

Prediksi Siswa SMK Al-Hidayah yang Masuk Perguruan Tinggi dengan Metode Klasifikasi

Khoirunnisa¹, Lia Susanti², Ira Tasfiyyutu Rokhmah³, Lilis Stianingsih⁴

^{1,2,3,4} Universitas Budi Luhur

Program Studi Magister Ilmu Komputer, Fakultas Pascasarjana, Universitas Budi Luhur
Jl.Raya Ciledug, Petukangan Utara, Kebayoran Lama, Jakarta Selatan 12260, Indonesia

e-mail: ¹khoirunnisa.1797@gmail.com, ²liasusanti.s4a.061@gmail.com, ³tasfy189.it@gmail.com,
⁴lilisstianingsih@gmail.com

Informasi Artikel Diterima: 16-10-2020 Direvisi: 11-11-2020 Disetujui: 15-12-2020

Abstrak

Salah satu aspek sebagai indikator kualitas sekolah menengah kejuruan adalah tingkat diterimanya siswa di perguruan tinggi baik itu negeri maupun swasta. Beberapa data siswa sekolah menengah kejuruan dianalisis untuk mengetahui tingkat diterimanya siswa di perguruan tinggi melalui penelusuran siswa. Dengan memprediksi siswa yang masuk perguruan tinggi, bisa dimanfaatkan oleh pihak sekolah untuk bahan promosi. Proses analisis data siswa tersebut menggunakan teknik *data mining*. Dengan tujuan penelitian ini untuk memprediksi siswa angkatan berikutnya yang masuk di perguruan tinggi negeri maupun swasta menggunakan hasil model klasifikasi yang terbentuk. Pada penelitian ini data yang digunakan adalah data siswa angkatan 2018 dan 2019 Sekolah Menengah Kejuruan AL-Hidayah 1. Yang berjumlah 503 data, setelah dilakukan pre-processing diperoleh data yang siap diolah berjumlah 158 data. Proses *data mining* dibantu oleh software Rapid Miner menggunakan klasifikasi *naïve bayes*. Proses evaluasi data membandingkan beberapa algoritma lain yaitu *Decision Tree* dan KNN. Dan dari hasil pengujian tingkat akurasi tertinggi diantara tiga algoritma yang telah diuji diantaranya *Decision Tree* dengan nilai akurasi 95.60%, sedangkan *naive bayes* dengan nilai akurasi 92.40% dan KNN nilai akurasinya 94.96%.

Kata Kunci: *Data Mining*, Klasifikasi, *Naïve Bayes*

Abstract

One aspect as an indicator of the quality of vocational high schools is the level of acceptance of students in both public and private universities. Some vocational high school student data were analyzed to determine the level of acceptance of students in higher education through student tracing. By predicting students who will enter college, the school can use it for promotional materials. The process of analyzing student data uses data mining techniques. With the aim of research to predict the next generation of students who will enter public or private universities using the results of the classification model that is formed. In this study, the data used were student data from the 2018 and 2019 batches of the AL-Hidayah 1 Jakarta Vocational High School. There are 503 data, after pre-processing, there are 158 data ready to be processed. The data mining process is assisted by Rapid Miner software using the naïve Bayes classification. The data evaluation process compares several other algorithms, namely Decision Tree and KNN. And the test results are accurate among the three algorithms that have been tested, including the Decision Tree with an accuracy value of 95.60%, while the naïve Bayes with an accuracy value of 92.40% and the KNN with an accuracy value of 94.96%.

Keywords: *Data Mining*, Classification, *Naïve Bayes*

1. Pendahuluan

Pemanfaatan teknologi komputer dapat menjadi sumber informasi penting bagi sekolah, salah satunya informasi penelusuran siswa. Informasi tersebut akan berguna bagi

sekolah dalam mendapatkan data siswa yang memilih melanjutkan pendidikan tinggi negeri atau swasta, sehingga dari data tersebut sekolah dapat mengukur kualitas pendidikan yang diberikan pada siswa selama tiga tahun



belajar di SMK. Kualitas pendidikan pada siswa merupakan salah satu tujuan dalam mempromosikan sekolah agar mendapatkan kesan yang positif dan dapat dikenal oleh masyarakat luas ataupun calon siswa baru. Aspek dalam indikator kualitas di sekolah adalah tingkat kelulusan siswa dan banyaknya lulusan siswa sekolah menengah kejuruan yang melanjutkan studi ke perguruan tinggi negeri maupun perguruan tinggi swasta, juga menunjukkan kualitas sekolah dalam memotivasi siswa untuk memberikan jaminan kesempatan kerja yang lebih baik di masa yang akan datang.

