PREDIKSI MASA STUDI MAHASISWA MENGGUNAKAN ALGORITMA NAÏVE BAYES PADA UNIVERSITAS HALMAHERA

Agustinus A. Botara[1*)], Ahmad Sabri[2)]
[1]Universitas Gunadarma, Depok Jln. Margonda Raya. Depok, 16424, Indonesia
E-mail: agustinusbotara@gmail.com
[2]Universitas Gunadarma, Depok Jln. Margonda Raya. Depok, 16424, Indonesia
E-mail: ahd_sabri@yahoo.com

Abstract

Student graduation is an important aspect of accreditation at a university. One of the problems of concern is that the number of new students is not balanced with the number of graduates (graduating on time) the problem can lead to potential dropouts. From these problems it is necessary to conduct an analysis to predict the study period at Halmahera University. The main objective of this research is to predict student study period using Naïve Bayes algorithm, by utilising student graduation data from 2014-2017. The approach used involves the stages of data selection, data preprocessing, analytical processing, and output, with evaluation of prediction results using confusion matrix. Prediction of study period is done by data mining method and using Naïve Bayes algorithm to find patterns (knowledge). The data used is student graduation data in 2014-2017 with a total of 1,157 records, the data is divided into 926 records for training data (80%) and 231 records as testing data (20%). The prediction results show the accuracy rate of the Naïve Bayes Algorithm is 82.9%.

Keywords : *Masa Studi, Kelulusan, Prediksi, Naïve Bayes*

1. PENDAHULUAN

Teknologi informasi sangat berkembang pesat, baik dalam dunia industri maupun dalam dunia Pendidikan, teknologi informasi memiliki kemampuan dalam mengolah dan menganalisis data dengan jumlah yang besar. Di era revolusi industri sekarang ini hampir semua aktivitas bergantung pada perangkat-perangkat pintar maupun komputer. Perguruan tinggi juga sudah beradapatasi dengan perkembangan teknologi informasi, beberapa perguruan memiliki sistem informasi akademi. Salah satunya Universitas Halmahera, dalam sistem informasi data dimanfaatkan untuk pengambilan keputusan, data dengan jumlah yang besar perlu digali untuk mendapat informasi baru yang belum diketahui sebelumnya. Saat ini perguruan tinggi juga menyimpan data dengan jumlah besar dan data tersebut menjadi sumber daya. Selain itu sistem informasi juga menjadi pendorong kemajuan suatu perguruan tinggi. Kelulusan tepat waktu merupakan salah satu indikator aspek pembelajaran dalam peraturan Badan Akreditasi Nasional Perguruan Tinggi Nomor 3 tahun 2019 tentang instrumen Akreditasi Perguruan Tinggi. Namun berdasarkan data kelulusan Universitas Halmahera dalam kurung waktu 2014-2017 masi terdapat mahasiswa yang masa studinya lebih dari 4 tahun dan berpotensi dropout, untuk itu perlu dilakukan penelitian yang lebih jauh agar dapat mengetahui pola kelulusan dan hasilnya akan menjadi dasar untuk prediksi pada tahun berikutnya.

Prediksi masa studi telah banyak dilakukan penelitian, salah satunya penelitian yang di lakukan oleh (Peling et al., 2017) didapatkan 86% namun masi terdapat kekurangan karena jumlah data yang digunakan hanya 80 record. Dalam memprediksi masa studi mahasiswa metode yang digunakan yaitu data mining, dimana metode tersebut dapat digunakan dalam menggali data/mining dengan tujuan mencari informasi (pengetahuan) yang belum diketahui sebelummnya. Terdapat beberapa algoritma dalam data mining untuk kasus prediksi, namun

pada penelitian ini digunakan algoritma Naïve Bayes. Naïve Bayes merupakan salah satu algoritma data mining yang digunakan dalam pengklasifikasian. Metode pengklasifikasian dengan mengunakan metode probabilitas dan statistik, yaitu memprediksi peluang masa yang akan datang berdasarkan pengalaman sebelunya.

Beberapa peneltian terdahulu (Ahmed et al., 2018; Asril & Isa, 2020; Jananto et al., 2021; Supriyanto et al., 2020; Yunita et al., 2020) yang telah dilakukan untuk memprediksi kelulusan mahasiswa atau kinerja mahasiswa, penelitian yang dilakukan (Azahari et al., 2020) menggunakan algoritma Niave Bayes dan Neural Network penelitian ini bertujuan memprediksi kelulusan mahasiswa, atribut yang digunakan yaitu; umur saat masuk kuliah, klasifikasi kota asal, sekolah menengah atas, pekerjaan ayah, program studi, kelas jumlah saudara, dan indeks prestasi akademik (IPK). Setelah prediksi dilakukan menggukan algoritma Naïve Bayes diperoleh tingkat akurasi hanya 57,63% sedangkan dengan menggunakan model Neural Network jauh lebih tinggi yakni 72%,68%.

