

IMPLEMENTASI DATA MINING UNTUK MEMPREDIKSI PEMESANAN DRIVER GO-JEK ONLINE DENGAN MENGGUNAKAN METODE NAIVE BAYES (STUDI KASUS: PT. GO-JEK INDONESIA)

Delisman Laia, Efori Buulolo, Matias Julyus Fika Sirait

Prodi Teknik Informatika STMIK Budi Darma, Medan, Indonesia
E-mail: ¹delismanlaia@gmail.com, ²buuloloefori21@gmail.com

Abstrak

PT. Go-Jek Indonesia merupakan perusahaan di bidang jasa. *Go-jek online* merupakan layanan ojek yang berbasis dengan teknologi di bidang jasa yang memimpin revolusi industri transportasi. Prediksi pemesanan *driver go-jek* dengan menggunakan algoritma data mining digunakan untuk menyelesaikan masalah yang dihadapi pihak perusahaan PT.Go-Jek Indonesia untuk memprediksi tingkat pemesanan *driver go-jek online*. Dalam menentukan waktu ramai dan sepi. Metode yang diusulkan adalah *naive bayes*. Algoritma *naive bayes* bertujuan untuk melakukan klasifikasi data pada kelas tertentu. Tujuan penelitian ini adalah untuk melihat pola prediksi dari setiap atribut-atribut yang terdapat pada data set dengan menggunakan algoritma *naive* dan melakukan pengujian data training terhadap data testing untuk melihat pola data baik atau belum. yang akan diprediksi adalah mengumpulkan data-data pemesan driver go-jek sebelumnya yaitu berdasarkan hari, waktu selama satu bulan. Algoritma *Naive Bayes* dipakai untuk memprediksi pemesanan *driver go-jek online* yang akan dialami setiap hari dengan cara melihat setiap pesanan seperti pagi, siang dan sore. Hasil penelitian ini adalah mempermudah pihak perusahaan dalam menganalisis data setiap pemesanan *driver go-jek* dalam mengambil kebijakan untuk mensejahterakan baik itu *driver go-jek* maupun konsumen atau pelanggan.

Kata kunci : *Driver Go-jek, Data Mining, Naive Bayes*

Abstract

PT. Go-Jek Indonesia is a service company. Go-jek online is a technology-based motorcycle taxi service that leads the transportation industry revolution. Predictions on ordering go-jek drivers using data mining algorithms are used to solve problems faced by the company PT. Go-Jek Indonesia to predict the level of ordering of online go-to drivers. In determining the crowded and lonely time. The proposed method is Naive Bayes. Naive Bayes algorithm aims to classify data in certain classes. The purpose of this study is to look at the prediction patterns of each of the attributes contained in the data set by using the naive algorithm and testing the training data on testing data to see whether the data pattern is good or not. what will be predicted is to collect the data of the previous driver ordering, which is based on the day, time for one month. The Naive Bayes algorithm is used to predict the ordering of online go-to-go drivers that will be experienced every day by seeing each order such as morning, afternoon and evening. The results of this study are to make it easier for the company to analyze the data of each go-jek driver booking in taking policies to ensure that both drivers and consumers or customers.

Keywords: *Go-jek Driver, Data Mining, Naive Bayes*

1. PENDAHULUAN

Perusahaan PT. *Go-jek* Indonesia adalah merupakan perusahaan di bidang jasa yang memimpin revolusi industri transportasi. *Go-jek online* merupakan layanan ojek yang berbasis dengan teknologi. Dengan kecanggihan teknologi yang ada saat ini, transportasi ojek ini dapat dipesan melalui online. Adanya ojek *online* ternyata memberikan solusi dan menjawab berbagai kekhawatiran masyarakat akan layanan transportasi umum. *Go-jek* memiliki tarif yang jelas dan transparan sebelum pengguna melakukan konfirmasi pemesanan. Selain itu juga, pelanggan juga bisa menggunakan layanan untuk berbicara langsung dengan menghubungi layanan tersebut. Pelayanan *go-jek* mempunyai banyak kecanggihan, *go-jek* tersedia sepanjang waktu selama 24 jam setiap hari. Dengan *driver go-jek* mempunyai batasan order penumpang di atas 25 km. Jarak terjauh para *driver go-jek* dalam mengantar para konsumennya yakni 25 km. Pelayan ojek ini pintu ke pintu serta mempunyai biaya yang lebih murah di banding dengan transportasi yang lain, misalnya *taxi*. *Go-jek* juga menjadi hal utama bagi mereka yang tinggal di pedalaman kota atau wilayah yang tidak terjangkau dengan angkutan umum.

