

ANALISIS PERBANDINGAN ALGORITMA DECISION TREE DAN NAÏVE UNTUK MENGEVALUASI PRESTASI BELAJAR SISWA STUDI KASUS: SMK AL-MUSYAWIRIN

Aldi Sholihin Fauzan, Ade Irma Purnama Sari, Irfan Ali
Teknik Informatika S1, STMIK IKMI Cirebon
Jalan Perjuangan No. 10B Majasem, Kota Cirebon, Indonesia
aldifauzan0405@gmail.com

ABSTRAK

Prestasi belajar siswa merupakan faktor kunci dalam dunia pendidikan. Suatu negara akan maju seiring dengan kemajuan dalam bidang pendidikannya. Prestasi belajar siswa dapat meningkatkan akreditasi sekolah dan merupakan salah satu faktor keberhasilan pendidikan. Analisis data akademik yang tepat dapat membantu mengenai unsur-unsur yang mempengaruhi hasil prestasi belajar siswa. Oleh karena itu, penggunaan algoritma klasifikasi seperti *Decision Tree* dan *Naïve Bayes* menjadi penting untuk membantu pengambilan keputusan di lingkungan pendidikan. Penelitian ini mendapatkan sumber data dari SMK Al-Musyawirin Weru Cirebon dengan jumlah data sebesar 202 dengan menggunakan teknik pengumpulan data melakukan observasi dan wawancara langsung. Atribut yang digunakan dalam penelitian ini adalah Nama Siswa, Mata Pelajaran Pendidikan Agama Islam, Pendidikan Kewarganegaraan, Bahasa Indonesia, Bahasa Inggris, Matematika, Sejarah, Seni Budaya, Pendidikan Jasmani Olahraga Kesehatan, Simulasi Digital, Fisika, Kimia dan Bahasa Sunda. Peneliti menerapkan kedua algoritma ini menggunakan teknik data mining. Pada pengujian *Cross Validation* dengan *5-fold* menggunakan metode *Naïve Bayes* menunjukkan hasil keakuratan tertinggi sebesar 85.93%. Sedangkan pengujian menggunakan *split* data medapatkan keakuratan 95.00%. Pada pengujian *Cross Validation* dengan *7-fold* menggunakan metode *Decision Tree* menunjukkan hasil keakuratan tertinggi sebesar 89,22%. Sedangkan pengujian menggunakan *split* data medapatkan keakuratan 100.00%. Menurut hasil yang telah diuji dengan menggunakan *tools RapidMiner*, menyatakan bahwa metode *Decision Tree* memiliki tingkat akurasi yang lebih tinggi daripada metode *Naïve Bayes* dalam menentukan prestasi akademik siswa. Dalam hal ini bagian akademik bisa mengantisipasi hasil pembelajaran siswa jika ada yang kurang berprestasi.

Kata kunci : *Data Mining, Decision Tree, Prestasi, Naïve Bayes*

1. PENDAHULUAN

Pendidikan memegang peranan penting dalam upaya menciptakan perkembangan siswa yang berkelanjutan dan memiliki mutu tinggi. Dengan pesatnya kemajuan teknologi informatika telah membawa transformasi dalam cara siswa belajar dan guru mengajar. Aspek penting dalam dunia pendidikan yaitu pengukuran dan evaluasi prestasi belajar siswa. Evaluasi ini menjadi landasan untuk mengidentifikasi potensi siswa, mengembangkan program pembelajaran yang lebih baik, dan mengukur efektifitas metode pengajaran. Khususnya di Sekolah Menengah Kejuruan (SMK), pemahaman yang mendalam tentang prestasi belajar siswa menjadi sangat penting, karena mereka mempersiapkan siswa untuk masuk ke dunia kerja yang seringkali membutuhkan keterampilan teknis yang kuat agar dapat bermanfaat bagi karier mereka dan dalam kehidupan sosial. Oleh karena itu, analisis yang tepat terhadap komponen yang mempengaruhi prestasi akademik siswa di SMK menjadi sangat penting. Kualitas guru merupakan salah satu komponen utama yang mempunyai dampak yang signifikan pada keberhasilan pembelajaran. Mengingat betapa pentingnya peran guru dalam proses pendidikan, maka sudah seharusnya untuk memperhatikan kualitas guru [1]. Permasalahan yang umum terjadi pada sekolah

saat ini khususnya di SMK Al-Musyawirin yaitu mengenai semangat belajar siswa yang cenderung menurun setiap tahunnya, banyak yang tidak mengerjakan tugas sekolah atau bahkan bolos sekolah, hasil belajar pun menjadi tidak maksimal sehingga tidak sedikit siswa yang prestasinya menurun. Di setiap sekolah pasti memiliki siswa dan siswi berprestasi dan kurang berprestasi. Oleh karena itu pihak sekolah harus mengambil tindakan pencegahan untuk meningkatkan kualitas pendidikan dalam mengatasi siswa yang kurang berprestasi, dengan tujuan menciptakan lulusan siswa yang unggul dan meningkatkan nama baik sekolah [2].

