Jurnal Teknika (Jurnal Fakultas Teknik Universitas Islam Lamongan) ISSN: 2085-0859
Volume 15, No.1, Tahun 2023 e-ISSN : 2620-4770
Tersedia Online http://www.jurnalteknik.unisla.ac.id/index.php/teknika/index
DOI:10.30736/jt.v15i1.929, Hal 61-66

PERBANDINGAN METODE DATA MINING UNTUK MEMPREDIKSI PRESTASI AKADEMIK SISWA

Fabiano Milan Almufqi[1], Apriade Voutama[2]
Sistem Informasi, Ilmu Komputer, Universitas Singaperbangsa Karawang
Jl. HS.Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat 41361
(0267) 641177
E-mail: fabiano.almufqi19043@student.unsika.ac.id[1], apriade.voutama@staff.unsika.ac.id[2]

ABSTRACT
In the world of education, institutions usually allocate scholarships in the form of appreciation for outstanding students. However, many problems that often arise, namely the lack of proper distribution of scholarships to students. The data collection method used to obtain the data needed in this research is using a questionnaire and literature study. From the data that has been collected, we get data of 102 student records who have filled out the questionnaire, the contents of the questionnaire such as parental support, distance from home to school, to test scores. If the test results are good then the student is eligible to get a scholarship. The first test uses the Naïve Bayes method and has an accuracy of 58.62%. Both use the decision tree method and have an accuracy of 65.52%. The third uses the random forest method and has an accuracy of 51.72%. Based on the results of tests that have been carried out using the RapidMiner application using 3 methods, the research results obtained show that the best method for predicting student academic achievement is the Decision Tree method with an accuracy rate of 65.52%.
Keywords*: Data Mining, Prediction, Naïve Bayes, Decision Tree, Random Forest.*

ABSTRAK
Pada dunia Pendidikan biasanya institusi mengalokasikan beasiswa dalam bentuk apresisasi untuk siswa-siswa yang berprestasi. Namun banyak masalah yang sering muncul, yaitu kurang terpatnya penyaluran beasiswa terhadap siswa. Metode pengumpulan data yang digunakan untuk memperoleh data-data yang dibutuhkan dalam penelitian ini adalah menggunakan kuesionar dan studi Pustaka. Dari data yang sudah dikumpulkan, mendapat data sebesar 102 record siswa yang sudah mengisi kuesioner tersebut, isi dari kuesioner tersebut seperti dukungan orangtua, jarak rumah dari sekolah, hingga ke nilai ulangan. Jika dari hasil ulangannya bagus maka siswa tersebut layak untuk mendapatkan beasiswa. Pengujian pertama menggunakan metode naïve bayes dan memiliki hasil accuracy 58.62%. Kedua menggunakan metode decision tree dan memiliki hasil accuracy 65.52%. Ketiga menggunakan metode random forest dan memiliki hasil accuracy 51.72%. Berdasarkan dari hasil uji yang telah dilakukan menggunakan aplikasi RapidMiner menggunakan 3 metode, hasil penelitian yang didapat menunjukan bahwa metode terbaik dari penelitian prediksi prestasi akademik siswa adalah metode Decision Tree dengan tingkat akurasi 65.52%.

Kata kunci: Data mining, Prediksi, Naïve Bayes, Decision Tree, Random Forest.

1. PENDAHULUAN

Pendidikan adalah salah satu komponen kehidupan yang dapat menunjang keberhasilan seseorang menuju kehidupan yang jauh lebih baik lagi. Terutama bagi anak yang ada dalam lingkup usia wajib belajar. Akan tetapi tidak semua anak wajib belajar dapat mengikuti pendidikan. Untuk mengatasi permasalahan yang ada, maka sekolah menyusun program bantuan dana pendidikan bagi siswa yang dianggap kurang mampu secara strata ekonomi. Dengan adanya syarat dan ketentuan yang berlaku dan semua kriteria pemilihan penentuan siswa memperoleh bantuan dana pendidikan itu dilakukan berdasarkan data siswa yang ada, kemudian dianalisis secara manual, namun terkadang hasil yang diperoleh tidak sesuai dan akurat. Maka dalam penelitian ini penulis membandingkan Algoritma Klarifikasi Data Mining untuk mendapatkan akurasi tertinggi diantara metode data mining lainnya. Sampel data diambil dari Sekolah Menengah Kejuruan (SMK) Taruna Karya 1 Karawang yang beralamat di Jl. Pangkal Perjuangan No. 76, Karawang Barat, Kab. Karawang, Jawa Barat 41316.

