



KLASIFIKASI RATING SEKOLAH MENENGAH ATAS BERDASARKAN JUMLAH SISWA SEKOLAH DI PROVINSI DKI JAKARTA MENGGUNAKAN METODE ALGORITMA NAIVE BAYES

Antika Zahrotul Kamalia¹, Hemdani Rahendra Herlianto², Zaenur Rozikin³, Andriani⁴

^{1,2,3}Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa

⁴Teknik Industri, Fakultas Teknik, Universitas Pelita Bangsa

antika.kamalia@pelitabangsa.ac.id¹,hemdani.rahendra@pelitabangsa.ac.id², zaenurrozikin@pelitabangsa.ac.id³

Abstract

Education has a crucial role in the development of a nation, and evaluating the quality of education is very important. One of the main indicators in assessing the quality of an educational institution is the school rating. In DKI Jakarta Province, a number of high schools provide a variety of student numbers and quality of education. This research aims to classify the ranking of senior high schools in DKI Jakarta Province based on the number of school students using the Naive Bayes algorithm method. This research method collects data on the number of school students and school ratings from various trusted sources. Then, the data is processed and analyzed using the Naive Bayes algorithm to classify school ratings based on the number of students. The analysis results show that the Naive Bayes algorithm can provide classification school ratings with a satisfactory level of accuracy.

Keywords: Classification, School Ratings, Total Students, Naive Bayes.

Abstrak

Pendidikan memiliki peran krusial dalam pembangunan suatu bangsa, dan evaluasi kualitas pendidikan merupakan hal yang sangat penting. Salah satu indikator utama dalam menilai kualitas suatu lembaga pendidikan adalah rating atau peringkat sekolah. Di Provinsi DKI Jakarta, sejumlah sekolah menengah atas menyajikan beragam jumlah siswa dan kualitas pendidikan. Penelitian ini bertujuan untuk memklasifikasi rating sekolah menengah atas di Provinsi DKI Jakarta berdasarkan jumlah siswa sekolah menggunakan metode algoritma Naive Bayes.

Metode penelitian ini mengumpulkan data jumlah siswa sekolah dan rating sekolah dari berbagai sumber terpercaya. Kemudian, data tersebut diproses dan dianalisis menggunakan algoritma Naive Bayes untuk memklasifikasi rating sekolah berdasarkan jumlah siswa. Hasil analisis menunjukkan bahwa algoritma Naive Bayes dapat memberikan klasifikasi rating sekolah dengan tingkat akurasi yang memuaskan.

Kata kunci: Klasifikasi, Rating Sekolah, Jumlah Siswa, Naive Bayes.

1. Pendahuluan

Pendidikan memegang peranan penting dalam pembangunan suatu bangsa. Evaluasi terhadap kualitas pendidikan menjadi hal yang krusial dalam menentukan arah dan keberhasilan sistem pendidikan suatu negara. Salah satu aspek penting dalam mengevaluasi kualitas pendidikan adalah rating atau peringkat sekolah. Rating sekolah memberikan gambaran tentang seberapa efektif lembaga pendidikan tersebut dalam memberikan pembelajaran kepada siswa.

Provinsi DKI Jakarta, sebagai salah satu pusat pendidikan terkemuka di Indonesia, memiliki sejumlah sekolah menengah atas yang beragam baik dalam hal jumlah siswa maupun kualitas pendidikan yang ditawarkan. Namun, untuk dapat menilai kualitas sekolah secara objektif, diperlukan metode analisis yang akurat dan efektif.

Dalam rangka meningkatkan rating sekolah di DKI Jakarta, diperlukan upaya yang holistik dan

terkoordinasi antara pemerintah, lembaga pendidikan, masyarakat, dan pihak terkait lainnya untuk mengatasi berbagai permasalahan yang ada. Hal ini termasuk peningkatan aksesibilitas, peningkatan kualitas fasilitas dan sumber daya, peningkatan kualitas tenaga pengajar, serta peningkatan manajemen dan administrasi sekolah secara keseluruhan.