Dalam menentukan prestasi siswa yang berbeda-beda, pihak sekolah SMK Al-Musyawirin memerlukan informasi yang akurat. Informasi ini akan menjadi panduan bagi pihak pendidikan untuk memahami lebih baik siswa yang mencapai prestasi tinggi dan yang memiliki prestasi yang lebih rendah. Dengan demikian, untuk mengatasi masalah ini, peneliti berencana untuk memanfaatkan teknik data mining. Tujuan dari penelitian ini adalah memberikan dukungan kepada sekolah dalam memprediksi prestasi siswa, sehingga dapat di lakukan langkah-langkah antisipatif dalam mengelola prestasi siswa di lingkungan sekolah. Selain itu, penelitian ini dapat bermanfaat bagi peneliti lain yang tertarik untuk

menggunakan algoritma yang dipakai dalam penelitian ini. Algoritma data mining yang digunakan peneliti untuk mengetahui perbandingan akurasi dalam menentukan Prestasi Belajar Siswa SMK adalah *Decision Tree* dan *Naïve Bayes*. Penelitian ini menggunakan pendekatan kuantitatif, dimana seluruh data disajikan dalam bentuk angka dan dianalisis menggunakan metode analisis statistik. Metode penelitian ini akan menggunakan data nilai akademik siswa SMK AL-Musyawirin Weru Cirebon yang akan menjadi sumber data untuk dianalisis. Data yang diperoleh akan diproses dan disiapkan untuk ditraining dan dilakukan pengujian algoritma *Decision Tree* dan *Naïve Bayes*. Pendekatan ini memberikan pemahaman yang lengkap mengenai algoritma *Decision Tree* dan *Naïve Bayes* untuk menentukan prestasi belajar siswa. Penelitian ini menggunakan data langsung dari bagian akademik SMK Al-Musyawirin, sedangkan pada penelitian terdahulu banyak menggunakan kuesioner untuk pengumpulan datanya. Penelitian ini juga menggunakan uji cross validation dengan nilai fold 5-fold, 7-fold dan 10-fold agar dapat menghasilkan hasil yang lebih akurat.

Hasil dari penelitian ini dapat memberikan kontribusi yang berarti dalam pemahaman kita tentang penerapan algoritma klasifikasi dalam konteks pendidikan, khususnya dalam meningkatkan proses pengelolaan dan pemahaman prestasi belajar siswa pada Sekolah Menengah Kejuruan (SMK). Hasil penelitian yang menunjukkan perbandingan akurasi antara Algoritma *Decision Tree* dan *Naïve Bayes* dapat menjadi dasar yang dalam pemilihan metode yang paling efektif dalam menentukan prestasi belajar siswa. Hal ini juga dapat mengetahui siswa-siswa yang berprestasi dan kurang berprestasi. Dampak umum penelitian ini juga dapat memberikan wawasan yang berharga dalam pengembangan dan peningkatan algoritma klasifikasi di bidang Informatika, terutama dalam konteks pemrosesan data pendidikan.

2. TINJAUAN PUSTAKA

2.1. Metode Literature Review

Penelitian ini dilakukan perbandingan metode data mining *Naïve Bayes*, *Decision Tree* dan *Random Forest* dan untuk memprediksi prestasi akademik siswa. Dalam penelitian ini menggunakan kuesioner dan studi pustaka untuk pengumpulan data. Data yang telah dikumpulkan mendapatkan sebanyak 102 data siswa yang mengisi kuesioner, kuesioner tersebut mencakup informasi tentang dukungan orangtua, jarak rumah ke sekolah dan nilai ujian. Pengujian metode *Naïve Bayes* menunjukkan akurasi 58.62%. Metode *Decision Tree* menunjukkan akurasi 65.52%. Metode *Random Forest* menunjukkan akurasi 51.72%. dari hasil penelitian tersebut metode yang paling akurat adalah metode *Decision Tree* [3]. Penelitian ini membahas perbedaan antara metode *Random Forest* dan *Naïve Bayes* dalam menganalisis dan memprediksi kemampuan belajar siswa. Penelitian ini diuji menggunakan *10-fold cross validation*. Dataset

dipisahkan menjadi 2 yaitu data *training* dan data *testing*. Hasil penelitian didapatkan metode algoritma *Random Forest* mempunyai akurasi yang paling besar yaitu 95,12% dibandingkan *Naïve Bayes* yang hanya memiliki akurasi 63.41% [4]. Pada penelitian terdahulu menjelaskan mengenai prediksi waktu lama studi mahasiswa dengan kategori lulus tepat waktu atau lulus tidak tepat waktu. Penelitian ini menggunakan metode *Naïve Bayes* dan *Decision Tree* dengan hasil yang menunjukkan bahwa *Decision Tree* memiliki persentase keakuratan yang lebih besar daripada *Naïve Bayes* yang hanya memiliki persentase 30% [5].