Data yang berlimpah membuka peluang diterapkannya data mining untuk pengelolaan Pendidikan yang lebih baik lagi dan data mining

Jurnal Teknika (Jurnal Fakultas Teknik Universitas Islam Lamongan) ISSN: 2085-0859
Volume 15, No.1, Tahun 2023 e-ISSN : 2620-4770
Tersedia Online http://www.jurnalteknik.unisla.ac.id/index.php/teknika/index
DOI:10.30736/jt.v15i1.929, Hal 61-66

dalam pelaksanaan pembelajaran dengan bantuan komputer yang lebih efektif (Andriyana and Nugroho 2015). Untuk meningkatkan akses dan minat belajar siswa serta memajukan lembaga pendidikan, lembaga pendidikan memberikan beasiswa kepada siswa berprestasi. Namun banyak permasalahan yang sering muncul yaitu pengalokasian beasiswa kepada siswa yang kurang tepat, seperti siswa yang tidak berhak mendapatkan beasiswa tetapi menerima beasiswa, begitu pula sebaliknya kepada siswa yang berhak mendapatkan beasiswa yaitu siswa yang berprestasi atau karena kurang mampu, tetapi tidak menerima beasiswa.

Sekolah Menengah Kejuruan Taruna Karya 1 Karawang (SMK Taruna Karya 1 Karawang) merupakan sekolah kejuruan yang mengutamakan hasil dan keterampilan siswa. Banyak faktor yang mempengaruhi hasil belajar siswa. Salah satu hal yang sangat penting adalah manajemen pembelajaran di sekolah yang baik, untuk mendapatkan hasil belajar yang baik manajemen sekolah perlu melakukan kerjasama antara orang tua siswa dan siswa. Sekali manajemen pembelajaran yang baik maka kinerja siswa sangat penting sehingga dianggap serius oleh manajemen sekolah, dikarenakan sulit untuk mengidentifikasi faktor atau variabel yang dapat mempengaruhi kinerja siswa. Karena ada beberapa faktor yang dapat mempengaruhi kemajuan akademik siswa yaitu faktor internal dan faktor eksternal.

Data Mining adalah metode untuk menemukan informasi dalam basis data besar. (Setiawan, 2021) *Data Mining* juga digunakan untuk analisis otomatis dan pengambilan informasi. *Data Mining* digunakan untuk menemukan pola yang diinginkan untuk tujuan mengekstraksi informasi yang berguna dari *database* besar. Pola-pola ini diidentifikasi menggunakan alat yang dapat memberikan analisis data yang berharga dan mendalam yang dapat dilanjutkan dengan menggunakan alat pendukung keputusan lainnya. *Data Mining* adalah salah satu teknik paling umum di KDD, tetapi ini adalah teknik yang sangat penting untuk menemukan pola yang bermakna dalam data besar.

2. METODE

2.1 Prediksi

Prediksi adalah memprediksi sesuatu akan terjadi sekali masa depan. (Suryaningrum & Wijaya, 2015) Prediksi juga bisa untuk klasifikasi, tidak hanya untuk memprediksi *time series*, karena sifat yang dapat dihasilkan class berdasarkan atribut yang ada. Perbedaan antara prediksi dengan perkiraan adalah jika perkiraan adalah hasil perhitungan berdasarkan data dan analisis ilmiah. Jika prediksi adalah hasil perhitungan berdasarkan data dan analisis apapun bisa analisis ilmiah ataupun nonilmiah.

2.2 *Knowledge Discovery in Database*

Knowledge Discovery in Database adalah proses komputasi dengan menggunakan perhitungan matematika yang bekerja untuk mengekstraksi informasi dan membuat kalkulasi probabilistic dari probabilitas Tindakan di masa yang akan datang (Bhatia, 2019). KDD yaitu aktivitas yang didalamnya termasuk pengumpulan, penggunaan data dalam menentukan keteraturan, pola dan hubungan data set yang banyak (Siregar, 2018).