Pemesanan *driver go-jek* ada waktunya ramai dan ada waktunya sepi. Namun, permasalahan yang dihadapi perusahaan adalah belum bisa memprediksi waktunya ramai dan sepi dalam tingkat pemesanan *driver go-jek*, sehingga *driver go-jek* mengalami kesulitan dalam menunggu orderan. Untuk mengatasi masalah tersebut maka dibutuhkan peran sebuah sistem prediksi pesanan yang dapat membantu memaksimalkan pemesanan dalam mensuplai data pemesan *go-jek* dari data sebelumnya, karena dengan mengetahui data pemesanan perusahaan mengetahui apa saja kendala dan masalah pada pemesan *driver go-jek* tersebut untuk mengambil keputusan dalam pemesanan yang akan datang.

Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data yang digabung dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi untuk pengenalan permasalahan pengambilan informasi dari database yang besar.

Algoritma *Naive Bayes* adalah merupakan salah satu *metode machine learning* yang menggunakan perhitungan probabilitas. Algoritma *Naive Bayes* memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema *Bayes*.

Berdasarkan penelitian dilapangan beberapa kasus yang terkait dengan metode Algoritma *Naive Bayes*. Untuk penelitian sebelumnya menggunakan metode *Naive Bayes* pernah di teliti oleh Effrida Manalu (STMIK Pelita Nusantara) dengan judul “penerapan algoritma *naive bayes* untuk memprediksi jumlah produksi barang berdasarkan data persediaan dan jumlah pemesanan pada cv. Papa dan mama pastries menunjukkan dalam mengelola data-data mengenai persediaan roti dan penerimaan pesanan roti [1].

2. TEORITIS

2.1 Data Mining

Data mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengintegrasikan suatu pengetahuan (*knowledge*) secara otomatis. Data mining di gunakan untuk mengekstrak (mengambil intisari) pengetahuan dari sekumpulan data sehingga didapatkan struktur yang dapat di mengerti manusia serta meliputi basisdata dan manajemen data, prapemrosesan data, pertimbangan model dan inferensi, ukuran ketertarikan, pertimbangan kompleksitas, pasca pemrosesan terhadap struktur yang di temukan, visualisasi, dan *online updating* [2].

2.2 Algoritma Naive Bayes

Algoritma *Naive Bayes* merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema bayes (aturan *bayes*) dengan ansumsi indenpendensi (ketidaktergantungan) yang kuat (naif). Dalam *bayes* (terutama *Naive Bayes*), indenpendensi yang kuat pada fitur adalah sebuah fitur pada sebuah data tidak berkaitan dengan adanya atau tidaknya fitur lain dalam data yang sama. [3]. Prediksi *bayes* didasarkan pada teorema *bayes* dengan formula umum sebagai berikut [4]:

$$P(H|X) = \frac{P(H|X).P(H)}{P(X)} E_2$$

Dimana :

X = Data dengan class yang belum di ketahui

H = Hipotesa data merupakan suatu slass spesifik

$P(H|X)$ = Probabilitas hipotesa H berdasar kondisi X (*posteriori probabilitas*)

$P(H)$ = probabilitas hipotesis H (*prior probabilitas*)

$P(X|H)$ = probabilitas X berdasarkan kondisi pada hipotesa H

$P(X)$ = probabilitas x

2.3 Prediksi

Prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki, agar kesalahannya (selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat di perkecil. Prediksi tidak harus memberikan jawaban secara pasti kejadian yang akan terjadi, melainkan berusaha untuk mencari jawaban sedekat mungkin yang akan terjadi. Kemampuan membuat prediksi adalah kemampuan perkiraan kejadian atau manfaat yang akan datang berdasarkan hasil observasi, pengukuran, data yang telah dikumpulkan maupun hasil penelitian yang di perlihatkan kecenderungan gejala tersebut [5]

2.4 Driver go-jek Online

Go-jek online merupakan layanan ojek yang berbasis dengan teknologi. *Go-jek* tidak hanya memberikan pelayanan dalam transportasi angkutan orang saja, uniknya perusahaan ini memiliki pelayanan lainnya yang ada di aplikasi *Go-Jek* seperti :

1 *Instant courier*

Go-Jek bisa dimanfaatkan sebagai pengantar barang secara "*real time*". Biaya yang dibayar tentu saja sesuai dengan jarak tempuh yang secara otomatis sudah tertera di aplikasi. Baik dokumen maupun barang bisa diantar. Dengan catatan untuk barang yang akan dikirimkan tidak boleh melebihi dari pada jarak stang motor dan tinggi pengemudi.

