



PENERAPAN DATA MINING PREDIKSI NILAI UJIAN NASIONAL (UN) SISWA SMP MENGGUNAKAN METODE NAIVE BAYES

Miftahul Jannah

AMIK Mitra Gama

Jalan Kayangan No.99 Kode Pos 28784

e-mail : miiftahuljannah32@gmail.com

ABSTRAK

Ujian Nasional (UN) dilakukan setiap tahun di semua tingkat pendidikan salah satunya SMP N 8 Mandau. Data nilai Ujian Nasional (UN) siswa di SMP N 8 Mandau setiap tahunnya bertambah dan disimpan dalam bentuk Soft file dan Hard file. Dari beberapa tahun belakangan nilai UN siswa di SMP N 8 tersebut tidak stabil. Nilai UN sangat berpengaruh terhadap kualitas sekolah. Tinggi rendahnya nilai UN siswa sangat berpengaruh pada kualitas sekolah. Untuk mengatasi hal ini diperlukan sebuah teknik data mining dengan metode naive bayes dalam memprediksi nilai UN siswa berdasarkan data siswa tahun 2016 dan 2017. Dari 10 data uji mendapatkan hasil prediksi 7 orang yang sesuai dengan data pada tahun 2017. Sehingga memiliki tingkat akurasi sebesar 70%. Sehingga metode ini dapat diterapkan dalam memprediksi nilai UN siswa dimasa mendatang. Dan pihak sekolah dapat terbantu dalam mensiasati dalam mencapai nilai UN yang tinggi.

Kata kunci : *Data Mining, Naive Bayes*

ABSTRACT

The National Examination (UN) is carried out annually at all levels of education, one of which is SMP N 8 Mandau. The data on the National Examination (UN) scores of students at SMP N 8 Mandau each year are added and stored in the form of soft files and hard files. From the last few years the UN scores of students at SMP N 8 are not stable. National examination scores greatly affect the quality of the school. The high and low students' national exam scores greatly affect the quality of schools. To overcome this, a data mining technique with the Naive Bayes method is needed in predicting student National Examination scores based on student data in 2016 and 2017. From 10 test data, the prediction results of 7 people match the data in 2017. So that it has an accuracy rate of 70%. So that this method can be applied in predicting student UN scores in the future. And the school can be helped in achieving UN high scores.

Keywords: *Soft file, Hard file, Data Mining, Naive Bayes*

Kata kunci : *Data Mining, Naive Bayes*

1. PENDAHULUAN

Untuk mempermudah kehidupan manusia saat ini maka komputer merupakan teknologi informasi yang banyak digunakan dalam memperoleh informasi. Komputer juga digunakan dalam mengolah data dan menyimpan data. Data disimpan dalam jangka panjang dalam komputer tersebut dapat

menyebabkan tumpukan data yang tidak bermanfaat karena data tersebut selalu bertambah.

Pertumbuhan yang sangat pesat dari akumulasi data telah menciptakan kondisi kaya akan data tapi minim informasi (Haryati, Sudarsono, & Suryana, 2016). proses penggalan informasi tersembunyi dalam suatu



basis data yang besar dinamakan Data Mining dan Knowledge Discovery in Database (KDD). Knowledge discovery in Database (KDD) didefinisikan sebagai ekstraksi informasi potensial, implisit dan tidak dikenal dari sekumpulan data. Proses knowledge discovery melibatkan hasil dari proses Data Mining (proses mengekstrak kecenderungan pola suatu data), kemudian mengubah hasilnya secara akurat menjadi informasi yang mudah dipahami. KDD sendiri diartikan sebagai keseluruhan proses non-trivial untuk mencari dan mengidentifikasi pola (pattern) dalam data, dimana pola yang ditemukan bersifat sah, baru, dapat bermanfaat dan dapat dimengerti. (Fajrin & Maulana, 2018)

Untuk mengatasi minimnya informasi dari data yang bertumpuk tersebut maka ditemukan sebuah ilmu baru yang dinamakan data mining. Data mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran computer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (knowledge) secara otomatis. Definisi lain diantaranya adalah pembelajaran berbasis induksi (*induction-based learning*) adalah proses pembentukan definisi-definisi konsep umum yang dilakukan dengan cara mengobservasi contoh-contoh spesifik dari konsep-konsep yang akan dipelajari (Juna, 2016). Data mining dan knowledge discovery in database (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalan informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD adalah data mining (F. Yunita, 2018)

Data Mining sendiri memiliki teknik di antaranya prediksi. Teknik prediksi memiliki beberapa metode salah satunya adalah algoritma Naive Bayes. Algoritma Naive Bayes digunakan untuk memprediksi peluang yang terjadi di masa depan berdasarkan pengalaman di masa sebelumnya. (Nuraeni, 2017).