Knowledge Discovery in Database (KDD) biasanya disamakan dengan Data Mining, berdasarkan tujuannya dalam mencari pengetahuan yang berguna dalam tumpukan data yang besar. *Data mining* merupakan salah satu bagian proses KDD yang bertugas untuk mengekstrak pola atau model dari data dengan menggunakan suatu algoritma yang spesifik. Pada penelitian ini, tahapan KDD yaitu:

a. Data *Selection*

Tahap seleksi data merupakan tahapan dimana data dilakukan seleksi atribut dan hasil dari seleksi data tersebut diintegrasikan menjadi sebuah dataset. Proses pembangunan dataset merupakan suatu proses yang penting karena proses pembelajaran data mining dan penemuan pola baru didasarkan pada dataset yang telah dibentuk.

b. Preprocessing *Data*

Pada tahapan ini dilakukan pembersihan data untuk meningkatkan keandalan data. Pembersihan data dilakukan dengan menangani nilai kosong, menangani baris data yang tidak relevan dan menghilangkan *noise* atau *outlier*. Proses awal ini dapat melibatkan metode spesifik yang kompleks atau menggunakan algoritma data mining yang spesifik.

c. *Transformation*

Tahapan ini dilakukan untuk pengembangan data sehingga data dipersiapkan dengan lebih baik dan siap untuk dilakukan pemodelan data mining. Hal yang dapat dilakukan untuk mempersiapkan data menjadi lebih baik adalah melakukan reduksi dimensi seperti pemilihan fitur dan ekstraksi sampel data.

Jurnal Teknika (Jurnal Fakultas Teknik Universitas Islam Lamongan) ISSN: 2085-0859
Volume 15, No.1, Tahun 2023 e-ISSN : 2620-4770
Tersedia Online http://www.jurnalteknik.unisla.ac.id/index.php/teknika/index
DOI:10.30736/jt.v15i1.929, Hal 61-66

d. *Analisis*

Pada proses ini analisis data yang pertama adalah dengan menggunakan metode *Naïve bayes*, *Decision Tree*, dan Random Forest dengan menggunakan aplikasi Rapidminer.

2.3 Naive Bayes

Naive Bayes adalah metode klasifikasi ini menggunakan probabilitas dan data statistik. Naïve Bayes bisa memprediksi peluang masa depan berdasarkan pengalaman masa lalu, sehingga disebut juga teorema Bayes (Andriansyah, Yusup, and Voutama 2021). Metode Naïve Bayes ini yang bagus untuk didalam pembelajaran berdasarkan data training, dengan dasar menggunakan probabilitas bersyarat. Metode Naïve Bayes ini memiliki sebuah kelemahan yaitu metode ini hanya bisa digunakan untuk persoalan klasifikasi dengan *supervised learning* dan juga data kategorikal (Triwidianti, Alfian, and Prasojo 2021).

2.4 Decision Tree

Decision Tree adalah struktur pohon di mana node pohon mewakili atribut yang diuji. Setiap cabang dari pohon ini merupakan distribusi hasil pengujian, dan simpul-simpul tersebut mewakili beberapa kelompok kelas. Pohon keputusan adalah alat pendukung dengan struktur pohon yang menampilkan kemungkinan hasil, biaya sumber daya, manfaat, dan hasil potensial. *Decision Tree* adalah salah satu metode penambangan data terbaik berdasarkan teknik pembelajaran. Metode pohon keputusan ini meningkatkan model prediktif, kemudahan interpretasi dan ketahanan. Metode ini efektif dalam menganalisis hubungan nonlinier karena dapat mengatasi tantangan pemasangan data seperti regresi dan klasifikasi.

