterbitkannya, serta lembaga dan profesi yang berkaitan dengan perdagangan efek” Adapun dari pengertian di atas dapat disimpulkan bahwa pasar modal memiliki kegiatan yang sangat luas, kegiatan tersebut mencakup kegiatan penawaran umum, kegiatan perdagangan efek, kegiatan yang berkaitan dengan perusahaan publik yang pada pembahasan ini pun tertera ketetapannya pada Pasal 1 Angka 8 Undang-Undang Nomor 40 Tahun 2007 tentang Perseroan Terbatas, dan terakhir adalah kegiatan yang berkaitan dengan lembaga atau profesi berkaitan dengan efek [10].

### **2.1.2. Peranan Pasar Modal**

Pasar modal memiliki peran yang cukup strategis bila dilihat dari kebutuhan pembangunan nasional yakni sebagai salah satu sumber pembiayaan bagi dunia usaha dan wahana investasi bagi masyarakat. Pasar modal juga memiliki peran yang cukup penting sebagai sarana investasi jangka panjang dalam perekonomian, di Indonesia sendiri Otoritas Jasa Keuangan (OJK) menaungi pasar modal Indonesia, hal ini dijelaskan bahwa pada rangkaian kegiatan dalam pasar modal terdapat pihak-pihak atau lembaga-lembaga yang terlibat diantaranya perusahaan emiten, investor, bank kustodian, dan lembaga penunjang [11].

Adapun pasar modal memiliki peranan lain, berikut menurut [12] dalam bukunya *Semakin Dekat dengan Pasar Modal Indonesia* diantaranya adalah:

1. Pasar modal sebagai sarana penambah modal untuk mempelancar kegiatan usaha. Hal ini dilakukan dengan cara perusahaan menjual saham ke pasar modal yang nantinya saham-saham ini akan dibeli oleh masyarakat umum, perusahaan-perusahaan lain, lembaga atau pemerintah untuk memperoleh tambahan dana.
2. Pasar modal sebagai sarana pemerataan pendapatan, dimana setelah jangka waktu tertentu, saham-saham yang telah dibeli akan memberikan sebuah keuntungan atau dividen kepada pemegang saham atau pemilik saham, oleh sebab itu pasar modal dapat dikatakan sebagai sarana pemerataan pendapatan.
3. Pasar modal sebagai sarana peningkatan kapasitas produksi, hal ini terjadi karena adanya penambahan modal yang diperoleh dari pasar modal sehingga membuat kegiatan produktivitas perusahaan tersebut menjadi meningkat.
4. Pasar modal sebagai sarana penciptaan tenaga kerja, hal ini terjadi karena adanya dorongan yang muncul dari pasar modal yang kemudian terjadinya perkembangan pada dunia industri lain yang selanjutnya menimbulkan terciptanya lapangan kerja baru.
5. Pasar modal sebagai sarana peningkatan pendapatan suatu negara, dapat dinyatakan membantu peningkatan pendapatan negara dikarenakan setiap dividen yang diberikan kepada pemilik saham akan dikenakan pajak oleh

pemerintah sehingga adanya pemasukan melalui pajak tersebut dan menjadikan pendapatan negara bertambah atau meningkat.

6. Pasar modal sebagai sarana indikator perekonomian negara, bila dilihat dari meningkatnya aktivitas dan tingkat volume penjualan atau pembelian yang dilakukan di pasar modal, hal ini memberikan indikasi bahwa aktivitas bisnis berbagai perusahaan dapat dikatakan berjalan dengan baik, hal ini berlaku sebaliknya.

### **2.1.3. Saham**

Pada pasar modal, transaksi jual beli saham selalu dilakukan oleh banyak pihak-pihak terkait. Saham adalah sebuah tanda bukti penyertaan kepemilikan modal atau dana pada suatu perusahaan yang bersangkutan. Saham juga merupakan bentuk kertas yang tercantum dengan jelas berisi beberapa informasi diantaranya nilai nominal, nama perusahaan dan diikuti dengan hak dan kewajiban yang dijelaskan kepada setiap pemegangnya atau pemilik saham, selain itu saham pun menjadi salah satu aset yang dimiliki oleh pemegang saham yang nantinya siap untuk dijual [13].

Bagi investor, investasi saham merupakan salah satu investasi yang paling banyak diminati, hal ini dikarenakan dapat memberikan tingkat *return* yang lebih tinggi dibandingkan dengan obligasi dan reksadana. Pendapatan *return* ini yang nantinya diharapkan bagi para investor, adapun pendapatan *return* ini terdiri dari dividen dan *capital gain* [14]. Letak perbedaan antara pendapatan dividen dengan *capital gain* adalah dimana investor yang melakukan investasi dalam jangka panjang akan mendapat dividen sehingga dalam hal ini investor lebih memfokuskan pada laporan keuangan perusahaan dibandingkan perubahan fluktuasi pada pasar modal, sedangkan bagi investor yang melakukan investasi dalam jangka pendek maka pendapatan yang diharapkan berupa *capital gain*.

### **2.1.4. Manfaat Saham**

Pada pemanfaatannya, beberapa diantara manfaat saham dapat dilihat dari dua aspek, yakni aspek emiten dan aspek pemodal. Menurut [15] manfaat dari kedua aspek tersebut yaitu :

1. Manfaat bagi emiten, saham digunakan sebagai alat penyandang dana. Adapun dana tersebut diperlukan olehnya atau emiten tersebut guna membantu dalam pelaksanaan pembangunan sarana usaha, pelebaran sayap perusahaan atau segala sesuatu yang berkaitan dengan kepentingan perusahaan atau pemerintah.
2. Manfaat saham bagi pemodal atau investor yakni diperuntukan untuk pemanfaatan dana dengan menanamkan modal atau dana yang dimiliki sebagai alternatif investasi yang nantinya akan menghasilkan keuntungan.

Adapun manfaat saham oleh [16] diantaranya adalah dividen, *capital gain* dan manfaat *non-material*. Dividen sendiri yaitu bagian dari keuntungan perusahaan yang nantinya dibagikan kepada pemilik saham, lalu *capital gain* yakni keuntungan yang diperoleh dari selisih jual dengan harga belinya dan terakhir yaitu manfaat *non-material* dimana manfaat ini menjadikan timbulnya kekuasaan atau memperoleh hak suara dalam menentukan kedepannya perusahaan tersebut berjalan.

#### **2.1.5. Harga Saham**

Menurut para ahli terkait harga saham diantaranya oleh [17] yang mendefinisikan harga saham adalah menunjukkan harga pada pasar riil dan merupakan harga yang paling mudah ditentukan karena merupakan harga suatu saham pada pasar modal yang sedang berjalan atau berlangsung, harga saham pun dapat dikatakan harga penutupnya jika pasar tutup. Adapun harga saham sebagai nilai saham, yang mana dijelaskan bahwa nilai tersebut terdiri dari tiga nilai harga saham sebagai berikut [18] :

1. Nilai buku (*Book Value*)

Nilai buku adalah nilai saham sesuai dengan nilai pembukuan emiten. Pada nilai buku ini dipengaruhi oleh nilai-nilai lain terdiri dari nilai nominal (*par*

*value*), nilai agio saham (*additional paid capital or in excess of par value*), nilai modal yang disetor (*paid in capital*) dan laba ditahan (*retained earnings*).

## 2. Nilai Pasar (*Market Value*)

Nilai pasar sendiri adalah harga saham yang terjadi di pasar bursa pada saat tertentu sesuai dengan yang ditentukan oleh pelaku pasar, adapun penentuan nilai pasar ini sesuai dengan permintaan dan penawaran saham yang terjadi di pasar modal tersebut.

## 3. Nilai Intrinsik (*Intrinsic Value*)

Nilai intrinsik atau nilai fundamental merupakan nilai *real* atau nilai yang sebenarnya suatu harga saham yang diberikan oleh perusahaan.

### **2.1.6. Faktor-Faktor Mempengaruhi Harga Saham**

Pada penelitian [19] menyatakan terdapat faktor-faktor yang mempengaruhi harga saham (per lembar saham) yakni adanya hukum permintaan dan penawaran, berita umum, indeks harga saham tingkat suku bunga, dana asing di bursa efek, mata uang dolar Amerika serta adanya kondisi fundamental.

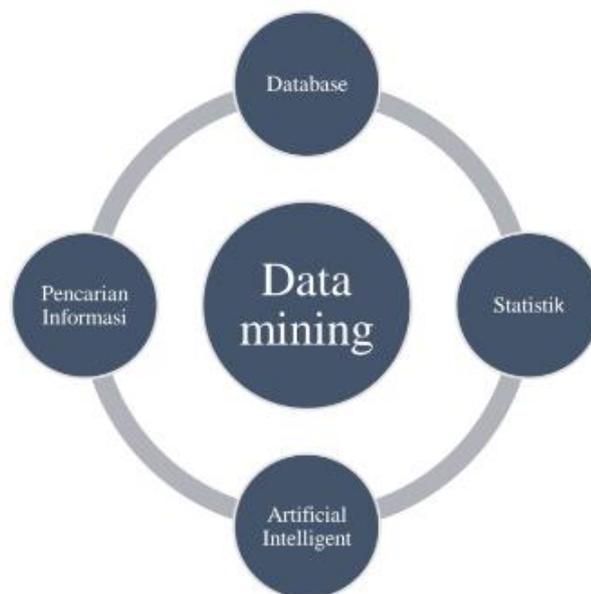
Adapun merujuk pada penelitian [20] menjelaskan bahwa harga saham bergerak secara fluktuatif baik harga saham agregat maupun satuan, bergerak secara fluaktif ini sebagai bentuk respon dari informasi yang tersebar pada pasar modal. Faktor-faktor yang mempengaruhi pergerakan harga saham diantaranya terbagi menjadi dua kelompok bagian yaitu faktor internal dan faktor eksternal perusahaan. Pada faktor internal yang cukup signifikan dalam mempengaruhi pergerakan harga saham diantaranya adalah proporsi utang, dividen serta profitabilitas perusahaan, sedangkan pada faktor eksternal diantaranya informasi publik, kondisi politis suatu negara, kebijakan makro ekonomi dan kondisi sosial.

### **2.1.7. Data mining**

Terdapat beberapa pengertian *data mining* menurut para ahli yang terdapat pada [21] diantaranya yaitu bahwa *data mining* disebutkan sebagai suatu istilah yang digunakan untuk menemukan hal baru berupa pengetahuan yang

tersembunyi dari suatu *database*. *Data mining* juga disebut sebagai proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan dan juga *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan yang terkait dengan berbagai *big databases*.

Adapun definisi *data mining* oleh [22] disebutkan bahwa *data mining* merupakan serangkaian aktivitas yang berkaitan dengan pengumpulan data, pemakaian data historis untuk mendapatkan pengetahuan baru serta informasi dan keteraturan, pola atau hubungan dalam data dengan jumlah yang besar, pada *output* yang dihasilkan dari *data mining* dapat digunakan sebagai alternatif dalam pengambilan keputusan serta untuk memperbaiki pengambilan keputusan pada masa yang akan datang. Adapun dijelaskan bahwa *data mining* bukanlah ilmu yang berdiri sendiri melainkan berkaitan dengan bidang ilmu lainnya seperti *database*, statistik, pencarian informasi dan *artificial intelligent* seperti pada Gambar 2.1 :



**Gambar 2.1** Bidang Ilmu *Data mining*

Kesatuan tersebut digambarkan sebagai keterkaitan *data mining* dengan bidang ilmu lainnya seperti *database* sebagai salah satu sumber kumpulan data yang digunakan dalam *data mining*, statistik dipergunakan sebagai sumber pengambilan keputusan adapun dalam statistik dibutuhkannya data dimulai dari

pengumpulan data, pengambilan data sampel data sampai dengan menganalisa dan mempresentasikan *outputnya* menggunakan teknik statistik, kemudian pencarian informasi yang merupakan salah satu kegiatan dalam proses *data mining* dan selanjutnya terdapat proses pembelajaran yang dilakukan oleh komputer dengan data *training* maupun *testing* sebagai pengimplementasian pada bidang ilmu *Artificial Intelligence*

### **2.1.8. Fungsi *Data mining***

Pada pengimplementasian *data mining* terdapat sebanyak 6 fungsi *data mining* menurut [23] diantaranya fungsi deskripsi (*description*), fungsi estimasi (*estimation*), fungsi prediksi (*prediction*), fungsi klasifikasi (*classification*), fungsi pengelompokan (*classifycation*), dan fungsi asosiasi (*association*). Adapun penjelasan dari masing-masing fungsi yaitu [24] :

#### 1. Fungsi Deskripsi (*Description*)

Pada fungsi ini, dikaitkan dengan peneliti ingin mencoba cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data, pada fungsi inilah didapat deskripsi dari pola dan kecenderungan yang terdapat pada data yang nantinya memberikan kemungkinan penjelasan suatu pola atau suatu kecenderungan.

#### 2. Fungsi Estimasi (*Estimation*)

Pada fungsi ini dikatakan mendekati sama dengan klasifikasi, namun pada variabel targetnya bersifat numerik dan model pada fungsi ini dibangun menggunakan *record* lengkap yang menyediakan nilai dari variabel targetnya bersifat sebagai nilai prediksi.

#### 3. Fungsi Prediksi (*Prediction*)

Pada fungsi prediksi dikatakan hampir menyamai dengan klasifikasi dan juga estimasi, hanya saja pada fungsi ini nilai prediksi akan ada pada masa mendatang sebagai contoh prediksi harga beras dalam tiga bulan mendatang.

#### 4. Fungsi Klasifikasi (*Classification*)

Pada fungsi klasifikasi terdapat target variabel berupa kategori sebagai contoh pada penggolongan pendapatan terdapat beberapa kategori diantaranya pendapatan tinggi, pendapatan sedang dan pendapatan dengan kategori rendah.

#### 5. Fungsi Pengelompokkan (*Classification*)

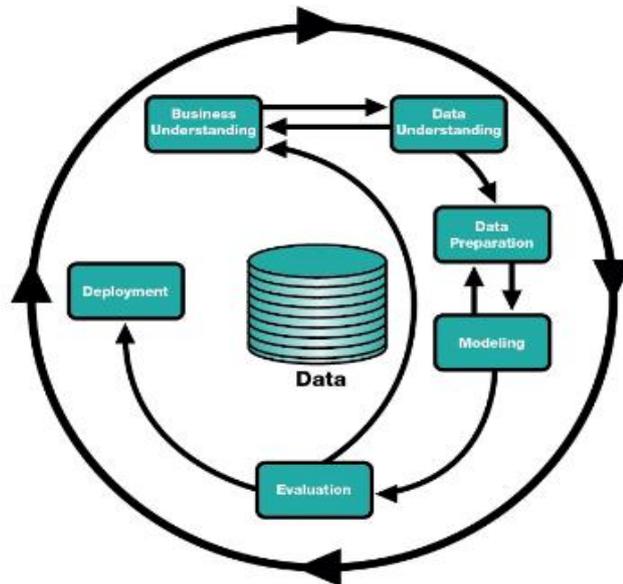
Fungsi ini seperti pengklasteran yang merupakan pengelompokkan *record*, pengamatan serta membentuk kelas objek–objek yang mempunyai kemiripan satu sama lain atau tidak memiliki kemiripan dengan *record-record* dalam klaster lain.

#### 6. Fungsi Asosiasi (*Association*)

Adapun fungsi asosiasi dalam *data mining* bertugas menemukan atribut yang muncul dalam satu waktu, umumnya dalam dunia bisnis disebut dengan analisis keranjang belanja, fungsi asosiasi pada analisis tersebut mencari kombinasi jenis barang yang akan terjual untuk kedepannya.

### **2.1.9. Tahapan *Data mining***

Pada tahapan yang dilakukan dalam proses *data mining* yang dikemukakan pada penelitian [25] yaitu diawali dengan seleksi data dari sumber data ke data target, setelah itu adanya proses *preprocessing* yang diperuntukkan untuk memperbaiki kualitas data, selanjutnya dilakukan transformasi data, *data mining* serta tahap interpretasi dan evaluasi yang akhirnya menghasilkan sebuah *output* berupa pengetahuan baru yang diharapkan nantinya akan memberikan kontribusi yang lebih baik. Salah satu model tahapan pendekatan pada *data mining* adalah CRISP-DM (*Cross-Industry Standard Process for Data mining*). Adapun secara detail dijelaskan seperti pada Gambar 2.2 :



**Gambar 2.2** Tahapan CIRSP-DM

Berikut keterangan atau penjelasan dari tahapan-tahapan dalam *data mining* pada Gambar 2.2 [26] sebagai berikut:

1. *Business Understanding*

Pada tahapan ini dilakukannya beberapa hal yaitu antara lain, menentukan tujuan bisnis, menilai situasi, dan menentukan tujuan *data mining*. Adapun pada fase pertama ini yaitu berfokus pada kegiatan dimana harus memahami objektif pelaksanaan sebuah *data mining* dan seluruh kebutuhan yang terkait yang nantinya akan ditinjau dari sudut pandang kebutuhan bisnis. Setelahnya informasi yang dihasilkan ini kemudian akan diproses menjadi definisi masalah dalam kegiatan penelitian *data mining* serta membuat rencana awal untuk mencapai tujuan yang telah diidentifikasi.

2. *Data Understanding*

Pada tahapan *data understanding*, ada beberapa hal yang dilakukan pada fase ini, dimulai dengan proses pengumpulan data awal, mendeskripsikan data dengan cara memahami data yang tersedia, mengeksplorasi data, memverifikasi masalah-masalah terkait kualitas data serta mendeteksi subset yang menarik untuk dijadikan hipotesis terkait informasi pada data yang tersembunyi.

3. *Data Preparation*

Pada tahapan data *preparation* ada beberapa hal yang dilakukan antara lain, deskripsi dataset, memilih data, membangun data, mengintegrasikan data dan membersihkan data. Lebih jelasnya pada fase ini mencakup seluruh aktivitas yang dibutuhkan untuk membangun dataset final yang akan digunakan dalam proses pemodelan. Adapun kegiatan yang dilakukan dalam fase ini biasanya dilakukan secara berulang kali tanpa urutan yang kaku. Kegiatan tersebut dapat diantaranya meliputi pembuatan tabel, pemilihan atribut, pembersihan data dan melakukan *transformation* data seperti normalisasi.

#### 4. *Modelling*

Pada tahapan *modelling*, fase ini dilakukannya pemilihan dan penerapan berbagai teknik pemodelan data dengan parameter yang sudah dikalibrasi dengan konfigurasi, hal ini bertujuan untuk membuat hasil pemodelan berada pada tingkat akurasi paling optimal dan akurat. Pada penerapan beberapa teknik pemodelan sendiri memiliki syarat-syarat terkait kondisi data yang digunakan yang perlu dipenuhi oleh masing-masing teknik pemodelan.

#### 5. *Evaluation*

Pada tahapan *evaluation* didefinisikan sebagai fase interpretasi terhadap hasil *data mining*. Pada fase ini dilakukannya evaluasi yang mendalam atas model-model yang telah ditentukan. Adapun evaluasi dilakukan untuk memastikan bahwa ada model yang telah memenuhi *business objective* yang telah ditentukan di awal. Pada akhir tahapan ini model akhir yang akan digunakan dalam *data mining* ditentukan berdasarkan hasil evaluasi tersebut seperti akurasi ataupun *error score*.

#### 6. *Deployment*

Adapun pada tahap terakhir ini dilakukannya penerapan model yang telah digunakan pada media yang dapat memungkinkan pengguna memahami cara menggunakan model yang telah dibuat. Pada bentuk kegiatannya didasarkan pada daftar kebutuhan yang telah ditentukan di fase awal. Bentuk kegiatan *deployment* sendiri bisa dilakukan dengan berbagai bentuk mulai dari sebatas laporan, hingga penerapan sistem untuk interaksi langsung antar pengguna.

#### **2.1.10. Data Time Series**

Data *time series* adalah susunan observasi berturut berdasarkan waktu, data *time series* sendiri merupakan serangkaian data pengamatan yang berasal dari satu sumber tetap yang terjadinya berdasarkan indeks waktu secara berurutan dengan interval waktu yang tetap. Pada data ini digunakan notasi untuk menyatakan nilai numerik dari pengamatan, dengan menunjukkan periode waktu terjadinya pengamatan. Adapun pada proses prediksi data *time series* tidak melibatkan variabel bebas lain selain indeks waktu itu sendiri. Data tersebut didasarkan pada urutan dari setiap titik-titik data yang memiliki jarak sama dalam waktu seperti mingguan, bulanan, kuartalan, dan lainnya [27].

#### **2.1.11. Prediksi**

Prediksi adalah suatu proses untuk memperkirakan atau meramalkan suatu variabel di masa yang akan datang. Dalam kasus prediksi biasanya data yang sering digunakan adalah data kuantitatif. Pada prediksi hasil yang didapatkan tidak harus berupa suatu jawaban yang pasti kejadiannya, melainkan dapat berusaha untuk mendapatkan jawaban yang sedekat mungkin dengan kejadian yang akan terjadi kedepannya [28]. Untuk mendapatkan hasil dari prediksi, tolak ukur yang biasa digunakan adalah dengan melihat hasil perhitungan kesalahan dari prediksi (*forecast error*), adapun dalam melihat hasil perhitungan kesalahan atau pengevaluasiannya dapat menggunakan diantaranya perhitungan *Mean Absolute Percentage Error* (MAPE), *Root Mean Square Error* (RMSE) atau *Mean Absolute Error* (MAE).

#### **2.1.12. Algoritma Linear Regression**

Algoritma *Linear Regression* sendiri merupakan jenis aturan *classification* dan *regression* pada kegiatan *data mining*, selain *Linear Regression* yang termasuk pada golongan ini, terdapat pula *Support Vector Machine*, *Logistic Regression* dan lain sebagainya. Analisis regresi linier adalah teknik *data mining* untuk menentukan bahwa terdapat hubungan antara variabel yang ingin

diramalkan dalam hal ini adalah variabel tidak bebas dengan variabel lain dalam hal ini adalah variabel bebas [29].

Adapun selanjutnya peramalan tersebut didasarkan pada sebuah asumsi mengenai pola pada pertumbuhan data sebuah historis yang bersifat linier, walaupun sebenarnya tidak mencapai 100% linier. Dan pola pertumbuhan ini didekati dengan suatu model yang menggambarkan hubungan-hubungan yang terkait dalam suatu keadaan. Selanjutnya dalam regresi linier, data dimodelkan dalam bentuk grafik dengan berbentuk garis *continues* dua dimensi. Oleh karena itu maka dibutuhkan variabel X dan Y, adapun dalam regresi linier variabel Y adalah sebagai variabel respon dan variabel X disebut sebagai variabel *predictor* [30]. Adapun formula dari kedua variabel tersebut dapat dilihat dengan rumus sebagai berikut [29] :

$$Y = a + bx \quad (2.1)$$

Keterangan:

Y = Variabel respon (variabel *dependent*)

a = Konstanta

b = Koefisien regresi (besaran respon yang dihasilkan oleh *predictor*)

x = Variabel prediktor (variabel bebas atau *independent*)

Adapun diperolehnya nilai a dan b dapat dihitung dengan rumus sebagai berikut:

$$\text{Nilai a} = \text{dengan cara } a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2} \quad (2.2)$$

$$\text{Nilai b} = \text{dengan cara } b = \frac{n(\sum XY) - (\sum X)(\sum Y)}{n(\sum X^2) - (\sum X)^2} \quad (2.3)$$

### 2.1.13. Random Forest Regression

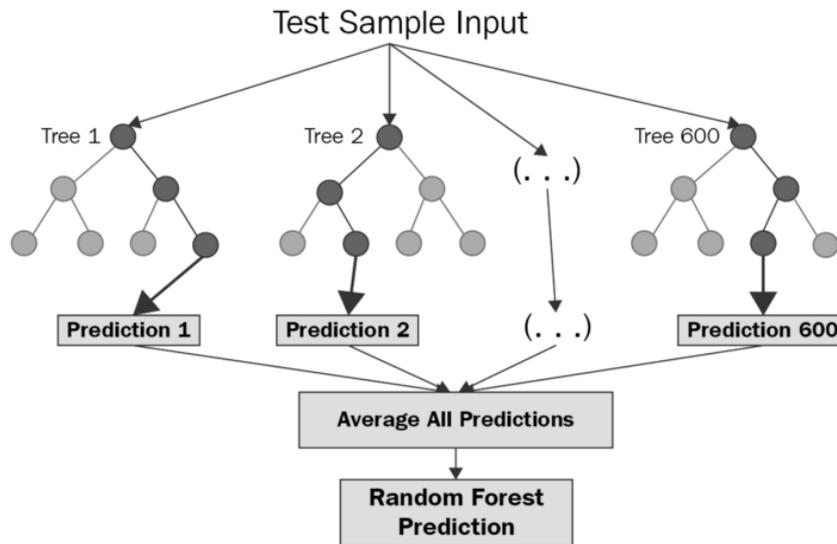
Metode *Random Forest* yang digunakan untuk pemodelan regresi disebut juga *Random Forest Regression* atau RFR. *Random Forest* dikenalkan oleh Breiman pada tahun 2001. *Random Forest* sendiri membangun *tree* menggunakan sampel *bootstrap* data yang berbeda dan mengubah cara regresi membangun pohon atau *tree*. Pada pohon standar, setiap node dibagi menggunakan *split*

terbaik di antara semua variabel, sedangkan pada *Random Forest* setiap node dibagi menggunakan yang terbaik diantara subset prediktor yang dipilih secara acak pada node tersebut. *Random Forest* mempunyai dua buah parameter, yaitu jumlah variabel dalam subset acak pada setiap node dan jumlah pohon [31].

Pada pemodelan ini banyak digunakan dalam pemodelan prediktif. Adapun metode ini memadukan percontohan ulang *bootstrap* dan seleksi peubah untuk mengurangi ragam galat prediksi akibat masalah prediktor berkorelasi tinggi dan meningkatkan ketepatan prediksi peubah respon dari metode regresi pohon. *Random Forest Regression* sendiri membangun banyak pohon regresi lalu menghitung nilai rata-rata hasil prediksi peubah respon dari semua pohon regresi tersebut. Gugus data asli yang terdiri dari  $p$  vektor prediktor  $x_1, x_1, \dots, x_p$  dan satu vektor peubah respon ( $y$ ) dinotasikan dengan  $Z$  gugus data baru yang dibuat dengan menggunakan percontohan ulang *bootstrap* dari  $Z$  dinotasikan dengan  $Z^*$ , dan fungsi pohon regresi pada gugus data hasil *bootraping* ke- $b$  dinotasikan dengan  $T_b(x^*)$ . Pada algoritma RFR, untuk memprediksi peubah respon dari percontohan ulang yang dilakukan sebanyak  $B$  adalah sebagai berikut [7]:

1. Pada percontohan ulang *bootstrap* ke  $b$ , dengan  $b = 1, 2, \dots, B$ , maka lakukan yang pertama, buat gugus data  $Z^*$  dari gugus asli menggunakan metode percontohan ulang *bootstrap*, dengan  $Z^*$  adalah gugus data yang terdiri dari  $p/3$  prediktor ( $x^*$ ) dan  $n$  amatan (jika  $p/3$  menghasilkan nilai desimal maka dibulatkan ke atas. Dan kedua buat pohon *Random Forest*,  $T_b(x^*)$  dari  $Z^*$  menggunakan metode pohon regresi.
2. Prediksi peubah respon dari metode RFR dengan  $B$  kali percontohan ulang *bootstrap*,  $\hat{f}_{rfr}^B(x)$  diperoleh dengan,  $\hat{f}_{rfr}^B(x) = (1/B) \sum_{b=1}^B T_b(x^*)$ .

Adapun diagram struktur pada algoritma random forest dapat dilihat pada Gambar 2.3 sebagai berikut:



**Gambar 2.3** Struktur Random Forest

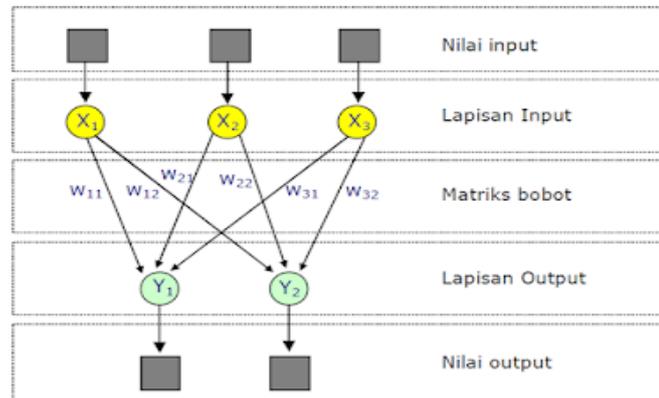
#### 2.1.14. Jaringan Syaraf Tiruan

Jaringan Syaraf Tiruan atau JST adalah paradigma pemrosesan suatu informasi yang terinspirasi oleh sistem sel saraf biologi. Pada jaringan ini biasanya diimplementasikan dengan menggunakan suatu komponen elektronik atau seperti dapat disimulasikan pada aplikasi komputer [32]. Selanjutnya Jaringan Syaraf Tiruan atau JST merupakan salah satu representasi buatan otak manusia yang selalu mencoba untuk mensimulasikan proses pembelajaran pada otak manusia. Adapun istilah buatan yang digunakan dikarenakan jaringan syaraf diimplementasikan dengan menggunakan program komputer yang mampu menyelesaikan sejumlah proses perhitungan selama proses pembelajaran atau *learning*. JST dimaksudkan untuk menghasilkan model sistem komputasi yang sesuai dengan cara kerja jaringan syaraf biologis [33].

Pada arsitekturnya, JST memiliki beberapa arsitektur jaringan yang cukup sering digunakan dalam berbagai aplikasi, arsitektur JST tersebut antara lain [14]:

#### 1. Jaringan Lapisan Tunggal (*Single Layer Network*)

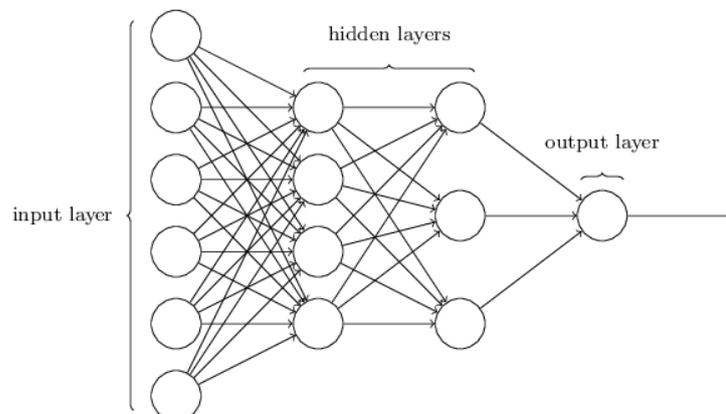
Pada arsitektur ini terdiri dari satu lapisan *input* dan satu lapisan *output*. Didalam lapisan *input* terdapat *neuron* yang selalu terhubung dengan setiap *neuron* yang terdapat pada lapisan *output*. Pada jaringan ini hanya menerima *input* kemudian secara langsung akan mengolahnya menjadi *output* tanpa harus melalui *hidden layer*. Berikut Gambar 2.4 yang merupakan arsitektur JST satu lapisan :



**Gambar 2.4** Arsitektur Lapisan Tunggal

## 2. Jaringan Banyak Lapisan (*Multilayer Network*)

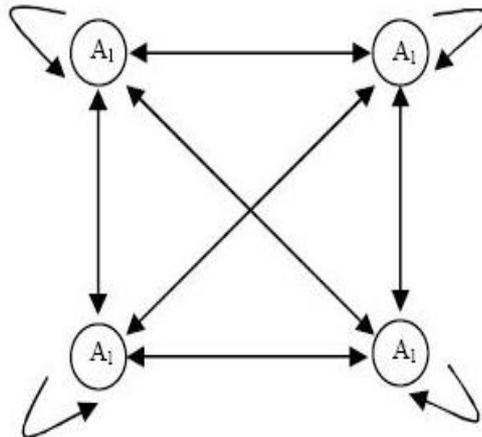
Pada jaringan ini memiliki ciri khas tersendiri yaitu memiliki tiga jenis lapisan yakni lapisan *input*, lapisan tersembunyi, dan lapisan *output*. Jaringan dengan banyak lapisan ini dapat menyelesaikan permasalahan yang lebih kompleks dibandingkan jaringan dengan *single layer*. Namun, pada proses pelatihannya sering membutuhkan waktu yang cukup lama. Berikut Gambar 2.5 yang merupakan arsitektur JST *Multilayer Network* :



**Gambar 2.5** Arsitektur *Multilayer Network*

### 3. Jaringan Lapisan Kompetitif (*Competitive Layer Network*)

Pada jaringan ini sekumpulan *neuron* bersaing untuk mendapatkan hak menjadi aktif. Umumnya, hubungan antar *neuron* pada lapisan kompetitif ini tidak diperlihatkan pada diagram arsitektur. Berikut salah satu contoh arsitektur jaringan lapisan kompetitif dapat dilihat pada Gambar 2.6 berikut::

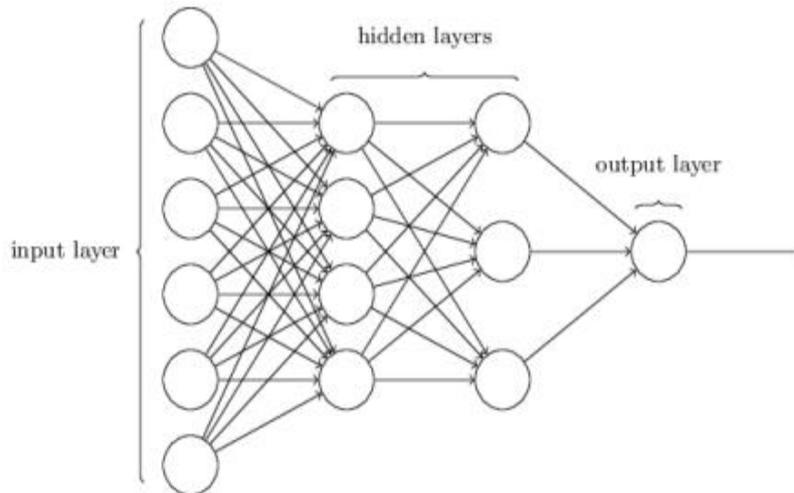


**Gambar 2.6** Arsitektur Lapisan Kompetitif

Adapun dari ketiga jenis arsitektur JST di atas, pada penelitian ini mengarah pada penggunaan arsitektur JST dengan *Multilayer Network* atau jaringan banyak lapisan.

#### 2.1.15. *Multilayer Perceptron*

Bentuk yang cukup umum digunakan pada peramalan jaringan syaraf tiruan adalah *Multilayer Perceptron*. *Multilayer Perceptron* atau MLP merupakan ANN dari *perceptron* yang berupa ANN *feedforward* dengan satu atau lebih *hidden layer*. Serta biasanya, jaringan terdiri dari satu lapisan *neuron* komputasi keluaran. Pada MLP, setiap lapisan memiliki fungsi khusus diantaranya lapisan masukan atau *input layer* berfungsi menerima sinyal atau vektor masukan dari luar dan mendistribusikannya ke semua *neuron* dalam *hidden layer*. Lalu pada lapisan keluaran menerima sinyal keluaran (atau dengan kata lain stimulus pola) dari *hidden layer* dan memunculkannya sinyal atau nilai atau kelas keluaran dari keseluruhan pada jaringan [34]. Arsitektur pada *Multilayer Perceptron* secara visual dapat dilihat pada Gambar 2.7 berikut:



**Gambar 2.7** Arsitektur *Multilayer Perceptron*

Pada visualisasi Gambar 2.7 arsitektur *Multilayer Perceptron* [35] dapat dijelaskan bahwa dalam melakukan peramalan, *output* yang dihasilkan hanya satu nilai dengan penggunaan dua *hidden layer* (adapun dalam penggunaannya, *hidden layer* dapat berjumlah lebih dari satu), yang kemudian pada setiap *hidden layer* terdapat 4 node untuk *hidden layer* pertama dan 3 node pada *hidden layer* kedua. Adapun pada *input layer* peramalan *one step ahead*,  $\hat{x}_{t+1}$  dihitung dengan menggunakan *input* yang berasal dari variabel lag  $k$  yang diteliti ( $x_{t-k}$ ) dan atau variabel eksogen lainnya. Adapun penjelasan pada setiap lapisan yang ada pada Gambar 2.6 arsitektur MLP adalah sebagai berikut [6]:

1. Lapisan *Input (input layer)*

Pada lapisan *input*, sebuah vektor dari nilai-nilai variabel prediktor ( $x_1 \dots x_p$ ) akan disajikan ke lapisan *input*. Lapisan *input* yaitu pengolahan sebelum *input layer* melakukan standarisasi nilai-nilai ini sehingga rentang dari setiap variabel adalah -1 ke 1. Selanjutnya lapisan masukan mendistribusikan nilai-nilai untuk setiap *neuron* pada lapisan tersembunyi. Adapun selain variabel prediktor, ada masukan konstan sebesar 1,0 yang disebut sebagai bias yang diumpankan ke masing-masing lapisan tersembunyi serta dikalikan dengan berat dan ditambahkan ke jumlah masuk ke *neuron*.

2. Lapisan tersembunyi (*hidden layer*)

Pada lapisan tersembunyi ini atau *hidden layer*, nilai dari setiap *neuron input* dikalikan dengan bobot  $w_{ji}$  dan nilai-nilai tertimbang yang dihasilkan ditambahkan bersama-sama menghasilkan nilai gabungan  $u_j$ . Jumlah tertimbang ( $u_j$ ) dimasukkan ke fungsi transfer, ( $\omega$ ) yang *output*  $h_j$ . Setelah itu keluaran yang dihasilkan dari lapisan tersembunyi didistribusikan ke *output layer*.

### 3. Lapisan keluaran (*output layer*)

Pada *output layer* ini, nilai yang terdapat dari setiap *neuron hidden layer* dikalikan dengan bobot ( $w_{kj}$ ) dan nilai-nilai tertimbang yang dihasilkan dijumlahkan bersama-sama yang menghasilkan nilai gabungan ( $v_j$ ). Adapun jumlah tertimbang ( $v_j$ ) diinput ke fungsi transfer ( $\omega$ ) yang *output* nilai  $y_k$ . selanjutnya nilai-nilai  $y$  adalah *output* yang terdapat dari jaringan.

Pada penerapan MLP, banyak algoritma pelatihan yang tersedia, namun yang cukup populer adalah *Backpropagation*. Adapun cara pelatihan yang dilakukan algoritma *Backpropagation* sama dengan *Perceptron*. Sejumlah data latih sebagai data masukan diberikan pada jaringan. Jaringan menghitung pada keluaran, dan bila ada *error* atau kesalahan pada kasus ini perbedaan antara target keluaran yang diinginkan dengan nilai keluaran yang didapatkan, maka bobot dalam jaringan akan diperbaharui untuk mengurangi kesalahan atau *error* tersebut. Adapun persamaan yang digunakan untuk pelatihan MLP *Backpropagation* [34] :

$$v = \sum_i^r x_i \cdot w_i \quad (2.4)$$

Keterangan :

$v$  = Nilai keluaran *hidden layer*.

$x_i$  = Nilai *input* atau fitur.

$w_i$  = Nilai bobot.

Pada nilai  $r$  adalah jumlah masukan (fitur) data masukan,  $x$  merupakan nilai fitur atau vektor, dan  $w$  adalah nilai bobot vektor. Nilai  $v$  tersebut kemudian di aktivasi untuk menghasilkan sinyal keluaran. Adapun fungsi aktivasi yang dapat digunakan terdapat dua fungsi yaitu *sigmoid biner* atau *sigmoid bipolar*. Persamaan dalam fungsi aktivasi sebagai berikut :

$$y = \frac{1}{1+e^{-v}} \quad (2.5)$$

Keterangan :

$y$  = Nilai sigmoid.

$e$  = Eksponen.

Untuk merambatkan sinyal *error*, dimulai dari *output layer* dan berjalan kembali ke *hidden layer*. Persamaan sinyal *error* di *neuron k* pada iterasi  $p$  sebagai berikut:

$$e_k(P) = y_{dk}(P) - y_k(P) \quad (2.6)$$

$e_k$  = Nilai selisih / *error*.

$y_{dk}$  = Nilai Sebenarnya.

$y_k$  = Nilai prediksi.

Adapun prosedur yang digunakan untuk memperbaharui nilai bobot pada koneksi *hidden layer* ke *output layer* sebagai berikut:

$$w_{jk}(p+1) = w_{jk}(p) + \Delta w_{jk} \quad (2.7)$$

Keterangan :

$\Delta w_{jk}$  = Koreksi nilai bobot

$w_{jk}$  = Nilai bobot

Adapun jika kondisi yang dialami adalah masukan *neuron* pada *output layer* berbeda dari *input neuron* pada *input layer*  $x_i$ . Oleh sebab itu yang dapat digunakan untuk menghitung koreksi bobot adalah sinyal *output neuron j* pada *hidden layer*  $y_j$  untuk menggantikan  $x_i$ , Persamaan koreksi bobot dalam MLP dapat dihitung dengan sebagai berikut :

$$\Delta w_{jk}(p) = \eta * x_i(p) * \delta_j(p) \quad (2.8)$$

Keterangan :

$\eta$  = *Learning rate*.

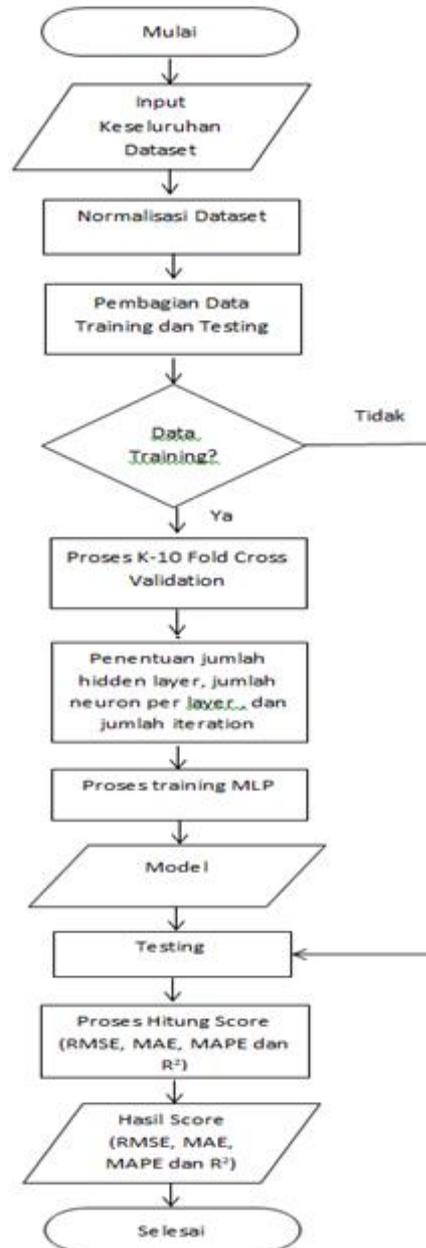
$\delta_j(p)$  = Gradien *error*.

$p$  = Iterasi

$\Delta w_{jk}(p)$  = Koreksi bobot.

Pada nilai  $\eta$  adalah laju pembelajaran, sedangkan  $\delta_j(p)$  adalah gradien *error* pada *neuron k* dalam *output layer* pada iterasi ke  $p$ .

Adapun pada penelitian ini dilakukannya proses ujicoba model MLP dengan struktur pengerjaan pada Gambar 2.8 sebagai berikut :



**Gambar 2.8** Diagram Model MLP Yang Diusulkan

\*Hasil Penelitian

### 2.1.16. Normalization

Pada tahapan pra proses suatu data, normalisasi merupakan proses data *transformation* yang dapat digunakan. Normalisasi adalah proses penskalaan nilai

atribut dari suatu data sehingga nilai atribut pada data tersebut dapat terletak pada rentang nilai tertentu atau yang sama. Terdapat beberapa normalisasi diantaranya *Min-Max Normalization*, *Z-Score Normalization* dan *Decimal Scaling Normalization*, adapun pada kasus penelitian yang penulis lakukan menggunakan *Min-Max Normalization*. *Min-Max Normalization* merupakan tahapan normalisasi dengan melakukan transformasi linier terhadap suatu data yang asli dengan tujuan menghasilkan keseimbangan serta standarisasi nilai perbandingan antar data saat sebelum ataupun sesudah proses [36]. Penskalaan data dapat dilakukan dengan rentang nilai antara -1 sampai 0 atau 0 sampai 1, adapun rumus pada metode normalisasi ini sebagai berikut [37] :

$$X' = \frac{x - x_{min}}{x_{max} - x_{min}} (new\ x_{max} - new\ x_{min}) + new\ x_{min} \quad (2.9)$$

Keterangan :

$X'$	= Perolehan normalisasi
$x$	= Nilai yang akan diproses
$x_{min}$	= Nilai data terkecil
$x_{max}$	= Nilai data terbesar
$new\ x_{max}$	= Nilai terbesar yang diinginkan
$new\ x_{min}$	= Nilai terkecil yang diharapkan

### 2.1.17. Evaluasi dan Model Validasi

#### 1. *Root Mean Square Error*

*Root Mean Square Error* atau RMSE merupakan salah satu parameter yang digunakan untuk mengevaluasi nilai hasil dari pengukuran terhadap nilai sebenarnya atau nilai dianggap benar. RMSE sebagai sebuah ukuran yang sering digunakan dari perbedaan antara nilai-nilai yang diprediksi oleh model atau estimator dan nilai-nilai yang benar-benar diamati.

Adapun perbedaan-perbedaan individual disebut residual ketika perhitungan tersebut dilakukan untuk sampel data yang digunakan buat estimasi, dan disebut juga sebuah kesalahan dalam melakukan prediksi ketika dihitung *out-of-sample*.

Nilai RMSE yang semakin rendah atau kecil membuktikan bahwa variasi nilai yang dihasilkan oleh suatu model prakiraan mendekati variasi nilai observasinya. Secara umum persamaan yang digunakan untuk menghitung nilai RMSE adalah sebagai berikut [38] :

$$RMSE = \sqrt{\frac{\sum(y_t - \hat{y}_t)^2}{n}} \quad (2.10)$$

Keterangan :

$RMSE$  = Root Mean Square Error

$n$  = Jumlah Data (Sampel)

$y_t$  = Nilai Aktual Indeks

$\hat{y}_t$  = Nilai Prediksi Indeks

## 2. Mean Absolute Error

*Mean Absolute Error* juga digunakan sebagai salah satu perhitungan evaluasi kesalahan dari peramalan, MAE merupakan rata-rata nilai *absolute error* dari kesalahan peramalan, adapun rumus yang dapat digunakan sebagai berikut [39] :

$$MAE = \frac{\sum |X_t - F_t|}{n} \quad (2.11)$$

Keterangan :

$X_t$  = Data aktual pada periode ke  $t$

$F_t$  = Peramalan untuk periode  $t$

$n$  = jumlah data yang diramalkan

## 3. Mean Absolute Percentage Error

*Mean Absolute Percentage Error* atau MAPE merupakan salah satu perhitungan evaluasi yang digunakan untuk mengukur seberapa tepat atau seberapa akurat suatu prediksi yang sering digunakan dalam penelitian, pada penggunaan MAPE kita bisa mendapatkan nilai selisih antara nilai aktual dengan nilai prediksi, adapun rumus perhitungan MAPE dapat dilihat sebagai berikut [40]:

$$MAPE = \frac{100}{n} \sum_{t=0}^n \left| \frac{\hat{y}_t - y_t}{y_t} \right| \quad (2.12)$$

Keterangan :

$\hat{y}_i$  = Hasil Prediksi

$y_i$  = Nilai Aktual

$n$  = Jumlah Data (sampel yang diuji)

Sebagai persentase dari nilai rata-rata *error rate absolute* periode data aktual, MAPE akan mengukur rata-rata dari *error* absolutnya . Nilai MAPE memiliki kriteria semakin kecil nilai MAPE maka nilai akurasi semakin baik. Kriteria nilai MAPE dapat dilihat pada Gambar 2.9 di bawah ini :

Nilai MAPE	Kriteria
<10	Sangat Baik
10-20	Baik
20-50	Cukup
>50	Buruk

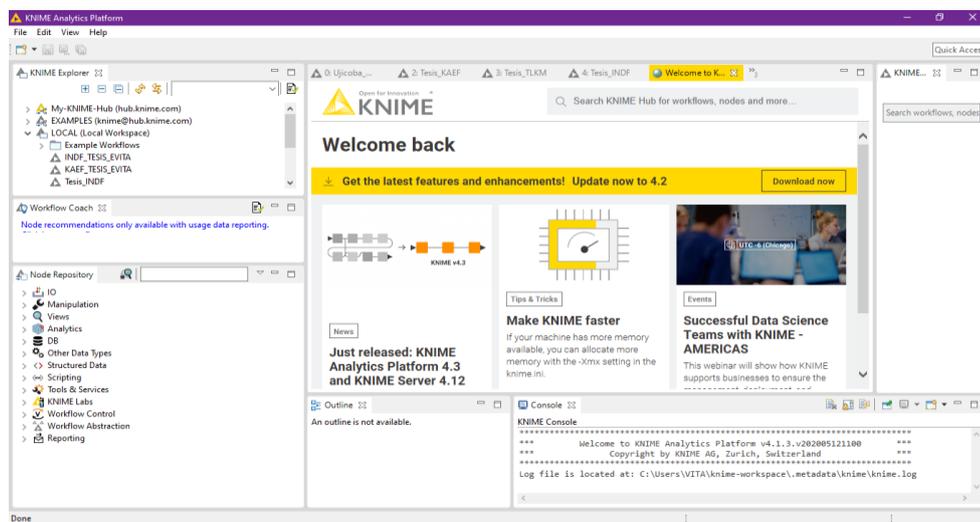
**Gambar 2.9** Kriteria Nilai MAPE

#### 4. Koefisien Determinasi ( $R^2$ )

Koefisien Determinasi atau  $R^2$  menurut [41] bahwa  $R^2$  pada intinya mengukur seberapa jauh kemampuan model prediksi yang diujikan dalam menerangkan variasi variabel aktual. Adapun nilai koefisien determinasi adalah antara nol (0) hingga satu (1). Pada nilai  $R^2$  yang berjumlah kecil berarti kemampuan variabel-variabel bebas dalam menjelaskan variasi variabel aktualnya amat terbatas. Dan pada nilai yang mendekati satu (1) berarti variabel-variabel bebas memberikan hampir semua informasi yang dibutuhkan untuk memprediksi variasi variabel aktualnya. Kelemahan pada  $R^2$  ini adalah bias terhadap jumlah variabel bebasnya, setiap ada tambahan variabel bebas maka nilai  $R^2$  meningkat tidak peduli apakah variabel tersebut cukup berpengaruh pada variabel aktualnya, oleh sebab itu banyak penelitian yang merekomendasikan untuk menggunakan  $R^2$  terhadap pengevaluasian suatu model.

