



Penerapan Metode Klasifikasi dan Seleksi Fitur untuk Memprediksi Minat Studi dan Karir Siswa (Studi Kasus SMK Se-Kabupaten Kudus)

Aditia Candra Kusuma ^{1*}, dan Eka Ardhiyanto ²

¹ Magister Teknologi Informasi, Fakultas Teknologi Informasi dan Industri, Universitas Stikubank; e-mail : aditiacandra0017@mhs.unisbank.ac.id

² Magister Teknologi Informasi, Fakultas Teknologi Informasi dan Industri, Universitas Stikubank; e-mail : ekaardhiyanto@edu.unisbank.ac.id

* Corresponding Author : Aditia Candra Kusuma

Abstract: Every year, vocational high schools in Kudus Regency produce competitive graduates in the national and international job markets, because vocational high school graduates must be equipped with technical skills and soft skills. However, students face a dilemma in determining the next step in choosing a study interest or career. In addition, several factors influence student interests, one of which is family background, so that students who were initially sure will become unsure of their interests. Therefore, the purpose of this study is to classify student interests. The use of classification methods with the naïve bayes algorithm and feature selection with forward selection was chosen in this study. Classification model testing was carried out three times with a comparison of training data and test data (90:10, 80:20, and 70:30), where the results obtained had an average accuracy value of 72.85%, an average precision value of 77.75%, and an average recall value of 86.43%. the classification results are continued with the application of forward selection to find influential attributes, and the results of feature selection show that major attribute, achievement attributes, and mathematics value attributes have an influence on student interest.

Keywords: Student interests; further studies; career; Naïve Bayes; Forward Selection

Abstrak: Setiap tahun, Sekolah Menengah Kejuruan (SMK) di Kabupaten Kudus menghasilkan lulusan yang kompetitif di pasar kerja nasional dan internasional, karena lulusan SMK harus dibekali dengan keterampilan teknis dan soft skill. Akan tetapi, siswa menghadapi dilema dalam menentukan langkah selanjutnya dalam memilih minat studi atau karier, disamping itu ada beberapa faktor-faktor yang mempengaruhi minat siswa salah satunya seperti latar belakang keluarga, sehingga siswa yang awalnya yakin akan menjadi tidak yakin dengan minat mereka. Oleh karena itu tujuan dari penelitian ini adalah mengklasifikasi minat siswa untuk lanjut studi atau berkarier dan atribut apa saja yang berpengaruh terhadap minat siswa. Penggunaan metode klasifikasi dengan algoritma *naïve bayes* dan seleksi fitur dengan *forward selection* yang dipilih dalam penelitian ini. Pengujian model klasifikasi dilakukan sebanyak tiga kali dengan perbandingan data training dan data testing (90:10, 80:20, dan 70:30), Dimana hasil yang diperoleh memiliki nilai rata-rata akurasi sebesar 72.85%, nilai rata-rata precision sebesar 77.75%, dan nilai rata-rata recall sebesar 86.43%. Hasil klasifikasi dilanjutkan dengan penerapan forward selection untuk mencari atribut yang berpengaruh dan hasil dari seleksi fitur menunjukkan atribut jurusan, atribut prestasi, dan atribut nilai matematika memiliki pengaruh terhadap minat siswa.

Kata kunci: Minat siswa; lanjut studi; karier; *Naïve Bayes*; *Forward Selection*

Received: July 27, 2025

Revised: August 14, 2025

Accepted: November 26, 2025

Published: November 29, 2025

Curr. Ver.: November 29, 2025



Copyright: © 2025 by the authors.
Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>)

1. Pendahuluan

Pendidikan kejuruan di Indonesia, khususnya Sekolah Menengah Kejuruan (SMK) dirancang untuk menghasilkan lulusan yang mampu bersaing di pasar kerja nasional dan internasional, karena lulusan SMK menekankan setiap siswa dibekali dengan keterampilan teknis dan *softs skill* sehingga memiliki kemampuan dalam bekerja. SMK di Kabupaten Kudus setiap tahunnya selalu memberikan lulusan terbaik, akan tetapi siswa pasca lulus selalu menghadapi dilema dalam menentukan minatnya untuk langkah berikutnya dalam pemilihan lanjut studi atau karir. Ini terjadi dikarenakan mereka tidak tahu apa yang harus dipilih dan bagaimana lingkungan mempengaruhi karir mereka, siswa yang awalnya yakin menjadi tidak yakin dengan minat mereka [1]. Siswa akan lebih cenderung memilih minat untuk lanjut studi atau karir berdasarkan program studi sesuai pemilihan kejuruan di SMK. Oleh karena sangat penting untuk diketahui mengenai faktor-faktor yang menarik minat siswa ketika mereka memutuskan untuk lanjut studi atau karir setelah mereka lulus.

Secara umum faktor-faktor yang mempengaruhi minat siswa dalam pemilihan untuk lanjut studi atau karir yaitu berdasarkan latar belakang keluarga, prestasi akademik, pengalaman magang, persepsi mereka tentang peluang kerja, dan pengaruh berdasarkan arahan dari guru berdasarkan bimbingan konseling karir. Oleh karena itu pengalihan data diperlukan untuk mengetahui faktor minat siswa lebih lanjut. Salah satu teknik pengalihan data yang paling populer pada basis data adalah teknologi data mining. Dalam bidang pendidikan, data mining berkonsentrasi pada penciptaan jenis data khusus yang dapat dieksplorasi. Ini dapat diciptakan dengan menggunakan metode dari algoritma data mining *machine learning*. Jenis penelitian dan algoritma yang dikenal sebagai *machine learning* berfokus pada penemuan pola dalam data, selanjutnya pola ini akan digunakan untuk mengklasifikasi, klusterisasi, asosiasi, prediksi, perkiraan, dan ekstraksi informasi berguna lainnya dari kumpulan data yang jumlahnya sangat besar [2]. Dalam proses klasifikasi, model dan fungsi yang dapat membedakan atau menjelaskan data yang telah ditemukan, hal ini digunakan untuk memperkirakan kelas objek yang labelnya tidak diketahui [3].