2.5 Random Forest

Random Forest adalah sebuah kombinasi dari masing-masing pohon yang baik kemudian dikombinasikan lagi kedalam satu model. *Random Forest* ini memiliki sifat ketergantungan pada sebuah nilai *vector random* dengan distribusi yang sama pada semua pohon yang masing-masing dari *decision tree* memiliki kedalaman yang maksimal. *Random Forest* biasanya digunakan untuk menyelesaikan masalah-masalah yang berhubungan dengan klasifikasi, regresi, dan sebagainya. Ada dua faktor yang membuat metode ini disebut random, yaitu :

1. Setiap tree/pohon tumbuh pada sampel *bootstrap* yang diambil dari data training secara acak.
2. Dalam setiap *node split* selama pembentukan *decision tree*, Sebagian sampel dari m variable dipilih dari kumpulan-kumpulan data yang asli, kemudian yang terbaik akan digunakan dalam *node* tersebut.

2.6 Metode Pengumpulan Data

Metode pengumpulan data yang digunakan untuk memperoleh data-data yang dibutuhkan dalam penelitian ini adalah :

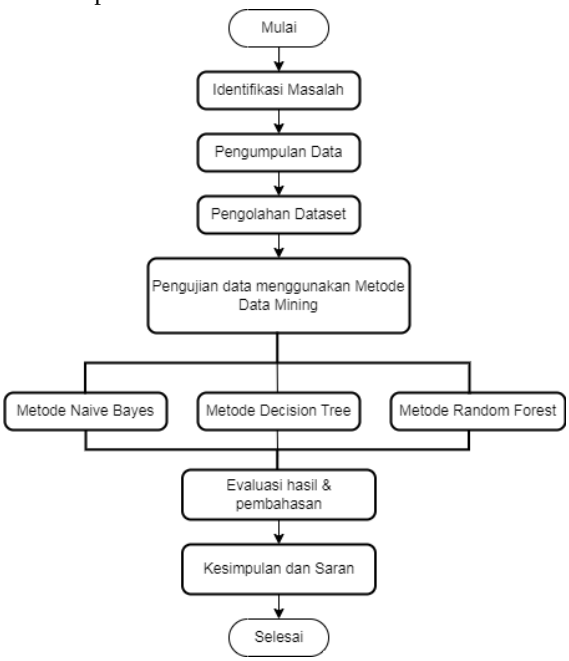
a. Kuesioner

Metode kuesioner adalah sebuah teknik menghimpun data dari sejumlah orang atau responden melalui seperangkat pertanyaan untuk dijawab. Dengan memberikan daftar pertanyaan tersebut, jawaban-jawaban yang diperoleh kemudian dikumpulkan sebagai data. Pertanyaan tersebut di jawab oleh orang yang mau diteliti yaitu siswa untuk mendapatkan suatu informasi.

b. Studi Pustaka

Merupakan metode pengumpulan data yang diperoleh dari hasil olahan orang lain berupa dokumen, buku pustaka, jurnal, dengan membaca berbagai bahan penulisan, mengenai permasalahan yang berhubungan dengan penulisan dan khususnya penelitian yang berkaitan karya ilmiah.

c. Tahapan Penelitian



Gambar 1. Alur Penelitian

Metodologi penelitian merupakan tahapan-tahapan yang sistematis dilakukan pada penelitian

Jurnal Teknika (Jurnal Fakultas Teknik Universitas Islam Lamongan)
Volume 15, No.1, Tahun 2023
Tersedia Online http://www.jurnalteknik.unisla.ac.id/index.php/teknika/index
DOI:10.30736/jt.v15i1.929, Hal 61-66
ISSN: 2085-0859
e-ISSN : 2620-4770

dan alur penelitian yang dilakukan adalah sebagai berikut :

Alur penelitian dimulai identifikasi masalah pencarian data set, data yang diperoleh tersebut diolah atau pre prosesing data menjadi data set, kemudian melakukan pengujian pada setiap metode, setelah itu peneliti melakukan evaluasi dan pembahasan yang terkahir peneliti menyimpulkan hasil.