#### 2.1.18. KNIME *Analytics Platform*

Dalam memproses atau melakukan *data mining*, banyak *tools* yang dapat digunakan untuk pemrosesannya, salah satu *tools* yang dapat digunakan adalah KNIME. Adapun pengembangan KNIME sendiri dimulai pada Januari 2004 oleh *developer* perangkat lunak di Konstanz University sebagai produk *proprietary*. Tim pengembang asli yang dipimpin oleh Michael Berthold berasal dari sebuah perusahaan di Silicon Valley yang menyediakan perangkat lunak untuk industri farmasi. Sejak awal melalui proses rekayasa perangkat lunak yang ketat KNIME sendiri telah dikembangkan dan sejak awal bertujuan untuk digunakan pada perusahaan-perusahaan berskala besar. Tujuan awalnya adalah untuk menciptakan sebuah *platform* pengolahan data modular, sangat sederhana dan *opensource* yang memungkinkan untuk integrasi yang dapat digunakan pada banyak *platform* seperti *windows*, *linux*, *Mac OS*. Adapun *platform* ini dimaksudkan sebagai media kolaborasi dan penelitian serta berperan sebagai *platform* integrasi untuk berbagai banyak proyek analisis data yang lainnya [42].



**Gambar 2.10** Tampilan Halaman Utama KNIME

Pada penggunaannya, KNIME menggabungkan ratusan node pengolahan data *input-output*, *preprocessing* dan pembersihan, pemodelan, analisis dan *data mining* ini yang disebut bahwa *KNIME Analytics* memungkinkan pengguna melakukan visualisasi aliran data masuk (*pipelines*), menjalankan tahapan analisis tertentu atau semua tahapan, dan melihat hasil dari pemodelan secara interaktif.

KNIME sendiri ditulis dengan bahasa Java (dengan *platform eclipse*) dan dilengkapi dengan mekanisme *plugin* untuk kemudahan pengembangan dan penambahan fungsionalitas dan juga mengintegrasikan semua modul analisis weka *data mining* terkenal serta *plugin* tambahan memungkinkan *R-script* dijalankan. Adapun pada versi utamanya, KNIME sudah meliputi ratusan modul integrasi data seperti *file I/O*, dukungan terhadap DBMS umum, dan juga transformasi data yang terdiri dari *convert*, *filter*, dan *combiner*, serta metoda umum yang digunakan dalam melakukan analisis dan visualisasi data penelitian [42].

## 2.2. Tinjauan Studi

Berikut adalah beberapa penelitian *data mining* terkait dalam memprediksi serta menganalisa dataset saham yang diantaranya adalah sebagai berikut :

Penelitian yang berjudul “*Prediksi Harga Saham Dengan SVM (Support Vector Machine) Dan Pemilihan Fitur F-Score*” melakukan prediksi harga saham berdasarkan analisa teknikal dengan menggunakan indikator teknikal sebagai fiturnya. Data yang dikumpulkan berupa data *time series*. Penelitian menggunakan metode SVM untuk prediksi dan melakukan pemilihan fitur (*feature selection*) dengan metode *F-Score*. Penelitian ini telah menyelesaikan pengembangan *prototype* yang dibutuhkan untuk melakukan prediksi saham. Pengumpulan data saham dari Bursa Efek Indonesia telah dilaksanakan dan proses prediksi yang dibutuhkan juga telah dilaksanakan. Metode *F-Score* membutuhkan proses yang jauh lebih sedikit dari pada proses maksimal yang mungkin dilakukan untuk mencari fitur terbaik terutama setelah jumlah fitur melebihi 5. Metode *F-Score* juga memberikan hasil 70% *F-Score* dan 71% akurasi. Hasil ini hanya 3% lebih buruk daripada pilihan fitur terbaik yang mungkin dicapai [43].

Pada penelitian dengan judul “*Analisis dan Perbandingan Algoritma Data mining Dalam Prediksi Harga Saham GGRM*”. Pada penelitian ini dilakukan prediksi harga saham dengan 5 atribut yaitu *date*, *open*, *high*, *low*, dan *close* sebagai labelnya, adapun kerangka pengerjaannya dimulai dari analisis dataset, implementasi algoritma prediksi terdiri dari *Neural Network*, *Linear Regression*, *SVM*, *Gaussian Process* dan *Polynomial Regression*, serta implementasi model

validasi dengan 10 *Fold Cross Validation* dan untuk evaluasinya dengan melihat nilai terendah dari RMSE nya. Hasil penelitian ini menyimpulkan bahwa data harga saham GGRM dapat diprediksi dengan menggunakan model algoritma *Neural Network*, dengan hasil akurasi prediksi RMSE 612.474 +/- 89.402 (mikro: 618.916 +/- 0.000) paling kecil dibandingkan dengan model algoritma lainnya, sehingga dengan prediksi ini dapat membantu dalam memprediksi harga saham GGRM di pasar modal [5].

Pada penelitian dengan judul “*Pengembangan Model Prediksi Harga Saham Berbasis Neural Network*” yang memprediksi harga saham dengan menggunakan lima atribut yakni *date, open, high, low, close* dan menggunakan model jaringan syaraf sebagai proses pelatihan, pembelajaran, dan *testingnya*. Setelah serangkaian proses penelitian mulai dari pelatihan, belajar, untuk melanjutkan proses pengujian hasilnya adalah bahwa dengan menggunakan satu set data 132 hasil *trend* akurasi prediksi 0.456 +/- 0,044 dan menggunakan dataset 256 akurasi prediksi meningkat menjadi 0481 +/- 0169. dengan menggunakan satu set data yang lebih maka dapat meningkatkan akurasi prediksi nilai *trend* yang dibuktikan dengan waktu percobaan pertama untuk menggunakan data sebanyak 132 dan pada kedua mencoba untuk menggunakan data sebanyak akurasi prediksi 256 *trend* meningkat 2,5% yaitu 0,025 +/- 0,125 [44].

Penelitian dengan judul “*Prediksi Indeks Harga Saham dengan Metode Vector Regression dan Jaringan Gabungan Support Syaraf Tiruan*” mengusulkan metode untuk memprediksi pergerakan harga saham dengan empat atribut (*open, high, low, close*) menggunakan metode gabungan *Support Vector Regression* (SVR) pada tahap pertama dan Jaringan Syaraf Tiruan (JST) pada tahap kedua. Pada penelitian ini, Algoritma Genetika atau *Genetic Algorithm* (GA) akan digunakan untuk melakukan optimasi parameter SVR. Prediksi dibuat untuk 1, 3, 5, 7, 10, 15, dan 30 hari kedepan. Dari serangkaian uji coba yang dilakukan, SVR-JST (SVR dioptimasi GA) memberikan tingkat kesalahan lebih kecil dibandingkan dengan metode JST. Metode gabungan SVR-JST (SVR dioptimasi GA) dapat digunakan sebagai model prediksi pergerakan harga saham harian Indonesia dan mampu menunjukkan performa yang cukup baik, hal ini dapat

ditunjukkan berdasarkan nilai selisih antara hasil prediksi dengan harga aktual MAPE yang relatif kecil yakni berkisar antara 1% hingga 10%. Kemudian berdasarkan hasil perbandingan MAPE dan RMSE dengan metode JST untuk memprediksi Indeks Harga Saham Gabungan, tingkat kesalahan metode SVR-JST lebih kecil dibandingkan metode JST (tanpa *hybrid*) pada 3, 7, 10, 15 dan 30 hari kedepan [45].

Dalam penelitian dengan judul “*Model Prediksi Rentet Waktu Neural Network Berbasis Particle Swarm Optimization Untuk Prediksi Harga Saham*” dilakukan pengujian model dengan menggunakan *Neural Network* dan *Neural Network* dengan *Particle Swarm Optimization* dengan menggunakan data harga saham terdiri dari lima atribut (*date, open, high, low, close*). Model yang dihasilkan diuji untuk mendapatkan nilai RMSE yang lebih rendah, didalam pengujian tersebut *Neural Network* mendapatkan nilai RMSE sebesar 0,466 sedangkan *Neural Network* dengan pemilihan fitur dengan menggunakan *Particle Swarm Optimization* mendapatkan nilai RMSE 0,373 atau 0,093 lebih kecil dari menggunakan *Neural Network* saja. Dengan demikian dapat disimpulkan bahwa *Neural Network* dengan *Particle Swarm Optimization* untuk pemilihan fitur pada harga saham lebih baik dari pada dengan menggunakan *Neural Network* saja serta dapat memecahkan masalah prediksi menjadi lebih akurat [46].

Selanjutnya penelitian dengan judul “*Penerapan Algoritma Artificial Neural Network Dalam Prediksi Harga Saham LQ45 PT. Bank Rakyat Indonesia, TBK*”. Pada penelitian ini menerapkan algoritma *Artificial Neural Network* untuk memprediksi harga saham LQ45 dengan studi kasus Bank BRI. Penelitian ini memakai data *time series* Dengan menggunakan empat atribut yaitu nilai *open, high, low* sebagai prediktor dan *close* sebagai *class*, penelitian ini berfokus pada penentuan nilai akurasi, *Root Mean Squared Error* (RMSE) dan *Normalized Mean Absolute Error* (NMAE), dengan mengoptimalkan nilai-nilai parameter. Hasil optimasi parameter tersebut akan diujikan, dengan pemilihan ukuran *hidden layer* ( 3, 9, 15 *neuron*) pada algoritma *Artificial Neural Network*. Dengan mengoptimalkan nilai-nilai parameter dan pemilihan ukuran *hidden layer*

memberikan hasil yang lebih baik, terbukti dengan hasil akurasi, RMSE dan NMAE masing-masing sebagai berikut : 81.80%, 22.042, dan 0.028, [47].

Penelitian dengan judul “*Prediksi Data Historis Saham PT.Bank Rakyat Indonesia Tbk (BBRI) Menggunakan Model Algoritma Artificial Neural Network*”. penelitian ini membahas peramalan saham menerapkan algoritma *Artificial Neural Network* untuk memprediksi harga saham BBRI. Dengan menggunakan empat atribut yaitu nilai *open, high, low* sebagai prediktor dan *close* sebagai *class*, penelitian ini berfokus pada penentuan *Root Mean Squared Error* (RMSE) dengan mengoptimalkan nilai-nilai parameter. Dengan mengoptimalkan nilai-nilai parameter dan pemilihan ukuran *hidden layer* memberikan hasil yang lebih baik, terbukti dengan hasil akurasi, RMSE masing-masing sebagai berikut : 0.266 +/- 0.000 Dan *hidden layer* pada *Artificial Neural Network* dengan jumlah *neuron* 16 mampu memberikan hasil terbaik. [4].

Penelitian dengan judul “*Stock Price Prediction Using Genetic Algorithms And Evolution Strategies*” dilakukan prediksi harga saham menggunakan algoritma genetika dan strategi evolusi yang terlihat sangat menjanjikan. Prediksi dilakukan dengan tujuh atribut yaitu *open, high, low, close, adj.close, volume* dan akurasi terbaik yang ditemukan dengan menggunakan algoritma genetika adalah 73.87% dan dengan menggunakan strategi evolusi adalah 71.77%. Algoritma genetika mampu memprediksi lebih baik dari pada strategi evolusi di lima kasus. Strategi evolusi mencapai akurasi 70% atau lebih baik dalam semua kasus. Pada penelitian ini menggunakan dua dataset yang berbeda untuk memprediksi harga saham yang pertama bertindak sebagai data pelatihan dan tindakan lainnya sebagai perangkat pengujian. [48].

Penelitian dengan judul “*Stock Price Prediction Using Machine Learning and Swarm Intelligence*”. Dalam penelitian ini, menggunakan pendekatan *machine learning* yang bekerja dalam dua tahap, diperkenalkan untuk memprediksi harga saham di hari berikutnya berdasarkan informasi yang diambil dari 26 hari terakhir. Pada fase pertama metode *Automatic Clustering Algorithm* mengelompokkan titik data ke dalam klaster yang berbeda, dan pada fase kedua model regresi hibrid, yang merupakan kombinasi dari *Particle Swarm*

*Optimization* dan SVR, dilatih untuk setiap klaster. Pada metode *hybrid* ini, algoritma *Particle Swarm Optimization* digunakan untuk penyetelan parameter dan pemilihan fitur. Hasilnya Akurasi metode yang diusulkan diukur dari lima dataset perusahaan yang aktif di pasar Bursa Efek Tehran melalui lima metrik yang berbeda. Secara rata-rata, metode yang diusulkan telah menunjukkan akurasi 82,6% dalam memprediksi harga saham dalam satu hari ke depan. Adapun metode yang digunakan adalah SVR, *Automatic Clustering Algorithm* (APSO-*Clustering*), PSO-SV dan atributnya berjumlah lima yaitu *open*, *high*, *low*, *close* dan *volume*. [49].

Dan pada penelitian dengan judul “*Stock Price Prediction Based on Machine Learning Approaches*” melakukan penelitian prediksi harga saham pada pasar China dengan dua atribut yaitu *opening* dan *closing* dengan periode Januari 2008 hingga Januari 2017 menggunakan model *Logistic Regression* dan SVR. Hasil penelitian menunjukkan bahwa regresi *Ridge* dan SVM dengan kernel linier lebih baik daripada regresi *Lasso* untuk akurasi, presisi, dan skor F1. Dengan *score* akurasi, *precision* dan *F1-Score* dari masing-masing model yaitu : *Lasso Regression* 0.71/0.73/0.72, *Ridge Regression* 0.74/0.78/0.76 dan SVM 0.77/0.79/0.78. [3].

Berikut merupakan ringkasan atau rangkuman dari penelitian terkait sebelumnya yang pernah dilakukan oleh beberapa peneliti lain dan dijadikan sebagai sumber referensi, adapun referensi didapat dari laman *google scholar* dan dapat dilihat pada Tabel 2.1 sebagai berikut:

Tabel 2.1 Rangkuman Perbandingan Jurnal Referensi

No	Judul	Sumber	Metode	Hasil
1	Prediksi Harga Saham Dengan SSM (Support Vector Machine) Dan Pemilihan Fitur	Jurnal Informatika UPGRIS Vol. 6, No. 1 Juni 2020 P/E-ISSN: 2460-	SVM Dan Pemilihan Fitur <i>F-Score</i>	Metode <i>F-Score</i> memberikan hasil 70% <i>F-Score</i> dan 71% akurasi.

	F-Score	4801/2447-6645		
2	Analisis dan Perbandingan Algoritma <i>Data mining</i> Dalam Prediksi Harga Saham GGRM	Jurnal Informatika Kaputama (JIK), Vol 3 No 1, Januari 2019	NN, <i>Linear Regression</i> , SVM, <i>Gaussian Process</i> , <i>Polynomial Regression</i>	Model algoritma <i>Neural Network</i> , dengan hasil akurasi prediksi RMSE 612.474 +/- 89.402 (mikro: 618.916 +/- 0.000)
3	Pengembangan Model Prediksi Harga Saham Berbasis <i>Neural Network</i>	Jurnal DISPROTEK, Th. 2016	NN	Menggunakan satu set data 132 hasil tren akurasi prediksi 0.456 +/- 0,044 dan menggunakan dataset 256 akurasi prediksi meningkat menjadi 0481 +/- 0169 dan percobaan kedua menggunakan data sebanyak 256 akurasi prediksi tren meningkat 2,5% yaitu 0,025 +/- 0,125.
4	Prediksi Indeks Harga Saham dengan Metode <i>Vector Regression</i> dan Jaringan	Ind. Journal on Computing, Th. 2017	Gabungan SVR-JST (SVR Dioptimasi oleh <i>Genetic Algorithm</i> )	Tingkat kesalahan metode SVR-JST lebih kecil dibandingkan metode JST (tanpa <i>hybrid</i> ) 1% hingga 10%

	Gabungan Support Syaraf Tiruan			
5	Model Prediksi Rentet Waktu <i>Neural Network</i> Berbasis <i>Particle Swarm Optimization</i> Untuk Prediksi Harga Saham	Telematika, Vol. 14, No. 02, Oktober, 2017, Pp. 100 – 106 ISSN 1829-667X	NN dan NN Menggunakan PSO ( <i>Particle Swarm Optimization</i> )	RMSE dengan NN sebesar 0,466, NN-PSO 0,373 atau 0,093
6	Penerapan Algoritma <i>Artificial Neural Network</i> Dalam Prediksi Harga Saham LQ45 PT. Bank Rakyat Indonesia, TBK	Jurnal Gerbang Vol.8 No.1 , Februari 2018	ANN	Akurasi, RMSE dan NMAE masing-masing sebagai berikut : sebesar 81.80%, 22.042, dan 0.028
7	Prediksi Data Historis Saham PT.Bank Rakyat Indonesia Tbk (BBRI) Menggunakan Model Algoritma	<i>The 10th University Research Colloqium</i> 2019 Sekolah Tinggi Ilmu Kesehatan Muhammadiyah	ANN	Hasil akurasi, RMSE 0.266 +/- 0.000 Dan <i>hidden layer</i> pada <i>Artificial Neural Network</i> dengan jumlah <i>neuron</i> 16 mampu memberikan hasil terbaik.

	<i>Artificial Neural Network</i>	ah Gombong		
8	<i>Stock Price Prediction Using Genetic Algorithms And Evolution Strategies.</i>	<i>International Conference on Electronics, Communication and Aerospace Technology</i> ICECA 2017	<i>Genetic Algorithm, Evolving Strategies</i>	Dengan algoritma genetika adalah 73.87% dan dengan menggunakan strategi evolusi adalah 71.77%.
9	<i>Stock Price Prediction Using Machine Learning and Swarm Intelligence</i>	<i>Journal of Electrical and Computer Engineering Innovations</i> (JECEI) Journal homepage: <a href="http://www.jecei.sru.ac.ir">http://www.jecei.sru.ac.ir</a> J. Electr. Comput. Eng. Innovations, 8(1): 31-40, 2020	SVR, <i>Automatic Clustering Algorithm</i> (APSO-Clustering), PSO-SVR	Secara rata-rata, metode yang diusulkan telah menunjukkan akurasi 82,6%
10	<i>Stock Price Prediction</i>	<i>DSIT 2020: Proceedings</i>	<i>Logistic Regression,</i>	Dengan <i>score</i> akurasi, <i>precision</i> dan <i>F1-Score</i>

	<i>Based on Machine Learning Approaches</i>	<i>of the 3rd International Conference on Data Science and Information Technology July 2020</i>	SVM	dari masing-masing model yaitu : <i>Lasso Regression</i> 0.71/0.73/0.72, <i>Ridge Regression</i> 0.74/0.78/0.76 dan SVM 0.77/0.79/0.78.
--	---	---	-----	---

Berdasarkan Tabel 2.1 perbedaan dari penelitian terkait dengan penelitian yang akan diujikan pada tesis ini adalah bahwa pada penelitian ini menganalisis historis harga saham untuk memprediksi tingkat kesalahan dalam setiap model yang digunakan dan menggunakan dataset historis harga saham dari tiga sektor industri di Indonesia yaitu perusahaan dengan kode INDF, KAEF dan TLKM. Adapun pada penelitian ini data atribut yang digunakan yaitu historis harga *open*, *high*, *low* dan *close* dan model prediksi yang digunakan adalah *Linear Regression*, *Random Forest Regression* dan *Multilayer Perceptron*.

### **2.3. Tinjauan Organisasi/Objek Penelitian**

Adapun tinjauan singkat dari objek yang diteliti adalah menggunakan dataset historis harga saham dari tiga sektor industri di Indonesia diantaranya sebagai berikut:

#### **2.3.1 PT Indofood Sukses Makmur Tbk (INDF)**

Berdasarkan profil Indofood pada laman resminya, PT Indofood Sukses Makmur didirikan tanggal 14 Agustus 1990 dengan nama PT Panganjaya Intikusuma dan memulai kegiatan usaha komersialnya pada tahun 1990. Indofood telah bertransformasi menjadi sebuah perusahaan *Total Food Solutions* dengan kegiatan operasional yang mencakup seluruh tahapan proses produksi makanan.

Mulai dari produksi dan pengolahan bahan baku hingga menjadi produk akhir yang tersedia di pasar. Kini Indofood dikenal sebagai perusahaan yang

mapan dan terkemuka di setiap kategori bisnisnya. Dalam menjalankan kegiatan operasionalnya, Indofood memperoleh manfaat dari skala ekonomis serta ketangguhan model bisnisnya yang terdiri dari empat kelompok usaha strategis ("Grup") yang saling melengkapi [50].

Induk usaha dari Indofood Sukses Makmur Tbk adalah *CAB Holding Limited* (miliki 50,07% saham INDF) Seychelles, sedangkan induk usaha terakhir dari Indofood Sukses Makmur Tbk adalah *First Pacific Company Limited* (FP) Hong Kong. Saat ini, Perusahaan memiliki anak usaha yang juga tercatat di Bursa Efek Indonesia (BEI), antara lain: Indofood CBP Sukses Makmur Tbk (ICBP) dan Salim Ivomas Pratama Tbk (SIMP) [51]

### **1.3.2. PT. Kimia Farma (KAEF)**

PT. Kimia Farma adalah perusahaan industri farmasi pertama di Indonesia yang didirikan oleh Pemerintah Hindia Belanda tahun 1817. Nama perusahaan ini pada awalnya adalah NV Chemicalien Handle Rathkamp & Co. Berdasarkan kebijaksanaan nasionalisasi atas eks-perusahaan Belanda di masa awal kemerdekaan, pada tahun 1958, Pemerintah Republik Indonesia melakukan peleburan sejumlah perusahaan farmasi menjadi PNF (Perusahaan Negara Farmasi) Bhinneka Kimia Farma.

Kemudian pada tanggal 16 Agustus 1971, bentuk badan hukum PNF diubah menjadi Perseroan Terbatas, sehingga nama perusahaan berubah menjadi PT Kimia Farma (Persero). Mencatatkan saham perdana untuk publik (IPO) pada tanggal 4 Juli 2001 dengan kode emiten KAEF dan komposisi saham 90,025% milik pemerintah dan 9,975% milik publik. Melalui proses *inbreng* yang dilaksanakan Pemerintah Republik Indonesia pada 28 Februari 2020, kepemilikan saham 4.999.999.999 saham seri B dialihkan kepada PT Biofarma [52].

### **1.3.3. PT Telkom Indonesia (Persero) Tbk (Telkom)**

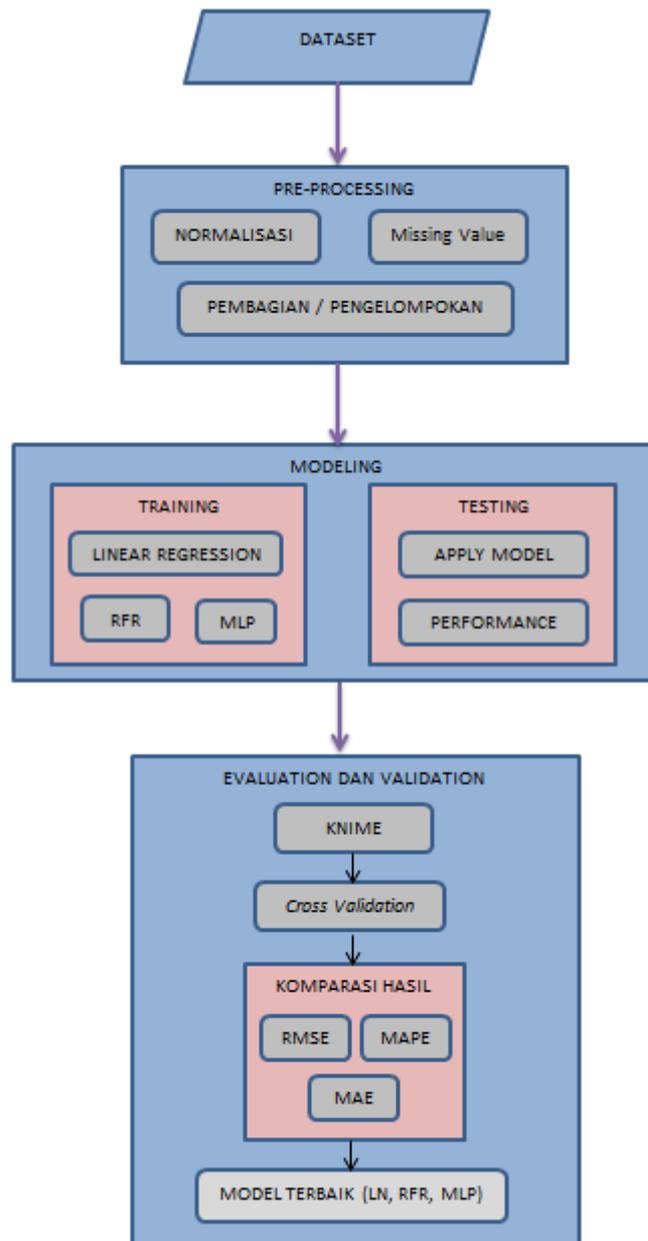
PT. Telkom adalah Badan Usaha Milik Negara (BUMN) yang bergerak di bidang jasa layanan teknologi informasi dan komunikasi (TIK) dan jaringan telekomunikasi di Indonesia. Pemegang saham mayoritas Telkom adalah

pemerintah Republik Indonesia sebesar 52.09%, sedangkan 47.91% sisanya dikuasai oleh publik. Saham Telkom diperdagangkan di Bursa Efek Indonesia (BEI) dengan kode “TLKM” dan *New York Stock Exchange* (NYSE) dengan kode “TLK”.

Dalam upaya bertransformasi menjadi *digital telecommunication company*, Telkom Group mengimplementasikan strategi bisnis dan operasional perusahaan yang berorientasi kepada pelanggan (*customer-oriented*). Transformasi tersebut akan membuat organisasi Telkom Group menjadi lebih *lean* (ramping) dan *agile* (lincah) dalam beradaptasi dengan perubahan industri telekomunikasi yang berlangsung sangat cepat. Organisasi yang baru juga diharapkan dapat meningkatkan efisiensi dan efektivitas dalam menciptakan *customer experience* yang berkualitas [53].

#### **1.4. Kerangka Pemikiran**

Adapun kerangka pemikiran dalam penelitian ini digambarkan pada Gambar 2.11 sebagai berikut :



Sumber : Hasil Penelitian

**Gambar 2.11** Kerangka Pemikiran

Pada Gambar 2.9 Kerangka Pemikiran di atas, terdapat tahap-tahapan yang dilakukan pada penelitian ini, adapun penjelasan dari masing-masing tahapan di atas adalah sebagai berikut:

1. Pada tahapan pertama dalam proses penelitian ini adalah pengumpulan data yang akan diujikan. Pengumpulan dataset didapatkan dari laman

<https://finance.yahoo.com> dengan mengambil data historis data harga saham yang telah disediakan, adapun rentang waktu data yang diambil adalah historis harga saham pada perusahaan INDF, KAEF dan TLKM periode 01 September 2015 sampai dengan 01 September 2020 dengan jumlah atribut sebanyak empat (*Open, High, Low, Close*).

2. Tahap kedua yakni tahapan *preprocessing* data, pengolahan data setelah dataset didapatkan. Pada tahap ini dilakukan proses pengolahan data diantaranya pengecekan *missing value* data yang pada pengerjaannya yaitu membersihkan data dari data yang kosong atau hilang, selanjutnya dilakukan normalisasi data sebagai proses data *transformation* untuk menyeimbangkan nilai pada setiap *record*, dikarenakan setiap data memiliki rentang data yang berbeda, dan terakhir yaitu adanya proses pembagian data yang terdiri dari data *training* dan data *testing*.
3. Selanjutnya tahap *modelling* prediksi, pada tahap ini prediksi dilakukan dengan menggunakan tiga model prediksi yaitu *Linear Regression, Random Forest Regression* dan *Multilayer Perceptron*.
4. Dan tahapan terakhir pada penelitian ini yaitu proses evaluasi, yang dilakukan menggunakan *tools* KNIME, dimana validasi data dilakukan pada data *training* dengan 10 *K-Fold Cross Validation* dan pada proses evaluasi ini diukur pada hasil perolehan RMSE, MAE dan MAPE yang lebih rendah dari masing-masing uji coba dataset dengan ketiga model prediksi (LR, RFR, ML) yang selanjutnya dibandingkan dan didapat hasil kesimpulan model prediksi mana yang cukup baik digunakan dalam prediksi historis harga saham diantara model LR, RFR dan MLP.

## BAB III METODE PENELITIAN

### 3.1. Analisis Kebutuhan

Pada analisis kebutuhan penelitian ini yang melakukan eksperimen pengujian dataset historis harga saham dengan model prediksi yang diusulkan yaitu LR, RFR dan MLP. Adapun *hardware* serta *software* yang digunakan dalam penelitian ini memiliki spesifikasi seperti Tabel 3.1 sebagai berikut :

Tabel 3.1. Spesifikasi Kebutuhan

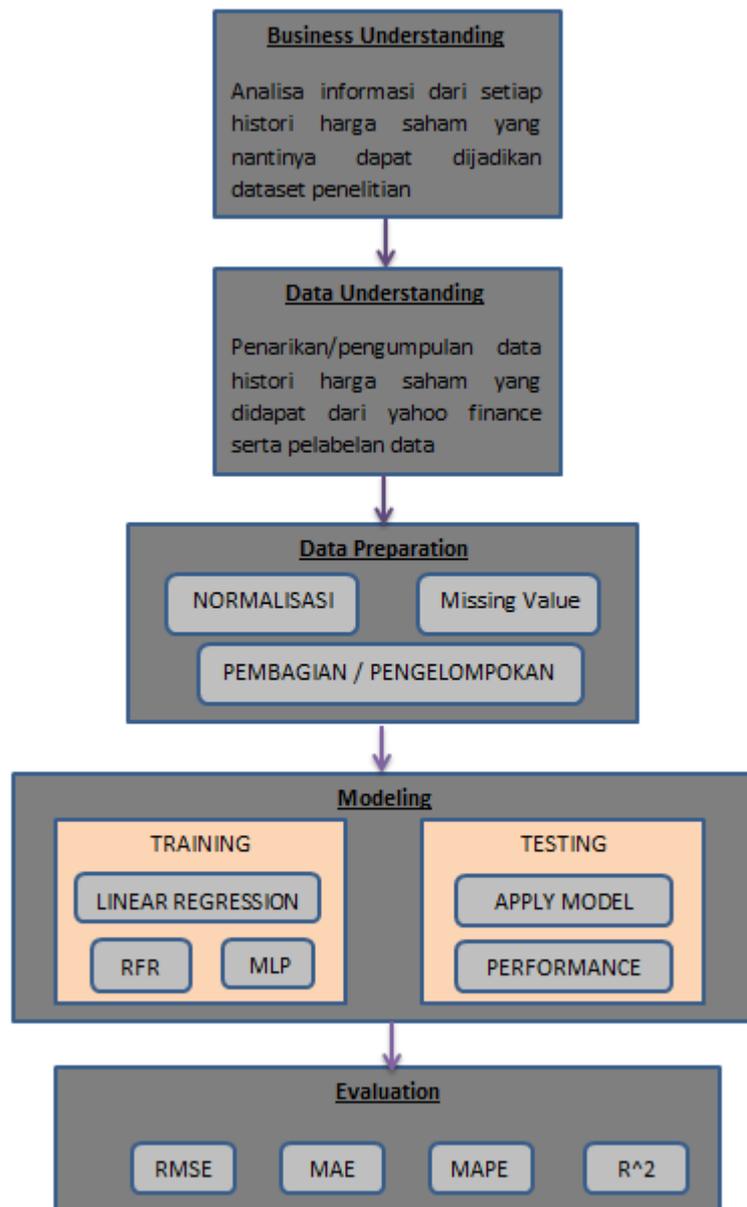
NO	<i>Hardware / Software</i>	Keterangan
1	Komputer	Processor : Intel® Pentium® Processor P6100 @2.00GHz, Memory (RAM) 2 GB dengan monitor 14.0" HD LED LCD dan sistem operasi Windows 10 Professional 32 bit Operating System, x64-based processor
2	KNIME Vers. 4.1.3	Software pengolahan <i>data mining</i>
3	Microsoft Excel 2010	Software penyimpanan dan pengecekan dataset (xls)

Dalam penelitian ini, pengolahan dataset menggunakan *framework* KNIME sebagai *tools* untuk memproses data historis harga saham dari mulai *preprocessing* dan *modelling* sehingga menghasilkan nilai yang dapat dievaluasi berupa tingkat kesalahan atau nilai *error* pada masing-masing model dengan lebih akurat.

### 3.2. Metode Penelitian

Pada penelitian ini, metode penelitian yang digunakan dalam eksperimen ini menggunakan salah satu metodologi standar dalam penelitian *Data Mining* yaitu model *Cross-Standard Industry for Data Mining* (CRISP-DM) yang terdiri dari 6

tahapan diantaranya adalah tahapan *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modelling*, *Evaluation* dan *Deployment*. Namun pada penelitian ini tidak diimplementasikan sampai dengan *Deployment*, melainkan proses *Data Understanding* hingga *Evaluation*. Berikut ini merupakan tahapan model penelitian yang diusulkan menggunakan metode CRISP-DM terlihat pada Gambar 3.1 sebagai berikut:



**Gambar 3.1** Metode Penelitian

Adapun penjelasan dari masing-masing tahapan dari Gambar 3.1 diatas adalah sebagai berikut :

### **1. Pemahaman Bisnis (*Business Understanding*)**

Tahapan ini merupakan proses pertama dari CRISP-DM dimana pada tahap ini terdiri dari menentukan latar belakang penulisan, mendefinisikan masalah penelitian, tujuan serta manfaat melakukan tinjauan pustaka dan tinjauan studi. Adapun dalam latar belakang menjelaskan tentang tujuan alasan memilih topik penelitian tersebut, selanjutnya memahami topik tersebut dan menghasilkan solusi yang tepat terkait topik masalah serta menjelaskan hal yang menjadi perhatian peneliti dan harapan peneliti yang akan datang terkait topik penelitian yang dilakukan. Pada penelitian ini topik penelitian yang diambil adalah menganalisis dan membandingkan model prediksi mana yang baik digunakan dalam memprediksi harga saham pada tiga data sektor industri di Indonesia menggunakan dataset dari data historis bersifat *time series* yang terdiri dari data harga saham sebelumnya.

### **2. Pemahaman Data (*Data Understanding*)**

Pada tahap data *understanding*, data yang akan diprediksi perlulah untuk dipahami sebelum penelitian dilakukan. Pada tahap ini penelitian bertujuan untuk mengumpulkan, mengidentifikasi serta memahami data yang dimiliki serta data yang nantinya akan digunakan dalam penelitian. Adapun dataset yang digunakan pada penelitian ini merupakan data historis harga saham yang didapat dari *website* <https://finance.yahoo.com> terdiri dari historis harga saham PT. Indofood Tbk, PT. Kimia Farma Tbk, dan PT. Telkom Indonesia Tbk, setelah data didapatkan maka akan dilakukannya pengurutan data dari data terlama ke data yang terbaru dan melakukan seleksi atribut dengan membuang atribut yang dianggap tidak diperlukan.

### **3. Persiapan Data (*Data Preparation*)**

Pada tahapan ini, setelah data yang akan digunakan sudah dipahami, maka akan dilakukan pengolahan data agar hasil yang diinginkan dapat tercapai. Terdapat beberapa tahap pengolahan data dimana langkah-langkah yang dilakukan adalah sebagai berikut :

a. *Data Cleaning*

*Cleaning* data dilakukan untuk menghilangkan *record* pada atribut yang tidak lengkap, tidak konsisten dan tidak rapi yang dapat mengakibatkan proses algoritma tidak sesuai serta dapat mengakibatkan program berjalan sangat lama. Adapun dalam penelitian ini menggunakan teknik *missing value* untuk proses *cleaning* data.

b. *Data Transformation*

Pada langkah data *transformation*, dilakukannya normalisasi data yang merupakan salah satu teknik yang dilakukan pada data *transformation*. Tujuan dilakukannya normalisasi adalah agar proses pengolahan data dapat berjalan dengan cepat dan lebih akurat. Pada penelitian ini akan melakukan normalisasi dengan metode *Min-Max Normalization*.

c. Pembagian Data

Pada pembagian data untuk penelitian ini dibagi menjadi dua bagian, yaitu data *training* dan data *testing*. Dan dari data saham per hari yang dikumpulkan selama lima tahun diperoleh 3789 *record* data. Adapun dalam pembagian datanya digunakan pembagian 80 : 20 menggunakan teknik *linear sampling*.

#### **4. Pemodelan (*Modelling*)**

Pada tahap ini juga dapat disebut sebagai tahap *learning*, dikarenakan pada tahap ini merupakan gambaran dari rangkaian kegiatan penelitian yang dilakukan untuk melakukan suatu pelatihan dan pengujian terhadap data yang sudah dikumpulkan dan diolah yang nantinya akan digunakan oleh peneliti. Adapun pada tahap ini, secara garis besar terdapat dua tahapan yang akan dilakukan yaitu :

- a. Pertama, tahapan pada pembuatan rancangan metode yang diusulkan, dalam kasus ini metode yang akan diusulkan adalah *Linear Regression*, *Random Forest Regression* dan *Multilayer Perceptron*.

- b. Tahapan kedua adalah akan dilakukannya proses eksperimen atau pengujian model yang diusulkan tersebut dengan menggunakan aplikasi KNIME Vers. 4.1.3.

## 5. Evaluasi (*Evaluation*)

Adapun langkah selanjutnya adalah evaluasi terhadap uji coba. Setelah diperoleh hasil pemodelan data *training*, maka akan dilakukan evaluasi terhadap model yang telah dilakukan. Pada proses evaluasi akan dilakukan validasi dengan menggunakan *Cross Validation*, dengan menguji model yang terbentuk menggunakan *10 K-Fold Cross Validation*. Untuk pengukuran akurasi (*Measurements*) akan menggunakan *Root Mean Squared Error* (RMSE),  $R^2$ , *Mean Absolute Percentage Error* (MAPE) dan *Mean Absolute Error* (MAE) dengan menguji model prediksi yang dianggap paling optimal yang terbentuk dengan algoritma *Linear Regression*, *Random Forest Regression* dan *Multilayer Perceptron*.

### 3.3. Sumber Data Penelitian

Pada penelitian ini, sumber data yang digunakan oleh peneliti adalah menggunakan data sekunder berupa data bersifat *time series* dari historis harga saham pada tiga perusahaan yang didapat dari website <https://finance.yahoo.com> dengan jumlah periode pengambilan data selama 5 tahun yaitu 01 September 2015 hingga 01 September 2020, adapun detail dataset yang digunakan dapat dilihat pada Gambar 3.2 sebagai berikut :

RINCIAN DATASET - STOCK PRICE PREDICTION						
Periode Dataset Di Ambil (Dikumpulkan) : 01 September 2015 s.d 01 Septembe 2020 (Rentang Waktu 5 Tahun)						
No	Sektor Industri	Nama Perusahaan	Kode Saham	Source Dataset	Atribut	Jumlah Instance
1	Industri Barang Konsumsi	PT Indofood Sukses Makmur Tbk	INDF.JK	<a href="https://finance.yahoo.com/quote/INDF.JK/?p=INDF.JK">https://finance.yahoo.com/quote/INDF.JK/?p=INDF.JK</a>	Open, High, Low, Close	1263
2	Transportasi & Infrastruktur	PT Telekomunikasi Indonesia Tbk	TLKM.JK	<a href="https://finance.yahoo.com/quote/TLKM.JK/">https://finance.yahoo.com/quote/TLKM.JK/</a>	Open, High, Low, Close	1263
3	Industri Bahan Dasar & Kimia	PT Kimia Farma Tbk	KAEF.JK	<a href="https://finance.yahoo.com/quote/KAEF.JK/history?p=KAEF.JK">https://finance.yahoo.com/quote/KAEF.JK/history?p=KAEF.JK</a>	Open, High, Low, Close	1263
<b>Total Instance Seluruh Dataset</b>						3789

**Gambar 3.2** Detail Sumber Dataset

### 3.4. Pengolahan Data Awal

Pada tahap pengolahan data awal atau data *input*, dataset awal merupakan data *time series* historis harga saham dengan sifat nilai *number* yang merupakan nilai harga per lembar saham, harga saham yang menjadi data *input* pada penelitian ini adalah harga pembuka (*open*), harga tertinggi (*high*), harga terendah (*low*) sebagai prediktornya dan harga penutup (*closing*) sebagai *classnya*. Data *input* ini pun terdiri dari tiga data harga saham perusahaan dalam sektor industri di Indonesia yang berbeda. Setelah data ini dikumpulkan maka dilakukan pengecekan data awal dengan proses *preprocessing* guna membersihkan data harga saham yang memiliki isi tidak sempurna seperti data yang hilang atau data yang tidak valid, pada penelitian ini, tahap ini disebut sebagai proses *missing value* data. Selain itu, pada tahap *preprocessing* ini dilakukan juga normalisasi data dari masing-masing historis harga saham, dimana normalisasi ini dilakukan dengan menggunakan *tools* KNIME, normalisasi sendiri digunakan atau diimplementasikan agar menjadi data yang bernilai sama yaitu dengan *range* 0 dan 1, hal ini dilakukan supaya pada proses *modelling* dengan data *training* dan *testing* dapat lebih efisien dan efektif apabila data-data yang diuji coba masuk

berada pada satu *range* yang sama. Adapun pada penelitian ini menggunakan *Min-Max Normalization* sebagai proses transformasi nya. Maka setelah dilakukannya proses pengecekan dengan *missing value* pada KNIME dan juga proses transformasi berupa normalisasi, selanjutnya data dapat dibagi menjadi dua bagian yaitu data *training* dan data *testing* dengan jumlah pembagian masing-masing data 80 : 20.

### **3.5. Metode Yang Diusulkan**

Adapun pada penelitian ini, metode yang diusulkan dalam analisa perbandingan model prediksi harga saham adalah menggunakan tiga jenis model prediksi diantaranya *Linear Regression (LR)*, *Random Forest Regression (RFR)* dan *Multilayer Perceptron (MLP)*. Pada pengujian dataset dengan ketiga model prediksi tersebut, maka dilakukan perbandingan yang dilihat dari hasil masing-masing yang diperoleh oleh ketiga model tersebut dengan tolak ukur nilai *error* atau tingkat nilai *error* terendah yang dihasilkan oleh ketiga model tersebut, sehingga diketahui model prediksi mana yang lebih baik digunakan dalam memprediksi harga saham dari ketiga model prediksi tersebut.

### **3.6. Eksperimen Dan Hasil Pengujian**

Pada proses eksperimen dari penelitian ini dilakukan menggunakan *tools* KNIME versi 4.1.3, dan pada pengujian model prediksi dilakukan menggunakan tiga dataset historis harga saham bersifat *time series* yang terdiri dari data harga saham PT. Indofood, PT. Kimia Farma dan PT. Telkom dari *website* Yahoo Finance. Adapun pengujian model prediksi dilakukan dengan dua eksperimen diantaranya yaitu pertama uji coba dataset dengan validasi *partitioning* (80 : 20) disertai *Cross Validation* pada pengujian disetiap ketiga model prediksi yang diusulkan, dan kedua uji coba dilakukan dengan model validasi *Cross Validation* tanpa *partitioning* data.

### **3.7. Evaluasi Dan Hasil Validasi**

Selanjutnya dalam tahapan ini pada pelaksanaan penelitian menganalisis data historis harga saham dari tiga data perusahaan yang berbeda menggunakan metode atau model prediksi LR, RFR dan MLP. Maka dalam evaluasi penelitian ini menggunakan uji validasi yaitu *cross-validation* yang digunakan untuk menguji keakuratan metode atau model prediksi. Selanjutnya proses yang dilakukan setelah proses pengujian yaitu mengukur *performance* dari masing-masing model prediksi yang diusulkan. Adapun dalam penelitian ini sebagai tolak ukur hasil *performance* yang baik, dilakukan dengan melihat hasil dari tingkat nilai *error* yang dihasilkan pada setiap model, diantaranya yang utama dilihat pada nilai *error* RMSE, selanjutnya barulah melihat nilai MAE, MAPE serta  $R^2$  yang menghasilkan nilai *error* terkecil, serta akan ditampilkannya *line plot* dari hasil masing-masing model prediksi yang diusulkan.

## **BAB IV**

### **PEMBAHASAN DAN HASIL**

Dalam penelitian ini, detail data *training* serta data *testing* yang digunakan pada saat pengujian didapatkan dari historis harga saham sebelumnya pada tiga sektor industri di Indonesia melalui *website Yahoo Finance*. Pada *website* tersebut, data yang dapat dikumpulkan atau diambil guna penelitian disediakan sebanyak tujuh atribut diantaranya adalah *date*, *open price*, *high price*, *low*, *close price*, *adj.close* dan *volume*. Namun dari tujuh atribut tersebut, atribut yang paling utama mempengaruhi *class (close price)* dalam penelitian ini adalah *open price*, *high price*, dan *low price* yang sekaligus dijadikan atribut prediktornya. Adapun data yang diambil sebanyak 3.789 *record* dari jumlah ketiga emiten yang diteliti. Pembagian antara data *training* dan *testing* dilakukan dengan teknik *linear sampling* dengan jumlah masing-masing 80 : 20. Kemudian pada data dilakukan *training* dataset sehingga didapat *accuracy* berupa nilai terendah *error*nya beserta gambaran *line plot* prediksi dari masing-masing metode yang diusulkan. Berikut penjelasan lebih rinci terkait pembahasan penelitian serta hasil penelitian yang diperoleh oleh penulis, sebagai berikut :

#### **4.1. *Business Understanding***

Tahapan pemahaman bisnis berfokus pada pemahaman tujuan kebutuhan berdasarkan penilaian bisnis yang kemudian pemahaman tersebut diubah menjadi sebuah plan atau rencana awal *data mining* yang dirancang untuk mencapai tujuan tersebut. Adapun tujuan bisnis pada penelitian ini adalah dilakukannya penelitian dengan mengimplementasikan beberapa model prediksi dalam mengenali pola historis harga saham untuk memprediksi kedepannya harga saham berdasarkan model prediksi yang diusulkan. Pada proses pemahaman penelitian ini dilakukan dengan mencari lebih detail mengenai data dan informasi melalui data historis harga saham di *website yahoo finance* dan juga pendekatan literatur dengan sumber artikel mengenai saham yang cukup diminati banyak investor. Adapun pemahaman model penelitian pada saat proses pengolahan data yaitu pendekatan regresi dengan model prediksi terbaik yang akan diujikan, dengan cara

membandingkan hasil dari algoritma dari masing-masing model prediksi tersebut. Prediksi dilakukan dengan menggunakan metode *Linear Regression* (LR), *Random Forest Regression* (RFR) dan *Multilayer Perceptron* (MLP).

#### 4.2. Data Understanding

Tahapan pemahaman data pada penelitian ini dimulai dengan pengumpulan data awal dan hasil kegiatan yang dilakukan dalam rangka untuk membiasakan diri dengan data yang akan digunakan dalam penelitian untuk mengidentifikasi mengenai masalah data, supaya nantinya dapat menentukan wawasan pertama ke data atau dapat mendeteksi *subset* menarik pada informasi yang tersembunyi. Adapun rincian pemahaman data yang dilakukan oleh peneliti pada penelitian ini adalah sebagai berikut:

##### 4.2.1. Pengumpulan Data Awal (*Collect The Initial Data*)

Pengumpulan data awal dilakukan dengan pengambilan data mentah bersifat data *time series* berupa historis harga saham dari tiga jenis emiten berbeda. Hal ini dilakukan karena masing-masing data harga setiap emiten berbeda, maka diambil tiga data harga saham sebagai sampel uji coba yang dilakukan pada penelitian ini, hal ini dapat menambah keakuratan model prediksi yang nantinya diuji cobakan pada masing-masing data harga saham emiten tersebut. Data harga saham yang bisa didapatkan di *website Yahoo Finance* diantaranya data *date*, *open price*, *high price*, *low price* dan *close price*. Adapun dataset yang dikumpulkan dari masing-masing emiten yang digunakan dalam penelitian ini sebagai berikut:

Tabel. 4.1 Tabel Daftar Historis Harga Saham PT. Indofood, Tbk

Date	Open	High	Low	Close
2015-09-01	5300	5300	5125	5200
2015-09-02	5125	5250	5075	5200
2015-09-03	5275	5350	5175	5325
2015-09-04	5400	5400	5250	5275

2015-09-07	5175	5200	4925	4940
2015-09-08	4850	4955	4845	4900
2015-09-09	4950	5000	4895	4965
2015-09-10	4900	5000	4890	4960
2015-09-11	5000	5200	4960	5175
2015-09-14	5200	5325	5150	5325
2015-09-15	5250	5325	5175	5300
2015-09-16	5325	5375	5150	5250
2015-09-17	5300	5350	5150	5350
2015-09-18	5350	5575	5225	5225
2015-09-21	5175	5325	5150	5225
2015-09-22	5300	5325	5200	5300
2015-09-23	5250	5250	5100	5175
2015-09-25	5075	5225	5050	5150
2015-09-28	5000	5150	4960	5100
2015-09-29	5125	5400	4990	5400
2015-09-30	5325	5525	5325	5500
2015-10-01	5525	5625	5425	5550
2015-10-02	5400	5550	5275	5275
2015-10-05	5325	5550	5325	5525
2015-10-06	5450	5625	5425	5525
2015-10-07	5525	5650	5500	5575
2015-10-08	5575	5650	5550	5575
2015-10-09	5675	5800	5625	5800
2015-10-12	5850	5900	5750	5850
2015-10-13	5850	5875	5500	5500
2015-10-15	5625	5925	5600	5900
2015-10-16	5900	6200	5750	6200
2015-10-19	6200	6400	6150	6400
Dst....	.....	.....	.....	.....

