

## PENERAPAN ADABOOST UNTUK MENINGKATKAN AKURASI NAIVE BAYES PADA PREDIKSI PENDAPATAN PENJUALAN FILM

Dini Nurlaela

Sistem Informasi  
Universitas Bina Sarana Informatika  
<http://www.bsi.ac.id>  
[dini.dur@bsi.ac.id](mailto:dini.dur@bsi.ac.id)



**Abstract**—For economists and financial experts predicting the success of doing business is very interesting. With the data analytics the prediction process has been facilitated by the past data stored to find out what will happen in the future. This research was conducted to facilitate the film industry players in considering the factors that can influence the income of the film to be produced. The naive bayes method is a popular machine learning technique for classification because it is very simple, efficient, and has good performance on many domains. But naive bayes has a disadvantage that is very sensitive to too many features, thus making the accuracy to be low, in this case the adaboost method to reduce bias so that it can and improve accuracy from naive bayes. Validation is done by using 10 fold cross validation while measuring accuracy using confusion matrix and kappa. The results showed an increase in the accuracy of Naive Bayes from 83.22% to 84.44% and the kappa value from 0.706 to 0.731. So that it can be concluded that the application of adaboost on 2014 & 2015 CSM film data is able to improve the accuracy of the Naive Bayes algorithm

**Keywords:** Adaboost, Naive Bayes, Predictions, Revenue, Films

**Abstrak**—Bagi para ekonom dan ahli keuangan prediksi dalam keberhasilan berbisnis merupakan hal sangat menarik. Dengan adanya data analytics proses prediksi sudah dimudahkan dengan adanya data masa lalu yang tersimpan untuk mengetahui bagaimana yang akan terjadi di masa depan. Penelitian ini dilakukan untuk mempermudah para pelaku industri film dalam mempertimbangkan faktor yang dapat mempengaruhi pendapatan film yang akan di produksi. Metode naive bayes adalah tehnik machine learning yang populer untuk klasifikasi karena karena sangat sederhana, efisien, dan memiliki performa yang baik pada banyak domain. Namun naive bayes memiliki kekurangan yaitu sangat sensitif pada fitur yang terlalu banyak, sehingga membuat akurasi menjadi rendah, pada hal ini metode adaboost untuk mengurangi bias agar dapat dan meningkatkan akurasi dari naive bayes. Validasi dilakukan dengan menggunakan 10 fold cross validation sedangkan pengukuran akurasinya menggunakan confusion matrix dan kappa. Hasil menunjukkan peningkatan akurasi Naive Bayes dari 83.22% menjadi 84.44% dan nilai kappa dari 0.706 menjadi 0.731. Sehingga dapat disimpulkan bahwa penerapan adaboost pada data film 2014 & 2015 CSM mampu meningkatkan akurasi algoritma Naive Bayes.

**Kata kunci:** Adaboost, Naive Bayes, Prediksi, Pendapatan, Film.

### PENDAHULUAN

Prediksi dalam keberhasilan berbisnis sangat menarik bagi para ekonom dan ahli keuangan. Dengan adanya *data analytics* proses prediksi sudah dimudahkan dengan adanya data masa lalu yang tersimpan untuk mengetahui bagaimana yang akan terjadi di masa depan. Penelitian ini telah dilakukan dalam memprediksi kesuksesan dan popularitas film yang diukur oleh *rating* (biasanya diwakili oleh angka 0-10) dan *gross* (Ahmed, Jahangir, Afzal, Majeed, & Siddiqi, 2015). Demikian juga dalam industri film ini

menunjukkan bahwa banyaknya opini berupa *text* saja tidak cukup mewakili sebuah film dapat dikatakan sukses dan populer.

Popularitas internet mendorong orang untuk mencari pendapat pengguna dari internet sebelum membeli produk atau melihat situs film (Tsou & Ma, 2011). Banyak situs yang menyediakan *review* tentang suatu produk yang dapat mencerminkan pendapat pengguna (Tsou & Ma, 2011). Salah satu contohnya adalah situs *Internet Movie Database* (IMDb). IMDb adalah situs yang berhubungan dengan film dan produksi film. Informasi yang diberikan IMDb sangat lengkap.

Siapa saja aktor/aktris yang main di film itu, sinopsis singkat dari film, link untuk trailer film, tanggal rilis untuk beberapa negara dan review dari user-user yang lain. Ketika seseorang ingin membeli atau menonton suatu film, komentar-komentar orang lain dan peringkat film biasanya mempengaruhi perilaku pembelian mereka.

Dari IMDb dapat dikembangkan untuk pencarian data berupa atribut jumlah *like*, *dislike*, serta jumlah komentar (yang diambil dari *trailer* film resmi) serta mengetahui jumlah *follower* 3 aktor/aktris paling terkenal yang terlibat di dalam film tersebut (Ahmed et al., 2015), yang dapat lebih banyak mempengaruhi para penikmat film tertarik untuk menonton film yang memiliki lebih banyak nilai dari atribut tersebut.