3. PEMBAHASAN

Dari data yang sudah dikumpulkan, didapatkan sebesar 102 record siswa yang sudah mengisi kuesioner tersebut. Kemudian data diproses untuk mengetahui apakah data tersebut layak atau tidak, dan setelah dilakukan proses data yang didapatkan layak untuk digunakan sebagai penelitian. Data yang akan dilakukan data training sebanyak 72 record data dan yang akan dijadikan data testing sebanyak 30 record data. Berikut adalah data testingnya :

Tabel 1. Record Kuesioner Kategori 1

Jenis Kelamin	Kondisi Fisik	Motivasi Belajar
Pria	Normal	Ada
Wanita	Normal	Ada
Pria	Normal	Ada
Pria	Normal	Ada
Wanita	Mata Minus	Ada

Tabel 2. Record Kuesioner Kategori 2

Jumlah Anggota Keluarga	Status Orangtua	Pendidikan Ayah
Lebih dari 3	Bersama	SMA/SMK
Lebih dari 3	Bersama	SMA/SMK
Lebih dari 3	Bersama	SMA/SMK
Lebih dari 3	Bersama	SMA/SMK
Lebih dari 3	Bersama	SD

Tabel 3. Record Kuesioner Kategori 3

Pendidikan Ibu	Pekerjaan Ayah	Pekerjaan Ibu
SMP	Tidak	IRT
SD	Pengusahasaha	IRT
SMA/SMK	Pegawai Swasta	IRT
SMA/SMK	Pegawai Swasta	IRT
SD	Petani	IRT

Tabel 4. Record Kuesioner Kategori 4

Dukungan Keluarga	Ekstrakulikuler	Internet Dirumah
Ya	Tidak	Ya
Ya	Ya	Tidak
Ya	Tidak	Ya
Ya	Ya	Tidak
Ya	Ya	Tidak

Tabel 5. Record Kuesioner Kategori 5

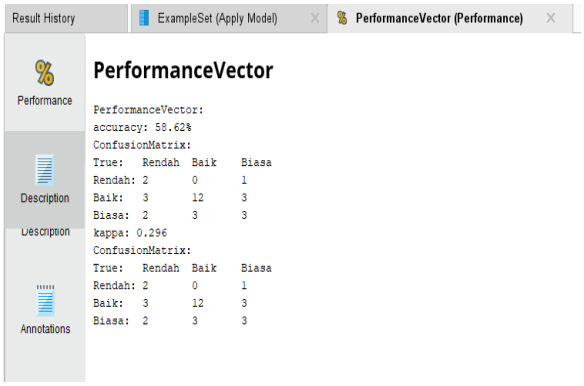
Jarak dari rumah	Waktu Belajar	Hasil Ulangan
Jauh	Rajin	Baik
Jauh	Biasa Saja	Baik
Jauh	Rajin	Baik
Jauh	Sangat Rajin	Biasa
Sangat Jauh	Sangat Jarang	Biasa

3.1 Hasil Uji

Hasil akurasi yang didapatkan sebagai berikut :

1. Naive Bayes

Pertama diuji dengan metode naïve bayes dan memiliki hasil accuracy sebesar 58.62%.



Gambar 2. *Performance Vector Naïve Bayes*

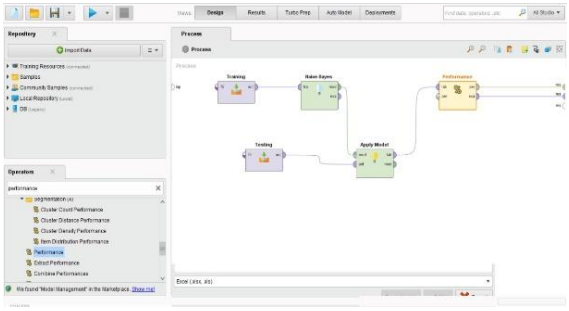
Tabel 5. Performance Naïve Bayes

	True. Rendah	True. Baik	True. Biasa	Class Presisi
Pred.Rendah	2	0	1	66.67%
Pred. Baik	3	12	3	66.67%
Pred. Biasa	2	3	3	37.50%
Class Recall	28.57%	80.00%	42.86%	

Accuracy = 58.62 %

Terlihat dari performance metode naïve bayes memiliki *accuracy* 58.62% dan pada *class recall* dan *precision* bisa dilihat pada tabel 5.

