

IMPLEMENTASI KLASIFIKASI NAIVE BAYES DALAM MEMPREDIKSI LAMA STUDI MAHASISWA (STUDI KASUS : UNIVERSITAS DHYANA PURA)

Kelvin Henry Loudry Malelak¹, I Made Dwi Ardiada², Gerson Feoh³

¹²³Teknik Informatika, Fakultas Kesehatan Sains dan Teknologi, Universitas Dhyana Pura
Jalan Raya Padang Luwih Tegaljaya, Dalung Kuta Utara, Kabupaten Badung, Bali, Indonesia

e-mail: hennrymalelak@gmail.com¹, dwiardiada@undhirabali.ac.id², gerson.feoh@undhirabali.ac.id³

Received : September, 2021

Accepted : October, 2021

Published : October, 2021

Abstract

Under normal conditions, undergraduate or undergraduate students from a university can complete their studies for 4 years or 8 semesters. In fact, many students complete their study period of more than 4 years. It is known that in fact in the 2015/2016 academic year there were 744 people who were accepted as students. Of the 744 people who were accepted, 405 people had completed a study period of about 4 years and the remaining 39 people completed their studies for 5 years and 300 of them did not continue their studies. Based on the problem, so this study implements a classification that can help Dhyana Pura University in predicting the length of study for students who are currently studying in various study programs at Dhyana Pura University. The author's method serves in the classification to predict long student study period is the Naive Bayes algorithm. By using the Java-based Rapid Miner tool to classify graduation data. Then the implementation of data mining which is divided into 968 training data and 193 data testing data with naive Bayes has succeeded in obtaining an accuracy rate of 100% which also has very good parameters.

Keywords: data mining, Naive Bayes Rapid Miner, training data, testing data.

Abstrak

Dalam kondisi normal, mahasiswa strata-1 atau S1 dari suatu Universitas, dapat menyelesaikan studinya selama 4 tahun atau 8 semester. Pada kenyataannya, banyak mahasiswa yang menyelesaikan masa studinya lebih dari 4 tahun. Diketahui bahwa faktanya pada tahun akademik 2015/2016 terdapat 744 orang yang diterima sebagai mahasiswa. Dari 744 orang yang diterima, 405 orang telah menyelesaikan masa studi sekitar 4 tahun dan sisanya 39 orang menyelesaikan studinya selama 5 tahun serta 300 orang diantaranya yang tidak melanjutkan studinya. Berdasarkan permasalahan di atas, maka penelitian ini melakukan implementasi klasifikasi yang dapat membantu pihak Universitas Dhyana Pura dalam memprediksi lama studi mahasiswa yang saat ini sedang kuliah di berbagai program studi pada Universitas Dhyana Pura. Metode yang penulis sajikan dalam klasifikasi untuk memprediksi lamanya masa studi mahasiswa adalah Algoritma naive bayes. Dengan menggunakan tool Rapid Miner yang berbasis java untuk mengklasifikasikan data kelulusan. Kemudian implementasi data mining yang terbagi atas data training sebanyak 968 data dan data testing 193 data dengan naive bayes berhasil memperoleh tingkat keakuratan sebesar 100% yang juga parameternya tergolong sangat baik.

Kata Kunci: data mining, data training, data testing, Naive Bayes, Rapid Miner.

1. PENDAHULUAN

Dalam kondisi normal, mahasiswa strata-1 atau S1 dari suatu Universitas, dapat menyelesaikan studinya selama 4 tahun atau 8 semester. Pada kenyataannya, dengan berbagai sebab, poly mahasiswa yang menuntaskan masa studinya lebih 4 tahun. Lama studi mahasiswa juga menjadi salah satu kriteria dalam penilaian akreditasi baik program studi maupun Universitas negeri atau swasta menurut lembaga Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT). Penelitian ini dilakukan pada mahasiswa di Universitas Dhyana Pura sebagai objek penelitian dan sampel, dengan mengambil data dari para alumni yang akan dibandingkan antara prediksi waktu kelulusan dengan kondisi kelulusan sebenarnya. Penulis memakai *Naive Bayes* pada metode penelitian ini lantaran metode ini pula dipakai pada prediksi berbagai kondisi pada penelitian-penelitian sebelumnya, diantaranya dalam penelitian Anggraini (2020) dan penelitian Handayani (2015). Tingkat keakuratan *Naive Bayes* dalam memprediksi kelulusan sebelumnya sudah di buktikan dalam penelitian Tambunan (2020) dan penelitian Silvi (2020) Berdasarkan penelitian-penelitian tersebut, maka penulis menggunakan *Naive Bayes* sebagai metode dalam penelitian ini untuk dapat digunakan oleh pihak Universitas Dhyana Pura dalam memprediksi kelulusan mahasiswanya.

