

PREDIKSI SISWA-SISWI SMK AZ-ZAHRA MLONGGO MASUK PASAR KERJA MENGGUNAKAN ALGORITMA NAIVE BAYES

Akhmad Khanif Zyen, Nur Aeni Widiastuti

Universitas Islam Nahdlatul Ulama Jepara

khanif.zyen@unisnu.ac.id, nuraeniwidiastuti@unisnu.ac.id

Abstract

Every year, most students of Azzahra Vocational School are expected to work directly according to their fields, but in processing the level of achievement scores of students who determine the criteria for entering the labor market, they still use the manual method which has many weaknesses and has the potential to cause errors. There is a need for a Classification Algorithm to solve this problem, there is one Classification Algorithm that has proven accuracy and good speed for solving these problems, namely the Naive Bayes Algorithm. Therefore the author analyzes the predictions of students entering the labor market using the Naïve Bayes method to determine whether they pass or not enter the labor market criteria. The author makes calculations for the Naïve Bayes method manually using Excel and using the Rapidminer 5.3 supporting application, where in this study the data used for 2018-2020 is 247 in total, the dataset is divided into 2 parts, namely 152 training data and 75 testing data, which predicts students who passed the criteria for entering the labor market were 57 people and those who did not pass the criteria for entering the labor market were 18 people. This is different from the student testing data where 46 people passed the criteria for entering the labor market and 29 people did not pass the criteria for entering the labor market and the accuracy obtained obtained a score of 77.33%. So, it can be concluded that testing using the Naïve Bayes method can support predicting whether or not students graduate in entering the job market.

Keywords: Rapidminer; Naïve Bayes; Enters the Job Market;

Abstrak

Siswa-siswi SMK Azzahra setiap tahunnya sebagian besar diharapkan langsung bekerja sesuai bidangnya, tapi dalam pengolahan tingkat nilai prestasi siswa yang menentukan kriteria masuk pasar kerja masih menggunakan cara manual yang terdapat banyak kelemahan dan berpotensi mengakibatkan kesalahan. Perlu adanya Algoritma Klasifikasi untuk memecahkan permasalahan ini, ada salah satu Algoritma Klasifikasi yang sudah terbukti akurasi dan kecepatan yang baik untuk memecahkan permasalahan tersebut yaitu Algoritma *Naive Bayes*. Oleh karena itu penulis menganalisis prediksi siswa-siswi masuk pasar kerja dengan metode *Naive Bayes* untuk menentukan lulus tidaknya masuk kriteria pasar kerja. Penulis membuat perhitungan metode *Naive Bayes* secara manual menggunakan Excel dan menggunakan aplikasi pendukung Rapidminer 5.3, dimana dalam penelitian ini data yang digunakan tahun 2018-2020 dengan total 247, dataset tersebut dibagi menjadi 2 bagian, yakni 152 data training dan 75 data testing, dimana memprediksi siswa yang lulus kriteria masuk pasar kerja sebanyak 57 orang dan yang tidak lulus kriteria masuk pasar kerja sebanyak 18 orang. Hal ini berbeda dengan data testing siswa dimana yang lulus kriteria masuk pasar kerja sebanyak 46 orang dan yang tidak lulus kriteria masuk pasar kerja sebanyak 29 orang dan akurasi yang diperoleh mendapatkan nilai sebesar 77,33%. Jadi, dapat disimpulkan bahwa pengujian menggunakan metode *Naive Bayes* dapat mendukung untuk memprediksi lulus tidaknya siswa-siswi dalam masuk pasar kerja.

Kata kunci: Rapidminer; Naïve Bayes; Masuk Pasar Kerja;

PENDAHULUAN

Pendidikan memiliki peran penting dalam meningkatkan sumber daya manusia serta menjadi kunci kemajuan suatu

bangsa, dimana pendidikan merupakan suatu usaha menciptakan proses pembelajaran agar peserta didik dapat mengembangkan potensi diri secara aktif

untuk mendapatkan ketrampilan, akhlak mulia, kecerdasan, pengendalian diri dan kekuatan spiritual keagamaan yang diperlukan oleh dirinya sendiri dan masyarakat.