Penelitian terkait penerapan Data Mining dalam memprediksi menggunakan metode Naive Bayes sebelumnya dilakukan oleh winda dkk dalam memprediksi hasil kelayakan tenaga kerja Indonesia dengan data uji sebanyak 17 orang dengan menggunakan dua kelas. klasifikasi dengan kelas Layak sebanyak 10 orang dan kelas Tidak Layak sebanyak 7 orang.

Pengujian data pada RapiMiner 5.3 dengan menggunakan metode Naive Bayes berhasil

menampilkan dua kelas dari hasil klasifikasi dengan persentase keakuratan sebesar 90,00%. Dengan class precision pada prediksi Layak memiliki nilai 90,91%, sedangkan pada prediksi Tidak Layak memiliki nilai 100 %. Class recall pada true Layak memiliki nilai 100%, sedangkan pada true Tidak Layak memiliki nilai 85,71%. (Purba, Suhendro, Damanik, & Syahputra, 2019). Selanjutnya dilakukan penelitian oleh yunita dkk mencoba menggabungkan 2 metode Naive Bayes dengan decision tree. Dari hasil uji dengan data asli dalam memprediksi prestasi siswa tidak seimbang dalam output tiap kelompok. Dilakukan uji data dengan re-sample maka didapatkan prediksi yang dihasilkan lebih akurat (D. Yunita, Amalia, & Rosyani, 2018). Jika dilihat dari beberapa studi – studi diatas yang merupakan penelitian terdahulu yang terkait, maka metode Naive Bayes dapat digunakan untuk memprediksi kejadian berdasarkan probabilitas dengan memberikan asumsi ketidak tergantungan yang sangat kuat. Algoritma Naive Bayes juga memiliki tingkat akurasi yang tinggi dalam menangani permasalahan prediksi.

Ujian Nasional (UN) dilakukan setiap tahun di semua tingkat pendidikan salah satunya SMP N 8 Mandau. Data nilai Ujian Nasional (UN) siswa di SMP N 8 Mandau setiap tahunnya bertambah dan disimpan dalam bentuk soft file dan hard file. Data yang disimpan setiap tahunnya tersebut dibiarkan bertumpuk tanpa menghasilkan suatu manfaat yang baru. Dari beberapa tahun belakangan nilai UN siswa di SMP N 8 tersebut tidak stabil. Tinggi rendahnya nilai UN siswa sangat berpengaruh pada kualitas sekolah.

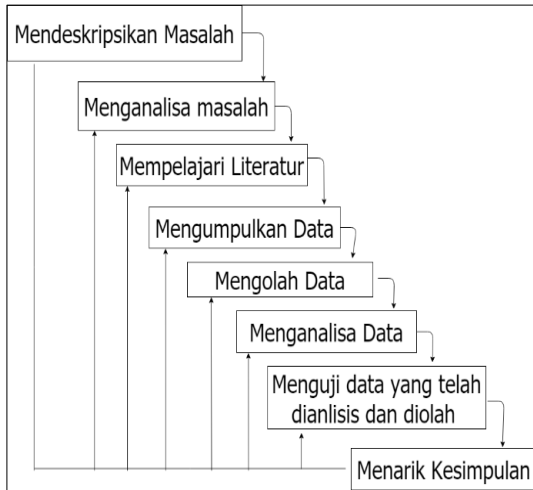
Untuk mengatasi hal ini diperlukan sebuah teknik di mana data yang ditumpuk banyak dapat menghasilkan suatu informasi baru yang disebut dengan Data Mining dengan prediksi menggunakan Algoritma Naive Bayes. Hasil Data Mining menggunakan metode Naive Bayes adalah berupa prediksi nilai rata-rata bidang studi UN yang dapat membantu pihak sekolah dalam menentukan strategi agar nilai UN mengalami peningkatan.