Klasifikasi merupakan sebuah penilaian secara sistematis yang memperkirakan masa depan berdasarkan pengetahuan sebelumnya dan saat ini, yang berguna untuk mengurangi kesalahan dan ketidaksesuaian antara apa yang diantisipasi dan apa yang sebenarnya terjadi [1]. Dalam penelitian Khoirunnisa pada tahun 2021 digunakan algoritma klasifikasi decision tree, naive bayes dan k-nearest neighbor (KNN) dalam memklasifikasi siswa sekolah menengah yang akan mendaftar ke perguruan tinggi berdasarkan jenis kelamin, usia, pendapatan, dan asuransi [2]. Dalam penelitiannya menunjukkan bahwa ketiga metode tersebut memiliki tingkat akurasi 96,60%, 92,40%, dan 94,96%. Penelitian lainnya yakni penelitian (Armansyah & Ramli, 2022) yang menggunakan metode naive bayes, menunjukkan bahwa metode model klasifikasi berkinerja sangat baik dan menampilkan akurasi hingga 96,50%. Nilai tersebut disebabkan oleh konfigurasi atribut data training yang relevan dengan rancangan model.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan prediktif untuk memklasifikasi rating sekolah menengah atas di Provinsi DKI Jakarta berdasarkan jumlah siswa sekolah. Metode yang digunakan adalah algoritma Naive Bayes, yang merupakan metode klasifikasi dari *Data Mining* yang efektif dalam memodelkan probabilitas berdasarkan data yang ada.

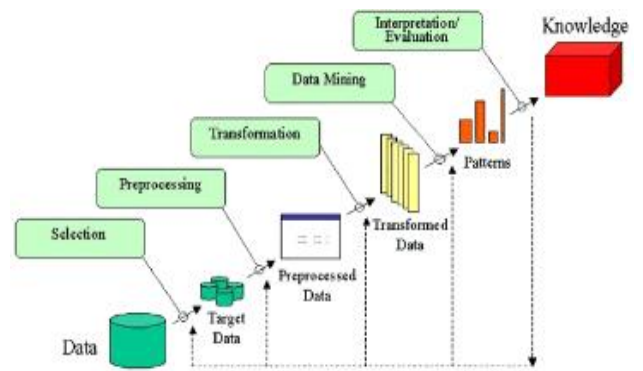
2.1. Data Mining

Data Mining disebut juga Knowledge Discovery in Database (KDD) didefinisikan sebagai ekstraksi informasi potensial. Dalam Data Mining terdapat dua pendekatan metode pelatihan implisit dan tidak dikenal dari sekumpulan data. Data Mining biasanya mengolah data dari database dengan ukuran yang besar. Dari data tersebut dilakukan pencarian pola atau tren sesuai dengan tujuan dari penerapan Data Mining tersebut. Hasil dari pengolahan Data Mining tersebut selanjutnya dapat digunakan untuk pengambilan keputusan maupun analisis yang dibutuhkan [3]. Data Mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. Data Mining merupakan serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual [4]. Proses Knowledge Discovery in Database

melibatkan hasil proses Data Mining (proses pengekstrak kecenderungan suatu pola data), kemudian mengubah hasilnya secara akurat menjadi informasi yang mudah dipahami, yaitu :

- a. *UnSupervised learning* metode ini diterapkan tanpa adanya latihan (training) dan tanpa ada guru (teacher). Guru di sini adalah label dari data.
- b. *Supervised learning* yaitu metode belajar dengan adanya latihan dan pelatih. Dalam pendekatan ini, untuk menemukan fungsi keputusan, fungsi pemisah atau fungsi regresi, digunakan beberapa contoh data yang mempunyai output atau label selama proses training.

Metode pengembangan data mining yang digunakan untuk menganalisis data dalam penerapan data mining ini menggunakan proses tahapan knowledge discovery in databases (KDD) yang terdiri dari Data, Data Cleaning, Data transformation, Data mining, Pattern evolution, knowledge.



Gambar 1. Tahapan KDD

Berikut merupakan hal – hal yang perlu di lakukan dalam penelitian berdasarkan tahapan knowledge discovery in databases :

1. Data

Data merupakan sekumpulan data operasional yang diperlu sebelum dilakukan sebelum tahap penggalian informasi dalam Knowledge Discovery Database (KDD) dimulai.

2. Data Cleaning

Proses data cleaning merupakan proses Pembersihan data yang bertujuan untuk menghilangkan data yang tidak memiliki nilai (null), data yang salah input, data yang tidak relevan, duplikat data dan data yang tidak konsisten

3. Data transformation

Data transformation dilakukan dengan memberikan inialisasi terhadap data yang memiliki nilai nominal menjadi bernilai numerik.

4. Data Mining.

Pada fase ini yang dilakukan adalah menerapkan algoritma atau metode pencarian pengetahuan. Ini adalah langkah penting di mana teknik kecerdasan diterapkan untuk mengekstrak pola informasi yang berpotensi berguna dari data yang dipilih.