Banyak data siswa yang dimiliki sekolah dari data profil siswa sampai nilai siswa memiliki informasi yang tersembunyi didalamnya, jika dikelola secara benar untuk kebutuhan sekolah. (Sidik et al., 2018). Tahun 2018 jumlah data siswa lulus pada SMK AL-Hidayah 1 kurang lebih sebanyak 250 siswa yang terdiri dari empat jurusan. Data tersebut belum dimanfaatkan untuk menggali sebuah informasi dalam memprediksi siswa yang melanjutkan studi ke perguruan tinggi negeri dan perguruan tinggi swasta. Teknik yang digunakan dalam memprediksi siswa yang melanjutkan studi perguruan tinggi negeri dan swasta ini menggunakan teknik klasifikasi dengan metode algoritma *naive bayes*. Sehingga informasi yang dihasilkan dari *data mining* ini akan dijadikan sebagai informasi untuk pihak sekolah SMK AL-Hidayah 1 dalam mengarahkan, membimbing, serta memotivasi siswanya agar melanjutkan studi ke perguruan tinggi dan juga digunakan untuk mempromosikan sekolah agar dapat meningkatkan kualitas SMK AL-Hidayah 1.

Permasalahan yang dapat diidentifikasi pada penelitian ini karena belum adanya suatu pola untuk memprediksi siswa yang akan melanjutkan studi ke perguruan tinggi antara negeri atau swasta di SMK AL-Hidayah 1 Jakarta. Dan faktor-faktor yang berpengaruh terhadap siswa dalam menentukan pilihan melanjutkan studi ke perguruan tinggi negeri atau swasta. Dari permasalahan tersebut perlu adanya suatu batasan untuk memberikan kemudahan dalam praktek dilapangan. Batasan masalah pada penelitian ini, yaitu data yang digunakan adalah data siswa tahun 2018 dan 2019 yang diperoleh dari bagian kurikulum, bagian tata usaha dan bagian bimbingan konseling Sekolah Menengah Kejuruan AL-Hidayah 1. Dalam pengolahan *data mining* ini peneliti menggunakan teknik klasifikasi dengan algoritma *naive bayes* dan membandingkan beberapa algoritma agar dapat mengetahui serta mengukur model untuk memprediksi penelitian ini dengan

algoritma dengan memiliki angka akurasi yang tinggi.

Dari penelitian ini dapat dibandingkan dengan beberapa penelitian sebelumnya antara lain penelitian yang dilakukan tentang Klasifikasi Profile Siswa SMA/SMK yang masuk PTN (Perguruan Tinggi Negeri) dengan *K-Nearest Neighbor* menggunakan 16 atribut yang terdiri dari profil siswa, algoritma yang di pakai yaitu *K-NN* dan memakai *WEKA* untuk melakukan proses *data mining* (Trisaputra, 2016). penelitian yang dilakukan tentang Sistem Prediksi Penerimaan SNMPTN menggunakan Algoritma *Decision Tree* C4.5. Sistem dibangun dalam bentuk website yang memanfaatkan *WEKA* CLI untuk proses prediksi (Tree et al., 2019). Penerapan teknik *data mining* untuk menentukan hasil seleksi masuk SMAN 1 Gibeber untuk siswa baru menggunakan *Decision Tree* dalam penelitian tersebut Uji yang dilakukan menggunakan 3 (tiga) jenis Algoritma, yaitu Algoritma C4.5 (*Decision Tree*), *Naive Bayes* dan *Naural Network*, dan akurasi tertinggi adalah algoritma C4.5 (Sugianto, 2017). penelitian yang dilakukan tentang seleksi calon mahasiswa pada penerima mahasiswa baru di Universitas pamulang. Menggunakan metode *Support Vector Machine* (SVM) untuk dataset yang digunakan terdiri dari enam atribut (Saifudin, 2018). penelitian yang dilakukan tentang klasifikasi data peserta didik menggunakan metode klasifikasi algoritma *Decision Tree* membuktikan dengan metode ini cukup mudah digunakan nilai akhir siswa sebagai atributnya. (Sutoyo, 2018). penelitian yang dilakukan tentang klasifikasi sumber belajar berbasis teks pada mata pelajaran produktif. Dalam pengklasifikasian teks menghasilkan sembilan kelompok mata pelajaran produktif dan pengujian menghasilkan nilai akurasi tertinggi 81,48%, sedangkan nilai akurasi terendah sebesar 79,63. Penelitian ini mengenai prediksi siswa yang masuk perguruan tinggi Negeri dan Swasta dengan algoritma *Naive Bayes* dan melakukan uji coba dengan tiga algoritma (Dwi Herlambang & Hadi wijoyo, 2019). penelitian yang dilakukan tentang Prediksi kemampuan lulusan SMK untuk dapat bersaing didunia kerja dengan menggunakan *Naive Bayes*. Dalam mengetahui pola dari variabel-variabel tersebut. Pengujian dilakukan dengan *Confusion Matrix* dan kurva ROC. Hasil akurasi yang diperoleh membuktikan bahwa *Naive Bayes* menghasilkan persentase akurasi 98% dan nilai AUC sebesar 0.980 (Kusuma, 2019).

