

**KLASIFIKASI ALGORITMA *DECISION TREE*, ALGORITMA *NAIVE BAYES* DAN *K-NEAREST NEIGHBORS* TERHADAP  
POPULARITAS PERFILMAN INDONESIA**



TESIS

DANING NUR SULISTYOWATI

14002092

PROGRAM STUDI MAGISTER ILMU KOMPUTER  
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER  
NUSA MANDIRI  
JAKARTA  
2018

**KLASIFIKASI ALGORITMA *DECISION TREE*, ALGORITMA *NAIVE BAYES* DAN *K-NEAREST NEIGHBORS* TERHADAP  
POPULARITAS PERFILMAN INDONESIA**



**TESIS**

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

**DANING NUR SULISTYOWATI**

14002092

**PROGRAM STUDI MAGISTER ILMU KOMPUTER  
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER**

**NUSA MANDIRI**

**JAKARTA**

**2018**

## SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini :

Nama : Daning Nur Sulistyowati  
NIM : 14002092  
Program Studi : Magister Ilmu Komputer  
Jenjang : Strata Dua (S2)  
Konsentrasi : Data Mining

Dengan ini menyatakan bahwa tesis yang telah saya buat dengan judul “Klasifikasi Algoritma *Decision Tree*, *Algoritma Naive Bayes* dan *K-Nearest Neighbors* Terhadap Popularitas Perfilman Indonesia” adalah hasil karya sendiri dan semua sumber baik yang kutip maupun yang rujuk telah saya nyatakan dengan benar dan tesis belum pernah diterbitkan atau dipublikasikan dimanapun dan dalam bentuk apapun.

Demikian surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila dikemudian hari ternyata saya memberikan keterangan palsu dan atau 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 Pascasarjana Magister Ilmu Komputer Sekolah Manajemen Informatika dan Komputer Nusa Mandiri dicabut/dibatalkan.

Jakarta, 07 Januari 2019

Yang menyatakan,



Daning Nur Sulistyowati

## PERSETUJUAN DAN PENGESAHAN TESIS

Tesis ini diajukan oleh:

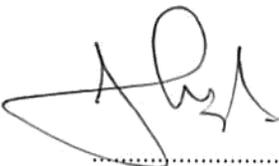
Nama : Daning Nur Sulistyowati  
NIM : 14002092  
Program Studi : Ilmu Komputer  
Jenjang : Program STRATA-2  
Judul Tesis : Klasifikasi Algoritma Decision Tree, Naive Bayes Dan K-nearest Neighbor Terhadap Popularitas Perfilman Indonesia

Telah dipertahankan pada periode 2018-2 dihadapan penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh MAGISTER ILMU KOMPUTER (M.Kom) pada Program STRATA-2 Program Studi Ilmu Komputer di Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri Pascasarjana.

Jakarta, 23 Januari 2019

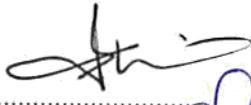
### PEMBIMBING TESIS

Dosen Pembimbing : Dr. Windu Gata, M.Kom

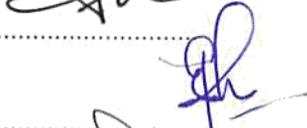


### DEWAN PENGUJI

Penguji I : Dr. Sfenrianto, M.Kom



Penguji II : Dr. Sularso Budilaksono, M.Kom



Penguji III / Dosen Pembimbing : Dr. Windu Gata, M.Kom



## SURAT PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan di bawah ini, saya :

Nama : Daning Nur Sulistyowati  
NIM : 14002092  
Program Studi : Magister Ilmu Komputer  
Jenjang : Strata Dua (S2)  
Konsentrasi : Data Mining  
Jenis Karya : Tesis

Demi pengembangan ilmu pengetahuan, dengan ini menyetujui untuk memberikan ijin kepada pihak Program Studi Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIK Nusa Mandiri) **Hak Bebas Royalti Non-Eksklusif (*Non-exclusive Royalti-Free Right*)** atas karya ilmiah kami yang berjudul : “Klasifikasi Algoritma *Decision Tree*, Algoritma *Naïve Bayes* dan *K-Nearest Neighbors* Terhadap Popularitas Perfilman 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, mengelolanya dalam pangkalan data (*database*), mendistribusikannya dan menampilkan atau mempublikasikannya di *internet* atau media lain untuk kepentingan akademis tanpa perlu meminta ijin 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, 07 Januari 2019

Yang menyatakan,



Daning Nur Sulistyowati

## KATA PENGANTAR

Syukur Alhamdulillah, penulis panjatkan kehadiran Allah SWT, yang telah melimpahkan rahmat dan karunia-Nya, sehingga pada akhirnya penulis dapat menyelesaikan tesis ini tepat pada waktunya. Dimana tesis ini penulis sajikan dalam bentuk buku yang sederhana. Adapun judul tesis, yang penulis ambil sebagai berikut : “**Klasifikasi Algoritma *Decision Tree*, Algoritma *Naive Bayes* dan *K-Nearest Neighbor* Terhadap Popularitas Perfilman Indonesia**”.

Tujuan penulisan tesis ini dibuat sebagai salah satu syarat untuk mendapatkan gelar Magister Ilmu Komputer (M.Kom) pada Program Studi Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIK Nusa Mandiri).

Tesis ini diambil berdasarkan hasil penelitian mengenai Klasifikasi Algoritma *Decision Tree*, Algoritma *Naive Bayes* dan *K-Nearest Neighbor* Terhadap Popularitas Perfilman Indonesia. Penulis juga melakukan pencarian dan menganalisa berbagai macam sumber referensi, baik dalam bentuk jurnal ilmiah, buku-buku literatur, internet dan lain lain yang terkait dengan pembahasan pada tesis ini.

Penulis menyadari bahwa tanpa bimbingan dan dukungan dari semua pihak dalam pembuatan tesis ini, maka penulis tidak dapat menyelesaikan tesis ini tepat pada waktunya. Untuk itu pada kesempatan ini penulis ingin sampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Direktur Program Studi Magister Ilmu Komputer STMIK Nusa Mandiri Jakarta.
2. Bapak Dr. Windu Gata, M.Kom selaku pembimbing tesis yang telah menyediakan waktu, pikiran, tenaga dan kesabaran dalam membimbing sehingga penulis dapat menyelesaikan tesis ini.
3. Orang Tua yang sangat penulis cintai yang telah memberikan dukungan material dan moral kepada penulis.
4. Seluruh staff pengajar (dosen) dan karyawan Program Studi Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa

Mandiri Jakarta yang telah memberikan pelajaran yang berarti bagi penulis selama menempuh studi.

5. Seluruh rekan kerja yang selalu memberikan bantuan dan semangat kepada penulis.
6. Seluruh rekan kuliah yang saling memberikan semangat untuk dapat menyelesaikan tesis ini.

Serta semua pihak yang terlalu banyak untuk penulis sebutkan satu persatu sehingga terwujudnya penulisan tesis ini. Penulis menyadari bahwa penulisan tesis ini masih jauh sekali dari sempurna, untuk itu penulis mohon kritik dan saran yang bersifat membangun demi kesempurnaan penulisan karya ilmiah yang penulis hasilkan untuk yang akan datang.

Akhir kata semoga tesis ini dapat bermanfaat bagi penulis khususnya dan bagi para pembaca yang berminat pada umumnya.

Jakarta, 07 Januari 2019

Daning Nur Sulistyowati  
Penulis

## ABSTRAK

Nama : Daning Nur Sulistyowati  
NIM : 14002092  
Program Studi : Magister Ilmu Komputer  
Jenjang : Strata Dua (S2)  
Konsentrasi : Data Mining

Perkembangan dunia perfilman yang sudah sangat pesat berdampak kepada para penikmat film yang membutuhkan film-film yang mempunyai kualitas gambar, suara, alur cerita dan nilai positif yang terdandung didalamnya. Hal itu didasarkan agar para penikmat film tetap antusias dalam mengikuti film-film yang terbaru. Agar dunia perfilman berkembang secara terus menerus dibutuhkan sebuah penilaian yang berasal dari masyarakat dan khususnya para penikmat film itu sendiri. Penilaian ini digunakan untuk mengetahui jenis film apa yang diinginkan. Maka dari itu dibutuhkan sebuah analisa untuk mengetahui minat dari penikmat film dengan menganalisis *rating* suatu film. Penelitian ini bertujuan untuk memprediksi tingkat kepopuleritasan film yang ada di Indonesia selama 10 tahun terakhir. Hasil yang didapatkan menggunakan algoritma *decision tree* menunjukkan 454 judul film populer dan 102 judul film tidak populer dengan nilai *accuracy* sebesar 76,45% dengan *precision* sebesar 82,78% dan *recall* sebesar 88,68%, *naive bayes* menunjukkan 446 judul film populer dan 110 judul film tidak populer dengan nilai *accuracy* sebesar 83,29% dengan *precision* sebesar 87,83% dan *recall* sebesar 91,24% dan *K-Nearest Neighbors* menunjukkan 443 judul film populer dan 113 judul film tidak populer dengan nilai *accuracy* sebesar 80,95% dengan *precision* sebesar 86,92% dan *recall* sebesar 88,90%. Hasil evaluasi dengan *ROC curve* menunjukkan nilai AUC menggunakan model *naive bayes* sebesar 0,836 model *K-Nearest Neighbors* sebesar 0,818 dan model *decision tree* sebesar 0,767. Dari ketiga model yang digunakan dua diantaranya termasuk kedalam klasifikasi *Good classification* yaitu model algoritma *naive bayes* dan *k-nearest neighbors*, dikarenakan memiliki nilai AUC diantara 0,80 – 0,90. Sedangkan model algoritma *decision tree* termasuk kedalam klasifikasi *Fair classification*, memiliki nilai AUC diantara 0,70 - 0,80.

**Kata kunci : Klasifikasi, Algoritma Decision Tree, Algoritma Naïve Bayes, K-Nearest Neighbors, Popularitas Film**

## **ABSTRACT**

*Name* : Daning Nur Sulistyowati  
*NIM* : 14002092  
*Study Of Program* : *Magister Ilmu Komputer*  
*Levels* : *Strata Dua (S2)*  
*Concentration* : *Data Mining*  
*Title* : “*Classification of Decision Tree Algorithms, Naïve Bayes Algorithms and K-Nearest Neighbors Against the Popularity of Indonesian Films*”

*The development of the film industry that has been very rapid has an impact on film lovers who need films that have image quality, sound, storyline and positive values contained therein. This was based on so that film lovers remained enthusiastic in following the latest films. In order for the world of film to develop continuously it needs an assessment that comes from the community and especially the audience of the film itself. This assessment is used to find out what type of film you want. Therefore we need an analysis to find out the interest of film lovers by analyzing the rating of a film. This study aims to predict the level of popularity of films in Indonesia for the past 10 years. The results obtained using a decision tree algorithm are 76.45% with precision of 82.78% and recall of 88.68%, naive bayes of 83.29% with precision of 87.83% and recall of 91.24% and K -Nearest Neighbors is 80.95% with precision of 86.92% and recall of 88.90%. The evaluation results with the ROC curve showed that the AUC value using the naive bayes model was 0.836, the K-Nearest Neighbors model was 0.818 and the decision tree model was 0.767. Of the three models used two of them are included in the classification of Good classification, namely the algorithm Naive Bayes model and the k-nearest neighbors, because it has an AUC value between 0.80 - 0.90. While the decision tree algorithm model included in the Fair classification classification, has an AUC value between 0.70 - 0.80.*

*Keywords: Classification, Decision Tree Algorithm, Naïve Bayes Algorithm, K-Nearest Neighbors, Film Popularity*

# DAFTAR ISI

	Halaman
<b>HALAMAN SAMPUL</b> .....	<b>i</b>
<b>HALAMAN JUDUL</b> .....	<b>ii</b>
<b>HALAMAN PERNYATAAN ORISINALITAS</b> .....	<b>iii</b>
<b>HALAMAN PERSETUJUAN DAN PENGESAHAN</b> .....	<b>iv</b>
<b>HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS</b> .....	<b>v</b>
<b>KATA PENGANTAR</b> .....	<b>vi</b>
<b>ABSTRAK</b> .....	<b>viii</b>
<b>ABSTRACT</b> .....	<b>ix</b>
<b>DAFTAR ISI</b> .....	<b>x</b>
<b>DAFTAR TABEL</b> .....	<b>xii</b>
<b>DAFTAR GAMBAR</b> .....	<b>xiii</b>
<b>DAFTAR LAMPIRAN</b> .....	<b>xiv</b>
<b>BAB 1. PENDAHULUAN</b> .....	<b>1</b>
1.1. Latar Belakang Penulisan .....	1
1.2. Identifikasi Masalah .....	2
1.3. Tujuan Penelitian .....	2
1.4. Ruang Lingkup Penelitian .....	2
1.5. Rumusan Masalah .....	3
1.6. Sistematika Penulisan .....	3
<b>BAB 2. LANDASAN TEORI</b> .....	<b>5</b>
2.1. Tinjauan Pustaka .....	5
2.2. Tinjauan Studi .....	16
2.3. Novelty Penelitian .....	29
2.4. Kerangka Pemikiran .....	29
2.5. Objek Penelitian .....	30
<b>BAB 3. METODOLOGI PENELITIAN</b> .....	<b>31</b>
3.1. Metode Penelitian .....	31
3.2. Metode Pemilihan Populasi dan Sampel .....	33
3.3. Pengumpulan Data .....	33
3.4. Instrumen Penelitian .....	34
3.5. Metode Analisis dan Pengujian Data .....	34
<b>BAB 4. HASIL PENELITIAN DAN PEMBAHASAN</b> .....	<b>35</b>
4.1. Business Understanding .....	35
4.2. Data Understanding .....	35
4.3. Data Preparation .....	37
4.4. Modeling .....	38
4.5. Evaluasi dan Validasi Hasil .....	41
4.6. Deployment .....	47

<b>BAB 5 PENUTUP</b> .....	<b>48</b>
5.1. Kesimpulan.....	48
5.2. Saran.....	49
<b>DAFTAR REFERENSI</b> .....	<b>50</b>
<b>DAFTAR RIWAYAT HIDUP</b> .....	<b>53</b>
<b>HALAMAN KONSULTASI BIMBINGAN</b> .....	<b>54</b>
<b>LAMPIRAN-LAMPIRAN</b> .....	<b>55</b>

## DAFTAR TABEL

	Halaman
Tabel 2.1. Model Confusion Matrix .....	14
Tabel 2.2. Ringkasan Penelitian .....	24
Tabel 3.1. Sampel Dataset .....	33
Tabel 4.1. Tabel Atribut .....	33
Tabel 4.2. Dataset Film Indonesia .....	34
Tabel 4.3. Perhitungan Probabilitas Posterior .....	41
Tabel 4.4. Hasil Evaluasi Nilai AUC .....	42
Tabel 4.5 Hasil Perhitungan Algoritma Decision Tree .....	50
Tabel 4.6. Konversi Confusion Matrix Algoritma Naive Bayes .....	51
Tabel 4.7. Hasil Perhitungan Algoritma Naive Bayes .....	52
Tabel 4.8. Konversi Cofusion Matrix Algoritma K- Nearest Neighbors.....	52
Tabel 4.9. Hasil Perhitungan Algoritma K-Nearest Neighbors .....	53
Tabel 4.10. Komparasi Nilai Akurasi.....	54
Tabel 4.11. Komparasi Nilai AUC.....	54

## DAFTAR GAMBAR

	Halaman
Gambar 2.1. Proses Data Mining.....	6
Gambar 2.2. Klasifikasi Sebagai Pemetaan Input Himpunan Atribut $x$ ke Dalam Kelas Label $y$ .....	9
Gambar 2.3. Model Grafik ROC .....	15
Gambar 2.4. Kerangka Pemikiran .....	30
Gambar 3.1. Tahapan Penelitian.....	31
Gambar 4.1. Pengujian Data Mining Dengan <i>Rapid Miner</i> .....	36
Gambar 4.2. Tampilan <i>Decision Tree</i> yang Terbentuk.....	36
Gambar 4.3. Model Validation Algoritma <i>Decision Tree</i> .....	37
Gambar 4.4. Model Validation Algoritma <i>Naive Bayes</i> .....	37
Gambar 4.5. Model Validation Algoritma <i>K-Nearest Neighbors</i> .....	38
Gambar 4.6. Konversi <i>Confusion Matrix</i> Algoritma <i>Decision Tree</i> .....	39
Gambar 4.7. Kurva ROC <i>Decision Tree</i> .....	39
Gambar 4.8. Konversi <i>Confusion Matrix</i> Algoritma <i>Naive Bayes</i> .....	40
Gambar 4.9 Kurva ROC <i>Naive Bayes</i> .....	40
Gambar 4.10. Konversi <i>Confusion Matrix</i> Algoritma <i>K-Nearest Neighbors</i> . .	40
Gambar 4.11 Kurva ROC <i>K-Nearest Neighbors</i> .....	41
Gambar 4.12 Tampilan Input Model Simulator .....	42
Gambar 4.13 Hasil Uji Coba Model Simulator.....	43
Gambar 4.14 Hasil Akurasi Model Simulator .....	43

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang Penulisan

Dunia perfilman menjadi salah satu industri terbesar dimuka bumi ini. Industri film merupakan salah satu industri yang tidak ada habisnya. Film merupakan salah satu media massa yang digunakan sebagai sarana hiburan. Selain itu film berperan sebagai sarana modern yang digunakan untuk menyebarkan informasi kepada masyarakat. Film menjadi salah satu media massa yang cukup efektif dalam menyampaikan suatu informasi.

Perkembangan dunia perfilman yang sudah sangat pesat berdampak kepada para penikmat film yang membutuhkan film-film yang mempunyai kualitas gambar, suara, alur cerita dan nilai positif yang terdandung didalamnya. Hal itu didasarkan agar para penikmat film tetap antusias dalam mengikuti film-film yang terbaru. Walaupun tidak semua masyarakat menyukai film atau hanya menyukai film dengan beberapa jenis (*genre*) tertentu saja.

Agar dunia perfilman berkembang secara terus menerus dibutuhkan sebuah penilaian. Penilaian-penilaian itu sendiri berasal dari masyarakat dan khususnya para penikmat film itu sendiri. Penilaian ini digunakan untuk mengetahui jenis film apa yang diinginkan. Maka dari itu dibutuhkan sebuah analisa untuk mengetahui minat dari penikmat film dengan menganalisis *rating* suatu film.