2. METODOLOGI PENELITIAN

Metodologi penelitian diperlukan sebagai kerangka dan panduan pada saat penelitian, sehingga penelitian yang dilakukan menjadi lebih terarah, teratur dan sistematis. Dalam menyelesaikan penelitian ini dibutuhkan kerangka penelitian agar penelitian dapat

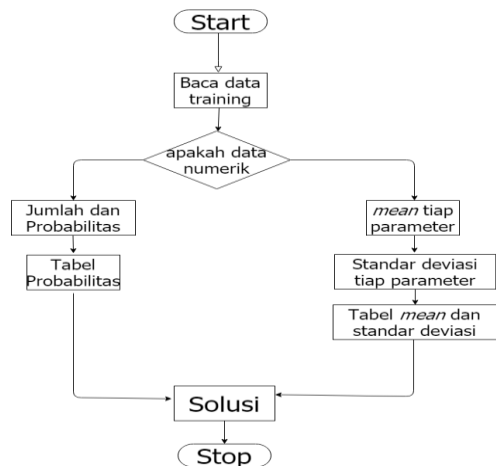


diselesaikan secara teratur, terarah dan sistematis maka diperlukan tahap-tahap. Adapun kerangka kerja penelitian dapat digambarkan sebagai berikut:



Gambar 1. Tahapan Penelitian

Langkah awal yang dilakukan dengan mendeskripsikan permasalahan yang ada dan dilakukan analisis. Selanjutnya mempelajari literatur berdasarkan penelitian terdahulu sehingga dapat dijadikan dasar dalam penelitian, lalu mengumpulkan data siswa di SMP N 8 Mandau yang akan mengikuti UN pada tahun 2017 dan data siswa yang telah melakukan UN di tahun 2016. Setelah data terkumpul dilanjutkan mengolah data dengan beberapa tahap di antaranya: data cleaning, data integrasi, data seleksi, data transformasi dan dilanjutkan dengan proses data mining dengan metode *Naive Bayes* seperti gambar dibawah ini



Gambar 2. Alur metode *naive bayes*

3. HASIL DAN PEMBAHASAN

Data Mining yang diterapkan dalam penelitian ini adalah berbentuk prediksi menggunakan metode Naive Bayes dalam melakukan prediksi terhadap nilai UN siswa di SMP N 8 Mandau. Prediksi yang dilakukan dalam penelitian ini menggunakan 3 atribut Input sebagai kriteria dalam memprediksi nilai UN. Atribut yang digunakan penelitian adalah jenis kelamin siswa, rata-rata nilai rapor siswa dari semester 1-5 dimana Bidang Studi yang diambil yang di UN kan B. Indonesia, B. Inggris, Matematika, IPA dan input yang terakhir adalah nilai rata-rata UAS siswa bidang studi yang di UN kan. Atribut output yang terakhir adalah nilai kelulusan UN. Ketiga atribut tersebut didapatkan dari data nilai siswa yang sudah melaksanakan ujian UN pada tahun 2015-2016 sebagai data training dan sebagai data testing menggunakan data siswa melaksanakan UN pada tahun 2016-2017.

Adapun hasil Output pada penelitian ini terbagi ke dalam 3 class, yaitu prediksi nilai Kurang untuk yang mendapatkan rata-rata nilai UN di bawah 55, Cukup untuk yang mendapatkan rata-rata nilai UN lebih 55 kurang 70 dan Baik untuk yang mendapatkan rata-rata nilai UN lebih dari 70 dan sama dengan 85. Agar tercapai tujuan dari penelitian ini, maka dilakukan pengumpulan data training yang diambil secara random adalah berjumlah 40 orang data siswa yang telah melaksanakan UN di tahun 2015-2016 pada tabel 1.

Tabel 1. Data Training

Nama Siswa	Jenis Kelamin	Nilai Rata-rata Rapo r	Nilai Rata-rata UAS	Status Nilai
Afni Yogeta	P	79	86	Cukup
Afni Yuli Shantia	P	79	86	Kurang
-	-	-	-	-
-	-	-	-	-
Muhamad Afif	L	82	86	Baik
Muhamad Fadhil Fatihil	L	83	89	Cukup



1. Baca Data Training
Menentukan data numerik atau tidak
2. Untuk data numerik menghitung nilai mean dan standar deviasi. Dan data yang tidak numerik menghitung probabilitasetiap kategori.
3. Persamaan yang digunakan untuk menghitung nilai rata-rata (mean) dapat digunakan rumus sebagai berikut :

$$\mu = \frac{x_1+x_2+x_3+\dots+x_n}{n} \dots\dots\dots(1)$$

Tabel 2. Perhitungan Nilai Mean Variabel Rata-Rata Nilai Rapor Untuk Status Nilai Kurang