Penerapan model klasifikasi *machine learning*, ada kemungkinan untuk mengembangkan strategi pengelolaan sumber daya manusia yang lebih efektif dan bermanfaat bagi individu maupun organisasi, terutama dalam rekrutmen calon karyawan dan calon mahasiswa baru berdasarkan hasil klasifikasi minat siswa pasca lulus SMK. Namun, hasil yang diberikan hanya berdasarkan minat siswa saja tidak selalu membuahkan hasil yang sesuai, tetapi harus bisa mengetahui faktor mana saja yang sangat berpengaruh perlu di perlukan. Oleh karena itu, pemilihan fitur atau *feature selection* adalah teknik data mining yang digunakan untuk mengetahui faktor-faktor yang paling berpengaruh terhadap minat siswa, dimana teknik ini dapat digunakan untuk pemilihan banyak fitur yang tidak penting maupun mengetahui subset fitur yang paling penting dari data set minat siswa [4].

Salah satu pendekatan yang paling populer dalam berbagai jenis permasalahan klasifikasi adalah menggunakan algoritma *naive bayes*. Dalam studi [5] menunjukkan dengan menerapkan metode *naive bayes* dalam mengklasifikasi kelanjutan siswa ke perguruan tinggi memberikan hasil akurasi sebesar 86.53%, Dimana hasil akurasi tersebut dikatakan baik dalam hal klasifikasi. Tidak hanya dalam kasus kelanjutan siswa ke perguruan tinggi saja, algoritma *naive bayes* juga dapat memberikan hasil akurasi lebih tinggi yaitu sebesar 90% dibandingkan *random forest* yang memiliki nilai akurasi sebesar 70% dalam hal kasus klasifikasi bahasa daerah [3]. Berbagai penelitian juga mencoba menggabungkan metode klasifikasi dengan metode *feature selection*, seperti penggabungan metode *naive bayes* dengan *forward selection* dalam studi kasus pemilihan fitur pada prediksi bimbingan konseling siswa dan hasil menunjukkan nilai akurasi sebesar 94.84%, dimana sebelum melakukan penggabungan metode *selection feature* memiliki nilai akurasi sebesar 94.55% [6].

Berdasarkan dari berbagai keterbatasan dan eksperimen temuan tersebut, penelitian ini bertujuan untuk melakukan klasifikasi terhadap minat siswa SMK untuk memilih lanjut studi atau karir, serta mengetahui atribut apa saja yang berpengaruh terhadap minat tersebut. Metode klasifikasi yang disarankan menggunakan algoritma *naive bayes* dan dilanjutkan pemilihan fitur dengan metode *forward selection*.

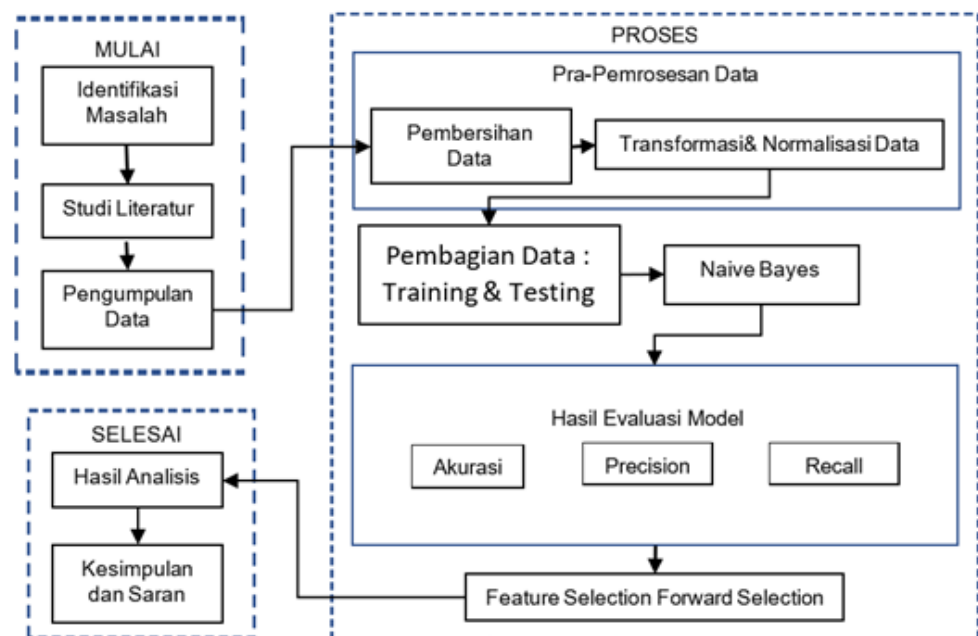
2. Kajian Pustaka atau Penelitian Terkait

Penelitian terdahulu sudah mengungkapkan potensi penggunaan algoritma *naïve bayes* sebagai metode klasifikasi dalam analisis minat siswa. Contohnya seperti penelitian yang telah dilakukan oleh [5] dalam penerapan algoritma *naïve bayes* untuk memprediksi kelanjutan studi siswa ke perguruan tinggi. Berdasarkan penerapan algoritma *naïve bayes*, penelitian ini memberikan hasil prediksi yang baik, karena algoritma *naïve bayes* dapat menghasilkan probabilitas di setiap kriteria dengan kelas yang berbeda. Oleh karena itu, nilai probabilitasnya dapat digunakan untuk mengoptimalkan prediksi dalam hal kelanjutan studi siswa ke perguruan tinggi [5].