Dari penelitian yang dilakukan oleh (Ahmed, Jahangir, Afzal, Majeed, & Siddiqi, 2015) dengan judul *Using crowd-source based features from social media and conventional features to predict the movies popularity* menjelaskan mengenai prediksi keberhasilan suatu film dengan menggunakan algoritma *Decision Tree* dengan tingkat akurasi sebesar 61%. Berdasarkan penelitian tersebut dilakukan penelitian yang serupa dengan berbeda kasus dan metode penelitian. Jika didalam penelitian sebelumnya kasus yang dibahas berupa film-film luar, dalam penelitian ini mengambil kasus perfilman Indonesia. Data yang diambil berdasarkan *website* IMDb ([www.imdb.com](http://www.imdb.com)) dan *filmindonesia* ([filmindonesia.co.id](http://filmindonesia.co.id)) dengan mengambil data perfilman 10 tahun terakhir. Untuk metode yang digunakan sama

dengan penelitian sebelumnya namun ditambahkan dengan Algoritma *Naïve Bayes* dan *K-Nearest Neighbors* (KNN) untuk mengetahui nilai akurasi tertinggi dalam memprediksi popularitas film, karena kedua algoritma tersebut memiliki keunggulan dalam hal mengklasifikasi sebuah data. Sehingga penulis mengambil judul “**Klasifikasi Algoritma *Decision Tree*, Algoritma *Naïve Bayes* dan *K-Nearest Neighbors* Terhadap Popularitas Perfilman Indonesia**”

### **1.2. Identifikasi Masalah**

Meskipun sudah ada bahkan banyak penelitian yang mengambil kasus berupa film namun dari banyaknya penelitian tersebut belum adanya penelitian yang mengambil data perfilman Indonesia, sebagian besar peneliti mengambil data film luar. Selain data film yang belum pernah dijadikan sebuah penelitian konsep prediksi popularitas film pun belum banyak diangkat dalam penelitian, lebih banyak penelitian mengenai *review* film atau popularitas sebuah film. Oleh sebab itu penulis mencoba membuat sebuah penelitian mengenai prediksi kepopuleran film di Indonesia.

### **1.3. Tujuan Penelitian**

Penelitian ini bertujuan untuk memprediksi kepopuleritasan film-film yang ada di Indonesia selama 10 tahun terakhir, dengan menggunakan metode algoritma *Decision Tree*, *Naïve Bayes* dan *K-Nearest Neighbors* untuk masing-masing kategori dengan nilai  $k=10$ .

### **1.4. Ruang Lingkup Penelitian**

Penelitian ini menggunakan data lokal yang ada di *website* IMDB dan data perfilman Indonesia. Data yang akan diolah sebanyak 556 data film dengan 9 atribut dimana sample data diambil dari data film yang populer dan tidak populer selama 10 tahun terakhir. Algoritma *Decision Tree*, *Naïve Bayes* dan *K-Nearest Neighbors* untuk menghasilkan nilai akurasi dari kategori yang ada.

### **1.5. Rumusan Masalah**

Berdasarkan latar belakang yang sudah dijelaskan sebelumnya, maka rumusan masalah dalam penelitian ini sebagai berikut:

1. Berapakan jumlah prediksi film populer dan tidak populer yang dihasilkan dengan menggunakan algoritma *decision tree*, algoritma *naive bayes* dan algoritma *K-Nearest Neighbors*?
2. Berapakah nilai *Confusion Matrix* yang didapat dengan menggunakan algoritma *decision tree*, algoritma *naive bayes* dan algoritma *K-Nearest Neighbors*?
3. Berapakah hasil yang didapat setelah dievaluasi menggunakan kurva ROC (*Receiver Operating Characteristic*) dan termasuk kedalam kategori apa berdasarkan hasil diagnosa kurva?

#### **1.6. Sistematika Penulisan**

Sistematika penulisan tesis ini terdiri dari 5 bab, dimana setiap bab terdiri dari beberapa sub bab, sebagai berikut :

##### **Bab I Pendahuluan**

Pada bab ini berisi latar belakang penulisan, identifikasi masalah, tujuan penelitian, ruang lingkup penelitian, rumusan masalah dan sistematika penulisan.

##### **Bab II Landasan dan Kerangka Pemikiran**

Pada bab ini membahas tinjauan pustaka mengenai data mining, algoritma *decision tree*, algoritma *naive bayes*, algoritma *K-Nearest Neighbors*, tinjauan studi yang berkaitan dengan penelitian yang sebelumnya dan kerangka pemikiran.

##### **Bab III Metode Penelitian**

Pada bab ini menjelaskan langkah-langkah dipenelitian menggunakan data mining dan algoritma-algoritma yang diterapkan.

##### **Bab IV Hasil Penelitian dan Pembahasan**

Pada bab ini menjelaskan mengenai pengujian model yang dihasilkan dengan menerapkan algoritma *decision tree*, *naive bayes* dan *K-Nearest Neighbors* yang kemudian dilakukan evaluasi dengan AUC (*Area Under Curve*).

## **Bab V Penutup**

Bab ini berisi mengenai hasil kesimpulan dari perhitungan berdasarkan pembahasan dibab sebelumnya dan memberikan saran untuk penelitian-penelitian selanjutnya.

## **BAB II**

### **LANDASAN TEORI**

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

##### **2.1.1. Review Film**

Film merupakan suatu selaput tipis yang dibuat dari seluloid dimana digunakan sebagai tempat gambar negatif yaitu tempat yang dibuat menjadi potret atau digunakan sebagai tempat gambar positif yaitu tempat untuk dimainkan di bioskop, film juga berarti sebagai lakon dalam cerita yaitu gambar hidup. *Review* merupakan sebuah teks yang digunakan untuk meninjau suatu karya baik film, buku dan karya lainnya yang memiliki tujuan untuk mengetahui kualitas, kelebihan dan kekurangan yang ada pada karya tersebut serta untuk melakukan kritik terhadap suatu peristiwa atau karya seni bagi khalayak (Pristiyanti, Fauzi, & Muflikhah, 2017).

##### **2.1.2. Data Mining**

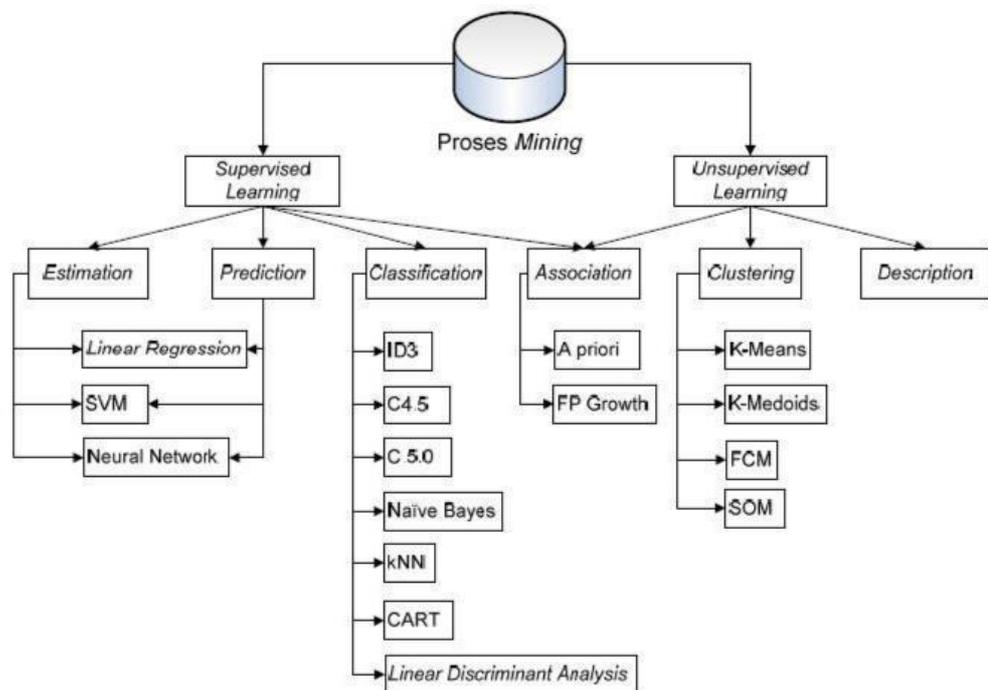
(McLeon, 2008) McLeod and Schell (2008:258) mengemukakan bahwa “*Data mining* adalah proses menemukan hubungan dalam data yang tidak diketahui oleh pengguna. Proses ini sama seperti seorang penambang yang mencari emas di aliran sungai pengunungan”.

*Data Mining* merupakan teknologi baru yang sangat berguna untuk membantu perusahaan-perusahaan menemukan informasi yang sangat penting dari gudang data mereka. Beberapa aplikasi *data mining* fokus pada prediksi, mereka meramalkan apa yang akan terjadi dalam situasi baru dari data yang menggambarkan apa yang terjadi dimasa lalu (Witten, I. H., Frank, E., & Hall, 2011).

Kakas *data mining* meramalkan *tren* dan sifat-sifat perilaku bisnis yang sangat berguna untuk mendukung pengambilan keputusan penting. Analisis yang diotomatisasi yang dilakukan oleh *data mining* melebihi yang dilakukan oleh sistem pendukung keputusan tradisional yang sudah banyak digunakan. Secara khusus, koleksi metode yang dikenal sebagai “*data mining*” menawarkan metodologi dan solusi teknik untuk mengatasi analisis data medis dan konstruksi prediksi model (Bellazzi, R., & Zupanb, 2008).

Berdasarkan tugas dan tujuan analisis, proses *data mining* dapat dibagi menjadi dua kategori utama tergantung pada adanya target variabel dan metode belajar (*learning*) yaitu antara proses belajar yang diawasi (*supervised*) dan tanpa pengawasan (*unsupervised*) (Vercillis, 2009).

1. Belajar yang diawasi (*supervised learning*)  
 Dalam analisis *supervised learning*, atribut target/label menggambarkan kelas yang dimiliki setiap catatan. Atau dengan kata lain metode belajar dengan adanya latihan (*training*) dan pelatihan. Contoh : *Regresi*, *Analisa Deskriminan*, *Artificial Neural Network* dan *Support Vector Machine*.
2. Belajar tanpa pengawasan (*unsupervised learning*)  
 Tanpa pengawasan analisis belajar tidak dipandu oleh atribut target. Oleh karena itu, data mining dalam hal ini ditunjukkan untuk menemukan pola berulang dan kedekatan dalam kumpulan data. Atau dengan kata lain metode belajar tanpa adanya latihan (*training*) dan pelatihan. Contoh *clustering* dan *Self Organization Map (SOM)*.



Sumber : (Ridwan, Suyono, & Sarosa, 2013)

Gambar 2.1. Proses Data Mining

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan :

1. Deskripsi (*Description*)  
Terkadang penelitian analisis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.
2. Estimasi (*Estimation*)  
Hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model dibangun menggunakan record lengkap yang menyediakan nilai variabel target sebagai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi.
3. Prediksi (*Prediction*)  
Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada dimasa datang. Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi. Dengan demikian, semakin kecil perbedaan antara apa yang diharapkan terjadi (hasil yang diharapkan) dan apa yang sebenarnya terjadi (diamati hasil), semakin baik prediksi.
4. Klasifikasi (*Classification*)  
Dalam klasifikasi, terdapat target variabel kategori. Pada klasifikasi, yang menjadi sasaran adalah variabel kategori, misalnya atribut penghasilan, yang bisa dikategorikan menjadi tiga kelas atau kategori yaitu, tinggi, sedang dan rendah. Model data mining membaca sejumlah besar *record*, dimana tiap *record* berisi informasi pada variabel target. Algoritma klasifikasi yang banyak digunakan secara luas untuk klasifikasi antara lain, *Decision Tree*, *Bayesian Classifier* dan *Neural Network* (Gorunescu, 2011).
5. Pengklusteran (*Clustering*)  
Merupakan pengelompokan *record*, pengamatan atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. *Cluster* adalah kumpulan *record* yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan *record-record* dalam *cluster* lain. Pengklusteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam pengklusteran. Pengklusteran tidak mencoba untuk

melakukan klasifikasi, mengestimasi atau memprediksi nilai dari variabel target. Akan tetapi, algoritma pengklusteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (*homogen*), yang mana kemiripan *record* dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan *record* dalam kelompok lain akan bernilai minimal.

6. Asosiasi (*Assosiation*)

Tugas Asosiasi dalam *data mining* adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja. Setiap asosiasi antara fitur-fitur yang dicari, bukan hanya satu yang memprediksi nilai kelas tertentu (Witten, I. H., Frank, E., & Hall, 2011). Pada prinsipnya, penemuan aturan asosiasi mempelajari aturan bagaimana kita memahami proses mengidentifikasi aturan antara ketergantungan yang berbeda dari fenomena kelompok (Gorunescu, 2011).

Proses data mining paling populer yaitu proses Cross-Industry Standard

Process for Data Mining (CRISP-DM), tahapan dalam CRISP-DM yaitu :

- a. *Bussiness Understanding*
- b. *Data Understanding*
- c. *Data Preparation*
- d. *Modelling*
- e. *Evaluation*
- f. *Deployment*

### 2.1.3. Klasifikasi

Klasifikasi adalah suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang ada yang telah diklasifikasi dan dengan menggunakan hasilnya untuk memberikan sejumlah aturan. Aturan-aturan tersebut digunakan pada data-data baru untuk diklasifikasi. Teknik ini menggunakan *supervised induction* yang memanfaatkan kumpulan pengujian dari data set yang terklasifikasi (Iskandar, D., & Suprpto, 2015).

Klasifikasi dapat diartikan proses menemukan kumpulan pola atau fungsi-fungsi yang mendeskripsikan dan memisahkan kelas data satu dengan lainnya, untuk dapat digunakan untuk memprediksi data yang belum memiliki kelas data tertentu. Jadi secara singkat, klasifikasi adalah proses untuk membedakan atau memisahkan kelas.



Sumber : Iskandar dan Suprpto (2013)

Gambar 2.2 Klasifikasi sebagai Pemetaan Input Himpunan Atribut  $x$  ke dalam Kelas Label  $y$

Data input untuk klasifikasi adalah sekumpulan *record* dikelompokkan sebagai *tupel*  $(x,y)$  dimana  $x$  adalah atribut dan  $y$  adalah kelas. Atribut *dataset* dapat berbentuk nilai *diskrit* atau kontinyu. Sedangkan kelas label harus berbentuk *diskrit*. Atribut kontinyu adalah atribut yang mengandung nilai angka riil, misalnya suhu, tinggi dan bobot. Sedangkan, atribut *diskrit* berupa kategori, misalnya kategori *mammal*, *reptile* dan *frog*. Atribut *diskrit* dapat juga bernilai *biner* (benar atau salah) model klasifikasi dibangun dengan data *training*. Model yang dibangun digunakan untuk menentukan kelas label yang belum diketahui.

#### 2.1.4. Algoritma C4.5

Algoritma C4.5 diperkenalkan oleh J. Ross Quinlan yang merupakan pengembang dari algoritma ID3, algoritma tersebut digunakan untuk membentuk pohon keputusan. Pohon keputusan dibuat dengan membagi nilai-nilai atribut menjadi cabang untuk setiap kemungkinan. Cara kerja pohon keputusan yaitu dengan melakukan penelusuran dari akar hingga ke cabang sampai class suatu

objek ditemukan. *Instance* diklasifikasikan dengan mengarahkan dari akar pohon sampai ke daun sesuai dengan hasil tes melalui node internal (Alfisahrin, 2014).

Beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5 (Ridwan et al., 2013) sebagai berikut :

1. Mempersiapkan data *training*, dapat diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon dengan menghitung nilai *gain* yang tertinggi dari masing-masing atribut atau berdasarkan nilai *index entropy* terendah.

Sebelum menghitung nilai *gain* terlebih dahulu mencari nilai *entropy* dengan rumus sebagai berikut :

$$Entropy(i) = - \sum_{j=1}^m f(i, j) \cdot \log_2 f(i, j) \dots \dots \dots (1)$$

Keterangan :

$i$  : himpunan kasus

$m$ : jumlah partisi  $i$

$f(i, j)$  : proporsi  $j$  terhadap  $i$

3. Kemudian menghitung nilai *gain* dengan rumus sebagai berikut :

$$Entropy(i) = - \sum_{i=1}^p \frac{n_i}{n} \cdot \log_2 \frac{n_i}{n} \dots \dots \dots (2)$$

Keterangan :

$P$  : jumlah partisi atribut

$n_i$  : proporsi  $n_i$  terhadap  $i$

$n$  : jumlah kasus dalam  $n$

4. Ulangi langkah ke 2 hingga semua nilai *record* terpartisi Proses partisi akan berhenti jika :
  - a. Semua *record* dalam *node*  $N$  mendapat kelas yang sama
  - b. Tidak ada atribut dalam *record* yang terpartisi kembali
  - c. Tidak ada *record* dalam cabang yang kosong

### 2.1.5. Naive Bayes

Klasifikasi adalah suatu bentuk analisis data yang ekstrak model menggambarkan kelas data penting. Klasifikasi memiliki berbagai aplikasi termasuk deteksi penipuan, pemasaran target, kinerja prediksi, manufaktur dan diagnosis media karena akurasi yang tinggi, tingkat prediksi dan metode otomatis untuk mencari hipotesis. Pada fase ini, *Naive Bayes* (NB) *classifier* digunakan untuk mengklasifikasikan Sinyal ECG sebagai normal dan abnormal

diimplementasikan dalam alat cepat *Miner*. *Naive Bayes* pengklasifikasi adalah statistik pengklasifikasi dapat memprediksi probabilitas keanggotaan kelas seperti probabilitas bahwa *tupel* diberikan milik kelas tertentu. (Padmavathi & Ramanujam, 2015).

Algoritma *naive bayes* merupakan suatu bentuk klasifikasi data dengan menggunakan metode probabilitas dan statistik. Metode ini pertama kali dikenalkan oleh ilmuwan Inggris Thomas Bayes, yaitu digunakan untuk memprediksi peluang yang terjadi dimasa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai *teorema bayes* (N. Nuraeni, 2017).

Metode *teorema bayes* kemudian dikombinasikan dengan *naive* yang diasumsikan dengan kondisi antar atribut yang saling bebas. Algoritma *naive bayes* dapat diartikan sebagai sebuah metode yang tidak memiliki aturan, *naive bayes* menggunakan cabang matematika yang dikenal dengan *teori probabilitas* untuk mencari peluang terbesar dari kemungkinan klasifikasi, dengan cara melihat frekuensi tiap klasifikasi pada *data training*.

Persamaan dari *teorema bayes* (Mukminin dan Riana, 2017) adalah :

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)}$$

Keterangan :

X : Data dengan *class* yang belum diketahui

Y: Hipotesis data yaitu suatu *class* spesifik

P(H|X) : Probabilitas hipotesis H berdasar kondisi X (*posteriori probability*)

P(H) : Probabilitas hipotesis H (*prior probability*)

P(X|Y) : Probabilitas X berdasarkan kondisi pada hipotesis H

P (X) : Probabilitas X Penjabaran

Algoritma *Naive Bayes* merupakan penyederhanaan metode *bayes*. Untuk mempermudah pemahaman, maka *Theorema Bayes* disederhanakan menjadi :

$$P(H|X) = P(X|H)P(H)$$

*Bayes rule* diterapkan untuk menghitung *posterior* dan *probabilitas* dari data sebelumnya. Dalam analisis *bayesian*, klasifikasi akhir dihasilkan dengan menggabungkan kedua sumber informasi (*prior* dan *posterior*) untuk menghasilkan *probabilitas* menggunakan aturan *bayes*.