5. evaluation

Pada tahap evaluasi, akan diketahui apakah hasil daripada tahap data mining dapat menjawab tujuan yang telah ditetapkan. Untuk itu akan dilakukan profilisasi pada setiap cluster yang telah terbentuk, untuk diketahui karakteristik pada kelompok tersebut. Disamping itu untuk diketahui kesesuaian dengan jalur perminatan akan dilakukan analisis lebih lanjut untuk dihubungkan dengan atribut perminatan, Sehingga diharapkan mendapatkan informasi atau pola yang berguna sebagai acuan pemutakhiran data.

6. Knowledge

Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat.

2.2. Naïve Bayes

Naive Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas [5]. Persamaan dari teorema bayes adalah :

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)}$$

Dimana :

X : Data dengan class yang belum diketahui

H : Hipotesis data merupakan suatu class spesifik

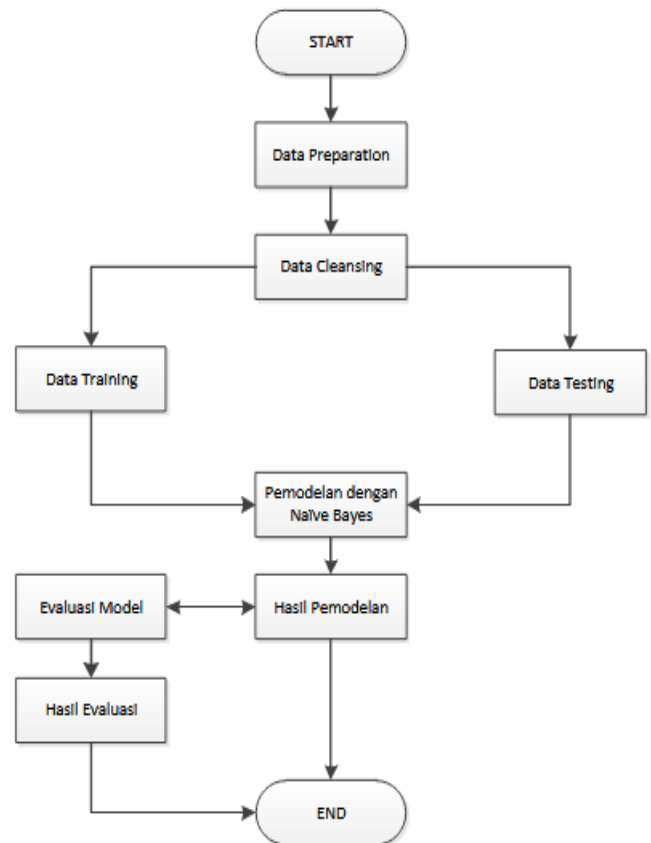
P(H|X) : Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)

P(H) : Probabilitas hipotesis H (prior probabilitas)

P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H

P(X) : Probabilitas X

Dalam melakukan penelitian ini peneliti menggunakan metode algoritma Naïve Bayes. Penelitian ini akan menghasilkan sebuah permodelan klasifikasi rating sekolah. Dalam membangun model klasifikasi untuk penelitian ini terdapat dua tahap proses yang saling berhubungan yaitu tahap training dan tahap testing yang digambarkan dalam gambar 1 berikut:



Gambar 2. Proses Pemodelan

3. Hasil dan Pembahasan

Data yang dipakai adalah data set yang diambil dari situs <https://dapo.kemdikbud.go.id/> yang akhirnya dipakai untuk data training dan data testing. Kemudian dilakukan preprocessing pada data tersebut dengan cara menerapkan metode cleaning. Lalu data tersebut melalui pengolahan, pengolahan data menggunakan software rapid miner yang nantinya bisa menghasilkan sebuah model klasifikasi rating sekolah. Data yang diperoleh berjumlah 114 Sekolah Menengah Atas di Provinsi DKI Jakarta.