Tujuan dari penelitian ini adalah Memprediksi siswa SMK Al Hidayah 1 yang melanjutkan studi Keperguruan Tinggi Negeri

atau Swasta. Kemudian mengetahui faktor apa saja yang mempengaruhi siswa dalam memilih untuk melanjutkan studi. Sehingga dapat digunakan sebagai acuan dalam mempromosikan sekolah dan meningkatkan kualitas sekolah agar lebih baik. Dan dapat diterapkan serta dijadikan suatu pola baru dalam memprediksi siswa dalam melanjutkan studi kejenjang yang lebih tinggi, sehingga dapat dikembangkan lagi dalam salah satu dasar meningkatkan suatu kualitas siswa siswi yang ada disekolah.

2. Metode Penelitian

A. Data Mining

Data mining adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Informasi yang dihasilkan dapat diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat dalam basis data (Nofriansyah et al., 2016).

Data mining juga banyak meliputi banyak bidang ilmu contohnya *artificial intelligent, machine learning, statistic* dan basis data (Annur, 2019). *Data mining* juga digunakan untuk menentukan pola serta informasi dalam suatu data (Mardi, 2017).

Secara garis besar *Knowledge Discovery in Database (KDD)* dapat dijelaskan sebagai berikut (Saikin & Kusri, 2019).

1. Data Selection

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD di mulai. Data hasil seleksi yang akan digunakan untuk proses *data mining*, disimpan dalam suatu berkas terpisah dari basis data operasional.

2. Pre-Processing/Cleaning

Proses *cleaning* antara lain membuang duplikasi data, memeriksa data yang inkonsisten dan memperbaiki kesalahan pada data. Pada proses ini dilakukan juga proses *enrichment*, yaitu proses memperkaya data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD.

3. Transformation

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining*.

4. Data Mining

Data Mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu.

5. Interpretation/Evaluation

Pola informasi yang dihasilkan dari proses *data mining* diterjemahkan menjadi bentuk yang lebih mudah dimengerti oleh pihak yang berkepentingan.

B. Klasifikasi

Dalam klasifikasi kita dapat menentukan orang atau objek kedalam suatu kategori tertentu, contoh untuk masalah klasifikasi adalah menentukan apakah seseorang pasien “mengidap” atau “tidak mengidap” penyakit tertentu. Informasi tentang pasien sebelumnya digunakan sebagai bahan untuk melatih algoritma untuk mendapatkan *rule* atau aturan. Salah satu tujuan klasifikasi adalah untuk meningkatkan kehandalan hasil yang diperoleh dari data (Kahramanli & Allahverdi, 2008).

C. Algoritma Naive Bayes

Algoritma *naive bayes* adalah salah satu algoritma yang terdapat pada teknik *data mining* klasifikasi. Algoritma *naive bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuan Inggris yaitu Thomas bayes, *naive bayes* memprediksi peluang dimasa depan berdasarkan pengalaman dimasa sebelumnya, sehingga dikenal dengan teorema bayes. Persamaan dari teorema bayes adalah (Kusuma, 2019): Penelitian dengan menggunakan *naive bayes* didasarkan pada nilai probabilitas (Junaidi et al, 2020

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad (1)$$

Keterangan :

X : Data dengan *class* yang belum diketahui

H : Hipotesis data X merupakan suatu *class* spesifik

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

$P(H)$: Probabilitas hipotesis H (*posteriori probability*)

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

$P(X)$: Probabilitas X

Berikut ini merupakan contoh kasus mengenai perhitungan *naive bayes*, kasus dituangkan dalam Tabel 1. Berdasarkan tabel 1 dapat dihitung klasifikasi data nasabah dapat ditentukan melalui langkah berikut :

1. Menghitung jumlah *class/label*.

$P(Y=Lancar) = 3/7$ “Jumlah data lancar pada data pelatihan dibagi dengan jumlah keseluruhan data”

$P(Y=\text{Kurang Lancar}) = 1/7$ "Jumlah data kurang lancar pada data pelatihan dibagi dengan jumlah keseluruhan data"
 $P(Y=\text{Tidak Lancar}) = 3/7$ "Jumlah tidak lancar pada data pelatihan dibagi dengan jumlah keseluruhan data".