### 2.1.6. *K-Nearest Neighbors*

Dalam pengenalan pola, *K-Nearest Neighbor* (KNN) adalah nonparametrik algoritma yang digunakan untuk klasifikasi belajar dan regresi. Karena merupakan jenis khas misalnya berbasis atau skema pembelajaran berbasis memori, semua perhitungan KNN ditangguhkan sampai klasifikasi dan tidak ada langkah pelatihan eksplisit diperlukan untuk membangun *classifier* KNN. Oleh karena itu, KNN adalah algoritma yang sangat sederhana namun efisien yang menunjukkan kompleksitas waktu  $O(1)$  ketika melatih *classifier* KNN dan  $O(mn + m \log 2m)$  ketika mengklasifikasikan contoh baru lebih pelatihan set dengan  $m$  contoh dan  $n$  atribut, dimana  $O(mn)$  adalah kompleksitas waktu untuk menghitung jarak antara instance baru dan masing-masing satu contoh pelatihan. Selain itu,  $O(m \log 2m)$  adalah kompleksitas waktu untuk menyortir jarak ketika menemukan  $k$  terdekat tetangga dari contoh baru (Wang, dan Ning., Et al, 2015).

Algoritma *k-nearest neighbor* merupakan salah satu metode untuk proses klasifikasi terhadap suatu objek berdasarkan data training yang memiliki jarak paling dekat dengan objek. Penentuan jarak dengan nilai terdekat atau terjauh dihitung berdasarkan jarak *Euclidean*. Proses perhitungan *K-Nearest Neighbor* :

*Preprocessing*, langkah pertama adalah menyiapkan data *training* dan

mendapatkan *tuple* himpunan  $DS = \left\{ \left( \begin{matrix} d \\ i \end{matrix}, \begin{matrix} c \\ j \end{matrix} \right) \mid \forall 0 \leq i < n, 0 \leq j < m \right\}$

dimana  $\begin{matrix} d \\ i \end{matrix}$  adalah *term vector* representation teks dokumen dan  $\begin{matrix} c \\ j \end{matrix}$  kategori label.

*Similarity Measure*, menggunakan TF x IDF untuk menghitung bobot setiap *term* dalam dokumen, sebagai variasi untuk meningkatkan akurasi yang signifikan. *Term Frequency* merupakan jumlah kemunculan sebuah *term* dalam sebuah dokumen. TF dapat dirumuskan pada persamaan berikut :

$$W_{t,d} = \log(tf_{t,d} + 1) \cdot \dots \dots \dots$$

*Inverse document Frequency* adalah log dari kebalikan probabilitas *term* yang ditemukan didalam dokumen. IDF dapat dirumuskan pada persamaan berikut :

$$idf_t = \log\left(\frac{n}{n_t}\right)$$

*Cosine similarity* adalah fungsi yang digunakan untuk menghitung kesamaan antara semua data *training* dengan dokumen X. *Cosine similarity* dapat dirumuskan pada persamaan berikut:

$$\cos(X, d_i) = \frac{\sum_{j=1}^m x_j \cdot d_{ij}}{\sqrt{\sum_{j=1}^m x_j^2} \cdot \sqrt{\sum_{j=1}^m d_{ij}^2}}$$

### 2.1.7. Evaluasi dan Validasi Metode

Evaluasi pada algoritma *Decision Tree* dan *naive bayes* menggunakan model *confusion matrix* dan kurva ROC (*Receiver Operating Characteristic*).

#### 1. Confusion Matrix

*Confusion Matrix* adalah alat yang sangat berguna untuk menganalisa seberapa baik pengklasifikasi bias mengenali *tuple* dari *class* yang berbeda (Utami, 2017). Evaluasi dengan menggunakan fungsi *confusion matrix* akan menghasilkan nilai *accuracy*, *precision* dan *recall*.

Kurva ROC akan digunakan untuk mengukur *Area Under Curve* (AUC). Kurva ROC membagi hasil positif dalam sumbu y dan hasil negative dalam sumbu x sehingga semakin besar area yang berada dibawah kurva, semakin baik pula hasil prediksi.

Tabel 2.1. Model Confusion Matrix

Classification	Predited Class	
	Class = Yes	Class = No

<i>Observed</i> <i>Class</i>	<i>Class = Yes</i>	A ( <i>true positive</i> –TP)	B ( <i>false negative</i> –FN)
	<i>Class = No</i>	C ( <i>false positive</i> –FP)	D ( <i>true negative</i> –TN)

Keterangan :

*True Positive* (TP) : proporsi positif dalam *dataset* yang diklasifikasikan positif

*True Negative* (TN) : proporsi *negative* dalam *dataset* yang diklasifikasikan *negative*

*False Positive* (FP) : proporsi negatif dalam *dataset* yang diklasifikasikan positif

*False Negative* (FN) : proporsi *negative* dalam *dataset* yang dikalsifikasikan negatif

Setelah data uji dimasukkan ke dalam *confusion matrix*, hitung nilai-nilai yang telah dimasukkan tersebut untuk dihitung jumlah *sensitivity* (*recall*), *specificity*, *precision* dan *accuracy*. *Sensitivity* digunakan untuk membandingkan jumlah TP terhadap jumlah *record* yang positif sedangkan *specificity* adalah perbandingan jumlah TN terhadap jumlah *record* yang negatif. Untuk menghitung digunakan persamaan dibawah ini (Irfansyah, 2016).

a.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \dots\dots\dots(8)$$

b.

$$\text{Sensitivity} = \frac{tp}{tp+fn} \dots\dots\dots(9)$$

c.

$$\text{Specitivity} = \frac{tn}{tn+fp} \dots\dots\dots(10)$$

d.

$$\text{PPV} = \frac{tp}{tp+fp} \dots\dots\dots(11)$$

e.

$$NPV = \frac{tn}{tn + fn} \dots\dots\dots(12)$$

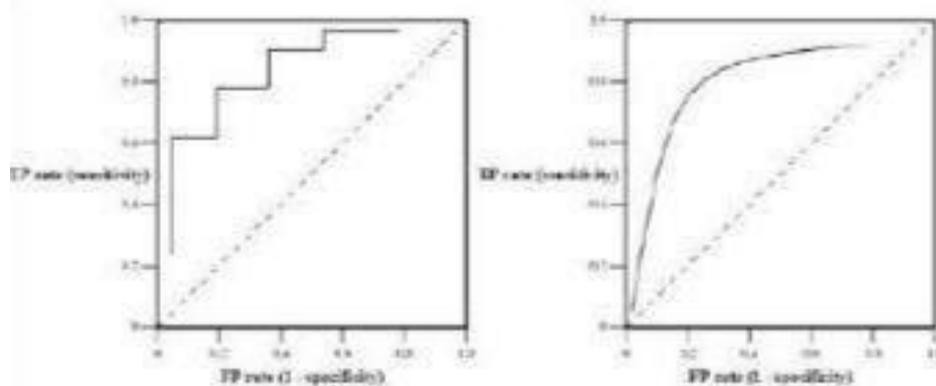
*Sensitivity* juga dapat dikatakan *true positive rate* (TP rate) atau *recall*.

Sebuah *sensitivity* 100% berarti bahwa pengklasifikasian mengikuti sebuah kasus yang diamati positif.

## 2. Kurva ROC

Fungsi kurva ROC adalah untuk menunjukkan akurasi dan membandingkan klasifikasi secara visual. ROC mengekspresikan *confusion matrix*, ROC adalah grafik dua dimensi dengan *false positive* sebagai garis *horizontal* dan *true positive* sebagai *garis vertical* (N. Nuraeni, 2017).

ROC (*Receive Operating Characteristic*) merupakan cara yang digunakan untuk menggambarkan akurasi diskriminasi dari suatu pengujian diagnosis untuk menentukan apakah seseorang menderita suatu penyakit tertentu atau tidak. Kurva ROC adalah plot dari *sensitifity* terhadap  $1 - \textit{specificity}$  untuk beberapa nilai *threshold* yang digunakan untuk menerangkan ketepatan uji dalam berbagai tingkatan titik potong dalam membaca *specificity* yang sesuai dengan *sensitivity* yang ada (Nugroho & Emiliyawati, 2017).



Sumber : Nuraeni, 2017

Gambar 2.3 Model Grafik ROC

Hasil perhitungan dapat divisualisasikan dengan kurva ROC (*Receiver Operating Characteristic*) atau AUC (*Area Under Curve*). Berikut tingkat nilai diagnosa dari ROC, yaitu : (Gorunescu, 2011).

- a. Akurasi bernilai 0.90 – 1.00 = *Excellent classification*
- b. Akurasi bernilai 0.80 – 0.90 = *Good classification*

- c. Akurasi bernilai  $0.70 - 0.80 = \textit{Fair classification}$
- d. Akurasi bernilai  $0.60 - 0.70 = \textit{Poor classification}$
- e. Akurasi bernilai  $0.50 - 0.60 = \textit{Failure}$

## 2.2. Tinjauan Studi

Berikut adalah beberapa penelitian yang menggunakan klasifikasi data mining untuk memprediksi dan menganalisa dataset yang diteliti diantaranya :

1. *Using Crowd-Source Based Features From Social Media And Conventional Features To Predict The Movies Popularity* (Ahmed, Jahangir, Afzal, Majeed, & Siddiqi, 2015). Memprediksi keberhasilan film telah menarik bagi para ekonom dan investor (media dan rumah produksi) serta analisis prediksi. Sejumlah atribut seperti pemeran, genre, anggaran, rumah produksi, peringkat PG mempengaruhi popularitas sebuah film. Media sosial seperti Twitter, YouTube, dll. Adalah platform utama tempat orang dapat berbagi pandangan tentang film. Makalah ini menjelaskan percobaan dalam analisis prediktif menggunakan algoritma pembelajaran mesin pada kedua fitur konvensional, yang dikumpulkan dari *database* film di Web serta fitur media sosial (komentar teks di YouTube, Tweet). Hasil menunjukkan bahwa sentimen yang dimanfaatkan dari media sosial dan fitur media sosial lainnya dapat memprediksi keberhasilan dengan lebih akurat daripada menggunakan fitur konvensional. Kami mencapai nilai terbaik 77% dan 61% masing-masing menggunakan fitur media sosial terpilih untuk Prediksi Rating dan Pendapatan, sedangkan fitur konvensional yang dipilih memberikan hasil masing-masing 76,2% dan 52%. Lebih lanjut ditemukan bahwa perpaduan kedua jenis atribut (konvensional dan yang dikumpulkan dari media sosial) dapat mengungguli pendekatan yang ada dalam domain ini.
2. *Role of Different Factors in Predicting Movie Success* (Bhave, Kulkarni, Biramane, & Kosamkar, 2015). Karena digitalisasi dan kemunculan media sosial yang cepat, industri film tumbuh dengan pesat. Jumlah rata-rata film yang diproduksi per tahun lebih besar dari 1000. Jadi untuk membuat film

ini menguntungkan, menjadi masalah yang membuat film ini berhasil. Dengan tingkat keberhasilan yang rendah, model dan mekanisme untuk memprediksi secara andal peringkat dan atau koleksi film *box office* dapat membantu mengurangi risiko bisnis secara signifikan dan meningkatkan pengembalian rata-rata. Model prediksi saat ini yang tersedia didasarkan pada berbagai faktor untuk penilaian film. Ini termasuk faktor-faktor klasik seperti pemain, produser, sutradara, dll. Atau faktor sosial dalam bentuk respon masyarakat pada berbagai *platform online*. Metodologi ini kurang memuaskan tingkat akurasi yang diperlukan. Makalah kami menunjukkan bahwa integrasi kedua faktor klasik dan sosial (antisipasi dan umpan balik pengguna) dan studi interelasi antara faktor-faktor klasik akan menyebabkan lebih akurat.

3. *Prediction of Movies Box Office Performance Using Social Media* (Apala et al., 2013). Dalam studi ini, kami menerapkan alat penambangan data untuk menghasilkan pola yang menarik untuk memprediksi kinerja film *box office* menggunakan data yang dikumpulkan dari berbagai media sosial dan sumber web termasuk Twitter, YouTube dan basis data film IMDb. Prediksi ini didasarkan pada faktor keputusan yang berasal dari basis data film historis, jumlah pengikut dari Twitter, dan analisis sentimen komentar pemirsa YouTube. Kami memberi label prediksi dalam tiga kelas, Hit, Netral, dan Flop, menggunakan alat pengelompokan K-Means Weka. Pola prediksi yang menarik dihasilkan oleh Jeka Weka. Karena prediksi kami adalah untuk film yang belum dirilis pada musim panas 2013, kinerja hasil akhir akan divalidasi oleh studi lanjutan.
4. *Movies Popularity Prediction Using Social Media and Conventional Features* (Jangid, Jadhav, Dhokate, Jadhav, & Bhandari, 2017). Sekarang sehari, penggunaan media sosial telah meningkat secara luas. Dengan menggunakan media sosial seperti Twitter, YouTube, dll. Pengguna dapat memposting ulasan mereka tentang film. Para ekonom, investor, dan analis prediksi sangat tertarik untuk memprediksi kesuksesan film mereka. Untuk memprediksi kesuksesan dan juga popularitas film, sejumlah faktor yang terpengaruh seperti aktor, aktris, anggaran yang diinvestasikan, rumah produksi, genre, peringkat PG. Dalam proyek ini, algoritma pembelajaran

mesin digunakan untuk analisis prediktif. Algoritma pembelajaran mesin diterapkan pada konvensional, yang dikumpulkan dari basis data film, dan fitur media sosial (komentar teks pada Tweet, YouTube). Menambang atribut dan konten media sosial memberi kita peluang untuk menemukan karakteristik struktur sosial, menganalisis pola tindakan secara kualitatif dan kuantitatif, dan kadang-kadang kemampuan untuk memprediksi peristiwa terkait manusia di masa depan. Hasil dari proyek ini yang memprediksi keberhasilan dengan kontrol dan penggunaan sentimen dari media sosial dan fitur media sosial lainnya.

5. Klasifikasi *Sentimen Review* Film Menggunakan Algoritma *Support Vector Machine* (Yulietha, Faraby, & Adiwijaya, 2017). Dengan kemajuan di bidang teknologi, seluruh informasi tentang semua film sudah tersedia di *Internet*. Jika informasi tersebut diolah dengan baik maka akan diperoleh kualitas dari informasi tersebut. Tugas Akhir ini bertujuan untuk menjelaskan klasifikasi sentimen pada dokumen *review* film. Satu hal yang penting dalam sebuah *review* atau ulasan yaitu opini yang terkandung didalamnya. Metode yang digunakan pada penelitian ini adalah *Support Vector Machine* (SVM). Metode ini dipilih karena mampu mengklasifikasikan data berdimensi tinggi sesuai dengan data yang digunakan pada Tugas Akhir ini yaitu berupa teks. Pengklasifikasi *Support Vector Machine* adalah teknik *machine learning* yang populer untuk klasifikasi teks karena dapat melakukan klasifikasi dengan cara belajar dari sekumpulan contoh dokumen yang telah diklasifikasi sebelumnya dan juga mampu memberikan hasil yang baik. Dari uji skenario yang dilakukan, dapat diketahui bahwa algoritma *Support Vector Machine* dapat digunakan untuk kasus *review* film dengan nilai *F1-Score* sebesar 84.9%.
6. Prediksi Rating Film Menggunakan Metode *Naïve Bayes* (Nugroho & Pratiwi, 2016). Pada saat ini perkembangan dunia perfilman sudah sangat pesat, contohnya dengan banyaknya film-film yang silih berganti untuk ditayangkan. Para penikmat film juga membutuhkan film-film yang mempunyai kualitas gambar, suara, alur cerita dan nilai positif yang baik dalam sebuah film, agar mereka tetap antusias dalam mengikuti film-film yang terbaru. Namun film-film yang ada tidak semuanya dapat dinikmati

dan tidak semua kalangan menyukai semua film. Agar suatu film dapat terus berkembang, tentunya membutuhkan penilaian-penilaian dari para penikmat film, untuk mengetahui selera film yang sesuai dengan para penikmat film. Untuk itu dibutuhkan analisis agar dapat mengetahui bagaimana minat penikmat film yaitu dengan membuat penilaian-penilaian yang nantinya digunakan untuk mengetahui rating suatu film menggunakan metode *naïve bayes* yaitu metode yang melakukan pendekatan statistika yang *fundamental* dalam pengenalan pola (*pattern recognition*). Pendekatan ini didasarkan pada kuantifikasi *trade-off* antara berbagai keputusan klasifikasi dengan menggunakan *probabilitas* dan resiko yang ditimbulkan dalam keputusan-keputusan tersebut. Metode tersebut merupakan salah satu metode dari *data mining*, dengan atribut yang sudah ditentukan, yaitu meliputi genre film, aktor film, bahasa, warna, durasi film, negara, dan lainnya yang dapat digunakan sebagai tolak ukur sutradara untuk membuat film..

7. *Sentiment Analysis* Peringkasan *Review* Film Menggunakan Metode *Information Gain* dan *K-Nearest Neighbor* (Pristiyanti et al., 2017), *Sentiment analysis* pada *review* film terbagi menjadi 2 yaitu berupa *review* positif dan *review* negatif. Pengelompokan hasil *sentiment analysis* dapat dipermudah dengan metode klasifikasi *k-nearest neighbor* dimana metode ini akan mencari dokumen yang memiliki kedekatan antara dokumen satu dengan yang lainnya. Penelitian ini menggunakan metode *information gain* untuk mengurangi banyak fitur yang digunakan pada saat proses klasifikasi. Metode ini akan memprediksi ada atau tidak adanya *term* dalam sebuah dokumen sehingga *term* yang sering muncul memiliki nilai *information gain* yang rendah sedangkan *term* yang jarang muncul atau hanya muncul pada salah satu kategori memiliki nilai *information gain* yang tinggi. *Term* dengan nilai *information gain* yang tinggi akan dapat digunakan untuk proses klasifikasi. Hasilnya penggunaan seluruh *term* untuk klasifikasi menghasilkan akurasi sebesar 92% dimana nilai akurasinya lebih baik dibandingkan dengan adanya *feature selection* karena adanya penghapusan *term* yang memiliki nilai *information gain* yang rendah.