Adapun salah satu prediksi yang dilakukan sebelumnya dengan memprediksi rating film menggunakan *naive bayes* dengan melibatkan atribut *genre* film, aktor film, bahasa, warna, durasi film, negara dan lainnya yang digunakan untuk tolak ukur sutradara untuk membuat film memiliki akurasi 55.80% , *precision* 32.41% dan *recall* 46.70% , dalam penelitian tersebut dihasilkan model bahwa rating film akan tinggi jika tahun filmnya baru dan durasinya panjang (Nugroho & Pratiwi, 2016). Prediksi tersebut tidak cukup untuk memotivasi para produser film untuk yakin bahwa mereka akan mendapatkan pendapatan yang baik.

Tujuan dari penelitian ini adalah untuk mengintegrasikan metode *adaboost* dan *naive bayes* dalam memprediksi film dari kumpulan data film yang memiliki kesuksesan dan popularitas.

1. Mengetahui algoritma klasifikasi apa yang paling akurat untuk untuk menangani masalah atribut yang kurang relevan pada prediksi kategori pendapatan film.
2. Mengetahui algoritma *adaboost* apakah memiliki performanya paling baik untuk menangani masalah atribut yang kurang relevan pada prediksi film.
3. Mengetahui seberapa meningkatnya *accuracy* algoritma *naive bayes* terbaik dan metode *adaboost* terbaik jika diintegrasikan pada prediksi film.
4. Memberikan kontribusi pengetahuan terhadap industri perfilman yang digarap berorientasi komersil agar tidak salah menentukan konsep film sehingga mengakibatkan kerugian.

Hipotesis pada penelitian ini adalah akan adanya peningkatan akurasi yang baik dalam penggunaan *naive bayes* dalam memprediksi data film dengan optimasi metode *adaboost*.

Algoritma klasifikasi machine learning sangat banyak. Oleh karena itu, penulis membatasi algoritma yang digunakan hanya *naive bayes* dan *adaboost* dengan menggunakan dataset yang sudah didapatkan di *UCI repository* sebagai *data training*.

IMDb review film, Youtube, Twitter, dan Wikipedia sebagai media penggalan data secara konvensional untuk mendapatkan *data testing*.

## BAHAN DAN METODE

Penelitian ini meliputi 5 tahap yang terdiri dari tahap pengumpulan data, pengolahan data awal, pembuatan desain model, pengujian model dan validasi serta evaluasi.

### Pengumpulan Data

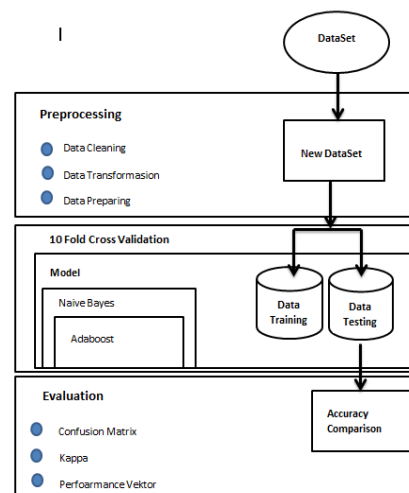
Data yang digunakan pada penelitian ini adalah delapan *dataset* uji dari *University of California Irvine (UCI) machine learning repository* yang diperoleh melalui situs <http://archive.ics.uci.edu/ml>.

### Pengolahan Data Awal

*Dataset* dari *UCI machine learning repository* tersebut telah dilakukan transformasi data ke dalam bentuk yang dibutuhkan untuk metode yang diusulkan. Dalam penelitian ini akan didapatkan nilai akurasi yang dihasilkan dari proses klasifikasi dengan menggunakan metode yang diusulkan, dan selanjutnya akan dilakukan perbandingan antara penggunaan *Naive bayes* dengan *Naive bayes* dan *Adaboost* untuk mendapatkan nilai akurasi yang terbaik.

### Metode yang diusulkan

Setelah data awal diolah dan ditransformasikan, selanjutnya akan dibuat desain sebuah model dengan melakukan pengaturan pemodelan untuk mengoptimalkan hasil. Model yang diusulkan dalam hal ini adalah *esemble* model *naive bayes* dengan penggabungan antara *naive bayes* dan *boosting (adaboost)*.



Sumber : (Nurlaela, 2020)

Gambar 1. Metode yang Diusulkan

**Eksperimen dan Pengujian Model**

Pengujian model yang diusulkan dalam eksperimen bertujuan untuk mendapatkan sebuah hasil (*rule*) yang akan digunakan dalam pengambilan keputusan. Dalam eksperimen dengan algoritma *naive bayes* yang digunakan untuk pelatihan data dan kemudian diulang untuk mendapatkan hasil yang terbaik. Setelah itu dilakukan pengujian data dan dilakukan *ensemble* dengan algoritma *adaboost*.

