



PENERAPAN DATA MINING UNTUK MEMPREDIKSI PEMBELIAN MOBIL BEKAS MENGGUNAKAN ALGORITMA NAÏVE BAYES

Diana Yusuf ^{1(*)}

¹ITB Ahmad Dahlan, Jakarta

Abstract

Database can also be interpreted as a data warehouse. The amount of data collected in the database can be processed to generate valuable knowledge for science. One popular and widely used technique for processing databases is data mining. Data mining is the process of extracting knowledge from large and complex data warehouse. Data mining encompasses various algorithm to generate knowledge, one of which is naïve bayes. The dataset used in this research, employing the naïve bayes algorithm, consists of attributes relevant to the purchase of used cars, such year, transmission, mileage, car condition, and brand. This research aims to produce patterns and additional knowledge for participants in the used car business to identify the supporting factors in purchasing used cars.

Kata Kunci: Database, Data Mining, Used Car, Naïve Bayes

Januari – Juni 2023, Vol 4 (1) : hlm 29-38
©2023 Institut Teknologi dan Bisnis Ahmad Dahlan.
All rights reserved.

(*) Korespondensi: dianayusuf01@gmail.com (Diana Yusuf)

PENDAHULUAN

Perusahaan mobil bekas merupakan salah satu bidang otomotif yang perkembangannya cukup pesat saat ini. Pesatnya perkembangan ini salah satunya disebabkan oleh ketertarikan konsumen untuk membeli kendaraan bekas. Keputusan membeli kendaraan/mobil bekas umumnya karena pertimbangan finansial dan preferensi pribadi. Bagi perusahaan yang bergerak dibidang otomotif, informasi yang sangat berharga ialah berkaitan dengan prediksi kecenderungan pembelian mobil bekas dan harga mobil bekas.

Perkembangan teknologi yang pesat juga berdampak jumlah data yang besar pula, sehingga dibutuhkan teknik atau metode yang efektif dan tepat untuk mengolah serta menganalisis data dengan jumlah yang besar. Salah satu metode yang cukup populer untuk mengolah dan menganalisis data ialah data mining. Data mining merupakan sebuah teknik ekstraksi pengetahuan dari kumpulan data yang besar (Budi, 2007). Pada pembelian mobil bekas dengan menggunakan data mining mampu membantu proses identifikasi pola dan hubungan yang sebelumnya tidak diketahui dari atribut-atribut mobil bekas.

Data mining adalah suatu proses pendukung pengambil keputusan dimana kita mencari pola informasi dalam data (Dana, 2003). Data mining dalam menemukan pola dan hubungan untuk membantu pengambilan keputusan dapat memanfaatkan beberapa algoritma, salah satunya ialah algoritma naïve bayes. Algoritma naïve bayes masuk ke dalam kelompok data mining klasifikasi yang mengadopsi teorema bayes. Algoritma ini memiliki keunggulan dalam mengatasi permasalahan yang jumlah atributnya banyak dan dapat memprediksi dengan cepat.

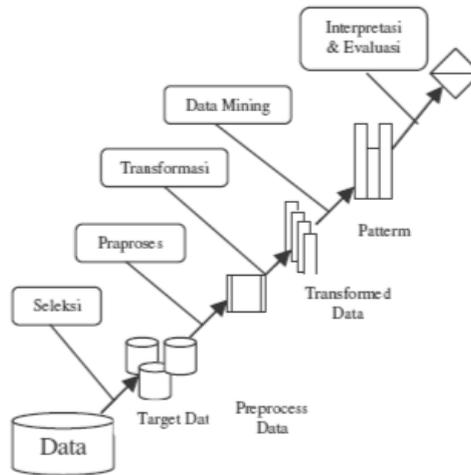
Data yang digunakan dalam penelitian ini berisi beberapa atribut seperti merk mobil, tahun, kilometer, kondisi, transmisi dan target/kelas. Data diperoleh dari Mobi Auto sebagai sampel penelitian ini. Data atribut ini akan diekstraksi menggunakan algoritma naïve bayes untuk memperoleh keputusan. Penelitian ini diharapkan dapat dimanfaatkan oleh pengusaha/pebisnis mobil bekas untuk membantu mereka dalam mengambil keputusan ketika membeli mobil bekas untuk stok dijual kembali. Adanya pengetahuan yang dihasilkan oleh data mining dengan memanfaatkan algoritma naïve bayes akan membantu pekerjaan para inspektur mobil bekas dalam menganalisis kelayakan mobil yang akan dibeli.