8. *Text Mining Untuk Analisis Sentimen Review Film Menggunakan Algoritma K-Means* (Budi, 2017), Kemudahan manusia didalam menggunakan *website* mengakibatkan bertambahnya dokumen teks yang berupa pendapat dan informasi. Dalam waktu yang lama dokumen teks akan bertambah besar. *Text mining* merupakan salah satu teknik yang digunakan untuk menggali kumpulan dokumen text sehingga dapat diambil intisarinnya. Ada beberapa algoritma yang digunakan untuk penggalian dokumen untuk analisis sentimen, salah satunya adalah K-Means. Didalam penelitian ini algoritma yang digunakan adalah K-Means. Hasil penelitian menunjukkan bahwa akurasi K-Means dengan *dataset* digunakan 300 positif dan 300 negatif akurasinya 57.83%, 700 dokumen positif dan 700 negatif akurasinya 56.71%, 1000 dokumen positif dan 1000 negatif akurasinya 50.40%. Dari hasil pengujian disimpulkan bahwa semakin besar *dataset* yang digunakan semakin rendah akurasi K-Means.
9. *Klasifikasi Sentimen pada Movie Review dengan Metode Multinomial Naïve Bayes* (Wisudawati, Adiwijaya, & Faraby, 2017). Opini orang lain terhadap suatu *movie review* dimedia sangat penting dalam membuat suatu keputusan. Untuk mengetahui opini seseorang terhadap suatu *movie review* dimedia diperlukan sistem yang dapat mempermudah dalam mengetahui opini seseorang. Klasifikasi sentimen dapat membantu dalam membangun sistem untuk mengetahui opini seseorang terhadap *movie review*. *Dataset* yang digunakan dalam proses klasifikasi sentimen ini adalah *Internet Movie Database* (IMDb). Namun yang menjadi permasalahan dalam mengetahui *polaritas* suatu opini dalam proses klasifikasi sentimen pada *dataset movie review* adalah adanya data yang tidak terstruktur, atribut data yang begitu banyak serta adanya negasi yang menyebabkan polaritas suatu kata akan berbeda pada konteks teks yang berbeda. Dengan permasalahan tersebut maka proses klasifikasi pada *dataset* tersebut akan di klasifikasikan ke dalam dua kelas polaritas yaitu positif dan negatif. Metode klasifikasi yang digunakan adalah dengan menggunakan metode *multinomial naïve bayes*. Untuk meningkatkan nilai akurasi metode *multinomial naïve bayes* dilakukan dengan memecahkan masalah

diatas. Dalam memecahkan permasalahan tersebut yang dilakukan adalah pertama, akan dilakukan proses *preprocessing* untuk menangani data *noisy*. Kedua, dilakukan penanganan negasi, adapun lingkup permasalahan negasi yang akan dilakukan adalah negasi dengan kata “not”, “n’t”, “no”. Ketiga, dilakukan penghitungan bobot setiap kata dengan menggunakan TF-IDF. Berdasarkan hasil pengujian yang telah dilakukan diperoleh nilai akurasi terbesar 85,16%. Hal tersebut dikarenakan *multinomial naïve bayes* dengan *negation handling* berdasarkan *punctuation, preprocessing* dan TF-IDF dapat meningkatkan nilai akurasi terhadap metode *multinomial naïve bayes*. Komparasi algoritma klasifikasi *machine learning* dan *feature selection* pada analisis *sentimen review* film (Chandani, Wahono, & Purwanto, 2015). Analisis sentimen adalah proses yang bertujuan untuk menentukan isi dari *dataset* yang berbentuk teks bersifat positif, negatif atau netral. Saat ini, pendapat khalayak umum menjadi sumber yang penting dalam pengambilan keputusan seseorang akan suatu produk. Algoritma *klasifikasi seperti Naïve Bayes (NB), Support Vector Machine (SVM) dan Artificial Neural Network (ANN)* diusulkan oleh banyak peneliti untuk digunakan pada analisis *sentimen review* film. Namun, klasifikasi sentimen teks mempunyai masalah pada banyaknya atribut yang digunakan pada sebuah *dataset*. *Feature selection* dapat digunakan untuk mengurangi atribut yang kurang *relevan* pada *dataset*. Beberapa algoritma *feature selection* yang digunakan adalah *information gain, chi square, forward selection dan backward elimination*. Hasil komparasi algoritma, SVM mendapatkan hasil yang terbaik dengan *accuracy* 81,10% dan AUC 0,904. Hasil dari komparasi *feature selection, information gain* mendapatkan hasil yang paling baik dengan *average accuracy* 84,57% dan *average AUC* 0,899. Hasil *integrasi* algoritma klasifikasi terbaik dan algoritma *feature selection* terbaik menghasilkan *accuracy* 81,50% dan AUC 0,929. Hasil ini mengalami kenaikan jika dibandingkan hasil eksperimen yang menggunakan SVM tanpa *feature selection*. Hasil dari pengujian algoritma *feature selection* terbaik untuk setiap algoritma klasifikasi adalah

*information gain* mendapatkan hasil terbaik untuk digunakan pada algoritma NB, SVM dan ANN.

10. Komparasi Algoritma Klasifikasi *Machine Learning* Dan *Feature Selection* pada Analisis Sentimen *Review* Film (Chandani et al., 2015). Analisis sentimen adalah proses yang bertujuan untuk menentukan isi dari dataset yang berbentuk teks bersifat positif, negatif atau netral. Saat ini, pendapat khalayak umum menjadi sumber yang penting dalam pengambilan keputusan seseorang akan suatu produk. Algoritma klasifikasi seperti *Naïve Bayes* (NB), *Support Vector Machine* (SVM), dan *Artificial Neural Network* (ANN) diusulkan oleh banyak peneliti untuk digunakan pada analisis sentimen *review* film. Namun, klasifikasi sentimen teks mempunyai masalah pada banyaknya atribut yang digunakan pada sebuah *dataset*. *Feature selection* dapat digunakan untuk mengurangi atribut yang kurang relevan pada *dataset*. Beberapa algoritma *feature selection* yang digunakan adalah *information gain*, *chi square*, *forward selection* dan *backward elimination*. Hasil komparasi algoritma, SVM mendapatkan hasil yang terbaik dengan *accuracy* 81.10% dan AUC 0.904. Hasil dari komparasi *feature selection*, *information gain* mendapatkan hasil yang paling baik dengan *average accuracy* 84.57% dan *average AUC* 0.899. Hasil integrasi algoritma klasifikasi terbaik dan algoritma *feature selection* terbaik menghasilkan *accuracy* 81.50% dan AUC 0.929. Hasil ini mengalami kenaikan jika dibandingkan hasil eksperimen yang menggunakan SVM tanpa *feature selection*. Hasil dari pengujian algoritma *feature selection* terbaik untuk setiap algoritma klasifikasi adalah *information gain* mendapatkan hasil terbaik untuk digunakan pada algoritma NB, SVM dan ANN.
11. Prediksi kepopuleran lagu berdasarkan tangga lagu billboard menggunakan *decision tree* dan k-means (Gumilar, Pudjiantoro, & Yuniarti, 2017). Billboard telah menjadi sumber terpercaya untuk peringkat popularitas lagu selama 60 tahun terakhir, dan sebagian besar label rekaman mengacu pada peringkat yang diberikan Billboard. Lagu hits biasanya tidak hanya dipengaruhi oleh lirik dan artis yang menyanyikannya, beberapa lagu hits dipengaruhi juga oleh faktor seperti

artis, genre, label rekaman dan lain sebagainya. Namun jika label rekaman dapat memprediksi sendiri apakah sebuah lagu dapat masuk ke dalam peringkat yang ada di Billboard tentunya akan sangat membantu. Dalam tangga lagu terdapat atribut yang dipertimbangkan seperti artis, judul, *genre* dan lainnya sehingga kombinasi berbagai atribut tersebut menjadi suatu pola dalam mengelompokkan sebuah lagu dalam tangga lagu. Penelitian ini bertujuan membangun sistem yang dapat memprediksi apakah suatu lagu dapat dikategorikan menjadi hits atau tidak menggunakan *Decision Tree* dan K-Means.

12. *Prediction of Movies popularity Using Machine Learning Techniques* (Latif & Afzal, 2016). Jumlah film dirilis setiap minggu. Ada sejumlah besar data yang terkait dengan film-film yang tersedia melalui internet, karena banyaknya data yang tersedia, ini adalah topik penambahan data yang menarik. Prediksi film adalah masalah yang kompleks. Setiap penonton, produser, rumah produksi sutradara semua penasaran dengan film yang akan diputar di teater. Banyak pekerjaan yang telah dilakukan berkaitan dengan film menggunakan jejaring sosial, artikel blog tetapi lebih sedikit telah dieksplorasi oleh data dan atribut yang terkait dengan film yang kontinu dan dalam dimensi yang berbeda. Kami telah menggunakan IMDB untuk eksperimen kami. Kami membuat *dataset* dan kemudian mengubahnya dan menerapkan pendekatan pembelajaran mesin untuk membangun model yang efisien yang dapat memprediksi popularitas film
13. *Implementasi Sentiment Analysis untuk Menentukan Tingkat Popularitas Tujuan Wisata* (Murnawan & Sinaga, 2017). Pemanfaatan dari *sentiment analysis* telah banyak diterapkan di berbagai bidang, bahkan beberapa perusahaan sudah mulai mengambil nilai lebih dari penerapan analisis tersebut untuk mendapat manfaat baru. Pada dasarnya, sebagian besar algoritma untuk *sentiment analysis* berdasarkan pada hasil klasifikasi data *training* yang menggunakan sekumpulan koleksi data teks. Sebelum data di latih, sekumpulan data teks tersebut diekstraksi terlebih dahulu. Pada penelitian ini, peneliti memutuskan untuk menggunakan strategi klasifikasi berdasarkan pada algoritma *naïve bayes* (NB) karena merupakan suatu

metode yang sederhana dan intuitif yang kinerjanya mirip dengan pendekatan lain. Selain itu, berdasarkan penelitian-penelitian terdahulu, NB menggabungkan efisiensi waktu kinerja yang optimal yang cukup akurat. Pada penelitian ini, sebuah prototipe aplikasi berbasis web dibangun untuk melakukan *crawling* informasi baik yang terdapat dari beberapa forum diskusi yang membahas tentang suatu tujuan wisata maupun dari media sosial facebook dan instagram, kemudian informasi tersebut dianalisis untuk dapat menghasilkan peringkat popularitas tempat pariwisata. Ada lima komponen yang akan dinilai untuk menghasilkan nilai akhir dari peringkat popularitas yaitu *comment count*, *facebook likes count*, *facebook were here count*, *facebook talking about* dan *instagram visitor*. Masing-masing komponen tersebut dianalisis menggunakan sentiment analisis untuk menentukan apakah suatu komentar atau opini bernilai positif, negatif atau netral. Berdasarkan hasil penelitian ini, tingkat akurasi *sentiment analysis* sangat tergantung dari banyaknya *frase* yang di dapat. Semakin banyak *frase* yang dimiliki sebagai *core* dari algoritma tersebut, maka semakin akurat *sentiment analysis* yang disajikan. Diharapkan dengan hasil implementasi ini, dapat menjadi media informasi yang bermanfaat baik bagi wisatawan maupun pemerintah.

14. Penentuan Kelayakan Kredit Dengan Algoritma *Naïve Bayes Classifier*: Studi Kasus Bank Mayapada Mitra Usaha Cabang PGC (Nia Nuraeni, 2017). Dalam menganalisis kredit kadang-kadang petugas kredit kurang akurat dalam analisis kredit, sehingga dapat menyebabkan peningkatan kredit macet. Klasifikasi algoritma *data mining* secara luas digunakan untuk menentukan kelayakan kredit dari salah satu *Naïve Bayes classifier*, BC unggul dalam meningkatkan nilai akurasi yang tinggi tetapi lemah dalam pemilihan atribut. Setelah menguji algoritma *Naïve Bayes* menghasilkan akurasi 89,33 % dan nilai AUC 0.955.

Tabel 2.2 Ringkasan Penelitian

No	Judul	Penulis	Kerangka Pemikiran	Hasil	Terbit
1.	<i>Using</i>	Mehreen	a) <i>Data Collection</i> ,	Akurasi 89%	Jurnal

No	Judul	Penulis	Kerangka Pemikiran	Hasil	Terbit
	<i>Crowd-Source Based Features From Social Media And Conventional Features To Predict The Movies Popularity</i>	Ahmed 1 , Maham Jahangir 2 , Dr. Hammad Afzal, Dr. Awais Majeed 4 , Dr. Imran Siddiqi 5	didapat dari website IMDB dna media sosial seperti youtube dan twitter. b) <i>preprocessing</i> dengan memfilter atribut yang digunakan. c) <i>Algorithms</i> menggunakan Linear, Polynominal Regression dan ID3 d) <i>Result</i> berupa presiksi rating.	dengan <i>Naive Bayes</i> dan 100% dengan J48	Internasional
2.	<i>Role of Different Factors in Predicting Movie Success</i>	Anand Bhave, Himanshu Kulkarni, Vinay Biramane, Pranali Kosamkar	a) <i>Data Acquisition</i> , b) <i>Data preprocessing and formatiom</i> of usable dataset c) <i>Feature Extraction</i> , terbagi menjadi 2 <i>platforms pre-production phase</i> dan <i>post-production phase</i> d) <i>Classification model</i> e) <i>Result Generation and Accuracy Estimation</i>	Nilai <i>R-squared</i> yang diperoleh adalah 0,7057	Konferensi Internasional
3.	<i>Prediction of Movies Box Office Performance Using Social Media</i>	Krushikanth R. Apala, Merin Jose, S upreme Motnam, C.- C. Chan, Kathy J. Liszka, and Federico de Gregorio	a) pengumpulan data, didapat dari <i>trailer</i> resmi film bulan mei, juni dan juli 2013 di AS b) eksperimen, menormalkan data penelitian, penerapan K-Means dan menghasilkan model prediksi c) hasil dan diskusi, dataset berlabel diuji dengan tools weka	Akurasi <i>naive bayes uniform weighted</i> 89% dan <i>non uniform weighted</i> 86%. Sedangkan dengan J48 <i>uniform weighted</i> 100% dan <i>non uniform weighted</i> 94%	Jurnal Internasional

No	Judul	Penulis	Kerangka Pemikiran	Hasil	Terbit
4.	<i>Movies Popularity Prediction Using Social Media and Conventional Features</i>	Babita M. Jangid, Chaitali K.Jadhav, Swati M. Dhokate, Grish M.Jadhav, Dr. G.M.Bhandari	a) <i>Data Collector</i> b) <i>feature extraction</i> c) <i>conversions</i> d) <i>prediction</i>	Perbandingan fitur konvensional dan media sosial menghasilkan fitur media sosial dapat memprediksi popularitas film	Jurnal Internasional
5.	Klasifikasi Sentimen Review Film Menggunakan Algoritma Support Vector Machine	Irene Mathilda Yulietha, Said Al Faraby, Adiwijaya	a) <i>Dataset</i> b) <i>preprocessing</i> c) ekstraksi fitur d) klasifikasi - <i>data training</i> - <i>learning method</i> “SVM” - <i>learning model</i> ) e) <i>validasi dan evaluasi (k-fold cross validation dan confusion matrix)</i>	Dari uji skenario diketahui algoritma SVM dapat digunakan untuk kasus <i>review film</i> dengan nilai sebesar 84.9%	Jurnal Nasional
6.	Prediksi Rating Film Menggunakan Metode <i>Naïve Bayes</i>	Riszki Wijayatun Pratiwi, Yusuf Sulistyono Nugroho	a) <i>dataset</i> diambil dari situs <a href="https://www.kaggle.com/">https://www.kaggle.com/</a> digunakan untuk data <i>training</i> dan data <i>testing</i> b) <i>preprocessing</i> dengan menerapkan metode <i>cleaning</i> c) pengolahan data dengan <i>software rapid miner</i>	Nilai <i>accuracy</i> 55,80%, <i>precision</i> 32,41% dan <i>recall</i> 46,70%	Jurnal Nasional
7.	<i>Sentiment Analysis</i> Peringkasan Review Film Menggunakan Metode	Ria Ine Pristiyanti, Mochammad Ali Fauzi, Lailil Muflikhah	Data <i>training preprocessing</i> dengan menggunakan <i>feature selection</i> untuk pengurangan fitur term klasifikasi dengan	Pengujian nilai k=5 menghasilkan nilai akurasi sebesar 91%	Jurnal Nasional

No	Judul	Penulis	Kerangka Pemikiran	Hasil	Terbit
	<i>Information Gain dan K-Nearest Neighbor</i>		metode KNN		
8.	<i>Text Mining Untuk Analisis Sentimen Review Film Menggunakan Algoritma K-Means</i>	Setyo Budi	a) <i>dataset review film</i> b) <i>preprocessing</i> dan pembobotan TF-IDF c) <i>clustering algoritma k-means</i> d) evaluasi ( <i>accuracy, precision, recall</i> )	Akurasi K-Means dengan <i>dataset</i> digunakan 300 positif dan 300 negatif akurasiya 57.83%, 700 dokumen positif dan 700 negatif akurasiya 56.71%, 1000 dokumen positif dan 1000 negatif akurasiya 50.40%	Jurnal Nasional
9.	Klasifikasi Sentimen pada <i>Movie Review</i> dengan Metode Multinomial <i>Naïve Bayes</i>	Jenepte Wisudawati, Adiwijaya, Said AL Faraby	Data training <i>negation handling preprocessing TF-IDF train clasifier using MNB trained clasifier classification</i> hasil klasifikasi	Nilai akurasi sebesar 85,16%	Prosiding Nasional
10	Komparasi algoritma klasifikasi <i>machine learning</i> dan <i>feature selection</i> pada analisis <i>sentimen</i>	Vinita Chandani, Romi Satria Wahono, Purwanto	a) <i>dataset</i> b) <i>text processing</i> c) komparasi <i>feature selection</i> d) <i>classifier</i> e) <i>evaluation, accuracy &amp; AUC</i>	Hasil a) nilai <i>accuracy ANN</i> 51,80%, SVM 81,10% dan NB 74% b) nilai AUC ANN 0,500, SVM 0.904	Jurnal Nasional

No	Judul	Penulis	Kerangka Pemikiran	Hasil	Terbit
	<i>review film</i>			dan NB 0,734	
11.	Prediksi kepopuleran lagu berdasarkan tangga lagu billboard menggunakan <i>decision tree</i> dan k-means	Desta Gumilar, Tacbir Hendro Pudjiantoro, Rezki Yuniarti	a) pengumpulan data b) proses data ( <i>praproses, proses prediksi</i> ) c) evaluasi	menghasilkan nilai <i>presicion</i> sebesar 80%, <i>recall</i> sebesar 70% dan akurasi sebesar 93%	Prosiding Nasional
12	<i>Prediction of Movies popularity Using Machine Learning Techniques</i>	Muhammad Hassan Latif, Hammad Afzal	a) <i>data extraction</i> b) <i>data preprocessing</i> c) <i>data integration and transformation</i> d) <i>feature selection</i> e) <i>classification</i>	Nilai akurasi <i>regresi logistik</i> 84,34% dan <i>logistik</i> sederhana 84,15%. pengklasifikasi jaringan saraf yaitu <i>multilayer perceptron</i> menghasilkan akurasi <i>pf</i> 79,07%, dan pohon keputusan hasil j48 adalah 82,42%	Jurnal Internasional
13	Implementasi <i>Sentiment Analysis</i> untuk Menentukan Tingkat Popularitas Tujuan Wisata	Murnawan, Ardiles Sinaga	a) pengumpulan data b) proses text <i>preprocessing</i> c) proses klasifikasi d) evaluasi	a) 100 <i>phrase accuracy</i> 65,65% b) 5000 <i>pharase accuracy</i> 82,67%	Prosiding Nasional
14	Penentuan Kelayakan	Nia Nuraeni	a) <i>bussiness understanding</i>	Nilai akurasi algoritma	Jurnal Nasional

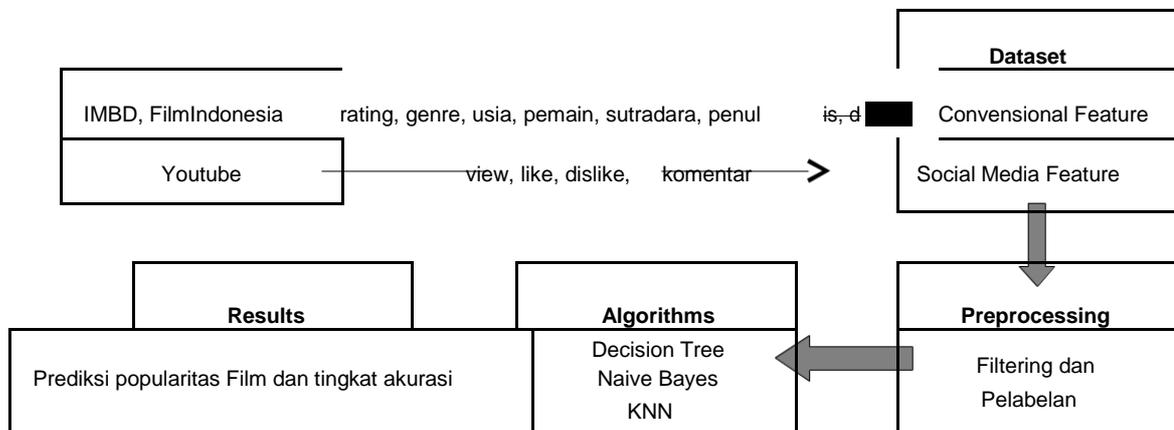
No	Judul	Penulis	Kerangka Pemikiran	Hasil	Terbit
	Kredit Dengan Algoritma <i>Naïve Bayes Classifier</i> : Studi Kasus Bank Mayapada Mitra Usaha Cabang PGC		b) <i>data understanding</i> c) <i>data preparation</i> d) <i>modelling</i> e) <i>evaluation</i> f) <i>deployment</i>	klasifikasi <i>naive bayes classifier</i> 89.33%. dan nilai AUC adalah 0,955 dengan tingkat diagnosa <i>Excellent Classification</i>	

**2.3. Novelti Penelitian**

Penelitian ini mengangkat kasus baru yang belum pernah diteliti sebelumnya yaitu prediksi popularitas perfilman Indonesia dengan menggunakan metode data mining seperti *Decision Tree*, *Naive Bayes* dan *K-Nearest Neighbors* untuk menghasilkan prediksi popularitas film Indonesia.