Tabel 1. Sampel Tabel Dataset

No	Nama Sekolah	Kategori	Jumlah Review	Rating	Jenis Rating	Jumlah Kelas	Total Siswa
1	Sekolah Menengah Atas Bukit Sion	Siswa	139	4.5	Sedang	12	475
2	Sekolah Menengah Atas PSKD 1 Jakarta	Negeri	75	4.8	Tinggi	11	269
3	Sekolah Menengah Atas	Negeri	86	4.7	Tinggi	10	707

Widya Manggala							
4	Sekolah Menengah Atas Al-HUDA	Negeri	102	38	Rendah	39	687
5	Sekolah Menengah Atas Al-Khairiyah Pagi	Negeri	63	44	Sedang	11	625
6	Sekolah Menengah Atas Muhammadiyah 4	Negeri	85	45	Tinggi	25	733
7	Sekolah Menengah Atas Negeri 109 Jakarta	Negeri	126	48	Tinggi	21	775
8	Sekolah Menengah Atas Negeri 2 Jakarta Barat	Negeri	178	47	Tinggi	20	753
...
1	Sekolah Menengah Kejuruan Negeri 16 Jakarta	Negeri	188	47	Tinggi	13	61
1	Sekolah Menengah Kejuruan Negeri 1 Jakarta	Negeri	67	47	Tinggi	49	77
4	Sekolah Menengah Kejuruan Negeri 1 Jakarta	Negeri	4	7	Tinggi	2	2

a. Data Selection

Data selection mencakup semua atribut dari kumpulan data asli, dipilih untuk mendapatkan atribut yang diperlukan untuk proses penambangan data selanjutnya. Berikut ini menjelaskan proses dimana tolos Rapidminer memilih atribut mana yang akan digunakan dalam tahapan data mining. Pada tahap ini adalah menyeleksi atribut mana saja yang dibutuhkan menggunakan oprator select attributes yang dapat memfilter atribut yang diperlukan. Dengan parameter subset, opsi ini memilih beberapa atribut yang akan digunakan. Jika metadata diketahui, semua atribut dicantumkan dan atribut yang diperlukan dapat dipilih dengan mudah. Berikut dibawah ini adalah hasil seleksi atribut.

Open in Turbo Prep Auto Model Interactive Analysis Filter (114 / 114 examples)

Row No.	Jenis Rating	Nama Sekol...	Kategori	Jumlah Rev...	Rating	Jumlah Kelas	Total Siswa
1	Sedang	Sekolah Men...	Swasta	139	4.500	12	475
2	Tinggi	Sekolah Men...	Negeri	75	4.800	11	269
3	Tinggi	Sekolah Men...	Negeri	86	4.700	10	707
4	Rendah	Sekolah Men...	Negeri	102	3.800	39	687
5	Sedang	Sekolah Men...	Negeri	63	4.400	11	625
6	Sedang	Sekolah Men...	Negeri	85	4.500	25	733
7	Tinggi	Sekolah Men...	Negeri	126	4.800	21	775
8	Tinggi	Sekolah Men...	Negeri	178	4.700	20	753
9	Tinggi	Sekolah Men...	Negeri	99	4.700	15	747
10	Rendah	Sekolah Men...	Swasta	65	3.900	16	667
11	Tinggi	Sekolah Men...	Swasta	160	4.800	14	415
12	Tinggi	Sekolah Men...	Swasta	114	4.800	6	435
13	Tinggi	Sekolah Men...	Swasta	142	4.900	16	637
14	Sedang	Sekolah Men...	Negeri	109	4.500	17	596

ExampleSet (114 examples, 1 special attribute, 6 regular attributes)

Gambar 3. Hasil Seleksi Atribut

b. Tranformasi

Pada tahapan ini, mengubah data menjadi model analisis data dan memodelkan data agar sesuai dengan analisis data mining yang diharapkan, tujuan transformasi adalah mengubah data yang dipilih ke dalam bentuk prosedur penambangan. Mengubah kode jenis rating menjadi label. Atribut dalam peran label berperan sebagai label dan atribut berlabel sebagai operator pembelajaran, label sering disebut sebagai variabel atau kelas. Berikut data yang diolah untuk data mining.

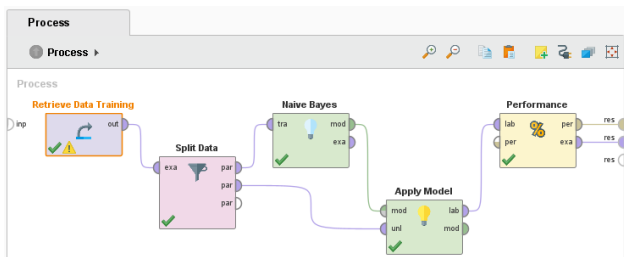
Open in Turbo Prep Auto Model Interactive Analysis Filter (23 / 23 exan)