SMK Az Zahra Mlonggo merupakan salah satu sekolah menengah kejuruan yang ada di kabupaten Jepara. Beralamat di jalan raya jepara-bangsri km.12 kuwasen kabupaten jepara. Lembaga pendidikan ini mengkombinasikan kurikulum dari pemerintah dengan kurikulum islami, sehingga diharapkan ketika lulus siswa mempunyai kompetensi keahlian yang baik sesuai bidang jurusannya serta memiliki akhlak yang baik. Sampai saat ini SMK Az Zahra memiliki 3 kompetensi keahlian, yaitu : multimedia, produksi dan siaran program televisi, teknik dan bisnis sepeda motor. SMK Az Zahra Mlonggo merupakan sekolah menengah kejuruan yang sangat mengedepankan nilai sikap (attitude), nilai prestasi siswa dan nilai skill (keterampilan) lulusan. Sehingga diharapkan lulusan SMK Az Zahra akan mudah diserap dan diterima di dunia industry atau lapangan pekerjaan.

Pengangguran merupakan penyebab terbesar dari kemiskinan. Pemerintah berupaya untuk menangani masalah ketenagakerjaan tersebut melalui kebijakan-kebijakan yang dikeluarkan. Kebijakan-kebijakan tersebut berupa kebijakan yang langsung bersinggungan dengan masalah ketenagakerjaan, maupun kebijakan pendukung berupa kebijakan pelayanan publik dan kebijakan tentang peningkatan kualitas pelayanan publik. Undang-Undang Nomor 13 Tahun 2003 menjadi payung hukum layanan ketenagakerjaan di Indonesia. Undang-undang tersebut mengatur masalah pelayanan terhadap pencari kerja meliputi pelatihan kerja, penempatan kerja, perluasan kerja, hubungan kerja, dan beberapa masalah ketenagakerjaan lainnya. (Zuhdiyaty & Kaluge, 2018).

Kelulusan siswa adalah suatu rangkaian tahapan proses yang harus dinilai oleh setiap siswa, diantaranya harus menyelesaikan sejumlah pembelajaran yang telah ditentukan, dan harus memenuhi syarat-syarat dan ketentuan yang telah ditetapkan oleh pihak sekolah, karena kelulusan menjadi acuan agar bisa

melanjutkan pendidikan ke jenjang yang lebih tinggi maupun bekerja di dunia usaha serta dunia industri. Maka dari itu, penelitian tersebut akan memanfaatkan data tentang kelulusan siswa dengan mengolahnya menggunakan Data Mining untuk mendapatkan informasi berupa prediksi kelulusan siswa. Algoritma yang akan digunakan adalah algoritma Naive Bayes. Atribut yang digunakan dalam memprediksi kelulusan siswa ini adalah Nilai Praktik, US (Ujian Sekolah) dan Perilaku Siswa. Pengujian Data Mining dengan menggunakan algoritma Naive Bayes menghasilkan accuracy 76.25%, precision 93.33%, recall 78.87% dan nilai AUC sebesar 0.751. Dapat disimpulkan bahwa, algoritma Naive Bayes memiliki akurasi yang sangat baik pada penelitian ini. (Suwarno et al., 2021).

Dalam mengatasi masalah tersebut, peneliti berniat untuk menggunakan Data Mining dalam Prediksi menggunakan algoritma Naive Bayes dalam rangka untuk memprediksi siswa-siswi yang masuk dalam kriteria yang diinginkan oleh pasar kerja, diharapkan dapat membantu menemukan informasi dalam memprediksi siswa-siswi masuk pasar kerja sehingga membantu pihak sekolah untuk memprediksi status siswa-siswanya dan dapat digunakan sebagai sarana untuk menentukan langkah tentang pembelajaran kedepannya.