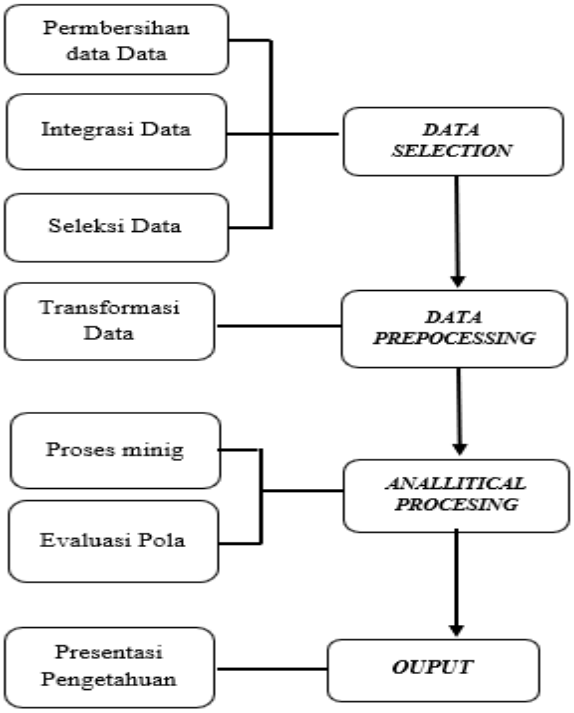
Penelitian yang dilakukan oleh (Hendra et al., 2020) memprediksi kelulusan mahasiswa tetapi menggunakan model Decision Tree dalam penelitian ini model *Decision Tree* saja tidak cukup untuk menghasilkan hasil yang optimal, untuk mendapatkan hasil yang optimal dengan model *Decision Tree* dibutuhkan metode pengoptimalan yaitu *Particle Swarm Optimization*. Hasil dari penelitian ini menunjukan model *Decision Tree* 86.55 % tingkat akurasi meningkat 01,01 % saat metode *Swarm Particle Optimization* di terapkan sehingga akurasi menjadi 86.56%.

Yulianto et al., (2020) melakukan penelitian untuk memprediksi kinerja siswa dengan menggunakan algoritma Decision tree c4.5 dan K-Nearest Neigbor. Hasil penelitian (Yulianto et al., 2020) menunjukan tingkat akurasi pada model K-Nearest Neigboar 59,32% sedangkan Decision Tree 54,80%. Selain itu, (Etriyanti et al., 2020) memprediksi kelulusan mahasiswa menggunakan algoritma Algoritma Naïve Bayes dan Algoritma C.45 dengan tujuan untuk mengetahui kinerja dari kedua algoritma dengan tingkat akurasi yang lebih besar. Hasil Penerapan kedua algoritma tersebut kemudian divalidasi menggunakan teknik k-Fold Cross Validation dan tahap evaluasi model dengan Coffusion Matrix digunakan untuk Validasi nilai akurasi hasil prediksi. Hasil Penelitin menunjukan metode algoritma C4.5 yang memiliki tingkat akurasi paling tinggi. Prediksi masa studi juga dilakukan oleh (Peling et al., 2017), hasil prediksi yang didapatkan tingkat akurasi 86% namun masi terdapat kekurangan karena jumlah data yang digunakan hanya 80 record.

Pada penelitian ini, menggunakan algoritma Naïve Bayes untuk memprediksi masa studi mahasiswa Universitas Halamhera data yang akan diprediksi angakatan 2014-2017 yang suda dinyatakan lulus. Atribut yang digunakan yaitu; pekerjaan orang tua, umur, asal sekolah, jurusan saat sekolah, program studi. Penelitian dilakukan dengan menambahkan jumlah dataset (1157 *record*) dan atribut atribut lain yang penelitinya sebelumnya belum gunakan dimana dengan penambahan jumlah dataset dan penggunaan atritubut baru diharapakan dapat memberikan pengatahuan baru, disinilah yang menjadi kebaruan penelitiannya.

2. METODE PENELITIAN

Metode yang digunakan dilakukan berdasarkan tahapan yang terdapat pada data minig yang disederhanakan oleh (Aggarwal, 2015) ditampilkan pada gambar 1.



Gambar 1 Metode Penelitian

2.1. Data *Selection*

Dalam tahapaan *data selection*, penulis akan mengumpulkan dataset mahasiswa Universitas Halmahera, data tersebut belum terdefinisi. Data yang diambil data kelulusan pada tahun 2014-2017 untuk diolah. Dalam tahapan ini juga ada beberapa langkah-langkah yang perlu dilakukan yaitu sebagai berikut:

a. Pembersihan Data
b. Integgrasi Data
c. Seleksi Data

2.2. Data *Prepocessing*

Setelah data mahasiswa dikumpulkan, tahapan selanjutnya dilakukan transformasi data agar data dapat dikenal oleh program komputervsebelum di mining, dari tahapan ini juga mencari kemungkinan ada data tidak konsisten, maka dari itu dilakukan data prepocesing agar dapat menghasilkan data berkualitas.

2.3. *Anallitical Procesing*

Tahapan ini adalah tahapan klasifikasi menggunakan Algoritma Naïve Bayes dan *confusion matrix* untuk melihat tingkat akurasi prediksi masa studi mahasiswa Universitas Halmahera. Beberapa langkah yang perlu dilakukan dalam tahapan ini yaitu sebagai berikut.

a. Proses Mining

Data yang telah dikumpulkan sesuai prosedur selanjutnya akan di mining menggunakan algoritma Naivey Bayes untuk menemukan pola serta pengetahuan.

b. Evaluasi Pola

Pada tahap ini evaluasi dilakukan dari hasil mining untuk mengetahui apakah algoritma bekerja secara baik atau tidak. Pada tahapan ini akan diterapkan metode confusion matrix