Selain itu, penelitian [7] menerapkan metode *feature selection* di metode klasifikasi yang tujuannya berupaya untuk mengembangkan dan meningkatkan kinerja algoritma *naïve bayes* dalam menangani relasi antar fitur menjadi suatu keharusan. Teknik seleksi fitur dapat mengurangi depensi, sehingga akurasi klasifikasi dapat meningkat dan teknik tersebut dapat lebih efektif dalam menangani ketergantungan antar fitur [7]. Kinerja algoritma klasifikasi dapat terbantu dalam hal mengurangi dimensi data dan meningkatkan efisien serta efektifitas dengan pemilihan metode *forward selection* yang bertugas untuk memilih fitur yang relevan terhadap hasil klasifikasi [6].

3. Metode yang Diusulkan

Penelitian ini disusun secara bertahap dan sistematis guna mencapai tujuan utama. Tujuan tersebut adalah untuk membuat model klasifikasi minat siswa SMK (lanjut studi atau karir/kerja) dengan menggunakan algoritma *naïve bayes* dan menemukan atribut yang paling relevan berdasarkan metode *feature selection* (*forward selection*). Langkah-langkah yang diambil berdasarkan tujuan yang ingin dicapai dapat digambarkan dalam gambar 1 berikut.



Gambar 1. Metodologi Penelitian

3.1. Identifikasi Masalah dan Studi Literatur

Penelitian ini berkonsentrasi pada pembuatan model klasifikasi yang dapat memprediksi minat siswa SMK di Kabupaten Kudus pasca lulus, apakah mereka memiliki minat lanjut studi atau karir. Salah satu masalah yang menjadi perhatian adalah bagaimana menganalisis untuk menemukan faktor yang relevan terhadap pengaruh minat siswa. Studi literatur dilakukan berguna untuk pemahaman konsep dasar algoritma klasifikasi, metode seleksi fitur, dan studi sebelumnya berkaitan dengan minat siswa.

3.2. Pengumpulan Data

Pemerolehan data dilakukan dengan pembagian angket yang di berikan kepada siswa/siswi SMK Kabupaten Kudus pada tahun 2023, dengan total record sejumlah 2119 data. Tujuan pengumpulan data dengan pembagian angket kepada siswa adalah salah satu cara untuk menggali informasi yang berkaitan dengan minat siswa.

3.3. Pra-pemrosesan Data

Pra-permosesan data dilakukan sebelum ke tahap analisis dan pengujian model berguna untuk memastikan konsistensi dan kualitas data. Kemudahan tidak jauh dari tahap pembersihan data guna menghapus data dimana record yang bersifat missing values, terjadinya data ganda atau duplikat, dan atribut yang tidak diperlukan. Selanjutnya tahap transformasi data dan normalisasi berguna untuk memungkinkan algoritma machine learning dapat memproses data kategorial seperti jurusan, prestasi, dan dukungan orang tua ditransformasikan menjadi format numerik, sehingga memastikan distribusi nilai berada dalam rentang tertentu.

Setelah data di lakukan pembersihan dan transformasi data, maka atribut yang ditetapkan pada penelitian ini berkaitan dengan minat siswa lanjut studi atau karir dapat diketahui pada tabel 1 berikut.

Tabel 1. Atribut Data Minat Siswa

Atribut	Keterangan
Nama	Nama siswa
Jenis Kelamin	Jenis Kelamin
Jurusan	Jurusan atau Program Studi
Prestasi	Prestasi / Penghargaan
Dukungan Orang Tua	Dukungan tinggal di kota kelahiran atau dibebaskan
Rata-Rata Nilai Raport	Nilai raport
Nilai Matematika	Nilai matematika
Nilai Bahasa Inggris	Nilai Bahasa Inggris
Nilai Bahasa Indonesia	Nilai Bahasa Indonesia
Nilai Ujian Praktek	Nilai Ujian Praktek
Minat	label

Berdasarkan tabel 1 diatas, maka dataset yang siap untuk di lanjutkan dalam pembentukan model klasifikasi dan pengujian model dapat dilihat pada tabel 2 berikut.

