

Penerapan Algoritma Naive Bayes dalam Evaluasi Tingkat Kenyamanan Mobil

Singgih Prayitno¹, Dinda Pradeka², Nina Verina³, Ari Prasetyo⁴, Tri Astuti⁵, Lusi Dwi Oktaviana⁶

^{1,2,3,4,6}Program Studi Sistem Informasi

⁵Program Studi Teknik Informatika

STMIK AMIKOM Purwokerto

Purwokerto, Indonesia

Email : Singgihpra001@gmail.com, dhexa123@gmsil.com, enverinasmoothy@gmail.com, ari13120193@gmail.com, lusiidwi@amikompurwokerto.ac.id

Abstrak— Penelitian ini dilatar belakangi dengan semakin banyaknya jenis-jenis mobil yang muncul seperti sekarang ini, dari sekian banyaknya jenis mobil yang beredar tentu setiap jenis mobil memiliki tingkat kenyamanan yang berbeda-beda. Atas dasar tersebut kami melakukan penelitian ini, agar dapat diketahui evaluasi tingkat kenyamanan dari sebuah mobil dari spesifikasi yang terdapat dalam mobil tersebut. Dataset yang kami gunakan yaitu dataset *car evaluation* hasil penelitian dari Marko Bohanec dengan metode data mining menggunakan algoritma naive bayes. Dari penelitian ini diharapkan mampu menghasilkan prediksi untuk menentukan tingkat kenyamanan dari sebuah mobil.

Kata kunci—*Component; Kenyamanan Mobil; Naive Bayes; Data Mining; Algoritma*

I. PENDAHULUAN

Seiring dengan laju pertumbuhan penduduk yang semakin tinggi serta gaya hidup manusia yang semakin modern, kendaraan terutama mobil merupakan suatu kebutuhan yang digunakan manusia untuk menopang aktivitas kehidupannya sehari-hari. Seseorang akan cenderung untuk memilih jenis mobil yang nyaman untuk dikendarai sehari-hari. Namun dengan semakin banyaknya jenis-jenis mobil yang semakin banyak, maka seseorang akan sulit menentukan pilihan mobil yang akan dikendarainya terutama dari sisi kenyamanan. Atas dasar itulah kami melakukan penelitian ini, yang bertujuan untuk membantu memberikan prediksi mobil mana yang memiliki tingkat kenyamanan baik, bagi orang yang hendak memiliki mobil.

II. TINJAUAN PUSTAKA

Data Mining adalah Serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basisdata dengan melakukan penggalian pola-pola dari data dengan tujuan untuk memanipulasi data menjadi informasi yang lebih berharga yang diperoleh dengan cara mengekstraksi dan mengenali pola

yang penting atau menarik dari data yang terdapat dalam basisdata.

Gartner Group data mining adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika [1] *An introduction to data mining*.

Data mining memiliki beberapa proses, yaitu:

1. Pembersihan data (untuk membuang data yang tidak konsisten dan *noise*). Pada umumnya data yang diperoleh, baik dari database suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa data mining yang kita miliki. Data-data yang tidak relevan itu juga lebih baik dibuang karena keberadaannya bisa mengurangi mutu atau akurasi dari hasil data mining nantinya. *Garbage in garbage out* (hanya sampah yang akan dihasilkan bila yang dimasukkan juga sampah) merupakan istilah yang sering dipakai untuk menggambarkan tahap ini. Pembersihan data juga akan mempengaruhi performansi dari sistem data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.
2. Integrasi data (penggabungan data dari beberapa sumber). Tidak jarang data yang diperlukan untuk data mining tidak hanya berasal dari satu database tetapi juga berasal dari beberapa *database* atau file teks. Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dsb. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan

pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada. Dalam integrasi data ini juga perlu dilakukan transformasi dan pembersihan data karena seringkali data dari dua database berbeda tidak sama cara penulisannya atau bahkan data yang ada di satu database ternyata tidak ada di database lainnya.

3. Transformasi data (data diubah menjadi bentuk yang sesuai untuk di-mining) Beberapa teknik data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa teknik standar seperti analisis asosiasi dan klastering hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut binning. Disini juga dilakukan pemilihan data yang diperlukan oleh teknik data mining yang dipakai. Transformasi dan pemilihan data ini juga menentukan kualitas dari hasil data mining nantinya karena ada beberapa karakteristik dari teknik-teknik data mining tertentu yang tergantung pada tahapan ini.
4. Aplikasi teknik data mining. Aplikasi teknik data mining sendiri hanya merupakan salah satu bagian dari proses data mining. Ada beberapa teknik data mining yang sudah umum dipakai. Kita akan membahas lebih jauh mengenai teknik-teknik yang ada di seksi berikutnya. Perlu diperhatikan bahwa ada kalanya teknik-teknik data mining umum yang tersedia di pasar tidak mencukupi untuk melaksanakan data mining di bidang tertentu atau untuk data tertentu.
5. Evaluasi pola yang ditemukan (untuk menemukan yang menarik/bernilai) Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti : menjadikannya umpan balik untuk memperbaiki proses data mining, mencoba teknik data mining lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat. Ada beberapa teknik data mining yang menghasilkan hasil analisa berjumlah besar seperti analisis asosiasi. Visualisasi hasil analisa akan sangat membantu untuk memudahkan pemahaman dari hasil data mining.
6. Presentasi pola yang ditemukan untuk menghasilkan aksi. Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisa yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami data mining. Karenanya presentasi hasil data mining dalam

bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses data mining. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil data mining [2].