METODE PENELITIAN

2.1. METODE PENGUMPULAN DATA

Pengumpulan data yang digunakan dalam menyusun serta melengkapi data adalah dengan cara observasi, wawancara dan studi pustaka.

1. Observasi

Pengamatan langsung diadakan untuk memperoleh data yang dilakukan pada sekolah terkait. Data yang digunakan adalah data nilai siswa-siswi kriteria yang diinginkan oleh pasar kerja, dimana nilai tersebut didapatkan dari penilaian oleh SMK Az-Zahra. Data yang didapatkan adalah data tahun 2018-2020 dengan total 247.

2. Wawancara

Wawancara dilakukan dengan cara berkomunikasi langsung dengan kepala

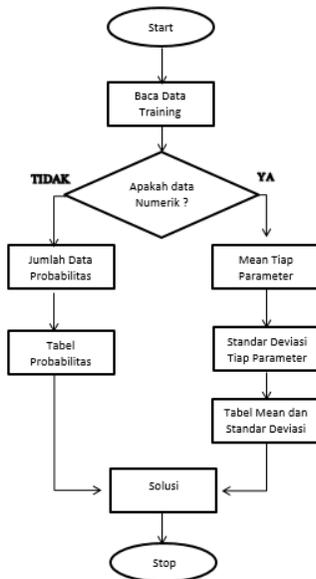
sekolah dan karyawan Tata Usaha SMK AZ-ZAHRA Mlonggo

3. Studi Pustaka

Penulis melakukan studi kepustakaan melalui literatur-literatur atas referensi-referensi yang ada di perpustakaan, maupun dari *internet* dan *e-book*.

2.2. METODE ANALISIS DATA

Metode untuk menganalisis data menggunakan Metode Naive Bayes yang merupakan metode pengklasifikasi probabilitas sederhana yang mengaplikasikan Teorema Bayes dengan berasumsi tidak ketergantungan independen yang tinggi. Keuntungan metode Naive Bayes adalah metode ini hanya memerlukan jumlah data pelatihan training data yang kecil untuk menentukan estimasi parameter yang dibutuhkan dalam proses pengklasifikasian.



Gamabr 1 Alur Metode Naïve Bayes

2.3 Evaluasi dan Validasi

Pada tahap terakhir model yang diusulkan pada penelitian ini pengujian dengan menggunakan *Confusion Matrix* untuk mengetahui tingkat akurasi. Pengujian dilakukan untuk menguji perhitungan dalam data training dengan menggunakan data testing. Pada tahap pengujian ini pula dilakukan perhitungan untuk menguji dan mengukur tingkat *precision*, *recall*, dan *accuracy* dari algoritma *naive bayes* (Rosandy, 2019).

Presisi didefinisikan sebagai rasio item relevan yang dipilih terhadap semua item yang terpilih. Presisi dapat diartikan

sebagai kecocokan antara permintaan informasi dengan jawaban terhadap permintaan tersebut (Rosandy, 2019). Rumus presisi adalah :

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

Recall didefinisikan sebagai rasio dari item relevan yang dipilih terhadap total jumlah item relevan yang tersedia. *Recall* dihitung dengan rumus :

$$Recall = \frac{TP}{TP+FN} \dots\dots\dots (3)$$

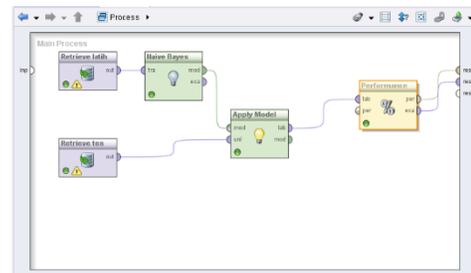
Accuracy merupakan perhitungan terhadap proporsi dari jumlah total prediksi yang benar. Rumus accuracy adalah :