Tabel 2. Dataset Minat Siswa

No	Nama	Jurusan	Jenis Kelamin	Prestasi	Dukungan Orang Tua	Rata-rata Raport	Nilai Matematika	Nilai Bahasa Inggris	Nilai Bahasa Indonesia	Nilai Ujian Praktik	Minat
1	Wahyu Ramandhani	APHP	P	0	2	98	69	84	95	64	KERJA
2	Silviyatul Auliya	APHP	P	0	2	88	74	93	80	95	KULIAH
3	Rahma Aulia Nurvaniza	FARMASI	P	0	2	74	76	62	65	95	KULIAH
4	Aditya Nugroho	FARMASI	L	0	2	67	77	83	62	62	KULIAH
5	Bagas Aditya Ribowo	FARMASI	L	1	2	80	68	68	69	83	KULIAH
6	syifa karima	FARMASI	P	0	2	98	84	85	64	76	KULIAH
7	Dwi Nailil Muna	FARMASI	P	0	2	78	73	79	85	98	KULIAH
8	Khoirunnisa'	FARMASI	P	1	1	82	84	69	80	85	KULIAH
9	Nelis Irawati Ningsih	FARMASI	P	0	2	70	81	78	85	63	KULIAH
10	firda tri fannya	FARMASI	P	0	2	70	89	62	97	71	KULIAH
11	Nida' Az zahra'	FARMASI	P	1	2	83	79	83	85	60	KERJA
12	Sofiya Himawati	APHP	P	0	1	95	80	64	67	80	KERJA
13	Naela Cahyana Putri	FARMASI	P	0	2	83	91	62	64	78	KULIAH
14	HILDA AULIYA RISMAYANTI	FARMASI	P	0	1	62	80	86	74	80	KULIAH
15	Indriyani Pemenangan	FARMASI	P	0	2	81	94	75	75	77	KULIAH
16	Elvina Damayanti	FARMASI	P	0	2	89	98	71	96	69	KERJA
17	Aris Lestariningsih	FARMASI	P	1	2	97	71	85	62	74	KULIAH
18	Nazarul Kayla A.R	FARMASI	P	0	2	61	77	93	90	64	KULIAH
19	Ghea Dayu Rahmadani	FARMASI	P	0	2	80	80	82	72	63	KULIAH
20	Alya Intan Khoirunnisa	FARMASI	P	0	2	92	72	90	75	87	KULIAH

21	Vania Yuven Salsabila	FARMASI	P	1	2	71	80	79	80	63	KULIAH
22	Desti Az Zahra	FARMASI	P	0	2	81	67	71	61	74	KERJA
23	Vera Irmayanti	FARMASI	P	0	1	84	66	82	86	76	KULIAH
24	Jesika Putri Damayanti	APHP	P	2	1	86	82	63	97	62	KERJA
25	Muhammad Abdul Aziz	APHP	L	1	2	87	91	82	85	69	KULIAH
26	Wulandari	APHP	P	0	2	75	70	60	97	82	KERJA
27	Nita Aurelia Putri	APHP	P	0	1	74	81	84	84	77	KERJA
28	Krida Taufany Bunga Panca Sakti	FARMASI	P	0	1	96	84	89	67	94	KERJA
29	Silviyatul Auliya	FARMASI	P	1	2	66	81	62	67	90	KULIAH
30	Muhammad Anif Nur Huda	FARMASI	L	1	2	80	87	80	92	60	KULIAH

Data diperoleh pada Tahun 2023

3.4. Pembagian Data

Dataset akan dibagi menjadi dua bagian yaitu data training digunakan untuk membangun model dan data testing digunakan untuk mengevaluasi kinerja model, dimana pelaksanaannya di lakukan sebanyak 3 kali, pertama perbandingan data training 90% dan data testing 10%, kedua perbandingan data training 80% dan data training 20%, ketiga perbandingan data training 70% dan data tesing 30% [8].

3.5. Algoritma *Naive Bayes*

Pada tahap ini penelitian menggunakan bantuan aplikasi rapidminer dalam menerapkan algoritma *naive bayes* sebagai proses klasifikasi dataset minat siswa. Pemilihan algoritma ini didasarkan pada prinsip asumsi independensi antar fitur dan probabilitas karena efisiensi, komputasi, kemampuan untuk menangani data kategorikal dan numerik, dan kompatibilitasnya dengan data Pendidikan multimediasi.

3.6. Evaluasi Model

Pengukuran performance seberapa baik model melakukan klasifikasi, evaluasi dilakukan secara otomatis dengan bantuan platform rapidminer, dimana aplikasi tersebut menyediakan pengujian dengan hasil confusion matrix dengan detail per kelas. Pengujian akurasi dievaluasi menggunakan tabel matrix yang hasilnya antara lain adalah akurasi yang merupakan persentase prediksi yang benar terhadap total data, *precision* yang merupakan hasil prediksi yang tepat bernilai positif, dan *recall* merupakan persentase prediksi yang benar dari kasus actual yang berhasil diprediksi.

3.7 Seleksi Fitur (*Forward Selection*)

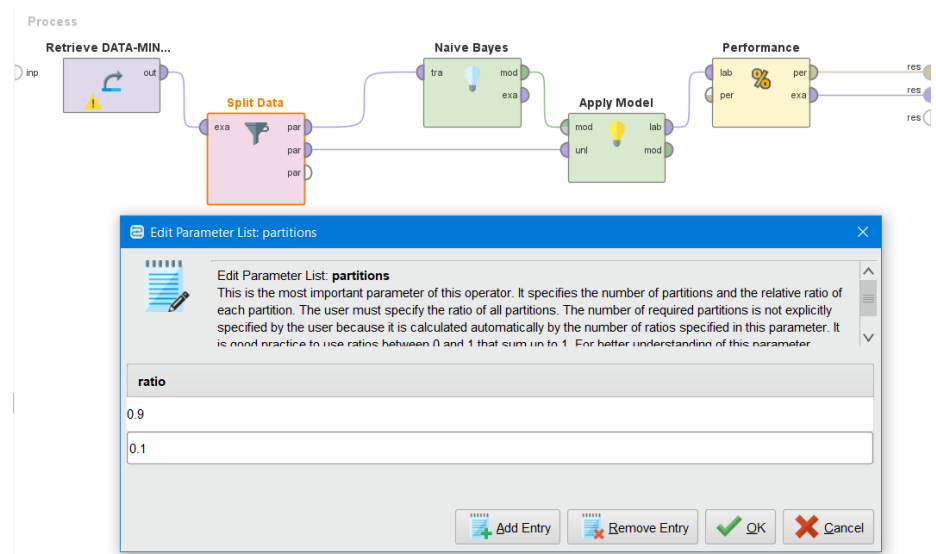
Setelah pembentukan dan evaluasi atau pengujian model *naive bayes*, langkah berikutnya adalah pemilihan fitur dengan menerapkan *forward selection*. Tujuan dari langkah ini adalah untuk menyederhanakan model, menemukan fitur yang relevan terhadap hasil klasifikasi yang sebelumnya sudah di lakukan, serta dapat meningkatkan akurasi dan efisiensi pembentukan model. Metode *forward selection* bekerja secara teratur dari awal hingga akhir dengan cara bertahap dalam menambahkan atribut untuk peningkatan kinerja model, dimulai dari kosong hingga tidak ada lagi peningkatan yang signifikan. Proses ini dilakukan dengan cara menghubungkan algoritma *naivebayes* kedalam operator seleksi pada aplikasi rapidminer. Kemudian berdasarkan akurasi, algoritma ini mengevaluasi kombinasi fitur terbaik.