Tabel 1 Klasifikasi Asuransi

JKL	Usia	Gaji	Masa Asuransi	Klasifikasi
L	30-40	25-50 Juta	>15	Tidak Lancar
P	30-40	< 25 Juta	5-10	Lancar
P	20-29	< 25 Juta	5-10	Tidak Lancar
P	30-40	< 25 Juta	5-10	Lancar
L	30-40	< 25 Juta	5-10	Kurang Lancar
L	30-40	> 50 Juta	11-15	Lancar
L	30-40	>25 Juta	11-15	Tidak Lancar
L	30-40	25-50 Juta	11-15	???

Sumber: Data Olah Pribadi

2. Menghitung jumlah kasus yang sama dengan class yang sama.

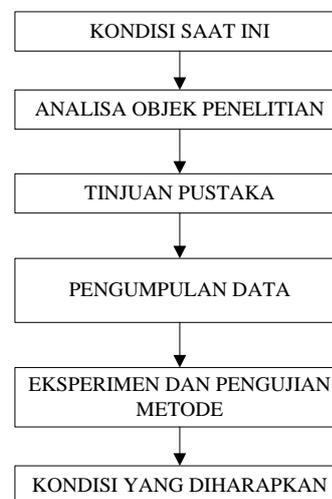
- $P(\text{Jenis Kelamin} = \text{Laki-laki} | Y=\text{Lancar}) = 1/3.$
- $P(\text{Jenis Kelamin} = \text{Laki-laki} | Y=\text{Kurang Lancar}) = 1/1.$
- $P(\text{Jenis Kelamin} = \text{Laki-laki} | Y=\text{Tidak Lancar}) = 2/3.$
- $P(\text{Usia} = 30 - 40 \text{ Tahun} | Y=\text{Lancar}) = 3/3.$
- $P(\text{Usia} = 30 - 40 \text{ Tahun} | Y=\text{Kurang Lancar}) = 1/1.$
- $P(\text{Usia} = 30 - 40 \text{ Tahun} | Y=\text{Tidak Lancar}) = 2/2.$
- $P(\text{Penghasilan} = 25 - 50 \text{ Juta} | Y=\text{Lancar}) = 0/3.$
- $P(\text{Penghasilan} = 25 - 50 \text{ Juta} | Y=\text{Kurang Lancar}) = 0/1.$
- $P(\text{Penghasilan} = 25 - 50 \text{ Juta} | Y=\text{Tidak Lancar}) = 1/2.$
- $P(\text{Masa_Asuransi} = 11 - 15 \text{ Tahun} | Y=\text{Lancar}) = 1/3.$
- $P(\text{Masa_Asuransi} = 11 - 15 \text{ Tahun} | Y=\text{Kurang Lancar}) = 0/1.$
- $P(\text{Masa_Asuransi} = 11 - 15 \text{ Tahun} | Y=\text{Tidak Lancar}) = 1/2.$

3. Kalikan semua hasil variabel Lancar, Kurang Lancar dan Tidak Lancar .
 $P(\text{Laki-Laki} \setminus \text{Lancar}) * P(30 - 40 \text{ Tahun} \setminus \text{Lancar}) * P(25-50 \text{ Juta} \setminus \text{Lancar}) * P(11 - 15 \text{ Tahun} \setminus \text{Lancar}) * P(\text{Lancar})$
 $= 1/3 \times 3/3 \times 0/3 \times 1/3 \times 3/7$
 $= 0,3333 \times 1 \times 0 \times 0,3333 \times 0,4285$
 $= 0$
 $P(\text{Laki - Laki} \setminus \text{Kurang Lancar}) * P(30 - 40 \text{ Tahun} \setminus \text{Kurang Lancar}) * P(25 - 50 \text{ Juta} \setminus \text{Kurang Lancar}) * P(11 - 15 \text{ Tahun} \setminus \text{Kurang Lancar}) * P(\text{Kurang Lancar})$
 $= 1/1 \times 1/1 \times 0/1 \times 0/1 \times 1/7$
 $= 1 \times 1 \times 0 \times 0 \times 0,1428$
 $= 0$
 $P(\text{Laki - Laki} \setminus \text{Tidak Lancar}) * P(30 - 40 \text{ Tahun} \setminus \text{Tidak Lancar}) * P(25 - 50 \text{ Juta} \setminus \text{Tidak Lancar}) * P(11 - 15 \text{ Tahun} \setminus \text{Tidak Lancar}) * P(\text{Tidak Lancar})$
 $= 2/3 \times 2/2 \times 1/2 \times 1/2 \times 3/7$
 $= 0,6666 \times 1 \times 0,5 \times 0,5 \times 0,4285$
 $= 0,0714$