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \dots\dots\dots (4)$$

HASIL DAN PEMBAHASAN

3.1 Uji Sistem Metode Naïve Bayes

Pada Pengujian ini penulis menggunakan Aplikasi pendukung berupa RapidMiner . Berikut adalah Gambar Proses pada RapidMiner:



Gambar 2 Proses metode Naive Bayes pada Rapidminer

Pada penelitian ini dataset siswa dibagi menjadi 2 (dua) bagian yaitu 70% digunakan sebagai *data training* dan 30% digunakan sebagai *data testing* dari 247 data siswa. Pembagian data dilakukan secara manual menggunakan excel.

3.2 Hasil ExampleSet Data Testing

Pada hasil prediksi menunjukkan bahwa ada 17 data yang tidak sesuai dengan data testing siswa. Dari pengujian aplikasi Rapi Miner 5.3 menggunakan algoritma Naive Bayes dalam memprediksi siswa-siswi SMK Az-Zahra masuk pasar kerja, memprediksi siswa yang lulus kriteria masuk pasar kerja sebanyak 57 orang dan yang tidak lulus kriteria masuk pasar kerja sebanyak 17 orang. Hal ini berbeda dengan data testing

siswa dimana siswa yang lulus kriteria masuk pasar kerja sebanyak 46 orang dan yang tidak lulus kriteria masuk pasar kerja sebanyak 29 orang. Hasil tersebut sama dengan perhitungan manual excel. Berikut Gambar Hasil Prediksi :

ID	Nama	Nilai	Status	Prediksi
1	Lulus	0.900	0.000	Lulus
2	Lulus	0.900	0.000	Lulus
3	Lulus	0.900	0.000	Lulus
4	Lulus	0.900	0.000	Lulus
5	Lulus	0.900	0.000	Lulus
6	Lulus	0.900	0.000	Lulus
7	Lulus	0.900	0.000	Lulus
8	Lulus	0.900	0.000	Lulus
9	Lulus	0.900	0.000	Lulus
10	Lulus	0.900	0.000	Lulus
11	Lulus	0.900	0.000	Lulus
12	Lulus	0.900	0.000	Lulus
13	Lulus	0.900	0.000	Lulus
14	Tidak Lulus	0.000	0.000	Tidak Lulus
15	Tidak Lulus	0.000	0.000	Tidak Lulus
16	Lulus	0.900	0.000	Lulus
17	Lulus	0.900	0.000	Lulus
18	Tidak Lulus	0.000	0.000	Tidak Lulus
19	Lulus	0.900	0.000	Lulus
20	Lulus	0.900	0.000	Lulus

Gambar 3 Hasil ExampleSet Data Testing
3.3 Evaluasi dan Validasi

Evaluasi dan Validasi hasil pada penelitian ini dilakukan menggunakan *Confusion Matrix* sebagai *performance* untuk mencari nilai *accuracy*, *Precision* dan *recall*. Selain itu, dilakukan pengujian menggunakan kurva ROC (*Receiver Operating Characteristic*) yang menghasilkan nilai AUC (*Area Under Curve*). Adapun penjelasan lebih lengkap terkait evaluasi dan validasi akan diuraikan sebagai berikut :

1. *Confusion Matrix*

Diketahui :

Table 1 Performance Vektor

PREDIKSI	True Lulus	True Tidak Lulus
Pred. Lulus	43	14
Pred. Tidak Lulus	3	15
ACCURACY	77,33%	