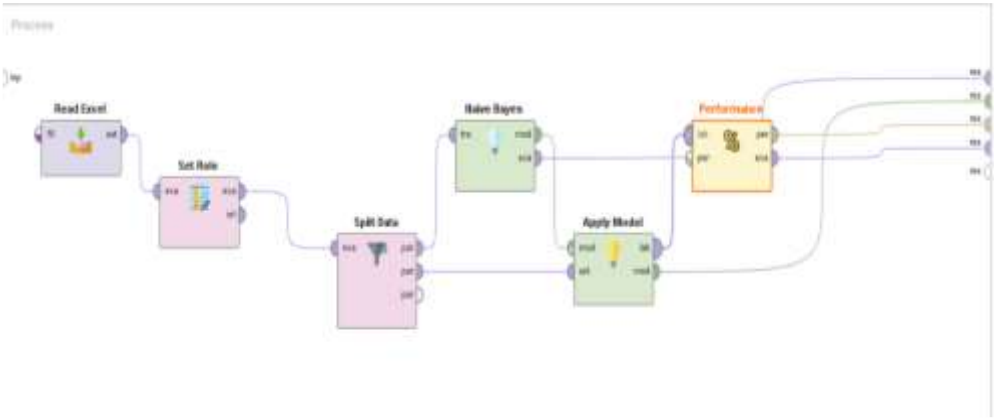
2.4. Output

Out menampilkan dari proses mining yang dilakukan dalam bentuk presentasi pengetahun (Visualisasi) dan output tersebut adalah hasil dari prediksi yang dilakukan.

3. HASIL DAN PEMBAHASAN

3.1. Hasil penelitian

Proses Klasifikasi

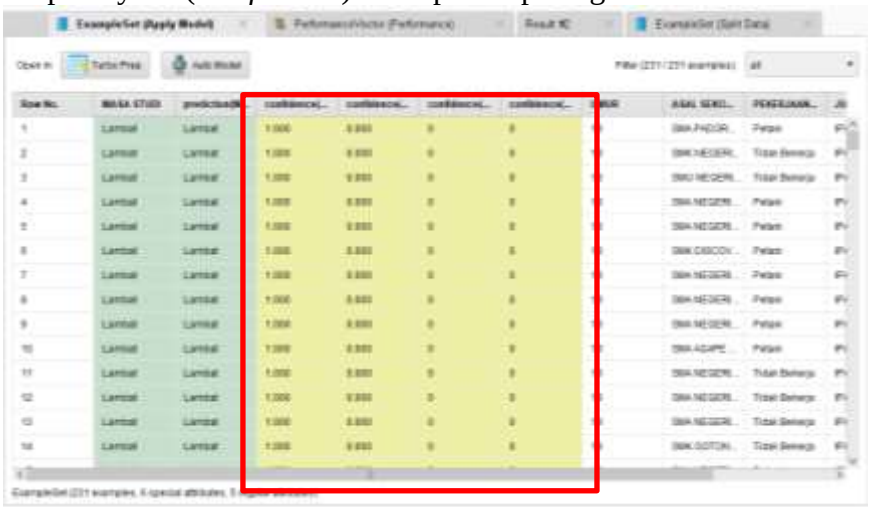


Gambar 2 Proses Klasifikasi

Dari gambar 2. diatas dapat dijelaskan beberapa operator-operator Rapidminer yang digunakan dalam memprediksi masa studi, seperti Read Exel, Split Role, Split Data, Naïve Bayes Apply Model dan Performance.

Apply Model

Hasil apply model pada data testing dengan jumlah 231 record, adalah hasil split data 20% dari 1.157 record data. Hasil tersebut menampilkan tingkat kepercayaan (*Confidence*) pada tiap record. Tingkat kepercayaan (*Confidence*) ditampilkan pada gambar 3



Gambar 3 Confidence

Perhitungan Confusion Matrix

Untuk membuktikan hasil pada Rapidminer sesuai dengan yang diharapkan, maka perlu dilakukan perhitungan confusion matix secara manual sebagai berikut :

$$Acuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (1)$$

TP : True Positive
TN : True Negative
True lambat : 52 orang
True Cepat : 105 orang
Ture Sangat Lambat : 13 orang
True Dropout : 20 orang

$$Acuracy = \frac{52 + 105 + 20 + 13}{52 + 105 + 20 + 13 + 21 + 1 + 2 + 3 + 14}$$

$$Acuracy = \frac{190}{231}$$

$$Acuracy = 0{,}822510823$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Precision\ Lambat = \frac{TP}{TP+FP} \quad (3)$$

$$Lambat = \frac{52}{52 + 21}$$

$$Lambat = \frac{52}{73} = 0{,}7123328767$$

$$Precision\ Cepat\ = \frac{TP}{TP+FP} \quad (4)$$

$$Cepat = \frac{105}{105 + 6}$$

$$Cepat = \frac{106}{111} = 0{,}945945946$$

$$Precision\ dropout = \frac{TP}{TP+FP} \quad (5)$$

$$dropout = \frac{20}{20 + 14}$$

$$dropout = \frac{20}{34} = 0{,}58822352994$$

$Precision\ Sangan\ lambat = \frac{TP}{TP+FP}$ (6)

$$sangat\ lambat = \frac{13}{13+0}$$

$$Sangat\ lambat = \frac{13}{13} = 1$$

$Recall\ = \frac{TP}{TP+FN}$ (7)

$Recall\ \ Lambat = \frac{TP}{TP+FP}$ (8)

$$lambat = \frac{52}{52+1}$$

$$lambat = \frac{20}{22} = 0{,}909090909$$

$Recall\ \ dropout = \frac{TP}{TP+FP}$ (9)

$$dropout = \frac{20}{20+2}$$

$$dropout = \frac{20}{22} = 0{,}909090909$$

$Recall\ \ Cepat = \frac{TP}{TP+FP}$ (10)

$$cepat = \frac{105}{105+140}$$

$$cepat = \frac{105}{245} = 0{,}75$$

$Recall\ \ Sangat\ Lambat = \frac{TP}{TP+FP}$ (11)