Berdasarkan latar belakang diatas penulis tertarik untuk melakukan penelitian dengan judul PENERAPAN DATA MINING UNTUK MEMREDIKSI PEMBELIAN MOBIL BEKAS MENGGUNAKAN ALGORITMA NAÏVE BAYES.

METODE

Knowledge Discovery in Databases (KDD)

Knowledge Discover in Databases (KDD) merupakan kegiatan yang meliputi pengumpulan dan pemakaian data historis untuk menemukan pola atau hubungan dalam set data berukuran besar. Salah satu tahapan dalam KDD adalah data mining yang dapat digunakan untuk membantu pengambilan keputusan pada masa depan (Fayyad, 2007).



Gambar 1. Tahapan KDD

Data Mining

Data mining merupakan sebagian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual (Pramudiono, 2007).

Data mining adalah teknik yang memanfaatkan data dalam jumlah yang besar untuk memperoleh informasi berharga yang sebelumnya tidak diketahui dan dapat dimanfaatkan untuk pengambilan keputusan penting (Dahlan, 2015). Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan yakni :

1. Deskripsi

Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

2. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih kearah numerik daripada kearah kategori. Model dibangun menggunakan record lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya pada penjualan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi.

3. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada dimasa mendatang.

4. Klasifikasi

Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh penggolongan pendapatan dapat dipisahkan dalam kategori yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

5. Pengklusteran

Pengklusteran merupakan pengelompokkan record, pengamatan atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Kluster adalah kumpulan record yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan record-record dalam kluster dalam kluster lain.

6. Asosiasi

Tugas asosiasi dalam data mining merupakan atribut yang muncul dalam satu waktu, dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

Algoritma Naïve Bayes

Algoritma naïve bayes merupakan salah satu algoritma pembelajaran mesin yang efektif dan efisien. Algoritma naïve bayes berasumsi bahwa tidak ada ketergantungan antar atribut, akan tetapi algoritma ini tetap mempunyai performa yang cukup kompetitif dalam pemrosesan klasifikasi pada data mining.

Algoritma naïve bayes menggunakan pendekatan statistik dan probabilitas. Pada algoritma ini, atribut numerik X diubah menjadi atribut nominal X^* pada tahap preprocessing data. Performa algoritma naïve bayes menjadi lebih baik saat atribut numerik diubah ke dalam bentuk atribut nominal. Dalam hal ini, nilai-nilai numerik akan dipetakan ke nilai-nilai nominal dengan memperhatikan interval tetap yang juga mempertimbangkan kelas dari masing-masing nilai numerik yang telah diubah ke bentuk nominal. Langkah-langkah dalam algoritma naïve bayes ialah sebagai berikut :

1. Persiapkan dataset dan data testing
2. Hitunglah jumlah kelas pada data training
3. Hitunglah jumlah kelas yang sama pada data testing yang akan dicari kelasnya dengan kelas yang sama pada data training
4. Kalikan semua hasil perhitungan jumlah kelas pada data testing
5. Bandingkan hasil perkalian per kelas, nilai tertinggi ditetapkan sebagai kelas yang baru

Teorema bayesia menghitung nilai posterior probability $P(H|X)$ menggunakan probabilitas $P(H)$, $P(X)$, dan $P(H|X)$. Rumus yang digunakan bisa seperti berikut ini :

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

Gambar 2. Rumus Naïve Bayes

Metodologi Penelitian

Agar hasil penelitian yang dilakukan maksimal maka penulis harus mengikuti kaidah-kaidah atau metode yang telah ditetapkan. Metodologi penelitian ini berisi kerangka kerja yang akan dilakukan oleh penulis terkait penelitian ini.



Gambar 3. Kerangka Kerja Penelitian

1. Mengidentifikasi masalah

Masalah yang diidentifikasi dalam penelitian ini ialah untuk mengklasifikasikan pembelian mobil bekas.

