



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

Diana Yusuf.¹

¹*ITB 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 2022, Vol 3 (1) : hlm 34-38
©2022 Institut Teknologi dan Bisnis Ahmad Dahlan.
All rights reserved.

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 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 inspektor mobil bekas dalam menganalisis kelayakan mobil yang akan dibeli.

Berdasarkan latar belakang diatas penulis tertarik untuk melakukan penelitian dengan judul PENERAPAN DATA MINING UNTUK MEMPREDIKSI 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).

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 jmlah 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 deskripsi, estimasi, prediksi, klasifikasi, clusterisasi, dan asosiasi.

Algoritma Naïve Bayes

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.

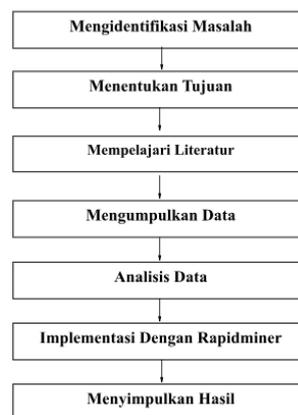
Teorema bayesia menghitung nilai posterior probability P (H|X) menggunakan probabilitas P(H), P(X), dan P(H|X). Rumus yang digunakan bisa seperti berikut ini :

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

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

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 |
| ... | | | | | | | | | |
| 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 sebagai contoh transformasi atribut:

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 |
| ... | | |
| 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 3. 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 |
| ... | | | | | | |
| 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 4. 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

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

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

Maka,

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

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

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, kilomaternya 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.