Terakreditasi SINTA Peringkat 2
Surat Keputusan Dirjen Penguatan Riset dan Pengembangan Ristek Dikti No. 10/E/KPT/2019
masa berlaku mulai Vol. 1 No. 1 tahun 2017 s.d Vol. 5 No. 3 tahun 2021

Terbit online pada laman web jurnal: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 3 No. 3 (2019) 364 - 370 ISSN Media Elektronik: 2580-0760

Penggunaan *Feature Selection* di Algoritma *Support Vector Machine* untuk Sentimen Analisis Komisi Pemilihan Umum

Imam Santoso[1], Windu Gata[2] Atik Budi Paryanti[3]
[1,2]Pascasarjana, Ilmu Komputer, STMIK Nusa Mandiri,[3] STIKOM Cipta Karya Informatika
[1]imanofani@yahoo.com, [2]windu@nusamandiri.ac.id,[3]atikbudiparyanti@gmail.com

Abstract

At this time sentiment analysis is very widely used by people to see the extent of people's sentiments towards an object. Objects that can be used in sentiment analysis can be various kinds, for example about the product regarding receipt by consumers, agencies or institutions regarding the performance of the agency. Whereas for this study taking sentiment analysis of the State Institution namely the General Election Commission (KPU) about the sentiments of the implementation of the ELECTION simultaneously and also the results of the implementation of the ELECTION which have become the subject of discussion by netizens on social media. So this research takes retweet data and retention comments from Twitter social media users. The algorithm used in this study is Support Vector Machine (SVM), with optimization of the use of Weight by Correlation Feature Selection (FS). The results of cross validation SVM without FS are 66.49% for accuracy and 0.716 for AUC. Whereas SVM with FS is 81.18% for accuracy and 0.943 for AUC. Very significant improvement with the use of Weight by Correlation Feature Selection (FS).

Keywords: KPU, Support Vector Machine, Feature Selection, Weight by Correlation

Abstrak

Saat ini analisis sentimen sangat banyak digunakan orang untuk melihat sejauh mana sentimen masyarakat terhadap suatu objek. Objek yang bisa digunakan dalam analisis sentimen bisa berbagai macam, misal tentang produk mengenai diterimanya oleh konsumen, instansi atau lembaga mengenai kinerja instansi tersebut. Sedangkan untuk penelitian ini mengambil analisis sentimen terhadap Lembaga Negara yaitu Komisi Pemilihan Umum (KPU) tentang sentimen pelaksanaan PEMILU serentak dan juga hasil dari pelaksanaan PEMILU tersebut yang telah menjadi bahan pembicaraan *netizen* di media sosial. Maka penelitian ini mengambil data *twit* maupun komen *retwit* dari para pengguna media sosial *twitter*. Algoritma yang digunakan dalam penelitian ini adalah *Support Vector Machine (SVM)*, dengan optimasi penggunaan *Feature Selection (FS) Weight by Correlation.* Hasil dari *cross validation* SVM tanpa FS adalah 66.49% untuk *accuracy* dan 0.716 untuk AUC. Sedangkan untuk SVM dengan FS adalah 81.18% untuk *accuracy* dan 0.943 untuk AUC. Peningkatan sangat signifikan dengan penggunaan *Feature Selection (FS) Weight by Correlation.*

Kata kunci: KPU, *Support Vector Machine, Feature Selection, Weight by Correlation*

© 2019 Jurnal RESTI

1. Pendahuluan

Bangsa Indonesia belum lama ini telah melakukan hajatan besar demokrasi yaitu PEMILU (Pemilihan Umum). Yang berbeda dengan Pemilu sebelumnya adalah pelaksanaan untuk Pemilihan Calon Legislatif maupun Presiden dilakukan secara bersamaan. Hal ini merupakan amanat Undang-Undang (UU) Nomor 42/2008 tentang Pilpres.

Namun dalam pelaksanaannya banyak sekali hal-hal yang dinilai oleh para pemilih bahwa Pemilu ini berjalan tidak sesuai dengan aturan yang ada. Mulai dari aroma kecurangan banyak disuarakan *netizen*, ketidaknetralan para penegak hukum terhadap salah satu calon pasangan presiden dan wakil presiden, serta pelaksanaan yang tidak JURDIL atau Jujur dan Adil. Semuanya banyak disuarakan lewat media sosial salah satunya adalah *twitter*. Ada juga dengan bentuk demonstrasi (unjuk rasa) sebagai bentuk kekecewaan

Diterima Redaksi : 27-07-2019 | Selesai Revisi : 05-11-2019 | Diterbitkan Online : 02-12-2019

dari para sebagian pemilih. KPU (Komisi Pemilihan Umum) selaku penyelenggara hajatan besar ini tak lepas dari tuduhan-tuduhan yang dilontarkan oleh sebagian pemilih yang merasa tidak puas dengan pelaksanaan Pemilu kali ini. Sentimen yang diberikan kepada KPU lewat media sosial merupakan gambaran kekecewaan tersebut, lewat *twit* yang di*share* oleh pengguna aplikasi *micro blogging* ini bisa kita ambil dan dijadikan analisis sentimen terhadap KPU selaku penyelenggara Pemilu.