**Validasi dan Evaluasi**

Validasi dan evaluasi hasil eksperimen untuk membahas dari model yang telah diusulkan, mengetahui kinerja dari algoritma yang telah dilakukan mengetahui kualitas dan efektifitas sebuah model, menentukan apakah model dalam kenyataannya telah mencapai tujuan.

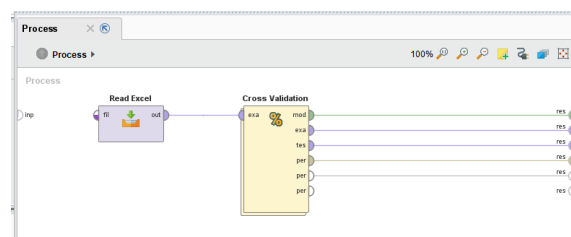
**HASIL DAN PEMBAHASAN**

**Menyiapkan Data Training**

Data *training* memiliki 11 atribut yang dibagi menjadi 2 atribut spesial yaitu *movie* sebagai judul film dan *gross* sebagai *class* yang menyatakan pendapatan yang masuk dalam kategori *block buster*, *Success*, *Average* dan *Flop* atas film yang sudah tayang.

1. *Validation*

*Validation* yaitu melakukan analisis berbagai model dan memilih model dengan kinerja prediksi yang baik. Pada gambar 2. merupakan proses validasi, setelah pembacaan file data, *blok read excel* dihubungkan dengan blok *validation*. Proses data *training* dan *testing* didalam blok proses (*validation*) dapat ditampilkan dengan melakukan klik 2x menggunakan mouse sehingga muncul proses *training* dan *testing*.

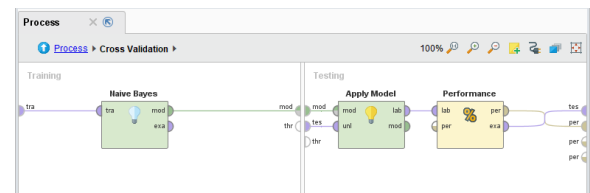


Sumber : (Nurlaela, 2020)  
Gambar 2. Validation untuk metode *Naive bayes*

2. Proses *Training* dan *Testing*

Proses *training* yaitu melakukan proses pelatihan data pada model (*naive bayes*). Sedangkan proses *testing* yaitu melakukan pengujian data yang akan menghasilkan pola. Pada gambar 2 dijelaskan bahwa proses *training* digunakan untuk blok model *naive bayes*,

dihubungkan dengan garis penghubung pada blok *apply model* dan blok *performance* dengan bagian *testing* sebagai penampil informasi hasil dari pengujian data, hasil pengujian ini akan menghasilkan arsitektur *naive bayes*. Dengan menggunakan Rapid Miner 8.2 dapat menganalisis dataset film menggunakan metode *naive bayes*.



Sumber : (Nurlaela, 2020)  
Gambar 3. *Training* dan *Testing* Metode *Naive bayes*

**Hasil Performance Vector *Naive bayes***

Proses klasifikasi dengan Rapid Miner 8.2 dengan metode *naive bayes* yang digunakan untuk mengklasifikasikan data film pada penelitian ini diperoleh nilai *Accuracy*, *Precision*, dan *Recall*.

1. *Accuracy*

*Accuracy* dalam klasifikasi adalah persentase ketepatan *record* data yang diklasifikasikan secara benar setelah dilakukan pengujian pada hasil klasifikasi (Gorunescu, 2011). Rumus *Accuracy* dapat didefinisikan pada persamaan (1), dimana TP merupakan jumlah *record/tuple* positif yang dilabelkan secara benar oleh model algoritma klasifikasi, TN merupakan jumlah *record/tuple* negatif yang dilabelkan secara salah oleh model algoritma klasifikasi, dan P merupakan total semua *record* yang dievaluasi. Semakin tinggi level *Accuracy*-nya, maka dapat dikatakan semakin efektif model algoritma klasifikasi tersebut (Mittal & Gill, 2014)

$$Accuracy = \frac{TP+TN}{P} \dots\dots\dots(1)$$

Akurasi memiliki tingkat diagnosa yaitu (Gorunescu, 2011):

- Akurasi bernilai 0.90 - 1.00 = *Excellent classification*
- Akurasi bernilai 0.80 - 0.90 = *Good classification*
- Akurasi bernilai 0.70 - 0.80 = *Fair classification*
- Akurasi bernilai 0.60 - 0.70 = *Poor classification*
- Akurasi bernilai 0.50 - 0.60 = *Failure*