Tabel. 4.2 Tabel Daftar Historis Harga Saham PT. Kimia Farma, Tbk

<b>Date</b>	<b>Open</b>	<b>High</b>	<b>Low</b>	<b>Close</b>
2015-09-01	705	765	700	730
2015-09-02	720	730	690	725
2015-09-03	725	730	710	715
2015-09-04	710	720	695	705
2015-09-07	680	695	670	695
2015-09-08	685	700	670	690
2015-09-09	690	710	690	705
2015-09-10	695	710	685	690
2015-09-11	690	700	645	665
2015-09-14	665	675	660	665
2015-09-15	665	680	660	665
2015-09-16	670	675	660	675
2015-09-17	670	710	670	710
2015-09-18	720	745	685	700
2015-09-21	690	700	660	690
2015-09-22	685	700	670	675
2015-09-23	675	675	645	655
2015-09-25	655	675	645	650
2015-09-28	655	655	615	615
2015-09-29	610	640	590	630
2015-09-30	635	650	625	640
2015-10-01	635	660	620	640

2015-10-02	645	645	625	625
2015-10-05	635	685	635	665
2015-10-06	680	695	660	660
2015-10-07	665	740	660	735
2015-10-08	740	795	720	735
2015-10-09	750	780	750	775
2015-10-12	780	815	780	815
2015-10-13	820	840	810	825
2015-10-15	825	835	810	820
2015-10-16	820	860	810	855
2015-10-19	850	905	845	900
dst...	.....	.....	.....	.....

Tabel. 4.3 Tabel Daftar Historis Harga Saham PT. Telkom Indonesia, Tbk

<b>Date</b>	<b>Open</b>	<b>High</b>	<b>Low</b>	<b>Close</b>
2015-09-01	2865	2875	2830	2830
2015-09-02	2790	2810	2765	2775
2015-09-03	2805	2820	2785	2800
2015-09-04	2800	2820	2800	2810
2015-09-07	2760	2765	2695	2700
2015-09-08	2680	2740	2680	2730
2015-09-09	2750	2755	2730	2730
2015-09-10	2720	2785	2705	2765
2015-09-11	2800	2800	2760	2760
2015-09-14	2800	2800	2765	2785
2015-09-15	2755	2770	2710	2710
2015-09-16	2720	2745	2700	2720
2015-09-17	2700	2710	2655	2680
2015-09-18	2690	2720	2675	2690
2015-09-21	2690	2735	2680	2730

2015-09-22	2735	2745	2695	2700
2015-09-23	2665	2700	2655	2655
2015-09-25	2720	2775	2640	2655
2015-09-28	2620	2645	2600	2600
2015-09-29	2550	2605	2485	2605
2015-09-30	2580	2645	2565	2645
2015-10-01	2640	2665	2610	2630
2015-10-02	2605	2635	2600	2620
2015-10-05	2620	2755	2615	2720
2015-10-06	2770	2830	2735	2790
2015-10-07	2775	2790	2700	2740
2015-10-08	2735	2745	2710	2710
2015-10-09	2730	2750	2695	2700
2015-10-12	2695	2720	2680	2695
2015-10-13	2695	2700	2635	2640
2015-10-15	2660	2730	2660	2685
2015-10-16	2725	2725	2675	2680
2015-10-19	2700	2730	2685	2720
Dst....	.....	.....	.....	.....

Pada ketiga tabel di atas (Tabel 4.1, Tabel 4.2, dan Tabel 4.3) dapat dilihat dari masing-masing historis harga saham setiap emiten memiliki *range* harga berbeda-beda baik harga *open*, *high*, *low* serta harga *closenya*. Data yang diambil merupakan harga saham dengan satuan rupiah, jumlah dari masing masing harga saham per emiten dengan *range* 01 september 2015 hingga 01 September 2020 adalah 1263 per emiten dengan total keseluruhan *instance* sebesar 3.789 data.

#### 4.2.2. Mendeskripsikan Data (*Describe The Data*)

Adapun deskripsi dari data yang telah didapat adalah dapat dilihat pada Tabel 4.4 sebagai berikut :

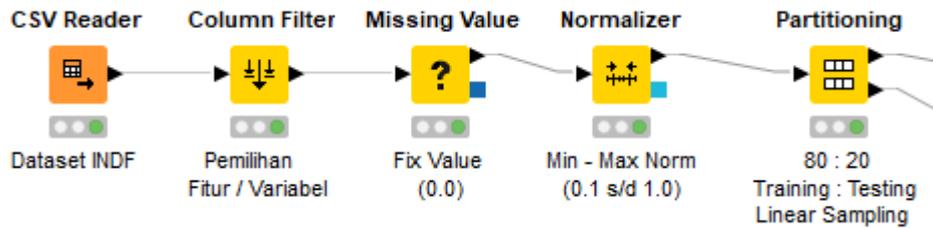
Tabel 4.4. Deskripsi Data

No.	Atribut	Deskripsi	Tipe Data
1	<i>Date</i>	Tanggal terjadinya transaksi jual beli saham, format (yyyy-mm-dd)	<i>Date</i>
2	<i>Open</i>	Harga pembukaan saham pada saat transaksi dimulai dalam periode hari yang bersangkutan.	<i>Numeric</i>
3	<i>High</i>	Harga tertinggi yang dicapai suatu saham pada periode hari yang bersangkutan.	<i>Numeric</i>
4	<i>Low</i>	Harga terendah yang dicapai suatu saham pada periode hari yang bersangkutan.	<i>Numeric</i>
5	<i>Close</i>	harga transaksi terakhir suatu saham pada periode hari yang bersangkutan (biasanya disebut dengan Last)	<i>Numeric</i>

Pada Tabel 4.4 menjelaskan dalam setiap atributnya memiliki data harga dengan fungsi yang berbeda, adapun catatan pada harga *open* bahwa harga *open* tidak harus sama dengan harga terakhir (*close* atau *last*) pada hari sebelumnya, walaupun terkadang terdapat harga *open* sama seperti harga *close* di hari sebelumnya.

### 4.3. Data Preparation

Pada tahap data *preparation* ini merupakan tahap dengan proses penyiapan atau dilakukannya pengolahan data yang bertujuan untuk mendapatkan data yang bersih dan siap untuk digunakan dalam penelitian. Data awal yang diperoleh kemudian dilakukan proses *cleansing* sebagai proses pembersihan data jika terdapat data yang tidak relevan yang mempengaruhi pemodelan dalam eksperimen nantinya, lalu transformasi data dan pembagian data menjadi data *training* dan data *testing*. Adapun *tools* KNIME digunakan dalam melakukan pengolahan data awal yang telah diperoleh, selanjutnya proses data *preparation* dari masing-masing proses di atas dapat dilihat seperti pada Gambar 4.1 berikut :



**Gambar 4.1** Proses *Preprocessing* Dataset

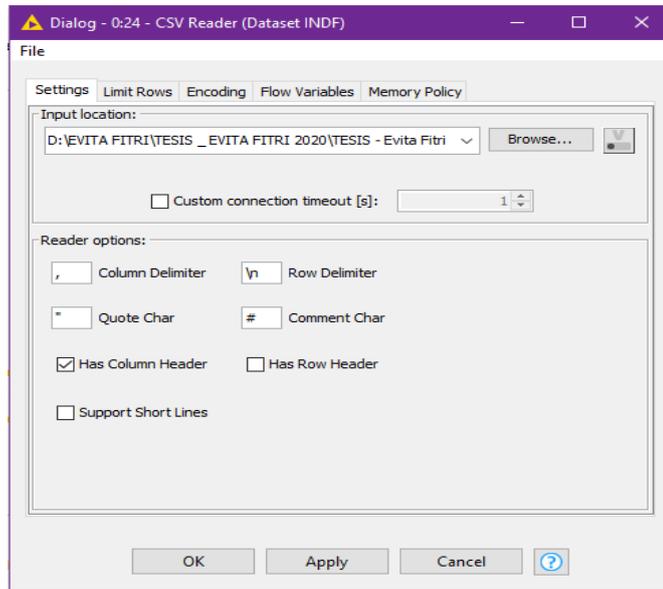
Pada Gambar 4.1 di atas menunjukkan desain model *preprocessing* data menggunakan beberapa node pada tools KNIME. Adapun tahapan yang dilakukan adalah menggunakan node diantaranya *file reader*, *column filtering*, *missing value*, *normalization* dan *partitioning*. Berikut adalah penjelasan dari setiap tahap data *preparation* atau *preprocessing* sebagai berikut :

#### 4.3.1. Data Cleaning

Pada tahap ini dilakukannya terlebih dahulu proses pembacaan data dengan node *file reader* dan pemilihan kolom atau fitur (atribut) dengan node *column filter* yang nantinya atribut tersebut akan digunakan dalam proses *modelling*, adapun dalam pengimplementasian metode atau model prediksi yang digunakan dalam penelitian ini hanya menggunakan empat atribut. Berikut merupakan penjelasan dari masing-masing proses pembersihan data :

##### 1. File Reader

Sebelum melakukan data *preprocessing*, maka langkah pertama yaitu pembacaan *file* dataset. Pada penelitian ini menggunakan dataset dengan format CSV. Node yang digunakan adalah *CSV Reader*, untuk konfigurasi dari node ini dimulai pada pencarian lokasi *file* dan penyesuaian seperti Gambar 4.2 sebagai berikut:



**Gambar 4.2** Konfigurasi CSV Reader

Pada Gambar 4.2 untuk *input location* disesuaikan dengan lokasi dataset yang ingin digunakan untuk uji coba, setelah ditentukan *file* dataset maka mensetting *file reader* pada bagian *reader option*, pada data ini menggunakan pilihan *Has Column Header* yang nantinya menggunakan nama kolom sesuai dengan *file* data CSVnya. Adapun hasil dari proses ini salah satunya dapat dilihat pada Gambar 4.3 data INDF sebagai berikut:

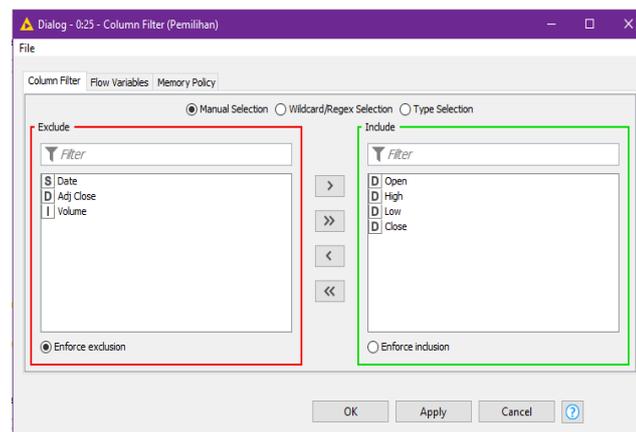
Row ID	S Date	D Open	D High	D Low	D Close	D Adj Close	I Volume
Row0	2015-09-01	5,300	5,300	5,125	5,200	4,414.674	4807200
Row1	2015-09-02	5,125	5,250	5,075	5,200	4,414.674	7089400
Row2	2015-09-03	5,275	5,350	5,175	5,325	4,520.797	8009900
Row3	2015-09-04	5,400	5,400	5,250	5,275	4,478.348	8186400
Row4	2015-09-07	5,175	5,200	4,925	4,940	4,193.941	8962600
Row5	2015-09-08	4,850	4,955	4,845	4,900	4,159.982	8962600
Row6	2015-09-09	4,950	5,000	4,895	4,865	4,215.165	9367300
Row7	2015-09-10	4,900	5,000	4,890	4,860	4,210.921	10809200
Row8	2015-09-11	5,000	5,200	4,960	5,175	4,393.451	6033400
Row9	2015-09-14	5,200	5,325	5,150	5,325	4,520.797	6211100
Row10	2015-09-15	5,250	5,325	5,175	5,300	4,499.572	5492000
Row11	2015-09-16	5,325	5,375	5,150	5,250	4,457.124	3758400
Row12	2015-09-17	5,300	5,350	5,150	5,350	4,542.021	4732900
Row13	2015-09-18	5,350	5,575	5,225	5,225	4,435.899	9971100
Row14	2015-09-21	5,175	5,325	5,150	5,225	4,435.899	3723500
Row15	2015-09-22	5,300	5,325	5,200	5,300	4,499.572	3339600
Row16	2015-09-23	5,250	5,250	5,100	5,175	4,393.451	5605100
Row17	2015-09-25	5,075	5,225	5,050	5,150	4,372.226	4787500
Row18	2015-09-28	5,000	5,150	4,960	5,100	4,329.777	5260300
Row19	2015-09-29	5,125	5,400	4,990	5,400	4,584.47	8548100
Row20	2015-09-30	5,325	5,525	5,325	5,500	4,669.367	11528200
Row21	2015-10-01	5,525	5,625	5,425	5,550	4,711.816	4430700
Row22	2015-10-02	5,400	5,550	5,275	5,275	4,478.348	5846200
Row23	2015-10-05	5,325	5,550	5,325	5,325	4,690.592	7035700
Row24	2015-10-06	5,450	5,625	5,425	5,325	4,690.592	8741700

**Gambar 4.3** Hasil Proses CSV Reader (Dataset INDF)

Pada Gambar 4.3 terdapat data mentah hasil penarikan data dari *Yahoo Finance* dengan tujuh atribut pada tiga *file* dataset harga saham dan nantinya akan ada pemilihan atribut yang selanjutnya akan digunakan dalam pemodelan.

## 2. *Column Filter*

Dalam penggunaan node *column filter* memungkinkan dalam menyaring dari tabel *input* (*file input*) atau berfungsi untuk memfilter kolom atau atribut mana yang akan digunakan dan mana yang tidak digunakan dengan memindahkan kolom di antara daftar *include* atau *exclude*. Pada penelitian ini digunakan 4 atribut dengan konfigurasi *column filter* yang dapat dilihat pada Gambar 4.4 sebagai berikut :



**Gambar 4.4** Konfigurasi *Column Filter*

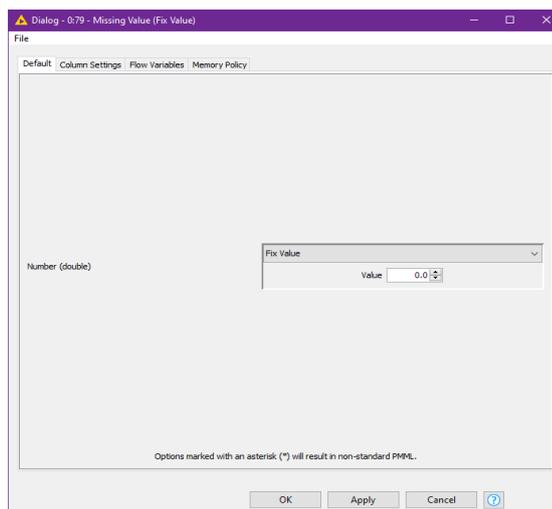
Pada Gambar 4.4 dapat dilihat atribut atau kolom yang digunakan adalah *open*, *high*, *low* dan *close*. Salah satu hasil dari proses *column filter* pada ketiga dataset emiten yang digunakan dapat dilihat pada Gambar 4.5 dengan dataset emiten INDF sebagai berikut :

Row ID	[D] Open	[D] High	[D] Low	[D] Close
Row0	5,300	5,300	5,125	5,200
Row1	5,125	5,250	5,075	5,200
Row2	5,275	5,350	5,175	5,325
Row3	5,400	5,400	5,250	5,275
Row4	5,175	5,200	4,925	4,940
Row5	4,850	4,955	4,845	4,900
Row6	4,950	5,000	4,895	4,965
Row7	4,900	5,000	4,890	4,960
Row8	5,000	5,200	4,960	5,175
Row9	5,200	5,325	5,150	5,325
Row10	5,250	5,325	5,175	5,300
Row11	5,325	5,375	5,150	5,250
Row12	5,300	5,350	5,150	5,350
Row13	5,350	5,575	5,225	5,225
Row14	5,175	5,325	5,150	5,225
Row15	5,300	5,325	5,200	5,300
Row16	5,250	5,250	5,100	5,175
Row17	5,075	5,225	5,050	5,150
Row18	5,000	5,150	4,960	5,100
Row19	5,125	5,400	4,990	5,400
Row20	5,325	5,525	5,325	5,500
Row21	5,525	5,625	5,425	5,550
Row22	5,400	5,550	5,275	5,275
Row23	5,325	5,550	5,325	5,525
Row24	5,450	5,625	5,425	5,525

**Gambar 4.5** Hasil Proses *Column Filter* (Dataset INDF)

### 3. *Missing Value*

Pada node *missing value* digunakan untuk proses pembersihan data, biasanya dalam suatu data terdapat kolom yang tidak sempurna seperti data hilang atau atribut yang tidak relevan, untuk itu node ini diperlukan untuk mengatasi permasalahan tersebut, adapun konfigurasinya dapat dilihat pada Gambar 4.6 sebagai berikut :



**Gambar 4.6** Konfigurasi *Missing Value*

Pada *type number* di node *missing value* menggunakan *Fix Value* dengan *value* 0.0 adapun hasil dari proses pembersihan data pada tiga *file* dataset emiten

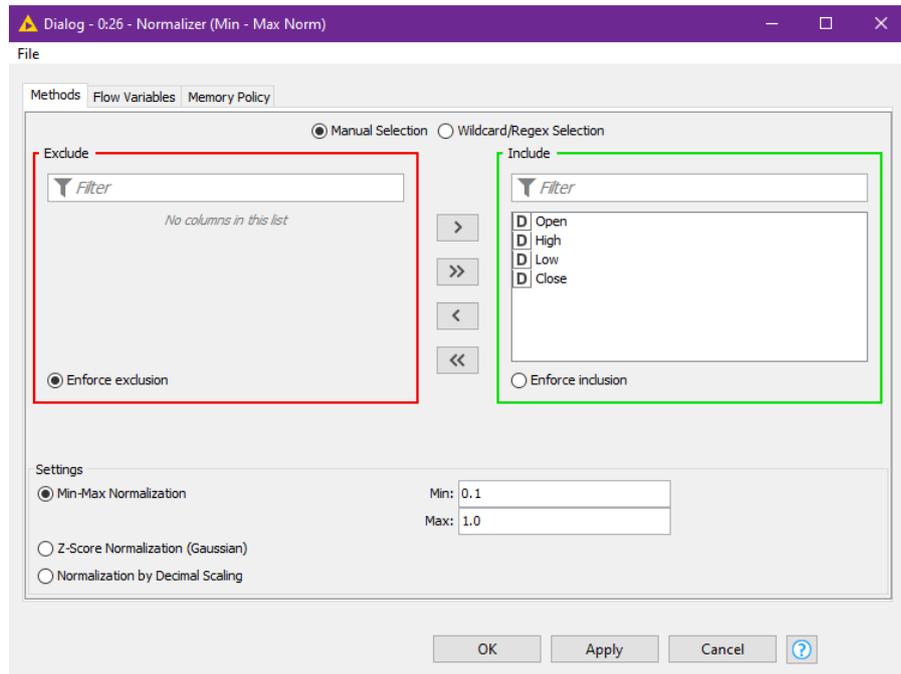
yang digunakan salah satunya dapat dilihat pada Gambar 4.7 dataset INDF sebagai berikut :

Row ID	Open	High	Low	Close
Row0	5,300	5,300	5,125	5,200
Row1	5,125	5,250	5,075	5,200
Row2	5,275	5,350	5,175	5,325
Row3	5,400	5,400	5,250	5,275
Row4	5,175	5,200	4,925	4,940
Row5	4,850	4,955	4,845	4,900
Row6	4,950	5,000	4,895	4,965
Row7	4,900	5,000	4,890	4,960
Row8	5,000	5,200	4,960	5,175
Row9	5,200	5,325	5,150	5,325
Row10	5,250	5,325	5,175	5,300
Row11	5,325	5,375	5,150	5,250
Row12	5,300	5,350	5,150	5,350
Row13	5,350	5,575	5,225	5,225
Row14	5,175	5,325	5,150	5,225
Row15	5,300	5,325	5,200	5,300
Row16	5,250	5,250	5,100	5,175
Row17	5,075	5,225	5,050	5,150
Row18	5,000	5,150	4,960	5,100
Row19	5,125	5,400	4,990	5,400
Row20	5,325	5,525	5,325	5,500
Row21	5,525	5,625	5,425	5,550
Row22	5,400	5,550	5,275	5,275
Row23	5,325	5,550	5,325	5,525
Row24	5,450	5,625	5,425	5,525

**Gambar 4.7** Hasil Proses Pada Node *Missing Value* (Dataset INDF)

#### 4.3.2. Data Transformation

Setelah dilakukan data *preprocessing*, maka selanjutnya akan menuju pada proses data transformasi, pada proses ini data yang telah dibersihkan akan diubah ke format yang sesuai untuk proses *data mining*. Pada proses transformasi ini menggunakan normalisasi data, dimana normalisasi dimaksudkan untuk menyeimbangkan nilai antara fitur-fitur yang ada sehingga tidak ada fitur atau kolom dengan nilai yang terlalu mendominasi. Node yang digunakan pada tahap ini yaitu node *Normalizer*. Adapun konfigurasinya dapat dilihat pada Gambar 4.8 sebagai berikut:



**Gambar 4.8** Konfigurasi Normalisasi Data

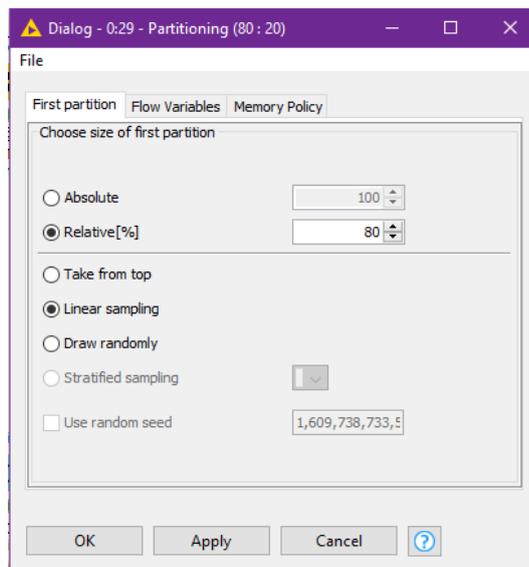
Pada Gambar 4.8 terdapat beberapa pilihan terkait teknik normalisasi yang ingin diimplementasikan diantaranya adalah *Min-Max Normalization*, *Z-Score Normalization (Gaussian)*, dan *Normalization by Decimal Scaling*. Pada penelitian ini digunakan normalisasi dengan tahnik *Min-Max Normalization* dimana *range* nilai terdiri dari 0 – 1, pada nilai kolom minimal ditentukan dengan nilai 0,1 dan nilai maksimal sebesar 1,0. Adapun salah satu hasil dari normalisasi tabel di atas dapat dilihat pada Gambar 4.9 dataset emiten INDF sebagai berikut :

Row ID	Open	High	Low	Close
Row0	0.195	0.176	0.161	0.168
Row1	0.159	0.165	0.15	0.168
Row2	0.19	0.187	0.172	0.195
Row3	0.216	0.197	0.188	0.184
Row4	0.169	0.155	0.118	0.114
Row5	0.102	0.103	0.101	0.105
Row6	0.123	0.113	0.112	0.119
Row7	0.112	0.113	0.111	0.118
Row8	0.133	0.155	0.126	0.163
Row9	0.174	0.181	0.166	0.195
Row10	0.185	0.181	0.172	0.189
Row11	0.2	0.192	0.166	0.179
Row12	0.195	0.187	0.166	0.2
Row13	0.205	0.234	0.182	0.174
Row14	0.169	0.181	0.166	0.174
Row15	0.195	0.181	0.177	0.189
Row16	0.185	0.165	0.156	0.163
Row17	0.149	0.16	0.145	0.158
Row18	0.133	0.144	0.126	0.147
Row19	0.159	0.197	0.132	0.211
Row20	0.2	0.224	0.204	0.232
Row21	0.241	0.245	0.225	0.242
Row22	0.216	0.229	0.193	0.184
Row23	0.2	0.229	0.204	0.237
Row24	0.226	0.245	0.225	0.237

Gambar 4.9 Hasil Proses Normalisasi (Dataset INDF)

### 4.3.3. Pembagian Data (*Partitioning*)

Node *partitioning* dapat memproses pembagian dataset sesuai kebutuhan, pada node ini pembagian dataset dapat dilakukan dengan tiga teknik diantaranya *Take From Top*, *Linear Sampling* atau dengan pembagian data *Draw Randomly*. Pada teknik pengambilan ukuran data (*size*) dapat menggunakan absolut atau relatif [%] sesuai kebutuhan. Adapun pada pembagian dataset penelitian ini dibagi menjadi dua data yaitu data *training* dan data *testing* dengan konfigurasi pada Gambar 4.10 sebagai berikut :



**Gambar 4.10** Konfigurasi Pembagian Data (*Partitioning*)

Pada Gambar 4.10 dapat dilihat konfigurasi pembagian data dilakukan dengan *size relative* [%] sebesar 80 : 20. Data *training* ditetapkan 80 % dan data *testing* ditetapkan 20 % dari jumlah keseluruhan dataset. Pada teknik pembagian datanya pun menggunakan *Linear Sampling* dengan hasil total data pada Tabel 4.5 sebagai berikut :

Tabel 4.5 Jumlah Pembagian Data (*Partitioning*)

No	Data	Jumlah Instance
1	<i>Training</i> (80 %)	1010
2	<i>Testing</i> (20 %)	253

Dari hasil *partitioning* data pada Tabel 4.5 di atas, dapat dilihat salah satu hasil dari node *partitioning* adalah pada dataset emiten INDF dengan rincian 80% data *training* dan 20% data *testing* pada Gambar 4.11 dan Gambar 4.12 sebagai berikut :

Row ID	Open	High	Low	Close
Row0	0.195	0.176	0.161	0.168
Row1	0.159	0.165	0.15	0.168
Row2	0.19	0.187	0.172	0.195
Row3	0.216	0.197	0.188	0.184
Row5	0.102	0.103	0.101	0.105
Row6	0.123	0.113	0.112	0.119
Row7	0.112	0.113	0.111	0.118
Row8	0.133	0.155	0.126	0.163
Row10	0.185	0.181	0.172	0.189
Row11	0.2	0.192	0.166	0.179
Row12	0.195	0.187	0.166	0.2
Row13	0.205	0.234	0.182	0.174
Row15	0.195	0.181	0.177	0.189
Row16	0.185	0.165	0.156	0.163
Row17	0.149	0.16	0.145	0.158
Row18	0.133	0.144	0.126	0.147
Row20	0.2	0.224	0.204	0.232
Row21	0.241	0.245	0.225	0.242
Row22	0.216	0.229	0.193	0.184
Row23	0.2	0.229	0.204	0.237
Row25	0.241	0.25	0.241	0.247
Row26	0.252	0.25	0.252	0.247
Row27	0.272	0.282	0.268	0.295
Row28	0.308	0.303	0.295	0.305
Row30	0.262	0.308	0.262	0.316

**Gambar 4.11** Hasil *First Partitioning* (*Training* INDF)

Second partition (remaining rows) - 0:29 - Partitioning (80 : 20)

File Hilite Navigation View

Table "default" - Rows: 253 Spec - Columns: 4 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close
Row4	0.169	0.155	0.118	0.114
Row9	0.174	0.181	0.166	0.195
Row14	0.169	0.181	0.166	0.174
Row19	0.159	0.197	0.132	0.211
Row24	0.226	0.245	0.225	0.237
Row29	0.308	0.298	0.241	0.232
Row34	0.407	0.414	0.38	0.384
Row39	0.35	0.356	0.359	0.353
Row44	0.288	0.345	0.295	0.358
Row49	0.21	0.192	0.166	0.179
Row54	0.2	0.213	0.193	0.211
Row59	0.205	0.202	0.204	0.2
Row64	0.179	0.171	0.172	0.163
Row69	0.149	0.129	0.15	0.142
Row74	0.131	0.113	0.113	0.124
Row79	0.127	0.113	0.115	0.117
Row84	0.195	0.229	0.198	0.226
Row89	0.252	0.298	0.246	0.289
Row94	0.267	0.298	0.273	0.3
Row99	0.262	0.261	0.246	0.237
Row104	0.396	0.387	0.385	0.389
Row109	0.448	0.509	0.46	0.5
Row114	0.463	0.451	0.46	0.447
Row119	0.515	0.504	0.466	0.474
Row124	0.582	0.625	0.594	0.605

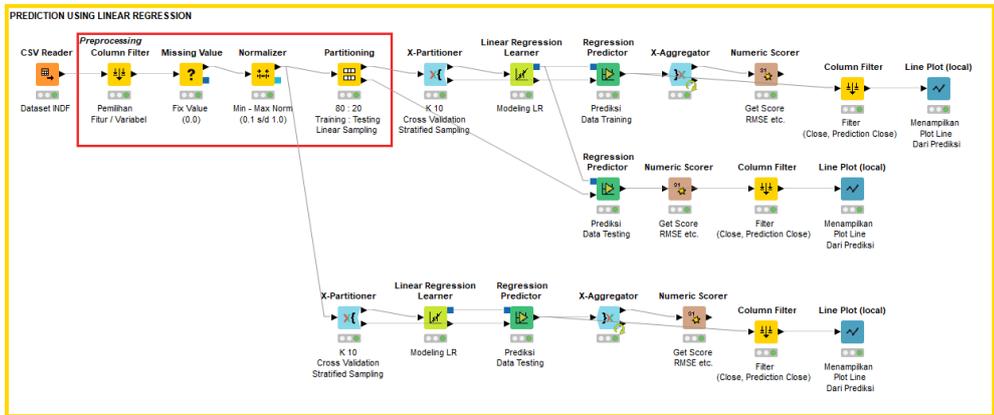
**Gambar 4.12** Hasil *Second Partitioning (Testing INDF)*

#### 4.4. *Modelling*

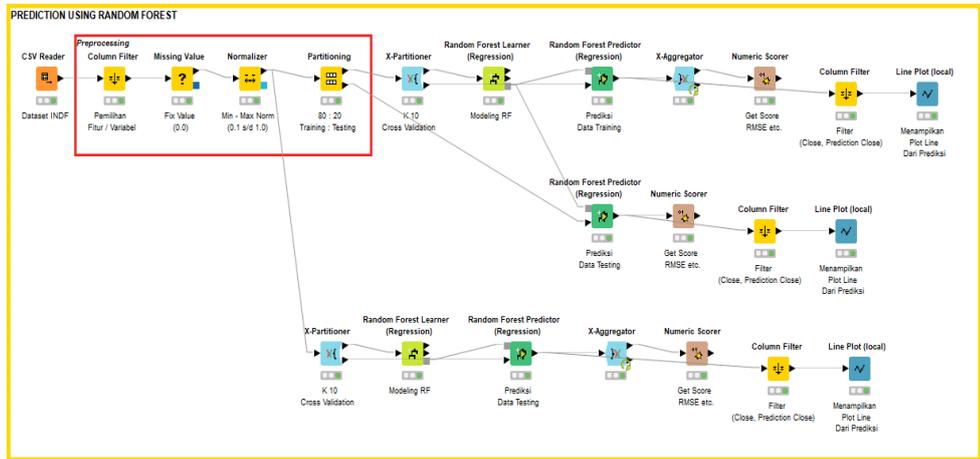
##### 4.4.1. *Design Modelling*

*Modelling* merupakan tahapan *data mining* dalam pemilihan teknik *mining* untuk menentukan metode atau model yang akan digunakan dalam mengolah dataset. Pada tahap *modelling* ini digunakan *tools* KNIME dengan hasil pengujian model yang dilakukan adalah memprediksi dataset historis harga saham pada tiga emiten dengan menggunakan algoritma atau model prediksi LR, RFR dan MLP.

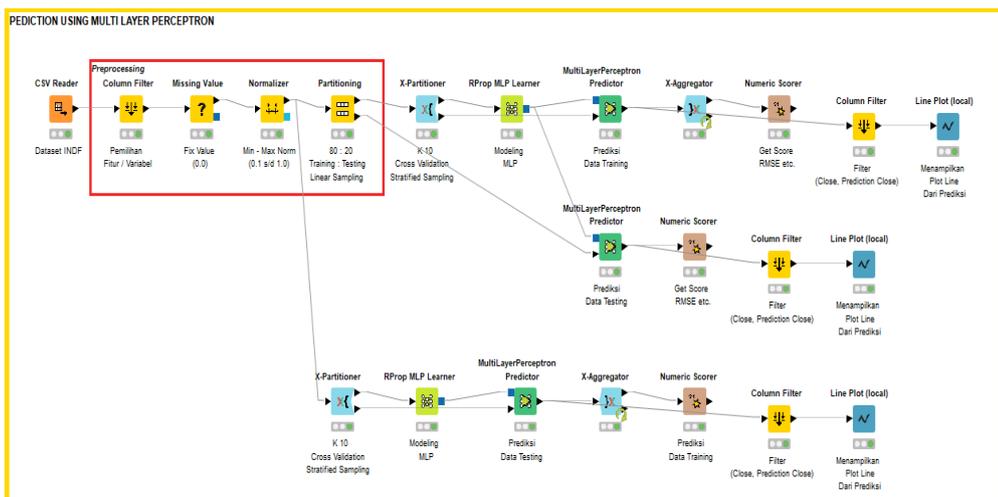
Adapun pengujian tersebut menggunakan model pembagian data *training-testing partitioning* dan *Cross Validation*. Berikut gambar di bawah ini (Gambar 4.12, Gambar 4.13 dan Gambar 4.14) adalah salah satu desain model pengolahan data historis harga saham yang mewakili hasil desain model yang diuji cobakan pada ketiga dataset emiten dengan menggunakan metode LR, RFR dan MLP sebagai berikut :



Gambar 4.13 Model Pengolahan Data Menggunakan *Linear Regression* (INDF)



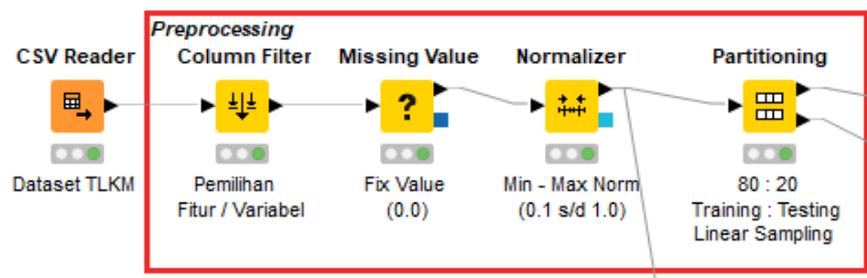
Gambar 4.14 Model Pengolahan Data Menggunakan RFR (INDF)



Gambar 4.15 Model Pengolahan Data Menggunakan MLP (INDF)

Berdasarkan pada Gambar 4.12, Gambar 4.13 dan Gambar 4.13 desain *modelling* di atas merupakan model penelitian yang dibuat penulis untuk pengolahan data historis harga saham dari ketiga emiten diantaranya INDF, KAEF dan TLKM yang didapat dari *yahoo finance*. Tertera dalam desain di atas yaitu menggunakan model prediksi LR, RFR dan MLP untuk menentukan hasil nilai *error prediction* pada setiap model tersebut.

Adapun dalam penelitian ini diawali dengan tahap *CSV reader*, *pre-processing*, *Cross Validation*, serta *partitioning* menggunakan *Cross Validation*, lalu prediksi menggunakan LR (*Linear Regression Learner, Regression Predictor*), RFR (*Random Forest Regression Learner, Random Forest Predictor Regression*) dan MLP (*Rprop MLP Learner, Multilayer Perceptron Predictor*), *numeric scoring*, *column filter* serta *plot line*. Adapun penjelasan dari tahapan pembacaan data sampai dengan *preprocessing* dapat dilihat pada Gambar 4.16 sebagai berikut :



**Gambar 4.16** Tahap Pembacaan Data dan *Preprocessing* Pada *Design Modelling*

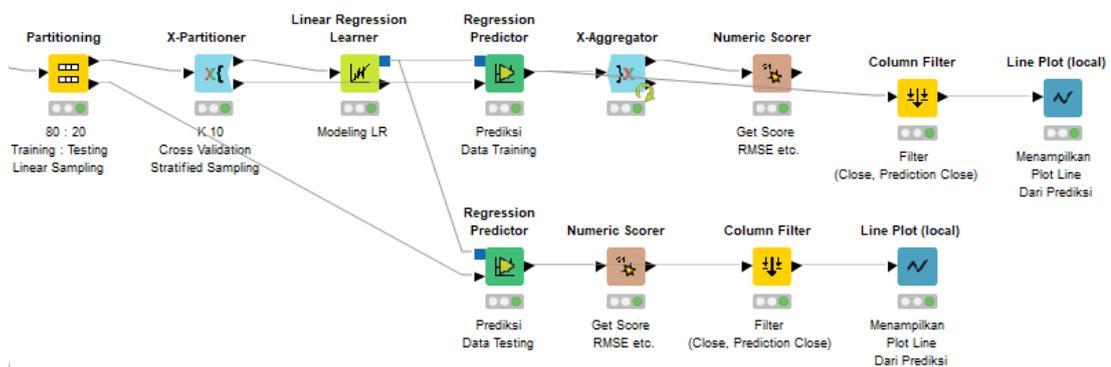
Pada Gambar 4.16 diketahui tahapan di atas merupakan tahapan *data preparation* yang sebelumnya dilakukan, terdiri dari pembacaan file CSV, penyeleksian atribut yang digunakan untuk *modelling* yaitu empat atribut diantaranya *open*, *high*, *low* dan *close*, pengecekan *missing value* atau pembersihan data yang kosong, normalisasi sebagai proses data transformasi yang pada penelitian ini menggunakan teknik *Min-Max Normalization* dengan nilai minimal 0,1 dan nilai maksimal 1,0 serta tahapan pembagian data yaitu data *training* dan data *testing* dengan node *partitioning* menggunakan *size* 80 : 20 dengan teknik pembagian data *Linear Sampling*, sehingga data siap dilanjutkan pada tahap pemodelan dengan tiga model prediksi yaitu RL, RFR dan MLP.

#### 4.4.2. Desain Model *Partitioning* Dengan *Cross Validation*

Pada penelitian ini, uji coba dilakukan dengan dua model uji coba, diantaranya uji coba pertama dengan *partitioning* data *training* dan *testing* sebesar 80 : 20 setelahnya dilakukan proses validasi dengan 10 *K-fold Cross Validation*, dan kedua dilakukan dengan cara pengujian model prediksi tanpa *partitioning*. Adapun berikut merupakan desain pengujian model menggunakan *partitioning* dengan validasi *Cross Validation* menggunakan metode LR, RFR dan MLP dengan *tools* KNIME sebagai berikut:

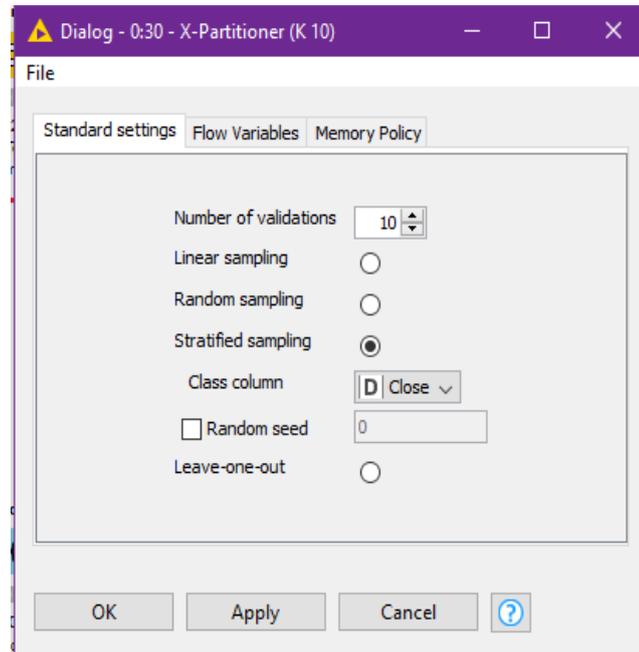
##### 1. Model Prediksi Metode *Linear Regression*

Model pengujian data dengan pembagian data *partitioning* 80 : 20 disertai proses validasi *Cross Validation* pada metode *Linear Regression* dengan dataset tiga emiten dapat dilihat pada Gambar 4.17 sebagai berikut:



**Gambar 4.17** *Partitioning* disertai *Cross Validation* LR (INDF, KAEF, TLKM)

Pada Gambar 4.17 menunjukkan beberapa tahapan diantaranya pertama diawali oleh pembagian data dengan node *partitioner* 80:20 teknik *Linear Sampling* yang menghasilkan data *training* sebanyak 1010 *record* dan data *testing* sebanyak 253 *record*. Pada masing-masing dataset emiten dilakukan proses yang sama dalam uji coba menggunakan *Linear Regression*. Setelah pembagian data dengan *partitioning* selanjutnya validasi data dengan *Cross Validation* (menggunakan node *X-Partitioner*) pada data *training* yang berjumlah jumlah 1010 dengan konfigurasi pada Gambar 4.18 berikut :



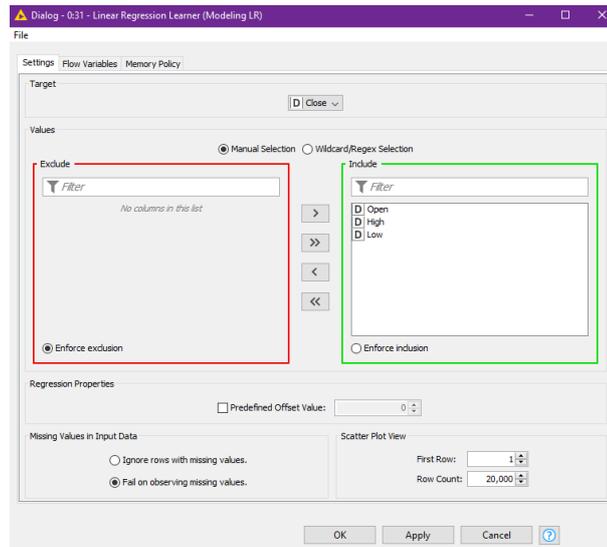
**Gambar 4.18** Konfigurasi *Cross Validation*

Pada Gambar 4.18 menunjukkan proses validasi menggunakan nilai K atau *number of validation* 10 dengan teknik *Stratified Sampling* berdasarkan atribut yang menjadi *class* prediksinya (*close*). Pada proses *Cross Validation* dihasilkan data validasi *training* dan data validasi *testing* yang dapat dilihat pada Tabel 4.6 sebagai berikut :

Tabel 4.6 Hasil Validasi Data *Training* Dengan *Cross Validation* Model LR

NO	<i>Training</i> data (1010) dengan 10 <i>Cross Validation</i>	Jumlah
1	Hasil <i>Cross Validation (Training)</i>	909 <i>Record</i>
2	Hasil <i>Cross Validation (Testing)</i>	101 <i>Record</i>

Dari Tabel 4.6 dapat dilihat setelah data *training* dilakukan proses *Cross Validation* dan menghasilkan data yang nantinya dimodelkan dengan *Linear Regression*. Adapun data yang menjadi masukkan atau data *input* pada node *Linear Regression Learner* adalah data hasil validasi di atas (data *training* 909). Konfigurasi dari proses *learner* LR dapat dilihat pada Gambar 4.19 sebagai berikut :



**Gambar 4.19** Konfigurasi *Linear Regression Learner*

Target atribut dari proses Gambar 4.19 adalah kolom *close* dan sebagai atribut *predictor* yang mempengaruhi kolom *class* pada proses ini adalah *open*, *high*, dan *low*. Proses ini adalah proses pemodelan, dimana data *training* yang sudah divalidasi maka akan dimodelkan dengan node LR *Learner* dan nantinya hasil dari model ini di jadikan model *input* untuk node LR *predictor* dengan data *testing*. Adapun pada proses pemodelan ini menghasilkan *Coefficients and Statistics* seperti pada ketiga tabel di bawah ini:

Tabel 4.7 *Coefficients and Statistic* Dari Model *Learner* LR (INDF)

No	Variabel	<i>Coeff.</i>	<i>Std. Error</i>	<i>t-value</i>
1	Open	-0.553	0.026	-21.193
2	High	0.736	0.021	34.303
3	Low	0.796	0.023	34.476
4	Intercept	0.010	0.001	7.223
5	Multiple R-Square	0.9961		
6	Adjusted R-Square	0.9961		

Tabel 4.8 *Coefficients and Statistic* Dari Model *Learner* LR (KAEF)

No	Variabel	<i>Coeff.</i>	<i>Std. Error</i>	<i>t-value</i>
1	Open	-0.486	0.031	-15.441
2	High	0.628	0.016	38.611

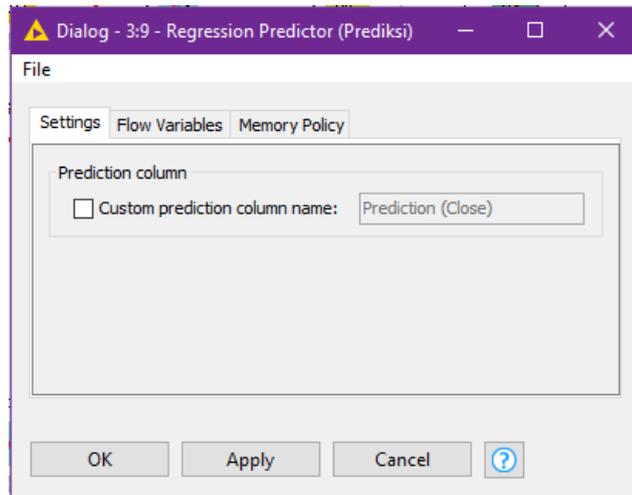
3	Low	0.850	0.027	30.907
4	Intercept	0	0.001	0.501
5	Multiple R-Square	0.9977		
6	Adjusted R-Square	0.9977		

Tabel 4.9 *Coefficients and Statistic* Dari Model *Learner LR* (TLKM)

No	Variabel	<i>Coeff.</i>	<i>Std. Error</i>	<i>t-value</i>
1	Open	-0.544	0.024	-22.316
2	High	0.756	0.020	37.530
3	Low	0.840	0.022	37.472
4	Intercept	-0.044	0.001	-37.432
5	Multiple R-Square	0.9973		
6	Adjusted R-Square	0.9973		

Pada tabel di atas (Tabel 4.7, Tabel 4.8 dan Tabel 4.9) dapat dilihat bahwa masing-masing model LR dari setiap dataset per emiten memiliki hasil *Coefficients* dan *Standard Error* yang berbeda-beda. *Standard Error* pada variabel *open* yang paling kecil ada pada model learner LR dengan dataset TLKM, pada variabel *high* dengan dataset KAEF dan pada variabel *low* nilai *standard error* terkecil dengan dataset TLKM, hal ini dikarenakan masing-masing *range* harga pada setiap variabel dari ketiga emiten berbeda serta dapat dilihat nilai R dari hasil tabel di atas juga berbeda pada masing-masing hasil pemodelan dengan tiga emiten tersebut, namun pada nilai multiple serta adjusted R pada setiap hasil dari masing-masing dataset emiten menghasilkan nilai yang sama.

Selanjutnya setelah pemodelan dengan data *training* yang telah divalidasi maka model tersebut dijadikan sebagai *input* model pada node *Predictor Regression* dengan data *recordnya* menggunakan data *testing* yang sudah divalidasi berjumlah 101 *record* dengan konfigurasi node *Predictor Regression* pada Gambar 4.20 sebagai berikut :



**Gambar 4.20** Konfigurasi Node *Predictor Regression*

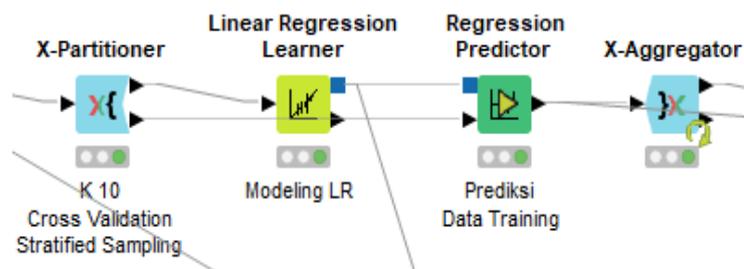
Pada Gambar 4.20 implementasi node di atas konfigurasi model *inputnya* merupakan hasil *modelling* dari node *Linear Regression Learner*, untuk *custom prediction column name* digunakan jika nama kolom prediksi ingin dirubah, dan salah satu hasil dari *predictor* dari proses di atas dapat dilihat pada Gambar 4.21 emiten INDF sebagai berikut :

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)
Row20	0.2	0.224	0.204	0.232	0.226
Row67	0.149	0.144	0.15	0.153	0.154
Row68	0.143	0.139	0.14	0.142	0.144
Row71	0.143	0.129	0.126	0.118	0.125
Row83	0.185	0.234	0.172	0.189	0.217
Row91	0.288	0.276	0.273	0.274	0.272
Row101	0.308	0.324	0.289	0.332	0.308
Row103	0.391	0.387	0.359	0.395	0.364
Row108	0.407	0.477	0.417	0.484	0.469
Row127	0.649	0.646	0.626	0.632	0.625
Row132	0.644	0.641	0.61	0.637	0.611
Row136	0.546	0.567	0.546	0.568	0.56
Row142	0.525	0.588	0.54	0.6	0.582
Row151	0.567	0.562	0.572	0.574	0.566
Row170	0.582	0.577	0.578	0.584	0.573
Row171	0.577	0.572	0.572	0.558	0.568
Row220	0.69	0.773	0.701	0.758	0.755
Row273	0.938	0.947	0.947	0.953	0.942
Row281	0.907	0.91	0.931	0.911	0.919
Row313	0.649	0.725	0.669	0.732	0.717
Row333	0.752	0.762	0.765	0.763	0.764
Row355	0.732	0.741	0.754	0.742	0.751
Row363	0.752	0.762	0.765	0.758	0.764
Row365	0.778	0.789	0.797	0.789	0.794
Row368	0.747	0.778	0.77	0.779	0.783

**Gambar 4.21** Hasil *Predicted Data* pada Node *Predictor Regression* (INDF)

Pada Gambar 4.21 merupakan hasil dari *Predictor Regression* pada dataset emiten INDF, terdapat kolom *open*, *high*, *low*, *close* dan hasil prediksi dengan nama kolom *Prediction (Close)*. Adapun setiap baris memiliki hasil prediksi berbeda beda, hasil nilai pada kolom *prediction close* didapatkan dari target *class (close)* dan variabel *predictor* yang mempengaruhi *class* terdiri dari *open*, *high* dan *low*.