Universitas Dhyana Pura (UNDHIRA) merupakan salah satu Universitas swasta di Kabupaten Badung yang terakreditasi B oleh BAN-PT. UNDHIRA memiliki 15 program studi dan terbagi atas 2 fakultas. Berdasarkan pengumpulan data awal yang penulis lakukan pada departemen Information And Communication Teknologi (ICT) di Universitas Dhyana Pura dan buku wisudawan yang diupload pada website Undhira, diketahui bahwa faktanya pada tahun akademik 2015/2016 terdapat 744 orang yang diterima sebagai mahasiswa. Dari 744 orang yang diterima, 405 orang telah menyelesaikan masa studi sekitar 4 tahun dan sisanya 39 orang menyelesaikan studinya selama 5 tahun serta 300 orang diantaranya yang tidak melanjutkan studinya. Tahun akademik 2016/2017 juga mendapat jumlah mahasiswa yang diterima ada 574 orang. Dari 574 orang mahasiswa yang diterima, 320 orang menyelesaikan studi 4 tahun dan sisanya 254 orang akan

menyelesaikan studinya 4,5 sampai 5 tahun. 2 (dua) faktor primer yang menghipnotis usang studinya mahasiswa yaitu faktor dalam diri mahasiswa dan luar diri mahasiswa. Faktor dalam diri mahasiswa merupakan faktor berdasarkan pada diri yang mencakup kecerdasan mahasiswa, penangkapan materi ketika dikelas dan kemampuan mahasiswa itu sendiri, sedangkan faktor luar diri mahasiswa merupakan faktor berdasarkan luar diri yang mencakup syarat ekonomi, syarat sosial dan pergaulan. Selain menghipnotis kelulusan, faktor tadi pula akan menghipnotis ekonomi mahasiswa misalnya wajib membayar uang semester lebih buat menanggung mata kuliah yang ketertinggalan mata kuliah (Moonallika, 2020). Permasalahan lainnya yaitu ketidakseimbangan rasio jumlah mahasiswa dengan jumlah lulusan setiap tahunnya sangat berpengaruh pada akreditasi program studi maupun perguruan tinggi itu sendiri. Berdasarkan konflik yang sudah disebutkan, maka penelitian ini melakukan implementasi klasifikasi yang dapat membantu pihak Universitas Dhyana Pura dalam memprediksi lama studi mahasiswa yang saat ini sedang kuliah di berbagai program studi pada Universitas Dhyana Pura. Maka dengan ini saya menggunakan perumusan masalah sebagai berikut Bagaimana mengimplementasikan metode *Naive Bayes* dalam prediksi lama studinya mahasiswa Undhira? serta Bagaimana tingkat keakuratan dari metode *Naive Bayes* dalam prediksi lama studinya mahasiswa Undhira jika dibandingkan dengan perhitungan manual dan *software*?, dengan tujuan Mengimplementasikan metode *Naive Bayes* dalam memprediksi lama studinya mahasiswa Undhira dan Mengetahui tingkat keakuratan dari metode *Naive Bayes* dalam memprediksi lama studinya mahasiswa Undhira dengan perhitungan manual yang dibandingkan dengan *Rapid Miner Studio*.

