
**REKOMENDASI MODA TRANSPORTASI MAHASISWA DENGAN
ALGORITMA NAIVE BAYES
(STUDI KASUS : FAKULTAS SAINS DAN TEKNOLOGI
UIN SYARIF HIDAYATULLAH JAKARTA)**

Rizki Fajar Febriansyah

Fakultas Sains dan Teknologi, Program Studi Sistem Informasi
Universitas Islam Negeri Syarif Hidayatullah Jakarta
Email: rizki.fajarfebriansyah16@mhs.uinjkt.ac.id

Fauzan Rivaldo Sukardi

Fakultas Sains dan Teknologi, Program Studi Sistem Informasi
Universitas Islam Negeri Syarif Hidayatullah Jakarta
Email: fauzan.rivaldo16@mhs.uinjkt.ac.id

Qurrotul Aini

Fakultas Sains dan Teknologi, Program Studi Sistem Informasi
Universitas Islam Negeri Syarif Hidayatullah Jakarta
Email: qurrotul.aini@uinjkt.ac.id

ABSTRAK

Semakin bertambahnya jumlah manusia mengakibatkan pertumbuhan akan kebutuhan transportasi semakin meningkat untuk bergerak dari suatu lokasi ke lokasi lainnya. Saat ini transportasi dibutuhkan oleh seluruh aspek dalam lini kehidupan, tanpa terkecuali mahasiswa-mahasiswa Fakultas Sains dan Teknologi (FST) Universitas Islam Negeri Syarif Hidayatullah Jakarta. Untuk menuju kampus, mereka dapat menggunakan moda transportasi umum maupun pribadi. Ada banyak faktor yang dapat dipertimbangkan untuk menentukan moda transportasi apa yang efektif dan efisien untuk digunakan oleh mahasiswa untuk mencapai kampus. Faktor-faktor tersebut seperti jumlah uang saku mingguan, alokasi untuk transportasi per minggu, jarak menuju kampus, waktu tempuh, kepemilikan kendaraan pribadi, serta kepemilikan SIM. Penelitian ini akan membahas tentang bagaimana cara memberikan rekomendasi moda transportasi yang efektif dan efisien untuk digunakan oleh mahasiswa menuju kampus berdasarkan faktor yang telah disebutkan sebelumnya. Untuk memberikan rekomendasi moda transportasi, dilakukan pengolahan data menggunakan algoritma *naïve bayes* berdasarkan kaidah-kaidah *data mining*. Penelitian ini menggunakan data yang diambil dari mahasiswa Fakultas Sains dan Teknologi (FST) UIN Syarif Hidayatullah Jakarta. Hasil pengujian menggunakan 45 *dataset* didapatkan nilai *Accuracy* sebesar 92.86%, yang dengan demikian dapat memberikan rekomendasi moda transportasi apa yang efektif dan efisien untuk digunakan oleh mahasiswa Fakultas Sains dan Teknologi (FST) untuk menuju UIN Syarif Hidayatullah Jakarta.

Kata kunci: *naïve bayes*; transportasi; *data mining*; *rapidminer*.

ABSTRACT

The increasing number of people has resulted in the increasing growth of transportation needs to move from one place to another. Currently transportation is needed by all levels of society, even for college students Science and technology faculty in UIN Syarif Hidayatullah Jakarta. College students can go to campus with public or private transportation. Many factors that can be used to determine what kind of transportation are needed for college students to reach effective and efficient to go to campus. These factors include the amount of weekly allowance, budget for weekly transportation, distance to campus, traveling time, private vehicle ownership, and driver license ownership. This study will discuss how to provide recommendations of effective and efficient kind of transportation for use by students to go to campus based on the factors previously mentioned. To provide transportation mode recommendations, data processing is performed using the Naïve Bayes algorithm based on the rules of data mining. This study uses data taken from students of the Faculty of Science and Technology UIN Syarif Hidayatullah Jakarta. Based on the results of testing using 45 datasets, obtained Accuracy value of 92.86%, which can thus provide recommendations for what modes of transportation are effective and efficient to be used by students of the Faculty of Science and Technology to go to UIN Syarif Hidayatullah Jakarta.

Keywords: *naïve bayes*; transportation; *data mining*; *rapidminer*.