Bandingkan hasil class Lancar, Kurang Lancar dan Tidak Lancar dari hasil di atas, terlihat bahwa nilai probabilitas tertinggi ada pada kelas (P| Tidak Lancar) sehingga dapat disimpulkan bahwa status calon nasabah tersebut masuk dalam klasifikasi "Tidak Lancar".

D. Kerangka Berfikir

Berdasarkan identifikasi masalah, tujuan penelitian dan kajian teori, studi dari penelitian sebelumnya, tinjauan obyek penelitian, maka dapat dibangun kerangka konsep penelitian. Peneliti akan menjelaskan langkah-langkah dalam memecahkan masalah dan metode yang digunakan. Berikut ini kerangka penelitian yang dibuat oleh penulis pada Gambar 1.



Sumber: Data Olah Pribadi

Gambar 1 Kerangka berfikir

Penjelasan pada Gambar 1 Kerangka berfikir, bahwa kondisi saat ini SMK AL Hidayah belum ada metode untuk memprediksi siswa yang akan melanjutkan kuliah di PTN (Perguruan Tinggi Negeri) ataupun di PTS (Perguruan Tinggi Swasta), kemudian dalam analisa objek penelitian ini dilakukan dengan mengumpulkan data siswa dan wawancara pihak sekolah yaitu kepala sekolah. Kemudian bagian tinjauan pustaka diperoleh dari berbagai sumber seperti jurnal ilmiah yang berkaitan dengan penelitian ini. Bagian pengumpulan data bersumber dari siswa SMK AL Hidayah dengan jumlah data sebanyak 158 *record* yang terdiri dari 9 atribut. Lalu pada bagian eksperimen dan pengujian data menggunakan rapid miner untutuk mendapatkan hasil AUC pada setiap algoritma yang sedang diuji. Kemudian kondisi yang diharapkan dalam mempromosikan SMK AL Hidayah agar dapat dikenal lebih luas oleh masyarakat.

E. Teknik Analisis

Proses Analisis dilakukan terhadap hasil tahapan pengumpulan data dengan wawancara, observasi dan studi pustaka. (Dwi Arum Ningtyas, 2019). Pada proses analisis yang akan dilakukan dengan analisis data dan informasi yang di dapatkan dari wawancara, observasi, dan studi pustaka. Dalam penelitian ini teknik klasifikasi *data mining* menggunakan algoritma *naive bayes*.

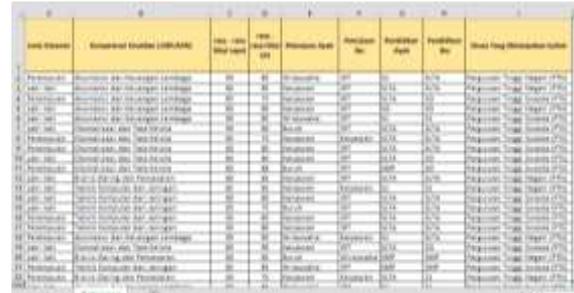
F. Pengujian Model

Data yang digunakan dalam percobaan ini adalah *data training* untuk mencari nilai *error rate* secara keseluruhan. Secara umum pengujian nilai k dilakukan sebanyak 10 kali untuk memperkirakan akurasi estimasi. Dalam penelitian ini nilai k yang digunakan berjumlah 10 atau *10-fold Cross Validation* (Suryadi, 2016)

3. Hasil dan Pembahasan

Pada bagian ini, akan dijelaskan hasil penelitian yang sudah di uji coba dengan model prediksi sesuai dengan pembahasan yang ada di bab sebelumnya mulai dari *cleaning* data sampai pengujian akurasi di aplikasi Rapid Miner (Nofriansyah et al., 2016). Penelitian ini menggunakan data siswa dengan jumlah 158 data.