Manfaat yang didapat dengan melakukan *sentiment analyst* adalah menemukan informasi berharga yang dibutuhkan orang lain dari data yang tidak terstruktur [1] Maka diharapkan dengan penelitian ini dapat diketahui sentimen para *netizen* atau pengguna twitter terhadap Komisi Pemilihan Umum. Selain itu diharapkan juga dengan penelitian ini akan mengetahui pengaruh *feature selection* terhadap peningkatan accuracy proses sentimen analisis.

Penelitian ini menggunakan dataset yang di *crawling* dari web twitter.com dengan *keyword* KPU. Data diambil untuk penelitian ini sebanyak 319 data, dengan sentiment positive sebanyak 142 dan sentiment negative sebanyak 177. Metode yang digunakan dalam penelitian ini adalah Support Vector Machine (SVM) dengan tambahan *feature selection* yakni *weight by correlation.* Penelitian akan membandingkan apakah metode SVM saja dengan SVM yang menggunakan *feature selection* dapat meningkatkan nilai *accuracy* dan juga membandingkan nilai AUC (*Area Under Curve*) untuk kedua metode tersebut.

Support Vector Machines (SVM) telah menjadi metode klasifikasi dan regresi yang sering digunakan untuk masalah linear dan nonlinear. Kelebihan dari algoritma *Support Vector Machines* adalah dari kemampuannya untuk menerapkan pemisahan linear pada input data non linear berdimensi tinggi, dan ini diperoleh dengan menggunakan fungsi kernel yang diperlukan. Efektivitas *Support Vector Machines* sangat dipengaruhi oleh jenis fungsi kernel yang dipilih dan diterapkan berdasarkan karakteristik data[2]. Banyak peneliti telah melaporkan bahwa *Support Vector Machines* metode yang paling akurat untuk teks klasifikasi[3].

Kemudian untuk *preprocessing* text hasil *crawling* yang digunakan dalam penelitian ini adalah menggunakan gataframework yang merupakan pengolah text berbasis *web* dengan bahasa pemograman php. Layanan ini tersedia di www.gataframework.com/textmining, banyak teknik yang tersedia pada halaman ini, mulai dari removal url, annotation removal dan lain sebagainya. Keunggulan dari *text preprocessing* ini adalah dalam *stemming* bahasa Indonesia, yang mana dalam tool yang digunakan dalam penelitian ini masih kesulitan untuk melakukan *stemming* bahasa Indonesia.

Penelitian sebelumnya yang membahas mengenai optimasi pada algoritma klasifikasi adalah dengan menggunakan algoritma Naïve Bayes dan optimasi *feature selection Information Gain* yang hanya menaikan sebesar *accuracy* 0.85% untuk sentimen analis pada *movie review* [4]. Diharapkan penelitian ini akan menghasilkan peningkatan yang signifikan dibandingkan penelitian sebelumnya.

2. Metode Penelitian

Machine learning saat ini sedang sangat ramai dibicarakan banyak orang, karena kemapuannya untuk bisa menjalankan kemampuan manusia oleh sebuah mesin. *Machine Learning* yang merupakan cabang dari Ilmu Komputer berhubungan dengan *Artificial Intelligent* atau ilmu kecerdasan buatan berfokus pada pembuatan/pengembangan dan studi suatu sistem dengan tujuan agar mampu belajar dari data-data yang diperolehnya. Menurut Arthur Samuel, *Machine learning* adalah bidang studi yang memberikan kemampuan pada program komputer untuk belajar tanpa secara eksplisit diprogram [5].

Dan salah satu dari kemampuan yang dikembangkan oleh para praktisi *Machine Learning* adalah *Sentiment Analyst*, seperti yang telah dijelaskan sebelumnya bahwa tujuan dari *sentiment analyst* adalah menyediakan informasi berharga bagi seseorang yang terkandung dari sebuah dataset yang tidak terstruktur. Kemampuan ini dipunyai oleh sebuah mesin setelah mempelajari data yang diberikan kepada mesin dan mencari pola dari data tersebut. Maka dataset yang diberikan kepada mesin terdiri dari 2 jenis [6]:

a. Data Training.