Dalam penelitian terdahulu membahas mengenai perbandingan dan juga evaluasi dari algoritma *C4.5* dan *Naïve Bayes* dalam memprediksi prestasi siswa di Sekolah Menengah Kejuruan. Data yang digunakan berasal dari arsip data siswa SMK Al-Ijtihad. Hasil penelitian menunjukkan algoritma *Naïve Bayes* lebih akurat daripada algoritma *C4.5* untuk memprediksi prestasi siswa [6]. Penelitian terdahulu menjelaskan mengenai implementasi algoritma *Decision Tree* untuk mengevaluasi prestasi akademik siswa. Penelitian tersebut menerapkan algoritma ini untuk mengklasifikasikan kinerja akademik dengan menggunakan struktur hirarki. Dari penelitian ini menghasilkan bahwa akurasi algoritma *Decision Tree* mencapai 68,75%. Data yang diuji menggunakan 144 data, menunjukkan hasil *recall* sebesar 58,54%, hasil presisi sebesar 54,84%, dan hasil kurasi sebesar 63,19%. Penurunan pada akurasi terjadi karena terdapat beberapa data memiliki kategori dan variasi yang sama tetapi memiliki kelas dan keputusan yang berbeda antara data pelatihan dan data pengujian [7].

Jurnal lainnya menjelaskan prediksi prestasi siswa menggunakan metode data mining. Penelitian ini mengkomparasi tiga metode data mining antara lain *Decision Tree*, *Naïve Bayes*, dan *Random Forest* untuk menilai tingkat akurasi prediksi prestasi siswa. Penelitian menekankan manajemen pembelajaran di sekolah, dunia kerja, guru, orang tua dan siswa sangat penting untuk meningkatkan prestasi siswa. Nilai ujian, ujian nasional, nilai test penerimaan, dan nilai raport adalah faktor yang dianalisis menggunakan data mining. Hasil penelitian didapatkan metode *Random Forest* memiliki tingkat akurasi tertinggi sebesar 97.3% [8]. Dalam paper yang ini menjelaskan mengenai implementasi algoritma *Decision Tree* dalam memprediksi prestasi siswa di kelas XII SMK Negeri 1 Selong. Masalah yang sering dihadapi sekolah menjadi subjek penelitian ini, yaitu belum adanya antisipasi terhadap siswa yang memungkinkan mengalami kesulitan dalam belajar atau kurang berprestasi. Data yang digunakan adalah data siswa kelas 10 tahun ajaran 2015/2016 dan kelas 11 tahun ajaran 2016/2017. Yang akan di prediksi dikelas 12. Dari penelitian tersebut dihasilkan akurasi sebesar 97,22% sedangkan persentase nilai AUC yang didapatkan adalah 0.889% dengan *good classification* [9]. Pada penelitian terdahulu ini menggunakan

metode yang bersifat kuantitatif. Kelas X (Sepuluh) Tahun Ajaran 2018–2020 adalah subjek dari kelompok ini. Test hasil belajar adalah metode pengumpulan data yang digunakan. Pembagian ini dilakukan menggunakan program Data Mining *Rapid Miner 5.3*. Hasilnya akan menunjukkan bahwa Algoritma *C4.5* digunakan untuk menilai prestasi belajar. Hasil perhitungan manual yang dilakukan dengan pengujian *RapidMiner* menghasilkan sebelas model. Aturan atau rule Tingkat Prestasi Siswa, dengan nilai akurasi sebesar 71,43% dan nilai *precision class* sebesar 63,89% dan nilai *precision class* sebesar 92,31%. Hasil ini menunjukkan bahwa aturan yang dihasilkan memiliki tingkat kebenaran mendekati 100% [10].

2.2. Naïve Bayes

Naïve Bayes merupakan metode klasifikasi data statistik yang dapat memprediksi probabilitas terkait keanggotaan suatu kelas. *Naïve Bayes* bersumber dari teorema Bayes, dan mempunyai kemampuan klasifikasi serupa dengan *Decision Tree* dan jaringan saraf. Metode *Naïve Bayes* ini sangat efektif dalam pembelajaran berdasarkan data pelatihan, karena mengandalkan probabilitas bersyarat. Namun, metode *Naïve Bayes* juga memiliki kekurangan, yakni hanya dapat digunakan dalam konteks klasifikasi dengan pembelajaran terawasi (*supervised learning*) dan hanya cocok untuk data yang bersifat kategorikal [11].

Dari segi statistik terdapat suatu teknik yang disebut *Naïve Bayes* yang bisa digunakan untuk memproyeksikan probabilitas kemunculan suatu kategori. Penerapan *Naïve Bayes* seringkali dilakukan untuk mengelompokkan data yang telah dikumpulkan dari proses pelatihan yang intensif. Selain itu juga dapat untuk klasifikasi peluang yang efektif yang menurunkan kemungkinan jumlah dengan menyatukan frekuensi dan nilai numerik dari gabungan data yang sudah ada [11].

2.3. Decision Tree

Decision Tree ialah salah satu cara yang digunakan untuk mengklasifikasikan pada suatu objek. Cara ini terdiri dari rangkaian *decision node* yang terkoneksi melalui cabang, berjalan ke arah bawah dari *root node* dan berakhir pada *node* paling ujung [7]. Metode *Decision Tree* ini meningkatkan model prediktif, membuatnya lebih mudah diinterpretasikan, dan membuatnya tahan lama. Selain itu, metode ini dapat menangani masalah pemasangan data seperti regresi dan klasifikasi yang membuatnya efisien dalam menganalisis hubungan nonlinier [12].