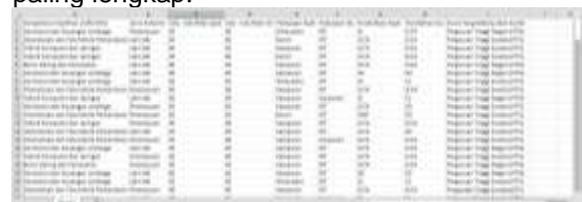
1. *Cleaning* Data



Sumber: Data Olah Pribadi

Gambar 2 *Cleaning* Data

Pada Gambar 2 *Cleaning* Data. Pada data penelusuran siswa yang didapat sebelumnya kurang lebih 503 data angkatan tahun 2018/2019. Data siswa ini di *cleaning* terlebih dahulu, karena banyak data yang tidak lengkap sehingga tidak bisa diikuti sertakan untuk penelitian lebih lanjut. Tujuan *cleaning* data agar peneliti dapat data sesuai atau lengkap untuk digunakan pada penelitian ini. Data siswa alumni ini digunakan untuk mengetahui informasi mereka setelah lulus dari SMK. Ada sebagian yang melanjutkan kuliah, dan ada yang sudah memiliki pekerjaan. Data yang diambil yaitu data siswa yang melanjutkan kuliah baik di PTN (Perguruan Tinggi Negeri) maupun di PTS (Perguruan Tinggi Swasta). Kemudian data yang terpilih tersebut adalah data siswa yang paling lengkap.

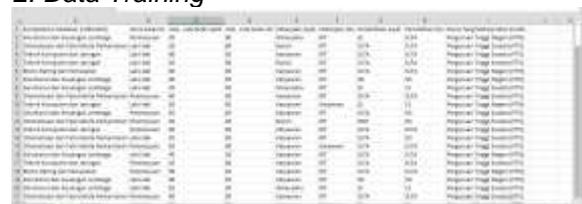


Sumber: Data Olah Pribadi

Gambar 3 Data Yang Sudah Di *Cleaning*

Pada gambar 3. Data yang sudah di *cleaning*, sebanyak 158 data kemudian, atribut yang dimiliki sebanyak 9, antara lain kompetensi keahlian (Jurusan siswa), Jenis Kelamin, Rata-rata nilai rapot, Rata-rata nilai UN, Pekerjaan ayah ibu, Pendidikan terakhir ayah ibu, dan siswa yang melanjutkan kuliah.

2. *Data Training*

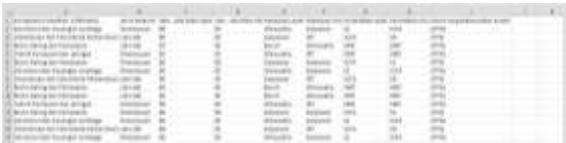


Sumber: Data Olah Pribadi

Gambar 4 *Data Training*

Pada Gambar 4 *data training* merupakan hasil data siswa yang sudah di *cleaning* dari data sebelumnya. Terdiri dari 158 data penelusuran siswa di SMK AL-Hidayah 1.

3. Data Testing



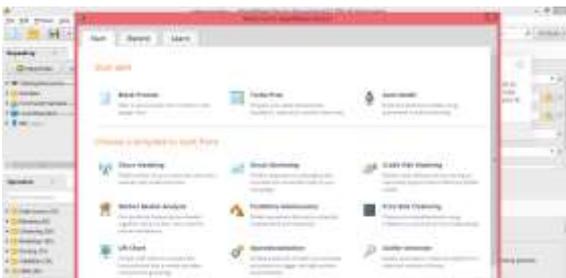
Sumber: Data Olah Pribadi

Gambar 5 Data Testing

Pada Gambar 5 *data testing* terdiri dari 14 data yang meliputi 9 atribut. Data testing dipakai untuk mengetahui performa algoritma yang sudah dilatih sebelumnya, untuk mendapatkan informasi baru.

4. Aplikasi RapidMiner

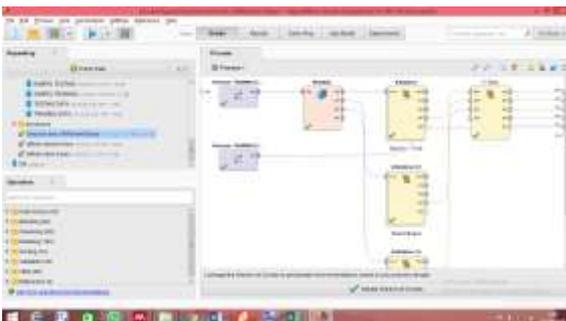
Dalam gambar 6 pengujian data, peneliti menggunakan aplikasi rapidminer studio versi 9.7 sebagai tools yang akan membantu mengukur tingkat akurasi dari algoritma yang sedang diuji.