**2.4. Kerangka Pemikiran**

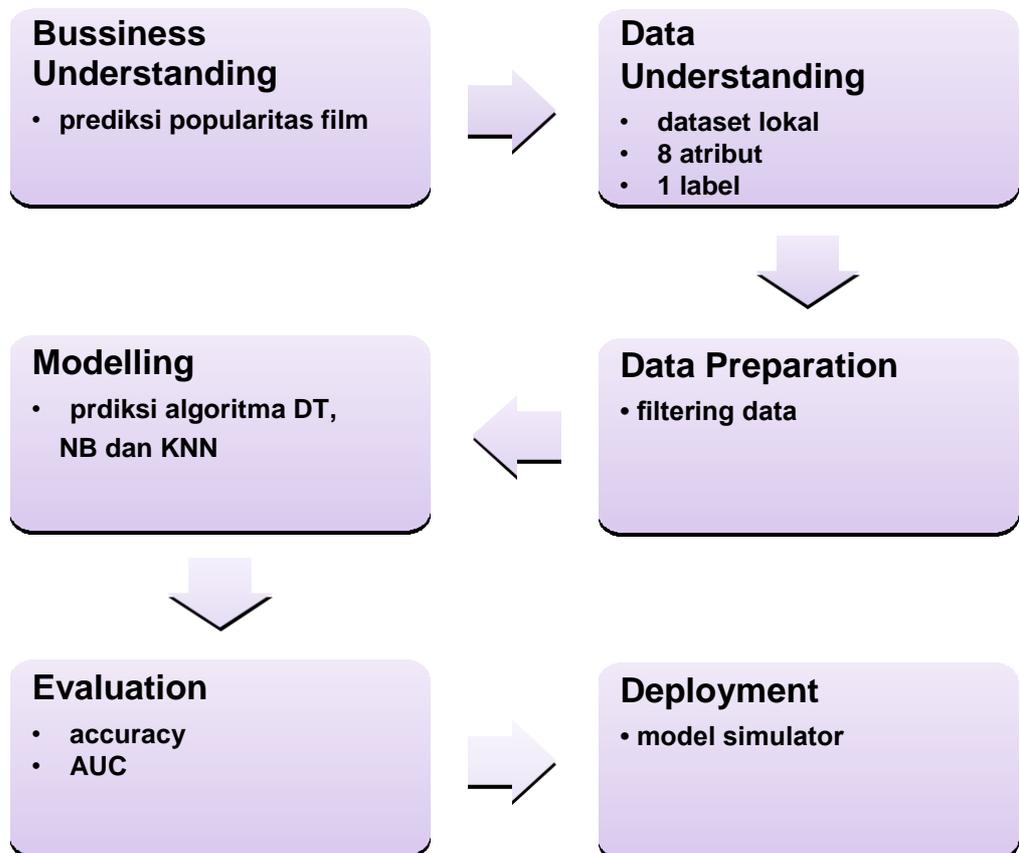
Kerangka pemikiran ini didasarkan pada tinjauan pustaka dan tinjauan studi diatas yang akan digunakan sebagai acuan penelitian ini. Penggambaran kerangka pemikiran terdiri dari beberapa tahap yang terlihat pada gambar 2.4. Permasalahan dalam penelitian ini yaitu belum adanya peneliti yang mengangkat kasus mengenai tingkat popularitas film-film Indonesia. Berikut penggambaran dari kerangka pemikiran penelitian ini.



Gambar 2.4 Kerangka Pemikiran

### **2.5.Objek Penelitian**

*Internet Movie Database* (IMDb) adalah situs *web* yang menyediakan informasi mengenai film dari seluruh dunia, termasuk orang-orang yang terlibat di dalamnya mulai dari aktor/aktris, sutradara, penulis sampai penata rias dan musikus.



klasifikasi data berdasarkan algoritma yang digunakan dan nilai akurasi yang dapat digunakan dalam memprediksi popularitas film-film yang ada di Indonesia.

### **3.1.2. Data Understanding**

Tahap *data understanding* meliputi pengumpulan data, melakukan analisis data dan mengevaluasi data tersebut. Sumber data yang digunakan yaitu data perfilman Indonesia dengan 8 atribut. Kemudian data dianalisis untuk memperkirakan banyaknya data yang akan diambil dan menghitung banyaknya data dengan keterangan populer dan tidak populer.

### **3.1.3. Data Preparation**

Tahap ini untuk menyiapkan data yang akan digunakan dalam proses *modeling*, data akan diseleksi dan dibersihkan sesuai dengan kebutuhan. Data yang akan digunakan sebanyak 556 data film terdiri dari judul film populer dan tidak populer dengan 8 atribut dan 1 atribut label yang didapat melalui fitur konvensional dan fitur media sosial.

### **3.1.4. Modeling**

Pemrosesan data dilakukan pada tahap ini dengan menggunakan teknik *data mining*. Model pendekatan yang digunakan berupa pendekatan algoritma *Decision Tree*, *Naive Bayes* dan *K-Nearest Neighbors* untuk menghasilkan sebuah nilai prediksi dan akan membentuk sebuah pohon keputusan.

### **3.1.5. Evaluation**

Tahap evaluasi sama juga dengan tahap klasifikasi dimana tahap ini ditentukan pengujian untuk akurasi. Tahap pengujian dikerjakan untuk melihat hasil dari penerapan algoritma yang diusulkan serta evaluasi menggunakan *confusion matrix* dan kurva ROC. Tujuan dari evaluasi untuk menentukan nilai dari model yang dibuat pada tahap sebelumnya. Penjelasan lebih detail mengenai pengujian model pendekatan metode algoritma yang diterapkan akan dibahas pada bab selanjutnya.

### 3.1.6. *Deployment*

Tahap *deployment* merupakan tahapan terakhir dalam model CRISP-DM dimana hasil dari seluruh tahapan yang telah dilakukan diimplementasikan secara nyata. Pada tahap ini dilakukan model simulator dalam pengimplementasian untuk memprediksi popularitas film ke dalam kategori populer atau tidak populer.

## 3.2. Metode Pemilihan Populasi dan Sampel

### 3.2.1. Populasi

Populasi adalah wilayah generalisasi yang terdiri atas obyek atau subyek yang mempunyai kualitas dan karakteristik tertentu yang ditetapkan oleh peneliti untuk dipelajari dan kemudian ditarik kesimpulannya (Sugiyono, 2013).

Populasi penelitian ini yaitu film-film Indonesia yang tayang dalam kurun waktu 10 tahun terakhir yang berasal dari data lokal sebuah *website* IMBD dan Film Indonesia serta media sosial seperti Official Youtube.

### 3.2.2. Sampel

Sampel dalam penelitian ini adalah data perfilman Indonesia yang populer dan tidak populer. Data ini bersifat lokal yang terpublikasi oleh pihak luar pada website IMDB dan Official youtube, jumlah sampel yang diambil sebagai berikut :

Tabel 3.1 Sampel *Dataset*

	Jumlah
Populer	437
Tidak Populer	119
Total Sampel	556

### 3.3. Pengumpulan Data

Metode pengumpulan data dibagi menjadi dua sumber data yaitu data primer dan data sekunder. Data primer yaitu data yang dikumpulkan pertama kali dan untuk melihat apa yang sesungguhnya terjadi melalui *observasi*, *interview*, *kuesioner* dan lainnya. Data sekunder adalah data yang telah dikumpulkan dan dianalisis oleh orang lain baik yang telah dipublikasikan maupun yang belum

dipublikasikan, misalnya dari dokumentasi, literatur, buku, jurnal dan informasi lainnya yang ada hubungannya dengan masalah yang diteliti (Sugiyono, 2013).

Metode pengumpulan untuk mendapatkan data-data yang akan digunakan dalam penelitian ini yaitu metode pengumpulan data primer. Dimana data yang akan digunakan yaitu *dataset* IMDb (*Internet Movie Database*) 10 tahun terakhir, yang terdiri dari 9 variabel dan 556 judul film.

### **3.4.Instrument Penelitian**

Instrumen penelitian dalam penelitian ini yaitu :

1. Menggunakan data primer yaitu data perfilman Indonesia yang tayang selama 10 tahun terakhir dengan rating tinggi dan rendah yang digunakan sebagai instrumen dalam memperoleh data dalam penentuan kepopuleran film tersebut.
2. Penyajian data dalam bentuk tabulasi model sebanyak 556 data judul film dan 8 variabel.
3. Penganalisan data menggunakan perangkat lunak berupa *RapidMiner*.

### **3.5.Metode Analisis dan Pengujian Data**

Analisis data dalam penelitian ini melalui data judul film yang memiliki rating tinggi dan rendah kemudian data tersebut diolah dan diuji pada algoritma *decision tree*, *naive bayes* dan *K-Nearest Neighbour*. Hasil pengujian dari algoritma tersebut diuji kembali menggunakan *confusion matrix* dan kurva AUC untuk mengetahui tingkat akurasi yang dihasilkan dari metode-metode tersebut.

## BAB IV

### HASIL PENELITIAN DAN PEMBAHASAN

#### 4.1. Business Understanding

Tujuan dari penelitian ini yaitu untuk memprediksi popularitas film dari data perfilman Indonesia menggunakan algoritma *decision tree*, *naive bayes* dan *k-nearest neighbors* menjadi populer atau tidak populer sebagai batasan penilaian untuk prediksi popularitas film yang akan datang.

#### 4.2. Data Understanding

Data yang diperoleh melalui pengumpulan data dengan fitur konvensional dengan mengunjungi *website* IMBD dan fitur sosial media dengan mengunjungi official youtube produksi film tersebut. Data yang terkumpul sebanyak 556 judul film yang pernah ditayangkan di Indonesia selama tahun 2009 sampai dengan tahun 2018. Berikut ini data-data film yang telah dikumpulkan :

Tabel 4.1. Data Film Indonesia

Movie	Year	Ratings	Genre	Usia	Durasi (menit)	Views	Likes	Dislikes	Komenta r
308	2013	6.7	Horror	R	120	149053	332	23	40
#66	2016	5.6	Action	D	116	60895	120	20	34
12 Menit: Kemenangan untuk	2014	8.6	Drama	R	112	100599	619	6	52
Selamanya 12:06 Rumah	2017	2.7	Horror	13	78	91492	127	26	19
Kucing 13 cara memanggil setan	2011	5.6	Horror	R	81	97418	20	28	17
13: The Haunted	2018	5.1	Horror	13	107	842614	8800	601	884
22 Menit 3 hati dua dunia, satu cinta	2018	7.2	Drama	13	71	234178	1000	88	103
3 Nafas	2010	7.1	Romance	R	108	3142	14	1	1
Likas 3 pejalan tanggung	2014	7.6	Drama	13	105	78199	246	3	18
3 Pocong idiot	2010	4.6	drama	R	80	23823	9	1	0
3 Pocong idiot	2012	4.8	Comedy	R	81	57146	75	12	8
3 Srikandi	2016	6.6	Biography	SU	121	101503	300	11	23
3600 Detik	2014	6.8	Drama	R	90	975340	4400	142	222
5 cm	2012	7.2	Drama	R	126	986172	2800	160	294

Movie	Year	Ratings	Genre	Usia	Durasi (menit)	Views	Likes	Dislikes	Komenta r
9 Summers	2013	8.4	Drama	R	114	119453	354	7	62
10 Autumns 99 Cahaya di Langit	2013	7.3	Drama	R	105	436321	1200	72	125
Eropa A Copy of	2015	7.3	Drama	R	118	413596	392	66	20
My Mind A: Aku,	2017	7.5	Drama	13	94	289904 5	22000	309	1216
Benci & cinta About a Woman	2015	9	Drama	D	76	22857	70	3	25

Sumber : (Sulistyowati & Gata, 2018)

Berdasarkan tabel *dataset* diatas dapat diketahui bahwa atribut yang digunakan sebanyak 8 atribut. Terdiri dari 4 atribut didapat dari *website* IMBD seperti tahun tayang, *genre*, batas usia penonton dan durasi film, serta 4 atribut lainnya didapat dari media sosial youtube, yaitu jumlah *views*, *likes*, *dislikes* dan komentar dengan rating sebagai label populer dan tidak populer. Berikut atribut yang digunakan dalam penelitian ini :

Tabel 4.2 Tabel Atribut

No	Atribut	Tipe	Keterangan
1	Year	Text	2009
			2010
			2011
			2012
			2013
			2014
			2015
			2016
			2017
2	Genre	Polinomial	ACTION
			ADVENTURE
			ANIMATION
			BIOGRAPHY
			COMEDY
			DOCUMENTARY
			DRAMA
			FAMILY
			HISTORY
			HORROR
			MUSICAL
MYSTERY			

No	Atribut	Tipe	Keterangan
			ROMANCE
			SPORT
			THRILLER
3	Usia	Polinomial	R (REMAJA)
			D (DEWASA)
			SU (SEMUA UMUR)
4	Durasi	Binominal	PANJANG
			PENDEK
5	Views	Polinomial	<1JT
			1JT-5JT
			>5JT
6	Likes	Polinomial	< 30K
			30K-60K
			> 60K
7	Dislikes	Polinomial	< 500
			500-1000
			> 1000
8	Komentar	Polinomial	< 500
			500-1000
			> 1000
9	Rating	Binominal	POPULER
			TIDAK POPULER

Sumber : (Sulistyowati & Gata, 2018)

### 4.3.Data Preparation

Data perfilman yang telah dikumpulkan selanjutnya dilakukan proses *filtering* data sehingga menghasilkan data seperti tabel berikut :

Tabel 4.3. Dataset Film

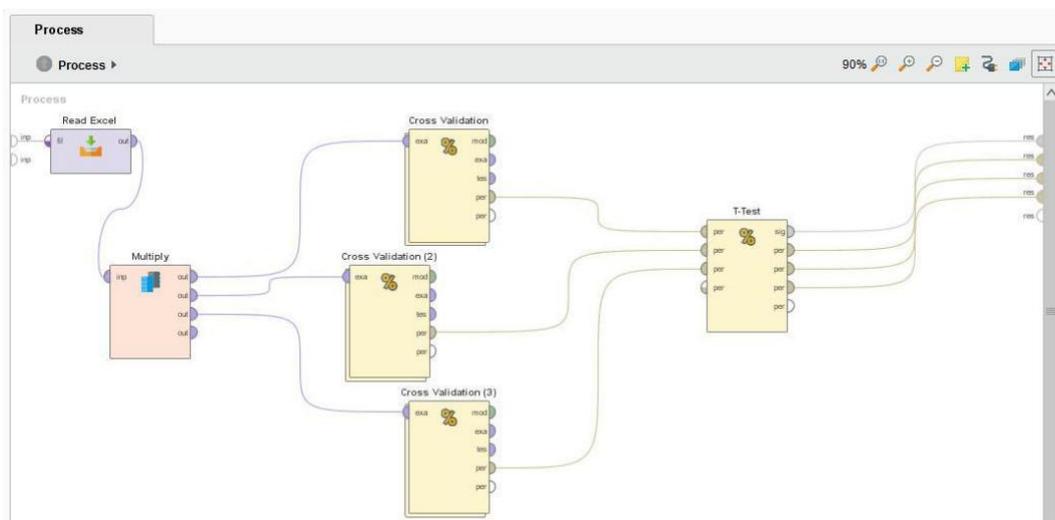
No	Movie	Year	Genre	Usia	Durasi	Views	Likes	Dislikes	Komentar	Rating
1	308	2013	Horror	R	PANJANG	<1JT	<30K	<500	<500	POPULER
2	#66	2016	Action	D	PANJANG	<1JT	<30K	<500	<500	POPULER
3	12 Menit: Kemenangan untuk Selamanya	2014	Drama	R	PANJANG	<1JT	<30K	<500	<500	POPULER
4	12:06 Rumah Kucing	2017	Horror	R	PANJANG	<1JT	<30K	<500	<500	TIDAK POPULER
5	13 cara memanggil setan	2011	Horror	R	PANJANG	<1JT	<30K	<500	<500	POPULER
6	13: The Haunted	2018	Horror	R	PANJANG	<1JT	<30K	500 sd 1000	500 sd 1000	POPULER
7	22 Menit	2018	Drama	R	PANJANG	<1JT	<30K	<500	<500	POPULER
8	3 hati dua dunia, satu cinta	2010	Romance	R	PANJANG	<1JT	<30K	<500	<500	POPULER

No	Movie	Year	Genre	Usia	Durasi	Views	Likes	Dislikes	Komentar	Rating
9	3 Nafas Likas	2014	Drama	R	PANJANG	<1JT	<30K	<500	<500	POPULER
10	3 pejalan tanggung	2010	Drama	R	PANJANG	<1JT	<30K	<500	<500	TIDAK POPULER
11	3 Pocong idiot	2012	Comedy	R	PANJANG	<1JT	<30K	<500	<500	TIDAK POPULER
12	3 Srikandi	2016	Biography	SU	PANJANG	<1JT	<30K	<500	<500	POPULER
13	3600 Detik	2014	Drama	R	PANJANG	<1JT	<30K	<500	<500	POPULER
14	5 cm	2012	Drama	R	PANJANG	<1JT	<30K	<500	<500	POPULER
15	9 Summers 10 Autumns	2013	Drama	R	PANJANG	<1JT	<30K	<500	<500	POPULER
16	99 Cahaya di Langit Eropa	2013	Drama	R	PANJANG	<1JT	<30K	<500	<500	POPULER
17	A Copy of My Mind	2015	Drama	R	PANJANG	<1JT	<30K	<500	<500	POPULER
18	A: Aku, Benci & cinta	2017	Drama	R	PANJANG	1JT sd 5JT	<30K	<500	>1000	POPULER
19	About a Woman	2015	Drama	D	PANJANG	<1JT	<30K	<500	<500	POPULER
20	Ada apa dengan pocong?	2011	Horror	R	PANJANG	<1JT	<30K	<500	<500	POPULER

Sumber : (Sulistyowati & Gata, 2018)

#### 4.4. Modeling

Tahap ini berisi eksperimen dan pengujian model dengan menghitung model algoritma yang digunakan. Model algoritma yang digunakan dalam penelitian ini berupa klasifikasi data mining yang terdiri dari algoritma *Decision Tree*, *Naive Bayes* dan *K-Nearest Neighbors*.