1. PENDAHULUAN

Aktivitas manusia yang begitu beragam terkadang membuat mereka dituntut untuk memiliki mobilitas yang tinggi dari satu lokasi ke lokasi lainnya. Untuk memenuhi kebutuhan akan mobilitas yang tinggi tersebut, manusia memerlukan alat transportasi. Transportasi dapat didefinisikan sebagai pemanfaatan kendaraan, baik digerakan oleh mesin atau manusia, untuk memindahkan manusia atau barang. [1]. Transportasi saat ini merupakan sebuah kebutuhan bagi keberlangsungan aktivitas hidup manusia. Tidak mengherankan jika banyak moda transportasi yang ada untuk menunjang kebutuhan manusia. Baik itu merupakan moda transportasi darat, udara maupun laut.

Universitas Islam Negeri (UIN) Syarif Hidayatullah Jakarta adalah Perguruan Tinggi Agama Islam Negeri (PTAIN) yang saat ini telah memiliki 12 fakultas. FST adalah salah satu fakultas dengan jurusan terbanyak di UIN Syarif Hidayatullah Jakarta. Ada banyak cara yang dapat dilakukan mahasiswa FST untuk mencapai kampus, seperti kendaraan umum dan kendaraan pribadi. Untuk menentukan moda transportasi, ada banyak faktor yang dapat dijadikan bahan pertimbangan. Contohnya seperti jarak yang ditempuh, jumlah uang saku, dan lamanya waktu yang diperlukan untuk mencapai kampus. Dari faktor-faktor yang telah disebutkan, kita dapat menentukan moda transportasi apa yang sesuai untuk digunakan mahasiswa untuk menunjang aktifitasnya menuju kampus secara efektif dan efisien.

Untuk melakukan klasifikasi dalam *data mining*, algoritma *naïve bayes* dapat dijadikan pilihan untuk tugas tersebut. [2]. Atas kemampuannya untuk memprediksi akan kemungkinan keanggotaan dalam suatu kelas, algoritma *Naïve Bayes* termasuk kedalam model klasifikasi statistik. Memanfaatkan teorema bayes untuk klasifikasi, *Naïve Bayes* memiliki fungsi yang serupa dengan algoritma klasifikasi lain seperti *SVM & k-NN*.

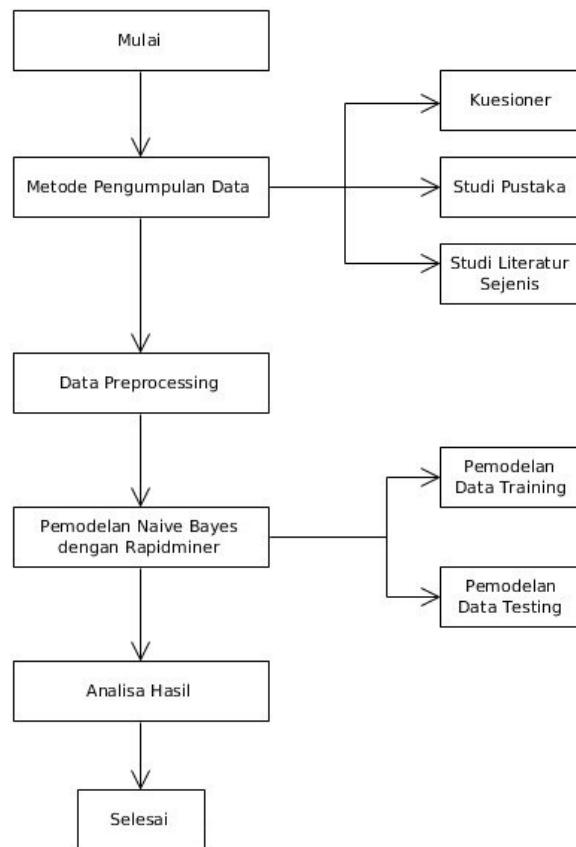
Sebelumnya, telah banyak dilakukan penelitian terkait penggunaan algoritma *Naïve Bayes* sebagai metode klasifikasi *data mining*. Contohnya seperti mendeteksi mana yang termasuk komentar spam bahasa indonesia pada aplikasi Instagram [3]. Kemudian mengevaluasi kinerja akademik mahasiswa di STMIK Dipanegara Makassar [4]. Algoritma *Naïve Bayes* juga mampu melakukan penentuan pemberian kredit [5].

Karena itu, penelitian ini dilakukan untuk memprediksi rekomendasi moda transportasi mahasiswa FST yang efektif dan efisien untuk menuju UIN Syarif Hidayatullah Jakarta.