4. Hasil dan Pembahasan

Berdasarkan tujuan penelitian ini yaitu untuk memprediksi minat studia tau karir pasca lulus sekloah SMK. Dataset yang berjumlah 2119 merupakan profile siswa yang akan dianalisis, masing-masing berisi atribut seperti jenis kelamin, jurusan, prestasi, dukungan orang tua, rata-rata nilai raport, nilai matematika, nilai Bahasa inggris, nilai ujian praktek, dan seagai label adalah minat. Berikut merupakan hasil implemntasi dan evaluasi berdasarkan penerapan metode klasifikasi dengan algoritma *naive bayes* dan seleksi fitur (*forward selection*) pada dataset minat siswa. Sesuai dengan pembahasan yang telah diuraikan sebelumnya, hasil penelitian yang sudah diuji coba dengan model prediksi akan dijelaskan di bagian ini, mencakup pengujian akurasi pada aplikasi rapidminer.

4.1. *Naive Bayes Classifier*

Pembentukan model dan pengujian model klasifikasi dengan algoritma *naive bayes* berdasarkan rasio perbandingan data training 90% dan data testing 10 % dapat dilihat pada gambar 2 dan gambar 3 berikut ini.



Gambar 2. Model *Naive Bayes* dengan Rasio 90% : 10%

Model klasifikasi dengan algoritma *naïve bayes* yang diuji akan memberikan hasil dalam bentuk *confusion matrix* dan *performance vector* seperti pada gambar 3 berikut.

accuracy: 73.11%

	true KERJA	true KULIAH	class precision
pred. KERJA	131	38	77.51%
pred. KULIAH	19	24	55.81%
class recall	87.33%	38.71%	

PerformanceVector

```

PerformanceVector:
accuracy: 73.11%
ConfusionMatrix:
True:  KERJA  KULIAH
KERJA: 131    38
KULIAH: 19    24
weighted_mean_recall: 63.02%, weights: 1, 1
ConfusionMatrix:
True:  KERJA  KULIAH
KERJA: 131    38
KULIAH: 19    24
weighted_mean_precision: 66.66%, weights: 1, 1
ConfusionMatrix:
True:  KERJA  KULIAH
KERJA: 131    38
KULIAH: 19    24
correlation: 0.295
squared_correlation: 0.087

```

Gambar 3. Hasil Pengujian *Naive Bayes* dengan Rasio 90% : 10%

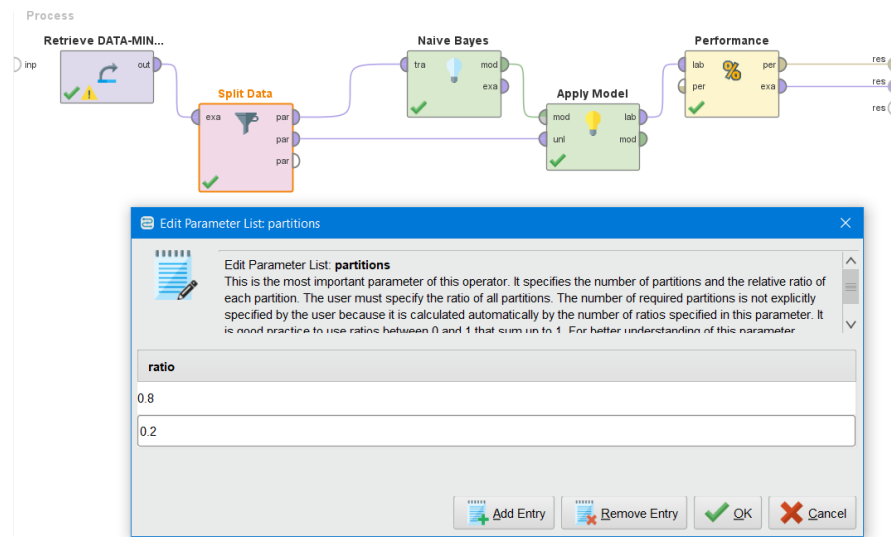
Confusion matrix yang terdiri dari *true positif* dan *true negative* disertakan dengan *performance* model yang di uji seperti akurasi, *precision*, dan *recall* dengan perhitungan berikut.

$$accuracy = \frac{24 + 131}{24 + 131 + 38 + 19} = 73.11\%$$

$$precision = \frac{131}{131 + 38} = 77.51\%$$

$$recall = \frac{131}{131 + 19} = 87.33\%$$

Pembentukan model dan pengujian model klasifikasi dengan algoritma naïve bayes berdasarkan rasio perbandingan data training 80% dan data testing 20% dapat dilihat pada gambar 4 dan gambar 5 berikut.