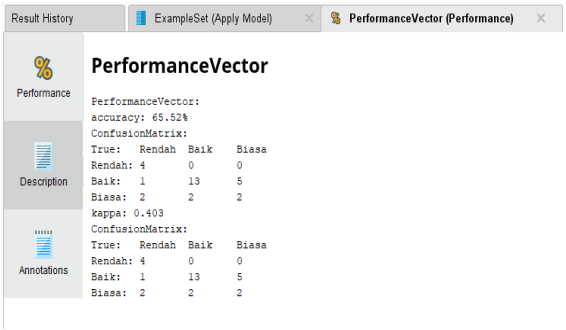
Jurnal Teknika (Jurnal Fakultas Teknik Universitas Islam Lamongan)
Volume 15, No.1, Tahun 2023
Tersedia Online http://www.jurnalteknik.unisla.ac.id/index.php/teknika/index
DOI:10.30736/jt.v15i1.929, Hal 61-66
ISSN: 2085-0859
e-ISSN : 2620-4770



Gambar 3. *Design Naïve Bayes*

2. Decision Tree

Kedua diuji menggunakan metode *Decision Tree* dan memiliki hasil *accuracy* sebesar 65.52%.



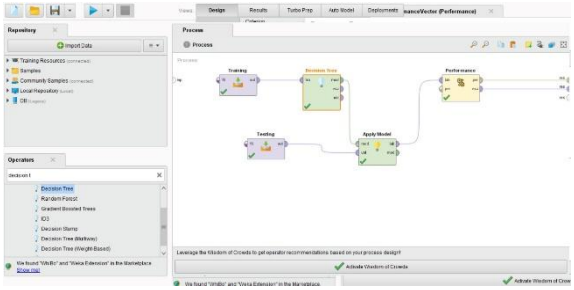
Gambar 4. *Performance Vector Decision Tree*

Tabel 6. Performance *Decision Tree*

	True. Rendah	True. Baik	True. Biasa	Class Presisi
Pred.Rendah	4	0	0	100.00%
Pred. Baik	1	13	5	68.42%
Pred. Biasa	2	2	2	33.33%
Class Recall	57.14%	86.67%	28.57%	

Accuracy = 65.52 %

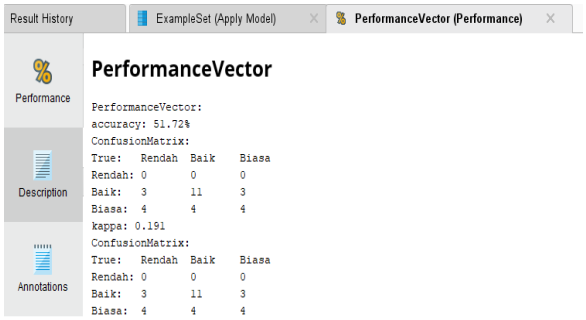
Terlihat dari performance metode *Decision Tree* memiliki *accuracy* sebesar 65.52% dan pada *class recall* dan *precision* bisa dilihat pada tabel 6 *performance decision tree*.



Gambar 5. *Design Decision Tree*

3. Random Forest

Ketiga diuji menggunakan metode *Random Forest* dan memiliki hasil *accuracy* sebesar 51.72%.



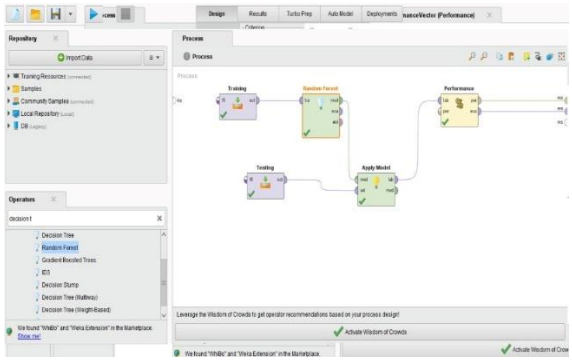
Gambar 6. *Performance Vector Random Forest*

Tabel 7. Performance *Random Forest*

	True. Rendah	True. Baik	True. Biasa	Class Presisi
Pred.Rendah	0	0	0	00.00%
Pred. Baik	3	11	3	64.71%
Pred. Biasa	4	4	4	33.33%
Class Recall	00.00%	73.33%	57.14%	

Accuracy = 51.72 %

Terlihat dari performance metode *Random Forest* memiliki *accuracy* sebesar 51.72% dan pada *class recall* dan *class precision* bisa dilihat pada tabel 7 *performance Random Forest*.