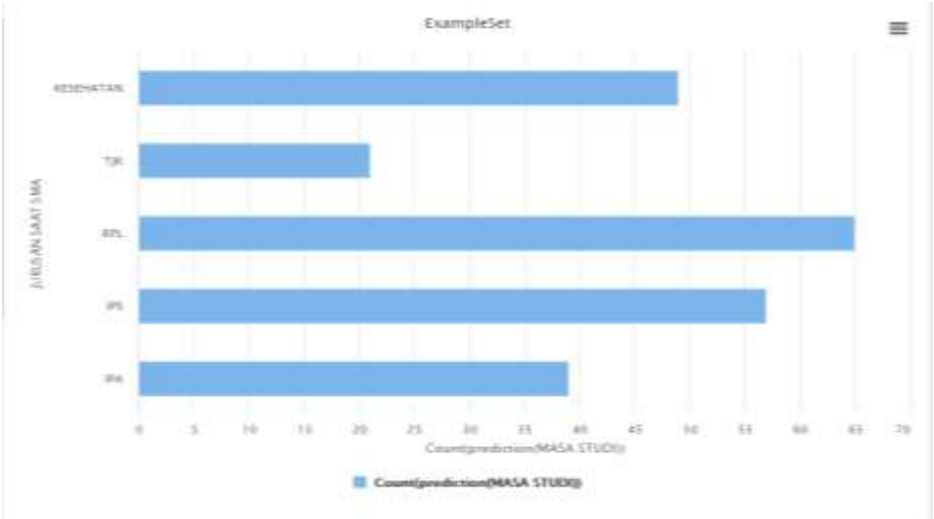
$$sangat\ lambat = \frac{13}{13+3}$$

$$Sangat\ lambat = \frac{13}{16} = 0{,}8125$$

Visualiasi

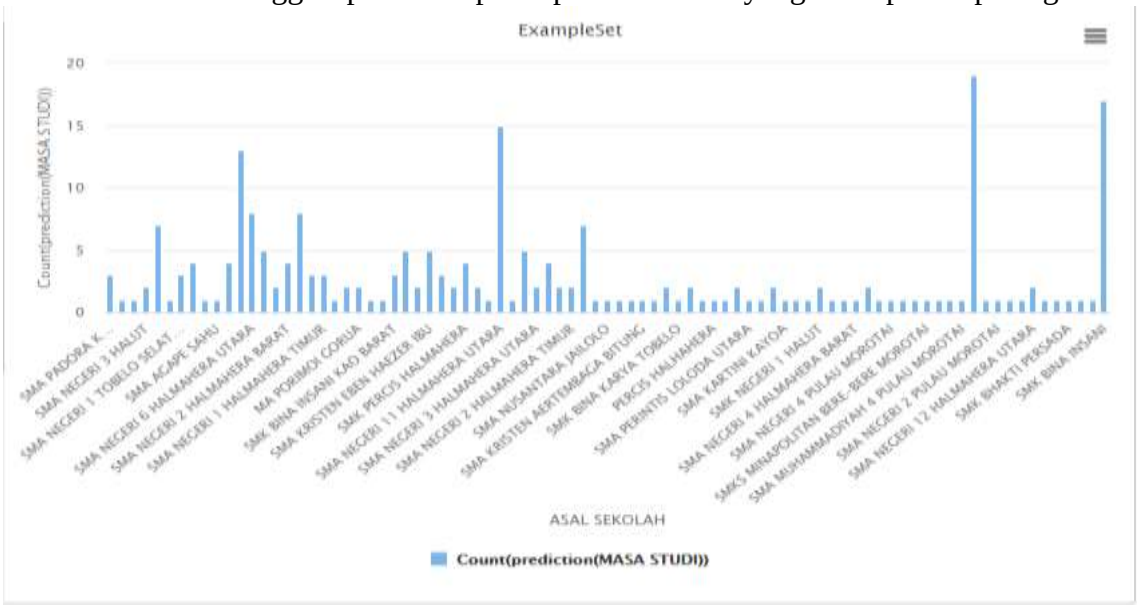
Jurusan saat di SMA merupakan pilihan yang diambil seorang siswa untuk menunjukan jurusan yang diminati. Sekolah Menengah Atas sebagai jenjang pendidikan yang wajib diikuti

sebelum melanjutkan ke jenjang perguruan tinggi. Dalam penelitian ini jurusan dijadikan sebagai atribut dengan kategori Kesehatan, TKJ, RPL, IPS dan IPA. Hasil prediksi menampilkan siswa yang mengambil jujursan RPL memiliki probabilitas kelulusan sangat tinggi dan yang kedua jurusan IPS. Visualisasi probabilitas kelulusan berdasarkan atribut jurusan dapat ditampilkan pada gambar 4.