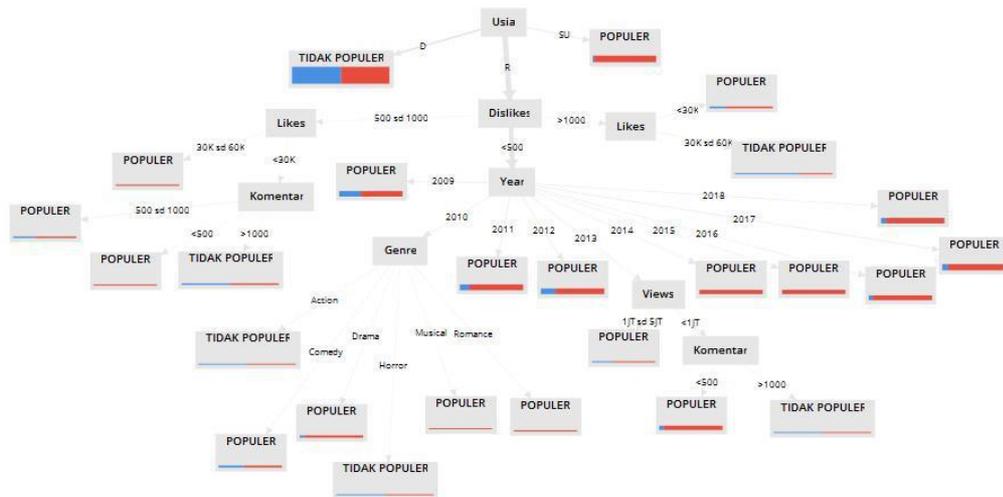


Gambar 4.1. Pengujian *Data Mining* Dengan *Rapid Miner*

Gambar diatas merupakan model algoritma prediksi dengan 1 *read excel* sebagai *operator* input data, *multiply* sebagai *parallel*, 3 operator *cross validation* dengan masing-masing algoritma data mining yang diujikan.

**4.2.1. Model Algoritma Decision Tree**

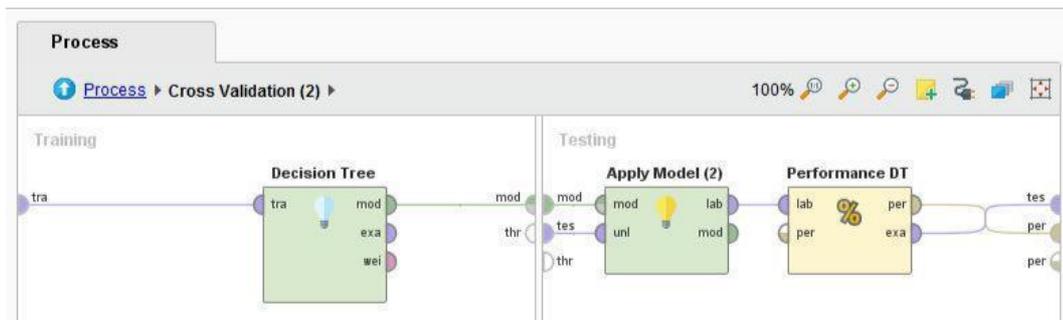
Pohon keputusan dibuat dengan menghitung jumlah *class* film yang populer dan tidak populer terlebih dahulu dari masing-masing *class* yang ada berdasarkan atribut yang digunakan. Jumlah data yang digunakan sebanyak 556 data film dengan 437 film yang populer dan 119 film yang tidak populer



Gambar 4.2. Tampilan *Decision Tree* yang Terbentuk

Gambar 4.2 merupakan bentuk dari pohon keputusan yang terbentuk dimana akar dari pohon keputusan tersebut adalah variabel usia. Tujuannya yaitu untuk mendapatkan rule yang akan digunakan dalam pengambilan keputusan pada kasus baru.

Berikut ini adalah tampilan hasil pengujian model algoritma *decisin tree* dengan menggunakan *software RapidMiner*.

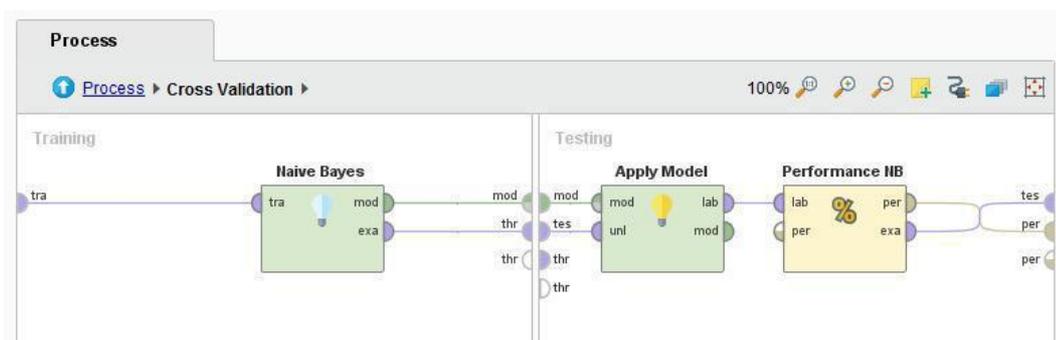


Gambar 4.3 Model *Validation* Algoritma *Decision Tree*

Model diatas merupakan operator algoritma yang terdapat didalam operator *validation* dimana *case* yang digunakan adalah model algoritma *decision tree*. Berdasarkan hasil pengujian diatas dapat diketahui bahwa nilai akurasi untuk algoritkan *decision tree* sebesar 76,45%.

#### 4.2.2. Model Algoritma *Naive Bayes*

Dataset film sebanyak 556 data film yang terdiri dari 437 data film populer dan 119 data film tidak populer digunakan sebagai masukan untuk klasifikasi algoritma *Naive Bayes Rapidminer*. Berikut ini adalah tampilan hasil pengujian model algoritma *naive bayes* dengan menggunakan *software RapidMiner*.

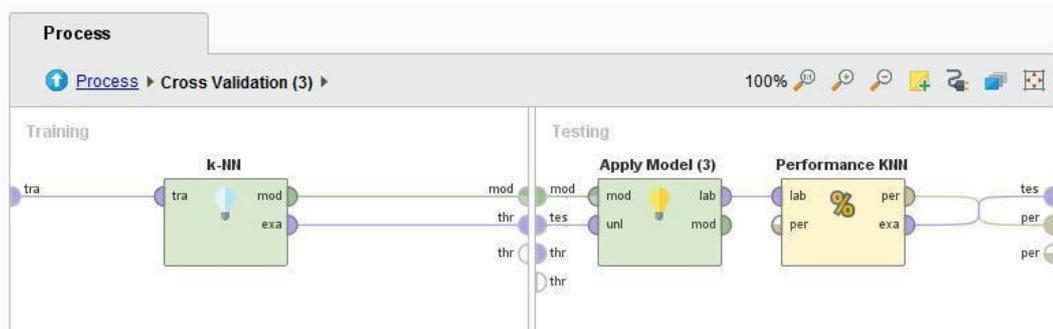


Gambar 4.4 Model *Validation* Algoritma *Naive Bayes*

Model diatas merupakan operator algoritma yang terdapat didalam operator *validation* dimana *case* yang digunakan adalah model *algoritma naive bayes*. Berdasarkan hasil pengujian diatas dapat diketahui bahwa nilai akurasi untuk algoritkan *naive bayes* sebesar 83,29%.

#### 4.2.3. Model Algoritma *K-Nearest Neighbors*

Dataset film sebanyak 556 data film yang terdiri dari 437 data film populer dan 119 data film tidak populer digunakan sebagai masukan untuk klasifikasi algoritma *K-Nearest Neighbors Rapidminer*. Berikut ini adalah tampilan hasil pengujian model algoritma *K-Nearest Neighbors* dengan menggunakan *software RapidMiner*.



Gambar 4.5 Model Validation Algoritma K-Nearest Neighbors

Model diatas merupakan operator algoritma yang terdapat didalam operator *validation* dimana *case* yang digunakan adalah model algoritma *k-nearest neighbors*. Berdasarkan hasil pengujian diatas dapat diketahui bahwa nilai akurasi untuk algoritman naive bayes sebesar 80,95%.

#### 4.5.Evaluasi dan Validasi Hasil

Metode yang digunakan pada penelitian ini yaitu algoritma klasifikasi *naive bayes*, *decision tree* dan *K-Nearest Neighbors* yang kemudian dilakukan evaluasi dengan AUC (*Area Under Curve*). Seperti yang telah dibahas pada bab sebelumnya, tujuan dari penelitian ini yaitu untuk memprediksi tingkat akurasi terhadap kepopularitas film berdasarkan variabel-variabel yang telah ditentukan.

Nilai akurasi model dengan menggunakan algoritma *decision tree* sebesar 76,45%, *naive bayes* sebesar 83,29% dan *K-Nearest Neighbors* sebesar 80,95%. Hasil dari pengujian model yang telah dilakukan adalah untuk mengukur tingkat akurasi dan AUC (*Area Under Curve*).

##### 4.5.1. Hasil Pengujian Model Algoritma Decision Tree

###### 1. Model Confusion Matrix

accuracy: 76.45% +/- 6.65% (micro average: 76.44%)

	true TIDAK POPULER	true POPULER	class precision
pred. TIDAK POPULER	41	49	45.56%
pred. POPULER	82	384	82.40%
class recall	33.33%	88.68%	

Gambar 4.6 Konversi Confusion Matrix Algoritma Decision Tree

Berdasarkan gambar diatas dari 556 data diketahui jumlah data yang sesuai dengan prediksi yang dikerjakan oleh algoritma *decision tree* terdapat 41 data diklasifikasi tidak populer, 49 data diprediksi tidak

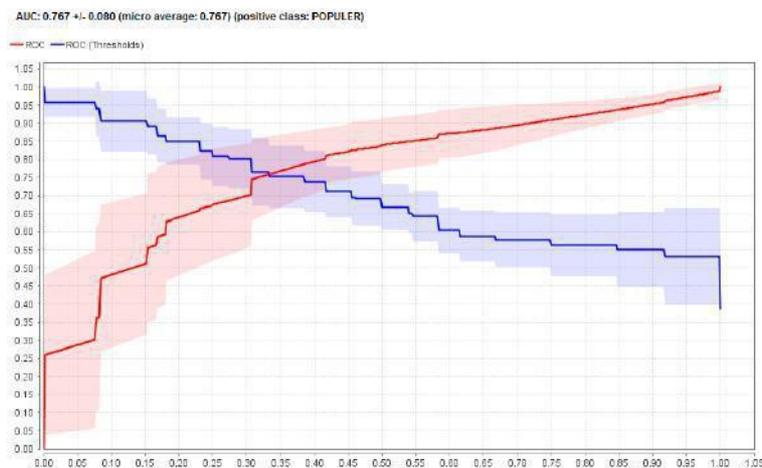
populer tetapi dalam *decision tree* dihasilkan populer, 82 data diprediksi populer tetapi oleh algoritma *decision tree* diklasifikasikan tidak populer dan 384 data diprediksi populer dan oleh *decision tree* diklasifikasikan sesuai. Prediksi yang dihasilkan dengan menggunakan algoritma *decision tree* adalah sebagai berikut :

Tabel 4.4. Hasil Prediksi Menggunakan Algoritma *Decision Tree*

NO	Movie	RATING	PREDICTION (RATING)	CONFIDENCE (TIDAK POPULER)	CONFIDENCE (POPULER)
1	Ai lop yu pul	TIDAK POPULER	POPULER	0,34375	0,65625
2	air terjun pengantin	TIDAK POPULER	TIDAK POPULER	0,50862069	0,49137931
3	anak setan	POPULER	TIDAK POPULER	0,612903226	0,387096774
4	asmara dua diana	POPULER	POPULER	0,361111111	0,638888889
5	benci disko	TIDAK POPULER	POPULER	0,333333333	0,666666667
6	bukan cinta biasa	POPULER	POPULER	0,34375	0,65625
7	bukan maling kundang	POPULER	POPULER	0,4	0,6
8	capres	POPULER	TIDAK POPULER	0,566666667	0,433333333
9	darah janda kolong wewe	TIDAK POPULER	TIDAK POPULER	0,504273504	0,495726496
10	Darah perawan bulan madu	TIDAK POPULER	TIDAK POPULER	0,50862069	0,49137931
11	dikejar setan	POPULER	POPULER	0,290322581	0,709677419
12	emak ingin naik haji	POPULER	POPULER	0,351351351	0,648648649
13	garuda di Dadaku	POPULER	POPULER	0	1
14	generasi biru	POPULER	TIDAK POPULER	0,50862069	0,49137931
15	get married 2	POPULER	POPULER	0,323529412	0,676470588
16	Glitch: Tersesat Dalam Waktu	TIDAK POPULER	POPULER	0,34375	0,65625
17	hantu biang kerok	TIDAK POPULER	TIDAK POPULER	0,504273504	0,495726496
18	Hantu binal jembatan semanggi	TIDAK POPULER	TIDAK POPULER	0,516666667	0,483333333
19	hantu jamu gendong	TIDAK POPULER	POPULER	0,495726496	0,504273504
20	hantu rumah ampera	TIDAK POPULER	TIDAK POPULER	0,566666667	0,433333333

Sumber: (Sulistyowati & Gata, 2018)

## 2. Model Kurva ROC

Gambar 4.7 Kurva ROC *Decision*

### Tree 4.5.2. Hasil Pengujian Model Algoritma Naive Bayes

#### 1. Model *Confusion Matrix*

accuracy: 83.29% +/- 3.60% (micro average: 83.27%)

	true TIDAK POPULER	true POPULER	class precision
pred. TIDAK POPULER	68	38	64.15%
pred. POPULER	55	395	87.78%
class recall	55.28%	91.22%	

Gambar 4.8 Konversi *Confusion Matrix* Algoritma Naive Bayes

Berdasarkan gambar diatas dari 556 data diketahui jumlah data yang sesuai dengan prediksi yang dikerjakan oleh algoritma *naive bayes* terdapat 68 data diklasifikasi tidak populer dan 395 data diprediksi populer. Kemudian 38 data diprediksi tidak populer tetapi dalam *naive bayes* dihasilkan populer serta 55 data diprediksi populer tetapi oleh algoritma *naive bayes* diklasifikasikan tidak populer. Prediksi yang dihasilkan dengan menggunakan algoritma *naive bayes* adalah sebagai berikut :

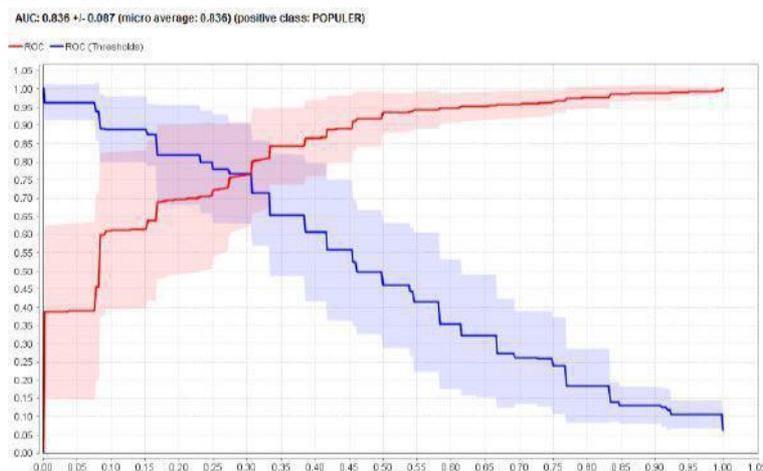
Tabel 4.5. Hasil Prediksi Menggunakan Algoritma Naive Bayes

NO	Movie	RATING	PREDICTION (RATING)	CONFIDENCE (TIDAK POPULER)	CONFIDENCE (POPULER)
1	Ai lop yu pul	TIDAK POPULER	POPULER	0,410849267	0,589150733
2	air terjun pengantin	TIDAK POPULER	TIDAK POPULER	0,925299954	0,074700046
3	anak setan	POPULER	TIDAK POPULER	0,936796019	0,063203981
4	asmara dua diana	POPULER	POPULER	0,149093644	0,850906356
5	benci disko	TIDAK POPULER	POPULER	0,382588082	0,617411918
6	bukan cinta biasa	POPULER	POPULER	1,90E-04	0,999809877
7	bukan maling kundang	POPULER	POPULER	0,440372031	0,559627969

NO	Movie	RATING	PREDICTION (RATING)	CONFIDENCE (TIDAK POPULER)	CONFIDENCE (POPULER)
8	capres	POPULER	POPULER	0,489211239	0,510788761
9	darah janda kolong wewe	TIDAK POPULER	TIDAK POPULER	0,930893557	0,069106443
10	Darah perawan bulan madu	TIDAK POPULER	TIDAK POPULER	0,925299954	0,074700046
11	dikejar setan	POPULER	TIDAK POPULER	0,672046262	0,327953738
12	emak ingin naik haji	POPULER	POPULER	0,168342562	0,831657438
13	garuda di Dadaku	POPULER	POPULER	0,027447951	0,972552049
14	generasi biru	POPULER	TIDAK POPULER	0,91143151	0,08856849
15	get married 2	POPULER	POPULER	0,396786504	0,603213496
16	Glitch: Tersesat Dalam Waktu	TIDAK POPULER	POPULER	0,363438327	0,636561673
17	hantu biang kerok	TIDAK POPULER	TIDAK POPULER	0,930893557	0,069106443
18	Hantu binal jembatan semanggi	TIDAK POPULER	TIDAK POPULER	0,943161297	0,056838703
19	hantu jamu gendong	TIDAK POPULER	TIDAK POPULER	0,92271979	0,07728021
20	hantu rumah ampera	TIDAK POPULER	TIDAK POPULER	0,925843289	0,074156711