2 *Transportation*

Ini merupakan fungsi utamanya yaitu sebagai sarana transportasi umum.

3 Go-Ride (Ojek Motor)

Layanan ini merupakan layanan sewa antar jemput setiap pemesan pada aplikasi *go-jek* [6]

3. ANALISA

Dalam menentukan waktu ramai dan sepi yang akan diprediksi adalah mengumpulkan data-data pemesan driver go-jek sebelumnya. Untuk mengatasi masalah dalam memprediksi pemesanan diperlukan suatu alat analisis bagi perusahaan untuk mengetahui tingkat tinggi, sedang atau rendahnya pemesanan dengan menggunakan atribut-atribut seperti hari, waktu dan jumlah pemesanan tiap minggu.

Tabel 1. Data Set Driver Go-Jek Kota Medan

No	Nama driver	Jenis K	BK Kendaraan	Jumlah Pemesan Tiap Hari (1 Bulan)																							
				Senin			Selasa			Rabu			Kamis			Jumat			Sabtu			Minggu					
				P	S	Sr	P	S	Sr	P	S	Sr	P	S	Sr	P	S	Sr	P	S	Sr	P	S	Sr			
1	Rio Martondi	L	BK 6388 CH	3	4	3	3	5	7	4	3	4	3	4	3	5	4	6	6	7	6	7	5	9			
2	M. Ferizal	L	BK 6737 AHZ	4	2	8	6	2	7	5	4	5	4	4	5	3	9	2	4	3	8	4	3	5			
3	Rahmad Julkifis	L	BK 6380 AEB	4	3	5	3	4	6	3	3	5	4	5	4	4	3	5	4	4	4	6	6	5			
4	Syahral Gunawan	L	BK 3331 AHI	3	5	4	4	2	6	4	5	3	5	4	5	2	5	2	5	4	3	6	3	6			
5	Akbar Aji S	L	BK 2481 AHA	4	5	3	5	3	5	7	4	5	4	3	4	4	5	3	5	4	4	4	5	6			
6	Amarullah	L	BK 6568 AFI	5	5	4	3	5	2	3	4	2	3	5	4	4	5	4	3	4	7	5	4	4			
7	Agus Saim	L	BK 3122 PAY	2	4	7	5	4	7	4	4	3	3	4	2	5	4	4	5	2	4	3	5	3			
8	Razky	L	BK 4116 AC	4	6	4	3	4	3	4	5	3	5	4	3	5	5	6	4	3	4	5	4	10			
9	Emmanuel	L	BK3363 AHM	5	4	6	4	5	5	4	3	6	5	6	6	6	4	3	4	3	8	3	2	4			
10	Agus Pohang	L	BK 6560 AFH	6	3	5	5	4	4	3	5	3	2	3	6	5	6	5	6	4	4	4	4	6			
11	Reza	L	BK 4610 ACQ	3	5	3	6	7	8	5	4	6	3	3	4	2	3	6	4	6	7	5	6	6			
12	Nanda Amrudin	L	BK 3772 AGK	6	4	5	4	6	7	4	3	6	4	3	5	3	6	5	7	4	6	4	3	5			
13	Sulaeman	L	BK 6800 ACC	7	3	4	6	4	2	3	6	4	4	3	5	5	6	6	3	3	7	5	6	6			
14	Dufid	L	BK 4664 AGP	5	7	5	5	6	4	5	3	6	6	3	4	7	5	6	6	5	8	4	6	5			
15	Deluman Lata	L	BK 5792 OW	3	5	4	4	7	4	6	5	6	5	4	4	5	7	6	4	7	4	4	7	8			
Total Jumlah Pemesan				64	65	70	78	68	77	64	60	67	61	57	59	66	75	67	71	63	83	71	68	88			

3.1 Mencari Rata-Rata Nilai

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

1. Mencari Standar Nilai Hari Senin

Mencari nilai dan standar variabel hari senin pagi

$$= \frac{\text{minggu 1} + \text{minggu 2} + \text{minggu 3} + \text{minggu 4}}{4 \text{ minggu}}$$

$$= \frac{64 + 68 + 72 + 73}{4}$$

$$= \frac{69,24}{15} = 4,61$$

Mencari nilai dan standar variabel hari senin siang

$$= \frac{\text{minggu 1} + \text{minggu 2} + \text{minggu 3} + \text{minggu 4}}{4 \text{ minggu}}$$

$$= \frac{65 + 66 + 56 + 65}{4}$$

$$= \frac{63}{15} = 4,2$$

Mencari nilai dan standar variabel hari senin sore

$$= \frac{\text{minggu 1} + \text{minggu 2} + \text{minggu 3} + \text{minggu 4}}{4 \text{ minggu}}$$