Gambar 4 Visualisasi probabilitas kelulusan berdasarkan atribut jurusan

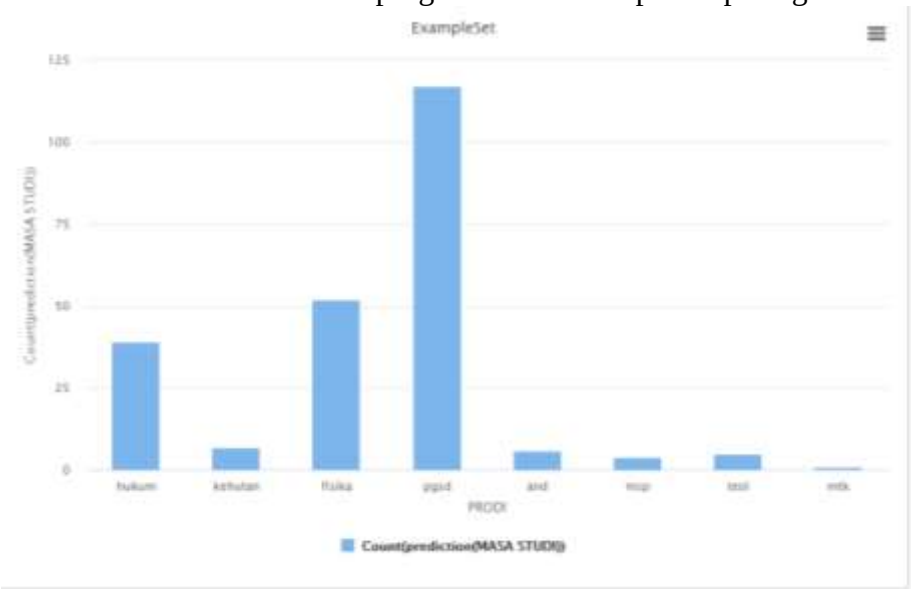
Asal sekolah menunjukan daerah, kabupaten/kota dari siswa saat menempuh pendidikan Sekolah Menengah Atas. Hasil prediksi menampilkan beberapa daerah asal sekolah memiliki probabilitas kelulusan tinggi seperti kabupaten pulau Morotai yang ditampilkan pada gambar 5.



Gambar 5 Visualiasi probabilitas kelulusan berdasarkan atribut asal sekolah.

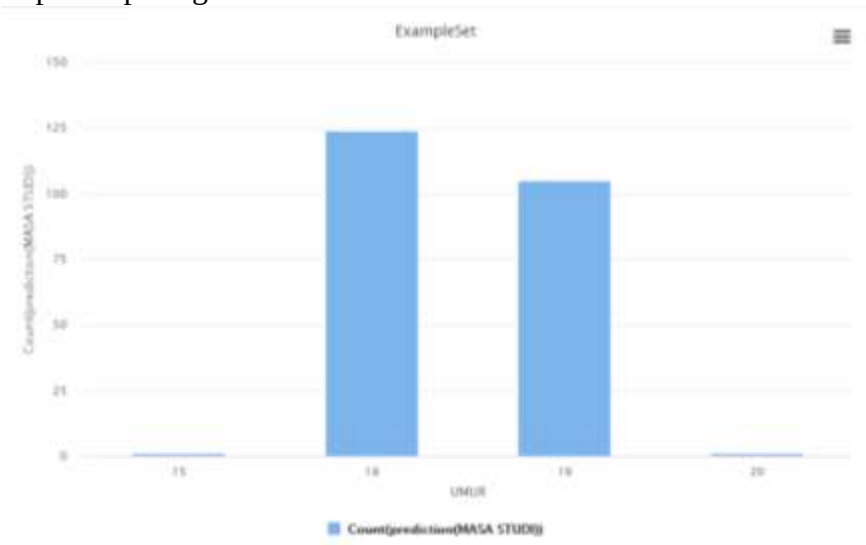
Terdapat beberapa program studi yang menjadi pilihan bagi mahasiswa yang menempuh pendidikan di Universitas Halmahera. Namun dari hasil prediksi manampilkan tidak semua jurusan memiliki probabilitas kelulusan yang sama. program studi Pendidikan Guru Sekolah Dasar (PGSD) salah satu program studi yang memiliki probabilitas tinggi nanum sebaliknya

program studi Matematika memiliki probabilitas kelulusan yang sangat rendah. Visualisasi probabilitas kelulusan berdasarkan atribut program studi ditampilkan pada gambar. 6



Gambar 6. Visualiasi probabilitas kelulusan berdasarkan atribut program studi.

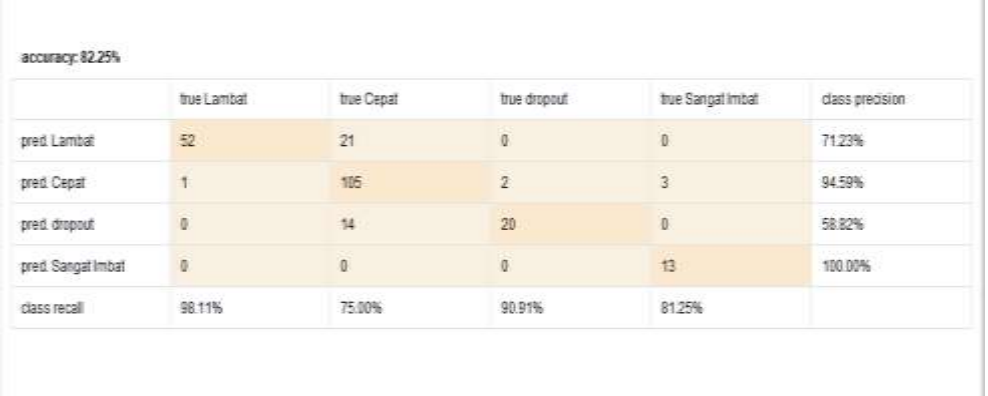
Dalam menempuh pendidikan faktor usia juga menjadi hal yang dapat mempengaruhi masa studi. Usia 18-19tahun memiliki probabilitas kelulusan yang hampir sama namun hasil prediksi menampilkan bahwa mahasiswa yang berumur 18 tahun (saat masuk perguruan tinggi) memiliki probabilitas kelulusan yang lebih tinggi. Visualisasi probabilitas kelulusan berdasarkan atribut umur dapat ditampilkan pada gambar 7.