Sumber: (Sulistyowati & Gata, 2018)

## 2. Model Kurva ROC



Gambar 4.9 Kurva ROC *Naive Bayes*

### 4.5.3. Hasil Pengujian Model *Algorithm K-Nearest Neighbors*

#### 1. Model *Confusion Matrix*

accuracy: 80.95% +/- 6.14% (micro average: 80.94%)

	true TIDAK POPULER	true POPULER	class precision
pred. TIDAK POPULER	65	48	57.52%
pred. POPULER	58	385	86.91%
class recall	52.85%	88.91%	

Gambar 4.10 Konversi *Confusion Matrix* Algoritma *K-Nearest Neighbors*

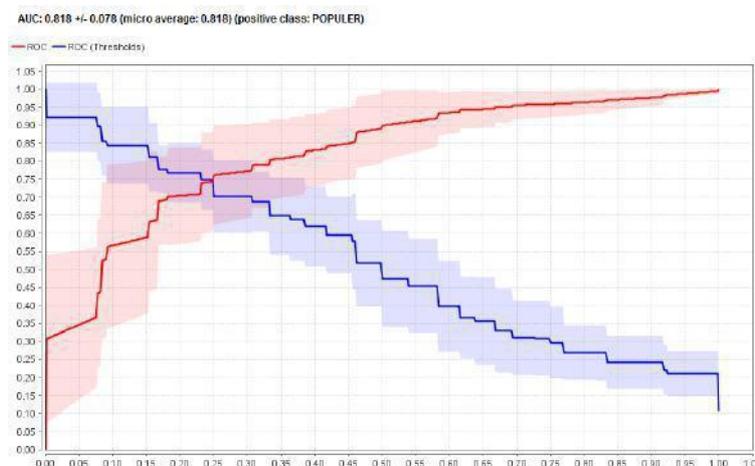
Gambar diatas menunjukkan dari 556 data diketahui jumlah data yang sesuai dengan prediksi yang dikerjakan oleh algoritma *k-nearest neighbors* terdapat 65 data diklasifikasi tidak populer dan 385 data diprediksi populer. Kemudian 65 data diprediksi tidak populer tetapi dalam *k-nearest neighbors* dihasilkan populer serta 48 data diprediksi populer tetapi oleh algoritma *k-nearest neighbors* diklasifikasikan tidak populer. Prediksi yang dihasilkan dengan menggunakan algoritma *k-nearest neighbors* adalah sebagai berikut :

Tabel 4.5. Hasil Prediksi Menggunakan Algoritma *K-Nearest Neighbors*

NO	Movie	RATING	PREDICTION (RATING)	CONFIDENCE (TIDAK POPULER)	CONFIDENCE (POPULER)
1	Ai lop yu pul	TIDAK POPULER	POPULER	0,4	0,6
2	air terjun pengantin	TIDAK POPULER	TIDAK POPULER	0,8	0,2
3	anak setan	POPULER	TIDAK POPULER	0,9	0,1
4	asmara dua diana	POPULER	POPULER	0,3	0,7
5	benci disko	TIDAK POPULER	POPULER	0,4	0,6
6	bukan cinta biasa	POPULER	POPULER	0,098765432	0,901234568
7	bukan maling kundang	POPULER	POPULER	0,444444444	0,555555556
8	capres	POPULER	POPULER	0,37037037	0,62962963
9	darah janda kolong wewe	TIDAK POPULER	TIDAK POPULER	0,8	0,2
10	Darah perawan bulan madu	TIDAK POPULER	TIDAK POPULER	0,8	0,2
11	dikejar setan	POPULER	TIDAK POPULER	0,507936508	0,492063492
12	emak ingin naik haji	POPULER	POPULER	0,3	0,7
13	garuda di Dadaku	POPULER	POPULER	0	1
14	generasi biru	POPULER	TIDAK POPULER	0,6	0,4
15	get married 2	POPULER	TIDAK POPULER	0,5	0,5
16	Glitch: Tersesat Dalam Waktu	TIDAK POPULER	POPULER	0,1	0,9
17	hantu biang kerok	TIDAK POPULER	TIDAK POPULER	0,8	0,2
18	Hantu binal jembatan semanggi	TIDAK POPULER	TIDAK POPULER	0,9	0,1
19	hantu jamu gendong	TIDAK POPULER	TIDAK POPULER	0,9	0,1
20	hantu rumah ampera	TIDAK POPULER	TIDAK POPULER	0,9	0,1

Sumber: (Sulistiyowati & Gata, 2018)

## 2. Model Kurva ROC



Gambar 4.11 Kurva ROC *K-Nearest*

### Neighbors 4.5.4. Analisa Hasil Pengujian

Hasil pengujian tiga algoritma klasifikasi data mining yang digunakan menunjukkan nilai sebagai berikut :

Tabel 4.6 Nilai Hasil Pengujian

	<b>Decision Tree</b>	<b>Naive Bayes</b>	<b><i>K-Nearest Neighbors</i></b>
<b><i>Accuracy</i></b>	76,45%	83,29%	80,95%
<b><i>Precision</i></b>	82,78%	87,83%	86,92%
<b><i>Recall</i></b>	88,68%	91,24%	88,90%

Hasil evaluasi dengan menggunakan ROC didapatkan melalui perbandingan hasil perhitungan nilai AUC untuk klasifikasi data mining yang digunakan. Hasil evaluasi dapat dilihat melalui tabel berikut :

Tabel 4.7 Hasil Evaluasi Nilai AUC

	<b>Decision Tree</b>	<b>Naive Bayes</b>	<b><i>K-Nearest Neighbors</i></b>
<b><i>AUC</i></b>	0,767	0,836	0,818

Hasil evaluasi nilai AUC menunjukkan nilai tertinggi menggunakan algoritma *naive bayes* sebesar 0,836 selanjutnya diposisi kedua yaitu algoritma *k-nearest neighbors* sebesar 0,818 dan nilai terendah yaitu pada algoritma sebesar

*decision tree* 0,767. Dari ketiga model yang digunakan dua diantaranya termasuk kedalam klasifikasi *Good classification* yaitu model algoritma *naive bayes* dan *k-nearest neighbors*, dikarenakan memiliki nilai AUC diantara 0,80 – 0,90. Sedangkan model algoritma *decision tree* termasuk kedalam klasifikasi *Fair classification*, memiliki nilai AUC diantara 0,70 - 0,80.

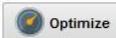
#### 4.6. Deployment

Penerapan GUI dalam penelitian ini menggunakan model simulator yang disediakan oleh Rapid Miner. Tujuan dilakukannya penerapan GUI untuk menguji model berdasarkan model algoritma yang terpilih dengan menggunakan data baru. Berikut ini merupakan tampilan hasil dari penerapan GUI :

**Input for Model**

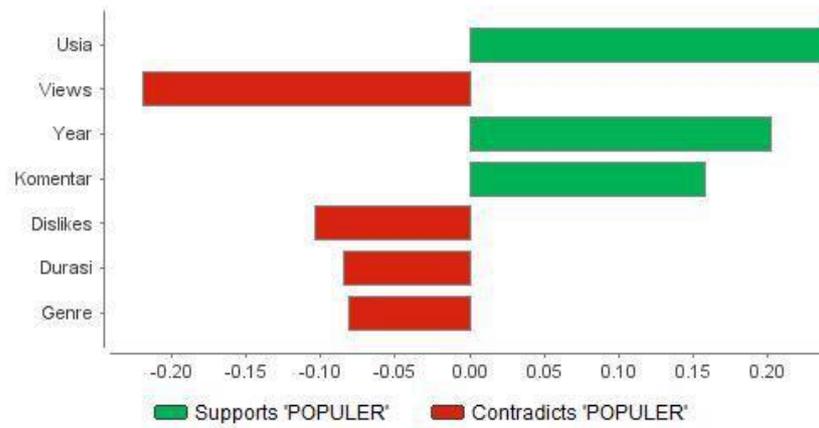
---

Year:	2018 ▼
Genre:	Thriller ▼
Usia:	D ▼
Durasi:	PANJANG ▼
Views:	<1JT ▼
Likes:	<30K ▼
Dislikes:	<500 ▼
Komentar:	500 sd 1000 ▼

 [What is this?](#)

Gambar 4.12 Tampilan Input Model Simulator

### Important Factors for POPULER



Gambar 4.13 Hasil Uji Coba Model Simulator

### Accuracy

78%

Sensitivity for POPULER: 99.77%

Precision for POPULER: 77.98%

Gambar 4.14 Hasil Akurasi Model Simulator

## **BAB V**

### **PENUTUP**

#### **5.1. Kesimpulan**

Berdasarkan penelitian yang telah dilakukan dengan menggunakan tiga klasifikasi data mining yaitu algoritma *Decision Tree*, *Naive Bayes* dan *K-Nearest Neighbors* yang kemudian dilakukan evaluasi dengan AUC (*Area Under Curve*) dengan menggunakan data perfilman Indonesia untuk memprediksi kepopuleran film. Nilai prediksi yang didapatkan setelah dilakukan perhitungan menggunakan tiga klasifikasi data mining, maka menghasilkan 454 judul film populer dan 102 judul film tidak populer dengan menggunakan algoritma *Decision Tree*, algoritma *Naive Bayes* menghasilkan 446 judul film populer dan 110 judul film tidak populer serta algoritma *K-Nearest Neighbors* menghasilkan 443 judul film populer dan 113 judul film tidak populer.

Nilai *Confusion Matrix* yang dihasilkan dengan menggunakan algoritma *decision tree* sebesar 76,45% nilai *accuracy* dengan *precision* sebesar 82,78% dan *recall* sebesar 88,68%, *naive bayes* sebesar 83,29% nilai *accuracy* dengan *precision* sebesar 87,83% dan *recall* sebesar 91,24% dan *K-Nearest Neighbors* sebesar 80,95% nilai *accuracy* dengan *precision* sebesar 86,92% dan *recall* sebesar 88,90%. Hasil evaluasi dengan *ROC curve* menunjukkan nilai AUC menggunakan model *naive bayes* sebesar 0,836 model *K-Nearest Neighbors* sebesar 0,818 dan model *decision tree* sebesar 0,767.

Sedangkan hasil untuk evaluasi menggunakan *ROC curve* untuk ketiga model menghasilkan nilai AUC tertinggi menggunakan algoritma *naive bayes* sebesar 0,836 selanjutnya diposisi kedua yaitu algoritma *k-nearest neighbors* sebesar 0,818 dan nilai terendah yaitu pada algoritma sebesar *decision tree* 0,767. Dari ketiga model yang digunakan dua diantaranya termasuk kedalam klasifikasi *Good classification* yaitu model algoritma *naive bayes* dan *k-nearest neighbors*, dikarenakan memiliki nilai AUC diantara 0,80 – 0,90. Sedangkan model algoritma *decision tree* termasuk kedalam klasifikasi *Fair classification*, memiliki nilai AUC diantara 0,70 - 0,80.

## 5.2.Saran

Data yang digunakan dalam penelitian ini merupakan data perfilman Indonesia selama 10 tahun terakhir sebanyak 556 judul film dengan 8 atribut variabel. Saran-saran yang dapat diusulkan penulis untuk penelitian selanjutnya, yaitu :

1. Untuk menghasilkan pengukuran yang lebih baik lagi disarankan menambahkan variabel baru yang lebih mempengaruhi dalam kepopuleran film.
2. Membangun sistem yang dapat digunakan dalam melakukan prediksi kepopuleran film-film yang akan tayang ataupun yang telah ditayangkan.
3. Melakukan perbandingan algoritma lain seperti *random forest*, *support vector machine* dan yang lainnya atau dapat melakukan optimasi algoritma.

## DAFTAR PUSTAKA

- Ahmed, M., Jahangir, M., Afzal, H., Majeed, A., & Siddiqi, I. (2015a). Using crowd-source based features from social media and conventional features to predict the movies popularity. *Proceedings - 2015 IEEE International Conference on Smart City, SmartCity 2015, Sustainable Computing and Communic*, (March 2016), 273–278.  
<https://doi.org/10.1109/SmartCity.2015.83>
- Alfisahrin, S. N. (2014). *Komparasi Algoritma C4.5, Naive Bayes dan Neural Network Untuk Memprediksi Penyakit Jantung*. Jakarta : Pascasarjana Magister Ilmu Komputer STMIK Nusa Mandiri.
- Apala, K. R., Jose, M., Motnam, S., Chan, C., Liszka, K. J., & Gregorio, F. De. (2013). Prediction of Movies Box Office Performance Using Social Media. *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, (July), 1209–1214.  
<https://doi.org/10.1145/2492517.2500232>
- Bellazzi, R., & Zupanb, B. (2008). Predictive Data Mining In Clinical Medicine: Current Issue And Guidelines. *International Journal Of Medical Informatics*, 81–97.
- Bhave, A., Kulkarni, H., Biramane, V., & Kosamkar, P. (2015). Role of different factors in predicting movie success. *2015 International Conference on Pervasive Computing (ICPC)*, 00(c), 1–4.  
<https://doi.org/10.1109/PERVASIVE.2015.7087152>
- Budi, S. (2017). Text Mining Untuk Analisis Sentimen Review Film Menggunakan Algoritma K-Means. *Techno.COM*, 16(1), 1–8.
- Chandani, V., Wahono, R. S., & Purwanto, . (2015). Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection pada Analisis Sentimen Review Film. *Journal of Intelligent Systems*, 1(1), 55–59. Retrieved from <http://journal.ilmukomputer.org/index.php/jis/article/view/10>
- Gorunescu, F. (2011). *Data Mining : Concepts, Models and Techniques*. Chennai: Springer.
- Gumilar, D., Pudjiantoro, T. H., & Yuniarti, R. (2017). Prediksi Kepopuleran Lagu Berdasarkan Tangga Lagu Billboard Menggunakan Decision Tree Dan K-Means. *Prosiding SNATIF Ke -4 Tahun 2017*, 187–192.  
<https://doi.org/10.1007/s12664-012-0191-3>
- Irfansyah, P. (2016). Kajian Komparasi Penerapan Algoritma Data Mining (C4.5, Bayesian Classifier Dan Neural Network) Dalam Menentukan Promosi Jabatan. *Prosiding Seminar Nasional*, 53–67.
- Iskandar, D., & Suprpto, Y. . (2015). Perbandingan Akurasi Klasifikasi Tingkat

- Kemiskinan Antara Algoritma C45 dan Naive Bayes. *Ilmiah NERO*, 2(1), 37–43.
- Jangid, B. M., Jadhav, C. K., Dhokate, S. M., Jadhav, G. M., & Bhandari, G. M. (2017). Movies Popularity Prediction Using Social Media and Conventional Features, (i), 5794–5798. <https://doi.org/10.15680/IJRSET.2017.0604192>
- Latif, M. H., & Afzal, H. (2016). Prediction of Movies popularity Using Machine Learning Techniques. *IJCSNS International Journal of Computer Science and Network Security*, 16(8), 127–131. Retrieved from [http://paper.ijcsns.org/07\\_book/201608/20160820.pdf](http://paper.ijcsns.org/07_book/201608/20160820.pdf)
- McLeon, R. G. P. S. (2008). *Sistem Informasi Manajemen. Salemba Empat*.
- Murnawan, & Sinaga, A. (2017). Implementasi Sentiment Analysis untuk Menentukan Tingkat Popularitas Tujuan Wisata. *Prosiding Seminar Nasional Teknologi Dan Rekayasa Informasi Tahun 2017*, (November), 12–18. Retrieved from <http://sentrin.filkom.ub.ac.id/event/sentrin17/jurnal/pimplementasi-sentiment-analysis-untuk-menentukan-tingkat-popularitas-tujuan-wisata-p/99e80f3>
- Nugroho, Y. S., & Emiliyawati, N. (2017). Sistem Klasifikasi Variabel Tingkat Penerimaan Konsumen Terhadap Mobil Menggunakan Metode Random Foresrt. *Jurnal Teknik Elektro*, 9 (1).
- Nugroho, Y. S., & Pratiwi, R. W. (2016). Prediksi Rating Film Menggunakan Metode Naïve Bayes. *Jurnal Teknik Elektro (ISSN 1411-0059)*, 8(2), 60–63.
- Nuraeni, N. (2017). Penentuan Calon Pegawai di PTPN 12 Kota Blater Tempurejo Jember dengan Metode Naive Bayes. *Seminar Nasional Teknologi Informasi Dan Multimetida*, 61–66.
- Nuraeni, N. (2017). Penentuan Kelayakan Kredit Dengan Algoritma Naïve Bayes Classifier : Studi Kasus Bank Mayapada Mitra Usaha Cabang PGC. *Jurnal Teknik Komputer AMIK BSI, III(1)*, 9–15. <https://doi.org/ISSN.2442-2436>
- Pristiyanti, R. I., Fauzi, M. A., & Muflikhah, L. (2017). Sentiment Analysis Peringkasan Review Film Menggunakan Metode Information Gain dan K-Nearest Neighbor, 2(3), 1179–1186.
- Ridwan, M., Suyono, H., & Sarosa, M. (2013). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Eeccis*, 7(1), 59–64. <https://doi.org/10.1038/hdy.2009.180>
- Utami, L. A. (2017). Melalui Komparasi Algoritma Support Vector Machine Dan K-Nearest Neighbor Berbasis Paerticle Swarm Optimization, 13((1)), 103–112.
- Vercillis, C. (2009). *Business Intelligence: Data Mining and Optimization for Decision Making*.