Decision Tree menjadi metode yang sangat populer dalam klasifikasi, pemodelan ini dibentuk berdasarkan data dengan struktur pohon. Untuk mengelompokkan data, setiap atribut harus melewati serangkaian *node* keputusan. Sesudah atribut terdapat pada ujung cabang pohon, data kemudian akan diklasifikasikan berdasarkan dengan kelas yang sudah ditentukan [11].

3. METODE PENELITIAN

Metode penelitian menjadi panduan esensial untuk menjalankan proses penelitian secara terarah, teratur, dan sistematis. Pada penelitian ini menggunakan jenis kuantitatif, penelitian yang berfokus pada data yang memberikan informasi yang berguna untuk menyimpulkan hasil. Metode penelitian kuantitatif dijelaskan sebagai suatu bentuk kajian sistematis terhadap suatu fenomena yang melibatkan observasi langsung di lapangan, serta pengumpulan data yang telah didapatkan selanjutnya dianalisis memanfaatkan teknik statistik untuk menemukan korelasi antar variabel yang diteliti dan juga membandingkan dua algoritma sebagai penilaian algoritma terbaik. Analisis yang diterapkan untuk memprediksi prestasi belajar siswa SMK Al-Musyawirin dengan metode algoritma *Naïve Bayes* dan algoritma *Decision Tree* dengan mengaplikasikan *tools rapidminer*. [6]

3.1. Tahapan Metode Penelitian

Tahapan penelitian ini dimulai dengan mengidentifikasi masalah yang ada, kemudian mulai untuk mengumpulkan data nilai-nilai siswa SMK Al-Musyawirin dalam bentuk excel. Setelah data terkumpul dilakukan pengolahan data awal dengan menghilangkan data yang tidak dibutuhkan atau data yang tidak sesuai. Selanjutnya dilakukan pengujian dan evaluasi memakai algoritma *Decision Tree* dan algoritma *Naïve Bayes* dengan *tools Rapidminer Studio* [12].

3.2. Teknik Pengumpulan Data

Teknik Pengumpulan data yang digunakan dalam penelitian ini adalah dengan cara mengamati langsung pada tempat penelitian yaitu SMK AL-Musyawirin Weru Cirebon, serta hasil wawancara langsung dengan bagian akademik sekolah [12].

3.3. Observasi

Melalui pengamatan, peneliti melakukan tinjauan langsung di SMK Al-Musyawirin Weru Cirebon untuk mendapatkan informasi terkait profil sekolah dan data yang dibutuhkan terkait penelitian ini mencakup nilai akademik siswa kelas 10 tahun ajaran 2019/2020 [9].

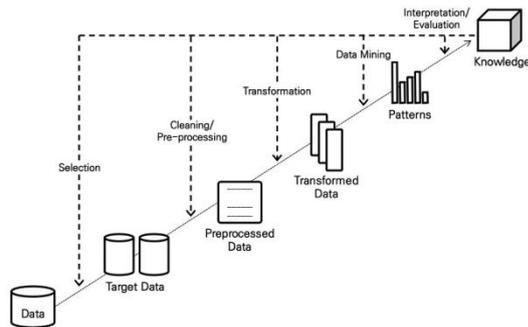
3.4. Studi Literature

Studi literatur adalah studi yang dikerjakan untuk menganalisis beberapa jurnal-jurnal yang ada, dan dipilih dari beberapa sumber untuk mendapatkan kesimpulan dan ide baru [12].

3.5. Teknik Analisis Data

Teknik kuantitatif digunakan untuk menganalisis data pada penelitian ini. Analisis data untuk mengkaji dan juga menguji variabel yang telah dipilih. Analisis bertujuan untuk memproses data mentah hingga menjadi data yang bisa di pahami. Penelitian ini memakai teknik analisis data *Knowledge Discovery in*

Database (KDD). merujuk pada proses yang mencakup mengumpulkan dan menggunakan data untuk mengidentifikasi pola, keteraturan, serta hubungan dalam dataset yang besar. Tujuan *Knowledge Discovery in Database* adalah untuk mengidentifikasi informasi yang terdapat pada dataset yang besar. Bagian dari proses KDD adalah data mining, dimana pola dan model diekstrak dari data dengan menggunakan algoritma tertentu. Pada penelitian ini tahapan *Knowledge Discovery in Database* yaitu [13].



Gambar 1. Tahapan proses KDD
Sumber:

<https://www.sciencedirect.com/science/article/pii/S0264275122005042>

Berikut ini adalah langkah-langkah proses dalam penerapan metode KDD yaitu:

a. *Data Selection*

Sebuah dataset dibentuk melalui proses pembangunan dataset, yang sangat penting karena pembelajaran data mining dan penemuan pola baru bergantung dengan dataset yang telah dibuat. Pada tahap seleksi data, data dipilih berdasarkan atributnya dan hasilnya diintegrasikan menjadi sebuah dataset. Seleksi data ini memiliki fungsi untuk menyeleksi data-data yang tidak dibutuhkan dengan cara menghapusnya. Dalam penelitian ini data yang digunakan adalah Nama, Nilai akhir mata pelajaran wajib, Nilai akhir mata pelajaran kejuruan, dan keterangan prestasi sebagai label kelas target [3].

b. *Data Preprocessing*

Pada tahap ini, adalah tahap *preprocessing* dalam pengolahan data antara lain menghilangkan data ganda, memvalidasi data yang tidak valid dan mengubah jika ada data yang tidak sesuai atau adanya typografi. Pada tahap ini, data dibersihkan untuk meningkatkan keandalan. Ini dilakukan dengan menghilangkan nilai kosong, baris data yang tidak sesuai atau nilai yang berbeda dengan nilai lainnya [3].

c. *Data Transformation*

Proses *data transformation* dilakukan untuk mengembangkan data sehingga lebih baik dipersiapkan dan siap untuk diproses data mining.