   Yakni sekumpulan data yang digunakan untuk melatih mesin mengetahui dan menerapkan algoritma yang digunakan. Gunanya adalah untuk menemukan model yang baik dari data tersebut agar bisa diterapkan pada data testing.

b. Data Testing.

   Sedangkan data testing merupakan sekumpulan data yang digunakan referensi dan mengevaluasi dari model yang telah ditemukan oleh mesin. Apakah model yang ada mempunyai nilai akurasi yang baik atau tidak.

2.1 Metode CRIS-DM

Metode *text mining* yang sering digunakan yaitu metode CRISP-DM (*Cross-Industry Standard Process for Data Mining*) yang terdiri 6 tahapan, yaitu [7]:

1. *Business Understanding.*

Pada tahapan ini yang dilakukan adalah memahami bagaimana proses yang akan dilalui dari penelitian agar menghasilkan hasil yang sesuai dengan diharapkan. Pengumpulan data adalah langkah awal agar penelitian

ini dapat dilakukan, kemudian waktu kapan penelitian ini dilaksanakan, serta yang terakhir adalah memastikan hasil yang akan diambil dari penelitian ini sesuai dengan yang diharapkan.

2. *Data Understanding*.

Tahapan selanjutnya adalah pemahaman data, artinya bagaimana kita mengetahui darimana data tersebut diambil dan bagaimana caranya. Penelitian ini mengambil data *twitt* dan komentar *retwitt* pengguna aplikasi sosial media Twitter.

3. *Data Preparation*.

Pada tahapan ini data yang telah diambil melalui tahap data understanding dilakukan proses preparation yang terdiri dari beberapa langkah, diantaranya adalah :

a. *Tokenize*.
Yakni proses memisahkan kata-kata pada tiap kalimat menjadi kata tersendiri. Sedangkan untuk tanda baca akan dihilangkan.
b. *Filter Tokens (By Length)*
Menghilangkan kata dengan panjang huruf tertentu. Misal untuk menghilangkan kata dengan panjang karakter 2 (dua huruf).
c. *Stopword Removal*
Yakni proses menghilangkan kata yang tidak mempunyai arti tidak tepat, yang biasanya merupakan kata hubung. Misal : dengan, karena, meskipun dsb.
d. *Transform Case*.
Tahapan ini hanya mengubah huruf pada text menjadi huruf kecil semua, guna menyeragamkan bentuk huruf.
e. *Stemming*.
Pada kebanyakan proses preparation, tahapa ini dilakukan pada akhir yang gunanya adalah mengubah kata yang didapat hasil *crawling* menjadi kata dasar. Akhir dari proses ini tidak ada lagi kata yang berimbuhan baik awalan maupun akhiran.

4. *Modelling*.

Pada tahap ini mulai ditentukan algoritma yang digunakan dan melakukan analisis data berdasarkan algoritma yang telah ditentukan.

5. *Evaluation*.

Tahap evaluasi adalah sebuah tahapan untuk melakukan evaluasi apakah model yang telah dibangun berdasarkan algoritma tertentu bisa menghasilkan sesuai dengan yang diharapakan. Evaluasi ini salah satunya bisa menggunakan teknik *10 fold- Cross Validation,* yakni melakukan dimana proses ini membagi data secara acak ke dalam 10 bagian dan kemudian Proses pengujian dimulai dengan pembentukan model dengan data pada bagian pertama. Model yang terbentuk akan diujikan pada 9 bagian data sisanya.

6. *Deployment*.

Tahap ini merupakan tahap akhir dari metode CRISP-DM, membuat atau membangun aplikasi berdasarkan model dan data yang dihasilkan pada tahapan sebelumnya.

2.2. *Feature Selection (Weight by Correlation)*.

Masalah yang sering timbul dalam klasifikasi khusus, dan *machine learning* pada umumnya, adalah menemukan cara untuk mengurangi dimensi *n* dari ruang fitur F untuk mengatasi risiko "Overfitting", overvitting sendiri adalah nilai yang disebabkan karena sedikitnya data training dibanding dengan jumlah *n* sebagai attribut[8].