Metode yang penulis sajikan dalam pembagian terstruktur mengenai untuk memprediksi lamanya masa studi mahasiswa yaitu Algoritma *Naive Bayes*. *Naive Bayes* merupakan sebuah pembagian terstruktur tentang menggunakan Metode probabilitas dan statistik oleh ilmuwan Inggris Thomas Bayes. Algoritma *Naive Bayes* memprediksi peluang pada masa depan berdasarkan pengalaman dalam masa sebelumnya menjadi akibatnya dikenal menjadi Teorema *Bayes* (Ratniasih,

2019). Penulis menggunakan *tool Rapid Miner* yang berbasis java untuk mengklasifikasikan data kelulusan. Dalam proses prediksi lamanya masa studi mahasiswa yang penulis buat, admin terlebih dahulu menyiapkan data-data yang diperlukan lalu memasukkannya ke dalam file dikomputer atau laptop. Setelah itu, maka akan dilakukan proses pengambilan data berupa *microsoft excel* yang ada pada file di laptop *admin* lalu mulai memilih label dan tipe data, kemudian melakukan pengukuran tingkat akurasi metode Naive Bayes. Selesai melakukan pengukuran tingkat akurasi metode Naive

2. METODE PENELITIAN

Tahapan metode yang dipakai yaitu Knowledge Discovery in Database (KDD) mempunyai arti yang berbeda, akan namun saling berkaitan satu menggunakan yang lainnya. Tahapan-tahapan KDD terlihat menjadi berikut (Silvi, 2020):

1. Data Selection

merupakan tahapan pemilihan data, dalam suatu *database* data yang dipakai hanya beberapa data yang sesuai untuk dianalisa.

2. Pre-processing/Cleaning

Pada proses ini dilakukan proses dimana data yang bermasalah dimusnahkan. Misalkan data yang diperoleh berdasarkan database eksklusif umumnya mempunyai data yang nir lengkap, terdapat beberapa isian yang hilang, data nir valid juga galat tulis. Maka data-data misalnya itu akan dibuang. Proses *cleaning* pula mensugesti kinerja menurut data mining karena data yang diproses akan berkurang jumlahnya dan kompleksitasnya.

3. Transformation Data

Tahap ini yaitu proses penyesuaian data menggunakan format yang terdapat di data mining. Misalkan mengganti data berdasarkan bentuk *numeric* sebagai kategorikal buat metode asosiasi dan *clustering*.

4. Data Mining

Yaitu tahapan yang dilakukan saat metode dipakai sebagai akibatnya mendapatkan keterangan baru menurut data yang ada. Pada tahapan ini juga perkumpulan data yang sudah melewati tahapan sebelumnya dianalisa menurut metode yang dipilih buat memperoleh output akhir.

5. Interpretation/Evaluation

yaitu proses berdasarkan data mining yang diperlihatkan pada pola informasi yang bisa

Bayes, lalu hasilnya akan sebagai sebuah tabel. Dengan proses klasifikasi prediksi lamanya masa studi mahasiswa yang penulis buat, diharapkan dapat memudahkan mahasiswa dalam meningkatkan kualitas di bidang pendidikan dari program studi maupun universitas.

Menurut latar belakang tersebut maka penulis tertarik mengajukan penelitian menggunakan judul "Implementasi Klasifikasi Naive Bayes Dalam Memprediksi Lama Studi Mahasiswa (Studi Kasus : Universitas Dhyana Pura)".

dimengerti dari pihak berkepentingan. Pada tahapan ini dilakukan pemeriksaan terhadap pola atau informasi apakah bertentangan memakai keterangan atau hipotesis sebelumnya.

Naive Bayes yaitu proses pengklasifikasian probabilistik memakai cara menghitung sekumpulan probabilitas menggunakan menjumlahkan frekuensi dan kombinasi nilai *data set* yang diberikan. Algoritma menggunakan Teorema *Bayes* memakai mengasumsikan seluruh atribut independen atau nir saling ketergantungan yang diberikan berdasarkan nilai dalam variabel kelas (Heryana, 2019). Penjabaran pembagian terstruktur tentang *Bayes* adalah pembagian terstruktur mengenai statistik yang sanggup memprediksi kelas suatu anggota probabilitas (Ratniasih, 2019).