Berdasarkan tabel diatas dapat disimpulkan bahwa dari dataset sebanyak 75 data siswa dengan kelas Lulus sebanyak 46 dan kelas Tidak Lulus sebanyak 29 data siswa. Didapati hasil prediksi benar "Lulus" sebanyak 43 dan prediksi kelas benar "Lulus" sebanyak 43 dan prediksi kelas "Tidak Lulus" sebanyak 14. Kemudian prediksi kelas "Tidak Lulus" sebanyak 29 dengan prediksi kelas "Lulus" sebanyak 3 dan prediksi kelas "Tidak Lulus" sebanyak 15 sehingga dihasilkan :

a. *Accuracy*

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \times 100\%$$

$$Accuracy = \frac{43+15}{43+14+3+15} \times 100\%$$

$$Accuracy = \frac{58}{75} \times 100\%$$

$$Accuracy = 77,33\%$$

Berdasarkan hasil *Accuracy* dari perhitungan diatas menghasilkan nilai akurasi dataset 77,33%. Berikut Gambar Hasil Akurasi pada RapidMiner 5.3 :

Classification Performance	True Lulus	True Tidak Lulus	Class Precision
pred Lulus	43	14	75.4%
pred Tidak Lulus	3	15	83.3%
class recall	83.0%	81.7%	

Gambar 4 Akurasi Metode Naive Bayes

b. *Precision*

1. Lulus

$$Precision = \frac{TP}{TP+FP} \times 100\%$$

$$Precision = \frac{43}{43+14} \times 100\%$$

$$Precision = \frac{43}{57} \times 100\%$$

$$Precision = 75,44\%$$

2. Tidak Lulus

$$Precision = \frac{TN}{TP+FP} \times 100\%$$

$$Precision = \frac{15}{15+3} \times 100\%$$

$$Precision = \frac{15}{18} \times 100\%$$

$$Precision = 83,33\%$$

Berdasarkan pengujian diatas dapat disimpulkan bahwa presentase *precision* yang dihasilkan adalah kelas "Lulus" sebesar 75,44% dan kelas "Tidak Lulus" sebesar 83,33%. Berikut hasil *performance precision* pada *software RapidMiner* :

Classification Performance	True Lulus	True Tidak Lulus	Class Precision
pred Lulus	43	14	75.4%
pred Tidak Lulus	3	15	83.3%
class recall	83.0%	81.7%	

Gambar 5 Hasil Performance Precision

c. *Recall*

1. Lulus

$$Recall = \frac{TP}{TP+FN} \times 100\%$$

$$Recall = \frac{43}{43+3} \times 100\%$$

$$Recall = \frac{43}{48} \times 100\%$$

$$Recall = 93,48\%$$

2. Tidak Lulus

$$Recall = \frac{TP}{TP+FN} \times 100\%$$

$$Recall = \frac{15}{15+14} \times 100\%$$

$$Recall = \frac{15}{29} \times 100\%$$

$$Recall = 51,72\%$$

Berdasarkan pengujian diatas dapat disimpulkan bahwa presentase *Recall* yang dihasilkan adalah kelas “Lulus” sebesar 93,48% dan kelas “Tidak Lulus” sebesar 51,72%. Berikut hasil *performance precision* pada *software RapidMiner* :

Class	True	False	Class Precision
class Lulus	43	14	75,44%
class Tidak Lulus	3	15	83,33%
class recall	93,48%	51,72%	

Gambar 6 Hasil Performance Recall

1. ROC (Receiver Operating Characteristic)

Berdasarkan pengujian yang telah dilakukan menggunakan *software RapidMiner*, dihasilkan nilai AUC (*Area Under Curve*) dari ROC (*Receiver Operating Characteristic*) sebesar 0,888 sehingga dapat disimpulkan bahwa algoritma mempunyai kinerja yang baik. Berikut adalah hasil *performance AUC* pada *software RapidMiner* :