2. Menentukan tujuan

Tujuan dari penelitian ini ialah :

- a. Untuk menghasilkan pengetahuan tentang seperti apa mobil bekas yang layak dibeli dan sesuai standar perusahaan
- b. Untuk memberikan masukan dan bantuan kepada inspektur dalam memutuskan pembelian mobil bekas agar sesuai standar perusahaan
- c. Meningkatkan daya saing perusahaan mobil bekas

3. Mempelajari literatur

Agar tujuan yang telah ditetapkan tercapai maka hal yang harus dilakukan berikutnya ialah mempelajari literatur yang akan digunakan sebagai pendukung penelitian. Melalui studi literatur dapat diperoleh informasi dan pelajaran dari teori-teori yang berhubungan dengan data mining, klasifikasi, dan algoritma naïve bayes. Teori tersebut dapat bersumber dari buku, jurnal, paper, ataupun situs internet yang berhubungan dengan data mining.

4. Mengumpulkan data

Metode pengumpulan data yang digunakan ialah melakukan wawancara dengan pihak perusahaan mobil bekas dan pihak lain yang terkait dengan penelitian ini.

5. Menganalisis data

Tahapan berikutnya ialah analisis data yang bertujuan agar penulis dapat mengetahui dan memperoleh gambaran jelas tentang bagaimana bentuk

penyelesaian dengan menggunakan algoritma naïve bayes dalam memprediksi pembelian mobil bekas.

6. Pengujian

Pada tahap ini penulis akan melakukan pengujian dengan memanfaatkan tools atau aplikasi data mining yang telah dimasukkan data-data yang akan diuji. Hasil pengujian diharapkan dapat memberikan informasi kepada perusahaan mobil bekas mengenai mobil bekas dengan merk, tahun, kondisi, kilometer, dan transmisi yang seperti apa yang layak dibeli.

HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini ialah data pembelian mobil bekas pada perusahaan otomotif Mobi Auto. Data mentah sebelum dilakukan proses apapun ialah berikut ini :

Tabel 1. Data Pembelian Mobil Bekas

No	State	Plate No	Tahun	Transmisi	KM	Kondisi	Manufacturer	Vehicle Model	UV Exterior Color
1	Canceled	xxxxx	2016	AT	13081	4465000	DAIHATSU	SIRION	Silver Metalik
2	Canceled	xxxxx	2016	MT	47090	Rp 1.665.000,00	TOYOTA	CALYA	Abu Metalik
3	Received	xxxxx	2016	AT	46941	Rp 4.015.000,00	DAIHATSU	SIGRA	Silver Metalik
4	Received	xxxxx	2017	AT	20010	Rp 1.850.000,00	HONDA	CR-V	Hitam Metalik
5	Canceled	xxxxx	2016	AT	50057	Rp 5.015.000,00	TOYOTA	AGYA	Putih Metalik
6	Received	xxxxx					HONDA	CR-V	Hitam Metalik
7	Canceled	xxxxx	2017	AT	30404	Rp 2.200.000,00	MITSUBISHI	PAJERO	Putih Metalik
8	Canceled	xxxxx	2016	AT	28765	Rp 6.800.000,00	HONDA	JAZZ	Putih Metalik
9	Received	xxxxx	2016	AT	46419	Rp 5.500.000,00	TOYOTA	CALYA	Abu Metalik
10	Canceled	xxxxx	2011	AT	11910 3	Rp 3.400.000,00	TOYOTA	AVANZA	Silver Metalik
11	Canceled	xxxxx	2017	AT	15168	Rp 2.200.000,00	TOYOTA	FORTUNER	Hitam Metalik
12	Canceled	xxxxx	2012	MT	46636	Rp 2.650.000,00	NISSAN	GRAND LIVINA	Hitam Metalik
13	Canceled	xxxxx	2012	MT	85305	Rp 3.900.000,00	TOYOTA	AVANZA	Hitam Metalik
14	Received	xxxxx	2016	AT	46941	Rp 4.015.000,00	DAIHATSU	SIGRA	Silver Metalik
...									
63									

Data mobil bekas ini masih mentah dan belum bisa dilakukan pengolahan menggunakan algoritma naïve bayes, sehingga perlu dilakukan transformasi data. Berikut hasil transformasi data yang diperoleh :