Tahap ini digunakan untuk mengubah tipe data atribut yang semua angka menjadi huruf [3].

d. *Data mining*

Data mining adalah proses mencari pola atau informasi dalam dataset. Tahapan ini adalah inti dari tahapan *knowledge discovery in database* (KDD) yang diterapkan untuk menganalisis data yang telah ditentukan menggunakan algoritma *Naïve Bayes* dan algoritma *Decision Tree*, dengan mengukur hasil akurasi masing-masing algoritma sebagai perbandingan [8].

e. *Evaluation*

Setelah memperoleh hasil dari proses data mining, pada tahap ini melibatkan evaluasi dan interpretasi hasil tersebut dengan tujuan memahami makna dan relevansi dari pola atau model yang telah diidentifikasi [8]. Hasil evaluasi dilandasi pada *Confusion Matri*. Evaluasi diperlukan untuk menampilkan hasil dengan memperlihatkan *accuracy*, *precision* dan *recall* [14]. Berikut ini adalah rumus untuk menghitung nilai memperlihatkan *accuracy*, *precision* dan *recall*. Terdapat pada rumus 1,2,3 [15].

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

4. HASIL DAN PEMBAHASAN

4.1. *Data*

Data yang dianalisis pada penelitian ini adalah nilai siswa kelas X Program Keahlian Teknik Komputer Jaringan (TKJ) dan Teknik Bisnis Sepeda Motor (TBSM) Tahun Ajaran 2019/2020. Data nilai siswa ini diperoleh dari bagian akademik sekolah SMK Al-Musyawirin Cirebon. Data tersebut berisi 20 atribut yaitu, No, Nama Siswa, Tempat Lahir, Tanggal Lahir, Nomor Induk Sekolah, Nama Orang Tua, Nomor Induk Siswa Nasional, Program Keahlian, Pendidikan Agama Islam, Pendidikan Kewarganegaraan, Bahasa Indonesia, Bahasa Inggris, Matematika, Sejarah, Seni Budaya, Pendidikan Jasmani Olahraga Kesehatan, Simulasi Digital, Fisika, Kimia dan Bahasa Sunda. Jumlah data penelitian ini berjumlah 202 record.

4.2. *Data Selection*

Tahap pertama pada metode *Knowledge Discovery in Database* (KDD) adalah seleksi data. Tujuan dari tahap ini yaitu untuk memilih data yang akan dipakai dalam metode data mining menggunakan *tools RapidMiner*, dalam data asli yang didapatkan memiliki 20 atribut, kemudian dilakukan seleksi data karena hanya atribut Nama Siswa dan Nilai Mata Pelajaran saja yang digunakan. Data yang tidak digunakan seperti atribut No, Tempat Lahir, Tanggal

Lahir, Nomor Induk Sekolah (NIS), Nama Orang Tua, Nomor Induk Siswa Nasional dan Program Keahlian akan dihapus. Dapat dilihat operator dataset pada gambar 2.



Gambar 2. Operator dataset

No	No	NAMA SISWA	AGAMA	PKN	B. INDONESIA	B. INGGRIS	MATEMATIKA	B. BUNGA	SEJARAH	SEM BUDI
1	1	ACHRIAD REFAL	B	C	C	C	C	C	B	C
2	2	ADY FARIS	B	B	C	B	C	C	B	C
3	3	ADIRINA	B	B	C	B	C	B	C	C
4	4	ATRI ALYANZI	B	C	C	C	B	C	C	C
5	5	ARHAB SANI	B	B	C	B	B	B	B	C
6	6	ARHAB LINTA	B	B	B	B	B	B	B	B
7	7	ALVIN SYARIL	B	B	B	B	B	B	B	B
8	8	ARIZONI SA.	B	B	C	C	C	C	C	C
9	9	ARWIDIANI	B	B	C	C	C	C	B	C
10	10	AUDA SYARIFA	B	B	C	B	B	C	B	C
11	11	ASFA LINDA BINA	B	B	B	B	B	B	B	C
12	12	ATRILIAJE TRIA	B	B	B	B	B	C	B	B
13	13	Ayu Winika	B	B	B	B	C	B	B	C
14	14	DEA SYALWA	B	B	B	B	B	C	B	B
15	15	DENY FREDAUS	B	B	C	C	C	C	B	C
16	16	Donia	B	B	B	B	C	B	B	C
17	17	Fitri Salsabilla	B	B	B	B	C	B	B	C

Gambar 3. Hasil read excel

4.3. Data Preprocessing

Tahap kedua adalah proses pembersihan data yang hilang, tidak lengkap atau tidak sesuai, dalam dataset yang digunakan sudah tidak data yang hilang, tidak lengkap atau tidak sesuai, maka data bisa digunakan ke proses selanjutnya.