$$= \frac{70 + 64 + 60 + 61}{4}$$

$$= \frac{63,75}{15} = 4,25$$

2. Mencari Standar Nilai Hari Selasa

Mencari nilai dan standar variabel hari selasa pagi

$$\begin{aligned}
 &= \frac{\text{minggu 1} + \text{minggu 2} + \text{minggu 3} + \text{minggu 4}}{4 \text{ minggu}} \\
 &= \frac{78 + 70 + 72 + 67}{4} \\
 &= \frac{71,75}{15} = 4,78
 \end{aligned}$$

Mencari nilai dan standar variabel hari selasa siang

$$\begin{aligned}
 &= \frac{\text{minggu 1} + \text{minggu 2} + \text{minggu 3} + \text{minggu 4}}{4 \text{ minggu}} \\
 &= \frac{68 + 62 + 64 + 56}{4} \\
 &= \frac{62}{15} = 4,13
 \end{aligned}$$

Mencari nilai dan standar variabel hari selasa sore

$$\begin{aligned}
 &= \frac{\text{minggu 1} + \text{minggu 2} + \text{minggu 3} + \text{minggu 4}}{4 \text{ minggu}} \\
 &= \frac{77 + 71 + 66 + 72}{4} \\
 &= \frac{71,5}{15} = 4,76
 \end{aligned}$$

3. Mencari Standar Nilai Hari Rabu

Mencari nilai dan standar variabel hari rabu pagi

$$\begin{aligned}
 &= \frac{\text{minggu 1} + \text{minggu 2} + \text{minggu 3} + \text{minggu 4}}{4 \text{ minggu}} \\
 &= \frac{64 + 67 + 72 + 61}{4} \\
 &= \frac{66}{15} = 4,4
 \end{aligned}$$

Mencari nilai dan standar variabel hari rabu siang

Tabel 2. Probabilitas Dan Kriteria

No	Hari	Rata-Rata	Waktu	Keterangan
1	Senin	4,61	Pagi	Tinggi
		4,2	Siang	Sedang
		4,25	Sore	Sedang
	Jumlah	13,6		
2	Selasa	4,78	Pagi	Tinggi
		4,13	Siang	Sedang
		4,76	Sore	Tinggi
	Jumlah	13,67		
3	Rabu	4,4	Pagi	Sedang
		3,96	Siang	Rendah
		4,61	Sore	Tinggi
	Jumlah	13,97		
4	Kamis	4,45	Pagi	Sedang
		4,1	Siang	Sedang
		4,46	Sore	Sedang
	Jumlah	13,1		
5	Jumat	4,3	Pagi	Sedang
		4,4	Siang	Sedang
		4,48	Sore	Sedang
	Jumlah	13,18		
6	Sabtu	4,51	Pagi	Tinggi
		4,15	Siang	Sedang
		4,73	Sore	Tinggi

	Jumlah	13,6		
7	Minggu	4,63	Pagi	Tinggi
		4,56	Siang	Tinggi
		5,48	Sore	Tinggi
	Jumlah	14,67		

3.2 Mencari probabilitas

- a. Menghitung Nilai $P(C_i)$
 - $P(\text{klasifikasi} = \text{"Rendah"}) = 1/21 = 0.047$
 - $P(\text{klasifikasi} = \text{"Sedang"}) = 11/21 = 0.52$
 - $P(\text{klasifikasi} = \text{"Tinggi"}) = 9/21 = 0.42$
- b. Menghitung Nilai $P(X|C_i)$
 - $P(\text{Pagi} = \text{"Hari"} \mid \text{jumlah} = \text{"rendah"}) = 0/7 = 0$
 - $P(\text{Pagi} = \text{"Hari"} \mid \text{jumlah} = \text{"sedang"}) = 3/7 = 0.42$
 - $P(\text{Pagi} = \text{"Hari"} \mid \text{jumlah} = \text{"tinggi"}) = 4/7 = 0.57$
 - $P(\text{Siang} = \text{"Hari"} \mid \text{jumlah} = \text{"rendah"}) = 1/7 = 0.14$
 - $P(\text{Siang} = \text{"Hari"} \mid \text{jumlah} = \text{"sedang"}) = 5/7 = 0.71$
 - $P(\text{Siang} = \text{"Hari"} \mid \text{jumlah} = \text{"tinggi"}) = 1/7 = 0.14$
 - $P(\text{Sore} = \text{"Hari"} \mid \text{jumlah} = \text{"rendah"}) = 0/7 = 0$
 - $P(\text{Sore} = \text{"Hari"} \mid \text{jumlah} = \text{"sedang"}) = 3/7 = 0.42$
 - $P(\text{Sore} = \text{"Hari"} \mid \text{jumlah} = \text{"tinggi"}) = 4/7 = 0.57$
 - $P(\text{klasifikasi} = \text{"Rendah"}) = 0 \times 0.14 \times 0 = 0$
 - $P(\text{klasifikasi} = \text{"Sedang"}) = 0.42 \times 0.71 \times 0.42 = 0.125$
 - $P(\text{klasifikasi} = \text{"Tinggi"}) = 0.57 \times 0.14 \times 0.57 = 0.045$
- c. Hitung Nilai $P(X|C_i) * P(C_i)$
 - $P(\text{klasifikasi} = \text{"Rendah"}) * P(\text{klasifikasi} = \text{"jumlah"}) = 0.047 \times 0 = 0$
 - $P(\text{klasifikasi} = \text{"Sedang"}) * P(\text{klasifikasi} = \text{"jumlah"}) = 0.52 \times 0.125 = 0.065$
 - $P(\text{klasifikasi} = \text{"Tinggi"}) * P(\text{klasifikasi} = \text{"jumlah"}) = 0.42 \times 0.045 = 0.0189$