Selanjutnya setelah pemodelan data *training* hasil validasi dengan *Linear Regression* dan prediksi data *testing* hasil validasi menggunakan *Predictor Regression*, maka dilakukannya implementasi dengan node *X-Aggregator* sebagai pelengkap penyelesaian proses *Cross Validation* yang dapat dilihat pada Gambar 4.22 sebagai berikut:



**Gambar 4.22** Penyelesaian Proses *Cross Validation* Menggunakan *X-Aggregator*

Pada Gambar 4.22 di atas, node *X-Aggregator* harus menjadi akhir dari *loop* validasi silang dan harus mengikuti node *X-Partitioner*. Pada proses ini mengumpulkan hasil dari node prediktor, membandingkan *class* yang diprediksi dan *real class* serta memberikan *output* prediksi untuk semua baris dan statistik iterasi. Adapun salah satu hasil dari proses *X-Aggregator* pada data *testing* (emiten INDF) dengan *Cross Validation* (101 *record*) dapat dilihat pada Gambar 4.23 sebagai berikut :

Prediction table - 2:33 - X-agggregator

File Hilite Navigation View

Table "default" - Rows: 1010 Spec - Columns: 5 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)
Row1	0.159	0.165	0.15	0.168	0.164
Row7	0.112	0.113	0.111	0.118	0.119
Row11	0.2	0.192	0.166	0.179	0.173
Row30	0.262	0.308	0.262	0.316	0.301
Row40	0.35	0.34	0.273	0.263	0.285
Row47	0.288	0.282	0.252	0.242	0.259
Row52	0.169	0.155	0.145	0.142	0.146
Row81	0.164	0.15	0.15	0.153	0.149
Row106	0.381	0.419	0.385	0.416	0.415
Row112	0.448	0.451	0.455	0.442	0.456
Row117	0.484	0.504	0.498	0.511	0.509
Row167	0.567	0.583	0.567	0.589	0.577
Row188	0.587	0.62	0.599	0.616	0.618
Row191	0.608	0.599	0.605	0.589	0.595
Row202	0.567	0.556	0.54	0.558	0.536
Row212	0.567	0.583	0.578	0.574	0.585
Row235	0.742	0.81	0.76	0.811	0.801
Row243	0.737	0.736	0.743	0.737	0.735
Row272	0.938	0.947	0.925	0.953	0.925
Row291	0.788	0.794	0.797	0.789	0.792
Row296	0.659	0.699	0.61	0.674	0.646
Row318	0.732	0.741	0.738	0.753	0.738
Row321	0.68	0.694	0.685	0.679	0.689
Row331	0.711	0.715	0.722	0.716	0.717
Row332	0.706	0.746	0.727	0.758	0.748

**Gambar 4.23** Prediction Table Dari Proses X-Aggregator LR (INDF)

Pada Gambar 4.23 menunjukkan hasil untuk setiap baris dari nilai yang diprediksi baik data *training* dan juga data *testing* yang sudah dievaluasi dengan validasi 10 *K-Fold Cross Validation*. Pada proses ini pun ditunjukkan hasil *error rates* pada setiap nilai *K-Fold* nya yang dapat dilihat pada Gambar 4.24 sebagai berikut :

Error rates - 2:33 - X-agggregator

File Hilite Navigation View

Table "default" - Rows: 10 Spec - Columns: 3 Properties Flow Variables

Row ID	D Total squared error	D Mean squared error	I Size of Test Set
fold 0	0.016	0	101
fold 1	0.013	0	101
fold 2	0.018	0	101
fold 3	0.014	0	101
fold 4	0.013	0	101
fold 5	0.015	0	101
fold 6	0.017	0	101
fold 7	0.015	0	101
fold 8	0.019	0	101
fold 9	0.015	0	101

**Gambar 4.24** Error Rate 10 K-Fold Cross Validation Model LR (INDF)

Pada hasil Gambar 4.24 *error rate* emiten INDF bahwa nilai total *squared error* terendah yaitu ada pada *k-fold* 4 sebesar 0.013 dan tertinggi ada pada *k-fold* 8 sebesar 0.019.

Row ID	Total squared error	Mean squared error	Size of Test Set
fold 0	0.009	0	101
fold 1	0.027	0	101
fold 2	0.011	0	101
fold 3	0.011	0	101
fold 4	0.014	0	101
fold 5	0.006	0	101
fold 6	0.013	0	101
fold 7	0.01	0	101
fold 8	0.016	0	101
fold 9	0.015	0	101

**Gambar 4.25** *Error Rate 10 K-Fold Cross Validation Model LR (KAEF)*

Pada hasil Gambar 4.25 *error rate* emiten KAEF bahwa nilai total *squared error* terendah yaitu ada pada *k-fold* 5 sebesar 0.006 dan tertinggi ada pada *k-fold* 1 sebesar 0.027.

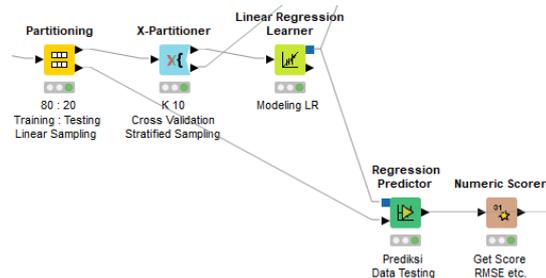
Row ID	Total squared error	Mean squared error	Size of Test Set
fold 0	0.009	0	101
fold 1	0.017	0	101
fold 2	0.007	0	101
fold 3	0.011	0	101
fold 4	0.007	0	101
fold 5	0.011	0	101
fold 6	0.01	0	101
fold 7	0.014	0	101
fold 8	0.012	0	101
fold 9	0.014	0	101

**Gambar 4.26** *Error Rate 10 K-Fold Cross Validation Model LR (TLKM)*

Pada hasil Gambar 4.26 *error rate* emiten TLKM bahwa nilai total *squared error* terendah yaitu ada pada *k-fold* 2 sebesar 0.007 dan tertinggi ada pada *k-fold* 1 sebesar 0.017.

Proses di atas merupakan proses terakhir dalam pemodelan data *training* hasil *partitioning* (1010 *record*) menggunakan evaluasi *Cross Validation* yang selanjutnya akan dilanjutkan dengan proses *scoring* dengan node *Numeric Score* agar mendapatkan hasil berupa nilai RMSE, MAE, MAPE dan  $R^2$  dari nilai prediksi yang didapatkan.

Sedangkan pada data *testing* (253 *record*) dari hasil *partitioning*, dilakukan prediksi langsung menggunakan model *Linear Regression Learner* yang sudah diimplementasikan dengan data *training* hasil proses evaluasi dengan 10 *K-Fold Cross Validation*, dengan menghubungkan node *Linear Regression Learner* sebagai model *input* ke node *Regression Predictor* seperti pada Gambar 4.27 berikut :



**Gambar 4.27** Prediksi Data *Testing* Dari *Partitioning* Dengan Model LR

Pada Gambar 4.27 di atas menunjukkan proses prediksi pada data *testing* hasil *partitioning* dengan jumlah *record* sebanyak 253 data, dan salah satu hasil dari proses di atas dapat dilihat pada Gambar 4.28 (emiten INDF) sebagai berikut :

Predicted data - 2:35 - Regression Predictor (Prediksi)

File Hilite Navigation View

Table "default" - Rows: 253 Spec - Columns: 5 Properties Flow Variables

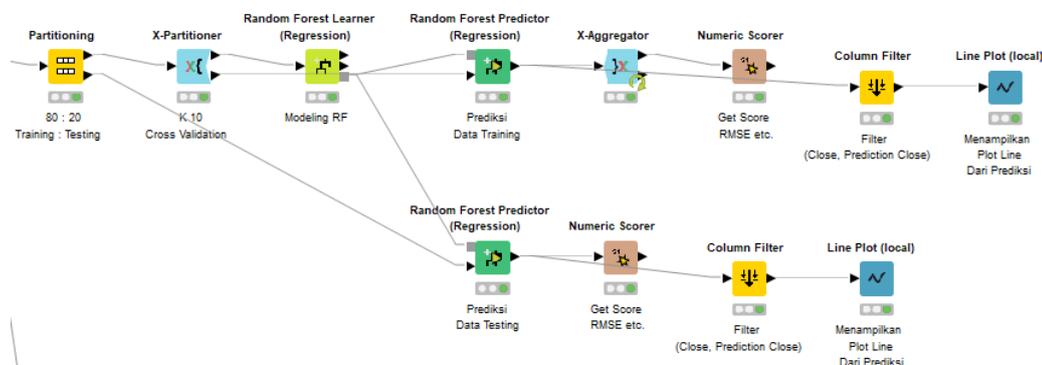
Row ID	D Open	D High	D Low	D Close	D Prediction (Close)
Row4	0.169	0.155	0.118	0.114	0.125
Row9	0.174	0.181	0.166	0.195	0.179
Row14	0.169	0.181	0.166	0.174	0.182
Row19	0.159	0.197	0.132	0.211	0.172
Row24	0.226	0.245	0.225	0.237	0.244
Row29	0.308	0.298	0.241	0.232	0.25
Row34	0.407	0.414	0.38	0.384	0.392
Row39	0.35	0.356	0.359	0.353	0.364
Row44	0.288	0.345	0.295	0.358	0.339
Row49	0.21	0.192	0.166	0.179	0.167
Row54	0.2	0.213	0.193	0.211	0.21
Row59	0.205	0.202	0.204	0.2	0.208
Row64	0.179	0.171	0.172	0.163	0.173
Row69	0.149	0.129	0.15	0.142	0.142
Row74	0.131	0.113	0.113	0.124	0.11
Row79	0.127	0.113	0.115	0.117	0.114
Row84	0.195	0.229	0.198	0.226	0.229
Row89	0.252	0.298	0.246	0.289	0.286
Row94	0.267	0.298	0.273	0.3	0.299
Row99	0.262	0.261	0.246	0.237	0.253
Row104	0.396	0.387	0.385	0.389	0.383
Row109	0.448	0.509	0.46	0.5	0.503
Row114	0.463	0.451	0.46	0.447	0.452
Row119	0.515	0.504	0.466	0.474	0.466
Row124	0.582	0.625	0.594	0.605	0.621

**Gambar 4.28** Predicted Data Testing Hasil Partitioning Model LR (INDF)

Proses pada Gambar 4.28 di atas merupakan proses terakhir dalam prediksi data *testing* hasil *partitioning* (253 record) menggunakan pemodelan *Linear Regression Learner* yang disertai dengan tahap evaluasi *Cross Validation* yang selanjutnya akan dilanjutkan dengan proses *scoring* dengan node *number score* agar mendapatkan hasil berupa nilai RMSE, MAE, MAPE dan  $R^2$  dari prediksi yang dilakukan.

## 2. Model Prediksi Metode *Random Forest Regression*

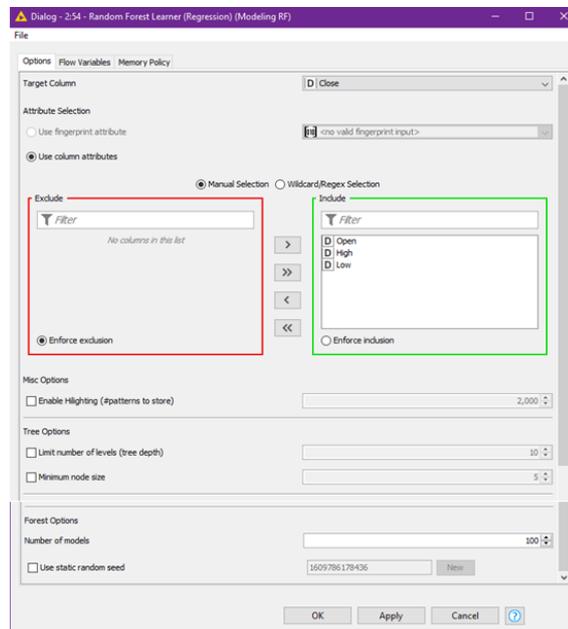
Model pengujian data dengan pembagian data *partitioning* 80 : 20 disertai proses validasi *Cross Validation* pada metode *Random Forest Regression* dengan dataset tiga emiten adalah sebagai berikut:



**Gambar 4.29** Partitioning dengan Cross Validation RFR (INDF, KAEF, TLKM)

Pada Gambar 4.29 di atas sama dengan pemodelan menggunakan *Linear Regression* sebelumnya, ada beberapa tahapan diantaranya diawali oleh pembagian data dengan node *partitioner* 80:20 teknik *Linear Sampling* yang menghasilkan data *training* sebanyak 1010 *record* dan data *testing* sebanyak 253 *record*. Pada masing-masing dataset emiten dilakukan proses yang sama dalam uji coba menggunakan *Random Forest Regression*. Setelah pembagian data dengan *partitioning* selanjutnya validasi data dengan *Cross Validation* pada data *training* yang berjumlah jumlah 1010 dengan konfigurasi yang sama dengan uji coba model regresi sebelumnya. Hasil tabel data dari proses node *X-Partitioner* dapat dilihat pada tabel sebelumnya yaitu Tabel 4.6.

Adapun perbedaan terletak pada pemodelan dengan node *Learnernya* yaitu menggunakan node *Random Forest Learner (Regression)*. Selanjutnya data yang menjadi data *input* pada node *Random Forest Learner (Regression)* adalah data hasil validasi (data *training* 909). Konfigurasi dari proses *learner* RFR dapat dilihat pada Gambar 4.30 sebagai berikut :



**Gambar 4.30** Konfigurasi *Random Forest Regression Learner*

Target kolom dari proses *leaner* RFR pada Gambar 4.30 di atas adalah kolom *close* dan sebagai atribut yang mempengaruhi kolom *class* pada proses ini adalah *open*, *high*, dan *low*. Pada *option forest* dalam pemodelan ini menggunakan

*number of model* sebanyak 100. Hal ini dikarenakan *number of model* pada *option* konfigurasi di atas untuk mempelajari jumlah pohon regresi. Biasanya jumlah model menggunakan nilai yang masuk akal dapat berkisaran dengan jumlah sangat sedikit (katakanlah 10) hingga ribuan, adapun nilai antara 100 dan 500 sudah cukup untuk sebagian besar kumpulan data. Proses ini adalah proses pemodelan, dimana data *training* yang sudah divalidasi lalu dimodelkan dengan node RFR *Learner* dan nantinya hasil dari model ini di jadikan model *input* untuk node RFR *predictor* menggunakan data *testing* hasil *partitioning*. Adapun pada proses pemodelan ini menghasilkan *Attribute Statistic* (emiten INDF) yang dapat dilihat pada Gambar 4.31 di bawah ini:

Row ID	#splits (level 0)	#splits (level 1)	#splits (level 2)	#candidates (level 0)	#candidates (level 1)	#candidates (level 2)
Open	39	68	155	39	68	155
High	36	64	130	36	64	130
Low	25	68	115	25	68	115

**Gambar 4.31** *Attribute Statistic* Hasil Pemodelan RFR *Partitioning* (INDF)

Pada Gambar 4.31 dapat dijelaskan bahwa tabel statistik tentang atribut yang digunakan di RF berbeda-beda. Setiap baris mewakili satu atribut pelatihan dengan statistik berikut: *#splits* (level x) sebagai jumlah model, yang menggunakan atribut sebagai pemisahan pada level x (dengan level 0 sebagai pemisahan *root*), *#candidates* (level x) adalah berapa kali atribut berada dalam sampel atribut untuk level x (dalam pengaturan RF, sampel ini berbeda dari satu node ke node lain). Jika tidak ada *sampling* atribut maka digunakan *#candidates* sebagai jumlah model. Selain hasil *attribute statistics* nya, terdapat pula hasil *Out Of Bag Prediction* pada pemodelan RF yang dapat dilihat pada Gambar 4.32 (emiten INDF) sebagai berikut:

Out-of-bag Predictions - 2:54 - Random Forest Learner (Regression) (Modeling RF)

File Hilit Navigation View

Table "default" - Rows: 909 Spec - Columns: 7 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close	D Close (Out-of-bag)	D Close (Out-of-bag) (Prediction Variance)	I model count
Row0	0.195	0.176	0.161	0.168	0.175	0.001	33
Row1	0.159	0.165	0.15	0.168	0.161	0	44
Row2	0.19	0.187	0.172	0.195	0.193	0	37
Row3	0.216	0.197	0.188	0.184	0.213	0.001	38
Row4	0.169	0.155	0.118	0.114	0.135	0	35
Row5	0.102	0.103	0.101	0.105	0.105	0	36
Row7	0.112	0.113	0.111	0.118	0.125	0	30
Row8	0.133	0.155	0.126	0.163	0.143	0	41
Row9	0.174	0.181	0.166	0.195	0.176	0	33
Row10	0.185	0.181	0.172	0.189	0.186	0	35
Row12	0.195	0.187	0.166	0.2	0.185	0	34
Row14	0.169	0.181	0.166	0.174	0.188	0	35
Row16	0.185	0.165	0.156	0.163	0.163	0	38
Row17	0.149	0.16	0.145	0.158	0.159	0	39
Row18	0.133	0.144	0.126	0.147	0.149	0	41
Row19	0.159	0.197	0.132	0.211	0.17	0	38
Row25	0.241	0.25	0.241	0.247	0.253	0	38
Row26	0.252	0.25	0.252	0.247	0.256	0	40
Row27	0.272	0.282	0.268	0.295	0.266	0	37
Row28	0.308	0.303	0.295	0.305	0.307	0	28
Row30	0.262	0.308	0.262	0.316	0.291	0	30
Row31	0.319	0.366	0.295	0.379	0.341	0.001	42
Row32	0.381	0.408	0.38	0.421	0.397	0	39
Row34	0.407	0.414	0.38	0.384	0.398	0	34
Row35	0.365	0.366	0.348	0.358	0.356	0	39

**Gambar 4.32** Hasil *Out of Bag Prediction* RFR *Partitioning* (INDF)

Pada Gambar 4.32 dapat dilihat hasil *Out of Bag prediction* dari data emiten INDF yang di prediksi. Data masukan dengan hasil *out of bag prediction* yaitu untuk setiap baris masukan *mean* dan varians keluaran dari semua model yang tidak menggunakan baris untuk pelatihan. Jika seluruh data digunakan untuk melatih model individu maka keluaran ini akan berisi data masukan dengan nilai respon dan varians respon yang hilang. Kolom yang ditambahkan setara dengan kolom yang ditambahkan oleh node prediktor yang sesuai. Terdapat satu tambahan model kolom *count*, yang berisi jumlah model yang digunakan untuk pemungutan jumlah suara (jumlah model yang tidak menggunakan baris selama pembelajaran) Hasil *Out-of-Bag* dapat digunakan untuk mendapatkan perkiraan (prediksi) estimasi dengan menghubungkan ke dalam node *Numeric Scorer*.

Selanjutnya setelah pemodelan dengan data *training* yang telah divalidasi maka model tersebut dijadikan sebagai *input* model pada node *Random Forest Predictor (Regression)* dengan data *recordnya* menggunakan data *testing* yang sudah divalidasi berjumlah 101 *record* dengan konfigurasi yang sama dengan proses prediksi pada node *Predictor Regression* (Gambar 4.20) sebelumnya. Adapun salah satu hasil dari proses prediksi pada node *Random Forest Predictor (Regression)* dapat dilihat pada Gambar 4.33 (dataset emiten INDF) sebagai berikut :

Prediction output - 2:55 - Random Forest Predictor (Regression) (Prediksi)

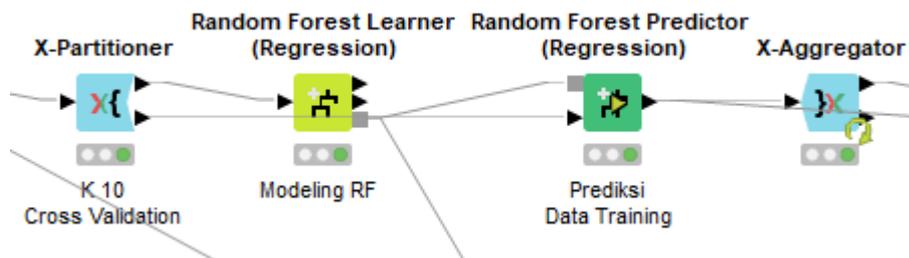
File Hiite Navigation View

Table "default" - Rows: 101 Spec - Columns: 6 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)	D Prediction (Close) (Prediction Variance)
Row24	0.226	0.245	0.225	0.237	0.24	0.002
Row29	0.308	0.298	0.241	0.232	0.279	0.001
Row53	0.169	0.181	0.166	0.184	0.181	0
Row59	0.205	0.202	0.204	0.2	0.211	0.001
Row60	0.221	0.213	0.209	0.226	0.208	0
Row89	0.252	0.298	0.246	0.289	0.286	0.001
Row149	0.577	0.583	0.594	0.595	0.586	0
Row159	0.597	0.593	0.61	0.605	0.598	0
Row177	0.525	0.519	0.53	0.521	0.527	0
Row186	0.551	0.572	0.562	0.574	0.568	0
Row190	0.608	0.609	0.605	0.616	0.6	0
Row200	0.541	0.572	0.556	0.568	0.563	0
Row237	0.794	0.789	0.754	0.742	0.769	0
Row242	0.732	0.731	0.727	0.742	0.732	0
Row261	0.917	0.915	0.914	0.916	0.921	0
Row298	0.69	0.715	0.701	0.695	0.711	0
Row300	0.69	0.704	0.701	0.711	0.703	0
Row305	0.628	0.646	0.637	0.637	0.647	0
Row315	0.752	0.768	0.765	0.768	0.763	0
Row356	0.732	0.746	0.754	0.747	0.745	0
Row407	0.773	0.773	0.781	0.779	0.771	0
Row415	0.856	0.868	0.866	0.853	0.862	0
Row425	0.902	0.915	0.914	0.926	0.917	0
Row442	0.856	0.857	0.845	0.826	0.846	0
Row477	0.799	0.815	0.813	0.805	0.816	0

**Gambar 4.33** Hasil Prediksi Data Dengan RFR Predictor (INDF)

Pada Gambar 4.33 di atas merupakan hasil dari *Random Forest Predictor (Regression)* pada salah satu dataset yaitu emiten INDF. terdapat kolom *open*, *high*, *low*, *close*, hasil prediksi dengan nama kolom *Prediction (Close)* dan *Prediction (Close) (Prediction Variance)*. Adapun setiap baris memiliki hasil prediksi berbeda-beda, hasil nilai pada kolom *prediction close* didapatkan dari target *class (close)* dan variabel *predictor* yang mempengaruhi *class* terdiri dari *open*, *high* dan *low*. Selanjutnya setelah pemodelan data *training* hasil validasi dengan node *Random Forest Regression Learner* dan prediksi data *testing* hasil validasi menggunakan node *Random Forest Predictor (Regression)* maka selanjutnya dilakukannya implementasi menggunakan node *X-Aggregator* sebagai pelengkap penyelesaian proses *Cross Validation* yang dapat dilihat pada Gambar 4.34 sebagai berikut:



**Gambar 4.34** Proses Terakhir *Cross Validation (X-Aggregator)* RFR

Seperti sebelumnya, pada Gambar 4.34 di atas dapat dijelaskan bahwa node *X-Aggregator* harus menjadi akhir dari *loop* validasi silang dan harus mengikuti node *X-Partitioner*. Pada proses ini mengumpulkan hasil dari node prediktor, membandingkan *class* yang diprediksi dan *real class* serta mengeluarkan atau memberikan *output* prediksi untuk semua baris dan statistik iterasi. Adapun salah satu hasil (emiten INDF) dari proses node *X-Aggregator* pada data *training* dengan *Cross Validation (1010 record)* dapat dilihat pada Gambar 4.35 sebagai berikut:

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)	D Prediction (Close) (Prediction Variance)
Row2	0.19	0.187	0.172	0.195	0.189	0
Row7	0.112	0.113	0.111	0.118	0.124	0
Row9	0.174	0.181	0.166	0.195	0.18	0
Row30	0.262	0.308	0.262	0.316	0.288	0.001
Row40	0.35	0.34	0.273	0.263	0.305	0.002
Row47	0.288	0.282	0.252	0.242	0.255	0
Row58	0.216	0.208	0.209	0.2	0.215	0
Row66	0.164	0.15	0.14	0.137	0.15	0
Row73	0.1	0.1	0.1	0.103	0.107	0
Row91	0.288	0.276	0.273	0.274	0.273	0
Row97	0.288	0.287	0.289	0.295	0.285	0
Row109	0.448	0.509	0.46	0.5	0.493	0.001
Row145	0.592	0.583	0.588	0.595	0.583	0
Row158	0.587	0.593	0.594	0.6	0.594	0
Row170	0.582	0.577	0.578	0.584	0.579	0
Row181	0.53	0.54	0.546	0.542	0.537	0
Row202	0.567	0.556	0.54	0.558	0.55	0
Row210	0.587	0.593	0.578	0.589	0.584	0
Row249	0.845	0.9	0.866	0.868	0.881	0
Row254	0.752	0.847	0.765	0.858	0.8	0.001
Row255	0.866	0.863	0.84	0.858	0.851	0
Row271	0.954	0.952	0.925	0.911	0.934	0.001
Row275	0.907	0.931	0.904	0.932	0.91	0.001
Row286	0.871	0.873	0.882	0.863	0.878	0
Row291	0.788	0.794	0.797	0.789	0.794	0

**Gambar 4.35** Prediction Table Dari Proses *X-Aggregator* RFR (INDF)

Pada Gambar 4.35 di atas menunjukkan hasil untuk setiap baris dari nilai yang diprediksi baik data *training* dan juga data *testing* hasil dari evaluasi dengan validasi *Cross Validation* berjumlah 1010 *record*. Pada proses ini pun ditunjukkan hasil *error rates* pada setiap nilai *K-Fold*nya yang dapat dilihat salah satunya adalah hasil data emiten INDF pada Gambar 4.36 sebagai berikut :

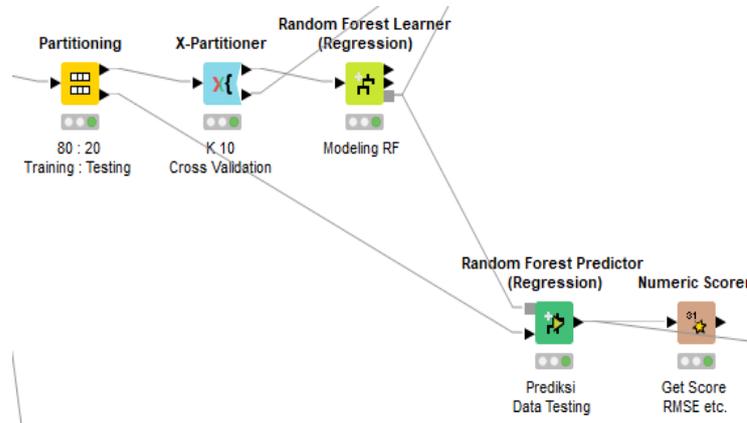
Row ID	Total squared error	Mean squared error	Size of Test Set
fold 0	0.025	0	101
fold 1	0.023	0	101
fold 2	0.021	0	101
fold 3	0.03	0	101
fold 4	0.016	0	101
fold 5	0.029	0	101
fold 6	0.028	0	101
fold 7	0.029	0	101
fold 8	0.028	0	101
fold 9	0.029	0	101

**Gambar 4.36** Error Rate 10 K-Fold Cross Validation Model RFR (INDF)

Pada hasil Gambar 4.36 di atas menunjukkan *error rate* emiten INDF menggunakan model RFR bahwa nilai total *squared error* terendah yaitu ada pada k-fold 4 sebesar 0.016 dan tertinggi ada pada k-fold 3 sebesar 0.030.

Proses di atas merupakan proses terakhir yang selanjutnya akan dilanjutkan dengan proses *scoring* dengan node *Numeric Scorer* agar mendapatkan hasil berupa nilai RMSE, MAE, MAPE dan  $R^2$  dari hasil prediksi yang dihasilkan.

Sedangkan pada data *testing* (253 record) dari hasil *partitioning*, dilakukan prediksi langsung menggunakan model hasil dari node *Random Forest Regression Learner* yang sudah dilakukan dengan menguji coba data *training* hasil proses validasi menggunakan *Cross Validation*, dengan menghubungkan node *learner* tersebut sebagai model *input* ke node *Random Forest Regression Predictor* yang dapat dilihat prosesnya pada Gambar 4.37 sebagai berikut :



**Gambar 4.37** Prediksi Data *Testing* Hasil *Partitioning* Dengan Model RFR

Pada Gambar 4.37 di atas menunjukkan proses prediksi pada data *testing* hasil *partitioning* dengan jumlah *record* sebanyak 253 data, dan dihasilkan nilai prediksi yang dapat dilihat pada Gambar 4.38 (emiten INDF) sebagai berikut :

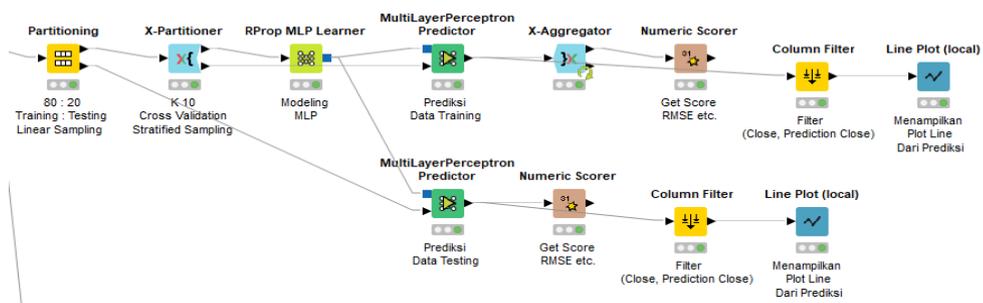
Row ID	Open	High	Low	Close	Prediction (Close)	Prediction (Close) (Prediction Variance)
Row6	0.123	0.113	0.112	0.119	0.121	0
Row11	0.2	0.192	0.166	0.179	0.192	0
Row13	0.205	0.234	0.182	0.174	0.214	0.002
Row15	0.195	0.181	0.177	0.189	0.189	0
Row20	0.2	0.224	0.204	0.232	0.214	0
Row21	0.241	0.245	0.225	0.242	0.23	0.001
Row22	0.216	0.229	0.193	0.184	0.209	0
Row23	0.2	0.229	0.204	0.237	0.214	0
Row33	0.422	0.408	0.417	0.405	0.411	0
Row41	0.257	0.255	0.246	0.237	0.25	0
Row43	0.314	0.313	0.305	0.3	0.303	0
Row44	0.288	0.345	0.295	0.358	0.342	0.001
Row50	0.169	0.176	0.172	0.163	0.168	0
Row57	0.247	0.245	0.225	0.221	0.232	0.001
Row69	0.149	0.129	0.15	0.142	0.14	0
Row75	0.174	0.155	0.145	0.153	0.15	0
Row90	0.314	0.308	0.305	0.305	0.3	0
Row96	0.319	0.303	0.268	0.263	0.287	0
Row108	0.407	0.477	0.417	0.484	0.442	0.001
Row118	0.51	0.504	0.508	0.5	0.503	0
Row132	0.644	0.641	0.61	0.637	0.631	0
Row134	0.582	0.593	0.594	0.589	0.594	0
Row136	0.546	0.567	0.546	0.568	0.558	0
Row141	0.52	0.525	0.524	0.532	0.524	0
Row152	0.572	0.562	0.588	0.574	0.574	0

**Gambar 4.38** Predicted Data Testing Partitioning Dengan Model RFR (INDF)

Proses pada Gambar 4.38 di atas merupakan proses terakhir dalam prediksi data *testing* hasil *partitioning* (253 record) menggunakan pemodelan *Random Forest Regression* dengan evaluasi *Cross Validation* yang selanjutnya akan dilanjutkan dengan proses *scoring* untuk mendapatkan hasil berupa nilai RMSE, MAE, MAPE dan  $R^2$  dari nilai prediksi yang didapatkan dari hasil node RFR Predictor.

### 3. Model Prediksi Metode *Multilayer Perceptron*

Desain model pengujian data dengan pembagian data *partitioning* 80 : 20 disertai proses validasi *Cross Validation* pada metode MLP dengan dataset tiga emiten dapat dilihat pada Gambar 4.39 sebagai berikut:

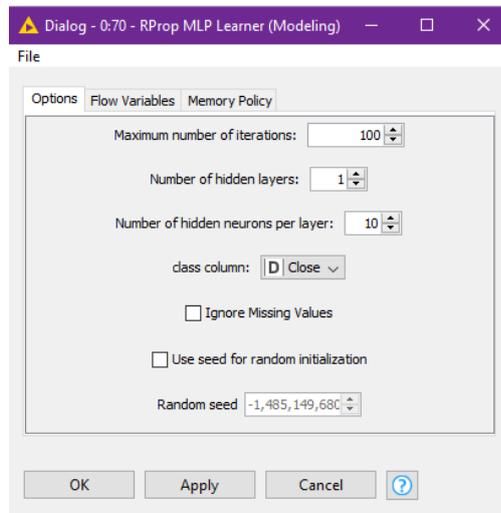


**Gambar 4.39** *Partitioning* Disertai Validasi *Cross Validation* Metode MLP  
(INDF, KAEF, TLKM)

Pada desain model Gambar 4.39 di atas, dibuat proses *modelling partitioning* sebesar 80 : 20 dengan hasil data 1010 *record* sebagai data *training* dan 253 *record* sebagai data *testing*. Proses *partitioning* ini pun sama dengan proses *partitioning* pada metode LR dan RFR sebelumnya, hasilnya pun memiliki jumlah *instance* yang sama dengan teknik pemilihan *Linear Sampling* pada dua metode sebelumnya yang dapat dilihat pada Tabel 4.6.

Adapun selanjutnya setelah tahap *partitioning* maka dilanjutkan dengan validasi data menggunakan node *X-Partitioner* sebagai proses awal *Cross Validation* yang selanjutnya salah satu *output* dari *X-Partitioner* (dalam kasus ini data *training* yang dihasilkan dari *X-Partitioner* yang digunakan) untuk data *input* pada pemodelan MLP *Learner* dengan konfigurasi *Number of Validation* nya sebanyak 10 dan menggunakan teknik *Stratified Sampling* pada *column class (close)*, adapun hasil jumlah data dari *X-Partitioner* dengan nilai K 10 adalah sebanyak 909 pada data *training* dan 101 pada data *testing*.

Kemudian data *training* hasil *partitioning* yang sudah divalidasi dengan *Cross Validation* dan menghasilkan jumlah data *training* baru berjumlah 909 dan data *testing* berjumlah 101 *record* sebagai data yang akan digunakan pada node prediksinya, maka data tersebut akan diinput pada node MLP *Learner* sebagai proses pembelajar dalam pemodelan MLP. Adapun data *input* untuk node pemodelannya menggunakan data *training* sebesar 909 *record* dengan konfigurasi yang dapat dilihat pada Gambar 4.40 sebagai berikut:



**Gambar 4.40** Konfigurasi MLP *Learner Data Partitioning* Dengan Validasi (INDF)

Dari Gambar 4.40 konfigurasi di atas, terdapat beberapa konfigurasi yang memungkinkan untuk diubah sesuai kebutuhan. Terdapat *Maximum Number of Iterations* yang dapat diubah sesuai kebutuhan, jumlah *hidden layer*, dan jumlah *neuron per layer*nya. Pada uji coba ini, peneliti melakukan percobaan pemodelan MLP terhadap data *training 909 record* yang sudah melalui tahap *Cross Validation* dengan jumlah *hidden layer* sebanyak 1 - 10 dan masing-masing *neuron* pada *layer*nya berjumlah 10 *neuron*. Hal ini dilakukan untuk menyeleksi hasil dari proses *scorer* yang dilihat pada nilai utama yaitu RMSE nya.

Penyeleksian dilakukan untuk menentukan setidaknya 3 model MLP dengan jumlah *hidden layer* berbeda yang menghasilkan nilai RMSE dengan tingkat nilai yang rendah. Adapun berikut tabel percobaan yang dilakukan oleh peneliti pada pemodelan menggunakan metode MLP sebagai berikut:

Tabel 4.10 Daftar Percobaan Pada Model Prediksi MLP

No	Percobaan	Hidden Layer	Neuron/Layer	Num. of Iterarions
1	Percobaan Ke - 1	1 Hidden Layer	10 Neuron	100
2	Percobaan Ke – 2	2 Hidden Layer	10 Neuron	100
3	Percobaan Ke – 3	3 Hidden Layer	10 Neuron	100
4	Percobaan Ke – 4	4 Hidden Layer	10 Neuron	100

5	Percobaan Ke – 5	5 Hidden Layer	10 Neuron	100
6	Percobaan Ke – 6	6 Hidden Layer	10 Neuron	100
7	Percobaan Ke – 7	7 Hidden Layer	10 Neuron	100
8	Percobaan Ke – 8	8 Hidden Layer	10 Neuron	100
9	Percobaan Ke – 9	9 Hidden Layer	10 Neuron	100
10	Percobaan Ke - 10	10 Hidden Layer	10 Neuron	100

Dilihat dari Tabel 4.10, Setelah dilakukan proses *learner* MLP menggunakan data *input training* sebanyak 909 *record* dengan menggunakan 10 kali percobaan dengan jumlah *hidden layer* berbeda, selanjutnya model tersebut disimpan dan diimplementasikan sebagai *input* model ke node *predictor* MLP. Adapun pada prediksi pertama menggunakan data *testing* yang sudah melalui tahap validasi yang berjumlah sebanyak 101 *record* (data ini merupakan kesatuan data *training* dari hasil *partitioning*), dan prediksi kedua menggunakan data *testing* dari hasil *partitioning* sebesar 253 *record* (prediksi dilakukan dengan node MLP *learner* yang sudah diuji coba pada tahap di atas dengan jumlah *hidden layer* yang berbeda). Adapun hasil prediksi data pertama dapat dilihat pada Gambar 4.41 sebagai berikut:

Row ID	Open	High	Low	Close	Prediction (Close)
Row15	0.195	0.181	0.177	0.189	0.163
Row20	0.2	0.224	0.204	0.232	0.188
Row30	0.262	0.308	0.262	0.316	0.277
Row43	0.314	0.313	0.305	0.3	0.308
Row68	0.143	0.139	0.14	0.142	0.143
Row71	0.143	0.129	0.126	0.118	0.141
Row81	0.164	0.15	0.15	0.153	0.148
Row97	0.288	0.287	0.289	0.295	0.285
Row167	0.567	0.583	0.567	0.589	0.572
Row180	0.536	0.546	0.54	0.553	0.541
Row181	0.53	0.54	0.546	0.542	0.538
Row195	0.556	0.572	0.572	0.584	0.566
Row198	0.551	0.546	0.551	0.558	0.549
Row201	0.567	0.567	0.562	0.547	0.565
Row206	0.592	0.593	0.588	0.6	0.59
Row222	0.814	0.873	0.818	0.826	0.844
Row227	0.825	0.831	0.829	0.842	0.833
Row228	0.845	0.852	0.856	0.863	0.855
Row237	0.794	0.789	0.754	0.742	0.779
Row248	0.788	0.826	0.813	0.837	0.816
Row255	0.866	0.863	0.84	0.858	0.858
Row278	0.881	0.884	0.888	0.874	0.88
Row291	0.788	0.794	0.797	0.789	0.796
Row293	0.845	0.842	0.792	0.8	0.829
Row311	0.649	0.657	0.658	0.658	0.653

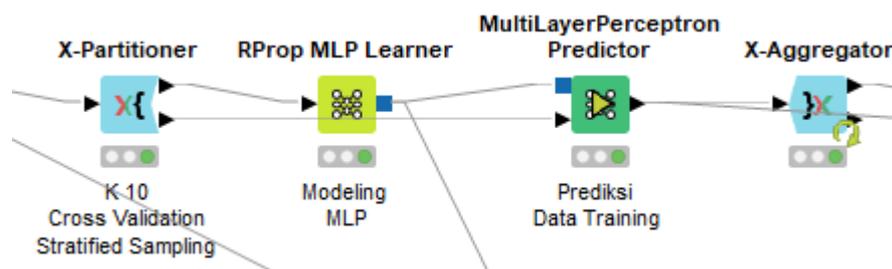
**Gambar 4.41** Hasil MLP Predictor Data Testing Partitioning Dengan Cross Validation (INDF)

Gambar 4.41 di atas merupakan salah satu hasil prediksi (emiten INDF) pada data *testing* 101 record menggunakan model prediksi MLP dengan jumlah *hidden layer* 5 dan *neuron per layer* sebanyak 10. Adapun prediksi data MLP Predictor pada data *testing* hasil *partitioning* yang berjumlah 253 record dapat dilihat pada Gambar 4.42 sebagai berikut:

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)
Row4	0.169	0.155	0.118	0.114	0.16
Row9	0.174	0.181	0.166	0.195	0.18
Row14	0.169	0.181	0.166	0.174	0.18
Row19	0.159	0.197	0.132	0.211	0.175
Row24	0.226	0.245	0.225	0.237	0.224
Row29	0.308	0.298	0.241	0.232	0.261
Row34	0.407	0.414	0.38	0.384	0.385
Row39	0.35	0.356	0.359	0.353	0.337
Row44	0.288	0.345	0.295	0.358	0.298
Row49	0.21	0.192	0.166	0.179	0.188
Row54	0.2	0.213	0.193	0.211	0.201
Row59	0.205	0.202	0.204	0.2	0.201
Row64	0.179	0.171	0.172	0.163	0.179
Row69	0.149	0.129	0.15	0.142	0.159
Row74	0.131	0.113	0.113	0.124	0.145
Row79	0.127	0.113	0.115	0.117	0.145
Row84	0.195	0.229	0.198	0.226	0.207
Row89	0.252	0.298	0.246	0.289	0.255
Row94	0.267	0.298	0.273	0.3	0.266
Row99	0.262	0.261	0.246	0.237	0.242
Row104	0.396	0.387	0.385	0.389	0.372
Row109	0.448	0.509	0.46	0.5	0.478
Row114	0.463	0.451	0.46	0.447	0.45
Row119	0.515	0.504	0.466	0.474	0.49
Row124	0.582	0.625	0.594	0.605	0.622

**Gambar 4.42** Hasil MLP Predictor Data Testing Partitioning (INDF)

Selanjutnya setelah pemodelan data *training* dengan MLP Learner dan prediksi data *testing* hasil validasi dan data *testing* hasil *partitioning* menggunakan MLP Predictor, maka dilanjutkan dengan dilakukannya implementasi node X-Aggregator sebagai pelengkap penyelesaian proses Cross Validation yang dapat dilihat pada Gambar 4.43 sebagai berikut:



**Gambar 4.43** Proses *Cross Validation* Menggunakan *X-Aggregator* MLP

Pada Gambar 4.43 dapat dijelaskan seperti sebelumnya, node *X-Aggregator* harus menjadi akhir dari *loop* validasi silang dan harus mengikuti node *X-Partitioner*. Pada proses ini mengumpulkan hasil dari node prediktor, membandingkan *class* yang diprediksi dan *real class* serta memberikan *output* prediksi untuk semua baris dan statistik iterasi. Adapun salah satu hasil dari proses *X-Aggregator* pada data *training* dengan *Cross Validation* (1010 *record*) dapat dilihat pada Gambar 4.44 (emiten INDF) sebagai berikut :

Row ID	Open	High	Low	Close	Predict...
Row1	0.159	0.165	0.15	0.168	0.15
Row7	0.112	0.113	0.111	0.118	0.13
Row11	0.2	0.192	0.166	0.179	0.184
Row32	0.381	0.408	0.38	0.421	0.391
Row47	0.288	0.282	0.252	0.242	0.274
Row52	0.169	0.155	0.145	0.142	0.148
Row75	0.174	0.155	0.145	0.153	0.149
Row98	0.298	0.287	0.279	0.274	0.288
Row101	0.308	0.324	0.289	0.332	0.307
Row105	0.381	0.366	0.375	0.368	0.376
Row115	0.463	0.498	0.466	0.511	0.471
Row130	0.634	0.625	0.653	0.637	0.646
Row133	0.628	0.62	0.605	0.595	0.618
Row186	0.551	0.572	0.562	0.574	0.56
Row213	0.587	0.588	0.588	0.6	0.589
Row218	0.608	0.63	0.626	0.632	0.628
Row226	0.804	0.847	0.813	0.811	0.826
Row233	0.794	0.789	0.743	0.737	0.762
Row256	0.85	0.857	0.861	0.842	0.864
Row273	0.938	0.947	0.947	0.953	0.914
Row296	0.659	0.699	0.61	0.674	0.652
Row308	0.665	0.683	0.669	0.668	0.676
Row321	0.68	0.694	0.685	0.679	0.689
Row333	0.752	0.762	0.765	0.763	0.757
Row347	0.768	0.762	0.765	0.758	0.76

**Gambar 4.44** *Prediction Table* Dari Proses *X-Aggregator* MLP (INDF)

Pada Gambar 4.44 di atas merupakan salah satu hasil percobaan MLP dengan jumlah *hidden layer* sebanyak 5 yang menunjukkan seluruh hasil untuk semua baris dari nilai yang diprediksi baik data *training* dan juga data *testing* yang sudah dievaluasi dengan validasi 10 *Cross Validation* berjumlah 1010 *record*. Pada proses ini pun ditunjukkan hasil *error rates* pada setiap nilai *K-Foldnya*, salah satunya adalah pada Gambar 4.45 hasil dari dataset emiten INDF sebagai berikut :

Row ID	Total squared error	Mean squared error	Size of Test Set
fold 0	0.027	0	101
fold 1	0.039	0	101
fold 2	0.041	0	101
fold 3	0.059	0.001	101
fold 4	0.036	0	101
fold 5	0.036	0	101
fold 6	0.032	0	101
fold 7	0.04	0	101
fold 8	0.037	0	101
fold 9	0.055	0.001	101

**Gambar 4.45** Error Rate Cross Validation Model MLP Partitioning (INDF)

Pada Gambar 4.45 dapat dilihat *error rate* emiten INDF menggunakan model MLP menunjukkan bahwa nilai total *squared error* terendah yaitu ada pada fold 0 sebesar 0.027 dan tertinggi ada pada fold 3 sebesar 0.059.

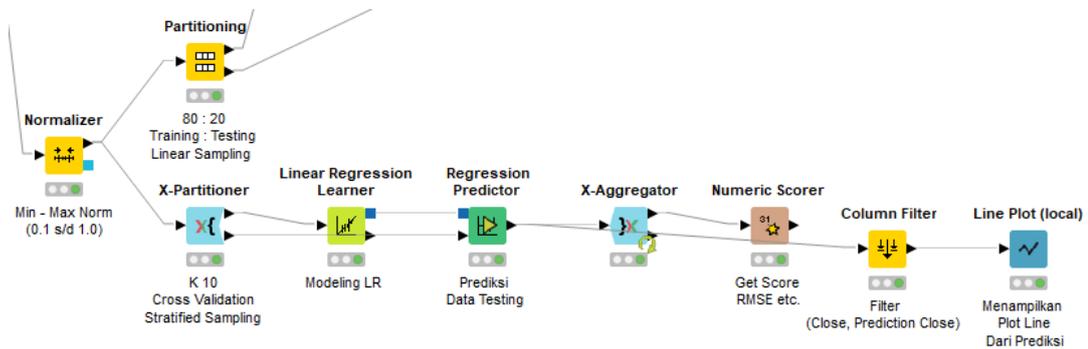
Proses di atas merupakan proses terakhir yang selanjutnya akan dilanjutkan dengan proses *scoring* dengan node *Numeric Scorer* agar mendapatkan hasil yang dapat dievaluasi berupa nilai RMSE, MAE, MAPE dan  $R^2$  dari nilai prediksi yang didapatkan.