4.4. Data Transformation

Tahap ketiga adalah dilakukan proses tranformasi data dimana atribut nilai mata pelajaran yang awalnya berbentuk numerik, diubah menjadi kategori Huruf A, B, C, dan D pada tabel 1 dengan ketentuan nilai yang dapat dilihat pada tabel 2. Serta diberi predikat akhir sebagai hasil nilai.

Tabel 1. Ketentuan data dengan kategori huruf

Nilai Angka	Nilai Huruf	Predikat
90-100	A	Sangat Baik
80-90	B	Baik
70-79	C	Cukup
>70	D	Kurang

Tabel 2. Ketentuan predikat akhir

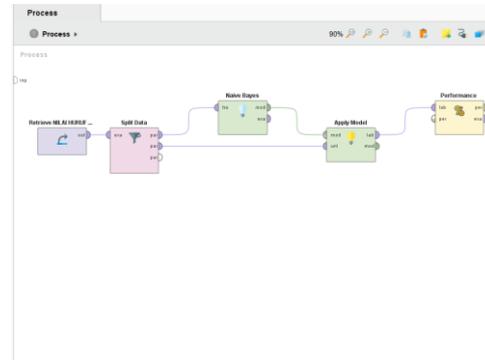
Nilai Angka	Predikat Akhir
83-100	BERPRESTASI
75-82	KURANG BERPRESTASI

4.5. Data Mining

4.5.1. Naïve Bayes

Pada tahap keempat adalah penerapan data minig algoritma Naïve Bayes menggunakan tools Rapidminer. Pengujian data sebanyak 202 record yang telah melalui proses data selection, data preprosessing dan data transformation. Berikut gambar model aturan untuk mengetahui prediksi prestasi siswa menerapkan

algoritma Naïve bayes pada gambar 4. Proses pertama, import dataset dengan pilih Import data, dalam dataset ini menggunakan jenis file excel.



Gambar 4. Model algoritma naïve bayes

Pada pengujian ini menggunakan split data, hasil dari pengujian split data dengan 90% data tranning dan 10% data testing, prediksi menggunakan Naïve Bayes mendapatkan hasil akurasi prediksi kurang berprestasi dan true kurang berprestasi sebesar 17 data, prediksi kurang berprestasi dan benar true berprestasi sebesar 0 data, prediksi berprestasi dan benar true kurang berprestasi sebesar 1 data, prediksi berprestasi dan benar true berprestasi sebesar 2 data. Dapat dilihat pada gambar 5.

	true KURANG BERPRESTASI	true BERPRESTASI	class precision
pred KURANG BERPRESTASI	17	0	100.00%
pred BERPRESTASI	1	2	66.67%
class recall	94.44%	100.00%	

Gambar 5. Hasil akurasi naïve bayes

Untuk mendapatkan hasil Confusion Matrix dari penerapan algoritma Naïve Bayes, Dapat dihitung untuk mencari nilai accuracy, precision, dan recall. Kinerja Confusion Matrix dapat diukur dengan True Positive, False Negative, False Positive dan True Negative.

True Positive adalah positif benar, yang berarti jumlah data benar yang benar memiliki nilai kebenaran benar. False Negative adalah jumlah data tidak benar dengan nilai kebenaran tidak benar.

False Positive adalah positif negatif, yang berarti jumlah data benar yang dipercaya sistem memiliki nilai kebenaran benar dan

True Negative adalah negative negatif, yang berarti jumlah data benar yang dipercaya sistem memiliki nilai kebenaran benar.

Hasil perhitungan Confusion Matrix accuracy, precision, recall pada algoritma Naïve Bayes :

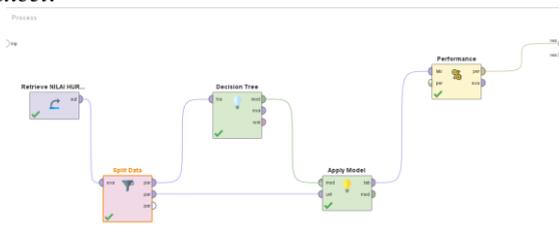
$$Accuracy = \frac{17 + 2}{17 + 0 + 1 + 2} = 0,95 (95,00\%)$$

$$Precision = \frac{17}{17 + 0} = 1 (100\%)$$

$$Recall = \frac{17}{17 + 1} = 0,9444 (94,44\%)$$

4.5.2. Decision Tree

Selanjutnya adalah penerapan data minig algoritma *Decision Tree* menggunakan *tools Rapidminer*. Pengujian data sebanyak 202 record yang telah melalui *proses data selection, data preprocessing* dan *data transformation*. Berikut gambaran model aturan dalam mengetahui prediksi prestasi siswa menggunakan algoritma *Decision Tree* pada gambar 4.1. Proses pertama, *import* dataset dengan pilih *Import data*, dalam dataset ini menggunakan jenis file *excel*.