Sumber: Data Olah Pribadi

Gambar 6 Pengujian data

5. Pengujian Data



Sumber: Data Olah Pribadi

Gambar 7 Pengujian Data

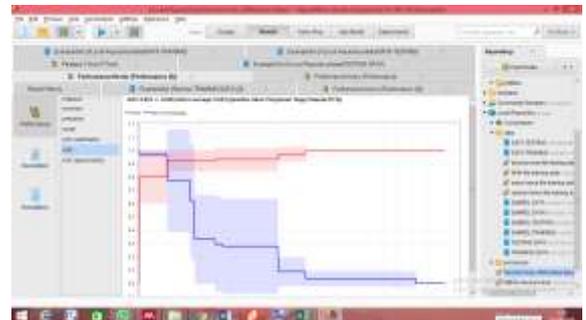
pada gambar 7 Pengujian data, bahwa peneliti menggunakan tiga validasi yang terdiri dari *decision tree*, *naive bayes*, dan *KNN* untuk diuji sehingga dapat mengetahui nilai akurasi tertinggi diantara algoritma tersebut.



Sumber: Data Olah Pribadi

Gambar 8 Hasil AUC Decision Tree

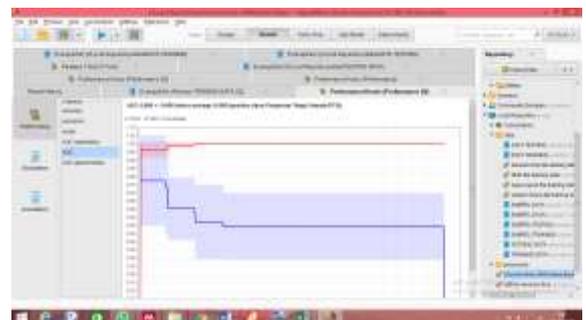
Pada gambar 8 hasil *accuracy decision tree* yaitu 95.60% dan AUC nya 0,975



Sumber: Data Olah Pribadi

Gambar 9 Hasil AUC Naive Bayes

Pada gambar 9 hasil *accuracy naive bayes* yaitu 92.40% dan AUC nya 0,953



Sumber: Data Olah Pribadi

Gambar 10 Hasil AUC KNN

Pada gambar 10 hasil *accuracy KNN* yaitu 94.96% dan AUC nya 0,996

Tabel 1 Perbandingan metode *Decision Tree*,
Naive Bayes dan KNN

Metode	Akurasi	Nilai AUC
<i>Naive Bayes</i>	92.40%	0.953
<i>Decision Tree</i>	95.60%	0.975
KNN	94.96%	0.996

Sumber: Data Olah Pribadi

Pada Tabel 1 Perbandingan metode *Decision Tree*, *Naive Bayes* dan KNN. Hasil pengujian data tersebut akurasi tertinggi terdapat di *decision tree* yaitu 95.60% dibandingkan dengan metode KNN dengan hasil akurasi 94.96% dan *naive bayes* dengan hasil akurasi 92.40%.

4. Kesimpulan

Penelitian ini untuk memprediksi siswa SMK AL Hidayah angkatan berikutnya yang masuk di perguruan tinggi negeri maupun swasta menggunakan hasil model klasifikasi yang terbentuk. Data yang digunakan adalah data siswa angkatan 2018 dan 2019 Sekolah Menengah Kejuruan AL-Hidayah 1. Yang berjumlah 503 data, setelah dilakukan *pre-processing* diperoleh data yang siap diolah berjumlah 158 data. Berdasarkan data yang diuji terhadap dataset siswa SMK Al-Hidayah melalui proses dengan menggunakan metode *naive bayes*, *decision tree*, dan KNN menggunakan aplikasi *rapid Miner*. Dalam penelitian ini terdapat 9 atribut yaitu antara lain kompetensi keahlian (Jurusan siswa), Jenis Kelamin, Rata-rata nilai raport, Rata-rata nilai UN, Pekerjaan ayah ibu, Pendidikan terakhir ayah ibu, dan siswa yang melanjutkan kuliah. Berdasarkan prediksi data mining pada algoritma *decision tree* menghasilkan keakurasian 95,60%, pada data mining menggunakan algoritma *Naive Bayes* Menghasilkan keaurasian 92,40%, dan pada data mining menggunakan KNN menghasilkan tingkat keakurasian mencapai 94,96%.