#### 4.4.3. Desain Model Cross Validation (Tanpa Partitioning)

Pada penelitian ini, uji coba kedua dilakukan dengan cara pengujian model prediksi tanpa *partitioning*. Dari keseluruhan jumlah *instance* dalam dataset (1263 *instance* per emiten) dimodelkan dengan LR, RFR dan MLP tanpa *partitioning* (pembagian data *training* dan *testing*) melainkan langsung pengujian dengan validasi *K-10 Fold* menggunakan *Cross Validation* (node *X-Partitioner* dan *X-Aggregator*). Adapun berikut merupakan desain pengujian model menggunakan *Cross Validation* dengan metode LR, RFR dan MLP sebagai berikut:

##### 1. Model Prediksi Metode *Linear Regression*

Desain model pengujian data dengan yang langsung dievaluasi dengan *Cross Validation* tanpa adanya *partitioning* (pembagian data *testing* dan data *training*) pada metode *Linear Regression* dengan dataset ketiga emiten yakni INDF, KAEF dan TLKM dapat dilihat pada Gambar 4.46 berikut:



**Gambar 4.46** Desain Model Dengan *Cross Validation* LR (INDF, KAEF, TLKM)

Pada Gambar 4.46 proses di atas menunjukkan beberapa tahapan, bila uji coba pertama dilakukan dengan proses *partitioning*, untuk uji coba kedua ini setelah proses data transformasi langsung dimodelkan dengan *Cross Validation* menggunakan node *X-Partitioner*. Adapun pada masing-masing dataset emiten dilakukan proses yang sama dalam uji coba menggunakan metode *Linear Regression*. Setelah dilakukannya evaluasi model menggunakan validasi data dengan *Cross Validation* pada keseluruhan dataset berjumlah 1263 per emiten selanjutnya yaitu pemodelan *learner* menggunakan node *Linear Regression Learner*. Adapun konfigurasi pada tahap *Cross Validation* sama seperti pada tahap satu yaitu menggunakan nilai K atau *Number of Validation* 10 dengan teknik *Stratified Sampling* berdasarkan atribut yang menjadi *class* prediksinya (*close*). Pada proses ini dihasilkan data validasi *training* dan data validasi *testing* sebagai berikut :

Tabel 4 11. *Output* Proses *Cross Validation* Dengan *X-Partitioner*

No	Hasil <i>Cross Validation</i>	Jumlah
1	First Table ( <i>Training</i> Data)	1137 Record
2	Second Table ( <i>Testing</i> Data)	126 Record

Dari Tabel 4.11 di atas, jumlah data yang dihasilkan per emiten berbeda dengan uji coba pertama yang sebelumnya menggunakan data *training* hasil *partitioning* yang diproses menggunakan *Cross Validation* (jumlah data *training* 909 record dan data *testing* 101 record). Pada uji coba kedua ini jumlah data

*training* dan *testing* hasil dari proses evaluasi dengan *Cross Validation* (tanpa *partitioning*) sebesar 1137 pada *training* dan 126 pada *testing* dan data yang dihasilkan ini nantinya dimodelkan dengan node *Linear Regression Learner*.

Adapun data yang menjadi data *input* pada node *Linear Regression Learner* adalah data hasil validasi di atas (data *training* 1137). Konfigurasi dari proses *learner* LR yaitu sama dengan pemodelan uji coba pertama yang menggunakan kolom *close* sebagai target atribut dari proses *learner* LR dan sebagai atribut *predictor* yang mempengaruhi kolom *class* pada proses *learner* adalah *open*, *high*, dan *low*. Proses ini adalah proses pemodelan, dimana data *training* (1137 *record*) hasil validasi dimodelkan dengan node LR *Learner* dan nantinya hasil dari model ini di jadikan model *input* untuk node LR *predictor* dengan menggunakan data *testing* hasil validasi (126 *record*). Adapun pada proses pemodelan ini menghasilkan *Coefficients and Statistics* sebagai berikut :

Tabel 4.12 *Coefficients and Statistic Model Learner LR Dengan Validasi (INDF)*

No	Variabel	<i>Coeff.</i>	<i>Std. Error</i>	<i>t-value</i>
1	Open	-0.5797	0.0231	25.1110
2	High	0.7691	0.0189	40.6163
3	Low	0.7887	0.0205	38.4036
4	Intercept	0.0104	0.0012	8.3512
5	Multiple R-Square	0.9963		
6	Adjusted R-Square	0.9962		

Tabel 4.13 *Coefficients and Statistic Model Learner LR Dengan Validasi (KAEF)*

No	Variabel	<i>Coeff.</i>	<i>Std. Error</i>	<i>t-value</i>
1	Open	-0.4985	0.0278	-17.9443
2	High	0.6759	0.0146	46.2461
3	Low	0.8143	0.0245	33.2355
4	Intercept	0.0011	0.0009	1.2159
5	Multiple R-Square	0.9977		
6	Adjusted R-Square	0.9977		

Tabel 4.14 *Coefficients and Statistic Model Learner LR Dengan Validasi(TLKM)*

No	Variabel	<i>Coeff.</i>	<i>Std. Error</i>	<i>t-value</i>
1	Open	-0.5467	0.0220	-24.8017
2	High	0.7941	0.0180	44.0682
3	Low	0.8005	0.0199	40.1420
4	Intercept	-0.0440	0.0011	-36.0201
5	Multiple R-Square	0.9973		
6	Adjusted R-Square	0.9973		

Pada data Tabel 4.12, Tabel 4.13 dan Tabel 4.14 di atas dapat dilihat bahwa masing-masing model LR dari setiap dataset per emiten memiliki hasil *Coefficients* dan *Standard Error* yang berbeda beda. *Standard Error* pada variabel *open* yang paling kecil ada pada model *learner LR* dengan dataset TLKM, pada variabel *high* dengan dataset KAEF dan pada variabel *low* dengan nilai terkecil pada dataset TLKM, hal ini dikarenakan masing-masing *range* harga pada setiap variabel dari ketiga emiten berbeda.

Selanjutnya dapat dilihat nilai R dari hasil tabel di atas pada masing-masing hasil pemodelan dengan tiga emiten yang berbeda, bahwa nilai R yang dihasilkan pun berbeda, dilihat pada dataset INDF menghasilkan nilai *multiple R* 0.9963 yang berbeda dengan *Adjusted R* nya sebesar 0.9962, adapun pada KAEF dan TLKM menghasilkan nilai *Multiple R-Square* dan *Adjusted R-Square* yang sama.

Setelah pemodelan di atas, maka model tersebut dijadikan sebagai *input* model pada node *Predictor Regression* dengan data *recordnya* menggunakan data *testing* hasil dari validasi berjumlah 126 *record* dengan konfigurasi node *Predictor Regression* yang sama seperti pada Gambar 4.20 sebelumnya, adapun salah satu hasil dari prediksi dengan data *testing* hasil pembagian data menggunakan validasi dengan *Cross Validation* menghasilkan prediksi data yang dapat dilihat pada Gambar 4.47 (menggunakan data emiten INDF) sebagai berikut :

Row ID	Open	High	Low	Close	Prediction (Close)
Row6	0.123	0.113	0.112	0.119	0.114
Row8	0.133	0.155	0.126	0.163	0.152
Row26	0.252	0.25	0.252	0.247	0.255
Row29	0.308	0.298	0.241	0.232	0.251
Row35	0.365	0.366	0.348	0.358	0.355
Row46	0.329	0.324	0.316	0.305	0.318
Row48	0.236	0.245	0.214	0.205	0.231
Row56	0.267	0.25	0.246	0.263	0.242
Row63	0.133	0.176	0.134	0.174	0.175
Row66	0.164	0.15	0.14	0.137	0.14
Row94	0.267	0.298	0.273	0.3	0.3
Row118	0.51	0.504	0.508	0.5	0.503
Row129	0.613	0.625	0.626	0.637	0.629
Row143	0.597	0.599	0.599	0.595	0.597
Row152	0.572	0.562	0.588	0.574	0.575
Row161	0.577	0.572	0.556	0.553	0.555
Row179	0.52	0.525	0.53	0.516	0.53
Row186	0.551	0.572	0.562	0.574	0.574
Row213	0.587	0.588	0.588	0.6	0.586
Row238	0.737	0.736	0.722	0.737	0.719
Row246	0.757	0.752	0.765	0.758	0.753
Row249	0.845	0.9	0.866	0.868	0.896
Row286	0.871	0.873	0.882	0.863	0.873
Row293	0.845	0.842	0.792	0.8	0.792
Row296	0.659	0.699	0.61	0.674	0.647

**Gambar 4.47** Hasil Node *Predictor Regression* Melalui *Cross Validation* (INDF)

Pada Gambar 4.47 di atas merupakan hasil dari node *Predictor Regression* pada setiap dataset emiten dengan data *testing* sebesar 162 *record*, terdapat kolom *open*, *high*, *low*, *close* dan hasil prediksi dengan nama kolom *Prediction (Close)*.

Adapun setiap baris memiliki hasil prediksi berbeda beda, hasil nilai pada kolom *Prediction Close* didapatkan dari target *class (close)* dan variabel *predictor* yang mempengaruhi *class* terdiri dari *open*, *high* dan *low*. Selanjutnya setelah pemodelan LR pada data *training* hasil *Cross Validation* sebesar 1137 *record* dengan node *LR Learner*, dan prediksi data *testing* hasil *Cross Validation* sebesar 126 *record* menggunakan node *Predictor Regression*, maka dilakukannya implementasi dengan node *X-Aggregator* sebagai pelengkap penyelesaian proses *Cross Validation*.

Seperti sebelumnya, node *X-Aggregator* harus menjadi akhir dari *loop* validasi silang dan harus mengikuti node *X-Partitioner*. Pada proses ini mengumpulkan hasil dari node *Predictor Regression*, membandingkan *class* yang diprediksi dan *real class* serta memberikan *output* prediksi untuk semua baris (data *training* 1137 dan data *testing* 126 *record*) dan statistik iterasi. Adapun hasil

dari proses pada node *X-Aggregator* dapat dilihat pada Gambar 4.48 (menggunakan dataset emiten INDF) sebagai berikut :

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)
Row0	0.195	0.176	0.161	0.168	0.16
Row14	0.169	0.181	0.166	0.174	0.183
Row21	0.241	0.245	0.225	0.242	0.236
Row69	0.149	0.129	0.15	0.142	0.141
Row78	0.133	0.113	0.114	0.119	0.11
Row81	0.164	0.15	0.15	0.153	0.149
Row83	0.185	0.234	0.172	0.189	0.219
Row91	0.288	0.276	0.273	0.274	0.272
Row93	0.257	0.266	0.257	0.268	0.269
Row100	0.247	0.292	0.252	0.305	0.291
Row103	0.391	0.387	0.359	0.395	0.365
Row110	0.468	0.514	0.482	0.489	0.514
Row132	0.644	0.641	0.61	0.637	0.611
Row140	0.515	0.525	0.519	0.521	0.525
Row156	0.572	0.562	0.588	0.574	0.575
Row164	0.52	0.567	0.535	0.574	0.567
Row182	0.541	0.556	0.551	0.563	0.559
Row204	0.556	0.583	0.562	0.584	0.579
Row209	0.587	0.599	0.572	0.589	0.582
Row219	0.639	0.673	0.647	0.684	0.668
Row227	0.825	0.831	0.829	0.842	0.826
Row231	0.794	0.794	0.743	0.768	0.748
Row237	0.794	0.789	0.754	0.742	0.752
Row245	0.752	0.762	0.765	0.758	0.764
Row247	0.768	0.789	0.781	0.784	0.788

**Gambar 4.48** Hasil *X-Aggregator* (Tanpa *Partitioning*) LR (INDF)

Pada Gambar 4.48 di atas menunjukkan hasil untuk setiap baris dari nilai yang diprediksi baik data *training* 1137 *record* dan juga data *testing* hasil dari validasi 10 *Cross Validation* berjumlah 126 *record* pada data emiten INDF yang diuji coba. Pada proses ini pun menghasilkan *error rates* pada setiap nilai *K-Fold*nya yang dapat dilihat pada Gambar 4.49 (dataset emiten INDF) sebagai berikut :

Row ID	D Total squared error	D Mean squared error	I Size of Test Set
fold 0	0.019	0	127
fold 1	0.024	0	127
fold 2	0.021	0	127
fold 3	0.016	0	126
fold 4	0.017	0	126
fold 5	0.017	0	126
fold 6	0.017	0	126
fold 7	0.022	0	126
fold 8	0.016	0	126
fold 9	0.021	0	126

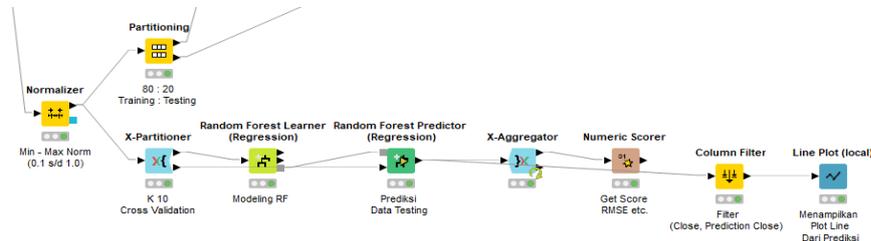
**Gambar 4.49** *Error Rate Cross Validation* (Tanpa *Partitioning*) Model LR (INDF)

Pada Gambar 4.49 di atas dapat dilihat hasil tabel view *error rate* emiten INDF bahwa nilai total *squared error* terendah yaitu ada pada *fold* 3 dan 8 sebesar 0.016 dan tertinggi ada pada *fold* 1 sebesar 0.024.

Proses pada Gambar 4.49 merupakan proses terakhir dalam pemodelan dataset menggunakan *Cross Validation* (tanpa *partitioning*) dengan hasil data *training* sebesar 1137 dan data *testing* sebesar 126 *record* yang selanjutnya akan dilanjutkan dengan proses *scoring* dengan node *Numeric Scorer* bertujuan untuk mendapatkan hasil berupa nilai yang akan dievaluasi berupa RMSE, MAE, MAPE dan  $R^2$  dari nilai prediksi yang dihasilkan.

## 2. Model Prediksi Metode *Random Forest Regression*

Desain model pengujian data kedua yaitu langsung dievaluasi dengan *Cross Validation* tanpa adanya *partitioning* (pembagian data *testing* dan data *training*) menggunakan metode MLP dengan dataset ketiga emiten yakni INDF, KAEF dan TLKM adalah sebagai berikut:



Gambar 4.50 Desain Model Dengan *Cross Validation* RFR (INDF, KAEF, TLKM)

Pada Gambar 4.50 proses di atas menunjukkan bahwa untuk uji coba kedua ini setelah proses transformasi data maka langsung dimodelkan dengan *Cross Validation* menggunakan node *X-Partitioner*. Adapun pada masing-masing dataset emiten dilakukan proses uji coba yang sama. Setelah dilakukannya evaluasi model menggunakan validasi data dengan *Cross Validation* pada keseluruhan dataset berjumlah 1263 per emiten selanjutnya yaitu pemodelan *learner* menggunakan node *Random Forest Regression Learner*.

Adapun konfigurasi pada tahap *Cross Validation* pada proses ini sama seperti pada Gambar 4.18 yaitu menggunakan nilai K atau *Number of Validation* 10

dengan teknik *Stratified Sampling* berdasarkan atribut yang menjadi *class* prediksinya (*close*). Pada proses ini dihasilkan data validasi *training* sebesar 1137 dan data validasi *testing* sebesar 126 *record*. Selanjutnya pada pemodelan dengan node *learnernya* yaitu menggunakan node *Random Forest Learner (Regression)*. Adapun data yang menjadi data *input* pada node *Random Forest Learner (Regression)*, adalah data hasil *Cross Validation* (data *training* 1137 *record*).

Konfigurasi dari proses *learner* RFR pada model ini tidak berbeda dengan pemodelan dataset dengan *partitioning* yang disertai *Cross Validation* diuji coba pertama yaitu pada Gambar 4.30, adapun pada proses pemodelan ini menghasilkan nilai *Attribute Statistic* yang berbeda dengan pemodelan RFR dengan *partitioning* data yang dapat dilihat pada Gambar 4.51 (dengan dataset emiten INDF) di bawah ini :

Row ID	#splits (level 0)	#splits (level 1)	#splits (level 2)	#candidates (level 0)	#candidates (level 1)	#candidates (level 2)
Open	35	70	130	35	70	130
High	35	61	127	35	61	127
Low	30	69	143	30	69	143

**Gambar 4.51** *Attribute Statistic* Hasil Modelling RFR *Cross Validation* (INDF)

Pada Gambar 4.51 menunjukkan hasil tabel statistik atribut yang terdapat perbedan dari masing-masing nilai *split level* dan juga *candidates levelnya*, hal ini dikarenakan adanya proses yang berbeda dengan uji coba pertama. Pada uji coba pertama menggunakan *partitioning* data dengan memasukkan data *input* untuk pemodelan sebesar 909 data *training* sedangkan pada Gambar 4.51 di atas menggunakan data *input* pemodelan dengan jumlah 1137 data *training* hasil *Cross Validation*. Selanjutnya terdapat pula hasil *Out Of Bag Prediction* pada pemodelan RF yang dapat dilihat pada Gambar 4.52 sebagai berikut:

Row ID	D Open	D High	D Low	D Close	D Close (Out-of-bag)	D Close (Out-of-bag) (Prediction Variance)	I model count
Row0	0.195	0.176	0.161	0.168	0.156	0.001	37
Row1	0.159	0.165	0.15	0.168	0.163	0	40
Row2	0.19	0.187	0.172	0.195	0.19	0	41
Row3	0.216	0.197	0.188	0.184	0.19	0	33
Row4	0.169	0.155	0.118	0.114	0.139	0	29
Row5	0.102	0.103	0.101	0.105	0.106	0	37
Row6	0.123	0.113	0.112	0.119	0.122	0	44
Row7	0.112	0.113	0.111	0.118	0.125	0	36
Row8	0.133	0.155	0.126	0.163	0.139	0	40
Row9	0.174	0.181	0.166	0.195	0.181	0	41
Row10	0.185	0.181	0.172	0.189	0.19	0	41
Row11	0.2	0.192	0.166	0.179	0.196	0	42
Row12	0.195	0.187	0.166	0.2	0.187	0	38
Row13	0.205	0.234	0.182	0.174	0.198	0	38
Row14	0.169	0.181	0.166	0.174	0.184	0	38
Row15	0.195	0.181	0.177	0.189	0.191	0	40
Row16	0.185	0.165	0.156	0.163	0.166	0	40
Row17	0.149	0.16	0.145	0.158	0.156	0	36
Row18	0.133	0.144	0.126	0.147	0.14	0	32
Row19	0.159	0.197	0.132	0.211	0.162	0.001	35
Row20	0.2	0.224	0.204	0.232	0.219	0	30
Row21	0.241	0.245	0.225	0.242	0.224	0	41
Row22	0.216	0.229	0.193	0.184	0.209	0	38
Row23	0.2	0.229	0.204	0.237	0.214	0	38
Row24	0.226	0.245	0.225	0.237	0.235	0.001	43

**Gambar 4.52** Output Out of Bag Prediction RFR Cross Validation (INDF)

Pada Gambar 4.52 dapat dilihat hasil *output Out of Bag Prediction* dari data emiten INDF yang diprediksi yaitu untuk setiap baris masukan *mean* dan varians keluaran yang tidak menggunakan baris untuk pelatihan. Sama halnya dengan uji coba pertama (*partitioning* dengan *Cross Validation*), jika seluruh data digunakan untuk melatih model individu maka keluaran ini akan berisi data masukan dengan nilai respon dan varians respon yang hilang. Kolom yang ditambahkan setara dengan kolom yang ditambahkan oleh node prediktor yang sesuai. Terdapat satu tambahan model kolom *count*, yang berisi jumlah model yang digunakan untuk pemungutan jumlah suara (jumlah model yang tidak menggunakan baris selama pembelajaran.) Prediksi *Out of Bag Prediction* dapat digunakan untuk mendapatkan perkiraan estimasi nilai akhir prediksi dengan menghubungkan ke dalam node *Numeric Scorer*.

Selanjutnya setelah pemodelan dengan data *training* yang telah divalidasi maka hasil tersebut dijadikan sebagai *input* model pada node *Random Forest Predictor (Regression)* dengan data *recordnya* menggunakan data *testing* berjumlah 126 *record* dengan konfigurasi yang sama dengan proses prediksi pada Gambar 4.20 sebelumnya. Adapun hasil dari proses prediksi pada node *Random Forest Predictor (Regression)* dapat dilihat pada Gambar 4.53 sebagai berikut :

Prediction output - 0.93 - Random Forest Predictor (Regression) (Prediksi)

File Hilite Navigation View

Table "default" - Rows: 126 Spec - Columns: 6 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)	D Prediction (Close) (Prediction Variance)
Row1262	0.675	0.704	0.695	0.679	0.708	0
Row1261	0.69	0.683	0.69	0.679	0.682	0
Row1260	0.685	0.688	0.69	0.689	0.69	0
Row1259	0.721	0.715	0.695	0.689	0.692	0
Row1258	0.649	0.715	0.669	0.726	0.711	0
Row1257	0.587	0.641	0.605	0.653	0.621	0
Row1256	0.597	0.599	0.599	0.589	0.598	0
Row1255	0.618	0.62	0.615	0.6	0.616	0
Row1254	0.577	0.604	0.583	0.616	0.585	0
Row1253	0.567	0.588	0.567	0.595	0.585	0
Row1252	0.556	0.567	0.562	0.558	0.563	0
Row1251	0.541	0.556	0.556	0.558	0.559	0
Row1250	0.52	0.535	0.535	0.542	0.534	0
Row1249	0.551	0.551	0.53	0.521	0.552	0
Row1248	0.567	0.572	0.551	0.553	0.554	0
Row1247	0.494	0.556	0.508	0.563	0.54	0
Row1246	0.474	0.493	0.455	0.495	0.479	0.001
Row1245	0.437	0.44	0.412	0.437	0.419	0
Row1244	0.443	0.43	0.433	0.432	0.435	0
Row1243	0.427	0.419	0.423	0.421	0.422	0
Row1242	0.443	0.43	0.433	0.426	0.435	0
Row1241	0.443	0.435	0.444	0.437	0.445	0
Row1240	0.448	0.435	0.444	0.437	0.443	0
Row1239	0.479	0.477	0.487	0.489	0.473	0
Row1238	0.489	0.493	0.487	0.479	0.491	0

**Gambar 4.53** Prediction Output Dari Prediksi RFR Predictor Cross Validation (INDF)

Pada Gambar 4.53 menunjukkan hasil dari node *Random Forest Predictor (Regression)* pada dataset emiten INDF. terdapat kolom *open*, *high*, *low*, *close*, hasil prediksi dengan nama kolom *Prediction (Close)* dan *Prediction (Close) (Prediction Variance)*. Adapun setiap baris memiliki hasil prediksi berbeda-beda, hasil nilai pada kolom *Prediction Close* didapatkan dari target *class (close)* dan variabel *predictor* yang mempengaruhi *class* terdiri dari *open*, *high* dan *low*.

Selanjutnya setelah pemodelan data *training* (1137 record) dengan node *Random Forest Regression Learner* dan prediksi data *testing* (126 record) menggunakan node *Random Forest Predictor (Regression)* maka dilakukannya implementasi dengan node *X-Aggregator* sebagai pelengkap penyelesaian proses *Cross Validation*, node tersebut harus menjadi akhir dari *loop* validasi silang dan harus mengikuti node *X-Partitioner*. Pada proses ini mengumpulkan hasil dari node prediktor, membandingkan *class* yang diprediksi dan *real class* serta memberikan *output* prediksi untuk semua baris dan statistik iterasi.

Adapun hasil dari proses *X-Aggregator* pada dataset yang sudah divalidasi dengan *Cross Validation* (data *training* 1137 dan *testing* 126 record) dapat dilihat pada Gambar 4.54 sebagai berikut :

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)	D Prediction (Close) (Prediction Variance)
Row0	0.195	0.176	0.161	0.168	0.206	0.005
Row1	0.159	0.165	0.15	0.168	0.189	0.004
Row2	0.19	0.187	0.172	0.195	0.208	0.005
Row3	0.216	0.197	0.188	0.184	0.211	0.004
Row4	0.169	0.155	0.118	0.114	0.189	0.004
Row5	0.102	0.103	0.101	0.105	0.189	0.004
Row6	0.123	0.113	0.112	0.119	0.189	0.004
Row7	0.112	0.113	0.111	0.118	0.189	0.004
Row8	0.133	0.155	0.126	0.163	0.189	0.004
Row9	0.174	0.181	0.166	0.195	0.206	0.005
Row10	0.185	0.181	0.172	0.189	0.208	0.005
Row11	0.2	0.192	0.166	0.179	0.206	0.005
Row12	0.195	0.187	0.166	0.2	0.206	0.005
Row13	0.205	0.234	0.182	0.174	0.208	0.005
Row14	0.169	0.181	0.166	0.174	0.206	0.005
Row15	0.195	0.181	0.177	0.189	0.208	0.005
Row16	0.185	0.165	0.156	0.163	0.189	0.004
Row17	0.149	0.16	0.145	0.158	0.189	0.004
Row18	0.133	0.144	0.126	0.147	0.189	0.004
Row19	0.159	0.197	0.132	0.211	0.189	0.004
Row20	0.2	0.224	0.204	0.232	0.207	0.005
Row21	0.241	0.245	0.225	0.242	0.212	0.003
Row22	0.216	0.229	0.193	0.184	0.212	0.004
Row23	0.2	0.229	0.204	0.237	0.207	0.005
Row24	0.226	0.245	0.225	0.237	0.213	0.003

**Gambar 4.54** Prediction Table Hasil Node X-Aggregator RFR Cross Validation (INDF)

Pada Gambar 4.54 di atas menunjukkan sebagian hasil untuk setiap baris dari nilai yang diprediksi baik data *training* dan juga data *testing* yang sudah dievaluasi dengan *Cross Validation* berjumlah 1263 record per emiten. Pada proses ini juga mendapatkan hasil *Error rates* pada setiap nilai *K-Fold*nya dari data *testing* yang dapat dilihat pada Gambar 4.55 (dataset emiten INDF) sebagai berikut :

Row ID	D Total squared error	D Mean squared error	I Size of Test Set
fold 0	0.146	0.001	127
fold 1	0.021	0	126
fold 2	0.051	0	126
fold 3	0.019	0	127
fold 4	0.019	0	126
fold 5	0.018	0	126
fold 6	0.027	0	127
fold 7	0.025	0	126
fold 8	0.028	0	126
fold 9	0.069	0.001	126

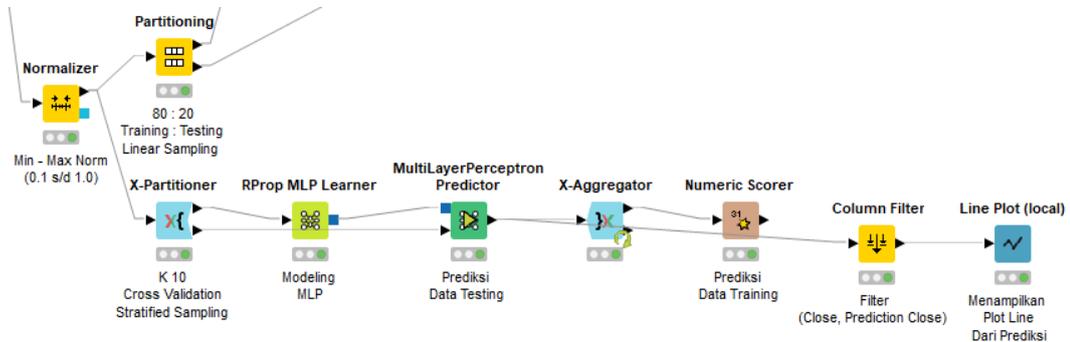
**Gambar 4.55** *Error Rates Cross Validation Model RFR (Tanpa Partitioning)*  
(INDF)

Pada Gambar 4.55 menunjukkan hasil tabel *view error ratse* emiten INDF menggunakan RFR bahwa nilai total *squared error* terendah yaitu ada pada *fold 5* sebesar 0.018 dan tertinggi ada pada *fold 0* sebesar 0.146.

Hasil proses pada Gambar 4.55 di atas merupakan proses terakhir yang selanjutnya akan dilanjutkan dengan proses *scoring* dengan node *Numeric Scorer* bertujuan untuk mendapatkan hasil evaluasi berupa nilai RMSE, MAE, MAPE dan  $R^2$  dari nilai prediksi yang dihasilkan.

### 3. Model Prediksi Metode *Multilayer Perceptron*

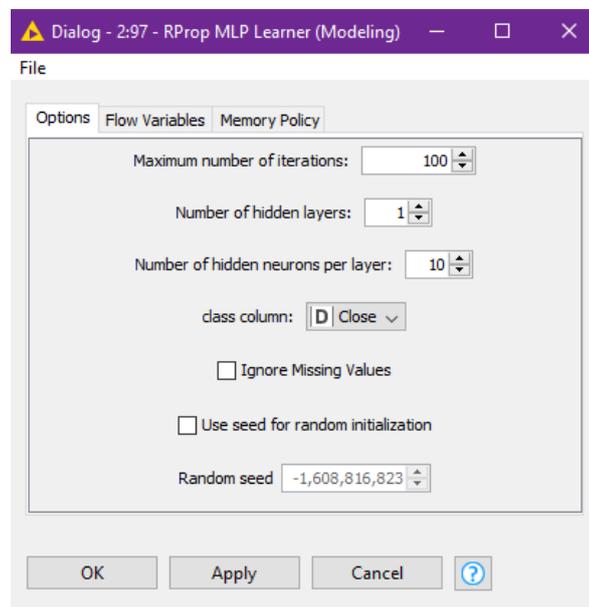
Desain model pengujian data yang langsung dievaluasi dengan *Cross Validation* tanpa adanya *partitioning* (pembagian data *testing* dan data *training*) menggunakan metode MLP dengan dataset ketiga emiten yakni INDF, KAEF dan TLKM adalah sebagai berikut:



**Gambar 4.56** Desain Model Dengan *Cross Validation* MLP (INDF, KAEF, TLKM)

Gambar 4.56 merupakan model MLP yang menunjukkan bahwa proses tahapan dilakukan sama seperti pada pemodelan *Linear Regression* dan RFR di atas sebelumnya. Tahapan untuk uji coba kedua ini setelah proses transformasi data langsung dimodelkan dengan *Cross Validation* menggunakan node *X-Partitioner*. Setelah dilakukannya evaluasi model menggunakan validasi data dengan *Cross Validation* pada keseluruhan dataset berjumlah 1263 per emiten selanjutnya yaitu pemodelan *learner* menggunakan node MLP Learner. Adapun

konfigurasi pada *X-Partitioner* menggunakan nilai K atau *Number of Validation* 10 dengan teknik *Stratified Sampling* berdasarkan atribut yang menjadi *class* prediksinya (*close*). Data yang dihasilkan oleh proses pada node *X-Partitioner* adalah data *training* sebesar 1137 *record* dan data validasi *testing* sebesar 126 *record* yang selanjutnya dilakukan pemodelan dengan node *learnernya* yaitu menggunakan node *MLP Learner*. Adapun data *input* pada node *MLP Learner* adalah data *training* hasil *Cross Validation* berjumlah 1137 *record* dengan konfigurasi yang bisa dilihat pada Gambar 4.57 sebagai berikut:



**Gambar 4.57** Konfigurasi *MLP Learner* Data Hasil *Cross Validation*

Pada Gambar 4.57 konfigurasi di atas, Terdapat beberapa pengaturan yang dapat diubah sesuai kebutuhan diantaranya *Maximum Number of Iterations*, jumlah *hidden layer*, dan jumlah *neuron* per *layernya*. Pada percobaan ini, peneliti juga melakukan percobaan dengan jumlah *hidden layer* sebanyak 1 - 10 dan masing-masing *neuron* pada *layernya* berjumlah 10 *neuron*. Seperti pada uji coba pertama, penggunaan percobaan *hidden layer* 1 hingga 10 sebagai proses menyeleksi hasil dari proses *scorer* yang dilihat pada nilai utama yaitu nilai RMSE.

Penyeleksian dilakukan untuk menentukan atau melihat hasil yang lebih baik setidaknya 3 model *MLP* dengan jumlah *hidden layer* berbeda yang menghasilkan

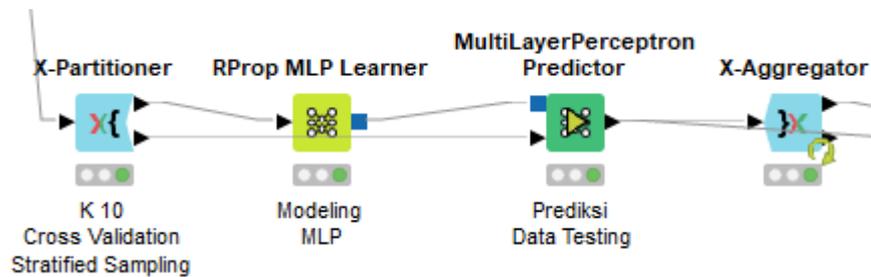
nilai RMSE dengan tingkat yang rendah. Adapun tabel percobaan *hidden layer* tersebut sama seperti pada Tabel 4.10.

Selanjutnya, setelah dilakukan pemodelan MLP menggunakan data *input training* sebanyak 1137 *record* dengan menggunakan 10 kali percobaan dengan jumlah *hidden layer* berbeda, maka model tersebut diimplementasikan sebagai *input* model ke node *predictor* MLP. Prediksi dilakukan menggunakan data *testing* hasil proses *Cross Validation* sebanyak 126 *record*. Adapun hasil prediksi dengan data *testing* 126 *record* emiten INDF dapat dilihat pada Gambar 4.58 sebagai berikut:

Row ID	Open	High	Low	Close	Prediction (Close)
Row14	0.169	0.181	0.166	0.174	0.194
Row20	0.2	0.224	0.204	0.232	0.212
Row25	0.241	0.25	0.241	0.247	0.238
Row28	0.308	0.303	0.295	0.305	0.292
Row36	0.381	0.408	0.375	0.384	0.375
Row42	0.226	0.287	0.23	0.3	0.242
Row48	0.236	0.245	0.214	0.205	0.231
Row51	0.159	0.165	0.15	0.163	0.189
Row66	0.164	0.15	0.14	0.137	0.187
Row78	0.133	0.113	0.114	0.119	0.178
Row86	0.252	0.255	0.246	0.263	0.244
Row93	0.257	0.266	0.257	0.268	0.251
Row101	0.308	0.324	0.289	0.332	0.299
Row103	0.391	0.387	0.359	0.395	0.369
Row133	0.628	0.62	0.605	0.595	0.628
Row151	0.567	0.562	0.572	0.574	0.577
Row156	0.572	0.562	0.588	0.574	0.583
Row162	0.515	0.535	0.519	0.516	0.53
Row167	0.567	0.583	0.567	0.589	0.584
Row171	0.577	0.572	0.572	0.558	0.586
Row175	0.546	0.54	0.54	0.553	0.553
Row189	0.628	0.62	0.626	0.621	0.632
Row231	0.794	0.794	0.743	0.768	0.791
Row237	0.794	0.789	0.754	0.742	0.791
Row252	0.757	0.789	0.77	0.795	0.774

**Gambar 4.58** Hasil Node MLP *Predictor* Data *Testing Cross Validation* (INDF)

Gambar 4.58 di atas merupakan salah satu hasil prediksi pada data *testing* 126 *record* menggunakan model prediksi MLP dengan jumlah *hidden layer* 5 dan *neuron* per *layer* sebanyak 10. Selanjutnya setelah pemodelan data *training* sebesar 1137 *record* dengan MLP *Learner* dan prediksi data *testing* hasil *Cross Validation* sebesar 126 *record* menggunakan MLP *Predictor*, maka terakhir dilakukannya implementasi dengan node *X-Aggregator* sebagai pelengkap penyelesaian proses *Cross Validation* yang dapat dilihat pada Gambar 4.59 berikut :



**Gambar 4.59** Proses *X-Aggregator* Tanpa *Partitioning* Model MLP

Adapun node ini menjadi akhir dari *loop* validasi silang dan harus mengikuti node *X-Partitioner*. Pada proses ini mengumpulkan hasil dari node prediktor, serta membandingkan *class* yang diprediksi dan *real class* dan memberikan *output* prediksi untuk semua baris dan statistik iterasi. Adapun hasil dari proses *X-Aggregator* pada dataset *training* dan *testing* hasil dari proses *X-Partitioner* (1263 *record*) sebagai berikut :

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)
Row 1	0.159	0.165	0.15	0.168	0.168
Row 6	0.123	0.113	0.112	0.119	0.157
Row 15	0.195	0.181	0.177	0.189	0.182
Row 37	0.396	0.393	0.391	0.384	0.4
Row 40	0.35	0.34	0.273	0.263	0.327
Row 47	0.288	0.282	0.252	0.242	0.272
Row 63	0.133	0.176	0.134	0.174	0.164
Row 68	0.143	0.139	0.14	0.142	0.163
Row 75	0.174	0.155	0.145	0.153	0.168
Row 91	0.288	0.276	0.273	0.274	0.28
Row 110	0.468	0.514	0.482	0.489	0.482
Row 118	0.51	0.504	0.508	0.5	0.502
Row 126	0.623	0.636	0.562	0.647	0.609
Row 135	0.597	0.588	0.535	0.521	0.572
Row 138	0.556	0.556	0.567	0.568	0.557
Row 141	0.52	0.525	0.524	0.532	0.518
Row 150	0.592	0.583	0.61	0.595	0.597
Row 168	0.556	0.577	0.551	0.537	0.558
Row 169	0.546	0.567	0.551	0.579	0.551
Row 170	0.582	0.577	0.578	0.584	0.579
Row 172	0.556	0.572	0.567	0.574	0.563
Row 181	0.53	0.54	0.546	0.542	0.534
Row 186	0.551	0.572	0.562	0.574	0.559
Row 199	0.556	0.551	0.54	0.547	0.546
Row 207	0.603	0.593	0.556	0.553	0.584

**Gambar 4.60** *Prediction Table* Hasil *X-Aggregator* MLP Model *Cross Validation* (INDF)

Pada Gambar 4.60 di atas merupakan salah satu hasil percobaan MLP (data *training* data *testing* hasil validasi 10 *Cross Validation* berjumlah 1263 *record*) dengan jumlah *hidden layer* sebanyak 5 yang menunjukkan hasil untuk setiap baris dari nilai yang diprediksi baik. Hasil *error rates* pada proses ini ditunjukkan sebagai berikut :

Row ID	Total squared error	Mean s...	Size of Test Set
fold 0	0.05	0	127
fold 1	0.047	0	127
fold 2	0.048	0	127
fold 3	0.043	0	126
fold 4	0.061	0	126
fold 5	0.043	0	126
fold 6	0.046	0	126
fold 7	0.066	0.001	126
fold 8	0.045	0	126
fold 9	0.065	0.001	126

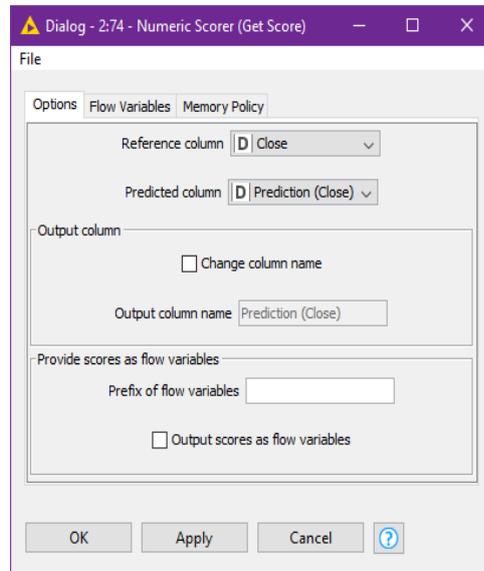
**Gambar 4.61** Error Rates X-Aggregator Model MLP (INDF)

Pada Gambar 4.61 menunjukkan hasil tabel *view error rate* emiten INDF menggunakan model MLP dengan data *testing* sebesar 126 *record* sebagai data prediksi menunjukkan bahwa nilai total *squared error* terendah yaitu ada pada *fold 3* sebesar 0.043 dan tertinggi ada pada *fold 7* sebesar 0.066.

Hasil pada Gambar 4.61 di atas merupakan hasil dari proses terakhir, selanjutnya akan dilanjutkan dengan proses *scoring* dengan node *Numeric Scorer* untuk memperoleh hasil yang dapat dievaluasi berupa nilai RMSE, MAE, MAPE dan  $R^2$  dari nilai prediksi yang dihasilkan.

#### 4.5. Evaluation

Pada penelitian ini, hasil pengujian model *partitioning* dengan model prediksi LR, RFR dan MLP digunakan untuk mengetahui nilai RMSE, MAE, MAPE dan  $R^2$ . *Cross Validation* juga digunakan sebagai model validasi. Untuk menghasilkan keseluruhan nilai *error* pada data yang sudah dimodelkan dengan model prediksi LR, RFR dan MLP di KNIME, node yang digunakan untuk mendapatkan hasil *score error* adalah *Numeric Score*, adapun konfigurasi dari node tersebut sebagai berikut :



**Gambar 4.62** Konfigurasi *Numeric Score*

Pada Gambar 4.62 di atas, untuk mendapatkan hasil perhitungan prediksi yang dapat dievaluasi berupa nilai *error* pada setiap model prediksi (LR, RFR, MLP), maka ditentukannya *reference column* sebagai atribut referensi yaitu atribut *class* nya (*Close*) dan ditetapkannya *Predicted Column* (kolom yang ingin diprediksi) yaitu *Prediction (close)*. Berikut adalah hasil dari RMSE, MAE, MAPE,  $R^2$  dan *line plot* dari masing-masing model prediksi:

#### 4.5.1. Hasil Evaluasi Menggunakan Model Prediksi *Linear Regression*

##### 1. Hasil Evaluasi Dengan Validasi *Partitioning (Cross Validation)*

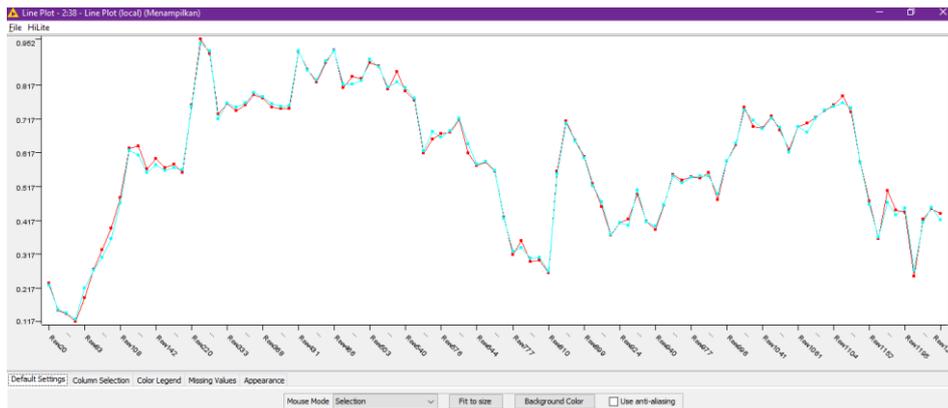
Hasil evaluasi pada tiga dataset menggunakan metode LR dengan model evaluasi validasi *partitioning* dapat dilihat pada tabel berikut :

Tabel 4.15 *Score Error* Dengan Metode LR Validasi *Partitioning*

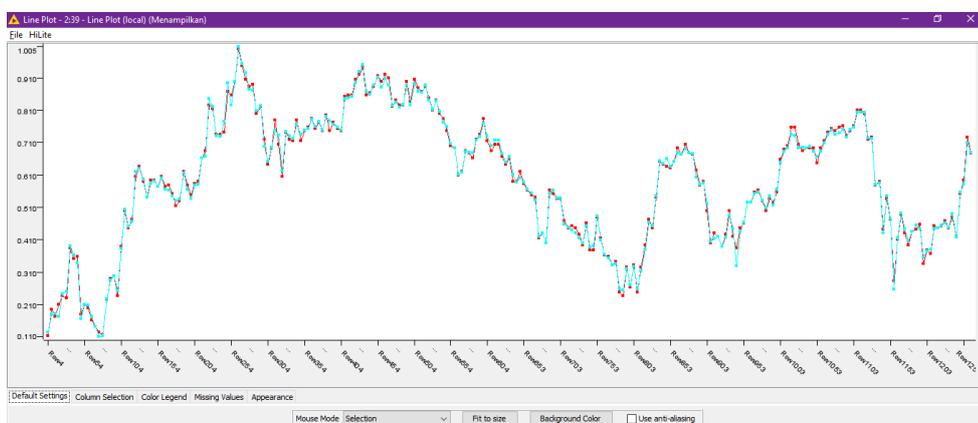
DATASET	DATA TRAINING				DATA TESTING			
	RMSE	MAPE	MAE	R2	RMSE	MAPE	MAE	R <sup>2</sup>
<b>INDF</b>	0.012	0.019	0.009	0.996	0.012	0.018	0.009	0.996
<b>KAEF</b>	0.011	0.013	0.007	0.998	0.010	0.012	0.006	0.998
<b>TLKM</b>	0.011	0.016	0.008	0.997	0.010	0.015	0.007	0.997

Dari Tabel 4.15 di atas, hasil evaluasi dari setiap dataset tiga emiten, menunjukkan bahwa pada hasil data *testing* memiliki nilai RMSE yang lebih rendah dibandingkan dengan hasil RMSE pada data *training*. Adapun nilai

terkecil RMSE ada pada uji coba LR dengan dataset emiten KAEF dan TLKM yang menghasilkan nilai sebesar 0.010. Artinya dengan menggunakan modelisasi pada data *training* menggunakan *Cross Validation* akan berpengaruh terhadap nilai *error* pada prediksi data *testing*. Berikut salah satu *Line Plot* pada hasil prediksi data *training* dan data *testing* pada emiten INDF sebagai berikut:



**Gambar 4.63** *Line Plot* Prediksi Data *Training* Metode *LR Partitioning* (INDF)



**Gambar 4.64** *Line Plot* Prediksi Data *Testing* Metode *LR Partitioning* (INDF)

Pada *line plot* Gambar 4.63 dan Gambar 4.64 di atas, garis dengan warna merah merupakan *line* dari *class* (*close*) dan garis warna hijau menunjukkan hasil prediksi, dan dapat dilihat dari tampilan *line plot* di atas bahwa pada hasil *line plot* data *testing* pada garis hijau menunjukkan prediksi yang tidak begitu jauh dari *line classnya*.