Sedangkan hasil yang didapat dengan mengetahui jumlah data yang diklasifikasikan secara benar maka dapat diketahui akurasi yaitu 83.22% dari hasil data *testing* maka masuk kedalam kategori *good clasification*.

accuracy: 83.22% +/- 5.69% (micro average: 83.24%)

	true Average	true Block-Buster	true Flop	true Success	class precision
pred. Average	18	0	7	1	69.23%
pred. Block-Buster	0	42	1	8	82.35%
pred. Flop	1	0	0	2	0.00%
pred. Success	3	7	1	94	89.52%
class recall	81.82%	85.71%	0.00%	89.52%	

Sumber : (Nurlaela, 2020)

Gambar 3. Accuracy metode Naive bayes

## 2. Kappa

*Kappa Statistic* adalah analisis statistik yang didasarkan pada kecocokan penafsiran atau derajat kesepakatan (*degree of agreement*) untuk data kualitatif. Pada dasarnya *kappa statistic* menunjukkan analisis diantara kelas-kelas yang berbeda. Semakin tinggi *kappa statistic*, maka akan dipertimbangkan sebagai performa yang mempunyai kriteria atau kinerja yang bagus (Mittal & Gill, 2014).

Kappa yang dihasilkan dari percobaan dengan menggunakan rapid miner 8.2 adalah sebesar 0,706.

kappa: 0.706 +/- 0.112 (micro average: 0.715)

	true Average	true Block-Buster	true Flop	true Success	class precision
pred. Average	18	0	7	1	69.23%
pred. Block-Buster	0	42	1	8	82.35%
pred. Flop	1	0	0	2	0.00%
pred. Success	3	7	1	94	89.52%
class recall	81.82%	85.71%	0.00%	89.52%	

Sumber : (Nurlaela, 2020)

Gambar 4. Kappa metode Naive bayes

## 3. Precision

*Precision* adalah jumlah data yang true positive (jumlah data pred. *Average*, pred.*Block-Buster*, pred.*Flop* dan pred.*Success*) dibagi dengan jumlah data yang dikenali sebagai Tinggi. Dari hasil pengujian nilai *Precision* yaitu 69.23% untuk class *Average*, 82.35% untuk class *Block-Buster* 83.35%, 00,00% untuk class *Block-Buster* dan 89.25% dan untuk class *Success*.

	true Average	true Block-Buster	true Flop	true Success	class precision
pred. Average	18	0	7	1	69.23%
pred. Block-Buster	0	42	1	8	82.35%
pred. Flop	1	0	0	2	0.00%
pred. Success	3	7	1	94	89.52%
class recall	81.82%	85.71%	0.00%	89.52%	

Sumber : (Nurlaela,2020)

Gambar 5. Precision metode Naive bayes

## 4. Recall

*Recall* adalah jumlah data yang true positive dibagi dengan jumlah data yang sebenarnya positif (true positive + true negative). Untuk nilai *Recall* yaitu *Precision* adalah jumlah data yang true positive (jumlah data pred. *Average*, pred.*Block-Buster*, pred.*Flop* dan pred.*Success*) dibagi dengan jumlah data yang dikenali sebagai Tinggi. Dari hasil pengujian nilai *Precision* yaitu 69.23% untuk class *Average*, 81.82% untuk class *Block-Buster* 85.71%, 00,00% untuk class *Block-Buster* dan 89.52% dan untuk class *Success*.

	true Average	true Block-Buster	true Flop	true Success	class precision
pred. Average	18	0	7	1	69.23%
pred. Block-Buster	0	42	1	8	82.35%
pred. Flop	1	0	0	2	0.00%
pred. Success	3	7	1	94	89.52%
class recall	81.82%	85.71%	0.00%	89.52%	

Sumber : (Nurlaela, 2020)

Gambar 6. Recall metode Naive bayes

Hasil *class Accuracy*, *Precision* dan *Recall* untuk memprediksi pendapatan film terhadap beberapa atribut terkait, dan dapat di lihat pada Tabel 1.

**Tabel 1. Hasil Class Accuracy, Precision, dan Recall metode Naive bayes**

PENERAPAN *ADABOOST* UNTUK MENINGKATKAN AKURASI *NAIVE BAYES* PADA PREDIKSI PENDAPATAN FILM

Penulis	Class	Accuracy	Precision	Recall
Dini Nurlaela	<i>Block-Buster</i>	83.22%	82.23%	85.71%
	<i>Success</i>		89.52%	89.52%
	<i>Average</i>		69.23%	81,82%
	<i>Flop</i>		0.00%	0.00%

Sumber : (Nurlaela, 2020)

## Hasil Klasifikasi Class dengan Metode Naive bayes

### 1. Simple Distribution Model

Dengan menggunakan Rapid Miner 8.2. untuk menganalisis tabel data film dalam memprediksi pendapatan film dengan menggunakan metode *naive bayes* dapat menghasilkan 2 (dua) kelas utama pembagian seperti pada gambar 6.