Gambar 7. *Design Random Forest*

Setelah dilakukan pengujian dengan tiga algoritma diatas, maka didapatkan hasil pengujiannya seperti di tabel 8 berikut:

Tabel 8. Hasil Akurasi

Metode Penelitian	Accuracy
Naïve Bayes	58.62%
Decision Tree	65.52%
Random Forest	51.72%

3.2 Pembahasan

Dari hasil tabel 8 tersebut didapatkan bahwa hasil pengujian prediksi terbaik dengan menggunakan algoritma *Decision Tree* sebesar 65.52%, diikuti dengan *Naïve Bayes* sebesar 58.62%, dan terakhir algoritma *Random Forest* sebesar 51.72%. Dari hasil tersebut dilihat bahwa model prediksi yang paling

sesuai dengan dataset ini yaitu menggunakan algoritma *Decision Tree.*

4. KESIMPULAN

Berdasarkan dari hasil uji yang telah dilakukan menggunakan aplikasi RapidMiner maka dapat dilihat tingkat akurasi menggunakan 3 metode, maka metode terbaik dari penelitian prediksi prestasi akademik siswa adalah metode *Decision Tree* dengan tingkat akurasi tinggi sebesar 65.52%.

PUSTAKA

Setiawan, R. (2021, October 30). *Apa itu Data Mining dan Bagaimana Metodenya?* Diambil kembali dari Dicoding: https://www.dicoding.com/blog/apa-itu-data-mining/

Suryaningrum, K. M., & Wijaya, S. P. (2015). Analisa dan Penerapan Metode Single Exponential Smoothing untuk Prediksi Penjualan pada Periode Tertentu (Studi Kasus: PT.Media Cemara Kreasi). *Prosiding SNATIF*, 259-266.

Andriansyah, Miftah Fariedh, Dadang Yusup, and *Apriade* Voutama. 2021. "MENGGUNAKAN METODE NAÏVE BAYES BERBASIS WEBSITE WEB-BASED EXPERT SYSTEM OF COVID-19 EARLY DETECTION USING NAÏVE BAYES METHOD." *Journal of Information Technology and Computer Science (INTECOMS)* 4(2): 446–55.

Andriyana, Veronica, and Yusuf Nugroho. 2015. *PERBANDINGAN 3 METODE DALAM DATA MINING UNTUK PREDIKSI PENERIMA BEASISWA BERDASARKAN PRESTASI DI SMA NEGERI 6 SURAKARTA.*

Muhammad, Fakhri, Nana Mulyana Maghfur, and Apriade Voutama. *2022.* "Sentiment Analysis Dataset on COVID-19 Variant News." *Systematics* 4(1): 382–91.

Ramadhan, Vicky, and Apriade Voutama. 2022. "Clustering Menggunakan Algoritma K-Means Pada *Penyakit* ISPA Di Puskesmas Kabupaten Karawang." *Jurnal Pendidikan dan Konseling* 4: 462–73.

Triwidianti, Jani, Firmansyah Yunialfi Alfian, and Margi Prasojo. 2021. "Perbandingan Metode Data Mining Untuk Prediksi Prestasi Siswa Tingkat Pendidikan Menengah Kejuruan Pada Sekolah Menengah Kejuruan Negeri (SMKN 1) Gadingrejo Pringsewu Lampung." 1: 126–33. https://jurnal.darmajaya.ac.id/ (March 30, 2022).

Yoga Pratama, Aditiya et al. 2021. "Analisis Sentimen Media Sosial Twitter Dengan Algoritma K-Nearest Neighbor Dan Seleksi Fitur Chi-Square (Kasus Omnibus Law Cipta Kerja)." *Jurnal Sains Komputer & Informatika (J-SAKTI* 5(2): 897–910.