Berdasarkan penelitian diatas, dari data data yang telah diuji bahwa tingkat akurasi *decision tree* lebih tinggi dari *naive bayes* dan KNN, akurasi pada algoritma *decision tree* tersebut yang mencapai 95.60%. sehingga dapat disimpulkan bahwa algoritma *decision tree* merupakan algoritma terbaik untuk memprediksi siswa SMK di AL Hidayah yang masuk ke Perguruan Tinggi Negeri (PTN) Atau Perguruan Tinggi Swasta (PTS)

Referensi

- Annur, H. (2019). Penerapan *Data Mining* Menentukan Strategi Penjualan Variasi Mobil Menggunakan Metode K-Means Clustering. *Jurnal Informatika Upgris*, 5(1).
- Dwi Arum Ningtyas, M. W. & N. N. (2019). Klasifikasi Siswa Smk Berpotensi Putus Sekolah Menggunakan Algoritma Decision Tree , Support Vector Machine Dan Naive Bayes. *Jurnal Khatulistiwa Informatika*, VII(2), 85–90.
- Dwi Herlambang, A., & Hadi wijoyo, S. (2019). Algoritma *Naive Bayes* Untuk Klasifikasi Sumber Belajar *Naive Bayes Algorithm for Text Based Learning Resources Classification in Productive Subject At Information and*. 6(4), 431–436. <https://doi.org/10.25126/jtiik.201961323>
- Junaidi, A., Dewi, N., Baidawi, T., Agustiani, S., Arifin, Y. T., & Sihotang, H. T. (2020, November). Expert System Of Syzygium Aqueum Disease Diagnose Using Bayes Method. In *Journal of Physics: Conference Series* (Vol. 1641, No. 1, p. 012097). IOP Publishing.
- Kahramanli, H., & Allahverdi, N. (2008). Design of a hybrid system for the diabetes and heart diseases. *Expert Systems with Applications*, 35(1–2), 82–89. <https://doi.org/10.1016/j.eswa.2007.06.004>
- Kusuma, L. W. (2019). *Prediksi Kemampuan Lulusan SMK untuk Dapat Bersaing Di Dunia Kerja dengan Menggunakan Naive Bayes: Studi Kasus SMK Buddhi Tangerang*. 1, 56–63.
- Mardi, Y. (2017). *Data Mining: Klasifikasi Menggunakan Algoritma C4.5*. *Jurnal Edik Informatika*, 2(2), 213–219.
- Nofriansyah, D., Erwansyah, K., & Ramadhan, M. (2016). Penerapan *Data Mining* dengan Algoritma Naive Bayes Clasifier untuk Mengetahui Minat Beli Pelanggan terhadap Kartu Internet XL (Studi Kasus di CV. Sumber Utama Telekomunikasi). *Jurnal Saindikom*, 15(2), 81–92.
- Saifudin, A. (2018). *Metode Data Mining Untuk Seleksi Calon Mahasiswa*. 10(1), 25–36.
- Saikin, S., & Kusriani, K. (2019). Model *Data Mining* Untuk Karakteristik Data Traveller Pada Perusahaan Tour and Travel. *Jurnal Manajemen Informatika Dan Sistem Informasi*, 2(2), 61. <https://doi.org/10.36595/misi.v2i2.105>
- Sidik, M., Rasminto, H., & Manongga, D. (2018). *Implementasi Data Mining Untuk Prediksi Kelulusan Menggunakan Metode Klasifikasi Naive Bayes*. 13–20.
- Sugianto, C. A. (2017). *Penerapan Teknik*

- Data Mining Untuk Menentukan Hasil Seleksi Masuk Sman 1 Gibeber Untuk Siswa Baru Menggunakan Decision Tree.* 39–43.
<https://doi.org/10.31227/osf.io/vedu7>
- Keputusan Seleksi Ujian Masuk Perguruan Tinggi Menggunakan Nbc (Naïve Bayes Classifier). *Kinetik*, 1(3), 173.
<https://doi.org/10.22219/kinetik.v1i3.120>
- Sutoyo, I. (2018). Implementasi Algoritma Decision Tree Untuk Klasifikasi Data Peserta Didik. *Jurnal Pilar Nusa Mandiri*, 14(2), 217.
<https://doi.org/10.33480/pilar.v14i2.926>
- Tree, C., Utomo, D. K., Supianto, A. A., & Purnomo, W. (2019). *Sistem Prediksi Penerimaan SNMPTN menggunakan Algoritme Decision.* 3(9), 9124–9131.
- Trisaputra, Y. (2016). *Klasifikasi Profil Siswa SMA / SMK yang Masuk PTN (Perguruan Tinggi Negeri) dengan k-Nearest Neighbor Klasifikasi Profil Siswa SMA / SMK yang Masuk PTN (Perguruan Tinggi Negeri) dengan k-Nearest Neighbor Yuandri Trisaputra , Indriyani , Shellafuri Mar. September 2015, 0–15.*