Maka dibutuhkan sebuah teknik untuk menghindari risiko overfitting , digunakanlah *feature selection. Weight by Corellation* adalah salah satu metode yang digunakan untuk menaikan nilai akurasi dari model *machine learning* yang dikembangkan. Sebenarnya cara pilihan dalam menggunakan *feature selection* dan salah satunya adalah *Weight by Corellation* yang artinya pembobotan pada attribut dengan cara menghubungkan (korelasi) satu attribute dengan attribut lainnya.[9]

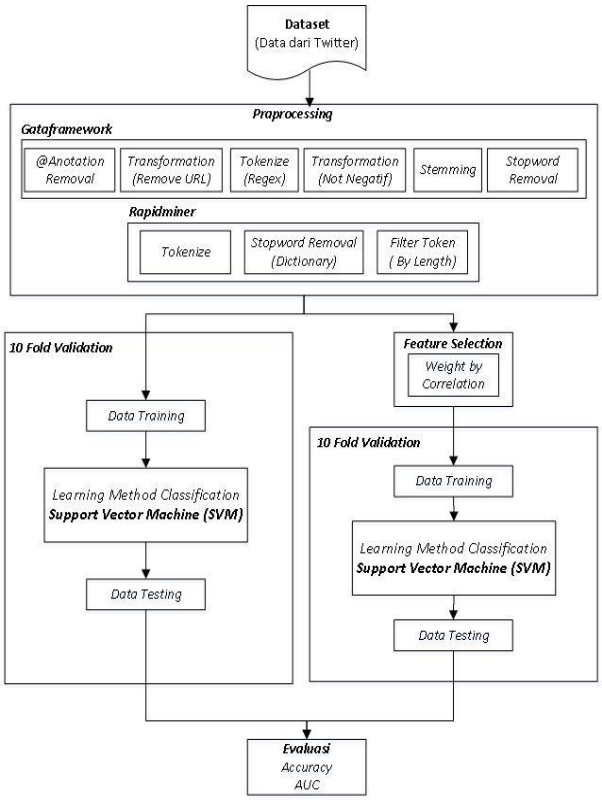
Feature selection yang merupakan salah satu optimasi dalam analis sentimen diletakan sebelum digunakannya algoritma sebagai model. Optimasi ada juga yang diletak pada akhir model analisis sentimen, salah satunya adalah *Particle Swarm Optimization (PSO)* namun PSO memiliki ketergantungan yang sensitif pada parameter yang digunakan [10].

2. 3. Kerangka Berpikir.

Dalam sebuah penelitian dibutuhkan kerangka berpikir, agar dalam pelaksanaan penelitian berjalan sesuai dengan apa yang telah direncanakan. Selain itu kerangka berpikir merupakan alur logika dalam suatu penelitian. Baik tidaknya suatu penelitian bisa dilihat dalam kerangka berpikir.

Dikerangka berpikir juga dapat dilihat apakah penelitian yang dilakukan merupakan suatu hal yang masuk logika, bila dilihat dari kerangka berpikir sudah merupakan hal yang tidak masuk dalam nalar, maka dapat dipastikan penelitian tersebut merupakan hal yang tidak mungkin berhasil. Adapun kerangka berpikir dalam penelitian ini dapat dilihat pada gambar 1.

Pada kerangka berpikir terlihat bahwa optimasi yang digunakan pada analisis sentimen penelitian ini diletakkan sebelum algoritma SVM. Adapun untuk jelasnya langkah-langkah pada kerangka berpikir diuraikan dibawah ini :



Gambar 1. Kerangka Berpikir dalam penelitian

1. Pengambilan Data.
Pengambilan data untuk penelitian ini termasuk dalam proses *Data Understanding*. Pengambilan dilakukan pada website *microbloging twitter*.

2. *Data Preparation.*
Sebelum dataset dimasukkan dalam model maka dilakukan proses *data preparation* yang meliputi yang pertama adalah *Anotation Removal*, yakni menghapus tanda *mention* (@) dan teks yang mengikutinya. Yang biasanya pada twitter digunakan untuk memberikan perhatian/ memanggil user yang lainnya. Setelah itu proses R*emove URL* atau menghilangkan *Uniform Resouce Locator* (URL) yang ada dalam dataset yang dihasikan dari proses crawling data. Kemudian selanjutnya adalah *Tokenize (Regex)*, proses ini adalah membelah kata-kata yang ada dalam dataset menjadi terpisah, bukan merupakan kalimat lagi. Setelah itu kata-kata tersebut akan disebut dengan menjadi istilah fitur.