Gambar 4. Model *Naïve Bayes* dengan Rasio 80% : 20%

accuracy: 72.64%

	true KERJA	true KULIAH	class precision
pred. KERJA	258	73	77.95%
pred. KULIAH	43	50	53.76%
class recall	85.71%	40.65%	

PerformanceVector

```

PerformanceVector:
accuracy: 72.64%
ConfusionMatrix:
True:  KERJA  KULIAH
KERJA: 258    73
KULIAH: 43    50
weighted_mean_recall: 63.18%, weights: 1, 1
ConfusionMatrix:
True:  KERJA  KULIAH
KERJA: 258    73
KULIAH: 43    50
weighted_mean_precision: 65.85%, weights: 1, 1
ConfusionMatrix:
True:  KERJA  KULIAH
KERJA: 258    73
KULIAH: 43    50
correlation: 0.289
squared_correlation: 0.084
  
```

Gambar 5. Hasil Pengujian Model *Naïve Bayes* dengan Rasio 80% : 20%

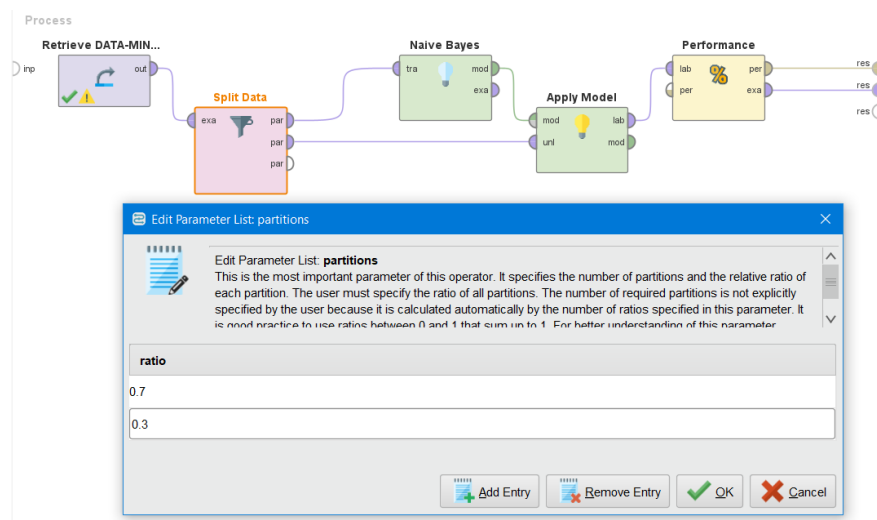
Confusion matrix pada gambar 5, dimana *performance* model yang di uji seperti akurasi, *precision*, dan *recall* dengan perhitungan berikut.

$$\text{accuracy} = \frac{258 + 50}{258 + 50 + 73 + 43} = 72.64\%$$

$$precision = \frac{258}{258 + 73} = 77.95\%$$

$$recall = \frac{258}{258 + 43} = 85.71\%$$

Pembentukan model dan pengujian model klasifikasi dengan algoritma naïve bayes berdasarkan rasio perbandingan data training 70% dan data testing 30% dapat dilihat pada gambar 6 dan gambar 7 berikut



Gambar 6. Model *Naive Bayes* dengan Rasio 70% : 30%

accuracy: 72.80%

	true KERJA	true KULIAH	class precision
pred. KERJA	389	111	77.80%
pred. KULIAH	62	74	54.41%
class recall	86.25%	40.00%	

PerformanceVector

```
PerformanceVector:
accuracy: 72.80%
ConfusionMatrix:
True:  KERJA  KULIAH
KERJA: 389    111
KULIAH: 62    74
weighted_mean_recall: 63.13%, weights: 1, 1
ConfusionMatrix:
True:  KERJA  KULIAH
KERJA: 389    111
KULIAH: 62    74
weighted_mean_precision: 66.11%, weights: 1, 1
ConfusionMatrix:
True:  KERJA  KULIAH
KERJA: 389    111
KULIAH: 62    74
correlation: 0.291
squared_correlation: 0.085
```

Gambar 7. Hasil Model *Naïve Bayes* dengan Rasio 70% : 30%

Confusion matrix pada gambar 7, dimana *performance* model yang di uji seperti akurasi, *precision*, dan *recall* berdsarakan perhitungan sebagai berikut.

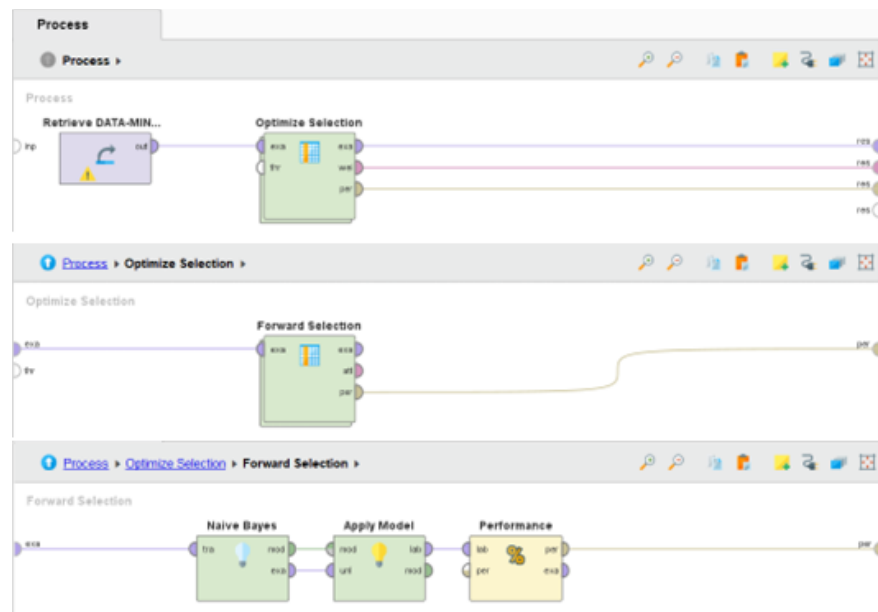
$$accuracy = \frac{389 + 74}{389 + 74 + 111 + 62} = 72.80\%$$