No	Nama Siswa	Kurang
1	Afni Yuli Shantia	79
2	Alvin Dwi Pratama	81
3	Apri Imam Padzli	80
4	Bagas Alfajri	80
5	Dea Putri Octami	80
6	Elya Ramayanti Napitupulu	81
7	Elza Fitri Rahayu	83
8	Fauzi Syaputra	79
9	Ferdi Juliandra	79
10	Hanafi Pramuja	82
11	Haris Suwandi	82
12	Heru Pratama	79
13	Ilyas Syaher	81
14	M.Agustian Renaldi	82
15	Muhammad Agustian	82

Perhitungan nilai mean dengan variabel nilai rapor kategori status nilai Kurang :

$$\mu = (79+81+80+80+80+81+83+79+79+82+82+79+81+82+82)/15$$

$$\mu = 1210/15$$

$$\mu = 80,6666666666667$$

Setelah mendapat nilai mean dari masing-masing kategori status nilai, langkah selanjutnya melakukan perhitungan standart deviasi dari masing-masing kategori status nilai. Dengan rumus perhitungan sebagai berikut

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n-1}} \dots\dots\dots(2)$$

Perhitungan standar deviasi dari kategori status nilai Kurang dengan variabel nilai rata-rata rapor:

$$\sigma = \sqrt{(25,33/(15-1))}$$

$$\sigma = \sqrt{1,8095}$$

$$\sigma = 1,3451854182691$$

Tabel 3. Mean dan standar deviasi variabel nilai rapor

	Kurang	Cukup	Baik
Mean	80,666 666666 6667	82,46 66666 66666 7	81,3
Standar t deviasi	1,3451 854182 691	3,270 35748 73762 8	1,41813649241 2181813649241 218

Tabel 4. Mean dan standar deviasi variabel nilai UAS

	Kurang	Cukup	Baik
Mean	86,86666 6666666 7	86,8	86,8
Standart deviasi	1,060098 8273786 2	1,3201 731488 1691	1,549193338 4829749193 33848297

Langkah selanjutnya adalah menghitung probabilitas untuk variabel bersifat diskret.

Tabel 5. Probabilitas Jenis Kelamin

Jenis Kela min	Kategori Status Nilai			Probabilitas Status Nilai		
	Kura ng	Cuk up	Ba ik	Kura ng	Cuk up	Ba ik
L	11	9	7	11/15	9/15	7/15
P	4	6	3	4/15	6/15	3/15
Jumla h	15	15	10	1	1	1

Tabel 6. Probabilitas Status Nilai

Kategori Status nilai			Probilitas status nilai		
Renda h	Sedan g	Ting gi	Renda h	Sedan g	Ting gi
15	15	10	15/40	15/40	10/40



3. Menghitung probabilitas diberikan input tertentu

Nama Siswa	Jenis Kelamin	Nilai rata-rata Rapor	Nilai rata-rata UAS	Status Nilai
Akbar Hakim	L	77	80	?

Dengan persamaan rumus sebagai berikut :

$$P(X_i=x_i|Y=Y_j) = \frac{1}{\sqrt{2\pi}\sigma} \exp - \frac{(X_i-\mu_{ij})^2}{2(\sigma_{ij})^2}$$

Dengan data input nilai rata-rata rapor = 77

f (rata-rata nilai rapor=77| status nilai=Kurang)

$$= \frac{1}{\sqrt{2\pi} (1,3451854182691)} \exp - \frac{(77-80,6666666666667)^2}{2(1,3451854182691)^2}$$

$$= 0,0072235506$$

f (rata-rata nilai rapor=77| status nilai=Cukup)

$$= \frac{1}{\sqrt{2\pi} (3,27035748737628)} \exp - \frac{(77-82,4666666666667)^2}{2(3,27035748737628)^2}$$

$$= 0,0301690272$$

f (rata-rata nilai rapor=77| status nilai=Baik)

=

$$\frac{1}{\sqrt{2\pi} (1,418136492412181813649241218)} \exp - \frac{(77-81,3)^2}{2(1,418136492412181813649241218)^2} \frac{0,000000000000058}{0,000000000000058 + 0,000000003555 + 0,000000008372}$$

$$= 0,0028362417$$

Untuk Nilai rata-rata UAS dengan data input 80 dapat dilakukan perhitungan sebagai berikut :

f (rata-rata UAS =80| status nilai=Kurang)

$$= \frac{1}{\sqrt{2\pi} (1,06009882737862)} \exp - \frac{(80-86,8666666666667)^2}{2(1,06009882737862)^2}$$

$$= 0,00000000029$$

f (rata-rata UAS=80| status nilai=Cukup)

$$= \frac{1}{\sqrt{2\pi} (1,32017314881691)} \exp - \frac{(80-86,8)^2}{2(1,32017314881691)^2}$$

$$= 0,0000005237$$

f (rata-rata UAS 80| status nilai=Baik)