Hasil klasifikasi dari data film dengan metode *naive bayes* membagi 3 kelas klasifikasi yaitu *Block-Buster*, *Success*, *Average* dan *Flop*. Untuk nilai *class Average* yaitu 0.119 , *class Block-Buster* yaitu 0,265, untuk kelas *Flop* 0.049 dan *class Success* 0,568.

## Simple Distribution

Distribution model for label attribute Class

Class Average (0.119)  
11 distributions

Class Block-Buster (0.265)  
11 distributions

Class Flop (0.049)  
11 distributions

Class Success (0.568)  
11 distributions

Sumber : (Nurlaela, 2020)

Gambar 7. Simple Distribution Model metode Naive bayes



2. Confusion Matrix

Nilai *Accuracy* dari *confusion matrix* tersebut adalah sebagai berikut:

$$Accuracy = \frac{(18+42+2+94)}{(18+0+0+3+0+42+0+7+7+1+2+1+1+8+0+94)}$$

$$Accuracy = \frac{156}{184} = 0,8478 = 84.78\%$$

**PerformanceVector**

```
PerformanceVector:
accuracy: 83.22% +/- 5.69% (micro average: 83.24%)
ConfusionMatrix:
True: Average Block-Buster Flop Success
Average: 18 0 7 1
Block-Buster: 0 42 1 8
Flop: 1 0 0 2
Success: 3 7 1 94
kappa: 0.706 +/- 0.112 (micro average: 0.715)
ConfusionMatrix:
True: Average Block-Buster Flop Success
Average: 18 0 7 1
Block-Buster: 0 42 1 8
Flop: 1 0 0 2
Success: 3 7 1 94
```

Sumber : (Nurlaela, 2020)

Gambar 8. *Confusion Matrix* metode *Naive bayes*

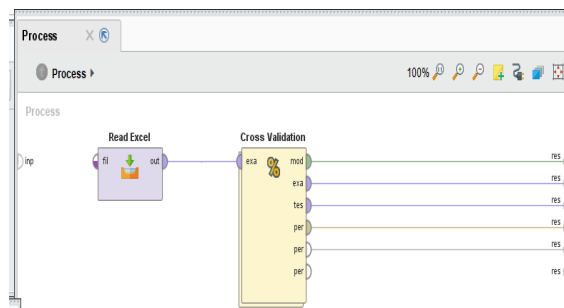
**Hasil Eksperimen dan Pengujian NB + Adaboost**

1. Menyiapkan Data *Training*

Data *training* memiliki 11 atribut yang dibagi menjadi 2 atribut spesial yaitu *movie* sebagai judul film dan *gross* sebagai *class* yang menyatakan pendapatan yang tinggi sedang dan rendah atas film yang sudah tayang.

2. *Validation*

*Validation* yaitu melakukan analisis berbagai model dan memilih model dengan kinerja prediksi yang baik. Pada gambar 8. merupakan proses validasi, setelah pembacaan file data, *blok read excel* dihubungkan dengan *blok validation*. Proses data *training* dan *testing* didalam *blok proses (validation)* dapat ditampilkan dengan melakukan klik 2x menggunakan mouse sehingga muncul proses *training* dan *testing*.

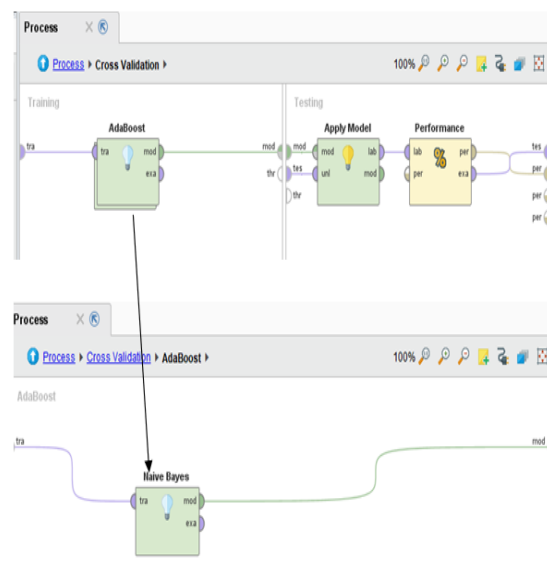


Sumber : (Nurlaela, 2020)

Gambar 9. *Validation* untuk metode *Naive bayes + Adaboost*

3. Proses *Training* dan *Testing*

Proses *training* yaitu melakukan proses pelatihan data pada model (*Adaboost* yg di dalamnya terdapat *naive bayes*). Sedangkan proses *testing* yaitu melakukan pengujian data yang akan menghasilkan pola. Pada gambar 9 dijelaskan bahwa proses *training* digunakan untuk *blok model Adaboost + naive bayes*, dihubungkan dengan garis penghubung pada *blok apply model* dan *blok performance* dengan bagian *testing* sebagai penampilan informasi hasil dari pengujian data, hasil pengujian ini akan menghasilkan arsitektur *naive bayes+Adaboost*. Dengan menggunakan Rapid Miner 8.2 dapat menganalisis dataset film menggunakan metode *naive bayes*.