- Wisudawati, J., Adiwijaya, & Faraby, S. Al. (2017). Klasifikasi Sentimen pada Movie Review dengan Metode Multinomial Naïve Bayes. *E-Proceeding of Engineering*, 4(2), 2978–2988. Retrieved from <http://libraryproceeding.telkomuniversity.ac.id/index.php/engineering/article/view/1384>
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining Practical Machine Learning Tools And Techniques*. Burlington, Usa: Morgan Kaufmann Publishes.
- Yulietha, I. M., Faraby, S. Al, & Adiwijaya. (2017). Klasifikasi Sentimen Review Film Menggunakan Algoritma Support Vector Machine. *E-Proceeding of Engineering*, 4(3), 4740–4750. Retrieved from <http://libraryproceeding.telkomuniversity.ac.id/index.php/engineering/article/view/5401>

## DAFTAR RIWAYAT HIDUP

### I. Biodata Mahasiswa

NIM : 14002092  
Nama Lengkap : Daning Nur Sulistyowati  
Tempat & Tanggal Lahir : Tangerang, 21 Desember 1993  
Jenis Kelamin : Perempuan  
Agama : Islam  
Alamat lengkap : Jl. Nusa Inda II RT/RW 002/007  
Kel.Curug Kec.Bojongsari Kota Depok  
No HP : 081213488757

### II. Pendidikan

1. SDN Priyang 1, lulus tahun 2006
2. SMPN 1 Kota Tangerang Selatan , lulus tahun 2009
3. SMKN 1 Kota Tangerang Selatan, lulus tahun 2012
4. Diploma Tiga (DIII) AMIK BSI, lulus tahun 2015
5. Strata 1 (S1) STMIK Nusa Mandiri, lulus tahun 2017

---

Jakarta, 07 Januari 2019

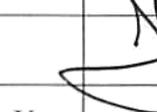
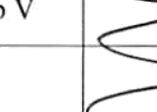
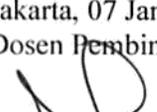
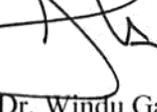
Foto

3 x 4

Daning Nur Sulistyowati

	<b>LEMBAR KONSULTASI BIMBINGAN TESIS</b>
	<b>Pascasarjana Magister Ilmu Komputer STMIK Nusa Mandiri</b>

Nama : Daning Nur Sulistyowati  
 NIM : 14002092  
 Dosen Pembimbing : Dr. Windu Gata, M.Kom  
 Judul Tesis : Klasifikasi Algoritma *Decision Tree*, *Algoritma Naïve Bayes* dan *K-Nearest Neighbors* Terhadap Popularitas Perfilman Indonesia

No	Tanggal Bimbingan	Materi Bimbingan	Paraf Dosen Pembimbing
1.	12-10-2018	Pengajuan Judul	
2.	19-10-2018	Revisi Judul dan Pengajuan Bab I	
3.	02-11-2018	Revisi Bab I dan Pengajuan Bab II	
4.	16-11-2018	Revisi Bab II	
5.	24-11-2018	Revisi Bab II dan Pengajuan Bab III	
6.	02-12-2018	Revisi Bab III dan Pengajuan Bab IV	
7.	07-12-2018	Revisi Bab IV	
8.	14-12-2018	Revisi Bab IV dan Pengajuan Bab V	
9.	21-12-2018	Acc Keseluruhan	
10.	24-12-2018	Acc Revisi	

Bimbingan dimulai pada tanggal : 12 Oktober 2018  
 Bimbingan diakhiri pada tanggal : 24 Desember 2018  
 Jumlah pertemuan : 10 kali

Jakarta, 07 Januari 2019  
 Dosen Pembimbing

  
 (Dr. Windu Gata, M.Kom)

## LAMPIRAN DATASET FILM INDONESIA

No	Movie	Year	Genre	Usia	Durasi	Views	Likes	Dislikes	Komentar	Rating
1	12:06 Rumah Kucing	2017	Horror	R	PANJANG	<1JT	<30K	<500	<500	TIDAK POPULER
2	A: Aku, Benci & cinta	2017	Drama	R	PANJANG	1JT sd 5JT	<30K	<500	>1000	POPULER
3	After School Horror 2	2017	Horror	R	PANJANG	<1JT	<30K	<500	<500	POPULER
4	Ayat-Ayat Cinta 2	2017	Romance	R	PANJANG	>5JT	30K sd 60K	500 sd 1000	>1000	POPULER
5	Baracas: Barisan Anti Cinta Asmara	2017	Drama	R	PANJANG	<1JT	<30K	<500	<500	POPULER
6	Bukaan 8	2017	Comedy	R	PANJANG	<1JT	<30K	<500	<500	POPULER
7	Chrisye	2017	Drama	R	PANJANG	1JT sd 5JT	<30K	<500	>1000	POPULER
8	Critical Eleven	2017	Romance	R	PANJANG	1JT sd 5JT	<30K	<500	<500	POPULER
9	Danur: I Can See Ghosts	2017	Horror	R	PANJANG	>5JT	<30K	>1000	>1000	POPULER
10	Dear Nathan	2017	Romance	R	PANJANG	1JT sd 5JT	30K sd 60K	<500	>1000	POPULER
11	Duka Sedalam Cinta	2017	Drama	R	PANJANG	<1JT	<30K	<500	<500	POPULER
12	Filosofi Kopi 2 : Ben & Jody	2017	Drama	R	PANJANG	<1JT	<30K	<500	<500	POPULER
13	Galih dan Ratna	2017	Drama	R	PANJANG	1JT sd 5JT	<30K	<500	<500	POPULER
14	Gasing Tengkorak	2017	Horror	R	PANJANG	1JT sd 5JT	<30K	<500	<500	TIDAK POPULER
15	Gerbang Neraka	2017	Horror	R	PANJANG	<1JT	<30K	<500	<500	POPULER
16	Hujan Bulan Juni	2017	Drama	R	PANJANG	1JT sd 5JT	<30K	<500	<500	POPULER
17	Insyallah Sah	2017	Comedy	R	PANJANG	1JT sd	<30K	500 sd	<500	POPULER

No	Movie	Year	Genre	Usia	Durasi	Views	Likes	Dislikes	Komenta r	Rating
						5JT		1000		
18	Jailangkung	2017	Horror	R	PANJANG	1JT sd 5JT	<30K	500 sd 1000	>1000	TIDAK POPULER
19	Jomblo	2017	Comedy	R	PANJANG	<1JT	<30K	<500	<500	POPULER
20	KAILI: Karena Aku Ingin Kembali	2017	Drama	R	PANJANG	<1JT	<30K	<500	<500	POPULER
21	Kartini	2017	Biography	SU	PANJANG	1JT sd 5JT	<30K	<500	<500	POPULER
22	Keluarga Tak Kasat Mata	2017	Horror	R	PANJANG	<1JT	<30K	<500	<500	TIDAK POPULER
23	Knight Kris	2017	Animatio n	SU	PANJANG	<1JT	<30K	<500	<500	POPULER
24	Labuan Hati	2017	Romance	R	PANJANG	<1JT	<30K	<500	<500	POPULER
25	Mantan	2017	Romance	R	PANJANG	<1JT	<30K	<500	<500	POPULER
26	Marlina the Murderer in Four Acts	2017	Thriller	D	PANJANG	<1JT	<30K	<500	<500	POPULER
27	Mata Batin	2017	Horror	R	PANJANG	1JT sd 5JT	<30K	500 sd 1000	500 sd 1000	POPULER
28	Merah Putih Memanggil	2017	Action	R	PANJANG	<1JT	<30K	<500	<500	POPULER
29	Mereka Yang Tak Terlihat	2017	Horror	R	PANJANG	1JT sd 5JT	<30K	<500	<500	POPULER
30	Moammar Emka's Jakarta Undercover	2017	Drama	D	PANJANG	<1JT	<30K	<500	<500	POPULER
31	Mooncake Story	2017	Drama	R	PANJANG	<1JT	<30K	<500	<500	POPULER
32	My Generation (II)	2017	Drama	R	PANJANG	<1JT	<30K	<500	<500	POPULER
33	Night Bus	2017	Thriller	R	PANJANG	<1JT	<30K	<500	<500	POPULER
34	One Fine Day	2017	Drama	R	PANJANG	1JT sd 5JT	<30K	<500	>1000	POPULER
35	Pengabdian Setan	2017	Horror	R	PANJANG	>5JT	<30K	>1000	>1000	POPULER

No	Movie	Year	Genre	Usia	Durasi	Views	Likes	Dislikes	Komenta r	Rating
36	Perfect dream (I)	2017	Drama	R	PANJANG	<1JT	<30K	<500	<500	POPULER
37	Pertaruhan	2017	Action	R	PANJANG	<1JT	<30K	<500	<500	POPULER
38	Petak Umpet Minako	2017	Horror	R	PANJANG	1JT sd 5JT	<30K	<500	500 sd 1000	POPULER
39	Posesif (I)	2017	Romance	R	PANJANG	1JT sd 5JT	<30K	<500	500 sd 1000	POPULER
40	Rafathar	2017	Adventure	R	PANJANG	1JT sd 5JT	30K sd 60K	>1000	>1000	TIDAK POPULER
41	Ruqyah: The Exorcism	2017	Horror	R	PANJANG	1JT sd 5JT	<30K	<500	500 sd 1000	TIDAK POPULER
42	Satu Hari Nanti	2017	Drama	D	PANJANG	<1JT	<30K	<500	<500	POPULER
43	Sekala Niskala	2017	Drama	R	PANJANG	<1JT	<30K	<500	<500	POPULER
44	Silariang the Movie	2017	Drama	R	PANJANG	<1JT	<30K	<500	<500	POPULER
45	Stip & Pensil	2017	Comedy	R	PANJANG	1JT sd 5JT	<30K	<500	<500	POPULER
46	Suhu Beku: The Movie	2017	Drama	R	PANJANG	<1JT	<30K	<500	<500	POPULER
47	Surat Cinta untuk Starla the Movie	2017	Drama	R	PANJANG	1JT sd 5JT	30K sd 60K	500 sd 1000	>1000	POPULER
48	Surat Kecil Untuk Tuhan	2017	Drama	R	PANJANG	1JT sd 5JT	<30K	<500	500 sd 1000	POPULER
49	Surau dan Silek	2017	Family	SU	PANJANG	<1JT	<30K	<500	<500	POPULER
50	Surga Yang Tak Dirindukan 2	2017	Drama	R	PANJANG	1JT sd 5JT	<30K	<500	<500	POPULER
51	Susah Sinyal	2017	Comedy	R	PANJANG	1JT sd 5JT	<30K	500 sd 1000	>1000	POPULER
52	Sweet 20	2017	Drama	R	PANJANG	1JT sd 5JT	<30K	<500	500 sd 1000	POPULER
53	The Curse (I)	2017	Horror	R	PANJANG	<1JT	<30K	<500	<500	POPULER
54	The Doll 2	2017	Horror	R	PANJANG	>5JT	<30K	>1000	>1000	POPULER

No	Movie	Year	Genre	Usia	Durasi	Views	Likes	Dislikes	Komenta r	Rating
55	The Gift (III)	2017	Drama	R	PANJANG	>5JT	<30K	<500	500 sd 1000	POPULER
56	The Guys	2017	Comedy	R	PANJANG	1JT sd 5JT	<30K	500 sd 1000	>1000	POPULER
57	The Promise	2017	Horror	D	PANJANG	<1JT	<30K	<500	<500	POPULER
58	The Underdogs	2017	Comedy	R	PANJANG	1JT sd 5JT	<30K	<500	<500	POPULER
59	Trinity, the Nekad Traveler	2017	Drama	R	PANJANG	<1JT	<30K	<500	<500	POPULER
60	Ular Tangga	2017	Horror	R	PANJANG	<1JT	<30K	<500	<500	POPULER
61	Valentine	2017	Action	R	PANJANG	<1JT	<30K	<500	<500	POPULER
62	Wage	2017	Biography	SU	PANJANG	<1JT	<30K	<500	<500	POPULER
63	Warkop DKI Reborn: Jangkrik Boss Part 2	2017	Comedy	R	PANJANG	1JT sd 5JT	<30K	500 sd 1000	>1000	POPULER
64	13: The Haunted	2018	Horror	R	PANJANG	<1JT	<30K	500 sd 1000	500 sd 1000	POPULER
65	22 Menit	2018	Drama	R	PANJANG	<1JT	<30K	<500	<500	POPULER
66	Aib #Cyberbully	2018	Horror	D	PANJANG	<1JT	<30K	<500	>1000	TIDAK POPULER
67	Alas Pati: Hutan Mati	2018	Horror	R	PANJANG	1JT sd 5JT	<30K	<500	<500	POPULER
68	Ananta	2018	Romance	R	PANJANG	1JT sd 5JT	<30K	<500	500 sd 1000	POPULER
69	Arini: Masih Ada Kereta yang Akan Lewat	2018	Drama	R	PANJANG	>5JT	<30K	<500	500 sd 1000	POPULER
70	Aruna & Lidahnya	2018	Drama	R	PANJANG	1JT sd 5JT	<30K	<500	<500	POPULER
71	Arwah Tumbal Nyai the Trilogy: part Arwah	2018	Horror	R	PANJANG	1JT sd 5JT	<30K	>1000	>1000	TIDAK POPULER
72	Asih	2018	Horror	R	PANJANG	1JT sd 5JT	<30K	<500	>1000	POPULER

No	Movie	Year	Genre	Usia	Durasi	Views	Likes	Dislikes	Komenta r	Rating
73	Belok Kanan Barcelona	2018	Drama	R	PANJANG	1JT sd 5JT	<30K	<500	<500	POPULER
74	Benyamin Biang Kerok	2018	Comedy	R	PANJANG	<1JT	<30K	>1000	>1000	TIDAK POPULER
75	Bisikan Iblis	2018	Horror	R	PANJANG	1JT sd 5JT	<30K	<500	<500	TIDAK POPULER
76	Bodyguard Ugal-Ugalan	2018	Comedy	R	PANJANG	<1JT	<30K	500 sd 1000	<500	POPULER
77	Buffalo Boys	2018	Action	R	PANJANG	<1JT	<30K	<500	>1000	POPULER
78	Danur 2: Maddah	2018	Horror	R	PANJANG	>5JT	<30K	>1000	>1000	POPULER
79	Dimsum Martabak	2018	Drama	R	PANJANG	1JT sd 5JT	30K sd 60K	>1000	>1000	TIDAK POPULER
80	DOA (Doyok-Otoy-Ali Oncom) : Cari Jodoh	2018	Comedy	R	PANJANG	1JT sd 5JT	<30K	<500	500 sd 1000	POPULER
81	Eiffel... I'm In Love 2	2018	Drama	R	PANJANG	1JT sd 5JT	<30K	<500	>1000	POPULER
82	Gentayangan	2018	Horror	R	PANJANG	1JT sd 5JT	<30K	<500	<500	TIDAK POPULER
83	Gila Lu Ndro	2018	Comedy	R	PANJANG	<1JT	<30K	500 sd 1000	500 sd 1000	TIDAK POPULER
84	Insya Allah Sah 2	2018	Comedy	R	PANJANG	1JT sd 5JT	<30K	<500	<500	POPULER
85	Jailangkung 2	2018	Horror	R	PANJANG	>5JT	<30K	500 sd 1000	500 sd 1000	TIDAK POPULER
86	Jelita Sejuba: Mencintai Kesatria Negara	2018	Drama	R	PANJANG	1JT sd 5JT	<30K	<500	<500	POPULER
87	Kafir: Bersekutu dengan Setan	2018	Horror	R	PANJANG	1JT sd 5JT	<30K	<500	500 sd 1000	POPULER
88	Kenapa Harus Bule?	2018	Drama	D	PANJANG	<1JT	<30K	<500	<500	POPULER
89	Koki-Koki Cilik	2018	Drama	SU	PANJANG	1JT sd	<30K	<500	<500	POPULER

No	Movie	Year	Genre	Usia	Durasi	Views	Likes	Dislikes	Komenta r	Rating
						5JT				
90	Kulari ke Pantai	2018	Family	SU	PANJANG	1JT sd 5JT	<30K	<500	<500	POPULER
91	Kuntilanak	2018	Horror	R	PANJANG	1JT sd 5JT	<30K	500 sd 1000	>1000	TIDAK POPULER
92	Lima	2018	Drama	R	PANJANG	<1JT	<30K	<500	<500	POPULER
93	Love for Sale	2018	Drama	D	PANJANG	1JT sd 5JT	<30K	<500	500 sd 1000	POPULER
94	Meet Me After Sunset	2018	Drama	R	PANJANG	<1JT	<30K	<500	<500	POPULER
95	Moonrise Over Egypt	2018	History	R	PANJANG	<1JT	<30K	<500	<500	POPULER
96	Pai Kau	2018	Drama	R	PANJANG	<1JT	<30K	<500	<500	POPULER
97	Partikelir	2018	Action	R	PANJANG	1JT sd 5JT	<30K	<500	<500	POPULER
98	Petualangan Menangkap Petir	2018	Family	SU	PANJANG	<1JT	<30K	<500	<500	POPULER
99	Rasuk	2018	Horror	R	PANJANG	1JT sd 5JT	<30K	<500	500 sd 1000	POPULER
100	Reuni Z	2018	Horror	R	PANJANG	1JT sd 5JT	<30K	<500	>1000	POPULER
101	Revan & Reina	2018	Drama	R	PANJANG	<1JT	<30K	<500	<500	POPULER
102	Rompis	2018	Drama	R	PANJANG	1JT sd 5JT	<30K	<500	500 sd 1000	POPULER
103	R-Raja, Ratu & Rahasia	2018	Drama	R	PANJANG	1JT sd 5JT	<30K	<500	500 sd 1000	POPULER
104	Rumah Belanda	2018	Horror	R	PANJANG	<1JT	<30K	<500	<500	TIDAK POPULER
105	Sabrina	2018	Horror	R	PANJANG	>5JT	30K sd 60K	>1000	>1000	POPULER
106	Sajen	2018	Horror	R	PANJANG	1JT sd 5JT	<30K	500 sd 1000	500 sd 1000	POPULER

No	Movie	Year	Genre	Usia	Durasi	Views	Likes	Dislikes	Komenta r	Rating
107	Sakral	2018	Horror	R	PANJANG	1JT sd 5JT	<30K	<500	<500	POPULER
108	Sebelum Iblis Menjemput	2018	Horror	R	PANJANG	1JT sd 5JT	<30K	<500	>1000	POPULER
109	Serendipity	2018	Drama	R	PANJANG	<1JT	<30K	<500	<500	POPULER
110	Sesat: Yang Sudah Pergi Jangan Dipanggil Lagi	2018	Horror	R	PANJANG	1JT sd 5JT	<30K	<500	500 sd 1000	POPULER
111	Si Doel the Movie	2018	Drama	R	PANJANG	1JT sd 5JT	30K sd 60K	500 sd 1000	>1000	POPULER
112	Some in Between	2018	Drama	R	PANJANG	1JT sd 5JT	<30K	<500	500 sd 1000	POPULER
113	Sultan Agung: Tahta, Perjuangan, Cinta	2018	Action	R	PANJANG	<1JT	<30K	<500	<500	POPULER
114	Target (II)	2018	Thriller	R	PANJANG	<1JT	<30K	<500	500 sd 1000	POPULER
115	Teman Tapi Menikah	2018	Drama	R	PANJANG	1JT sd 5JT	<30K	>1000	>1000	POPULER
116	Terbang: Menembus Langit	2018	Drama	R	PANJANG	<1JT	<30K	<500	500 sd 1000	POPULER
117	The Origin of Santet	2018	Horror	R	PANJANG	<1JT	<30K	<500	<500	POPULER
118	The Perfect Husband	2018	Drama	R	PANJANG	1JT sd 5JT	<30K	<500	500 sd 1000	POPULER
119	Wiro Sableng: Pendekar Kapak Maut Naga Geni 212	2018	Comedy	R	PANJANG	1JT sd 5JT	<30K	<500	>1000	POPULER
120	Yowis Ben	2018	Comedy	R	PANJANG	1JT sd 5JT	<30K	<500	>1000	POPULER