Setelah dilanjutkan dengan tahapan *Stemming, p*ada tahapan ini berfungsi untuuk mengubah kata perkata menjadi kata dasar tanpa diberi imbuhan baik di awalan maupun akhiran. Dilanjutkan dengan *Transformation Not Negatif*, digunakan untuk mengubah kata-kata bermakna negatif yang akan disatukan dengan tanda *underscore* (_). Setelah itu dilakukan proses *Indonesia Stopword* yaitu membuang kata-kata yang diabaikan pada sentimen analis. Biasanya berupa kata keterangan dan kata sambung. Untuk kemudian masuk ke proses *Filter Token by Length*, yakni menghilangkan kata-kata dengan panjang karakter tertentu. Karena biasanya kata yang hanya memiliki hanya 2 karakter tidak mempunyai arti. Terakhir dalam tahapan *data preparation* adalah F*eature Selecion (Weight by Correlation)* yaitu proses menghitung bobot untuk masing-masing fitur (*text*) yang ada dalam dataset. Kemudian dilakukan pemilihan fitur dengan kriteria bobot terurut dari besar ke kecil.

3. Modelling.
Pada tahapan ini dilakukan 10 fold cross validation yakni proses membagi dataset menjadi 10 bagian yang mana 1 diantara bagian lainnya menjadi data testing, dan yang lainnya menjadi data training. Kemudian dimasukkan kedalam model algoritma *Support Vector Machine (SVM).* Ini dilakukan bergantian pada tiap bagian data sampai mendapat nilai terbaik dari model ini.

4. Evaluasi.
Setelah tahapan modelling selesai maka selanjutnya adalah melakukan evaluasi terhadap hasil dari pemodelan tersebut. Membandingkan dua hasil dari pemodelan yang *berupa accuracy, precision, recall* maupun AUC antara model dengan *Feature Selecion (Weight by Correlation)* maupun tidak.

3. Hasil dan Pembahasan

Pada bab ini akan dijelaskan mengenai proses yang dijalankan dalam penelitian ini. Ada beberapa langkah yang diterapkan pada penelitian ini.

3.1. Pengambilan Data

Pengambilan data pada penelitian ini menggunakan add-on dari google chrome yakni Data Miner. Penggunaan Data Miner sangat memudahkan bagi peneliti yang ingin mendapatkan data dari sebuah halaman website. Layanan yang diberikan oleh Data Miner juga memungkinkan pengambilan data di export kedalam bentuk file yang diinginkan.

3.2. *Data Preparation*

Setelah pengambilan data dari website, data tersebut tidak bisa langsung dimasukan dalam pengolahan untuk sentimen analis, maka dilanjutkan dengan tahapan *Data Preparation*. Dalam langkah ini terbagi dalam beberapa langkah :

1. *Anotation Removal*.
   Langkah ini adalah menghilangkan tanda anotation yang ada dalam text. Perbedaan dari sebelum dan sesudah proses anotation removal ada Tabel 1.

2. *Removal URL*.
   Langkah selanjutnya dalam data preparation adalah *removal URL* langkah ini adalah menghilangkan URL yang ada pada text. Berikut akan diberikan

contoh perbedaan antara sebelum dan sesudah proses ini. Lihat Tabel 2.

Tabel 1.Hasil dari Anotation Removal

Text	Anotation Removal
Alangkah eloknya jika @KPU_ID , @bawaslu_RI dan Polri dalam bulan yang penuh ampunan ini menyampaikan permohonan maaf yang tulus kepada bangsa Indonesia.	alangkah eloknya jika , dan polri dalam bulan yang penuh ampunan ini menyampaikan permohonan maaf yang tulus kepada bangsa indonesia.

Tabel 2.Hasil dari Removal URL

Text	Removal URL
Bawaslu Apresiasi Kinerja Bawaslu Provinsi dan KPU https://bawaslu.go.id/id/berita/bawaslu-apresiasi-kinerja-bawaslu-provinsi-dan-kpu.	bawaslu apresiasi kinerja bawaslu Provinsi dan KPU.

3. *Tokenize*.
Tahapan ini adalah memisahkan kata-kata dari tiap kalimat yang akan diproses. Sehingga kata-kata akan dilanjutkan ke tahapan berikutnya.

4. *Transformation Not (Negative)*.
Langkah ini akan menggabungkan dua kata yang mengandung kata negatif. Misal kata tidak ada, apabila setelah di-*tokenize* kata tersebut akan terpisah dan tidak memiliki arti dikarena berdiri masing-masing. Maka untuk menghindari itu lakukan proses *transformation Not (Negatif)*. Tabel 3 memperlihatkan hasil dari tahapan ini.