$$precision = \frac{389}{389 + 111} = 77.80\%$$

$$recall = \frac{389}{389 + 62} = 86.25\%$$

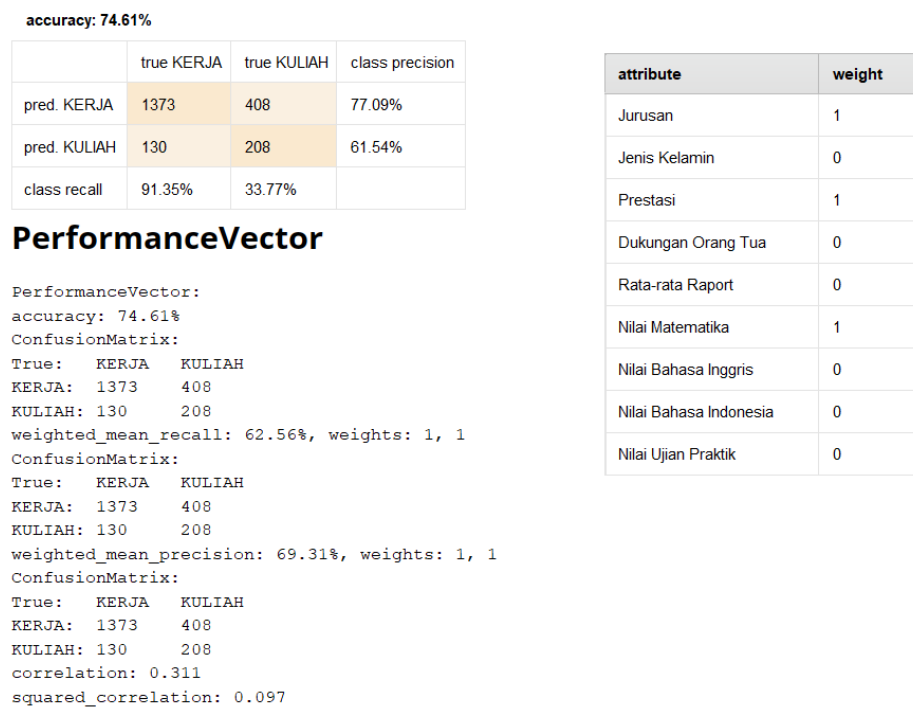
4.2. Feature Selection (Forward Selection)

Setelah data minat siswa diinnalisis dengan model klasifikasi algoritma naïve bayes, hasil analisis ini dilanjutkan dengan pemilihan fitur menggunakan metode forward selection yang diimplementasikan di aplikasi rapidminer, dimana model atau desain penerapannya dapat dilihat pada gambar 8 sebagai berikut.



Gambar 8. Model Feature Selection (Forward Selection)

Berdasarkan gambar 8, pada lembar kerja pertama terdapat data minat siswa dan *optimize selection* yang saling terhubung, kemudian di dalam bagian *optimize selection* terdapat lembar kerja yang berisikan *forward selection* yang tujuannya untuk mencari fitur atau atribut terbaik. Dimana di dalam *forward selection* terdapat lembar kerja yang berisikan model *naive bayes*, *apply model* dan *performance classification* sebagai hasil klasifikasi.

Gambar 9. Hasil Model *Feature Selection* (Forward Selection)

Gambar 9 menunjukkan hasil dari

4.3. Pembahasan

a. Penerapan Metode *Naïve Bayes* Dalam Klasifikasi Minat Siswa

Tabel 3. Hasil Pengujian Klasifikasi Model *Naïve Bayes* Pada Data Minat Siswa SMK Se-Kabupaten Kudus Tahun 2023

	Perbandingan Data Training : Data Testing			Rata-Rata
	90% : 10%	80% : 20%	70% : 30%	
Accuracy	73.11%	72.64%	72.80%	72.85%
Precision	77.51%	77.95%	77.80%	77.75%
Recall	87.33%	85.71%	86.25%	86.43%

Hasil dari proses klasifikasi pada tabel 3., menunjukkan bahwa model *naïve bayes* mampu mengashilkan rata-rata akurasi sebesar 72.85%, *precision* dengan rata-rata sebesar 77.75% dan hasil dari *recall* dengan rata-rata sebesar 86.43%.

b. Penerapan Metode *Feature Selection* Terhadap Hasil Klasifikasi Minat Siswa

Berdasarkan gambar 9 hasil pemilihan fitur dengan menerapkan forward selction, menunjukkan terdsapat 3 atribut yang paing berpengaruh terhadap klasifikasi minat siswa antara lain sebagai berikut.