Pada teorema *Bayes* terdapat bentuk umum,

(Silvi, 2020):

$$P(C_i|X) = \frac{P(X|C_i) \times P(C_i)}{P(X)} \quad (1)$$

Keterangan:

X : Data menggunakan group yang belum diketahui

C_i : Hipotesis Data X adalah suatu group spesifik

P(C_i|X) : Probabilitas Hipotesis C_i menurut syarat X

P(C_i) : Probabilitas Hipotesis C_i

P(X|C_i) : Probabilitas X menurut syarat tersebut

P(X) : Probabilitas menurut X

Langkah awal, dilakukan pengelompokan data training untuk target yang dicapai yaitu lulus

atau tidaknya seorang mahasiswa, lalu langkah selanjutnya yaitu menghitung nilai probabilitas menurut pengelompokan data pelatihan yang ada. Nilai probabilitas tertinggi setelah penerapan contoh sebagai penentu prediksi kelulusan mahasiswa cepat atau terlambat.

3. HASIL DAN PEMBAHASAN

Pada proses ini dilakukan dengan tahap pengambilan data yang sudah didapatkan pada lokasi penelitian sebanyak 988 data mahasiswa dan mahasiswi yang lulus tahun periode 2019-2020, kemudian pada 988 data tersebut di dilakukan sesuai dengan tahapan analisa metode *Naive Bayes*.

3.1 Data Selection

Proses menunjukkan bahwa atribut NIM bisa diketahui tahun masuk mahasiswa, kemudian atribut JENIS KELAMIN diketahui jenis kelamin mahasiswa, lalu juga atribut Indeks prestasi Kumulatif akan mengetahui ipk mahasiswa tergolong 3 atau 2, dan atribut jalur masuk akan disesuaikan dengan data masuknya mahasiswa.

3.2 Data Pre-processing

Pembersihan data dilakukan menggunakan memperbaiki data kelulusan mahasiswa sebelum dilakukan proses mining, misalnya adanya missing values dalam data. Dalam hal ini, missing value pada data mahasiswa dari berdasarkan data-data yang atributnya tidak mempunyai nilai atau informasi.

3.3 Data Transformation

Selanjutnya dilakukan transformasi data, yaitu perubahan data nim menjadi tahun masuk. Transformasi data ini berguna untuk memudahkan pada saat penghitungan di dalam sebuah sistem. Hasil transformasi dari 988 data terdapat 968 data setelah dilakukan *data pre-processing*.

3.4 Proses Pengolahan Data Mining

Pada proses ini dilakukan dengan tahap pengambilan data yang sudah dibersihkan terlebih dahulu yaitu bisa disebut dengan Data Mining sebanyak 968 data mahasiswa dan mahasiswi yang lulus tahun periode 2019-2020. Kemudian pada 968 data tersebut di *Split data* lagi sebagai 80 % *data training* dan 20% *data testing* pada ambil berdasarkan mahasiswa yang belum lulus. Penelitian ini juga menjadi perbandingan dari proses *split data* 80:20 dan 70:30 untuk dijadikan yang terbaik.

Tabel 3.1 Potongan *Data Training*

Tahun Masuk	Jenis Kelamin	Jalur Masuk	IPK	Keputusan
2014	Pria	Reguler	3.17	Terlambat
2014	Wanita	Reguler	3.75	Terlambat
2014	Wanita	Reguler	3.67	Terlambat
2015	Wanita	Reguler	3.62	Cepat
2015	Wanita	Reguler	3.54	Cepat
2015	Wanita	Reguler	3.48	Cepat
2015	Pria	Reguler	3.65	Cepat
2015	Pria	Reguler	3.51	Cepat
2015	Wanita	Reguler	3.64	Cepat
2015	Pria	Reguler	3.68	Cepat
2015	Pria	Reguler	3.74	Cepat
2015	Wanita	Reguler	3.61	Cepat
2015	Pria	Reguler	3.77	Cepat
2015	Wanita	Reguler	3.56	Cepat
2015	Pria	Reguler	3.52	Cepat
2015	Pria	Reguler	3.32	Cepat
2015	Wanita	Reguler	3.76	Cepat
2015	Pria	Reguler	3.60	Cepat
2015	Pria	Reguler	3.68	Cepat
2015	Pria	Reguler	3.52	Cepat
2015	Wanita	Reguler	3.50	Cepat
2015	Pria	Reguler	3.42	Cepat
2015	Pria	Reguler	3.67	Cepat
2015	Wanita	Reguler	3.55	Cepat
2015	Wanita	Reguler	3.53	Cepat
2015	Wanita	Reguler	3.68	Cepat
2015	Pria	Reguler	3.44	Cepat
2015	Wanita	Reguler	3.89	Cepat