Sumber : (Nurlaela, 2020)

Gambar 10. *Training* dan *Testing* Metode *Naive bayes + Adaboost*

**Hasil Performance Vector *Naive bayes+Adaboost***

Proses klasifikasi dengan Rapid Miner 8.2 dengan metode *naive bayes* yang digunakan untuk mengklasifikasikan data film pada penelitian ini diperoleh nilai *Accuracy*, *Precision*, dan *Recall*.

1. *Accuracy*

Akurasi memiliki tingkat diagnosa yaitu (Gorunescu, 2011) :

Akurasi bernilai 0.90 - 1.00 = *Excellent classification*

Akurasi bernilai 0.80 - 0.90 = *Good classification*

Akurasi bernilai 0.70 - 0.80 = *Fair classification*

Akurasi bernilai 0.60 - 0.70 = *Poor classification*

Akurasi bernilai 0.50 - 0.60 = *Failure*

Dengan mengetahui jumlah data yang diklasifikasikan secara benar maka dapat diketahui akurasi dengan menggunakan algoritma *naive bayes* dan di boosting dengan menggunakan *Adaboost* yaitu 84.24% dari hasil data *testing* dan masuk kedalam kategori *good classification*.

accuracy: 84.24% +/- 7.67% (micro average: 84.32%)

	true Average	true Block-Buster	true Flop	true Success	class precision
pred. Average	19	0	8	4	61.29%
pred. Block-Buster	0	43	1	5	87.76%
pred. Flop	0	0	0	2	0.00%
pred. Success	3	6	0	94	91.26%
class recall	86.36%	87.76%	0.00%	89.52%	

Sumber : (Nurlaela, 2020)

Gambar 11. Accuracy metode Naive bayes + Adaboost

## 2. Kappa

Kappa yang dihasilkan dari percobaan dengan metode *naive bayes+Adaboost* menggunakan rapid miner 8.2 adalah sebesar 0,731.

kappa: 0.731 +/- 0.123 (micro average: 0.736)

	true Average	true Block-Buster	true Flop	true Success	class precision
pred. Average	19	0	8	4	61.29%
pred. Block-Buster	0	43	1	5	87.76%
pred. Flop	0	0	0	2	0.00%
pred. Success	3	6	0	94	91.26%
class recall	86.36%	87.76%	0.00%	89.52%	

Sumber : (Nurlaela, 2020)

Gambar 12. Kappa metode Naive bayes + Adaboost

## 3. Precision

Dari hasil pengujian nilai Precision yaitu 61.29% untuk *class Average*, 87.76% untuk *class Block-Buster*, 0.00% untuk *Flop* dan 91.26% untuk *class Success*.

	true Average	true Block-Buster	true Flop	true Success	class precision
pred. Average	19	0	8	4	61.29%
pred. Block-Buster	0	43	1	5	87.76%
pred. Flop	0	0	0	2	0.00%
pred. Success	3	6	0	94	91.26%
class recall	86.36%	87.76%	0.00%	89.52%	

Sumber : (Nurlaela, 2020)

Gambar 13. Precision metode Naive bayes + Adaboost

## 4. Recall

Recall adalah jumlah data yang true positive dibagi dengan jumlah data yang sebenarnya positif (true positive + true negative). Untuk nilai Recall yaitu 86.36 % pada *class Average* ,87.76% pada *class blok\_buster*, 0.00% pada *class Flop* dan 89.52% pada *class Success*.

	true Average	true Block-Buster	true Flop	true Success	class precision
pred. Average	19	0	8	4	61.29%
pred. Block-Buster	0	43	1	5	87.76%
pred. Flop	0	0	0	2	0.00%
pred. Success	3	6	0	94	91.26%
class recall	86.36%	87.76%	0.00%	89.52%	

Sumber : (Nurlaela, 2020)

Gambar 14. Recall metode Naive bayes + Adaboost

Hasil *class Accuracy*, *Precision* dan *Recall* untuk memprediksi pendapatan film terhadap beberapa atribut terkait dengan penggunaan metode *naive bayes* dan diboosting menggunakan *Adaboost*, dan dapat di lihat pada Tabel .2.