=

$$\frac{1}{\sqrt{2\pi} (1,549193338482974919333848297)} \exp - \frac{(80-86,3)^2}{2(1,549193338482974919333848297)^2}$$

$$= 0,0000168683$$

Setelah mendapatkan nilai gaussian maka langkah selanjutnya melakukan perhitungan likelihood dari masing-masing status nilai untuk mendapatkan nilai probabilitas.

Likelihood Kurang =

$$0,0072235506 \times 0,00000000029 \times 11/15 \times 15/40 = 0,0000000000058$$

Likelihood Cukup =

$$0,0301690272 \times 0,0000005237 \times 9/15 \times 15/40 = 0,000000003555$$

Likelihood Baik =

$$0,0028362417 \times 0,0000168683 \times 7/10 \times 10/40 = 0,000000008372$$

Setelah mendapatkan nilai likelihood maka nilai probabilitas dapat dihitung dengan melakukan normalisasi terhadap likelihood tersebut sehingga nilai yang diperoleh mendekati 1.

Probabilitas Kurang

$$\frac{0,000000000000058}{0,000000000000058 + 0,000000003555 + 0,000000008372} = 0,000048$$

Probabilitas Cukup =

$$\frac{0,000000003555}{0,000000000000058 + 0,000000003555 + 0,000000008372} = 0,298027$$

Probabilitas Baik =

$$\frac{0,000000008372}{0,000000000000058 + 0,000000003555 + 0,000000008372} = 0,701925$$

Dari hasil tersebut, terlihat bahwa nilai probabilitas tertinggi ada pada status nilai Baik. Sehingga dapat disimpulkan bahwa Akbar Hakim memiliki status nilai Baik.

**4. KESIMPULAN**

238–249.

Kesimpulan Dari hasil pengujian yang telah dilakukan oleh peneliti dapat disimpulkan bahwa

1. Prediksi hasil Ujian Nasional (UN) tahun 2017 dapat diselesaikan dengan teknik data mining menggunakan metode Naive Bayes, sehingga dapat membantu pihak sekolah mengetahui hasil Ujian Nasional (UN) yang diperoleh siswa dan meningkatkan strategi menghadapi Ujian Nasional ditahun yang akan datang.
2. Dari hasil percobaan prediksi hasil nilai Ujian Nasional (UN) tahun 2017 menggunakan Rapidminer metode Naive Bayes diperoleh akurasi sebesar 70,00% dan error rate sebesar 10,00%

5. REFERENSI

- Fajrin, A. A., & Maulana, A. (2018). Penerapan Data Mining Untuk Analisis Pola Pembelian Konsumen Dengan Algoritma Fp- Growth Pada Data Transaksi Penjualan. *Kumpulan Jurnal Ilmu Komputer (Klik)*, 5(1), 27–36.
- Haryati, S., Sudarsono, A., & Suryana, E. (2016). *Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4 . 5*. 11(2), 130–138.
- Juna, E. (2016). Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4 . 5 Stmik Royal , Ksieran. *Jurteks*, 2(2), 9–13.
- Nuraeni, N. (2017). Penentuan Kelayakan Kredit Dengan Algoritma Naïve Bayes Classifier : Studi Kasus Bank Mayapada Mitra Usaha Cabang Pgc. *Jurnal Teknik Komputer Amik Bsi, Iii*(1), 9–15.
- Purba, W. H., Suhendro, D., Damanik, I. S., & Syahputra, I. (2019). *Penerapan Algoritma Naive Bayes Pada Penentuan Kelayakan Calon Tenaga Kerja Indonesia*. (September), 771–779.
- Yunita, D., Amalia, R., & Rosyani, P. (2018). *Analisa Prestasi Siswa Berdasarkan Kedisiplinan , Nilai Hasil Belajar , Sosial Ekonomi Dan Aktivitas Organisasi Menggunakan Algoritma Naïve Bayes*. 3(4), 2–5.
- Yunita, F. (2018). Yunita, Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru (Studi Kasus: Universitas Islam Indragiri) 1 238. *Sistemasi*, 7(September),