Tabel 3.2 Potongan *Data Testing*

Tahun Masuk	Jenis Kelamin	Jalur Masuk	IPK	Keputusan
2017	Pria	Reguler	3.7	Cepat
2017	Wanita	Reguler	3.55	Cepat
2017	Wanita	Reguler	3.33	Cepat
2017	Pria	Reguler	3.8	Cepat
2017	Pria	Reguler	3.9	Cepat
2017	Pria	Reguler	3.5	Cepat
2017	Wanita	Reguler	3.65	Cepat
2017	Pria	Reguler	3.45	Cepat
2017	Pria	Reguler	3.65	Cepat
2017	Pria	Reguler	3.57	Cepat
2017	Wanita	Reguler	3.71	Cepat
2017	Wanita	Reguler	3.35	Cepat

2017	Wanita	Reguler	3.68	Cepat
2017	Wanita	Reguler	3.39	Cepat
2017	Wanita	Reguler	3.25	Cepat
2017	Wanita	Reguler	3.49	Cepat
2017	Pria	Reguler	3.75	Cepat
2017	Wanita	Reguler	3	Cepat
2017	Wanita	Reguler	3.5	Cepat
2017	Pria	Reguler	3.8	Cepat
2017	Pria	Reguler	3.55	Cepat
2017	Pria	Reguler	3.69	Cepat
2017	Wanita	Reguler	3.85	Cepat
2017	Wanita	Reguler	3.88	Cepat
2017	Wanita	Reguler	3.55	Cepat
2017	Wanita	Reguler	3.7	Cepat
2017	Wanita	Reguler	3.89	Cepat
2017	Wanita	Reguler	3.3	Cepat
2017	Wanita	Reguler	3.5	Cepat

3.5 Proses Perhitungan Naive Bayes

Dengan menghitung probabilitas masing-masing kelas pencarian tersebut merupakan kelayakan dari metode *Naive Bayes*. Prediksi kelulusan ditentukan lewat dua kelas yaitu "Cepat" dan "Terlambat". Dimana ada tahapan-tahapannya sebagai berikut.

1. Proses Menghitung Jumlah Kelas

$$P(C_i) = \frac{\text{Kemunculan data masuk}}{\text{Jumlah keseluruhan data}} \quad (2)$$

Tabel 3.3 Probabilitas Kelas

	Kelas	
	Cepat	Terlambat
	894/968	74/968
HasilProbabilitas	0.923553	0.076446

2. Proses Menghitung Jumlah Kelas

$$P(X|C) = \frac{\text{jumlah data dengan atribut}}{\text{jumlah data masuk}} \quad (3)$$

Tabel 3.4 Probabilitas Atribut Jenis Kelamin

Jenis Kelamin	Kelas	
	Cepat	Terlambat
Pria	294/894	39/74
HasilProbabilitas	0.328859	0.527027
Wanita	600/894	35/74
HasilProbabilitas	0.671140	0.472972

Tabel 3.5 Probabilitas Atribut IPK

IPK	Kelas	
	Cepat	Terlambat
2	19/894	16/74
HasilProbabilitas	0.021252	0.216216
3	875/894	58/74
HasilProbabilitas	0.978747	0.783783

Tabel 3.6 Probabilitas Tahun Masuk

Tahun Masuk	Kelas	
	Cepat	Terlambat
2014	0/894	36/74
HasilProbabilitas	0	0.486486
2015	401/894	38/74
HasilProbabilitas	0.448545	0.513513
2016	493/894	0/74
HasilProbabilitas	0.551454	0