Gambar 7. Visualiasi probabilitas kelulusan berdasarkan atribut umur

Hasil confusion matrix Menampilkam mahasiswa yang lulus dalam kategori cepat 105 (benar) dan salah prediksi 6 orang. Hasil prediksi untuk kategori mahasiswa yang lulus sangat lambat adalah 13 orang hasil atau 100%, untuk mahasiswa dalam kategori lambat hanya diprediksi 53 orang, tetapi pada kenyataanya total secara kesuluruhan ada 73 sedangkan untuk

mahasiswa yang termasuk dalam kategori dropout diprediksi 20 orang tetapi pada kenyataannya 34 orang droput. Hasil Confusion matrix ditampilkan pada 8.

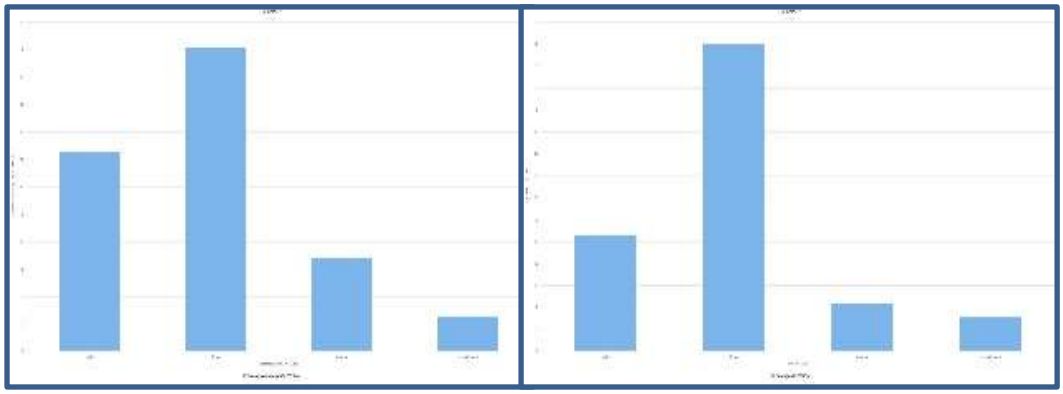


accuracy: 82.25%

	true Lambat	true Cepat	true dropout	true Sangat lmbat	class precision
pred. Lambat	52	21	0	0	71.23%
pred. Cepat	1	105	2	3	94.59%
pred. dropout	0	14	20	0	58.82%
pred. Sangat lmbat	0	0	0	13	100.00%
class recall	98.11%	75.00%	90.91%	81.25%	

Gambar 8 Confusion Matrix

Pebandingan hasil prediksi dan data (*real*) juga dilakukan untuk melihat seberapa besar peluang hasil prediksi. Grafik Perbandingan dapat dilihat pada gambar 9.



Gambar 9 Grafik Perbandingan masa studi yang sebenarnya (real) dan hasil prediksi

3.2. Pembahasan

Data mining

Data mining adalah proses menemukan korelasi, pola, dan tren yang baru dengan menggali/menambang (mining) data dengan jumlah yang besar yang disimpan dalam data werehouse, mengunakan statistic, *machine learning*, *Artficial Intelegence* dan teknik Visualiasi (Sumathi & Sivanandam, 2006) . Definisi lain tentang data mining juga ditambahkan oleh (Larose, 2015) data mining adalah proses menemukan pola dan tren yang bermanfaat dalam dataset yang besar. Data yang digunakan dalam peneltian ini merupakan data primer yang di ambil dari Sistem Informasi Akademik (SIA) Universitas Halmahera.

Tahapan Data mining

Tahapan -tahapan data mining menurut (Aggarwal, 2015):

a. *Data Collection*

Tahapan Pertama yang dilakaukan yaitu pengumpulan data, setelah itu data tersebut disimpan untuk diolah.

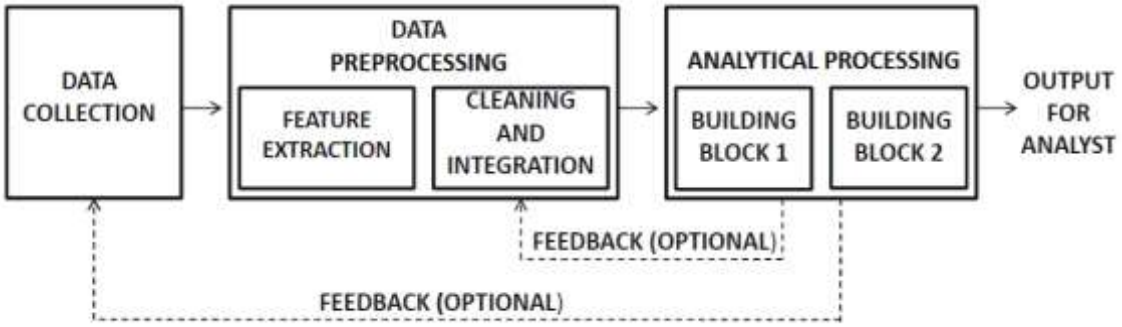
b. *Feature Extraxtions and cleaning*

Tahapan ini dilakukan untuk pembersihan data, jika terdapat data yang hilang atau salah dapat diperbaiki.

c. *Analititical Proccessing and Algorithms*

Bagian Akhir dari tahap data mining adalah mendesain metode analsis yang efektif dari data yang diolah.

Tahapan-tahapan tersebut dapat dilihat pada gambar 10.



Gambar 10 Tahapan Dalam Data mining
Sumber: (Aggarwal, 2015)

Klasifikasi

Menurut (Bustami, 2013) klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang lebelnya tidak diketahui.