2. METODOLOGI PENELITIAN

2.1 Metode Penelitian

Pendekatan kuantitatif digunakan sebagai acuan dalam penelitian ini. Penelitian ini menetapkan mahasiswa-mahasiswa FST UIN Syarif Hidayatullah Jakarta sebagai objek penelitian. Berikut adalah *flowchart* dari penelitian ini



Gambar 1. Flowchart Penelitian

2.1.1 Pengumpulan Data/Informasi

Sebelum melaksanakan penelitian, dilakukan pembelajaran terhadap studi literatur dan studi penelitian sejenis terkait penelitian *data mining* serta pemanfaatan algoritma *naïve bayes* sebagai algoritma klasifikasi. Kemudian pengumpulan data yang digunakan sebagai bahan penelitian didapat dengan menggunakan metode kuesioner yang disebarakan kepada sejumlah mahasiswa di FST.

Beberapa variable yang dikumpulkan dalam kuesioner adalah sebagai berikut:

- a. Jumlah uang saku mingguan.
- b. Alokasi/pengeluaran untuk transportasi per minggu.
- c. Jarak ke kampus.
- d. Waktu tempuh.
- e. Kepemilikan kendaraan pribadi.
- f. Kepemilikan SIM.
- g. Keputusan penggunaan jenis kendaraan umum / pribadi.

2.1.2 Pengolahan Data/Data Preprocessing

Data hasil kuesioner akan diolah dengan metode *data preprocessing* agar sesuai dengan kebutuhan penelitian. Data hasil proses ini akan dipecah menjadi 2 bagian dengan perbandingan 70:30, yang artinya sebanyak 70% sebagai *training set* dan 30% sebagai *testing set*.

2.1.3 Pemodelan Naïve Bayes Dengan RapidMiner

Setelah *data preprocessing* maka langkah selanjutnya adalah membuat model *naïve bayes* menggunakan *software RapidMiner*. Pemodelan yang dibangun adalah model untuk memprediksi rekomendasi moda transportasi mahasiswa FST untuk menuju kampus. Pada model juga dilakukan *performance analysis* untuk menentukan tingkat akurasi pemodelan *naïve bayes* yang dibuat.

2.1.4 Analisa Hasil

Setelah membuat selesai membuat pemodelan maka dilakukan analisa terkait hasil penelitian yang didapat.

2.2 Pengertian Data Mining

Data mining merupakan serangkaian proses mulai dari mengumpulkan data hingga mengolah data masa historis dengan tujuan untuk mendapatkan suatu keterkaitan, keteraturan, atau pola yang tersembunyi pada suatu data berukuran besar. Berdasarkan fungsinya, *data mining* dibagi menjadi enam kelompok. Keenam kelompok tersebut adalah deskripsi (mendeskripsikan pola), estimasi (memperkirakan nilai), prediksi (memperkirakan masa depan), klustering (pegelompokkan), klasifikasi (penentuan kategori), dan asosiasi (keterkaitan) [6].

Klasifikasi (*classification*) merupakan proses untuk menempatkan objek tertentu (konsep) berdasarkan tiap – tiap *property* yang dimiliki objek (konsep) tersebut kedalam suatu kategori yang sesuai. Ada empat macam komponen yang mendasari proses klasifikasi, yaitu *class*, *predictor*, *training set*, dan *dataset testing*. Beberapa algoritma klasifikasi yang populer digunakan dalam *data mining* adalah *Naïve Bayes*, *ID3*, *Support Vector Machines (SVM)*, *Neural Networks*, dan *K-Nearest Neighbor (k-NN)*.

2.3 Klasifikasi Naïve Bayes

Teorema *Bayes* adalah sebuah teorema yang menggambarkan peluang bersyarat (*conditional probability*) dari suatu kejadian berdasarkan data dan informasi atau keyakinan mengenai kondisi yang berhubungan dengan kejadian tersebut [7]. Secara umum, berikut adalah persamaan teorema *Naïve Bayes*.

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} \quad (1)$$

Data yang nilai *class* belum diketahui adalah Variabel X. Variabel Y adalah suatu *class* spesifik yang menjadi hipotesis data. Variabel $P(Y|X)$ adalah peluang sesuai kondisi X (*posteriori probability*). Variabel $P(Y)$ adalah peluang Hipotesis Y (*prior probability*). Variabel $P(X|Y)$ adalah peluang X saat kondisi hipotesis Y. Variabel $P(X)$ adalah peluang X. Teorema bayes dikemukakan oleh Thomas Bayes yang merupakan seorang ilmuwan berkebangsaan Inggris. Dikarenakan akurasi & kecepatannya yang baik, *Naïve Bayes* dapat diaplikasikan pada *database Big Data*.