Tabel 3.7 Data Uji

Tahun Masuk	Jenis Kelamin	Jalur Masuk	IPK
2015	Wanita	Reguler	3

Kategori	Nilai	Hasil
Cepat	$(0.448545 * 0.671140 * 0.978747)^*$ 0.923553	0.272114
Terlambat	$(0.513513 * 0.472972 * 0.783783)^*$ 0.076446	0.014552
Prediksi	CEPAT	0.272114

Tabel 3.8 Hasil Data Uji

3.6 Interpretation/Evaluation

yaitu proses evaluasi menggunakan pengujian *confusion matrix*. *Confusion matrix* itu tabel yang memberikan *output* kerja berdasarkan suatu klasifikasi. Menghitung taraf akurasi adalah hal terpenting pada sebuah penelitian supaya bisa diketahui taraf keberhasilan dan kegagalan terhadap penelitian (Silvi, 2020).

Tabel 3.7 Tabel Confusion Matrix

Aktual	Prediksi	
	Cepat	Terlambat
Cepat	TP	FN
Terlambat	FP	TN

Keterangan :

TN = nilai true negative

TP = nilai true positive

FP = nilai false positive

FN = nilai false negative

TN merupakan jumlah lulus cepat dengan berhasil diprediksi menjadi lulus cepat. TP merupakan jumlah lulus terlambat dengan berhasil diprediksi lulus terlambat. FP merupakan jumlah lulus cepat dengan diprediksi lulus terlambat. Sedangkan, FN menyatakan lulus terlambat dengan diprediksi lulus cepat (Fahrudy, 2020).

1. Pengujian Dengan Data 80:20

Pada tahap pengujian ini memberikan hasil dari proses split data 80:20 ada pada 80% dari 968 data kelulusan mahasiswa serta 20% dari 193 data mahasiswa yang belum lulus.

	Real Cepat	Real Terlambat	Class prediction
pred Cepat	39	0	100%
pred Terlambat	0	0	0%
class real	100.0%	0.0%	

Gambar 1. Akurasi 80:20
[Sumber: pengelola data]

Hasil Akurasi Split data 80:20 yang telah menunjukkan hasil akurasi dari *metode Naive Bayes*. Rumusnya bisa dilihat pada dalam bagian bawah ini.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$Accuracy = \frac{39 + 0}{39 + 0 + 0 + 0} \times 100\%$$

$$Accuracy = \frac{39}{39} \times 100\%$$

$$Accuracy = 1$$

$$Accuracy = 100\%$$

Dengan *precision*nya juga mencapai 100% dapat dilihat hasil perhitungan ini.

$$Precision = \frac{TP}{TP + FP} \times 100\%$$

$$Precision = \frac{39}{39 + 0} \times 100\%$$

$$Precision = \frac{39}{39} \times 100\%$$

$$Precision = 1$$

$$Precision = 100\%$$

Serta *Recall*nya juga 100% dapat dilihat hasil perhitungan ini.

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

$$Recall = \frac{39}{39 + 0} \times 100\%$$

$$Recall = \frac{39}{39} \times 100\%$$

$$Recall = 1$$

$$Recall = 100\%$$

2. Pengujian Dengan Data 70:30

Pada tahap pengujian ini memberikan hasil dari proses split data 70:30 ada pada 70% dari 968 data kelulusan mahasiswa dan 30% dari data mahasiswa yang belum lulus. Hasil 100% mempengaruhi parameter yang dikategorikan sangat baik.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$Accuracy = \frac{58 + 0}{58 + 0 + 0 + 0} \times 100\%$$

$$Accuracy = \frac{58}{58} \times 100\%$$

$$Accuracy = 1$$

$$Accuracy = 100\%$$

Perhitungan *precision* berdasarkan rumus ini.

$$Precision = \frac{TP}{TP + FP} \times 100\%$$

$$Precision = \frac{58}{58 + 0} \times 100\%$$

$$Precision = \frac{58}{58} \times 100\%$$

$$Precision = 1$$

$$Precision = 100\%$$

Perhitungan *recall* berdasarkan rumus ini.