Gambar 7 Hasil Performance AUC

KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan mengenai penentuan masuk pasar

kerja menggunakan metode Naïve Bayes pada SMK AZ-ZAHRA Mlonggo, maka dapat diambil kesimpulan bahwa Algoritma *Naïve Bayes* yang digunakan menghasilkan proses penyeleksian yang cepat dan algoritma yang mudah dipahami dengan tingkat akurasi yang bagus. Pada prediksi siswa-siswi smk az-zahra mlonggo masuk pasar kerja dapat dilakukan dengan mempertimbangkan dataset siswa di masa lalu menggunakan algoritma yaitu Naïve Bayes. Algoritma Naïve Bayes dapat diterapkan untuk membantu analisis siswa-siswi masuk pasar kerja, dimana dalam penelitian ini data yang digunakan tahun 2018-2020 dengan total 247. Dataset tersebut dibagi menjadi 2 bagian, yakni 152 data training dan 75 data testing, dimana memprediksi siswa yang lulus kriteria masuk pasar kerja sebanyak 57 orang dan yang tidak lulus kriteria masuk pasar kerja sebanyak 18 orang. Hal ini berbeda dengan data testing siswa dimana yang lulus kriteria masuk pasar kerja sebanyak 46 orang dan yang tidak lulus kriteria masuk pasar kerja sebanyak 29 orang dan akurasi yang diperoleh mendapatkan nilai sebesar 77,33% sedangkan untuk precision mendapatkan nilai sebesar 75,44% class Lulus dan 83,33% class Tidak Lulus dan untuk recall mendapatkan nilai sebesar 93,48% class Lulus dan 51,72% class Tidak Lulus. Nilai akurasi tersebut sama besarnya saat dihitung manual maupun menggunakan Aplikasi RapidMiner 5.3.

DAFTAR PUSTAKA

Kurniawan, Y. I. (2018). Perbandingan Algoritma Naive Bayes dan C.45 dalam Klasifikasi Data Mining. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 5(4), 455. <https://doi.org/10.25126/jtiik.201854803>

Lumapelumey, I., Studi, P., Ekonomi, I., Universitas, P., & Urimesing, K. (2019). *Analisis Faktor - Faktor Yang Mempengaruhi Lama Mencari. XIII(1)*.

Orpa, E. P. K., Ripanti, E. F., & Tursina, T. (2019). Model Prediksi Awal Masa Studi Mahasiswa Menggunakan Algoritma Decision Tree C4.5. *Jurnal Sistem Dan Teknologi Informasi (JUSTIN)*, 7(4), 272. <https://doi.org/10.26418/justin.v7i4.33163>

Rajagukguk, Z. (2010). Pasar Kerja Fleksibel Versus Perlindungan Pekerja di Indonesia. *Jurnal Kependudukan*, 5, No. 2(2), 1–28.

Rosandy, T. (2019). PERBANDINGAN METODE NAIVE BAYES CLASSIFIER DENGAN METODE DECISION TREE (C4.5) UNTUK MENGANALISA KELANCARAN PEMBIAYAAN (Study Kasus : KSPPS / BMT AL-FADHILA. *Jurnal Teknologi Informasi Magister Darmajaya*, 2(01), 52–62.

- Srisulistiwati, D. B., Khaerudin, M., & ... (2021). Sistem Informasi Prediksi Penjualan Alat Tulis Kantor Dengan Metode Fp-Growth (Studi Kasus Toko Koperasi Sekolah Bina Mulia). *JSI (Jurnal ...)*, 243–256.
- Suwarno, A., Ferawati, N., Sari, P. A., Studi, P., Informatika, T., Teknik, F., Pelita, U., Orientasi, O., & Barang, S. (2021). *Jurnal Teknologi Pelita Bangsa*. 12(4), 33–40.
- Yuli Mardi. (2019). Data Mining: Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . *Jurnal Edik Informatika. Jurnal Edik Informatika*, 2.
- Zuhdiyaty, N., & Kaluge, D. (2018). Analisis Faktor - Faktor Yang Mempengaruhi Kemiskinan Di Indonesia Selama Lima Tahun Terakhir. *Jurnal Ilmiah Bisnis Dan Ekonomi Asia*, 11(2), 27–31.
<https://doi.org/10.32812/jibeka.v11i2.42>