Tabel 3.Hasil dari *Transformation Not (Negative)*

Text	Transformation Not (Negative)
Azyumardi: Demo KPU bukan Jihad	azyumardi demo kpu bukan_jihad

5. *Stemming*.
Stemming adalah proses mengubah seluruh kata yang ada dalam dataset menjadi kata dasar yang selanjutnya akan diproses dengan menggunakan algoritma klasifikasi. Berikut adalah tabel 4 contoh hasil penggunaan stemming pada text.

Tabel 4.Hasil dari *Stemming*

Text	Stemming
Iya, kita mengakui kemenangan Pak Jokowi...Soal pilpres, kami akui hasil KPU..., kata Ketua Umum PAN Zulkifli Hasan.	kaku menang jokowi soal pilpres akui hasil kpu ketua pan zulkifli hasan

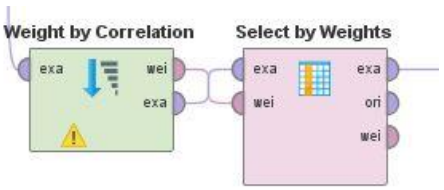
6. *Stop Word Removal*.
Tahapan ini merupakan tahapan akhir pada proses Data Preparation. Yaitu menghapus kata-kata yang tidak mempunyai arti yang biasanya merupakan kata sambung, kata keterangan dan sebagainya.

Tabel 5.Hasil dari *Stop Word Removal*

Text	Stemming
Beberapa massa meneriak-kan kata BESOK bawaslu KPU	massa teriak besok bawaslu kpu

3.3. *Feature Selection (Weight by Correlation)*

Untuk langkah selanjutnya digunakan operator *Weight by Correlation* yang berfungsi untuk menghitung bobot berdasarkan korelasi antara satu fitur dengan fitur yang ada dalam dataset dengan mempertimbangkan nilai label. Setelah fitur dilakukan penghitungan bobot maka dilakukan *sort ascending* (diurutkan dari nilai terkecil ke nilai terbesar) fitur berdasarkan nilai bobot itu sendiri. Kemudian setelah proses *sort* maka dilakukan pengambilan fitur yang mempunyai urutan 1 sampai dengan 300 agar tidak semua fitur masuk dalam model. Hal ini dilakukan dengan menggunakan operator *select by weight* dengan parameter *weight relation = top k.* Dengan nilai k=300. Penggunaan kedua operator ini dapat dilihat pada gambar 2.



Gambar 2. Operator Weight by Correlation

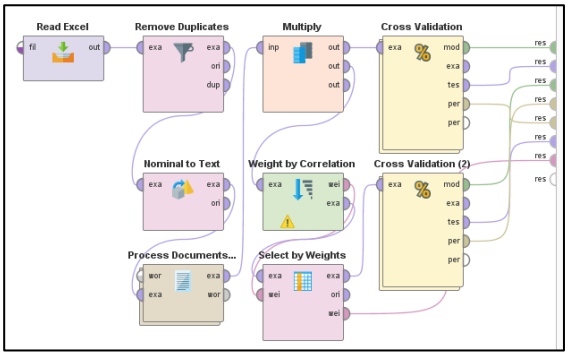
Tabel 6 adalah menampilkan contoh atrribute/ fitur yang telah di bobotkan dan kemudian di urutan mulai dari bobot terbesar ke terkecil (Descending) pada software *rapidminer*. Attribute/ Fitur yang tampil pada tabel pada software *rapidminer* adalah 300 attribute, namun tabel berikut hanya menampilkan contoh attribute / fitur tersebut.

Tabel 6. Contoh hasil dari pembobotan dengan operator *Weight by Correlation*

Attribute / Fitur	Weight
rekapitulasi	0.27
selamat	0.163
situng	0.162
baca	0.149
tanggap	0.146
bonjol	0.139
nasional	0.138

3.4 *Modelling*.

Adapun pemodelan untuk penelitian ini digambarkan secara lengkap terlihat pada gambar 3. Yang merupakan keseluruhan operator yang digunakan dalam penelitian pada *software rapidminer*.