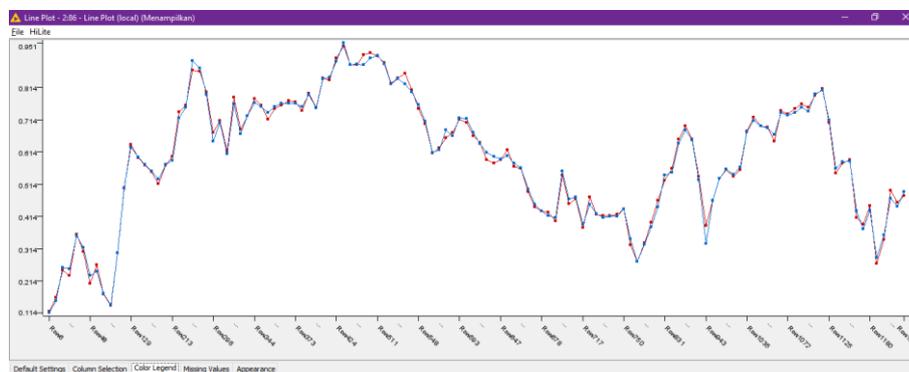
## 2. Hasil Evaluasi Dengan Validasi *Cross Validation* (Tanpa *Partitioning*)

Hasil evaluasi pada tiga dataset menggunakan metode LR dengan model evaluasi validasi *Cross Validation* (tanpa *partitioning* data) dapat dilihat pada Tabel 4.16 berikut :

Tabel 4.16 *Score Error* Dengan Metode LR Dengan *Cross Validation*

DATASET	DATA TESTING			
	RMSE	MAPE	MAE	R2
INDF	0.012	0.019	0.009	0.996
KAEF	0.011	0.013	0.007	0.998
TLKM	0.010	0.016	0.008	0.997

Adapun pada hasil evaluasi menggunakan model prediksi LR dengan validasi *Cross Validation* (tanpa *partitioning*) yang ditunjukkan pada Tabel 4.16 di atas, menunjukkan bahwa hasil yang bila dibandingkan dengan hasil *error score* pada data *testing* model validasi *partitioning* menunjukkan pada emiten INDF dan TLKM memiliki kesamaan *score* RMSE, dan emiten KAEF dengan nilai *error* yang naik dari sebelumnya 0.010 menjadi 0.011 sehingga jika dibandingkan maka hasil *score* data *testing* pada model validasi *partitioning* lebih baik jika dibandingkan dengan hasil evaluasi nilai *error* pada model validasi *Cross Validation* tanpa *partitioning*. Adapun nilai *score* pada MAPE dan MAE mengalami perubahan antar kedua hasil uji coba di atas. Berikut *line plot* pada hasil prediksi data *testing* dengan validasi *Cross Validation* (tanpa *partitioning*) sebagai berikut:



**Gambar 4.65** *Line Plot* Prediksi Data Testing Metode LR *Cross Validation* (INDF)

#### 4.5.2. Hasil Evaluasi Menggunakan Model Prediksi RFR

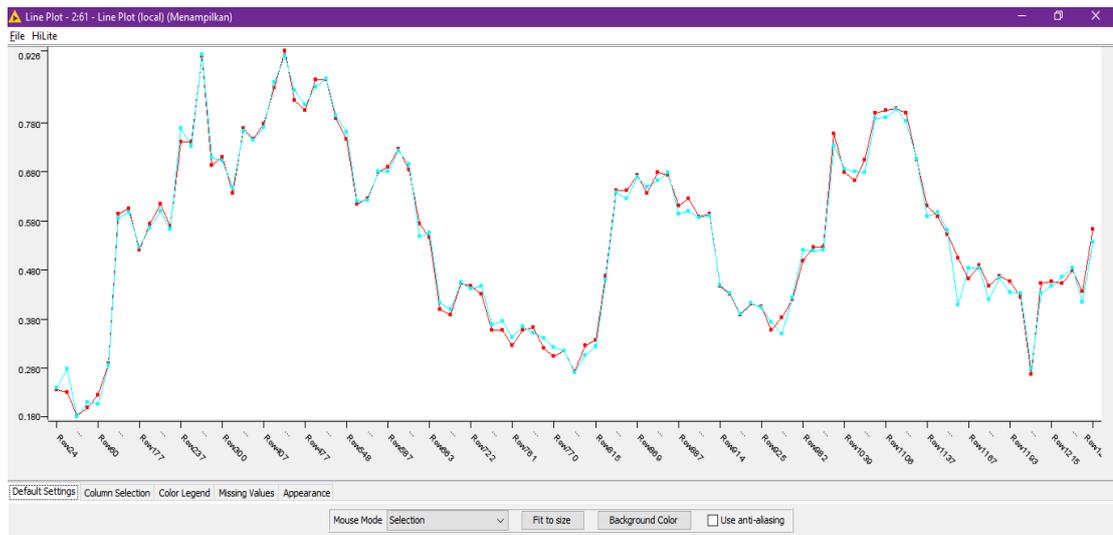
##### 1. Hasil Evaluasi Dengan Validasi *Partitioning* (*Cross validation*)

Hasil evaluasi pada tiga dataset menggunakan metode RFR dengan model evaluasi validasi *partitioning* dapat dilihat pada Tabel 4.17 berikut :

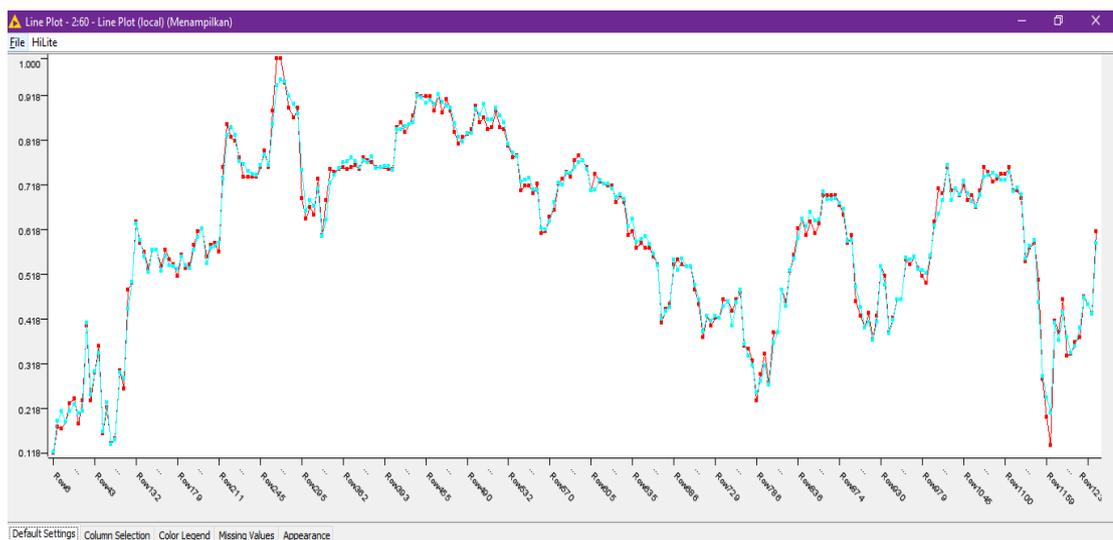
Tabel 4.17 *Score Error* Dengan Metode RFR Validasi *Partitioning*

DATASET	DATA TRAINING				DATA TESTING			
	RMSE	MAPE	MAE	R2	RMSE	MAPE	MAE	R <sup>2</sup>
INDF	0.016	0.025	0.012	0.993	0.017	0.025	0.012	0.992
KAEF	0.013	0.017	0.008	0.997	0.012	0.016	0.007	0.997
TLKM	0.014	0.021	0.010	0.995	0.012	0.020	0.009	0.996

Dari hasil Tabel 4.17 di atas, merupakan hasil evaluasi dari setiap dataset ketiga emiten yang menunjukkan bahwa pada hasil data *testing* memiliki nilai RMSE yang lebih rendah pada emiten KAEF dan TLKM sebesar 0.012 dibandingkan dengan hasil RMSE pada data *training* sebesar 0.013 dan 0.014. Sedangkan pada emiten INDF nilai *error* terendah ada pada hasil pengujian data *training*. Namun dapat dikatakan, pada metode RFR dengan mengujikan modelisasi pada data *training* menggunakan *cross validation* dapat berpengaruh terhadap nilai *error* pada prediksi data *testing*. Berikut salah satu tampilan *Line Plot* pada hasil prediksi data *training* dan data *testing* pada emiten INDF sebagai berikut:



**Gambar 4.66** Line Plot Prediksi Data Training Metode RFR Partitioning (INDF)



**Gambar 4.67** Line Plot Prediksi Data Testing Metode RFR Partitioning (INDF)

Pada Gambar 4.66 dan Gambar 4.67 di atas, garis dengan warna merah merupakan *line* dari *class* (*close*) dan garis warna hijau menunjukkan hasil prediksi, dan dapat dilihat dari tampilan *plot* di atas bahwa pada hasil *line plot* data *testing* pada garis hijau menunjukkan prediksi yang tidak begitu jauh dari *line classnya*.

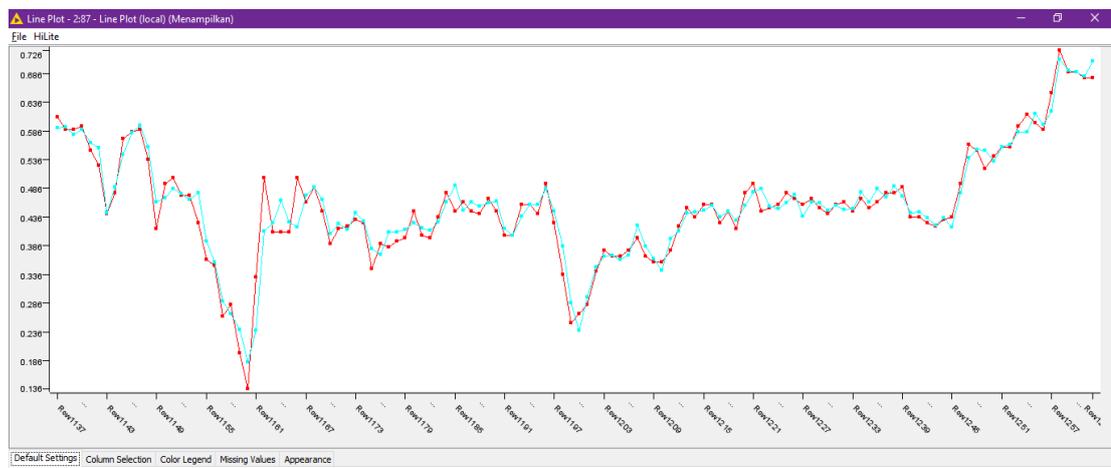
## 2. Hasil Evaluasi Dengan Validasi *Cross Validation* (Tanpa *Partitioning*)

Hasil evaluasi pada tiga dataset menggunakan metode RFR dengan model evaluasi validasi *Cross Validation* (tanpa *partitioning* data) dapat dilihat pada Tabel 4.18 sebagai berikut :

Tabel 4.18 *Score Error* Dengan Metode RFR Validasi *Cross Validation*

DATASET	DATA TESTING			
	RMSE	MAPE	MAE	R2
INDF	0.018	0.034	0.013	0.991
KAEF	0.013	0.017	0.008	0.997
TLKM	0.014	0.021	0.010	0.995

Pada Tabel 4.18 di atas, adapun hasil evaluasi menggunakan model prediksi RFR dengan validasi *Cross Validation* (tanpa *partitioning*) memberikan *output* hasil yang jika dibandingkan dengan hasil *error score* pada data *testing* model validasi *partitioning*, menunjukkan bahwa hasil *error score* tiga emiten memiliki nilai *error* yang rendah pada RMSE nya pada pemodelan RFR dengan *partitioning*, maka dapat dikatakan hasil *score* data *testing* pada model validasi *partitioning* lebih baik jika dibandingkan dengan hasil evaluasi nilai *error* pada model validasi *cross validation* tanpa *partitioning*. Dan pada nilai *score* MAPE, MAE serta R<sup>2</sup> pun mengalami perubahan hasil uji coba sebelumnya yang dapat dilihat pada tabel di atas. Berikut *line plot* pada hasil prediksi data *testing* dengan validasi *Cross Validation* (tanpa *partitioning*) sebagai berikut:



**Gambar 4.68** Line Plot Prediksi Data Testing Metode RFR Cross Validation (INDF)

### 4.5.3. Hasil Evaluasi Menggunakan Model Prediksi MLP

#### 1. Hasil Evaluasi Dengan Validasi *Partitioning* (Cross validation)

Hasil evaluasi pada ketiga dataset menggunakan metode *Multilayer Perceptron* dengan model evaluasi validasi *partitioning* dapat dilihat pada Gambar 4.69 berikut :

Data	Hidden Layer	Data Training				Data Testing			
		RMSE	MAPE	MAE	R2	RMSE	MAPE	MAE	R <sup>2</sup>
INDF	5 Hidden Layer	0.020	0.032	0.014	0.990	0.020	0.032	0.015	0.990
KAEF		0.017	0.025	0.011	0.994	0.016	0.027	0.010	0.995
TLKM		0.019	0.030	0.014	0.991	0.018	0.028	0.013	0.992
INDF	6 Hidden Layer	0.020	0.031	0.014	0.990	0.020	0.031	0.015	0.990
KAEF		0.017	0.024	0.011	0.994	0.015	0.022	0.009	0.996
TLKM		0.019	0.030	0.014	0.991	0.017	0.030	0.013	0.993
INDF	7 Hidden Layer	0.021	0.033	0.015	0.988	0.020	0.033	0.015	0.990
KAEF		0.019	0.026	0.011	0.993	0.014	0.023	0.009	0.996
TLKM		0.020	0.033	0.015	0.990	0.019	0.031	0.014	0.991

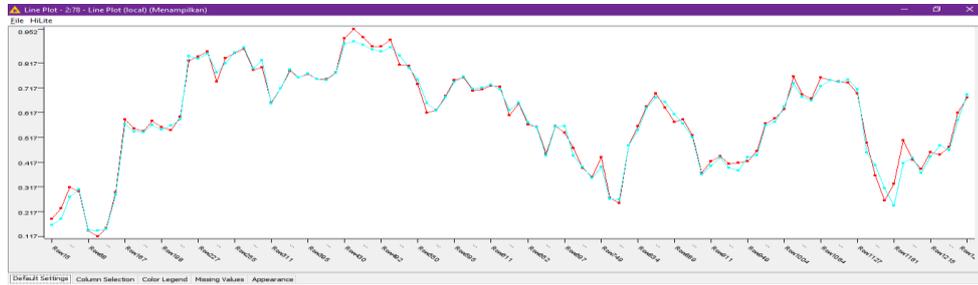
**Gambar 4.69** Score Error Dengan Metode MLP Validasi *Partitioning*

Dari Gambar 4.69 di atas, merupakan hasil *error score* pada pemodelan MLP menggunakan percobaan penginputan 1 – 10 hidden layer. Dari percobaan yang dilakukan, didapat hasil dengan tingkat *error* yang rendah pada pemodelan dengan hidden layer berjumlah 5, 6 dan 7 hidden layer. hasil evaluasi pada Gambar 4.69 merupakan hasil dari setiap dataset dari tiga emiten.

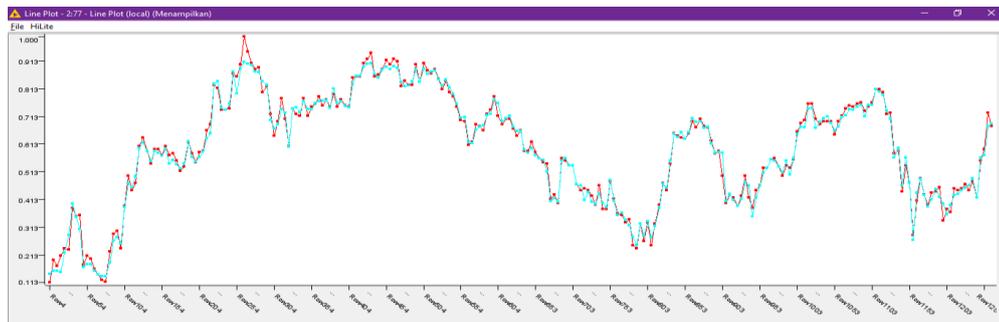
Pada Gambar 4.69 di atas menunjukkan bahwa pada hasil data *testing* rata-rata memiliki nilai RMSE yang lebih rendah dibandingkan dengan hasil RMSE pada data *training*. Adapun nilai terkecil RMSE ada pada uji coba MLP dengan dataset emiten KAEF yang menghasilkan nilai sebesar 0.014 dengan jumlah *hidden layer* sebesar 7.

Dapat diartikan bahwa dengan melakukan uji coba modelisasi pada data *training* menggunakan *Cross Validation* akan berpengaruh terhadap nilai *error* pada prediksi data *testing*. Berikut merupakan salah satu tampilan *line plot* pada

hasil prediksi data *training* dan data *testing* pada emiten INDF menggunakan metode MLP sebagai berikut:



**Gambar 4.70** Line Plot Prediksi Data Training Metode *Partitioning* MLP Dengan *Hidden Layer* 5 (INDF)



**Gambar 4.71.**Line Plot Prediksi Data *Testing* Metode *Partitioning* MLP Dengan *Hidden Layer* 5 (INDF)

Pada kedua gambar di atas (Gambar 4.69 dan Gambar 4.70), garis dengan warna merah merupakan *line* dari *class* (*close*) dan garis warna hijau menunjukkan hasil prediksi, dapat dilihat dari tampilan *plot line* di atas pada hasil data *testing* sedikit terjadi perbedaan dimana ada bagian *line* hijau berada sedikit jauh pada titik *line* merah pada prediksi data di *range* antara *row* 254 ke *row* 304.

## 2. Hasil Evaluasi Dengan Validasi *Cross Validation* (Tanpa *Partitioning*)

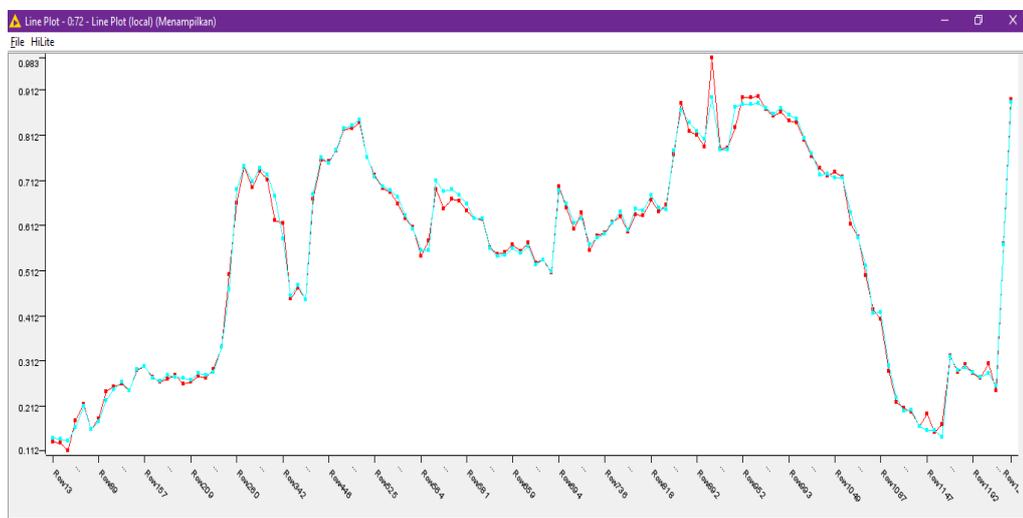
Hasil evaluasi pada tiga dataset menggunakan metode MLP dengan model evaluasi validasi *Cross Validation* (tanpa *partitioning* data) dapat dilihat pada Tabel 4.19 berikut :

Tabel 4.19 *Score Error* Dengan Metode MLP Validasi *Cross Validation*

Dataset	<i>Hidden Layer</i>	Data <i>Testing</i>
---------	---------------------	---------------------

		RMSE	MAPE	MAE	R2
INDF	5 Hidden Layer	0.020	0.033	0.015	0.990
KAEF		0.017	0.025	0.011	0.995
TLKM		0.018	0.029	0.013	0.992
INDF	6 Hidden Layer	0.020	0.032	0.015	0.989
KAEF		0.017	0.023	0.011	0.995
TLKM		0.019	0.029	0.014	0.991
INDF	7 Hidden Layer	0.021	0.033	0.015	0.989
KAEF		0.017	0.026	0.011	0.994
TLKM		0.019	0.030	0.014	0.991

Pada Tabel 4.19 di atas, percobaan pun dilakukan sama seperti validasi *partitioning* menggunakan MLP yaitu dengan jumlah *hidden layer* 1 hingga 10 *hidden layer*, dan nilai RMSE yang lebih kecil terdapat pada hasil prediksi dengan 5, 6 dan 7 hidden layer. Pada hasil evaluasi Tabel 4.19 di atas cukup mengalami perubahan pada masing-masing *score error* nya sebesar 0.01 hingga 0.02 lebih tinggi jika dibandingkan dengan hasil *error score* pada data *testing* model validasi *partitioning*. hal ini dapat menunjukkan bahwa hasil *error score* pada tiga emiten memiliki nilai *error* yang rendah pada RMSE nya di pemodelan MLP dengan validasi *partitioning*. Adapun pada nilai *score* MAPE, MAE serta  $R^2$  mengalami perubahan dengan hasil uji coba sebelumnya yang dapat dilihat pada Tabel 4.19 di atas. Berikut *line plot* pada hasil prediksi data *testing* dengan validasi *Cross Validation* (tanpa *partitioning*) menggunakan metode MLP sebagai berikut:



**Gambar 4.72** *Line Plot* Prediksi Data *Testing Cross Validation* (Tanpa *Partitioning*) MLP Dengan 7 *Hidden Layer* (KAEF)

#### 4.5.4. Rangkuman Hasil Evaluasi

Adapun hasil dari percobaan dengan menggunakan 3 model *prediction* yaitu *Linear Regression*, *Random Forest Regression* dan *Multilayer Perceptron* yang menghasilkan RMSE, MAPE, MAE dan  $R^2$  yang telah dilakukan dengan penggunaan dua uji coba model validasi yaitu uji coba pertama model *partitioning* yang divalidasikan dengan *Cross Validation* dan uji coba kedua dengan validasi *Cross Validation* tanpa *partitioning* dapat ditarik kesimpulan sebagai berikut:

##### 1. Hasil Evaluasi Dengan Metode *Partitioning* dan *Cross Validation*

Berikut rangkuman hasil dari uji coba pertama menggunakan *partitioning* yang divalidasikan dengan *Cross Validation* dapat dilihat pada Gambar 4.73 sebagai berikut:

Data	Metode	Data Training				Data Testing				Keterangan	
		RMSE	MAPE	MAE	R2	RMSE	MAPE	MAE	R2		
INDF	LR	0.012	0.019	0.009	0.996	0.012	0.018	0.009	0.996		
	RFR	0.016	0.025	0.012	0.993	0.017	0.025	0.012	0.992		
	MLP		0.020	0.032	0.014	0.990	0.020	0.032	0.015	0.990	5 Hidden Layer
			0.020	0.031	0.014	0.990	0.020	0.031	0.015	0.990	6 Hidden Layer
			0.021	0.033	0.015	0.988	0.020	0.033	0.015	0.990	7 Hidden Layer
KAEF	LR	0.011	0.013	0.007	0.998	0.010	0.012	0.006	0.998		
	RFR	0.013	0.017	0.008	0.997	0.012	0.016	0.007	0.997		
	MLP		0.017	0.025	0.011	0.994	0.016	0.027	0.010	0.995	5 Hidden Layer
			0.017	0.024	0.011	0.994	0.015	0.022	0.009	0.996	6 Hidden Layer
			0.019	0.026	0.011	0.993	0.014	0.023	0.009	0.996	7 Hidden Layer
TLKM	LR	0.011	0.016	0.008	0.997	0.010	0.015	0.007	0.997		
	RFR	0.014	0.021	0.010	0.995	0.012	0.020	0.009	0.996		
	MLP		0.019	0.030	0.014	0.991	0.018	0.028	0.013	0.992	5 Hidden Layer
			0.019	0.030	0.014	0.991	0.017	0.030	0.013	0.993	6 Hidden Layer
			0.020	0.033	0.015	0.990	0.019	0.031	0.014	0.991	7 Hidden Layer

Gambar 4.73 Perbandingan RMSE, MAPE, MAE dan  $R^2$

## 2. Hasil Evaluasi Dengan Validasi *Cross validation* (Tanpa *Partitioning*)

Berikut rangkuman hasil dari uji coba kedua menggunakan *Cross Validation* tanpa *partitioning* dapat dilihat pada Tabel 4.20 sebagai berikut:

Tabel 4.20 Perbandingan RMSE, MAPE, MAE dan  $R^2$

DATASET	METODE	DATA TESTING				KET
		RMSE	MAPE	MAE	$R^2$	
INDF	LR	0.012	0.019	0.009	0.996	
	RFR	0.018	0.034	0.013	0.991	
	MLP	0.020	0.033	0.015	0.990	5 Hidden Layer
		0.020	0.032	0.015	0.989	6 Hidden Layer
		0.021	0.033	0.015	0.989	7 Hidden Layer
KAEF	LR	0.011	0.013	0.007	0.998	
	RFR	0.013	0.017	0.008	0.997	
	MLP	0.017	0.025	0.011	0.995	5 Hidden Layer
		0.017	0.023	0.011	0.995	6 Hidden Layer
		0.017	0.026	0.011	0.994	7 Hidden Layer
TLKM	LR	0.010	0.016	0.008	0.997	
	RFR	0.014	0.021	0.010	0.995	
	MLP	0.018	0.029	0.013	0.992	5 Hidden Layer
		0.019	0.029	0.014	0.991	6 Hidden Layer
		0.019	0.030	0.014	0.991	7 Hidden Layer

Hasil rangkuman perbandingan *score error* pada Gambar 4.73 di atas menghasilkan kesimpulan dari percobaan dengan tiga model prediksi (LR, RFR, MLP) menggunakan model validasi *partitioning* sebagai pembagian data *training* dan data *testing* yang divalidasi kan dengan *Cross Validation* yaitu bahwa pada ketiga model algoritma yang diuji cobakan pada setiap dataset ditemukan nilai *error* terendah ada pada model prediksi *Linear Regression* sebesar 0.012 untuk emiten INDF, 0.011 untuk emiten KAEF dan TLKM, dan selanjutnya pemodelan dengan metode RFR menghasilkan RMSE sebesar 0.013 pada emiten KAEF, sebesar 0.014 pada dataset TLKM dan sebesar 0.016 pada dataset INDF, hasil

prediksi tersebut cukup baik berupa nilai *error* yang cukup rendah dibandingkan dengan metode MLP.

Adapun pada rangkuman perbandingan *score error* pada tabel 4.20 di atas menghasilkan kesimpulan dari percobaan dengan tiga model prediksi (LR, RFR, MLP) menggunakan model validasi dengan *Cross Validation* tanpa adanya pembagian data melalui *partitioning* yaitu bahwa pada ketiga model algoritma yang diuji cobakan pada setiap dataset emiten ditemukan nilai *error* terendah ada pada model prediksi *Linear Regression* sebesar 0.010 untuk emiten TLKM, 0.011 untuk emiten KAEF dan RMSE sebesar 0.012 pada emiten INDF, dan selanjutnya pemodelan dengan metode RFR menghasilkan RMSE sebesar 0.013 pada emiten KAEF, sebesar 0.014 pada dataset TLKM dan sebesar 0.018 pada dataset INDF, hasil prediksi tersebut cukup baik karena nilai *error* yang cukup rendah dibandingkan dengan metode MLP.

Dalam pemodelan yang sudah dilakukan dan analisa selama proses uji coba, jumlah dataset yang digunakan pada percobaan dengan model prediksi dapat mempengaruhi hasil *error score* pada penelitian. Dilihat dari hasil evaluasi di atas, jumlah data *testing* yang cukup banyak *recordnya* dapat menghasilkan nilai *error* lebih rendah pada suatu model prediksi. Namun *error score* juga dapat dipengaruhi oleh proses validasi menggunakan *Cross Validation*, hal ini dapat membuat hasil *error score* menjadi lebih akurat.

## BAB V PENUTUP

### 5.1. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan dengan tema pembahasan analisa perbandingan model *prediction* dalam memprediksi historis harga saham pada tiga sektor industri di Indonesia, maka dapat disimpulkan bahwa :

1. Model *prediction* yang telah digunakan dalam memprediksi histori harga saham adalah *Linear Regression*, *Random Forest Regression* dan *MultiLayer Perceptron*. Pada tiga model prediksi ini mampu memberikan nilai *error* prediksi yang cukup baik pada uji coba dataset pada setiap jenis emitennya dengan rata-rata nilai *Root Mean Square Error*nya (RMSE) di range 0.010 sebagai nilai terendah hingga sebesar 0.020 sebagai nilai tertingginya.
2. Pada penelitian ini digunakan 2 model validasi sebagai proses uji coba dari masing-masing model *prediction*nya. Pertama menggunakan validasi *partitioning* (pembagian data *training* dan *testing* sebesar 80 : 20) yang dalam uji coba ini juga disertakan *Cross Validation* sebagai pengevaluasi kinerja setiap model atau algoritma dengan tujuan untuk menemukan kombinasi data terbaik serta memperoleh hasil akurasi berupa *error score* yang lebih maksimal. Kedua menggunakan validasi *Cross Validation* langsung terhadap jumlah kesatuan dataset tanpa adanya *partitioning*.
3. Dalam keseluruhan uji coba pada dataset dengan model prediksi LR, RFR, MLP serta kedua model validasi yang dilakukan pada setiap model prediksi. Dapat dilihat dari rangkuman hasil penelitian, bahwa metode atau model prediksi *Linear Regression* mampu memberikan nilai RMSE yang cukup rendah pada setiap jenis dataset emitennya dibandingkan dengan metode atau model prediksi RFR dan juga MLP, dengan nilai RMSE sebesar 0.010 paling rendah dan 0.012 paling tinggi. Adapun *error score* lainnya yaitu nilai MAPE terendah sebesar 1,2% dan tertinggi sebesar 1,9 %, nilai MAE terendah 0.006 dan tertinggi sebesar 0.009, serta nilai koefisien determinasi ( $R^2$ ) tertinggi sebesar 99,8% dan terendah sebesar 99,6% pada model prediksi LR.

4. Pada model prediksi RFR memiliki nilai *error* yang cukup kecil dibandingkan dengan MLP. Pada model prediksi RFR nilai RMSE terkecil sebesar 0.012 dan tertinggi sebesar 0.018, nilai MAPE terendah sebesar 1,6% dan tertinggi sebesar 3,4%, nilai MAE terkecil sebesar 0.007 dan tertinggi sebesar 0.013, nilai  $R^2$  tertinggi sebesar 99,7% dan terendah sebesar 99,1%. Sedangkan pada model prediksi MLP nilai RMSE terkecil sebesar 0.014 dan tertinggi sebesar 0.020, nilai MAPE terendah sebesar 2,2% dan tertinggi sebesar 3,3%, nilai MAE terkecil sebesar 0,009 dan tertinggi sebesar 0,015, nilai  $R^2$  tertinggi sebesar 99,6% dan terendah sebesar 98,9%.

## 5.2. Saran

Adapun saran yang dapat diberikan berdasarkan penelitian yang sudah dilakukan diantaranya sebagai berikut :

1. Dalam menganalisa harga saham kedepannya dapat menggunakan perbandingan model prediksi yang lainnya seperti menggunakan *Logistic Linear*, *SVM Regression* atau metode lain selain LR, RFR dan MLP supaya dapat diketahui *score error* yang berbeda dari penelitian ini.
2. Dapat menggunakan metode pemilihan fitur seperti *F-Score* atau PSO untuk mengetahui seberapa berpengaruhnya pemilihan fitur ini terhadap *score error* yang dihasilkan nantinya.
3. Menggunakan dataset yang berbeda serta penggunaan dataset dengan jumlah yang lebih besar, hal ini dikarenakan pada jumlah *instance* dataset yang lebih besar dapat mempengaruhi hasil *error score* cukup baik yang dapat dievaluasi dibandingkan dengan jumlah dataset yang sedikit.
4. Pada penelitian selanjutnya diharapkan dapat dilakukannya tahap *deployment* program terkait hasil evaluasi yang didapatkan berdasarkan metode atau model prediksi yang digunakan, hal ini dikarenakan pada penelitian yang dilakukan ini hanya pada sampai tahap evaluasi.

## DAFTAR PUSTAKA

- [1] Reza Maulana and Devy Kumalasari, “Analisis Dan Perbandingan Algoritma *Data mining* Dalam Prediksi Harga Saham Ggrm,” *J. Inform. Kaputama*, vol. 3, no. 1, pp. 22–28, 2019, [Online]. Available: <https://finance.yahoo.com/quote/GGRM.J>.
- [2] W. Y. Rusyida and V. Y. Pratama, “Prediksi Harga Saham Garuda Indonesia di Tengah Pandemi Covid-19 Menggunakan Metode ARIMA,” *Sq. J. Math. Math. Educ.*, vol. 2, no. 1, p. 73, 2020, doi: 10.21580/square.2020.2.1.5626.
- [3] H. Wang, “Stock price prediction based on machine learning approaches,” *ACM Int. Conf. Proceeding Ser.*, 2020, doi: 10.1145/3414274.3414275.
- [4] Saifuddin and A. Hermawan, “Prediksi Data Historis Saham PT . Bank Rakyat Indonesia Tbk ( BBRI ) Menggunakan Model Algoritma Artificial Neural Network,” pp. 139–144, 2019.
- [5] R. Maulana and D. Kumalasari, “Analisis Dan Perbandingan Algoritma *Data mining* Dalam Prediksi Harga Saham Ggrm,” *J. Inform. Kaputama*, vol. 3, no. 1, pp. 22–28, 2019, [Online]. Available: <https://finance.yahoo.com/quote/GGRM.J>.
- [6] R. I. Nurachim, “Pemilihan Model Prediksi Indeks Harga Saham Yang Dikembangkan Berdasarkan Algoritma Support Vector Machine ( Svm ) Atau Multilayer Perceptron ( Mlp ) Studi Kasus : Saham Pt Telekomunikasi Indonesia Tbk,” vol. 5, no. 1, pp. 29–35, 2019.
- [7] A. A. Fauzi, A. M. Soleh, and A. Djuraidah, “Kuadrat Terkecil Parsial , Support Vector Machine Dan Random Forest,” pp. 203–215, 2020.
- [8] H. W. Herwanto, T. Widiyaningtyas, and P. Indriana, “Penerapan Algoritme *Linear Regression* untuk Prediksi Hasil Panen Tanaman Padi,” *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 8, no. 4, p. 364, 2019, doi: 10.22146/jnteti.v8i4.537.
- [9] L. . Dr. Mas Rahmah, S.H., M.H., *Hukum Pasar Modal*. Rawamangun, Jakarta: Prenada Media Group, 2019.
- [10] R. Indonesia, “2. Capital Markets,” *Cover. Glob.*, pp. 17–31, 2016, doi:

10.7312/schi13174-003.

- [11] S. Bakhri, “Minat Mahasiswa Dalam Investasi Di Pasar Modal,” *Al-Amwal J. Ekon. dan Perbank. Syari’ah*, vol. 10, no. 1, p. 146, 2018, doi: 10.24235/amwal.v10i1.2846.
- [12] F. P. P. Abi, *Semakin Dekat dengan Pasar Modal Indonesia*. Yogyakarta: Deepublish (Group Penerbit CV Budi Utama), 2016.
- [13] M. Dr. Sri Handini, *Buku Ajar: Manajemen Keuangan*. Surabaya: Scopindo Media Pustaka, 2020.
- [14] I. Angraini and I. Yusra, “Pendekatan Data Panel Terhadap Return Saham : Studi Kasus Pada Perusahaan LQ45,” *Japanese J. Allergol.*, vol. 49, 2019.
- [15] D. Gibtiah, *Fikih Kontemporer*. Jakarta: Prenada Media Group, 2016.
- [16] H. M. Hasanuddin and H. J. Mubarok, *Perkembangan Akad Musyarakah*. Jakarta: Kencana Prenada Media Group, 2018.
- [17] M. S. Dr. Musdalifah Azis, S.E., M. S. Prof. Dr. Sri Mintarti, and M. S. Maryam Nadir, S.E., *Manajemen Investasi Fundamental, Teknikal, Perilaku Investor dan Return Saham*. Yogyakarta: Deepublish (Group Penerbit CV Budi Utama), 2015.
- [18] J. Hartono, *Teori Portofolio dan Analisis Investasi*. Yogyakarta: BPFE-Yogyakarta, 2016.
- [19] N. Christian and F. Frecky, “Analisis Pengaruh Faktor-Faktor Yang Mempengaruhi Harga Saham Pada Perusahaan Yang Terdaftar Di Bursa Efek Indonesia,” *J. Benefita*, vol. 1, no. 1, p. 115, 2019, doi: 10.22216/jbe.v1i1.3417.
- [20] D. Purwaningsih and T. Sulistiyani, “Analisis Faktor Faktor Yang Mempengaruhi Indeks Harga Saham Gabungan Di Bursa Efek Indonesia Periode 2012-2014,” *J. Manaj. Bisnis*, vol. 13, no. 1, pp. 1–18, 2018.
- [21] N. A. Hasibuan *et al.*, “Implementasi *Data mining* Untuk Pengaturan Layout,” vol. 4, no. 4, pp. 6–11, 2017.
- [22] E. Buulolo, *Data mining Untuk Perguruan Tinggi*. Yogyakarta: Deepublish (Group Penerbit CV Budi Utama), 2020.
- [23] A. Bastian, H. Sujadi, and G. Febrianto, “Penerapan Algoritma K-Means

- Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka),” no. 1, pp. 26–32.
- [24] M. Fauzi *et al.*, “Penerapan Algoritma K-Means Clustering Untuk Mendeteksi Penyebaran Penyakit TBC ( Studi Kasus : Di Kabupaten Deli Serdang ) Abstrak,” vol. 1, no. 2, 2017.
- [25] K. Fatmawati and A. P. Windarto, “*Data mining*: Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (Dbd) Berdasarkan Provinsi,” *Comput. Eng. Sci. Syst. J.*, vol. 3, no. 2, p. 173, 2018, doi: 10.24114/cess.v3i2.9661.
- [26] Nafisah Nurul Hakim, “Implementasi Machine Learning pada Sistem Prediksi Kejadian dan Lokasi Patah Rel Kereta Api di Indonesia,” *J. Sist. Cerdas*, vol. 3, no. 1, pp. 25–35, 2020, doi: 10.37396/jsc.v3i1.58.
- [27] A. Izzah, “Prediksi Harga Saham Menggunakan Improved Multiple *Linear Regression* untuk Pencegahan Data Outlier,” *Kinetik*, vol. 2, no. 3, pp. 141–149, 2017, doi: 10.22219/kinetik.v2i3.268.
- [28] B. Putro, M. T. Furqon, and S. H. Wijoyo, “Prediksi Jumlah Kebutuhan Pemakaian Air Menggunakan Metode Exponential Smoothing ( Studi Kasus : PDAM Kota Malang ),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 11, pp. 4679–4686, 2018.
- [29] R. Yanto, “Implementasi *Data mining* Estimasi Ketersediaan Lahan Pembuangan Sampah menggunakan Algoritma Simple *Linear Regression*,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 2, no. 1, pp. 361–366, 2018, doi: 10.29207/resti.v2i1.282.
- [30] F. Nurzaman, “Penerapan Algoritma Regresi Linier Untuk Prediksi Jumlah Klaim Pada Asuransi Kesehatan,” *Semin. Nas. Teknol. Inf.*, pp. 105–114, 2017.
- [31] L. H. Tresnawati, W. A. Kusuma, S. H. Wijaya, and L. S. Hasibuan, “Asosiasi Single Nucleotide Polymorphism pada Diabetes Mellitus Tipe 2 Menggunakan *Random Forest Regression*,” *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 8, no. 4, p. 320, 2019, doi: 10.22146/jnteti.v8i4.531.
- [32] S. Solikhun, M. Safii, and A. Trisno, “Jaringan Saraf Tiruan Untuk

- Memprediksi Tingkat Pemahaman Sisiwa Terhadap Matapelajaran Dengan Menggunakan Algoritma Backpropagation,” *J-SAKTI (Jurnal Sains Komput. dan Inform.*, vol. 1, no. 1, p. 24, 2017, doi: 10.30645/j-sakti.v1i1.26.
- [33] A. Revi, S. Solikhun, and M. Safii, “Jaringan Syaraf Tiruan Dalam Memprediksi Jumlah Produksi Daging Sapi Berdasarkan Provinsi,” *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 297–304, 2018, doi: 10.30865/komik.v2i1.941.
- [34] T. F. Hadimarta, R. R. Muhima, and M. Kurniawan, “Implementasi Multilayer Perceptron Pada Jaringan Saraf Tiruan Untuk Memprediksi Nilai Valuta Asing,” *INTEGER J. Inf. Technol.*, vol. 5, no. 1, pp. 56–63, 2020, doi: 10.31284/j.integer.2020.v5i1.909.
- [35] M. Fajar and U. Padjadjaran, “Meningkatkan Akurasi Peramalan dengan Menggunakan Metode Hybrid Singular Spectrum Analysis-Multilayer Perceptron Neural Networks,” no. February, 2018, doi: 10.13140/RG.2.2.34999.01443.
- [36] D. A. Nasution, H. H. Khotimah, and N. Chamidah, “Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN,” *Comput. Eng. Sci. Syst. J.*, vol. 4, no. 1, p. 78, 2019, doi: 10.24114/cess.v4i1.11458.
- [37] F. C. Anggian, N. Hidayat, and M. T. Furqon, “Implementasi Metode Modified K-Nearest Neighbor untuk Klasifikasi Status Gunung Berapi,” vol. 3, no. 12, pp. 11027–11033, 2019.
- [38] A. Bode, “K-Nearest Neighbor Dengan Feature Selection Menggunakan Backward Elimination Untuk Prediksi Harga Komoditi Kopi Arabika,” *Ilk. J. Ilm.*, vol. 9, no. 2, pp. 188–195, 2017, doi: 10.33096/ilkom.v9i2.139.188-195.
- [39] H. Ihsan, R. Syam, and F. Ahmad, “Peramalan Penjualan dengan Metode Exponential Smoothing (Studi Kasus : Penjualan Bakso Kemasaan/Kiloan Rumah Bakso Bang Ipul),” *J. Math. Comput. Stat.*, vol. 1, no. 1, p. 1, 2019, doi: 10.35580/jmathcos.v1i1.9168.

- [40] C. V. Hudyanti, F. A. Bachtiar, and B. D. Setiawan, "Perbandingan Double Moving Average dan Double Exponential Smoothing untuk Peramalan Jumlah Kedatangan Wisatawan Mancanegara di Bandara Ngurah Rai," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 3, pp. 2667–2672, 2019.
- [41] H. T. Manurung and A. M. Haryanto, "Analisis Pengaruh ROE , EPS , NPM Dan MVA Terhadap Harga Saham ( Studi Kasus Pada Perusahaan Manufaktur Go Public Sektor Food Dan Beverages Di BEI Tahun 2009-2013 )," *Manajemen*, vol. 4, no. 1999, pp. 1–16, 2015.
- [42] Falahah and D. D. A. Nur, "Pengembangan Aplikasi Sentiment Analysis Menggunakan Metode Naïve Bayes," *Semin. Nas. Sist. Inf. Indones.*, no. November, pp. 335–340, 2015.
- [43] V. G. Utomo, N. Wakhidah, and A. N. Putri, "Prediksi Harga Saham Dengan Svm (Support Vector Machine) Dan Pemilihan Fitur F-Score," *J. Inform. Upgris*, vol. 6, no. 1, 2020, doi: 10.26877/jiu.v6i1.5306.
- [44] A. K. Zyen and R. H. Kusumodestoni, "Pengembangan Model Prediksi Harga Saham Berbasis Neural Network," *J. DISPROTEK*, vol. 7, no. 1, pp. 74–83, 2016.
- [45] L. E. Siahaan, "Prediksi Indeks Harga Saham dengan Metode Gabungan Support Vector Regression dan Jaringan Syaraf Tiruan," *Indones. J. Comput.*, vol. 2, no. 1, p. 21, 2017, doi: 10.21108/indojc.2017.2.1.45.
- [46] A. Pramuntadi, "Model Prediksi Rentet Waktu Neural Network Berbasis Particle Swarm Optimization Untuk Prediksi Harga Saham," *Telematika*, vol. 14, no. 2, pp. 100–106, 2017, doi: 10.31315/telematika.v14i2.2097.
- [47] Y. Umaidah, "Penerapan Algoritma Artificial Neural Network Dalam Prediksi Harga Saham Lq45 Pt. Bank Rakyat Indonesia, Tbk," *J. Gerbang*, vol. 8, no. 1, pp. 57–64, 2018.
- [48] S. Sable, A. Porwal, and U. Singh, "Stock price prediction using genetic algorithms and evolution strategies," *Proc. Int. Conf. Electron. Commun. Aerosp. Technol. ICECA 2017*, vol. 2017-Janua, pp. 549–553, 2017, doi: 10.1109/ICECA.2017.8212724.

- [49] S. M. R. Behravan, "Stock Price Prediction Using Machine Learning and Deep Learning and Swarm Intelligence," *J. Electr. Comput. Eng. Innovations*, vol. 8(1), no. December 2019, pp. 31–40, 2020, doi: 10.22061/JECEI.2020.6898.346.
- [50] "Gambaran Umum PT. Indofood Sukses Makmur." <https://www.indofood.com/company/indofood-at-a-glance>.
- [51] "Sejarah dan Profil Singkat Indofood." <https://britama.com/index.php/2012/11/sejarah-dan-profil-singkat-indf/>.
- [52] "Profil Kimia Farma Tbk," 2020. <https://kimiafarma.co.id/index.php?lang=id>.
- [53] "Profil Telkom Group." [https://www.telkom.co.id/sites/about-telkom/id\\_ID/page/profil-dan-riwayat-singkat](https://www.telkom.co.id/sites/about-telkom/id_ID/page/profil-dan-riwayat-singkat).

# DAFRAT LAMPIRAN

Lampiran 1. Hasil Proses CSV Reader (Dataset KAEF, TLKM)

File Table - 2:1 - CSV Reader (Dataset KAEF)

File Hilite Navigation View

Table "KAEF.JK.csv" - Rows: 1263 Spec - Columns: 7 Properties Flow Variables

Row ID	S Date	D Open	D High	D Low	D Close	D Adj Close	I Volume
Row0	2015-09-01	705	765	700	730	730	3036100
Row1	2015-09-02	720	730	690	725	725	2584300
Row2	2015-09-03	725	730	710	715	715	1561200
Row3	2015-09-04	710	720	695	705	705	779400
Row4	2015-09-07	680	695	670	695	695	746900
Row5	2015-09-08	685	700	670	690	690	2471400
Row6	2015-09-09	690	710	690	705	705	2471900
Row7	2015-09-10	695	710	685	690	690	998800
Row8	2015-09-11	690	700	645	665	665	2563400
Row9	2015-09-14	665	675	660	665	665	905000
Row10	2015-09-15	665	680	660	665	665	690700
Row11	2015-09-16	670	675	660	675	675	897700
Row12	2015-09-17	670	710	670	710	710	3457900
Row13	2015-09-18	720	745	685	700	700	6033200
Row14	2015-09-21	690	700	660	690	690	2248100
Row15	2015-09-22	685	700	670	675	675	612700
Row16	2015-09-23	675	675	645	655	655	1988700
Row17	2015-09-25	655	675	645	650	650	2021200
Row18	2015-09-28	655	655	615	615	615	1479800
Row19	2015-09-29	610	640	590	630	630	1101700
Row20	2015-09-30	635	650	625	640	640	1259700
Row21	2015-10-01	635	660	620	640	640	12573100
Row22	2015-10-02	645	645	625	625	625	1030800
Row23	2015-10-05	635	685	635	665	665	11365100
Row24	2015-10-06	680	695	660	660	660	20627000

File Table - 3:1 - CSV Reader (Dataset TLKM)

File Hilite Navigation View

Table "TLKM.JK.csv" - Rows: 1263 Spec - Columns: 7 Properties Flow Variables

Row ID	S Date	D Open	D High	D Low	D Close	D Adj Close	I Volume
Row0	2015-09-01	2,865	2,875	2,830	2,830	2,632.869	47176700
Row1	2015-09-02	2,790	2,810	2,765	2,775	2,581.7	84858200
Row2	2015-09-03	2,805	2,820	2,785	2,800	2,604.959	29988900
Row3	2015-09-04	2,800	2,820	2,800	2,810	2,614.262	38478400
Row4	2015-09-07	2,760	2,765	2,695	2,700	2,511.925	53535900
Row5	2015-09-08	2,680	2,740	2,680	2,730	2,539.835	35420600
Row6	2015-09-09	2,750	2,755	2,730	2,730	2,539.835	77359400
Row7	2015-09-10	2,720	2,785	2,705	2,765	2,572.397	64256500
Row8	2015-09-11	2,800	2,800	2,760	2,760	2,567.745	28864200
Row9	2015-09-14	2,800	2,800	2,765	2,785	2,591.004	34577600
Row10	2015-09-15	2,755	2,770	2,710	2,710	2,521.228	28528700
Row11	2015-09-16	2,720	2,745	2,700	2,720	2,530.532	32023700
Row12	2015-09-17	2,700	2,710	2,655	2,680	2,493.318	110693400
Row13	2015-09-18	2,690	2,720	2,675	2,690	2,502.621	119461900
Row14	2015-09-21	2,690	2,735	2,680	2,730	2,539.835	39502900
Row15	2015-09-22	2,735	2,745	2,695	2,700	2,511.925	52305600
Row16	2015-09-23	2,665	2,700	2,655	2,655	2,470.06	36091900
Row17	2015-09-25	2,720	2,775	2,640	2,655	2,470.06	63639600
Row18	2015-09-28	2,620	2,645	2,600	2,600	2,418.89	20695500
Row19	2015-09-29	2,550	2,605	2,485	2,605	2,423.542	69456600
Row20	2015-09-30	2,580	2,645	2,565	2,645	2,460.756	89608600
Row21	2015-10-01	2,640	2,665	2,610	2,630	2,446.801	53417800
Row22	2015-10-02	2,605	2,635	2,600	2,620	2,437.497	41421500
Row23	2015-10-05	2,620	2,755	2,615	2,720	2,530.532	71562100
Row24	2015-10-06	2,770	2,830	2,735	2,790	2,595.656	119008800

Lampiran 2. Hasil Proses *Column Filter* (Dataset KAEF, TLKM)

**Filtered table - 2:2 - Column Filter (Pemilihan)**

File [Hilite](#) [Navigation](#) [View](#)

Table "default" - Rows: 1263 Spec - Columns: 4 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close
Row0	705	765	700	730
Row1	720	730	690	725
Row2	725	730	710	715
Row3	710	720	695	705
Row4	680	695	670	695
Row5	685	700	670	690
Row6	690	710	690	705
Row7	695	710	685	690
Row8	690	700	645	665
Row9	665	675	660	665
Row10	665	680	660	665
Row11	670	675	660	675
Row12	670	710	670	710
Row13	720	745	685	700
Row14	690	700	660	690
Row15	685	700	670	675
Row16	675	675	645	655
Row17	655	675	645	650
Row18	655	655	615	615
Row19	610	640	590	630
Row20	635	650	625	640
Row21	635	660	620	640
Row22	645	645	625	625
Row23	635	685	635	665
Row24	680	695	660	660

**Filtered table - 3:2 - Column Filter (Pemilihan)**

File [Hilite](#) [Navigation](#) [View](#)

Table "default" - Rows: 1263 Spec - Columns: 4 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close
Row0	2,865	2,875	2,830	2,830
Row1	2,790	2,810	2,765	2,775
Row2	2,805	2,820	2,785	2,800
Row3	2,800	2,820	2,800	2,810
Row4	2,760	2,765	2,695	2,700
Row5	2,680	2,740	2,680	2,730
Row6	2,750	2,755	2,730	2,730
Row7	2,720	2,785	2,705	2,765
Row8	2,800	2,800	2,760	2,760
Row9	2,800	2,800	2,765	2,785
Row10	2,755	2,770	2,710	2,710
Row11	2,720	2,745	2,700	2,720
Row12	2,700	2,710	2,655	2,680
Row13	2,690	2,720	2,675	2,690
Row14	2,690	2,735	2,680	2,730
Row15	2,735	2,745	2,695	2,700
Row16	2,665	2,700	2,655	2,655
Row17	2,720	2,775	2,640	2,655
Row18	2,620	2,645	2,600	2,600
Row19	2,550	2,605	2,485	2,605
Row20	2,580	2,645	2,565	2,645
Row21	2,640	2,665	2,610	2,630
Row22	2,605	2,635	2,600	2,620
Row23	2,620	2,755	2,615	2,720
Row24	2,770	2,830	2,735	2,790

Lampiran 3. Hasil Proses Node *Missing Value* (Dataset KAEF, TLKM)

**Output table - 2:4 - Missing Value (Fix Value)**

File Hilite Navigation View

Table "default" - Rows: 1263 Spec - Columns: 4 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close
Row0	705	765	700	730
Row1	720	730	690	725
Row2	725	730	710	715
Row3	710	720	695	705
Row4	680	695	670	695
Row5	685	700	670	690
Row6	690	710	690	705
Row7	695	710	685	690
Row8	690	700	645	665
Row9	665	675	660	665
Row10	665	680	660	665
Row11	670	675	660	675
Row12	670	710	670	710
Row13	720	745	685	700
Row14	690	700	660	690
Row15	685	700	670	675
Row16	675	675	645	655
Row17	655	675	645	650
Row18	655	655	615	615
Row19	610	640	590	630
Row20	635	650	625	640
Row21	635	660	620	640
Row22	645	645	625	625
Row23	635	685	635	665
Row24	680	695	660	660