Naïve Bayes

Naïve Bayes juga didefinsikan sebagai salah satu algoritma yang terdapat pada tehnik klasifikasi. Pengaklasifikasian dengan menggunakan metode probabilitas dan statistik untuk memprediksi peluang dimasa depan berdasarkan pengalaman di masa sebelumnya yang dikemukaan oleh Thomas Bayes. Persamaan dari teorama bayes dapat dilihat pada persamaan berikut (Bustami, 2013)

$$P\ (H|X) = \frac{p(X|H)P\ (H)}{P(X)} \qquad (12)$$

X : Data dengan kelas yang belum diketahui
H : Hipotesis data X merupakan suatu kelas (class) spesifik
P (H|X) : Probabilitas H berdasarkan kondisiX (Pasteriori Probability)
P(H) : Probabilitas Hipotesis H (prior probability)
P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H
P(X) : Probabilitas X

Tahapan Naïve Bayes

Adapun skema dalam metode Naïve Bayes menurut (Bustami, 2013) adalah sebagai berikut :

a. Baca data traning.
b. Hitung jumah dan probabilitas, namum apabila data numerik maka cari nilai mean dan standar deviasi dan nilai cari probabilistik.
c. Mendapatkan nilai dalam tabel.

Evaluasi Performance

Dalam data minig hasil dari model yang digunakan perlu dilakukan evaluasi dengan tujuan mengukur *performance* dari model tersebut (Naïve Bayes). Salah metode yang dikenal yaitu Confusion matrix. Dalam Confusion Matrix terdapat 3 perhitungan yang perlu dilakukan yaitu akurasi, presisi dan recall. Akurasi adalah metode yang digunakan untuk melakukan perhitungan tingkat akurasi dari model yang digunakan. Presisi adalah proporsi kasus yang diprediksi posifit yang juga positif benar pada data yang sebenarnya. Sedangkan recall proporsi kasus posifit yang sebernarnya yang diprediksi posifif secara benar (Rosandy, 2016).

$$Acuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (13)$$

$$Precision = \frac{TP}{TP+FP} \quad (14)$$

$$Recall = \frac{TP}{TP+FN} \quad (15)$$

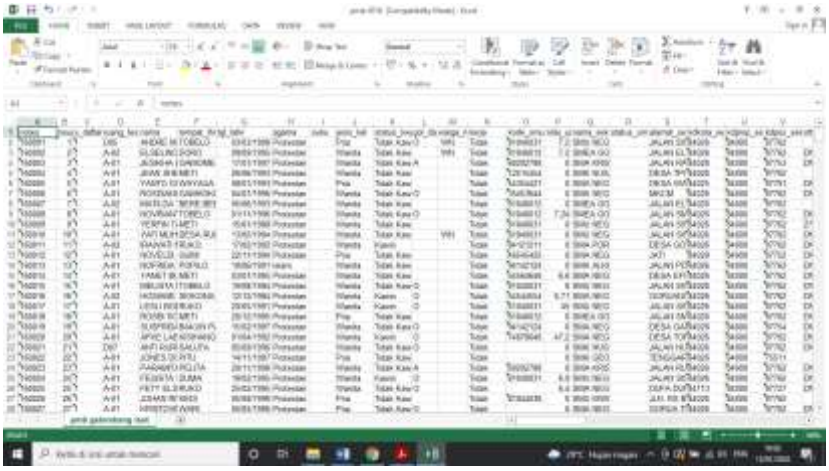
True Positive (*TP*): Prediksi dilakukan untuk data yang bernilai posifit dan itu benar posifit.
True Negative (*TP*): Prediksi dilakukan untuk data yang bernilai negatif dan hasilnya benar negatif
False Pofitive (*FP*): Prediksi dilakukan posifit dan itu salah
False Negative (*FN*): Prediksi dilakukan negatif dan itu salah.

Rapidminer

Rapidminer adalah perangkat lunak yang berfungsi sebagi alat pembelajaran dalam ilmu data, pembelajaran mesin, pembelajaran mendalam, penambangan teks dan analisis prediktif yang digunkan untuk bisnis dan komersial dan juga untuk penelitian, pendidikan, pelatihan, rapid prototyping dan pengembangan aplikasi serta mendudukung semua langkah- langkah dalam proses pembelajaran mesin termasuk persiapan data, hasil visualisasi, validasi model, dan optimasi (Hofmann & Klinkenberg, 2016)

Data

Data diambil dari Sistem Informasi Akademik Universitas berupa data kelulusan mahasiswa yang dimana data tersebut masi berupa kumpulan data yang belum terdefinisi secara baik. Selanjutnya data dikumpulkan untuk tahapan *Cleaning* atau pembersihan. data kelulusan mahasiswa pada tahun 2014-2017. Jumlah data yang akan digunakan setelah preprocesing sebanyak 1157 record dan akan di gunakan 926 untuk data traning dan 231 untuk data testing. Data dibagi dengan ratio 0.8 data traning dan 0.2 data testing. Daset yang belum sebelum procesing dapat ditampilkan pada gamabar.3



Gambar 11 Dataset Mahasiswa

Atribut dan Kelas

Atribut digunakan sebagai ciri atau karakter dalam membedahkan antara entitas yang satu dengan entitas yang lain. Dalam memprediksi masa studi mahasiswa atribut-atriut yang dibutuhkan yaitu umur saat masuk kuliah, asal sekolah, jurusan saat dibangku sekolah SMA/sederajat, dan pekerjaan orang tua dan program studi. Atribut dan kelas dapat ditampilkan pada tabel 1.