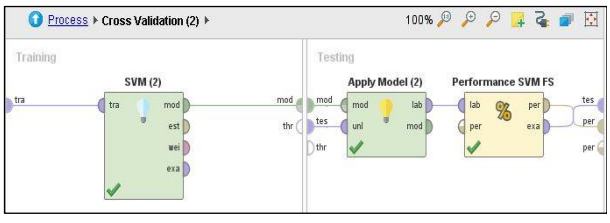
Gambar 3. Pemodelan Penelitian di *Rapidminer*

Pada pemodelan diatas diketahui bahwa penggunaan algoritma SVM yang ada dalam Cross Validation terdapat 2 buah, yang mana perbedaannya adalah salah satunya tanpa dilakukan pembobotan dengan korelasi (hasil dari *data preparation* langsung ke *Cross Validation*) dan yang satunya menggunakan pembobotan korelasi.

Untuk membedakan kedua proses diatas digunakan operator *Multiply* yang berfungsi untuk membagi *exampleset* keluaran dari *data preparation* masuk ke *cross validation* untuk dikenakan algoritma SVM dan dilakukan pengukuran evaluasi.

3.5. Evaluasi.

Dan untuk melakukan evaluasinya penelitian ini menggunakan *Accuracy* dan AUC *(Area Under Curve).* Hasil dari SVM tanpa dan dengan *feature selection* akan dibandingkan untuk mengukur seberapa peningkatan yang dihasilkan. Untuk mendapatkan nilai yang akan dievaluasi tersebut maka digunakan *10 fold cross validation* terhadap data yang telah dihitung pembobotannya (Gambar 4).



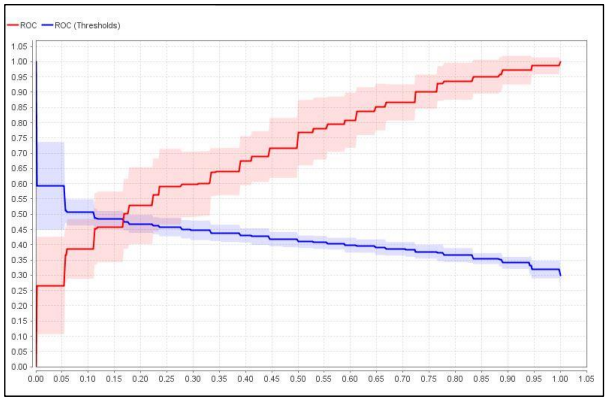
Gambar 4. *Cross Validation* untuk mendapatkan nilai *Accurcy* dan AUC.

Dari tahapan-tahapan yang dilakukan maka menghasilkan Accuracy dan AUC dari algoritma SVM tanpa feature selection sebagai berikut :

Tabel 7. Hasil *Cross Validation* dari SVM tanpa *feature selection*

	SVM tanpa *feature selection*
Accuracy	66,49 %
AUC	0.716

Diketahui bahwa dari cross validation SVM tanpa feature selection hanya mendapatkan 66.49% sedangkan untuk nilai AUC nya sebesar 0.716. Sedangkan bentuk kurva dari ROC yang dihasilkan bisa dilihat pada Gambar 5 dibawah.



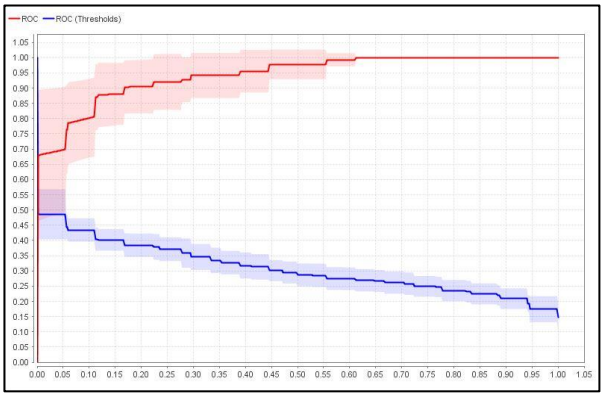
Gambar 5. Bentuk AUC untuk SVM tanpa *feature selection*

Sedangkan untuk nilai accuracy dan AUC untuk cross validation algoritma SVM dengan *feature selection* terlihat di tabel 8.

Tabel 8. Hasil *Cross Validation* dari SVM dengan *feature selection*

	SVM tanpa *feature selection*
Accuracy	81.18 %
AUC	0.943

Dari tabel bisa terlihat bahwa hasil cross validation dari algoritma SVM dengan *feature selection* naik menjadi 81,18 % dan untuk nilai AUC menjadi 0.943. Detail dari gambar kurva ROC dapat dilihat pada gambar 6.



Gambar 6. Bentuk AUC untuk SVM dengan *feature selection*

Dari kedua hasil *cross validation* diatas maka dapat kita cari peningkatan selisih peningkatan dengan

penggunaan *feature selection* untuk nilai accuracy dan AUC. Dengan melihat tabel 8 kita bisa lihat peningkatan nilai accuracy adlah sebesar 14.69% dan untuk AUC sebesar 0.227. Perbedaan sangat terlilhat signifikan untuk kenaikkan presentasenya.