Tabel 2. Transformasi Data Atribut Tahun

No	Tahun	Tranformasi (Tahun)
1	2016	MUDA
2	2016	MUDA
3	2016	MUDA
4	2017	TUA
5	2016	MUDA
6	2017	TUA
7	2016	MUDA
8	2016	TUA
9	2011	TUA
10	2017	TUA
11	2012	TUA
12	2012	TUA
13	2016	TUA
14	2015	TUA
...		
63		

Tabel 3. Transformasi Data Atribut Kilometer

No	Kilometer	Tranformasi (Kilometer)
1	13081	RENDAH
2	47090	SEDANG
3	46941	RENDAH
4	20010	TINGGI
5	50057	RENDAH
6	30404	SEDANG
7	28765	SEDANG
8	46419	SEDANG
9	119103	SEDANG
10	15168	TINGGI
11	46636	SEDANG
12	85305	SEDANG
13	62013	SEDANG
14	74621	RENDAH
...		
63		

Tabel 4. Transformasi Data Atribut Kondisi

No	Kondisi	Tranformasi (Kondisi)
1	4.465.000,00	BAGUS
2	1.665.000,00	CUKUP
3	4.015.000,00	CUKUP
4	1.850.000,00	BAGUS
5	5.015.000,00	BAGUS
6	2.200.000,00	BAGUS
7	6.800.000,00	BAGUS
8	5.500.000,00	KURANG
9	3.400.000,00	KURANG
10	2.200.000,00	BAGUS
11	2.650.000,00	CUKUP
12	3.900.000,00	CUKUP
13	3.515.000,00	BAGUS
14	17.515.000,00	KURANG

...		
63		

Tabel 5. Transformasi Data Atribut Kelas

No	Kelas	Tranformasi (Kelas)
1	Canceled	TIDAK BELI
2	Canceled	TIDAK BELI
3	Received	BELI
4	Canceled	TIDAK BELI
5	Canceled	TIDAK BELI
6	Canceled	TIDAK BELI
7	Canceled	TIDAK BELI
8	Canceled	TIDAK BELI
9	Canceled	TIDAK BELI
10	Canceled	TIDAK BELI
11	Received	BELI
12	Received	BELI
13	Canceled	TIDAK BELI
14	Canceled	TIDAK BELI
...		
63		

Proses algoritma naïve bayes untuk memprediksi pembelian mobil bekas sudah dapat dimulai, karena seluruh data sudah ditransformasi seperti yang terlihat pada tabel data training dibawah ini :

Tabel 6. Data Training

No	Tahun	Transmisi	KM	Kondisi	Merk	Kelas
1	MUDA	MT	RENDAH	BAGUS	TOYOTA	TIDAK BELI
2	MUDA	AT	SEDANG	CUKUP	TOYOTA	TIDAK BELI
3	MUDA	AT	RENDAH	CUKUP	TOYOTA	BELI
4	TUA	AT	TINGGI	BAGUS	TOYOTA	TIDAK BELI
5	MUDA	AT	RENDAH	BAGUS	TOYOTA	TIDAK BELI
6	TUA	MT	SEDANG	BAGUS	TOYOTA	TIDAK BELI
7	MUDA	AT	SEDANG	BAGUS	TOYOTA	TIDAK BELI
8	TUA	AT	SEDANG	KURANG	TOYOTA	TIDAK BELI
9	TUA	AT	SEDANG	KURANG	TOYOTA	TIDAK BELI
10	TUA	MT	TINGGI	BAGUS	TOYOTA	TIDAK BELI
11	TUA	AT	SEDANG	CUKUP	TOYOTA	BELI
12	TUA	AT	SEDANG	CUKUP	TOYOTA	BELI
13	TUA	MT	SEDANG	BAGUS	TOYOTA	TIDAK BELI
14	TUA	AT	RENDAH	KURANG	TOYOTA	TIDAK BELI
...						
63						

Langkah untuk menemukan pengetahuan atau pola baru berdasarkan data diatas menggunakan algoritma naïve bayes ialah :

1. Tentukan data testing terhadap data training diatas

Tabel 7. Data Testing

No	Tahun	Transmisi	KM	Kondisi	Merk	Kelas
1	Muda	AT	Sedang	Cukup	Honda	?