Tabel 3. Probabilitas Kriteria Hasil Pemesan

Waktu	Jumlah kategori			Probabilitas		
	Rendah	Sedang	Tinggi	Rendah	Sedang	Tinggi
	1/21	11/21	9/21	0	0.065	0.0189

Dari hasil tersebut, terlihat bahwa nilai probabilitas tertinggi ada pada status pemesanan driver go-jek tinggi sedang dan rendah. Sehingga dapat disimpulkan bahwa prediksi pemesanan *driver go-jek* pada PT. *Go-Jek Indonesia* memiliki status sedang.

4. KESIMPULAN

Adpun yang menjadi kesimpulan dari penelitian ini adalah sebagai berikut :

1. Aplikasi data mining dengan algoritma naive bayes ini dapat digunakan untuk memprediksi hari, waktu, ramai dalam melakukan pemesanan pada PT.Go-Jek Indonesia.
2. Aplikasi data mining ini dapat mempermudah untuk menentukan memprediksi tingkat pemesanan *driver go-jek* diselesaikan dengan menggunakan metode *naive bayes*
3. Data training pemesanan driver go-jek dapat mempengaruhi hasil pengujian, karena pola data training tersebut akan dijadikan sebagai *rule* untuk menentukan kelas pada data pemesanan *driver go-jek* sehingga tinggi, sedang, rendah tingkat akurasi yang didapat dapat mempengaruhi oleh penentuan data *training*.
4. Proses implementasi algoritma *naive bayes* untuk memprediksi tingkat pemesanan *driver go-jek* menggunakan aplikasi *Visual Basic Net 2008*.

5. REFERENSI

- [1] manalu, Effrida; Sianturi, Ariwisanto Fricles; Manalu, Rofendy Mamed;, "PENERAPAN ALGORITMA NAIVE BAYES UNTUK MEMPREDIKSI JUMLAH PRODUKSI BARANG BERDASARKAN DATA PERSEDIAAN DAN JUMLAH PEMESANAN," *Jurnal Mantik Penusa*, vol. 1, no. Jurnal Manajemen Dan Informatika Pelita Nusantara, p. 17, 2017.
- [2] M.sc, Dr. Suyanto, S.T.,; DATA MINING untuk klasifikasi dan klasterisasi data, Jawa Timur: INFORMATIKA, 2017.
- [3] Eko, Prasetyo, DATA MINING- KOnsep Dan Aplikasi Menggunakan MATLAB, 2012, p. 434.
- [4] Bustami;, "PENERAPAN ALGORITMA NAIVE BAYES UNTUK MENGLASIFIKASI DATA NASABAH ASURANSI," *INFORMATIKA*, vol. 8, p. 887, 2014.
- [5] Sulistiono, Eko; Mujilawati, Siti;, "SISTEM PREDIKSI PENJUALAN DENGAN SINGLE MOVING EVERAGE," *JOUTICA-PRESS*, no. Universitas Islam Lamongan, p. 45, 2016.
- [6] Amajida, Fania Darma;, "KREATIVITAS DIGITAL DALAM MASYARAKAT RISIKO PERKOTAAN," *INFORMASI Kajian Ilmu Komunikasi*, vol. 46, no. Unifersitas Indonesia, p. 118, 2016.