Tabel 1 Atribut dan Kelas yang dibutuhkan

Umur Kuliah	Asal Sekolah	Jurusan (SMA/Sederajat)	Prodi	Pekrjaaan Orangtua	Masa studi
20	Tobelo	RPL	Kehutan	Petani	Lambat
23	Tobelo	Kesehatan	Fisika	Petani	Cepat
20	Tobelo	IPA	Matematika	Petani	Sangat Lambat
18	Galela	IPS	Teologi	Petani	Cepat
18	Medan	IPS	Pemerintahan	Pegawai Negeri	Cepat
19	Tobelo	TKJ	ADN	Petani	Sangat Lambat

Dari tabel diatas dapat dijelaskan, nomor induk mahasiswa sebagai id, asal sekolah adalah kota/kabupaten tempat mahasiswa waktu dijenjang Sekolah Menengah Atas/sederajat, jurusan adalah minat yang diambil saat masih sekolah; seperti RPL, TKJ, IPS, IPA sedangkan pekerjaan orang tua dibagi lagi dalam beberapa kategori yaitu, Pegawai Negeri Sipil/PNS, TNI-Polri, Karywan Swasta ,Wiraswata, Petani dan lain-lain, dan Program studi adalah Program studi yang terdapat di universitas Halmahera.

Clas Masa studi di kelompokan sebagai berikut :

Cepat : 3,6-4 Tahun
Lambat : 4,6-5 Tahun
Sangat Lambat : 5,6- 6 Tahun
Dropout : dikeluarkan karena melebihi masa studi maksimal

4. KESIMPULAN

Penerapan algoritma Naïve Bayes dalam memprediksi masa studi mahasiswa dibagi dalam 4 (empat) kategori kelulusan yaitu cepat, lambat, sangat lambat dan droput. Evaluasi model dengan confusion matrix untuk mengetahui accuracy, recall dan precision dari algoritma Naïve Bayes. Berdasarkan prediksi yang telah dilakukan, maka dapat disimpulkan tingkat akurasi Naïve Bayes dalam memprediksi masa studi sebesar 82,25%. Dalam data mining, untuk kasus prediksi jumlah atribut yang digunakan, jumlah kelas yang gunakan, dan jumlah data sangat berpengaruh pada tingkat akurasi suata prediksi. Semakin besar jumlah data testing maka semakin baik tingkat kepercayaan (confidence) terhadapat suatu data yang diprediksi.

DAFTAR PUSTAKA

Aggarwal, C. C. (2015). *Data mining: the textbook* (Vol. 1). Springer.

Ahmed, R. M., Omran, N. F., & Ali, A. A. (2018). Predicting and analysis of students' academic performance using data mining techniques. *International Journal of Computer Applications*, *181*(48), 36–43.

Asril, T., & Isa, S. M. (2020). Prediction of students study period using K-Nearest Neighbor algorithm. *International Journal*, *8*(6).

Azahari, A., Yulindawati, Y., Rosita, D., & Mallala, S. (2020). Komparasi Data Mining Naive Bayes dan Neural Network memprediksi Masa Studi Mahasiswa S1. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *7*(3), 443–452.

Bustami, B. (2013). Penerapan algoritma Naive Bayes untuk mengklasifikasi data nasabah asuransi. *TECHSI-Jurnal Teknik Informatika*, *5*(2).

Etriyanti, E., Syamsuar, D., & Kunang, N. (2020). Implementasi Data Mining Menggunakan Algoritme Naive Bayes Classifier dan C4. 5 untuk Memprediksi Kelulusan Mahasiswa. *Telematika*, *13*(1), 56–67.

Hendra, H., Azis, M. A., & Suhardjono, S. (2020). Analisis Prediksi Kelulusan Mahasiswa Menggunakan Decission Tree Berbasis Particle Swarm Optimization. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, *9*(1), 102–107.

Hofmann, M., & Klinkenberg, R. (2016). *RapidMiner: Data mining use cases and business analytics applications*. CRC Press.

Jananto, A., Sulastri, S., Wahyudi, E. N., & Sunardi, S. (2021). Data Induk Mahasiswa sebagai Prediktor Ketepatan Waktu Lulus Menggunakan Algoritma CART Klasifikasi Data Mining. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, *10*(1), 71–78.

Larose, D. T. (2015). *Data mining and predictive analytics*. John Wiley & Sons.

Peling, I. B. A., Arnawan, I. N., Arthawan, I. P. A., & Janardana, I. G. N. (2017). Implementation of Data Mining To Predict Period of Students Study Using Naive Bayes

Algorithm. *Int. J. Eng. Emerg. Technol*, *2*(1), 53.

Rosandy, T. (2016). Perbandingan Metode Naive Bayes Classifier Dengan Metode Decision Tree (C4. 5) Untuk Menganalisa Kelancaran Pembiayaan (Study Kasus: KSPPS/BMT Al-Fadhila. *Jurnal Teknologi Informasi Magister*, *2*(01), 52–62.

Sumathi, S., & Sivanandam, S. N. (2006). *Introduction to data mining and its applications* (Vol. 29). Springer.

Supriyanto, A., Maryono, D., & Liantoni, F. (2020). Predicted student study period with C4. 5 data mining algorithm. *IJIE (Indonesian Journal of Informatics Education)*, *4*(2), 94–100.

Yulianto, L. D., Triayudi, A., & Sholihati, I. D. (2020). Implementation Educational Data Mining For Analysis of Student Performance Prediction with Comparison of K-Nearest Neighbor Data Mining Method and Decision Tree C4.5. *Jurnal Mantik*, *4*(1), 441–451.

Yunita, Y., Puspita, A., & Lestari, A. F. (2020). Student Performance Analysis Using C4. 5 Algorithm To Optimize Selection. *Jurnal Pilar Nusa Mandiri*, *16*(2), 149–154.