Gambar 6. Model algoritma Decision Tree

Pada pengujian ini menggunakan *split* data, hasil dari pengujian *split* data dengan 90% data training dan 10% data testing, prediksi menggunakan *Decision Tree* mendapatkan hasil akurasi prediksi kurang berprestasi dan *true* kurang berprestasi sebesar 18 data, prediksi kurang berprestasi dan benar *true* berprestasi sebesar 0 data, prediksi berprestasi dan benar *true* kurang berprestasi sebesar 0 data, prediksi berprestasi dan benar *true* berprestasi sebesar 2 data.

	true KURANG BERPRESTASI	true BERPRESTASI	class precision
pred KURANG BERPRESTASI	18	4	84.44%
pred BERPRESTASI	5	4	44.44%
class recall	85.10%	50.00%	

Gambar 7. Hasil akurasi decision tree

Untuk mendapatkan hasil *Confusion Matrix* dari penerapan algoritma *Decision Tree*, Dapat dihitung untuk mencari nilai *accuracy, precision*, dan *recall*. Kinerja *Confusion Matrix* dapat diukur dengan *True Positive, False Negative, False Positive* dan *True Negative*.

True Positive adalah positif benar, yang berarti jumlah data benar yang benar memiliki nilai kebenaran benar. *False Negative* adalah jumlah data tidak benar dengan nilai kebenaran tidak benar.

False Positive adalah positif negatif, yang berarti jumlah data benar yang dipercaya sistem memiliki nilai kebenaran benar dan

True Negative adalah negative negatif, yang berarti jumlah data benar yang dipercaya sistem memiliki nilai kebenaran benar.

Hasil perhitungan *Confusion Matrix accuracy, precision, recall* pada algoritma *Decision Tree* :

$$Accuracy = \frac{18 + 2}{18 + 0 + 0 + 2}$$

$$Precision = \frac{18}{18 + 0} = 1 (100\%)$$

$$Recall = \frac{18}{18 + 0} = 1 (100\%)$$

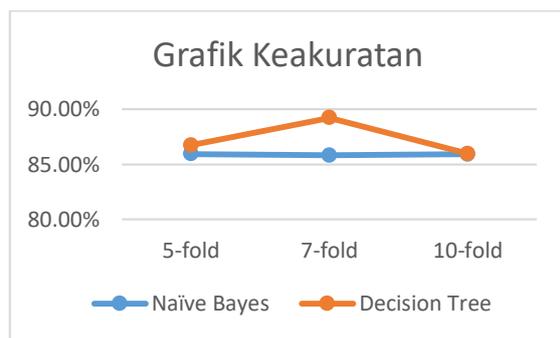
4.5.3. Uji Cross-Validation

Pada pengujian *Cross-Validation* fol yang digunakan adalah nilai *fold 5-fold, 7-fold, dan 10-fold*. Terdapat pada tabel 3. dibawah ini

Tabel 3. Uji cross-validation

Cross-Validation	Keakuratan	
	Naïve Bayes	Decision Tree
5-fold	85,93%	86,73%
7-fold	85,81%	89,22%
10-fold	85,90%	85,96%

Berdasarkan rincian tabel diatas dapat diketahui bahwa nilai keakuratan tertinggi terdapat pada *10-fold*, untuk *Naïve Bayes* sebesar 85.90% dan keakuratan tertinggi terdapat pada *7-fold, Decision Tree* sebesar 89.22%.



Gambar 8. Uji cross-validation

Grafik diatas menjelaskan bahwa nilai *fold* bisa mempengaruhi hasil keakuratan, dimana keakuratan dengan nilai tertinggi pada *Naive Bayes* sebesar 85.93% dengan *5-fold*. Sedangkan keakuratan dengan nilai tertinggi pada *Decision Tree* sebesar 89,22% dengan *7-fold*. Dengan ini metode *Decision Tree* lebih baik dibandingkan metode *Naive Bayes* dalam pengujian *cross-validation*.

5. KESIMPULAN DAN SARAN

Berdasarkan proses hasil data mining perbandingan 2 metode algoritma *Naive Bayes* dan *Decision Tree* menggunakan dataset nilai siswa SMK AL-Musyawirin Weru Cirebon, dapat disimpulkan algoritma *Decision Tree* dan *Naive Bayes* dapat dimanfaatkan dalam memprediksi prestasi belajar siswa, dengan menggunakan *tools Rapidminer*, pengujian dengan menggunakan *split data*. Hasil analisis dan uji *Cross-Validation* menggunakan algoritma *Decision Tree* lebih tinggi dalam akurasi dibandingkan dengan algoritma *Naive Bayes*. Dengan

nilai akurasi algoritma *Decision Tree* sebesar 89,22%. Dan algoritma *Naïve Bayes* sebesar 85,93%. Saran untuk peneliti selanjutnya dapat dikembangkan dengan menggunakan perbandingan algoritma *data mining* lainnya seperti *Random Forest*, *K-Nearest Neighbor*, *Nueral Network* dan *Super Vector Machines* (SVM). Dan saran untuk SMK Al-Muysawirin agar bisa mengajarkan siswanya lebih baik lagi, agar prestasi belajar siswa dapat selalu meningkat.