**Output table - 3:4 - Missing Value (Fix Value)**

File Hilite Navigation View

Table "default" - Rows: 1263 Spec - Columns: 4 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close
Row0	2,865	2,875	2,830	2,830
Row1	2,790	2,810	2,765	2,775
Row2	2,805	2,820	2,785	2,800
Row3	2,800	2,820	2,800	2,810
Row4	2,760	2,765	2,695	2,700
Row5	2,680	2,740	2,680	2,730
Row6	2,750	2,755	2,730	2,730
Row7	2,720	2,785	2,705	2,765
Row8	2,800	2,800	2,760	2,760
Row9	2,800	2,800	2,765	2,785
Row10	2,755	2,770	2,710	2,710
Row11	2,720	2,745	2,700	2,720
Row12	2,700	2,710	2,655	2,680
Row13	2,690	2,720	2,675	2,690
Row14	2,690	2,735	2,680	2,730
Row15	2,735	2,745	2,695	2,700
Row16	2,665	2,700	2,655	2,655
Row17	2,720	2,775	2,640	2,655
Row18	2,620	2,645	2,600	2,600
Row19	2,550	2,605	2,485	2,605
Row20	2,580	2,645	2,565	2,645
Row21	2,640	2,665	2,610	2,630
Row22	2,605	2,635	2,600	2,620
Row23	2,620	2,755	2,615	2,720
Row24	2,770	2,830	2,735	2,790

## Lampiran 4. Hasil Proses Normalisasi (Dataset KAEF, TLKM)

Normalized table - 2:3 - Normalizer (Min - Max Norm)

File Hilite Navigation View

Table "default" - Rows: 1263 Spec - Columns: 4 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close
Row0	0.133	0.138	0.146	0.142
Row1	0.138	0.128	0.143	0.141
Row2	0.139	0.128	0.149	0.138
Row3	0.135	0.125	0.144	0.135
Row4	0.127	0.118	0.137	0.133
Row5	0.128	0.12	0.137	0.131
Row6	0.129	0.122	0.143	0.135
Row7	0.131	0.122	0.141	0.131
Row8	0.129	0.12	0.13	0.124
Row9	0.122	0.113	0.134	0.124
Row10	0.122	0.114	0.134	0.124
Row11	0.124	0.113	0.134	0.127
Row12	0.124	0.122	0.137	0.137
Row13	0.138	0.132	0.141	0.134
Row14	0.129	0.12	0.134	0.131
Row15	0.128	0.12	0.137	0.127
Row16	0.125	0.113	0.13	0.121
Row17	0.12	0.113	0.13	0.12
Row18	0.12	0.107	0.121	0.11
Row19	0.107	0.103	0.114	0.114
Row20	0.114	0.106	0.124	0.117
Row21	0.114	0.108	0.123	0.117
Row22	0.117	0.104	0.124	0.113
Row23	0.114	0.115	0.127	0.124
Row24	0.127	0.118	0.134	0.123

Normalized table - 3:3 - Normalizer (Min - Max Norm)

File Hilite Navigation View

Table "default" - Rows: 1263 Spec - Columns: 4 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close
Row0	0.226	0.209	0.247	0.194
Row1	0.196	0.183	0.222	0.172
Row2	0.202	0.187	0.229	0.182
Row3	0.2	0.187	0.235	0.186
Row4	0.184	0.164	0.195	0.141
Row5	0.152	0.154	0.189	0.153
Row6	0.18	0.16	0.208	0.153
Row7	0.168	0.172	0.198	0.167
Row8	0.2	0.179	0.22	0.165
Row9	0.2	0.179	0.222	0.176
Row10	0.182	0.166	0.2	0.145
Row11	0.168	0.156	0.197	0.149
Row12	0.16	0.142	0.179	0.133
Row13	0.156	0.146	0.187	0.137
Row14	0.156	0.152	0.189	0.153
Row15	0.174	0.156	0.195	0.141
Row16	0.146	0.138	0.179	0.122
Row17	0.168	0.168	0.173	0.122
Row18	0.128	0.116	0.158	0.1
Row19	0.1	0.1	0.114	0.102
Row20	0.112	0.116	0.144	0.118
Row21	0.136	0.124	0.162	0.112
Row22	0.122	0.112	0.158	0.108
Row23	0.128	0.16	0.164	0.149
Row24	0.188	0.191	0.21	0.178

Lampiran 5. Hasil *Partitioning* (Training dan Testing KAEF)

▲ First partition (as defined in dialog) - 2:6 - Partitioning (80 : 20)

File Hilite Navigation View

Table "default" - Rows: 1010 Spec - Columns: 4 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close
Row0	0.133	0.138	0.146	0.142
Row1	0.138	0.128	0.143	0.141
Row2	0.139	0.128	0.149	0.138
Row3	0.135	0.125	0.144	0.135
Row5	0.128	0.12	0.137	0.131
Row6	0.129	0.122	0.143	0.135
Row7	0.131	0.122	0.141	0.131
Row8	0.129	0.12	0.13	0.124
Row10	0.122	0.114	0.134	0.124
Row11	0.124	0.113	0.134	0.127
Row12	0.124	0.122	0.137	0.137
Row13	0.138	0.132	0.141	0.134
Row15	0.128	0.12	0.137	0.127
Row16	0.125	0.113	0.13	0.121
Row17	0.12	0.113	0.13	0.12
Row18	0.12	0.107	0.121	0.11
Row20	0.114	0.106	0.124	0.117
Row21	0.114	0.108	0.123	0.117
Row22	0.117	0.104	0.124	0.113
Row23	0.114	0.115	0.127	0.124
Row25	0.122	0.131	0.134	0.144
Row26	0.143	0.146	0.151	0.144
Row27	0.146	0.142	0.16	0.155
Row28	0.154	0.152	0.169	0.167
Row30	0.167	0.157	0.177	0.168

▲ Second partition (remaining rows) - 2:6 - Partitioning (80 : 20)

File Hilite Navigation View

Table "default" - Rows: 253 Spec - Columns: 4 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close
Row4	0.127	0.118	0.137	0.133
Row9	0.122	0.113	0.134	0.124
Row14	0.129	0.12	0.134	0.131
Row19	0.107	0.103	0.114	0.114
Row24	0.127	0.118	0.134	0.123
Row29	0.166	0.159	0.177	0.169
Row34	0.191	0.179	0.191	0.186
Row39	0.189	0.179	0.194	0.182
Row44	0.18	0.171	0.193	0.185
Row49	0.206	0.217	0.217	0.223
Row54	0.209	0.198	0.214	0.208
Row59	0.213	0.2	0.224	0.213
Row64	0.21	0.2	0.224	0.212
Row69	0.195	0.182	0.21	0.198
Row74	0.17	0.164	0.184	0.174
Row79	0.181	0.168	0.184	0.174
Row84	0.174	0.165	0.184	0.181
Row89	0.184	0.174	0.191	0.183
Row94	0.198	0.189	0.206	0.2
Row99	0.192	0.188	0.203	0.199
Row104	0.245	0.253	0.261	0.263
Row109	0.244	0.235	0.251	0.242
Row114	0.259	0.249	0.274	0.261
Row119	0.241	0.232	0.251	0.247
Row124	0.244	0.253	0.26	0.268

Lampiran 6 Hasil *Partitioning* (Training Dan Testing TLKM)

▲ First partition (as defined in dialog) - 3:43 - Partitioning (80 : 20)

File Hilite Navigation View

Table "default" - Rows: 1010 Spec - Columns: 4 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close
Row0	0.226	0.209	0.247	0.194
Row1	0.196	0.183	0.222	0.172
Row2	0.202	0.187	0.229	0.182
Row3	0.2	0.187	0.235	0.186
Row5	0.152	0.154	0.189	0.153
Row6	0.18	0.16	0.208	0.153
Row7	0.168	0.172	0.198	0.167
Row8	0.2	0.179	0.22	0.165
Row10	0.182	0.166	0.2	0.145
Row11	0.168	0.156	0.197	0.149
Row12	0.16	0.142	0.179	0.133
Row13	0.156	0.146	0.187	0.137
Row15	0.174	0.156	0.195	0.141
Row16	0.146	0.138	0.179	0.122
Row17	0.168	0.168	0.173	0.122
Row18	0.128	0.116	0.158	0.1
Row20	0.112	0.116	0.144	0.118
Row21	0.136	0.124	0.162	0.112
Row22	0.122	0.112	0.158	0.108
Row23	0.128	0.16	0.164	0.149
Row25	0.19	0.174	0.197	0.157
Row26	0.174	0.156	0.2	0.145
Row27	0.172	0.158	0.195	0.141
Row28	0.158	0.146	0.189	0.139
Row30	0.144	0.15	0.181	0.135

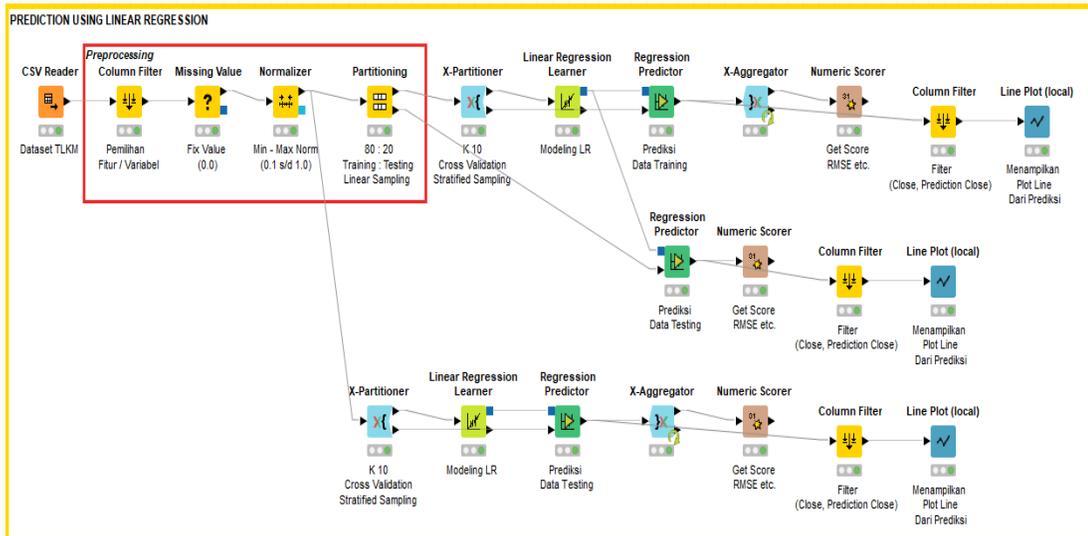
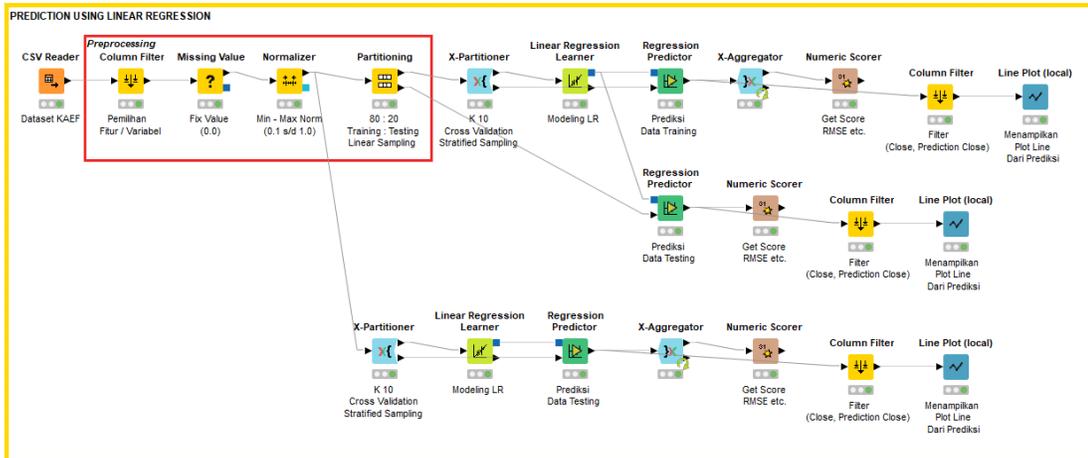
▲ Second partition (remaining rows) - 3:43 - Partitioning (80 : 20)

File Hilite Navigation View

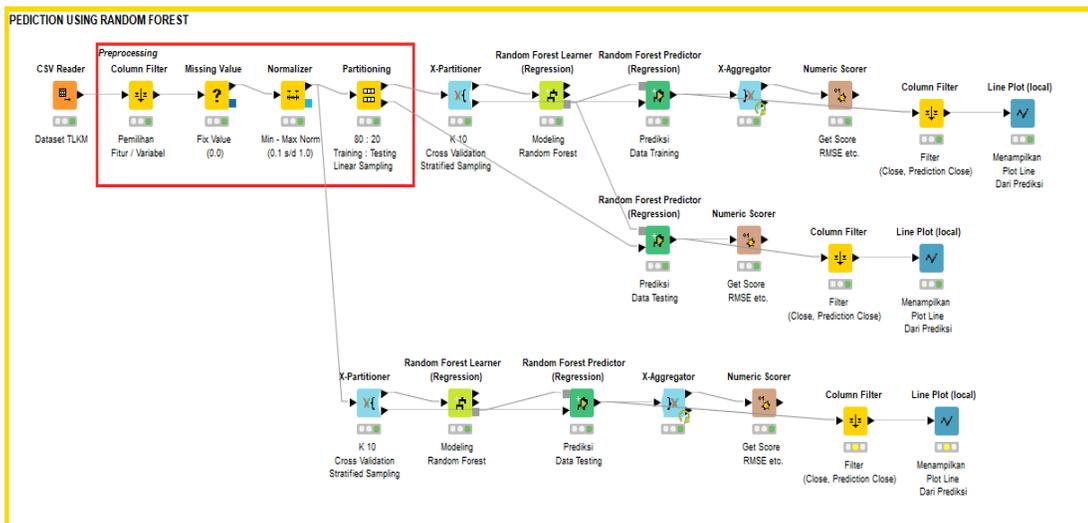
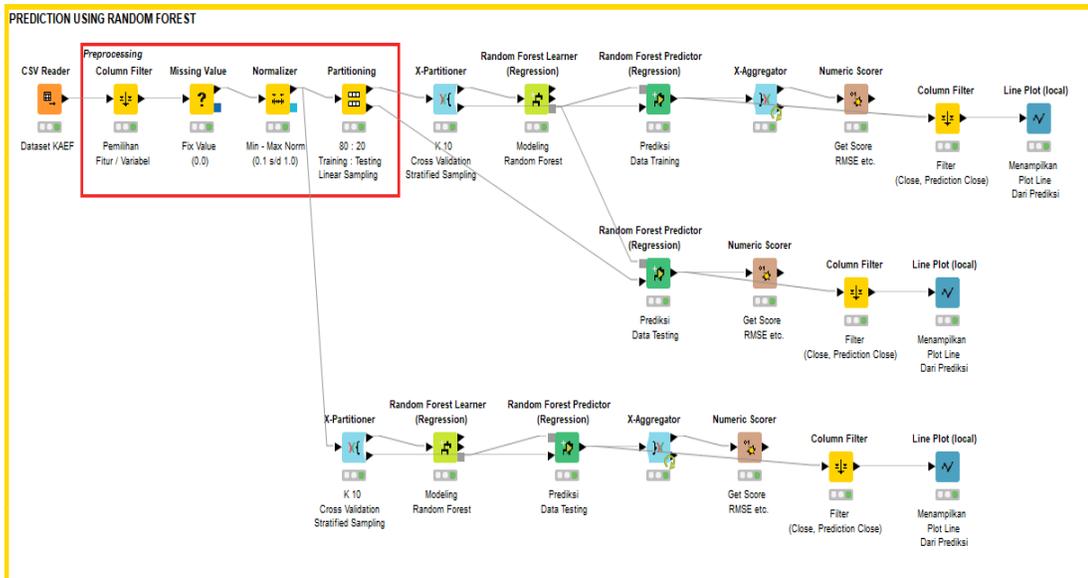
Table "default" - Rows: 253 Spec - Columns: 4 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close
Row4	0.184	0.164	0.195	0.141
Row9	0.2	0.179	0.222	0.176
Row14	0.156	0.152	0.189	0.153
Row19	0.1	0.1	0.114	0.102
Row24	0.188	0.191	0.21	0.178
Row29	0.158	0.138	0.171	0.116
Row34	0.178	0.164	0.202	0.153
Row39	0.186	0.181	0.222	0.18
Row44	0.176	0.16	0.206	0.153
Row49	0.152	0.138	0.181	0.127
Row54	0.216	0.205	0.245	0.202
Row59	0.236	0.229	0.258	0.223
Row64	0.254	0.251	0.287	0.251
Row69	0.286	0.265	0.318	0.27
Row74	0.314	0.293	0.297	0.28
Row79	0.31	0.317	0.32	0.309
Row84	0.33	0.35	0.361	0.343
Row89	0.314	0.342	0.343	0.323
Row94	0.31	0.313	0.341	0.309
Row99	0.362	0.352	0.386	0.356
Row104	0.392	0.388	0.403	0.36
Row109	0.408	0.416	0.436	0.421
Row114	0.408	0.402	0.432	0.401
Row119	0.358	0.356	0.38	0.358
Row124	0.416	0.408	0.438	0.403

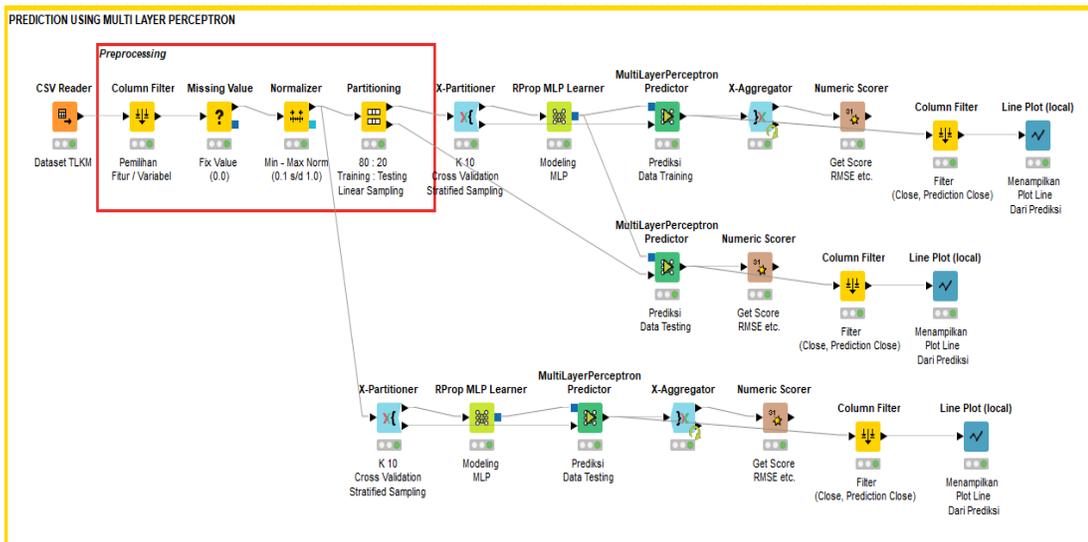
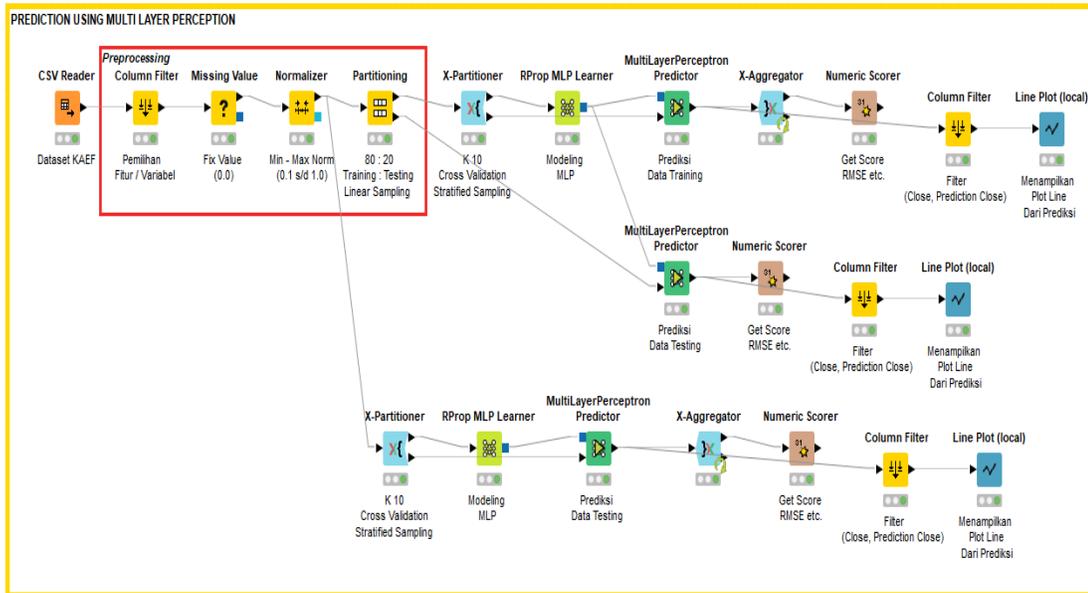
Lampiran 7. Model *Linear Regression* (KAEF, TLKM)



Lampiran 8. Model RFR (KAEF, TLKM)



Lampiran 9. Model MLP (KAEF, TLKM)



Lampiran 10. Hasil Prediksi Node *Predictor Regression* (KAEF, TLKM)

**Predicted data - 0:9 - Regression Predictor (Prediksi)**

File [Hilite](#) [Navigation](#) [View](#)

Table "default" - Rows: 101 Spec - Columns: 5 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)
Row8	0.129	0.12	0.13	0.124	0.123
Row18	0.12	0.107	0.121	0.11	0.113
Row51	0.203	0.191	0.207	0.2	0.197
Row57	0.206	0.203	0.22	0.217	0.215
Row65	0.209	0.2	0.221	0.213	0.213
Row73	0.16	0.156	0.173	0.169	0.167
Row78	0.182	0.172	0.193	0.181	0.184
Row85	0.174	0.178	0.189	0.189	0.188
Row116	0.258	0.245	0.264	0.254	0.254
Row122	0.241	0.228	0.251	0.24	0.24
Row127	0.272	0.276	0.286	0.287	0.284
Row140	0.302	0.289	0.303	0.295	0.293
Row147	0.293	0.283	0.307	0.294	0.297
Row160	0.291	0.296	0.306	0.314	0.305
Row172	0.274	0.264	0.274	0.277	0.266
Row191	0.269	0.288	0.286	0.297	0.294
Row208	0.262	0.255	0.266	0.266	0.259
Row218	0.29	0.28	0.304	0.291	0.294
Row231	0.339	0.341	0.354	0.343	0.351
Row241	0.553	0.632	0.577	0.643	0.619
Row242	0.64	0.738	0.666	0.734	0.719
Row248	0.676	0.677	0.614	0.601	0.619
Row252	0.615	0.613	0.574	0.558	0.574
Row267	0.679	0.682	0.686	0.669	0.682
Row270	0.623	0.621	0.637	0.626	0.629

**Predicted data - 3:9 - Regression Predictor (Prediksi)**

File [Hilite](#) [Navigation](#) [View](#)

Table "default" - Rows: 101 Spec - Columns: 5 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)
Row18	0.128	0.116	0.158	0.1	0.107
Row28	0.158	0.146	0.189	0.139	0.14
Row32	0.16	0.15	0.191	0.149	0.143
Row40	0.198	0.177	0.204	0.153	0.154
Row55	0.226	0.211	0.237	0.188	0.192
Row72	0.24	0.241	0.274	0.235	0.238
Row77	0.304	0.305	0.328	0.309	0.297
Row103	0.396	0.41	0.424	0.399	0.407
Row112	0.404	0.396	0.421	0.397	0.389
Row122	0.416	0.396	0.407	0.366	0.371
Row138	0.432	0.412	0.428	0.388	0.392
Row197	0.636	0.63	0.645	0.628	0.627
Row198	0.628	0.617	0.633	0.611	0.613
Row201	0.624	0.609	0.633	0.615	0.609
Row213	0.752	0.738	0.741	0.73	0.728
Row243	0.736	0.738	0.745	0.742	0.74
Row250	0.76	0.746	0.761	0.738	0.746
Row251	0.74	0.726	0.71	0.693	0.699
Row255	0.736	0.75	0.73	0.71	0.736
Row266	0.824	0.819	0.818	0.828	0.814
Row270	0.768	0.758	0.764	0.755	0.754
Row281	0.76	0.758	0.772	0.763	0.764
Row290	0.728	0.726	0.737	0.734	0.728
Row308	0.608	0.621	0.61	0.583	0.607
Row310	0.608	0.646	0.625	0.636	0.639

## Lampiran 11. Hasil Prediksi Proses X-Aggregator LR (KAEF,TLKM)

**Prediction table - 0:10 - X-Aggregator**

File Hilite Navigation View

Table "default" - Rows: 1010 Spec - Columns: 5 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close	D Predict...
Row0	0.133	0.138	0.146	0.142	0.147
Row20	0.114	0.106	0.124	0.117	0.117
Row31	0.166	0.164	0.177	0.178	0.174
Row42	0.182	0.175	0.187	0.176	0.181
Row56	0.202	0.196	0.217	0.208	0.21
Row58	0.216	0.209	0.227	0.216	0.22
Row86	0.185	0.181	0.199	0.186	0.193
Row90	0.185	0.179	0.197	0.188	0.191
Row96	0.203	0.193	0.204	0.195	0.197
Row112	0.244	0.248	0.254	0.261	0.254
Row123	0.238	0.23	0.249	0.246	0.24
Row141	0.291	0.284	0.309	0.297	0.3
Row162	0.295	0.284	0.306	0.295	0.295
Row165	0.273	0.266	0.289	0.277	0.28
Row173	0.273	0.271	0.284	0.28	0.28
Row183	0.272	0.263	0.286	0.275	0.277
Row205	0.266	0.257	0.277	0.267	0.268
Row210	0.262	0.256	0.277	0.266	0.27
Row255	0.57	0.593	0.549	0.539	0.562
Row258	0.5	0.568	0.523	0.59	0.56
Row263	0.62	0.629	0.62	0.629	0.621
Row265	0.618	0.632	0.64	0.635	0.641
Row275	0.637	0.632	0.654	0.641	0.643
Row280	0.587	0.593	0.606	0.604	0.602
Row325	0.673	0.688	0.7	0.7	0.7

**Prediction table - 3:13 - X-Aggregator**

File Hilite Navigation View

Table "default" - Rows: 1010 Spec - Columns: 5 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close	D Predict...
Row0	0.226	0.209	0.247	0.194	0.198
Row7	0.168	0.172	0.198	0.167	0.161
Row20	0.112	0.116	0.144	0.118	0.103
Row48	0.16	0.148	0.195	0.139	0.144
Row56	0.202	0.227	0.237	0.229	0.216
Row63	0.262	0.255	0.283	0.239	0.244
Row86	0.36	0.36	0.372	0.35	0.344
Row105	0.36	0.37	0.382	0.376	0.36
Row123	0.38	0.378	0.409	0.384	0.378
Row140	0.41	0.396	0.421	0.388	0.385
Row143	0.418	0.402	0.428	0.397	0.392
Row150	0.458	0.438	0.484	0.446	0.445
Row166	0.464	0.468	0.482	0.444	0.462
Row178	0.552	0.553	0.567	0.554	0.55
Row181	0.6	0.581	0.591	0.579	0.565
Row182	0.58	0.565	0.591	0.575	0.564
Row190	0.636	0.626	0.587	0.587	0.575
Row202	0.632	0.617	0.598	0.599	0.581
Row203	0.596	0.609	0.614	0.607	0.607
Row207	0.684	0.682	0.679	0.66	0.67
Row216	0.76	0.742	0.745	0.722	0.729
Row227	0.816	0.807	0.815	0.816	0.806
Row246	0.728	0.746	0.745	0.755	0.75
Row268	0.812	0.803	0.807	0.795	0.799
Row275	0.76	0.762	0.768	0.759	0.764

Lampiran 12. Data *Testing* Hasil *Partitioning* LR (KAEF,TLKM)

▲ Predicted data - 0:12 - Regression Predictor (Prediksi)

File Hilite Navigation View

Table "default" - Rows: 253 Spec - Columns: 5 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)
Row4	0.127	0.118	0.137	0.133	0.13
Row9	0.122	0.113	0.134	0.124	0.126
Row14	0.129	0.12	0.134	0.131	0.127
Row19	0.107	0.103	0.114	0.114	0.11
Row24	0.127	0.118	0.134	0.123	0.127
Row29	0.166	0.159	0.177	0.169	0.17
Row34	0.191	0.179	0.191	0.186	0.183
Row39	0.189	0.179	0.194	0.182	0.186
Row44	0.18	0.171	0.193	0.185	0.185
Row49	0.206	0.217	0.217	0.223	0.221
Row54	0.209	0.198	0.214	0.208	0.205
Row59	0.213	0.2	0.224	0.213	0.213
Row64	0.21	0.2	0.224	0.212	0.215
Row69	0.195	0.182	0.21	0.198	0.199
Row74	0.17	0.164	0.184	0.174	0.178
Row79	0.181	0.168	0.184	0.174	0.175
Row84	0.174	0.165	0.184	0.181	0.176
Row89	0.184	0.174	0.191	0.183	0.183
Row94	0.198	0.189	0.206	0.2	0.198
Row99	0.192	0.188	0.203	0.199	0.197
Row104	0.245	0.253	0.261	0.263	0.263
Row109	0.244	0.235	0.251	0.242	0.243
Row114	0.259	0.249	0.274	0.261	0.264
Row119	0.241	0.232	0.251	0.247	0.243
Row124	0.244	0.253	0.26	0.268	0.262

▲ Predicted data - 3:16 - Regression Predictor (Prediksi)

File Hilite Navigation View

Table "default" - Rows: 253 Spec - Columns: 5 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)
Row4	0.184	0.164	0.195	0.141	0.144
Row9	0.2	0.179	0.222	0.176	0.169
Row14	0.156	0.152	0.189	0.153	0.145
Row19	0.1	0.1	0.114	0.102	0.073
Row24	0.188	0.191	0.21	0.178	0.175
Row29	0.158	0.138	0.171	0.116	0.119
Row34	0.178	0.164	0.202	0.153	0.154
Row39	0.186	0.181	0.222	0.18	0.178
Row44	0.176	0.16	0.206	0.153	0.155
Row49	0.152	0.138	0.181	0.127	0.13
Row54	0.216	0.205	0.245	0.202	0.199
Row59	0.236	0.229	0.258	0.223	0.218
Row64	0.254	0.251	0.287	0.251	0.249
Row69	0.286	0.265	0.318	0.27	0.268
Row74	0.314	0.293	0.297	0.28	0.257
Row79	0.31	0.317	0.32	0.309	0.296
Row84	0.33	0.35	0.361	0.343	0.344
Row89	0.314	0.342	0.343	0.323	0.332
Row94	0.31	0.313	0.341	0.309	0.311
Row99	0.362	0.352	0.386	0.356	0.349
Row104	0.392	0.388	0.403	0.36	0.375
Row109	0.408	0.416	0.436	0.421	0.415
Row114	0.408	0.402	0.432	0.401	0.401
Row119	0.358	0.356	0.38	0.358	0.35
Row124	0.416	0.408	0.438	0.403	0.406

Lampiran 13. Hasil *Attribute Statistic RFR Partitioning* (KAEF, TLKM)

Attribute Statistics - 0:25 - Random Forest Learner (Regression) (Modeling)

File Hilitte Navigation View

Table "Tree Ensemble Column Statistic" - Rows: 3 Spec - Columns: 6 Properties Flow Variables

Row ID	#splits (level 0)	#splits (level 1)	#splits (level 2)	#candidates (level 0)	#candidates (level 1)	#candidates (level 2)
Open	35	71	136	35	71	136
High	37	57	140	37	57	140
Low	28	72	124	28	72	124

Attribute Statistics - 3:29 - Random Forest Learner (Regression) (Modeling)

File Hilitte Navigation View

Table "Tree Ensemble Column Statistic" - Rows: 3 Spec - Columns: 6 Properties Flow Variables

Row ID	#splits (level 0)	#splits (level 1)	#splits (level 2)	#candidates (level 0)	#candidates (level 1)	#candidates (level 2)
Open	31	60	128	31	60	128
High	24	66	125	24	66	125
Low	45	74	147	45	74	147

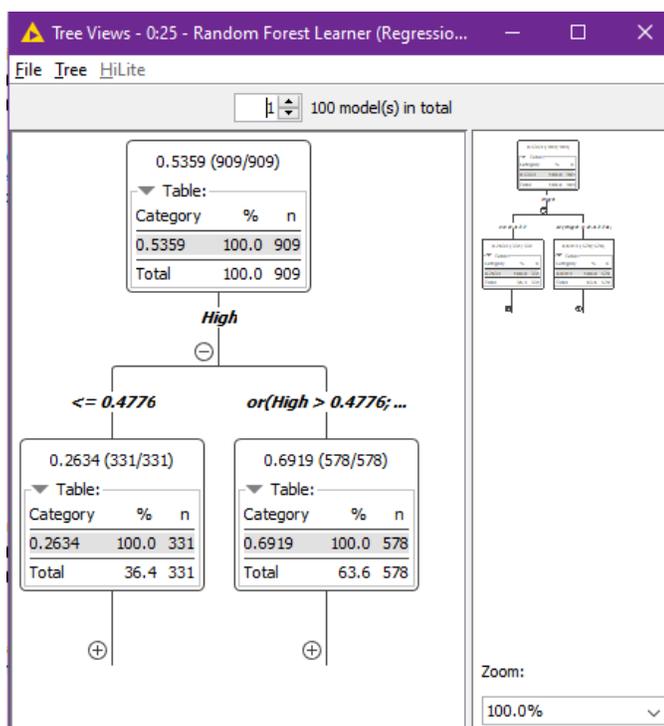
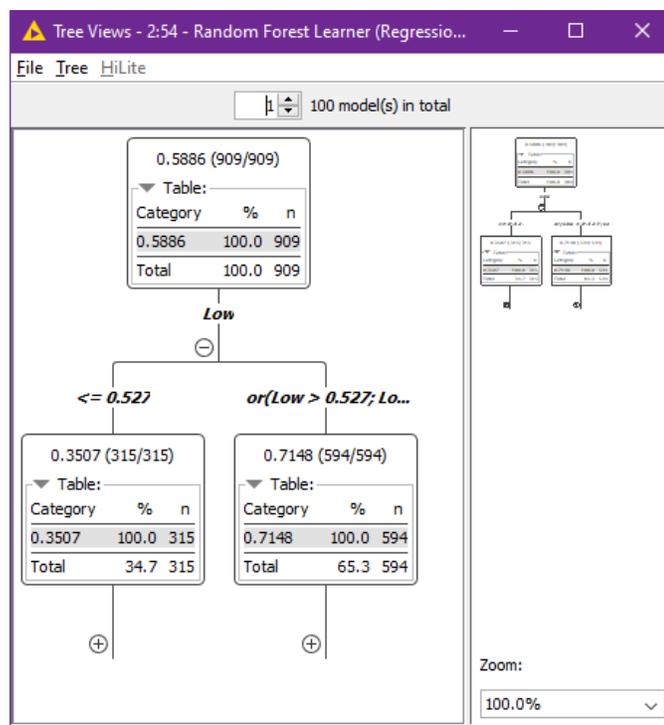
Lampiran 14. *Output Out of Bag Prediction RFR Partitioning* (KAEF)

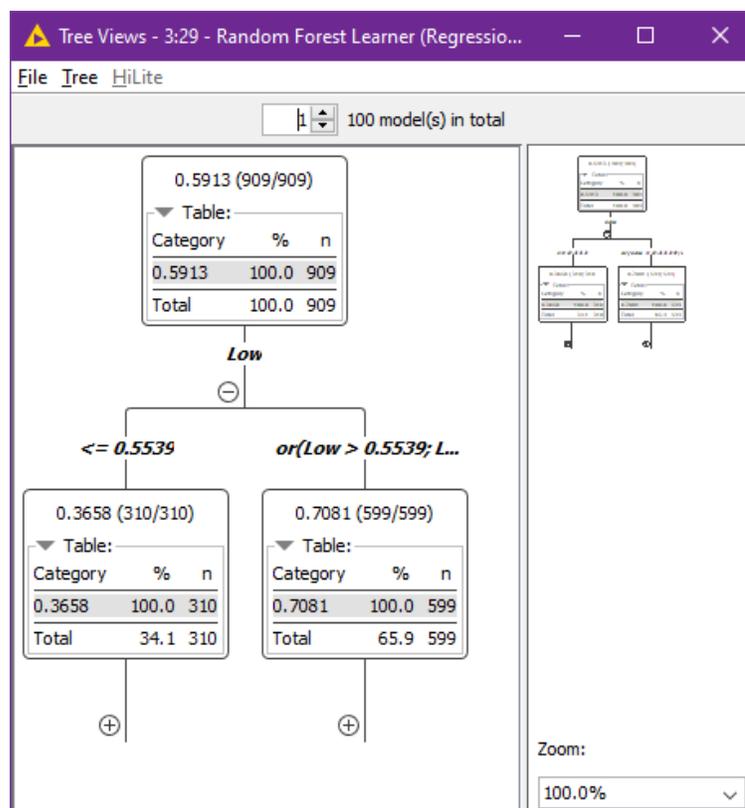
Out-of-bag Predictions - 0:25 - Random Forest Learner (Regression) (Modeling)

File Hilitte Navigation View

Table "default" - Rows: 909 Spec - Columns: 7 Properties Flow Variables

Row ID	Open	High	Low	Close	Close (Out-of-bag)	Close (Out-of-bag) (Prediction Variance)	model count
Row0	0.133	0.138	0.146	0.142	0.138	0	38
Row1	0.138	0.128	0.143	0.141	0.133	0	39
Row2	0.139	0.128	0.149	0.138	0.138	0	35
Row3	0.135	0.125	0.144	0.135	0.134	0	35
Row6	0.129	0.122	0.143	0.135	0.133	0	31
Row7	0.131	0.122	0.141	0.131	0.135	0	36
Row8	0.129	0.12	0.13	0.124	0.121	0	39
Row10	0.122	0.114	0.134	0.124	0.13	0	32
Row11	0.124	0.113	0.134	0.127	0.128	0	43
Row12	0.124	0.122	0.137	0.137	0.129	0	41
Row13	0.138	0.132	0.141	0.134	0.135	0	41
Row15	0.128	0.12	0.137	0.127	0.131	0	35
Row16	0.125	0.113	0.13	0.121	0.122	0	34
Row17	0.12	0.113	0.13	0.12	0.119	0	40
Row18	0.12	0.107	0.121	0.11	0.116	0	29
Row20	0.114	0.106	0.124	0.117	0.113	0	33
Row22	0.117	0.104	0.124	0.113	0.117	0	31
Row23	0.114	0.115	0.127	0.124	0.116	0	38
Row25	0.122	0.131	0.134	0.144	0.13	0	37
Row26	0.143	0.146	0.151	0.144	0.151	0	31
Row27	0.146	0.142	0.16	0.155	0.155	0	40
Row28	0.154	0.152	0.169	0.167	0.161	0	40
Row30	0.167	0.157	0.177	0.168	0.169	0	35
Row31	0.166	0.164	0.177	0.178	0.168	0	33
Row32	0.174	0.177	0.187	0.191	0.184	0	38

Lampiran 15. Hasil *Output Tree Views RFR Partitioning* (INDF, KAUF, TLKM)



Lampiran 16. Hasil Prediksi RFR *Predictor Partitioning* (KAEF, TLKM)

▲ Prediction output - 0:26 - Random Forest Predictor (Regression) (Prediksi)

File [Hilite](#) [Navigation](#) [View](#)

Table "default" - Rows: 101 Spec - Columns: 6 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)	D Prediction (Close) (Prediction Variance)
Row5	0.128	0.12	0.137	0.131	0.129	0
Row21	0.114	0.108	0.123	0.117	0.115	0
Row36	0.2	0.195	0.211	0.2	0.205	0
Row61	0.213	0.202	0.217	0.21	0.209	0
Row62	0.207	0.195	0.203	0.21	0.203	0
Row90	0.185	0.179	0.197	0.188	0.191	0
Row91	0.182	0.174	0.191	0.188	0.182	0
Row103	0.233	0.245	0.247	0.249	0.25	0
Row117	0.251	0.239	0.26	0.25	0.249	0
Row177	0.274	0.267	0.267	0.268	0.264	0
Row185	0.273	0.262	0.287	0.275	0.277	0
Row201	0.27	0.26	0.284	0.274	0.273	0
Row207	0.266	0.253	0.274	0.266	0.268	0
Row220	0.279	0.271	0.289	0.275	0.277	0
Row251	0.662	0.663	0.62	0.624	0.637	0.001
Row258	0.5	0.568	0.523	0.59	0.569	0.002
Row260	0.654	0.76	0.643	0.663	0.702	0.002
Row261	0.662	0.668	0.649	0.675	0.653	0
Row270	0.623	0.621	0.637	0.626	0.629	0
Row277	0.618	0.621	0.623	0.615	0.623	0
Row287	0.542	0.596	0.56	0.595	0.565	0
Row301	0.545	0.579	0.569	0.595	0.566	0
Row308	0.671	0.71	0.689	0.708	0.686	0.001
Row326	0.69	0.688	0.714	0.697	0.705	0
Row353	0.514	0.546	0.537	0.553	0.535	0

▲ Prediction output - 3:30 - Random Forest Predictor (Regression) (Prediksi)

File [Hilite](#) [Navigation](#) [View](#)

Table "default" - Rows: 101 Spec - Columns: 6 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)	D Prediction (Close) (Prediction Variance)
Row6	0.18	0.16	0.208	0.153	0.149	0
Row18	0.128	0.116	0.158	0.1	0.113	0
Row23	0.128	0.16	0.164	0.149	0.122	0
Row48	0.16	0.148	0.195	0.139	0.144	0
Row53	0.18	0.199	0.212	0.188	0.161	0
Row72	0.24	0.241	0.274	0.235	0.24	0
Row77	0.304	0.305	0.328	0.309	0.304	0
Row103	0.396	0.41	0.424	0.399	0.393	0
Row121	0.416	0.396	0.424	0.39	0.395	0
Row122	0.416	0.396	0.407	0.366	0.39	0
Row140	0.41	0.396	0.421	0.388	0.394	0
Row143	0.418	0.402	0.428	0.397	0.399	0
Row152	0.44	0.42	0.467	0.427	0.426	0
Row171	0.564	0.573	0.583	0.575	0.57	0
Row175	0.576	0.561	0.591	0.566	0.57	0
Row178	0.552	0.553	0.567	0.554	0.541	0
Row183	0.584	0.577	0.583	0.55	0.553	0
Row200	0.632	0.617	0.641	0.62	0.621	0
Row216	0.76	0.742	0.745	0.722	0.735	0
Row227	0.816	0.807	0.815	0.816	0.808	0
Row258	0.744	0.75	0.737	0.755	0.736	0
Row266	0.824	0.819	0.818	0.828	0.817	0
Row273	0.732	0.726	0.745	0.73	0.726	0
Row290	0.728	0.726	0.737	0.734	0.727	0
Row308	0.608	0.621	0.61	0.583	0.606	0

Lampiran 17. Hasil Proses *X-Aggregator* RFR (KAEF, TLKM)

▲ Prediction table - 0:27 - X-Aggregator

File Hilit Navigation View

Table "default" - Rows: 1010 Spec - Columns: 6 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)	D Prediction (Close) (Prediction Variance)
Row30	0.167	0.157	0.177	0.168	0.17	0
Row42	0.182	0.175	0.187	0.176	0.184	0
Row53	0.21	0.203	0.221	0.209	0.213	0
Row57	0.206	0.203	0.22	0.217	0.212	0
Row67	0.205	0.195	0.213	0.206	0.204	0
Row78	0.182	0.172	0.193	0.181	0.185	0
Row82	0.182	0.171	0.189	0.182	0.182	0
Row100	0.199	0.243	0.214	0.244	0.228	0.001
Row107	0.233	0.241	0.249	0.254	0.247	0
Row121	0.238	0.228	0.253	0.242	0.241	0
Row153	0.309	0.296	0.327	0.314	0.313	0
Row158	0.295	0.287	0.303	0.292	0.294	0
Row163	0.291	0.281	0.289	0.275	0.288	0
Row202	0.272	0.259	0.263	0.271	0.257	0
Row208	0.262	0.255	0.266	0.266	0.258	0
Row212	0.263	0.253	0.277	0.267	0.266	0
Row231	0.339	0.341	0.354	0.343	0.344	0
Row236	0.364	0.457	0.383	0.476	0.408	0.002
Row247	0.704	0.872	0.674	0.672	0.726	0.004
Row265	0.618	0.632	0.64	0.635	0.634	0
Row276	0.634	0.627	0.643	0.626	0.631	0
Row278	0.609	0.604	0.609	0.595	0.6	0
Row293	0.615	0.632	0.617	0.643	0.627	0
Row296	0.637	0.641	0.643	0.638	0.643	0
Row297	0.632	0.627	0.626	0.609	0.62	0

▲ Prediction table - 3:28 - X-Aggregator

File Hilit Navigation View

Table "default" - Rows: 1010 Spec - Columns: 6 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)	D Prediction (Close) (Prediction Variance)
Row0	0.226	0.209	0.247	0.194	0.189	0
Row7	0.168	0.172	0.198	0.167	0.156	0
Row20	0.112	0.116	0.144	0.118	0.115	0.001
Row28	0.158	0.146	0.189	0.139	0.142	0
Row41	0.18	0.162	0.189	0.133	0.147	0
Row56	0.202	0.227	0.237	0.229	0.191	0
Row63	0.262	0.255	0.283	0.239	0.241	0
Row86	0.36	0.36	0.372	0.35	0.337	0
Row105	0.36	0.37	0.382	0.376	0.343	0
Row112	0.404	0.396	0.421	0.397	0.388	0
Row123	0.38	0.378	0.409	0.384	0.369	0
Row138	0.432	0.412	0.428	0.388	0.4	0
Row150	0.458	0.438	0.484	0.446	0.446	0
Row166	0.464	0.468	0.482	0.444	0.456	0
Row167	0.464	0.46	0.475	0.464	0.45	0
Row182	0.58	0.565	0.591	0.575	0.572	0
Row188	0.608	0.613	0.625	0.624	0.617	0
Row220	0.8	0.803	0.799	0.812	0.792	0
Row236	0.78	0.77	0.745	0.73	0.75	0.001
Row281	0.76	0.758	0.772	0.763	0.759	0
Row287	0.772	0.758	0.776	0.755	0.76	0
Row292	0.748	0.783	0.764	0.787	0.784	0.001
Row307	0.612	0.613	0.633	0.607	0.616	0
Row312	0.652	0.646	0.66	0.636	0.646	0
Row321	0.58	0.581	0.587	0.562	0.568	0

Lampiran 18. Hasil *Error Rate* Dari *X-Aggregator* RFR (KAEF, TLKM)

**Error rates - 0:27 - X-Aggregator**

File [Hilite](#) [Navigation](#) [View](#)

Table "default" - Rows: 10 Spec - Columns: 3 Properties Flow Variables

Row ID	D Total squared error	D Mean squared error	I Size of Test Set
fold 0	0.018	0	101
fold 1	0.022	0	101
fold 2	0.011	0	101
fold 3	0.021	0	101
fold 4	0.02	0	101
fold 5	0.016	0	101
fold 6	0.023	0	101
fold 7	0.013	0	101
fold 8	0.016	0	101
fold 9	0.017	0	101

**Error rates - 3:28 - X-Aggregator**

File [Hilite](#) [Navigation](#) [View](#)

Table "default" - Rows: 10 Spec - Columns: 3 Properties Flow Variables

Row ID	D Total squared error	D Mean squared error	I Size of Test Set
fold 0	0.048	0	101
fold 1	0.047	0	101
fold 2	0.048	0	101
fold 3	0.046	0	101
fold 4	0.046	0	101
fold 5	0.047	0	101
fold 6	0.052	0.001	101
fold 7	0.047	0	101
fold 8	0.048	0	101
fold 9	0.049	0	101

Lampiran 19. Hasil Data *Testing Partitioning* RFR (KAEF, TLKM)

**▲ Prediction output - 0:29 - Random Forest Predictor (Regression) (Prediksi)**

File [Hilite](#) [Navigation](#) [View](#)

Table "default" - Rows: 253 Spec - Columns: 6 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)	D Prediction (Close) (Prediction Variance)
Row4	0.127	0.118	0.137	0.133	0.128	0
Row9	0.122	0.113	0.134	0.124	0.127	0
Row14	0.129	0.12	0.134	0.131	0.129	0
Row19	0.107	0.103	0.114	0.114	0.112	0
Row24	0.127	0.118	0.134	0.123	0.129	0
Row29	0.166	0.159	0.177	0.169	0.169	0
Row34	0.191	0.179	0.191	0.186	0.183	0
Row39	0.189	0.179	0.194	0.182	0.191	0
Row44	0.18	0.171	0.193	0.185	0.181	0
Row49	0.206	0.217	0.217	0.223	0.214	0
Row54	0.209	0.198	0.214	0.208	0.209	0
Row59	0.213	0.2	0.224	0.213	0.212	0
Row64	0.21	0.2	0.224	0.212	0.212	0
Row69	0.195	0.182	0.21	0.198	0.192	0
Row74	0.17	0.164	0.184	0.174	0.172	0
Row79	0.181	0.168	0.184	0.174	0.175	0
Row84	0.174	0.165	0.184	0.181	0.174	0
Row89	0.184	0.174	0.191	0.183	0.184	0
Row94	0.198	0.189	0.206	0.2	0.198	0
Row99	0.192	0.188	0.203	0.199	0.198	0
Row104	0.245	0.253	0.261	0.263	0.259	0
Row109	0.244	0.235	0.251	0.242	0.244	0
Row114	0.259	0.249	0.274	0.261	0.266	0
Row119	0.241	0.232	0.251	0.247	0.241	0
Row124	0.244	0.253	0.26	0.268	0.257	0