**Tabel 2. Hasil Class Accuracy, Precision, dan Recall metode Naive bayes+Adaboost**

PENERAPAN ADABOOST UNTUK MENINGKATKAN AKURASI NAIVE BAYES PADA PREDIKSI PENDAPATAN PENJUALAN FILM				
Penulis	Class	Accuracy	Precision	Recall
Dini Nurlaela	<i>Block-Buster</i>	84.24%	87.76%	87.76%
	<i>Success</i>		91.26%	89.52%
	<i>Average</i>		61.29%	86.36%
	<i>Flop</i>		0.00%	0.00%

Sumber : (Nurlaela, 2020)

## Hasil Klasifikasi Class dengan Metode Naive bayes+Adaboost

### 1. Simple Distribution Model

Dengan menggunakan Rapid Miner 8.2. untuk menganalisis tabel data film dalam memprediksi pendapatan film dengan menggunakan metode *naive bayes* dan di boosting dengan *Adaboost* dapat menghasilkan 4 (tiga) kelas utama pembagian.

Hasil klasifikasi dari data film dengan metode *naive bayes* membagi 4 kelas klasifikasi yaitu *class Block-Buster*, *class Success* , *class Average* dan *class Flop* sebanyak 7 model, dan di jelaskan dalam bentuk tabel perbandingan 7 model.

**Tabel 3. Distribution Model metode Naive bayes+Adaboost**

Model ke-	Class				Weight
	<i>Block-Buster</i>	<i>Success</i>	<i>Average</i>	<i>Flop</i>	
1	0.265	0.568	0.119	0.265	3.236
2	0.275	0.432	0.130	0.163	2.627
3	0.186	0.464	0.205	0.145	1.384
4	0.136	0.407	0.134	0.323	1.908
5	0.103	0.530	0.089	0.278	2.268
6	0.192	0.463	0.049	0.296	2.634
7	0.167	0.307	0.368	0.159	0.129

Sumber : (Nurlaela, 2020)

2. Confusion Matrix

Nilai *Accuracy* dari *confusion matrix* tersebut adalah sebagai berikut:

$$Accuracy = \frac{(19+43+2+94)}{(19+0+0+3+0+43+0+6+8+1+2+0+4+5+0+94)}$$

$$Accuracy = \frac{158}{185} = 0,8541 = 85.41\%$$

PerformanceVector

```
PerformanceVector:
accuracy: 84.24% +/- 7.67% (micro average: 84.32%)
ConfusionMatrix:
True: Average Block-Buster Flop Success
Average: 19 0 8 4
Block-Buster: 0 43 1 5
Flop: 0 0 0 2
Success: 3 6 0 94
kappa: 0.731 +/- 0.123 (micro average: 0.736)
ConfusionMatrix:
True: Average Block-Buster Flop Success
Average: 19 0 8 4
Block-Buster: 0 43 1 5
Flop: 0 0 0 2
Success: 3 6 0 94
```

Sumber : (Nurlaela, 2020)

Gambar 15. *Confusion Matrix* metode *Naive bayes + Adaboost*

3. Distribution Tabel

Dapat dilihat tabel distribusi hasil analisa dengan metode *naive bayes* dan *Adaboost* terhadap tabel data film, dalam memprediksi pendapatan film pada tahun 2014 dan 2015.

Row No.	Class	prediction(C...	confidence_...	confidence_...	confidence_...	confidence_...
1	Average	Success	0.000	1.000	0.000	0.000
2	Average	Success	0.000	1.000	0.000	0.000
3	Average	Success	0.000	1.000	0.000	0.000
4	Block-Buster	Block-Buster	1.000	0.000	0.000	0.000
5	Block-Buster	Block-Buster	0.921	0.078	0.000	0.000
6	Block-Buster	Block-Buster	1.000	0.000	0.000	0.000
7	Block-Buster	Block-Buster	1.000	0.000	0.000	0.000
8	Block-Buster	Block-Buster	1.000	0.000	0.000	0.000
9	Flop	Average	0.000	0.000	0.998	0.002
10	Success	Success	0.000	1.000	0.000	0.000
11	Success	Success	0.003	0.996	0.000	0.000
12	Success	Success	0.000	1.000	0.000	0.000
13	Success	Success	0.000	1.000	0.000	0.000
14	Success	Success	0.000	1.000	0.000	0.000
15	Success	Success	0.000	1.000	0.000	0.000
16	Success	Success	0.000	1.000	0.000	0.000