DAFTAR PUSTAKA

- [1] A. Nurjana, A. P. Windarto, and H. Qurniawan, "Implementasi data mining dalam memprediksi prestasi siswa dengan algoritma c4.5," *SmartEDU*, vol. 1, no. 4, pp. 171–180, 2022, [Online]. Available: <https://ejournal.abivasi.id/index.php/SmartEDU/article/view/45>
- [2] S. Marpaung, S. -, and I. -, "Penerapan Metode Naïve Bayes Dalam Memprediksi Prestasi Siswa Di SMA Negeri 1 Panombeian Panei," *J. Sist. Inf. dan Ilmu Komput. Prima(JUSIKOM PRIMA)*, vol. 4, no. 2, pp. 8–13, 2021, doi: 10.34012/journalsisteminformasidanilmukomputer.v4i2.1522.
- [3] F. M. Almufqi and A. Voutama, "Perbandingan Metode Data Mining Untuk Memprediksi Prestasi Akademik Siswa," *J. Tek.*, vol. 15, no. 1, pp. 61–66, 2023, doi: 10.30736/jt.v15i1.929.
- [4] Bh. Hayadi, "Perbandingan Metode Naive Bayes Dan Random Forest Untuk Menentukan Prestasi Belajar Siswa Pada Jurusan RPL (Studi Kasus SMK Swasta Siti Banun Sigambal)," *J. Comput. Sci. Inf. Technol. Progr. Stud. Teknol. Inf.*, no. 2, p. 2022, 2022, [Online]. Available: <http://jurnal.ulb.ac.id/index.php/JCoInT/index>
- [5] I. C. Wibowo, A. C. Fauzan, M. Dwi, P. Yustiana, and F. A. Qhabib, "Komparasi Algoritma Naive Bayes dan Decision Tree Untuk Memprediksi Lama Studi Mahasiswa," vol. 1, no. 2, pp. 65–74, 2019.
- [6] A. Noviriandini and N. Nurajijah, "Analisis Kinerja Algoritma C4.5 Dan Naïve Bayes Untuk Memprediksi Prestasi Siswa Sekolah Menengah Kejuruan," *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 5, no. 1, pp. 23–28, 2019, doi: 10.33480/jitk.v5i1.607.
- [7] Salsabila Citra Esananda, B. Nugroho, and F. Anggraeny, "Penerapan Algoritma Decision Tree Dalam Menentukan Prestasi Akademik Siswa," *J. Inform. dan Sist. Inf.*, vol. 2, no. 2, pp. 413–424, 2021, doi: 10.33005/jifosi.v2i2.311.
- [8] J. Triwidianti, F. Y. Alfian, and ..., "Perbandingan Metode Data Mining Untuk Prediksi Prestasi Siswa Tingkat Pendidikan Menengah Kejuruan Pada Sekolah Menengah Kejuruan Negeri (SMKN 1) ...," *Pros. Semin.*, vol. 1, no. Smkn 1, pp. 126–133, 2021, [Online]. Available: <https://jurnal.darmajaya.ac.id/index.php/PSND/article/view/2945%0Ahttps://jurnal.darmajaya.ac.id/index.php/PSND/article/viewFile/2945/1265>
- [9] M. Nasution and M. Sofyan, "Penerapan Metode Decision Tree Untuk Memprediksi Prestasi Siswa Kelas XII Dilihat dari Nilai Akhir Semester di SMK Negeri 1 Selong Tahun Pelajaran 2017/2018," *Infotek J. Inform. dan Teknol.*, vol. 3, no. 1, pp. 58–65, 2020, doi: 10.29408/jit.v3i1.1825.
- [10] A. M. Farid, A. Pratama, I. Sholihin, A. R. Rinaldi, and Faturrhohman, "JURNAL DATA SCIENCE & INFORMATIKA (JDSI)," vol. 1, no. 1, pp. 16–20, 2021.
- [11] E. Novianto, A. Hermawan, and D. Avianto, "KLASIFIKASI ALGORITMA K-NEAREST NEIGHBOR, NAIVE BAYES, DECISION TREE UNTUK PREDIKSI STATUS KELULUSAN MAHASISWA S1," *J. Teknol. dan Sist. Inf. Univrab*, vol. 8, pp. 146–154, 2023, [Online]. Available: <https://jurnal.univrab.ac.id/index.php/rabbit/article/view/3434/1538>
- [12] F. M. Almufqi, A. Voutama, S. Informasi, I. Komputer, U. S. Karawang, and T. Timur, "PERBANDINGAN METODE DATA MINING UNTUK MEMPREDIKSI PRESTASI," vol. 15, no. 1, pp. 61–66, 2023, doi: 10.30736/jt.v15i1.929.
- [13] S. E. Choi, J. Kim, and D. Seo, "Travel patterns of free-floating e-bike-sharing users before and during COVID-19 pandemic," *Cities*, vol. 132, no. January 2022, p. 104065, 2023, doi: 10.1016/j.cities.2022.104065.
- [14] W. A. Firmansyach, U. Hayati, and Y. Arie Wijaya, "Analisa Terjadinya Overfitting Dan Underfitting Pada Algoritma Naive Bayes Dan Decision Tree Dengan Teknik Cross Validation," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 262–269, 2023, doi: 10.36040/jati.v7i1.6329.
- [15] U. Pujianto and P. Y. Ristanti, "Perbandingan kinerja metode C4.5 dan Naive Bayes dalam klasifikasi artikel jurnal PGSD berdasarkan mata pelajaran," vol. 29, no. 1, pp. 50–67, 2019.