Row No.	Jenis Rating	prediction(L...	confidence(L...	confidence(L...	confidence(L...	Nama Sekol...	Kategori
6	Tinggi	Tinggi	0.051	0.948	0	Sekolah Men...	Negeri
7	Tinggi	Tinggi	0.380	0.620	0	Sekolah Men...	Negeri
8	Sedang	Sedang	0.933	0.067	0	Sekolah Men...	Negeri
9	Sedang	Sedang	0.999	0.001	0	Sekolah Men...	Negeri
10	Tinggi	Tinggi	0.011	0.989	0	Sekolah Men...	Negeri
11	Tinggi	Tinggi	0.077	0.923	0	Sekolah Men...	Swasta
12	Sedang	Sedang	0.952	0.048	0	Sekolah Men...	Negeri
13	Sedang	Sedang	0.999	0.001	0	Sekolah Men...	Negeri
14	Tinggi	Tinggi	0.249	0.751	0	Sekolah Men...	Negeri
15	Tinggi	Tinggi	0.280	0.720	0	Sekolah Men...	Negeri
16	Tinggi	Tinggi	0.028	0.972	0	Sekolah Men...	Negeri
17	Sedang	Sedang	0.932	0.068	0	Sekolah Men...	Negeri
18	Tinggi	Tinggi	0.004	0.996	0	Sekolah Men...	Negeri
19	Sedang	Sedang	1.000	0.000	0	Sekolah Men...	Swasta

Gambar 4. Hasil Transformasi Data

c. Data Mining

Pada tahapan data mining ini ialah menggunakan algoritma naïve bayes yang berfungsi memecahkan masalah klasifikasi. aplikasi yang digunakan pada proses ini ialah rapidminer versi 10.3. Dengan menggunakan operator Retrieve, Split Data, Naive Bayes dan Performance. Pada gambar dibawah ini ialah proses Data mining dengan memakai algoritma naïve bayes untuk memklasifikasi rating sekolah berdasarkan jumlah siswa sekolah di Provinsi DKI Jakarta.



Gambar 5. Proses Algoritma Naive Bayes

Berdasarkan gambar di atas, Operator Split Data digunakan untuk membagi data menjadi data latih dan data uji. pada subprocess training menggunakan operator naive bayes sebagai penerapan model algoritma.

d. Interpretation/Evaluation

Pada fase ini dilakukan proses evaluasi hasil data mining untuk memudahkan pendataan, salah satunya adalah penggunaan Confusion Matrix. Confusion matrix ialah alat untuk menganalisis seberapa baik sebuah klasifikasi jenis rating sedang, tinggi dan rendah pada sekolah yang terdapat di Provinsi DKI Jakarta.

PerformanceVector

```

PerformanceVector:
accuracy: 95.65%
ConfusionMatrix:
True:   Sedang  Tinggi  Rendah
Sedang: 8       0       1
Tinggi: 0       14      0
Rendah: 0       0       0
    
```

Gambar 6. Performance Vector

Dari hasil pengujian dengan menggunakan algoritma naive bayes didapatkan hasil akurasi mencapai 95,65%.

4. Kesimpulan

Algoritma Naive bayes terbukti dapat memklasifikasi rating sekolah dengan baik berdasarkan jumlah siswa sekolah di Provinsi DKI Jakarta. Hasil dari pengujian pada Rapid Miner menunjukkan nilai akurasi mencapai 95,65%. Dengan melihat hasil pengujian, dapat disimpulkan bahwa algoritma naive bayes cukup baik dalam memklasifikasi rating sekolah berdasarkan jumlah siswa sekolah di Provinsi DKI Jakarta dan dapat digunakan untuk membantu pemangku kepentingan, seperti pemerintah daerah dan lembaga pendidikan, dalam pengambilan keputusan terkait peningkatan kualitas pendidikan dan dapat memberikan kontribusi positif dalam upaya meningkatkan mutu pendidikan di tingkat sekolah menengah atas.

Referensi

- [1] Suwandi, "Klasifikasi Harga Emas Menggunakan Metode Single Moving Average.," *JiTEKH*, 2020.
- [2] K. S. L. R. I. T. & Khoirunnisa, "Klasifikasi Siswa Smk Al Hidayah Yang Masuk Perguruan Tinggi Dengan Metode Klafikasi," *Jurnal Informatika*, 2021.
- [3] Widaningsih, "Perbandingan Metode Data Mining Untuk," *Jurnal Tekno*, p. 16–25, 2019.
- [4] Idris, "Implementasi Data Mining Dengan Algoritma Naive Bayes Untuk Memklasifikasi Angka Kelahiran," *Pelita Informatika*, p. 160–167, 2019.
- [5] T. Patil and S. Sheekar, "Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification," *International Journal of Computer Science and Application*, vol. 6, 2013.