Tabel 9. Perbandingan *Cross Validation* dari SVM tanpa *feature selection* dan dengan *feature selection (FS)*

	SVM tanpa *FS*	SVM + *FS*	Selisih
Accuracy	66.49 %	81.18 %	**14.69%**
AUC	0.716	0.943	**0.227**

4. Kesimpulan

Setelah melakukan penelitian dengan model algoritma SVM tanpa *feature selection* dan dibandingkan dengan model SVM dengan *feature selection* dapat kita simpulkan bahwa penggunaan *feature selection Weight by Correlation* dapat meningkatkan nilai dari Accuracy dan AUC. Peningkatan yang didapatkan sangat signifikan yang sebelumnya model SVM tanpa *feature selection* hanya menghasilkan 66.49% dan nilai AUC 0.716 setelah ditambahkan penggunaan *feature selection* menjadi 81.18% untuk accuracy dan nilai AUC 0.943.

Jelas hal ini memberikan peningkatan yang sangat tajam dengan nilai selisih untuk accuracy sebesar 14.69 dan 0.227 untuk AUC. Sehingga dapat disimpulkan penggunaan *feature selection Weight by Corellation* sangat baik apabila digunakan pada algoritma *Support Vector Machine (SVM).*

Untuk penelitian selanjutnya dapat disarankan menggunakan *feature selection Weight* lainnya selain *feature selection Weight by Correlation.* Karena masih ada beberapa operator untuk *feature selection* contoh seperti *feature selection Weight by SVM* yang menghitung pembobotan dengan metode SVM. Kemudian pengambilan data juga dapat dilakukan dengan lebih banyak jumlahnya untuk mendapatkan nilai akurasi lebih baik.

Daftar Rujukan

[1] Chaovalit, Pimwadee and Lina Zhou, 2005. Movie Review Mining: a Comparison between Supervised and Unsupervised Classification Approaches, *IEEE*, pp. 1-9, .

[2] Haddi, E., Liu, X., & Shi, Y., 2013. The Role of Text Pre-processing in Sentiment Analysis. *First International Conference on Information Technology and Quantitative Management*, 17, 26–32. https://doi.org/10.1016/j.procs.2013.05.05

[3] Moraes, R., Valiati, J. F., & Gavião Neto, W. P., 2013. Document-level sentiment classification: An empirical comparison between SVM and ANN. *Expert Systems with Applications*, 40(2), 621–633. https://doi.org/10.1016/j.eswa.2012.07.05.

[4] Andilala, 2016, Movie Review Sentimen Analis dengan Metode Naïve Bayes Base On Feature Selection, *Jurnal Pseudocode*, Volume III Nomor 1, Februari 2016, ISSN 2355 – 5920.

[5] Fikriya, Zulfa Afiq; Irawan, Mohammad Isa; Soetrisno, 2017. *"*Implementasi Extreme Learning Machine untuk Pengenalan Objek Citra Digital*"*, *Jurnal Sains dan Seni ITS*, Vol. 6, No. 1. 2337-3520,

[6] Rahmansyah A., Dewi O., Andini P., Hastuti PN, Triana and Eka Suryana, Muhammad. 2016, Membandingkan Pengaruh Feature Selection Terhadap Algoritma Naïve Bayes dan Support Vector Machine. *Seminar Nasional Aplikasi Teknologi Informasi (SNATi)* , 2018 p. A1 - A7.

[7] Ana Azevedo, Manuel Filipe Santos, 2008., KDD, SEMMA and CRISP-DM: A Parallel Overview, *IADIS European Conference Data Mining*, 2008, pp 182 - 185

[8] Guyon, I., Weston, J., and Barnhill, S. (2002), *Machine Learning*, Gene Selection for Cancer Classification using Support Vector Machines, Netherland , Kluwer Academic Publishers.

[9] Rapidminer Documentation, 2019, *Weight by Correllation*, [Online] tersedia di : https://docs.rapidminer.com/latest/studio/operators/modeling/feature_weights/weight_by_correlation.html [Accessed : 24 Juli 2019]

[10] Achyani, Yuni Eka. 2018, Penerapan Metode Particle Swarm Optimization Pada Optimasi Prediksi Pemasaran Langsung, *Jurnal Informatika*, Vol.5 No.1 April 2018, pp. 1~11