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

$$Recall = \frac{58}{58 + 0} \times 100\%$$

$$Recall = \frac{58}{58} \times 100\%$$

$$Recall = 1$$

$$Recall = 100\%$$

Dari hasil perhitungan tersebut terdapat 100% dari kedua rasio perbandingan yang dimana proses 80:20 dan 70:30 karena label keputusan Cepat lebih dominan.

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan dengan topik data mining pada proses klasifikasi dan prediksi lama studi mahasiswa di Undhira Bali terkhususnya tahun ajaran kelulusan 2019-2020 dapat diambil kesimpulannya sebagai berikut:

1. Implementasi metode Naive Bayes pada prediksi usang studinya mahasiswa Undhira berhasil dengan *data training* sebesar 968 data dan *data testing* 193 data bahwa mahasiswa dan mahasiswi Undhira studinya **Cepat** tidak Terlambat.
2. Hasil implementasi metode *Naive Bayes* dalam prediksi lama studinya mahasiswa Undhira berhasil dengan memperoleh tingkat keakuratan sebesar 100% dari *software Rapid Miner* yang juga parameternya tergolong sangat baik dari data mining dengan proses split data 80:20 dan 70:30 diatas, maka yang terbaik adalah proses split data 80:20 dan 70:30. Dari perhitungan yang dilakukan *Rapid Miner* dan perhitungan manual maka hasilnya tidak jauh berbeda.

DAFTAR PUSTAKA

- [1] Anggraini. M. "Implementasi Algoritma Naive Bayes Dalam

Penentuan Rating Buku", Jurnal sistemasi., vol. 9 no 3, pp. 557-566. 2020.

- [2] Tambunan. R. H. "Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Berdasarkan Kinerja Akademis Mahasiswa Menggunakan Algoritma Naive Bayes dengan Implementasi Data Mining Studi Kasus : Departemen Teknik Industri USU", Repositori institusi Universitas Sumatera Utara., 1-115. Internet: <http://repositori.usu.ac.id/handle/123456789/27825> 21 September, 2020 [Nov. 4, 2020] .
- [3] Silvi. M. "Penerapan Metode Naive Bayes Classifier Untuk Memprediksi Lama Masa Studi Mahasiswa", UIN Suska Riau Repositori., 1-100. Internet: <http://repository.uinsuska.ac.id/27494> 22 Juni, 2020 [Okto. 25, 2020].
- [4] Moonallika. P. S. "Penerapan Data Mining Untuk Memprediksi Kelulusan Mahasiswa Menggunakan Algoritma Naive Bayes", Jurnal Ilmiah Komputer., 16, 47-56.2020.
- [5] Ratniasih. N. L. "Optimasi Data Mining Menggunakan Algoritma Naive Bayes Dan C4.5 Untuk Klasifikasi Kelulusan Mahasiswa", Jurnal Teknologi Informasi dan Komputer., 28-34.2019.
- [6] Heryana. D. "Data mining untuk memprediksi kelulusan mahasiswa pendidikan matematika uin Raden Intan Lampung Menggunakan Naive Bayes", repository UIN Raden Intan Lampung, 1-104. Internet:<http://repository.radenintan.ac.id/6430/> 30 April, 2019 [Nov. 18, 2020] .
- [7] Fahrudy. D. "Penerapan Feature Selection Information Gain Pada Algoritma Naive Bayes Untuk Prediksi Kelulusan Mahasiswa (Studi Kasus: Teknik Informatika UIN SUSKA RIAU)", UIN Suska Riau Repositori, 1-110. Internet: <http://repository.uin-suska.ac.id/id/eprint/30623> 29 September, 2020 [Des. 18, 2020].
- [8] Siyoto. S. Dasar Metodologi Penelitian., yogyakarta: literasi media publishing, 2015.
- [9] Subawa. I. G. "Prediksi Kelulusan Mahasiswa Menggunakan Teorema Bayes (Studi Kasus di Universitas

Pendidikan Ganesha)", JANAPATI., vol. 8 no 3, pp. 227-236.2020.

- [10] Widaningsih. S. "Perbandingan Metode Data mining Untuk Prediksi Nilai Dan Waktu kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C45, Naive Bayes, KNN, Dan SVN", Tekno Insentif., vol. 13 no 1, pp 16-25.2019.