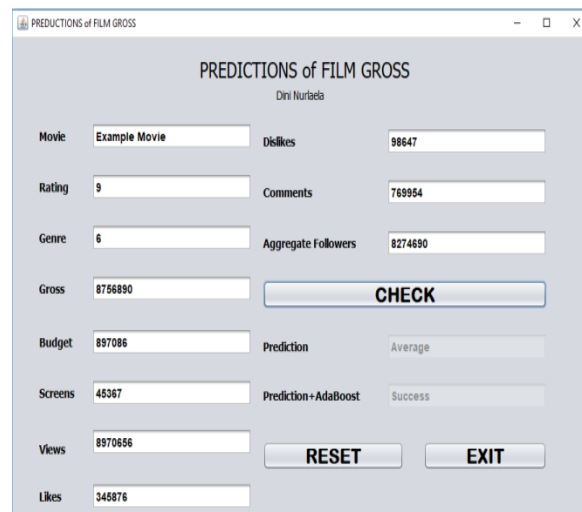
Sumber : (Nurlaela, 2020)

Gambar 16. *Distribution Tabel* metode *Naive bayes + Adaboost*

Implementasi Aplikasi Model dari *Naive bayes* dengan *Naive bayes* dan *Adaboost* (NBBoost)

Berdasarkan hasil eksperimen dari model yang sudah ditemukan, maka di dapatkan model yang dapat di implemnetasikan dalam bentuk aplikasi sederhana. Dalam pengujiannya dapat

dilakukan menggunakan sample yang baru atau yg sudah ada, aplikasi pengujiannya dapat dilihat pada gambar 17.



Sumber : (Nurlaela, 2020)

Gambar 17. Aplikasi Pengujian NB dan NB + *Adaboost*

Hasil pengujian dari aplikasi dapat dilihat pada gambar 16. dengan masukan (input) yang tertera pada gambar awalnya prediksi menggunakan NB saja menunjukkan hasilnya film tersebut masuk kedalam *class Flop* dan prediksi dengan menggunakan NB+*Adaboost* menghasilkan *class Average*.

KESIMPULAN

Penelitian ini dilakukan dengan menerapkan *adaboost* untuk menyelesaikan ketidak seimbangan kelas pada penentuan pendapatan film dengan metode *naive bayes*. Eksperimen telah dilakukan untuk mendapatkan sebuah model arsitektur yang cukup optimal dan cukup akurat. Hasil eksperimen dengan *naive bayes* sebesar 83,22% , kappa 0,706 , class precision pred. *Block-Buster* 82.35% , class precision pred. *Success* 89.52% , class precision pred. *Average* 69.23% dan untuk class flop 00.00%. Dan untuk class recall masing-masing 87.76%, 80.52%, 89.52% dan 00.00%. Setelah dilakukan *boosting* dengan *adaboost* akurasi meningkat sebesar 84.24% kappa 0.731, class class precision pred. *Block-Buster* 87.76%, class precision pred. *Success* 91.26% , class precision pred. *Average* 61.29% dan untuk class flop 00.00%. Dan untuk class recall masing-masing 87.76%, 80.52%, 89.52% dan 00.00% dapat ditarik kesimpulan bahwa kenaikan akurasi sebanyak 1.02%. Hasil pengujian diatas dapat disimpulkan bahwa metode *adaboost* terbukti efektif dalam menyelesaikan ketidak

seimbangan kelas pada penentuan pendapatan film dengan metode naive bayes.

Analysis of Classification Techniques on Medical Data Sets, 454–460.

#### REFERENSI

- Ahmed, M., Jahangir, M., Afzal, H., Majeed, A., & Siddiqi, I. (2015). Using crowd-source based features from social media and conventional features to predict the movies popularity. *Proceedings - 2015 IEEE International Conference on Ahmed, Mehreen Jahangir, Maham Afzal, Hammad Majeed, Awais Siddiqi, Imran Smart City, SmartCity 2015, Held Jointly with 8th IEEE International Conference on Social Computing and Networking, SocialCom 2015*, (December 2015), 273–278. <https://doi.org/10.1109/SmartCity.2015.83>
- Gorunescu, F. (2011). *Data Mining: Concepts, Models and Techniques. Data mining - Concepts, Models and Technique.* <https://doi.org/10.1007/978-3-642-19721-5>
- Mittal, P., & Gill, N. S. (2014). a Comparative Analysis of Classification Techniques on Medical Data Sets, 454–460.
- Nugroho, Y. S., & Pratiwi, R. W. (2016). Prediksi Rating Film Menggunakan Metode Naïve Bayes. *Jurnal Teknik Elektro (ISSN 1411-0059)*, 8(2), 60–63.
- Nurlaela, D. (2020). *LAPORAN AKHIR PENELITIAN PDY: PENERAPAN ADABOOST UNTUK MENINGKATKAN AKURASI NAIVE BAYES PADA PREDIKSI PENDAPATAN PENJUALAN FILM.* Bogor.
- Tsou, B. K., & Ma, M. (2011). Aspect Based Opinion Polling from Customer Reviews. *IEEE Transactions on Affective Computing*, 2(1), 37–49. <https://doi.org/10.1109/T-AFFC.2011.2>