**▲ Prediction output - 3:31 - Random Forest Predictor (Regression) (Prediksi)**

File [Hilite](#) [Navigation](#) [View](#)

Table "default" - Rows: 253 Spec - Columns: 6 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)	D Prediction (Close) (Prediction Variance)
Row4	0.184	0.164	0.195	0.141	0.142	0
Row9	0.2	0.179	0.222	0.176	0.172	0
Row14	0.156	0.152	0.189	0.153	0.146	0
Row19	0.1	0.1	0.114	0.102	0.147	0.004
Row24	0.188	0.191	0.21	0.178	0.166	0
Row29	0.158	0.138	0.171	0.116	0.126	0
Row34	0.178	0.164	0.202	0.153	0.149	0
Row39	0.186	0.181	0.222	0.18	0.167	0
Row44	0.176	0.16	0.206	0.153	0.151	0
Row49	0.152	0.138	0.181	0.127	0.133	0
Row54	0.216	0.205	0.245	0.202	0.194	0
Row59	0.236	0.229	0.258	0.223	0.226	0
Row64	0.254	0.251	0.287	0.251	0.238	0
Row69	0.286	0.265	0.318	0.27	0.269	0
Row74	0.314	0.293	0.297	0.28	0.272	0
Row79	0.31	0.317	0.32	0.309	0.299	0
Row84	0.33	0.35	0.361	0.343	0.34	0
Row89	0.314	0.342	0.343	0.323	0.317	0
Row94	0.31	0.313	0.341	0.309	0.304	0
Row99	0.362	0.352	0.386	0.356	0.339	0
Row104	0.392	0.388	0.403	0.36	0.39	0.001
Row109	0.408	0.416	0.436	0.421	0.398	0
Row114	0.408	0.402	0.432	0.401	0.395	0
Row119	0.358	0.356	0.38	0.358	0.34	0
Row124	0.416	0.408	0.438	0.403	0.41	0

Lampiran 20. Hasil MLP Predictor Data Testing Partitioning Cross Validation  
(KAEF, TLKM)

Classified Data - 2:44 - MultiLayerPerceptron Predictor (Prediksi)

File Hilite Navigation View

Table "default" - Rows: 101 Spec - Columns: 5 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)
Row18	0.12	0.107	0.121	0.11	0.149
Row51	0.203	0.191	0.207	0.2	0.201
Row57	0.206	0.203	0.22	0.217	0.209
Row65	0.209	0.2	0.221	0.213	0.209
Row70	0.193	0.181	0.19	0.181	0.193
Row73	0.16	0.156	0.173	0.169	0.177
Row85	0.174	0.178	0.189	0.189	0.189
Row107	0.233	0.241	0.249	0.254	0.233
Row122	0.241	0.228	0.251	0.24	0.231
Row127	0.272	0.276	0.286	0.287	0.264
Row146	0.295	0.284	0.309	0.297	0.279
Row155	0.309	0.296	0.327	0.314	0.293
Row170	0.259	0.253	0.269	0.263	0.248
Row192	0.293	0.287	0.303	0.294	0.278
Row200	0.274	0.262	0.286	0.274	0.259
Row205	0.266	0.257	0.277	0.267	0.254
Row208	0.262	0.255	0.266	0.266	0.249
Row216	0.288	0.284	0.306	0.295	0.277
Row231	0.339	0.341	0.354	0.343	0.328
Row253	0.523	0.532	0.517	0.516	0.533
Row292	0.615	0.621	0.634	0.618	0.641
Row295	0.646	0.638	0.643	0.646	0.659
Row298	0.601	0.604	0.606	0.598	0.62
Row301	0.545	0.579	0.569	0.595	0.581
Row303	0.752	0.744	0.703	0.7	0.741

Classified Data - 3:47 - MultiLayerPerceptron Predictor (Prediksi)

File Hilite Navigation View

Table "default" - Rows: 101 Spec - Columns: 5 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)
Row18	0.128	0.116	0.158	0.1	0.135
Row28	0.158	0.146	0.189	0.139	0.152
Row47	0.176	0.154	0.197	0.153	0.158
Row52	0.172	0.154	0.204	0.149	0.16
Row55	0.226	0.211	0.237	0.188	0.192
Row62	0.24	0.251	0.274	0.235	0.221
Row85	0.38	0.414	0.397	0.366	0.366
Row103	0.396	0.41	0.424	0.399	0.381
Row107	0.428	0.464	0.455	0.468	0.431
Row112	0.404	0.396	0.421	0.397	0.372
Row138	0.432	0.412	0.428	0.388	0.388
Row155	0.44	0.42	0.467	0.427	0.415
Row160	0.58	0.585	0.591	0.583	0.589
Row166	0.464	0.468	0.482	0.444	0.453
Row192	0.596	0.581	0.602	0.57	0.595
Row206	0.66	0.666	0.679	0.665	0.681
Row221	0.816	0.815	0.822	0.812	0.801
Row238	0.756	0.75	0.73	0.763	0.744
Row246	0.728	0.746	0.745	0.755	0.747
Row251	0.74	0.726	0.71	0.693	0.725
Row260	0.772	0.762	0.764	0.767	0.762
Row293	0.78	0.779	0.741	0.722	0.76
Row296	0.62	0.613	0.56	0.587	0.59
Row297	0.612	0.642	0.606	0.607	0.629
Row311	0.64	0.662	0.656	0.656	0.666

Lampiran 21. Hasil MLP *Predictor Data Testing Partitioning* (KAEF, TLKM)

Classified Data - 2:43 - MultiLayerPerceptron Predictor (Prediksi)

File Hilite Navigation View

Table "default" - Rows: 253 Spec - Columns: 5 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)
Row4	0.127	0.118	0.137	0.133	0.156
Row9	0.122	0.113	0.134	0.124	0.153
Row14	0.129	0.12	0.134	0.131	0.156
Row19	0.107	0.103	0.114	0.114	0.146
Row24	0.127	0.118	0.134	0.123	0.155
Row29	0.166	0.159	0.177	0.169	0.179
Row34	0.191	0.179	0.191	0.186	0.192
Row39	0.189	0.179	0.194	0.182	0.193
Row44	0.18	0.171	0.193	0.185	0.188
Row49	0.206	0.217	0.217	0.223	0.213
Row54	0.209	0.198	0.214	0.208	0.206
Row59	0.213	0.2	0.224	0.213	0.21
Row64	0.21	0.2	0.224	0.212	0.21
Row69	0.195	0.182	0.21	0.198	0.198
Row74	0.17	0.164	0.184	0.174	0.183
Row79	0.181	0.168	0.184	0.174	0.186
Row84	0.174	0.165	0.184	0.181	0.184
Row89	0.184	0.174	0.191	0.183	0.19
Row94	0.198	0.189	0.206	0.2	0.2
Row99	0.192	0.188	0.203	0.199	0.198
Row104	0.245	0.253	0.261	0.263	0.244
Row109	0.244	0.235	0.251	0.242	0.234
Row114	0.259	0.249	0.274	0.261	0.248
Row119	0.241	0.232	0.251	0.247	0.233
Row124	0.244	0.253	0.26	0.268	0.243

Classified Data - 3:48 - MultiLayerPerceptron Predictor (Prediksi)

File Hilite Navigation View

Table "default" - Rows: 253 Spec - Columns: 5 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)
Row4	0.184	0.164	0.195	0.141	0.161
Row9	0.2	0.179	0.222	0.176	0.175
Row14	0.156	0.152	0.189	0.153	0.154
Row19	0.1	0.1	0.114	0.102	0.118
Row24	0.188	0.191	0.21	0.178	0.173
Row29	0.158	0.138	0.171	0.116	0.145
Row34	0.178	0.164	0.202	0.153	0.163
Row39	0.186	0.181	0.222	0.18	0.174
Row44	0.176	0.16	0.206	0.153	0.163
Row49	0.152	0.138	0.181	0.127	0.148
Row54	0.216	0.205	0.245	0.202	0.192
Row59	0.236	0.229	0.258	0.223	0.207
Row64	0.254	0.251	0.287	0.251	0.228
Row69	0.286	0.265	0.318	0.27	0.249
Row74	0.314	0.293	0.297	0.28	0.253
Row79	0.31	0.317	0.32	0.309	0.274
Row84	0.33	0.35	0.361	0.343	0.31
Row89	0.314	0.342	0.343	0.323	0.296
Row94	0.31	0.313	0.341	0.309	0.282
Row99	0.362	0.352	0.386	0.356	0.327
Row104	0.392	0.388	0.403	0.36	0.357
Row109	0.408	0.416	0.436	0.421	0.392
Row114	0.408	0.402	0.432	0.401	0.383
Row119	0.358	0.356	0.38	0.358	0.326
Row124	0.416	0.408	0.438	0.403	0.39

Lampiran 22. Hasil Dari Proses *X-Aggregator* MLP (KAEF, TLKM)

**Prediction table - 2:45 - X-Aggregator**

File [Hilite](#) [Navigation](#) [View](#)

Table "default" - Rows: 1010 Spec - Columns: 5 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)
Row0	0.133	0.138	0.146	0.142	0.155
Row20	0.114	0.106	0.124	0.117	0.14
Row23	0.114	0.115	0.127	0.124	0.143
Row31	0.166	0.164	0.177	0.178	0.174
Row42	0.182	0.175	0.187	0.176	0.182
Row56	0.202	0.196	0.217	0.208	0.2
Row58	0.216	0.209	0.227	0.216	0.209
Row86	0.185	0.181	0.199	0.186	0.187
Row91	0.182	0.174	0.191	0.188	0.183
Row96	0.203	0.193	0.204	0.195	0.195
Row112	0.244	0.248	0.254	0.261	0.236
Row123	0.238	0.23	0.249	0.246	0.226
Row143	0.294	0.283	0.307	0.295	0.277
Row172	0.274	0.264	0.274	0.277	0.254
Row173	0.273	0.271	0.284	0.28	0.26
Row195	0.276	0.263	0.286	0.274	0.258
Row207	0.266	0.253	0.274	0.266	0.248
Row223	0.284	0.281	0.301	0.291	0.272
Row255	0.57	0.593	0.549	0.539	0.589
Row268	0.673	0.66	0.663	0.652	0.681
Row285	0.556	0.549	0.549	0.542	0.565
Row311	0.682	0.685	0.683	0.672	0.699
Row318	0.721	0.721	0.731	0.723	0.733
Row326	0.69	0.688	0.714	0.697	0.711
Row331	0.699	0.688	0.717	0.7	0.714

**Prediction table - 3:45 - X-Aggregator**

File [Hilite](#) [Navigation](#) [View](#)

Table "default" - Rows: 1010 Spec - Columns: 5 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)
Row0	0.226	0.209	0.247	0.194	0.179
Row7	0.168	0.172	0.198	0.167	0.147
Row20	0.112	0.116	0.144	0.118	0.115
Row41	0.18	0.162	0.189	0.133	0.146
Row48	0.16	0.148	0.195	0.139	0.137
Row56	0.202	0.227	0.237	0.229	0.18
Row67	0.28	0.283	0.312	0.276	0.236
Row77	0.304	0.305	0.328	0.309	0.258
Row86	0.36	0.36	0.372	0.35	0.316
Row105	0.36	0.37	0.382	0.376	0.323
Row123	0.38	0.378	0.409	0.384	0.339
Row133	0.436	0.436	0.455	0.444	0.408
Row140	0.41	0.396	0.421	0.388	0.365
Row143	0.418	0.402	0.428	0.397	0.373
Row150	0.458	0.438	0.484	0.446	0.422
Row182	0.58	0.565	0.591	0.575	0.576
Row197	0.636	0.63	0.645	0.628	0.645
Row205	0.624	0.65	0.641	0.66	0.654
Row235	0.756	0.766	0.761	0.775	0.762
Row241	0.76	0.754	0.764	0.759	0.758
Row250	0.76	0.746	0.761	0.738	0.754
Row252	0.7	0.698	0.691	0.693	0.709
Row256	0.732	0.758	0.749	0.755	0.753
Row258	0.744	0.75	0.737	0.755	0.751
Row262	0.756	0.774	0.764	0.787	0.766

Lampiran 23. Hasil *Error Rate X-Aggregator MLP Partitioning* (KAEF,TLKM)

**Error rates - 2:45 - X-Aggregator**

File Hilite Navigation View

Table "default" - Rows: 10 Spec - Columns: 3 Properties Flow Variables

Row ID	D Total squared error	D Mean squared error	I Size of Test Set
fold 0	0.065	0.001	101
fold 1	0.079	0.001	101
fold 2	0.065	0.001	101
fold 3	0.063	0.001	101
fold 4	0.071	0.001	101
fold 5	0.048	0	101
fold 6	0.048	0	101
fold 7	0.071	0.001	101
fold 8	0.072	0.001	101
fold 9	0.075	0.001	101

**Error rates - 3:45 - X-Aggregator**

File Hilite Navigation View

Table "default" - Rows: 10 Spec - Columns: 3 Properties Flow Variables

Row ID	D Total squared error	D Mean squared error	I Size of Test Set
fold 0	0.107	0.001	101
fold 1	0.08	0.001	101
fold 2	0.142	0.001	101
fold 3	0.084	0.001	101
fold 4	0.111	0.001	101
fold 5	0.106	0.001	101
fold 6	0.096	0.001	101
fold 7	0.096	0.001	101
fold 8	0.097	0.001	101
fold 9	0.074	0.001	101

Lampiran 24. Hasil Node *Predictor Regression Cross Validation* (KAEF,TLKM)

▲ Predicted data - 2:56 - Regression Predictor (Prediksi)

File Hilite Navigation View

Table "default" - Rows: 126 Spec - Columns: 5 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)
Row7	0.131	0.122	0.141	0.131	0.134
Row22	0.117	0.104	0.124	0.113	0.114
Row30	0.167	0.157	0.177	0.168	0.168
Row56	0.202	0.196	0.217	0.208	0.21
Row57	0.206	0.203	0.22	0.217	0.215
Row78	0.182	0.172	0.193	0.181	0.184
Row81	0.181	0.171	0.193	0.183	0.184
Row93	0.195	0.189	0.21	0.199	0.203
Row96	0.203	0.193	0.204	0.195	0.197
Row112	0.244	0.248	0.254	0.261	0.254
Row119	0.241	0.232	0.251	0.247	0.243
Row123	0.238	0.23	0.249	0.246	0.24
Row144	0.294	0.289	0.309	0.304	0.301
Row147	0.293	0.283	0.307	0.294	0.296
Row149	0.295	0.299	0.309	0.307	0.307
Row173	0.273	0.271	0.284	0.28	0.28
Row181	0.263	0.255	0.276	0.267	0.266
Row188	0.27	0.257	0.283	0.273	0.271
Row199	0.28	0.271	0.287	0.274	0.279
Row210	0.262	0.256	0.277	0.266	0.269
Row214	0.272	0.28	0.287	0.288	0.289
Row252	0.615	0.613	0.574	0.558	0.576
Row257	0.523	0.521	0.523	0.508	0.518
Row264	0.626	0.635	0.64	0.626	0.639
Row267	0.679	0.682	0.686	0.669	0.682

▲ Predicted data - 3:57 - Regression Predictor (Prediksi)

File Hilite Navigation View

Table "default" - Rows: 126 Spec - Columns: 5 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)
Row9	0.2	0.179	0.222	0.176	0.169
Row30	0.144	0.15	0.181	0.135	0.145
Row40	0.198	0.177	0.204	0.153	0.154
Row41	0.18	0.162	0.189	0.133	0.141
Row53	0.18	0.199	0.212	0.188	0.188
Row61	0.264	0.247	0.291	0.247	0.244
Row76	0.312	0.295	0.332	0.288	0.288
Row78	0.32	0.313	0.351	0.319	0.314
Row86	0.36	0.36	0.372	0.35	0.346
Row95	0.318	0.311	0.347	0.3	0.31
Row101	0.388	0.372	0.394	0.37	0.357
Row114	0.408	0.402	0.432	0.401	0.401
Row117	0.354	0.344	0.367	0.337	0.332
Row137	0.456	0.444	0.461	0.435	0.432
Row146	0.426	0.42	0.452	0.417	0.421
Row154	0.44	0.42	0.467	0.427	0.426
Row176	0.568	0.553	0.536	0.521	0.517
Row189	0.632	0.634	0.641	0.636	0.63
Row194	0.58	0.573	0.591	0.579	0.57
Row195	0.596	0.585	0.602	0.595	0.58
Row196	0.616	0.613	0.633	0.624	0.616
Row211	0.74	0.73	0.73	0.718	0.718
Row221	0.816	0.815	0.822	0.812	0.818
Row230	0.812	0.811	0.803	0.804	0.802
Row237	0.744	0.73	0.741	0.738	0.725

Lampiran 25. Hasil Prediksi Node *X-Aggregator* (Tanpa *Partitioning*) LR  
(KAEF, TLKM)

**▲ Prediction table - 2:53 - X-Aggregator**

File [Hilite](#) [Navigation](#) [View](#)

Table "default" - Rows: 1263 Spec - Columns: 5 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)
Row0	0.133	0.138	0.146	0.142	0.146
Row13	0.138	0.132	0.141	0.134	0.137
Row24	0.127	0.118	0.134	0.123	0.127
Row31	0.166	0.164	0.177	0.178	0.174
Row35	0.184	0.184	0.199	0.199	0.195
Row43	0.175	0.17	0.19	0.181	0.183
Row46	0.184	0.191	0.197	0.202	0.199
Row50	0.22	0.214	0.219	0.208	0.214
Row60	0.212	0.203	0.226	0.216	0.217
Row80	0.173	0.168	0.187	0.182	0.181
Row100	0.199	0.243	0.214	0.244	0.24
Row109	0.244	0.235	0.251	0.242	0.243
Row114	0.259	0.249	0.274	0.261	0.264
Row116	0.258	0.245	0.264	0.254	0.253
Row124	0.244	0.253	0.26	0.268	0.262
Row157	0.304	0.294	0.311	0.3	0.302
Row158	0.295	0.287	0.303	0.292	0.294
Row170	0.259	0.253	0.269	0.263	0.262
Row172	0.274	0.264	0.274	0.277	0.266
Row189	0.269	0.259	0.283	0.273	0.272
Row211	0.263	0.255	0.276	0.267	0.267
Row218	0.29	0.28	0.304	0.291	0.294
Row221	0.272	0.266	0.289	0.281	0.28
Row231	0.339	0.341	0.354	0.343	0.351
Row242	0.64	0.738	0.666	0.734	0.722

**▲ Prediction table - 3:58 - X-Aggregator**

File [Hilite](#) [Navigation](#) [View](#)

Table "default" - Rows: 1263 Spec - Columns: 5 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)
Row0	0.226	0.209	0.247	0.194	0.199
Row6	0.18	0.16	0.208	0.153	0.155
Row22	0.122	0.112	0.158	0.108	0.108
Row26	0.174	0.156	0.2	0.145	0.149
Row31	0.17	0.148	0.187	0.133	0.134
Row37	0.196	0.189	0.224	0.184	0.181
Row54	0.216	0.205	0.245	0.202	0.2
Row70	0.26	0.267	0.293	0.268	0.263
Row72	0.24	0.241	0.274	0.235	0.238
Row89	0.314	0.342	0.343	0.323	0.333
Row108	0.452	0.432	0.452	0.417	0.417
Row113	0.404	0.396	0.43	0.401	0.397
Row118	0.356	0.36	0.378	0.354	0.353
Row124	0.416	0.408	0.438	0.403	0.406
Row128	0.432	0.426	0.44	0.423	0.413
Row140	0.41	0.396	0.421	0.388	0.386
Row144	0.412	0.4	0.43	0.407	0.396
Row155	0.44	0.42	0.467	0.427	0.426
Row184	0.56	0.561	0.571	0.55	0.556
Row187	0.608	0.597	0.602	0.591	0.583
Row193	0.568	0.569	0.591	0.575	0.573
Row204	0.624	0.621	0.614	0.599	0.603
Row235	0.756	0.766	0.761	0.775	0.763
Row246	0.728	0.746	0.745	0.755	0.75
Row273	0.732	0.726	0.745	0.73	0.732

Lampiran 26. Hasil *Error Rate Cross Validation (Tanpa Partitioning)* LR (KAEF, TLKM)

 Error rates - 2:53 - X-Aggregator

File Hilite Navigation View

Table "default" - Rows: 10 Spec - Columns: 3 Properties Flow Variables

Row ID	D Total squared error	D Mean squared error	I Size of Test Set
fold 0	0.012	0	127
fold 1	0.008	0	127
fold 2	0.012	0	127
fold 3	0.014	0	126
fold 4	0.017	0	126
fold 5	0.014	0	126
fold 6	0.019	0	126
fold 7	0.028	0	126
fold 8	0.019	0	126
fold 9	0.011	0	126

 Error rates - 3:58 - X-Aggregator

File Hilite Navigation View

Table "default" - Rows: 10 Spec - Columns: 3 Properties Flow Variables

Row ID	D Total squared error	D Mean squared error	I Size of Test Set
fold 0	0.009	0	127
fold 1	0.018	0	127
fold 2	0.016	0	127
fold 3	0.019	0	126
fold 4	0.013	0	126
fold 5	0.013	0	126
fold 6	0.015	0	126
fold 7	0.01	0	126
fold 8	0.012	0	126
fold 9	0.015	0	126

Lampiran 27. Hasil *Attribute Statistic RFR Cross Validation* (KAEF, TLKM)

Attribute Statistics - 2:64 - Random Forest Learner (Regression) (Modeling)

File Hilite Navigation View

Table "Tree Ensemble Column Statistic" - Rows: 3 Spec - Columns: 6 Properties Flow Variables

Row ID	#splits (level 0)	#splits (level 1)	#splits (level 2)	#candidates (level 0)	#candidates (level 1)	#candidates (level 2)
Open	35	56	144	35	56	144
High	28	73	126	28	73	126
Low	37	71	130	37	71	130

Attribute Statistics - 3:65 - Random Forest Learner (Regression) (Modeling)

File Hilite Navigation View

Table "Tree Ensemble Column Statistic" - Rows: 3 Spec - Columns: 6 Properties Flow Variables

Row ID	#splits (level 0)	#splits (level 1)	#splits (level 2)	#candidates (level 0)	#candidates (level 1)	#candidates (level 2)
Open	33	67	115	33	67	115
High	34	73	140	34	73	140
Low	33	60	145	33	60	145

Lampiran 28. Hasil *Out of Bag Prediction RFR Cross Validation* (KAEF, TLKM)

Out-of-bag Predictions - 2:64 - Random Forest Learner (Regression) (Modeling)

File Hilite Navigation View

Table "default" - Rows: 1137 Spec - Columns: 7 Properties Flow Variables

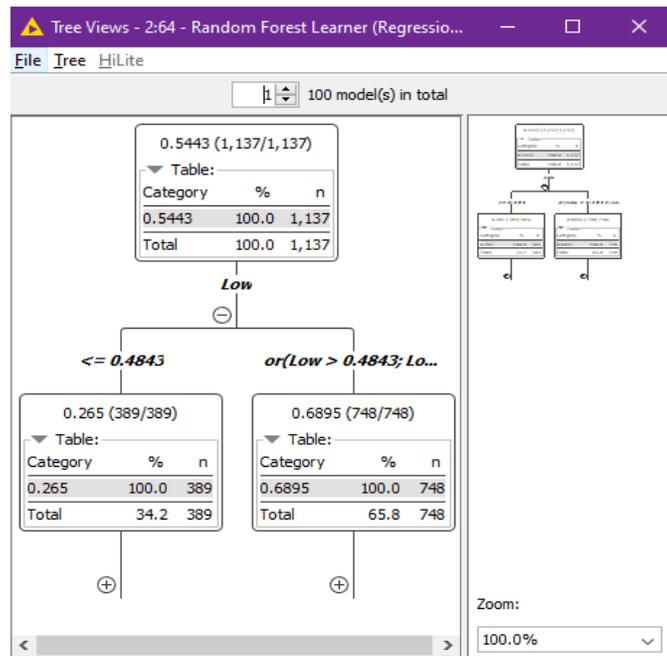
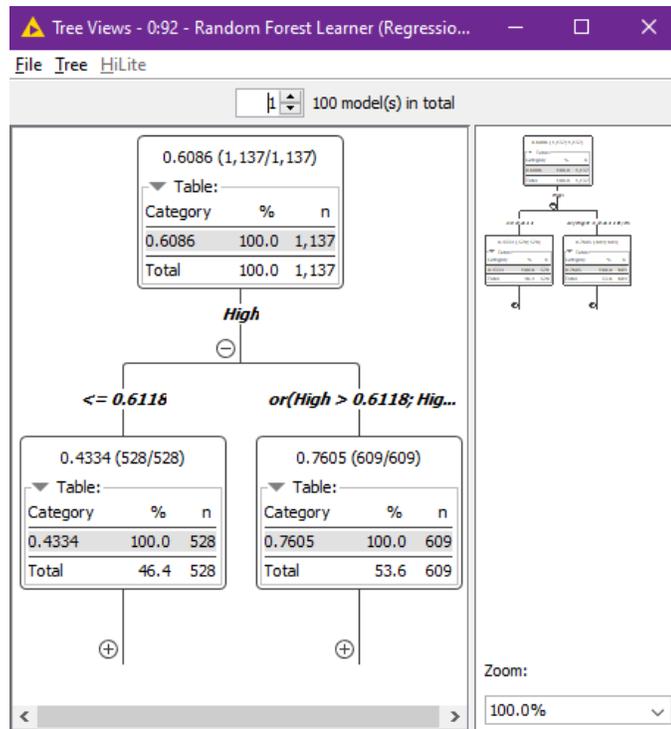
Row ID	Open	High	Low	Close	Close (Out-of-bag)	Close (Out-of-bag) (Prediction Variance)	model count
Row0	0.133	0.138	0.146	0.142	0.142	0	43
Row1	0.138	0.128	0.143	0.141	0.135	0	33
Row2	0.139	0.128	0.149	0.138	0.142	0	39
Row3	0.135	0.125	0.144	0.135	0.134	0	33
Row4	0.127	0.118	0.137	0.133	0.127	0	36
Row5	0.128	0.12	0.137	0.131	0.128	0	41
Row6	0.129	0.122	0.143	0.135	0.133	0	35
Row7	0.131	0.122	0.141	0.131	0.136	0	41
Row8	0.129	0.12	0.13	0.124	0.123	0	32
Row9	0.122	0.113	0.134	0.124	0.126	0	34
Row10	0.122	0.114	0.134	0.124	0.126	0	38
Row11	0.124	0.113	0.134	0.127	0.124	0	41
Row12	0.124	0.122	0.137	0.137	0.131	0	47
Row15	0.128	0.12	0.137	0.127	0.131	0	37
Row16	0.125	0.113	0.13	0.121	0.121	0	34
Row17	0.12	0.113	0.13	0.12	0.121	0	38
Row18	0.12	0.107	0.121	0.11	0.118	0	33
Row19	0.107	0.103	0.114	0.114	0.112	0	20
Row20	0.114	0.106	0.124	0.117	0.113	0	43
Row21	0.114	0.108	0.123	0.117	0.114	0	34
Row23	0.114	0.115	0.127	0.124	0.116	0	35
Row24	0.127	0.118	0.134	0.123	0.13	0	43
Row25	0.122	0.131	0.134	0.144	0.131	0	33
Row26	0.143	0.146	0.151	0.144	0.149	0	41
Row27	0.146	0.142	0.16	0.155	0.152	0	39

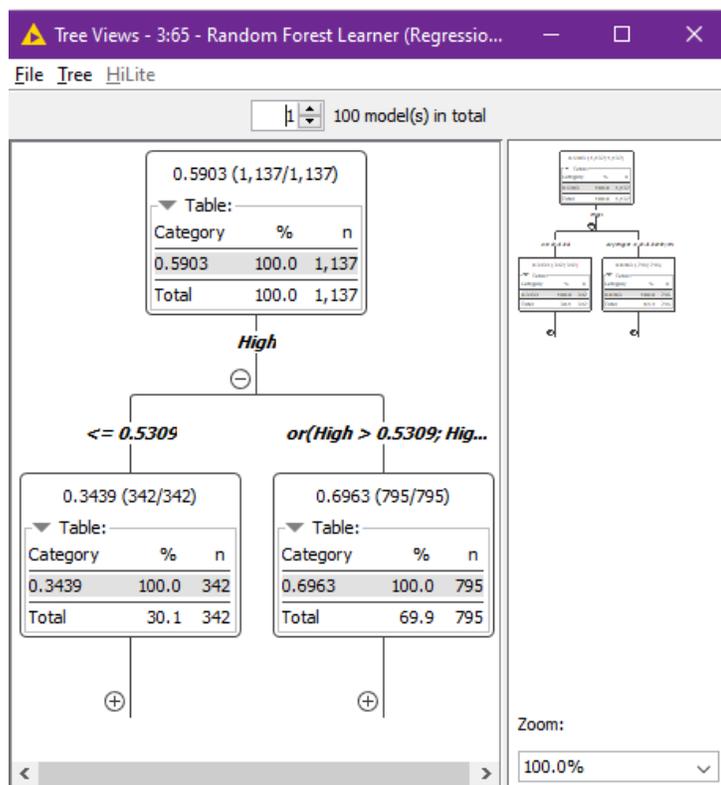
Out-of-bag Predictions - 3:65 - Random Forest Learner (Regression) (Modeling)

File HiLite Navigation View

Table "default" - Rows: 1137 Spec - Columns: 7 Properties Flow Variables

Row ID	D] Open	D] High	D] Low	D] Close	D] Close (Out-of-bag)	D] Close (Out-of-bag) (Prediction Variance)	I] model count
Row0	0.226	0.209	0.247	0.194	0.193	0.001	34
Row1	0.196	0.183	0.222	0.172	0.169	0	37
Row2	0.202	0.187	0.229	0.182	0.19	0.001	36
Row3	0.2	0.187	0.235	0.186	0.186	0	38
Row4	0.184	0.164	0.195	0.141	0.143	0	39
Row5	0.152	0.154	0.189	0.153	0.147	0	38
Row6	0.18	0.16	0.208	0.153	0.152	0	48
Row7	0.168	0.172	0.198	0.167	0.154	0	46
Row8	0.2	0.179	0.22	0.165	0.175	0	36
Row10	0.182	0.166	0.2	0.145	0.151	0	35
Row11	0.168	0.156	0.197	0.149	0.151	0	42
Row13	0.156	0.146	0.187	0.137	0.141	0	35
Row14	0.156	0.152	0.189	0.153	0.146	0	37
Row15	0.174	0.156	0.195	0.141	0.144	0	48
Row16	0.146	0.138	0.179	0.122	0.126	0	39
Row17	0.168	0.168	0.173	0.122	0.136	0	27
Row18	0.128	0.116	0.158	0.1	0.115	0	44
Row19	0.1	0.1	0.114	0.102	0.151	0.003	36
Row20	0.112	0.116	0.144	0.118	0.12	0.002	41
Row21	0.136	0.124	0.162	0.112	0.114	0	34
Row22	0.122	0.112	0.158	0.108	0.109	0	41
Row23	0.128	0.16	0.164	0.149	0.125	0	41
Row24	0.188	0.191	0.21	0.178	0.175	0	40
Row25	0.19	0.174	0.197	0.157	0.157	0	40

Lampiran 29. Hasil *Tree Views* RFR Cross Validation (INDF, KAEF, TLKM)



Lampiran 30. Hasil Prediksi Node RFR *Predictor Cross Validation* (KAEF, TLKM)

▲ Prediction output - 2:65 - Random Forest Predictor (Regression) (Prediksi)

File Hilite Navigation View

Table "default" - Rows: 126 Spec - Columns: 6 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)	D Prediction (Close) (Prediction Variance)
Row13	0.138	0.132	0.141	0.134	0.136	0
Row14	0.129	0.12	0.134	0.131	0.127	0
Row22	0.117	0.104	0.124	0.113	0.116	0
Row37	0.202	0.191	0.211	0.199	0.201	0
Row54	0.209	0.198	0.214	0.208	0.204	0
Row57	0.206	0.203	0.22	0.217	0.212	0
Row72	0.17	0.159	0.17	0.161	0.167	0
Row84	0.174	0.165	0.184	0.181	0.17	0
Row89	0.184	0.174	0.191	0.183	0.185	0
Row98	0.191	0.185	0.206	0.195	0.194	0
Row100	0.199	0.243	0.214	0.244	0.229	0.001
Row113	0.262	0.249	0.271	0.261	0.263	0
Row123	0.238	0.23	0.249	0.246	0.242	0
Row161	0.309	0.296	0.309	0.3	0.304	0
Row170	0.259	0.253	0.269	0.263	0.262	0
Row194	0.279	0.273	0.286	0.273	0.276	0
Row198	0.269	0.266	0.286	0.281	0.277	0
Row199	0.28	0.271	0.287	0.274	0.276	0
Row205	0.266	0.257	0.277	0.267	0.267	0
Row217	0.291	0.283	0.304	0.294	0.294	0
Row222	0.28	0.277	0.297	0.288	0.284	0
Row231	0.339	0.341	0.354	0.343	0.342	0
Row238	0.456	0.485	0.46	0.505	0.476	0.001
Row242	0.64	0.738	0.666	0.734	0.71	0.001
Row251	0.662	0.663	0.62	0.624	0.64	0.001

▲ Prediction output - 3:66 - Random Forest Predictor (Regression) (Prediksi)

File Hilite Navigation View

Table "default" - Rows: 126 Spec - Columns: 6 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)	D Prediction (Close) (Prediction Variance)
Row9	0.2	0.179	0.222	0.176	0.171	0
Row12	0.16	0.142	0.179	0.133	0.131	0
Row30	0.144	0.15	0.181	0.135	0.139	0
Row44	0.176	0.16	0.206	0.153	0.152	0
Row55	0.226	0.211	0.237	0.188	0.192	0.001
Row70	0.26	0.267	0.293	0.268	0.26	0
Row76	0.312	0.295	0.332	0.288	0.288	0
Row91	0.32	0.346	0.351	0.319	0.338	0
Row101	0.388	0.372	0.394	0.37	0.352	0.001
Row108	0.452	0.432	0.452	0.417	0.425	0
Row113	0.404	0.396	0.43	0.401	0.393	0
Row117	0.354	0.344	0.367	0.337	0.332	0
Row137	0.456	0.444	0.461	0.435	0.425	0
Row148	0.448	0.432	0.467	0.427	0.429	0
Row167	0.464	0.46	0.475	0.464	0.449	0
Row208	0.712	0.762	0.71	0.701	0.722	0
Row217	0.74	0.722	0.733	0.718	0.72	0
Row221	0.816	0.815	0.822	0.812	0.814	0
Row232	0.792	0.787	0.764	0.767	0.772	0
Row236	0.78	0.77	0.745	0.73	0.748	0.001
Row237	0.744	0.73	0.741	0.738	0.729	0
Row244	0.744	0.758	0.757	0.759	0.756	0
Row253	0.7	0.682	0.679	0.66	0.66	0
Row309	0.608	0.621	0.625	0.615	0.617	0
Row315	0.652	0.654	0.664	0.656	0.651	0

Lampiran 31. Hasil Prediksi Node X-Aggregator RFR Cross Validation  
(KAEF, TLKM)

▲ Prediction table - 2:62 - X-Aggregator

File Hilite Navigation View

Table "default" - Rows: 1263 Spec - Columns: 6 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)	D Prediction (Close) (Prediction Variance)
Row0	0.133	0.138	0.146	0.142	0.142	0
Row23	0.114	0.115	0.127	0.124	0.118	0
Row24	0.127	0.118	0.134	0.123	0.131	0
Row31	0.166	0.164	0.177	0.178	0.17	0
Row39	0.189	0.179	0.194	0.182	0.189	0
Row41	0.166	0.167	0.18	0.181	0.171	0
Row46	0.184	0.191	0.197	0.202	0.194	0
Row56	0.202	0.196	0.217	0.208	0.207	0
Row58	0.216	0.209	0.227	0.216	0.225	0
Row93	0.195	0.189	0.21	0.199	0.2	0
Row112	0.244	0.248	0.254	0.261	0.251	0
Row120	0.244	0.235	0.254	0.242	0.245	0
Row126	0.277	0.267	0.286	0.273	0.275	0
Row132	0.3	0.287	0.3	0.291	0.295	0
Row135	0.305	0.292	0.31	0.298	0.302	0
Row136	0.294	0.284	0.306	0.292	0.296	0
Row148	0.293	0.283	0.309	0.3	0.298	0
Row163	0.291	0.281	0.289	0.275	0.286	0
Row165	0.273	0.266	0.289	0.277	0.279	0
Row167	0.267	0.255	0.276	0.268	0.268	0
Row179	0.263	0.253	0.274	0.267	0.266	0
Row186	0.26	0.266	0.277	0.274	0.271	0
Row191	0.269	0.288	0.286	0.297	0.291	0
Row192	0.293	0.287	0.303	0.294	0.294	0
Row206	0.272	0.263	0.276	0.263	0.273	0

▲ Prediction table - 3:63 - X-Aggregator

File Hilite Navigation View

Table "default" - Rows: 1263 Spec - Columns: 6 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)	D Prediction (Close) (Prediction Variance)
Row0	0.226	0.209	0.247	0.194	0.194	0
Row22	0.122	0.112	0.158	0.108	0.109	0
Row26	0.174	0.156	0.2	0.145	0.15	0
Row37	0.196	0.189	0.224	0.184	0.178	0
Row40	0.198	0.177	0.204	0.153	0.164	0
Row54	0.216	0.205	0.245	0.202	0.188	0
Row72	0.24	0.241	0.274	0.235	0.239	0
Row83	0.312	0.33	0.343	0.321	0.314	0
Row87	0.348	0.336	0.363	0.323	0.325	0
Row102	0.396	0.396	0.401	0.403	0.373	0
Row114	0.408	0.402	0.432	0.401	0.4	0
Row120	0.378	0.384	0.394	0.386	0.365	0
Row128	0.432	0.426	0.44	0.423	0.418	0
Row138	0.432	0.412	0.428	0.388	0.399	0
Row146	0.426	0.42	0.452	0.417	0.425	0
Row156	0.44	0.42	0.467	0.427	0.426	0
Row158	0.55	0.531	0.542	0.521	0.519	0
Row179	0.556	0.565	0.579	0.562	0.559	0
Row181	0.6	0.581	0.591	0.579	0.571	0
Row187	0.608	0.597	0.602	0.591	0.582	0
Row228	0.824	0.823	0.83	0.836	0.824	0
Row230	0.812	0.811	0.803	0.804	0.799	0
Row261	0.76	0.762	0.772	0.771	0.76	0
Row271	0.752	0.754	0.757	0.738	0.752	0
Row280	0.76	0.758	0.772	0.759	0.76	0

Lampiran 32. *Error Rate Cross Validation Model RFR Tanpa Partitioning*  
(KAEF, TLKM)

 Error rates - 2:62 - X-Aggregator

File Hilite Navigation View

Table "default" - Rows: 10 Spec - Columns: 3 Properties Flow Variables

Row ID	D Total squared error	D Mean squared error	I Size of Test Set
fold 0	0.023	0	127
fold 1	0.015	0	127
fold 2	0.019	0	127
fold 3	0.022	0	126
fold 4	0.025	0	126
fold 5	0.039	0	126
fold 6	0.014	0	126
fold 7	0.014	0	126
fold 8	0.015	0	126
fold 9	0.023	0	126

 Error rates - 3:63 - X-Aggregator

File Hilite Navigation View

Table "default" - Rows: 10 Spec - Columns: 3 Properties Flow Variables

Row ID	D Total squared error	D Mean squared error	I Size of Test Set
fold 0	0.026	0	127
fold 1	0.019	0	127
fold 2	0.016	0	127
fold 3	0.026	0	126
fold 4	0.028	0	126
fold 5	0.021	0	126
fold 6	0.02	0	126
fold 7	0.023	0	126
fold 8	0.023	0	126
fold 9	0.018	0	126

Lampiran 33. Hasil Node MLP Predictor Data Testing Cross Validation  
(KAEF, TLKM)

Classified Data - 0:70 - MultiLayerPerceptron Predictor (Prediksi)

File Hilite Navigation View

Table "default" - Rows: 126 Spec - Columns: 5 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)
Row7	0.131	0.122	0.141	0.131	0.145
Row22	0.117	0.104	0.124	0.113	0.135
Row30	0.167	0.157	0.177	0.168	0.168
Row57	0.206	0.203	0.22	0.217	0.2
Row66	0.207	0.195	0.217	0.208	0.197
Row72	0.17	0.159	0.17	0.161	0.168
Row81	0.181	0.171	0.193	0.183	0.179
Row84	0.174	0.165	0.184	0.181	0.174
Row96	0.203	0.193	0.204	0.195	0.193
Row99	0.192	0.188	0.203	0.199	0.188
Row114	0.259	0.249	0.274	0.261	0.244
Row119	0.241	0.232	0.251	0.247	0.227
Row123	0.238	0.23	0.249	0.246	0.225
Row126	0.277	0.267	0.286	0.273	0.26
Row138	0.291	0.285	0.307	0.3	0.277
Row144	0.294	0.289	0.309	0.304	0.28
Row149	0.295	0.299	0.309	0.307	0.285
Row164	0.272	0.26	0.271	0.277	0.251
Row198	0.269	0.266	0.286	0.281	0.256
Row200	0.274	0.262	0.286	0.274	0.256
Row208	0.262	0.255	0.266	0.266	0.245
Row214	0.272	0.28	0.287	0.288	0.263
Row217	0.291	0.283	0.304	0.294	0.275
Row238	0.456	0.485	0.46	0.505	0.478
Row243	0.732	0.76	0.746	0.779	0.747

Classified Data - 3:73 - MultiLayerPerceptron Predictor (Prediksi)

File Hilite Navigation View

Table "default" - Rows: 126 Spec - Columns: 5 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)
Row5	0.152	0.154	0.189	0.153	0.128
Row9	0.2	0.179	0.222	0.176	0.154
Row12	0.16	0.142	0.179	0.133	0.125
Row22	0.122	0.112	0.158	0.108	0.106
Row30	0.144	0.15	0.181	0.135	0.123
Row55	0.226	0.211	0.237	0.188	0.174
Row70	0.26	0.267	0.293	0.268	0.22
Row91	0.32	0.346	0.351	0.319	0.291
Row95	0.318	0.311	0.347	0.3	0.277
Row101	0.388	0.372	0.394	0.37	0.347
Row108	0.452	0.432	0.452	0.417	0.425
Row113	0.404	0.396	0.43	0.401	0.381
Row128	0.432	0.426	0.44	0.423	0.409
Row141	0.404	0.394	0.421	0.386	0.375
Row152	0.44	0.42	0.467	0.427	0.423
Row162	0.54	0.543	0.558	0.55	0.553
Row192	0.596	0.581	0.602	0.57	0.606
Row193	0.568	0.569	0.591	0.575	0.587
Row195	0.596	0.585	0.602	0.595	0.607
Row203	0.596	0.609	0.614	0.607	0.619
Row205	0.624	0.65	0.641	0.66	0.651
Row211	0.74	0.73	0.73	0.718	0.732
Row229	0.844	0.835	0.826	0.824	0.796
Row252	0.7	0.698	0.691	0.693	0.703
Row264	0.772	0.799	0.788	0.791	0.768

Lampiran 34. Hasil Prediksi Node X-Aggregator Model MLP Cross Validation  
(KAEF, TLKM)

**Prediction table - 0:67 - X-Aggregator**

File Hilite Navigation View

Table "default" - Rows: 1263 Spec - Columns: 5 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)
Row0	0.133	0.138	0.146	0.142	0.156
Row8	0.129	0.12	0.13	0.124	0.148
Row13	0.138	0.132	0.141	0.134	0.155
Row24	0.127	0.118	0.134	0.123	0.149
Row31	0.166	0.164	0.177	0.178	0.176
Row39	0.189	0.179	0.194	0.182	0.188
Row43	0.175	0.17	0.19	0.181	0.182
Row46	0.184	0.191	0.197	0.202	0.191
Row56	0.202	0.196	0.217	0.208	0.202
Row60	0.212	0.203	0.226	0.216	0.208
Row93	0.195	0.189	0.21	0.199	0.196
Row100	0.199	0.243	0.214	0.244	0.214
Row112	0.244	0.248	0.254	0.261	0.237
Row121	0.238	0.228	0.253	0.242	0.229
Row158	0.295	0.287	0.303	0.292	0.279
Row161	0.309	0.296	0.309	0.3	0.288
Row170	0.259	0.253	0.269	0.263	0.247
Row177	0.274	0.267	0.267	0.268	0.255
Row178	0.269	0.26	0.279	0.267	0.255
Row189	0.269	0.259	0.283	0.273	0.256
Row218	0.29	0.28	0.304	0.291	0.276
Row220	0.279	0.271	0.289	0.275	0.265
Row221	0.272	0.266	0.289	0.281	0.261
Row231	0.339	0.341	0.354	0.343	0.33
Row257	0.523	0.521	0.523	0.508	0.532

**Prediction table - 3:70 - X-Aggregator**

File Hilite Navigation View

Table "default" - Rows: 1263 Spec - Columns: 5 Properties Flow Variables

Row ID	D Open	D High	D Low	D Close	D Prediction (Close)
Row0	0.226	0.209	0.247	0.194	0.184
Row26	0.174	0.156	0.2	0.145	0.145
Row37	0.196	0.189	0.224	0.184	0.165
Row47	0.176	0.154	0.197	0.153	0.144
Row54	0.216	0.205	0.245	0.202	0.18
Row72	0.24	0.241	0.274	0.235	0.206
Row79	0.31	0.317	0.32	0.309	0.272
Row83	0.312	0.33	0.343	0.321	0.286
Row87	0.348	0.336	0.363	0.323	0.305
Row102	0.396	0.396	0.401	0.403	0.366
Row114	0.408	0.402	0.432	0.401	0.383
Row120	0.378	0.384	0.394	0.386	0.351
Row135	0.462	0.46	0.479	0.452	0.449
Row138	0.432	0.412	0.428	0.388	0.395
Row146	0.426	0.42	0.452	0.417	0.405
Row154	0.44	0.42	0.467	0.427	0.415
Row164	0.514	0.499	0.517	0.489	0.501
Row182	0.58	0.565	0.591	0.575	0.58
Row213	0.752	0.738	0.741	0.73	0.736
Row220	0.8	0.803	0.799	0.812	0.776
Row225	0.82	0.815	0.803	0.783	0.783
Row226	0.808	0.803	0.807	0.804	0.778
Row256	0.732	0.758	0.749	0.755	0.741
Row261	0.76	0.762	0.772	0.771	0.752
Row262	0.756	0.774	0.764	0.787	0.754

Lampiran 35. *Error Rate Cross Validation Model MLP Tanpa Partitioning*  
(KAEF, TLKM)

 Error rates - 0:67 - X-Aggregator

File Hilite Navigation View

Table "default" - Rows: 10 Spec - Columns: 3 Properties Flow Variables

Row ID	D Total squared error	D Mean squared error	I Size of Test Set
fold 0	0.092	0.001	127
fold 1	0.095	0.001	127
fold 2	0.091	0.001	127
fold 3	0.064	0.001	126
fold 4	0.096	0.001	126
fold 5	0.098	0.001	126
fold 6	0.076	0.001	126
fold 7	0.077	0.001	126
fold 8	0.079	0.001	126
fold 9	0.103	0.001	126

 Error rates - 3:70 - X-Aggregator

File Hilite Navigation View

Table "default" - Rows: 10 Spec - Columns: 3 Properties Flow Variables

Row ID	D Total squared error	D Mean squared error	I Size of Test Set
fold 0	0.147	0.001	127
fold 1	0.127	0.001	127
fold 2	0.109	0.001	127
fold 3	0.138	0.001	126
fold 4	0.157	0.001	126
fold 5	0.115	0.001	126
fold 6	0.128	0.001	126
fold 7	0.116	0.001	126
fold 8	0.12	0.001	126
fold 9	0.175	0.001	126

## DAFTAR RIWAYAT HIDUP

### A. Biodata Mahasiswa

NIM : 14002313  
Nama Lengkap : Evita Fitri  
Tempat & Tanggal Lahir : Bekasi, 01 Februari 1997  
Alamat Lengkap : Ujung Harapan, Kav. Bumimas Utama  
Rt.006/004 Kel. Bahagia, Kec. Babelan,  
Kab. Bekasi

### B. Riwayat Pendidikan Formal

1. MI Attaqwa 02 Annur, lulus Tahun 2008
2. SMP Attaqwa Pusat, lulus tahun 2011
3. SMK Bina Karya Utama Jakarta Timur, lulus tahun 2014
4. Diploma III AMIK BSI Bekasi, lulus tahun 2017
5. Strata I STMIK Nusa Mandiri, lulus tahun 2019

### C. Riwayat Pengalaman Organisasi / Pekerjaan

1. SMP Attaqwa Pusat sebagai Anggota Departemen Kesenian (2009 s/d 2010)
2. SMK Bina Karya Utama sebagai Ketua OSIS (2012 s/d 013)
3. Senat Mahasiswa BSI Kaliabang sebagai Staff.PSDM (2014 s/d 2015) dan Sekretaris Umum (2015 s/d 2016)
4. Staff. Guru Akuntansi SMK Bina Karya Utama Jakarta (2017 s/d Sekarang)



Bekasi, 17 Januari 2021

**Evita Fitri**