Naive Bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. Naive Bayes didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network*.

Naïve Bayesian Classifier Adalah metode *classifier* yang berdasarkan probabilitas dan Teorema Bayesian dengan asumsi bahwa setiap variabel X bersifat bebas (*independence*) Dengan kata lain, *Naïve Bayesian Classifier* mengansumsikan bahwa keberadaan sebuah atribut (variabel) tidak ada kaitannya dengan beradaan atribut (variabel) yang lain

Karena asumsi atribut tidak saling terkait (*conditionally independent*), maka rumus/formula yang dipakai yaitu:

$$P(X|C_i) = \prod_{k=1}^n P(x_k|C_i)$$

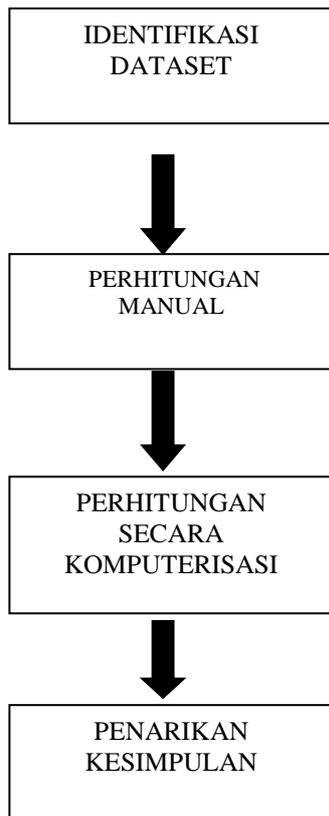
Weka (*Waikato Environment for Knowledge Analysis*) adalah perangkat lunak data mining yang memiliki sekumpulan algoritma standar data mining, Weka dapat dijalankan berbasis GUI (*Graphical user interface*) dan secara langsung melalui *Command Line* (*advanced users*). Weka dapat digunakan untuk melakukan *pre-processing*, klasifikasi, *clustering* (pengelompokan), regresi, *association rule mining* (ARM) dan visualisasi.



Gambar 1. Tampilan Interface WEKA

III. METODE PENELITIAN

Dalam melakukain penelitian ini, ada beberapa tahapan/proses yang kami lakukan:



Gambar 2. Alur/ tahapan dalam penelitian

1. Identifikasi dataset

Proses yang pertama kali kami lakukan yaitu mengidentifikasi dataset, yaitu dengan meneliti secara mendetail dataset tersebut.

2. Perhitungan manual

Dilakukannya perhitungan manual adalah sebagai pembanding dengan perhitungan secara komputerisasi. Adapun kasusnya adalah sebagai berikut:

Mobil X= (*buying=high, maint=med, doors=2, persons=2, lug_boot=small, safety=low*)

3. Perhitungan secara komputerisasi

Dalam penelitian ini, kami menggunakan software WEKA untuk melakukan perhitungan menggunakan komputer.

IV. HASIL DAN PEMBAHASAN

Hasil dari identifikasi dataset, menghasilkan data sebagai berikut:

- Jumlah *attribut* 6
- Jumlah *instances* 100
- Jumlah *missing values* 0

Atribut class memiliki 4 variabel yaitu:

- *UNNAC*: Mobil tidak nyaman
- *ACC*: Mobil cukup nyaman
- *GOOD*: Mobil nyaman
- *VGOOD*: Mobil sangat nyaman

Dengan menggunakan perhitungan manual, didapatkan hasil dari kasus tersebut memiliki *class UNACC* dikarenakan memiliki nilai maksimum tertinggi yaitu 0,005

Dengan menggunakan aplikasi WEKA didapatkan hasil seperti berikut:

No.	1: buying	2: maint	3: doors	4: persons	5: lug_boot	6: safety	7: prediction margin	8: predicted class	9: class
	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Numeric	Nominal	Nominal
1	high	med	2	2	small	low	0.221154	unacc	

Gambar 3. Hasil Weka

Dari perbandingan perhitungan manual dan menggunakan aplikasi, menunjuka hasil variabel unacc untuk kasus Mobil X= (*buying=high, maint=med, doors=2, persons=2, lug_boot=small, safety=low*).

V. KESIMPULAN DAN SARAN

Dari penjelasan diatas dapat ditarik kesimpulan bahwa, penggunaan data mining dapat membantu manusia dalam berbagai prediksi. Seperti contoh diatas, penerapan data mining dapat membantu menilai kualitas kenyamanan dari sebuah mobil.

Untuk mendapatkan data yang valid kami menyarankan penggunaan dataset hasil penelitian yang baru, karena hasil penelitian terbaru dapat merupakan hasil penyempurnaan penelitian-penelitian sebelumnya.

Referensi

- [1] Larose (2006): An introduction to data mining
- [2] Yudho Giri Sucahyo (2003): Penerapan Data Mining Untuk Mengali Tingkat Resiko Calon Pemegang Polis: Studi Kasus Pada PT CAR.