2. Hitung jumlah kelas pada data training
Jumlah kelas Beli ialah 14

Jumlah kelas Tidak Beli 49

Maka,

$$P(C = \text{"Beli"}) = 14/63 = 0,22$$

$$P(C = \text{"Tidak Beli"}) = 49/63 = 0,78$$

3. Hitung jumlah kasus yang sama dengan kelas yang sama

$$P(\text{Tahun} = \text{"Muda"} \mid C = \text{"Beli"}) = 5/14 = 0,36$$

$$P(\text{Tahun} = \text{"Muda"} \mid C = \text{"Tidak Beli"}) = 17/49 = 0,35$$

$$P(\text{Transmisi} = \text{"AT"} \mid C = \text{"Beli"}) = 14/14 = 1$$

$$P(\text{Transmisi} = \text{"AT"} \mid C = \text{"Tidak Beli"}) = 38/49 = 0,77$$

$$P(\text{KM} = \text{"Sedang"} \mid C = \text{"Beli"}) = 7/14 = 0,5$$

$$P(\text{KM} = \text{"Sedang"} \mid C = \text{"Tidak Beli"}) = 20/49 = 0,41$$

$$P(\text{Kondisi} = \text{"Cukup"} \mid C = \text{"Beli"}) = 9/14 = 0,64$$

$$P(\text{Kondisi} = \text{"Cukup"} \mid C = \text{"Tidak Beli"}) = 20/49 = 0,41$$

$$P(\text{Merk} = \text{"Honda"} \mid C = \text{"Beli"}) = 9/14 = 0,64$$

$$P(\text{Merk} = \text{"Honda"} \mid C = \text{"Tidak Beli"}) = 11/49 = 0,22$$

4. Kalikan semua hasil sesuai dengan data testing yang akan dicari kelasnya

$$P(X \mid C = \text{"Beli"}) = 0,36 * 1 * 0,5 * 0,64 * 0,64 = 0,074$$

$$P(X \mid C = \text{"Tidak Beli"}) = 0,35 * 0,77 * 0,41 * 0,41 * 0,22 = 0,01$$

Maka,

$$P(C = \text{"Beli"} \mid X) = 0,074 * 0,22 = 0,016$$

$$P(C = \text{"Tidak Beli"} \mid X) = 0,01 * 0,78 = 0,007$$

5. Bandingkan hasil per kelas

Dari perhitungan probabilitas Beli dan probabilitas Tidak Beli, maka dapat disimpulkan bahwa data dengan tahun=muda, transmisi=AT, kilometer=sedang, kondisi=cukup, merk=honda masuk ke dalam kelas Beli. Hal ini karena probabilitas Beli lebih tinggi dibandingkan dengan probabilitas Tidak Beli.

KESIMPULAN

Kesimpulan dari penelitian ini ialah bahwa memprediksi data pembelian mobil bekas menggunakan algoritma naive bayes dapat dilakukan dan telah ditemukan hasil pengetahuan baru. Berdasarkan sampel data sebanyak 63 dan data testing seperti pada bagian diatas diketahui bahwa perusahaan mobil bekas atau inspektor dari perusahaan mobil bekas dapat menemukan pola bahwa mobil dengan tahun produksi muda, transmisinya automatic, kilometernya tidak terlalu tinggi, kondisi perbaikinnnya cukup, dan merknya honda merupakan mobil bekas yang layak untuk dibeli.

DAFTAR PUSTAKA

- Budi Susanto (2007). “*Data Mining Menggali Pengetahuan dari Bongkahan Data*”. Edisi I. Yogyakarta. C.V ANDI OFFSET.
- Dana Sulistiyo Kusumo (2003), “ Data Mining dengan *Algoritma Apriori* pada RDBMS ORACLE”
- Defid Sarjon. 2012. “*Data Mining*”. (Bahan Kuliah *Advance Database*). Padang.
- Kusrini dan Emha Taufik Luthfi (2009). “*Algoritma Data Mining*”. Edisi I. Yogyakarta. C.V ANDI OFFSET.
- Sani Susanto dan Dedy Suryadi (2010). “*Pengantar Data Mining Teknik pemanfaatan data untuk Keperluan Bisnis*”. GRAHA ILMU.
- Feri Sulianta dan Dominikus Juju (2010). “*Data Mining Meramalkan Bisnis Perusahaan*”. Edisi I. Yogyakarta. C.V ALEX MEDIA KOMPUTINDO.