1) Jurusan

Selama Pendidikan di Sekolah Menengah Kejuruan (SMK), program Pendidikan atau program studi memberikan pengaruh terhadap minat siswa. Misalnya, siswa dari jurusan tenkik yang bersifat keterampilan langsung aplikatif, seperti teknik listri, pada umunya cenderung merasa siap untk bekerja lngangsung karena meraka memiliki keahlian yang sesuai dengan kebutuhan industry. Sebaliknya, siswa yang memilih program studi bersifat administrative biasanya akan memiliki minat untuk lanjut studi di perguruan tinggi. Oleh karena itu, jurusan bukan hanya menjadi identitas akademik bagi siswa, akan tetapi juga mempengaruhi tujuan masa depan, seperti kesiapan kerja atau lanjut studi.

2) Prestasi

Siswa dengan prestasi yang baik cenderung lebih termotivasi dan lebih percaya diri untuk maju ke jenjang pendidikan berikutnya. Siswa yang berprestasi sering mendapatkan dukungan lebih besar dari orang-orang disekitar mereka, seperti orang tua, teman, dan guru untuk meningkatkan keterampilan atau keahlian di perguruan tinggi. Prestasi juga dapat menunjukkan keinginan untuk belajar dalam jangka Panjang, yang mencakup keinginan untuk terus meningkatkan diri melalui Pendidikan formal.

3) Nilai Matematika

Matematika dianggap sebagai salah satu bidang penting yang membutuhkan kemampuan berpikir sistematis dan logika. Siswa dengan nilai matematika yang baik cenderung memiliki kemampuan analisis yang baik, yang diperlukan untuk melanjutkan Pendidikan, khususnya dalam bidang bisnis, sains, dan teknik. namun, perlu diingat bahwa minat siswa tidak terbatas pada nilai matematika, akan tetapi hanya berfungsi sebagai penilaian tambahan tentang seberapa siap siswa dalam untuk melanjutkan pendidikan mereka.

5. Kesimpulan

Bersarakan penggunaan data minat siswa SMK Kabupaten Kudus, algoritma *naïve bayes* telah terbukti efektif dalam mengklasifikasi minat siswa setelah lulus sekolah, termasuk penilaian mereka untuk melanjutkan studi ke perguruan tinggi atau langsung memasuki dunia kerja. Model *naïve bayes* klasifikasi minat siswa mampu memberikan hasil akurasi rata-rata sebesar 72.85%, *precision* rata-rata sebesar 77.75%, dan *recall* dengan rata-rata sebesar 86.43%. nilai-nilai ini menunjukkan bahwa *naïve bayes* mampu mengenali pola data secara konsisten dan efektif, bahkan pada data yang heterogen seperti data minat siswa. oleh karena itu, metode ini dapat diandalkan sebagai langkah awal dalam sistem prediksi keputusan karir atau studi siswa berbasis data. Guna menilai kontribusi masing-masing atribut terhadap akurasi model klasifikasi, penelitian ini juga menggunakan metode seleksi fitur dengan forward selection. Hasil dari prosedur ini menunjukkan bahwa tidak semua fitur memiliki dampak yang signifikan. Jurusan, prestasi, dan nilai matematika adalah atribut yang terbukti paling berpengaruh dalam menentukan minat siswa. selain itu, penerapan *forward selection* dengan cara bertahap dalam pemilihan fitur mampu meningkatkan nilai akurasi yang signifikan terhadap model klasifikasi.

Daftar Pustaka

- [1] D. Nofitasary, Y. Rahmawati, and Suprianto, "Klasifikasi Minat Karir Siswa Sekolah Menengah Atas Menggunakan Algoritma C4.5," *Jutisi*, vol. 13, pp. 711–724, 2024.
- [2] A. Doahir and A. N. Qolbi, "Analisis Potensi Siswa Untuk Kuliah Dengan Classification Menggunakan Metode Decision Tree," *J. POROS Tek.*, vol. 14, no. 1, 2022.
- [3] G. M. Momole, "Perbandingan Naïve Bayes dan Random Forest Dalam Klasifikasi Bahasa Daerah," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, 2022, doi: 10.35957/jatisi.v9i2.1857.
- [4] N. I. P. N. M. F. B. B. M. A. Nazal, "Analisis Perbandingan Metode Feature Selection Backward Method dan Stepwise Method," *J. Stat. Data Sci.*, 2024.
- [5] G. W. N. Wibowo, "PREDIKSI KELANJUTAN STUDI SISWA KE PERGURUAN TINGGI DENGAN NAIVE BAYES," *J. DISPROTEK*, vol. 11, no. 1, 2020, doi: 10.34001/jdpt.v11i1.1159.
- [6] M. R. Fanani, "Penggabungan Forward Selection untuk Pemilihan Fitur pada Prediksi Bimbingan Konseling Siswa dengan Menggunakan Algoritma Naive Bayes," *Smart Comp Jurnalnya Orang Pint. Komput.*, vol. 9, no. 2, 2020, doi: 10.30591/smartcomp.v9i2.1924.
- [7] Suprpto, "Improvement Naive Bayes Menggunakan Forward Selection, Information Gain dan Gain Ratio untuk Penanganan Independensi Fitur," *J. Sos. dan Teknol.*, vol. 5 No. 4, 2025.

-
- [8] L. Pebrianti, E. Simamora, U. Manullang, N. Taufiq, and Chairunisah, “PERBANDINGAN METODE ALGORITMA SUPERVISED NAiVE BAYES DAN SUPPORT VECTOR MACHINE (SVM) DALAM KLASIFIKASI PENDERITA STUNTING DI KABUPATEN DELI SERDANG,” *Mhs. Tek. Inform.*, vol. 9 No. 3, 2025.