2.4 Data Preprocessing

Data preprocessing merupakan sebuah proses awal yang harus dilakukan terlebih dahulu sebelum *data mining* dilakukan. *Data preprocessing* dapat diartikan sebagai proses pengolahan data mentah agar sesuai dengan proses *data mining* yang dilakukan. Berikut adalah beberapa teknik/metode yang digunakan dalam *data preprocessing* [8].

- Data Cleaning*, merupakan proses untuk menghapus nilai yang memiliki kesalahan, mengoreksi kesalahan yang ada pada data dan mencari inkonsistensi pada data.
- Data Integration*, merupakan proses penggabungan data dari sumber-sumber yang berbeda dan menyimpannya di penampungan data yang tepat.
- Data Transformation*, biasa disebut proses normalisasi merupakan proses untuk mengubah data agar memiliki kesamaan bentuk dengan data yang lain.
- Data Reduction*, merupakan proses memecah data menjadi bentuk-bentuk yang lebih sederhana tanpa menghilangkan nilai analitis dari data itu sendiri.
- Data Discrete*, merupakan proses yang termasuk kedalam *data reduction*, namun lebih dominan untuk menganalisis data-data numerik (angka).

2.5 RapidMiner

RapidMiner mulai diciptakan pada tahun 2001 oleh Ralf Klinkenberg bersama dua orang rekannya yaitu Ingo Mierswa dan Simon Fischer di Universitas Dortmund. Perangkat lunak ini menjadi populer karena fungsinya untuk melakukan hal-hal yang berkaitan dengan *data science*. Umumnya *RapidMiner* digunakan untuk melakukan proses olah data sekaligus analisis data.

RapidMiner bekerja menggunakan konsep-konsep yang ada dalam *data mining*. Mulanya *RapidMiner* mengambil pola-pola dari *dataset* yang memiliki ukuran besar kemudian memadukannya menggunakan kaidah statistika, *artificial intelligence*, dan menyusunnya dalam bentuk *database*.

RapidMiner memiliki operator-operator yang dapat memudahkan penggunanya untuk menghitung data-data yang jumlahnya banyak. Operator-operator ini juga mampu untuk memodifikasi data. Hasil yang ditampilkan oleh *RapidMiner* pun dapat dilihat dalam bentuk visual menggunakan grafik.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Digunakan sebanyak 45 data yang berasal dari mahasiswa FST UIN Syarif Hidayatullah Jakarta. *Dataset* berisi berbagai atribut terkait untuk membuat sistem rekomendasi kendaraan untuk menuju ke UIN Syarif Hidayatullah Jakarta, seperti uang saku, uang untuk biaya transport, kepemilikan kendaraan, jarak ke kampus, dan lain sebagainya. Kuesioner disebar secara online dengan menggunakan *Google Form*. Berikut adalah beberapa data hasil kuesioner.

Tabel 1. Sample Dataset

<i>No</i>	<i>Us</i>	<i>Pt</i>	<i>Jk</i>	<i>Wt</i>	<i>Pk</i>	<i>Ps</i>	<i>Ds</i>
1	100000	45000	22	60	YA	YA	PRIBADI
2	50000	20000	18	45	YA	YA	PRIBADI
3	150000	55000	7	20	TIDAK	TIDAK	UMUM
4	200000	20000	5	15	YA	YA	PRIBADI
5	250000	35000	16	42	TIDAK	TIDAK	UMUM

Catatan:

Us : Uang Saku tiap minggu (Rupiah)

Pt : Alokasi biaya transportasi mingguan (Rupiah)

Jk : Jarak menuju kampus (Kilometer)

Wt : Waktu tempuh (Menit)

Pk : Kepemilikan kendaraan pribadi (Ya/Tidak)

Ps : Kepemilikan SIM (Ya/Tidak)

Ds : Variabel Keputusan (Kendaraan Pribadi / Kendaraan Umum)

3.2 Data Preprocessing

Setelah data kuesioner didapat, maka langkah selanjutnya adalah melakukan proses *Data Preprocessing*. Beberapa proses yang dilakukan pada tahap ini adalah menghapus kolom yang memiliki *missing value*, memperbaiki kesalahan penulisan pada kolom, serta menyeragamkan perbedaan jawaban yang bermakna sama.

3.3 Pemodelan Naïve Bayes dengan RapidMiner

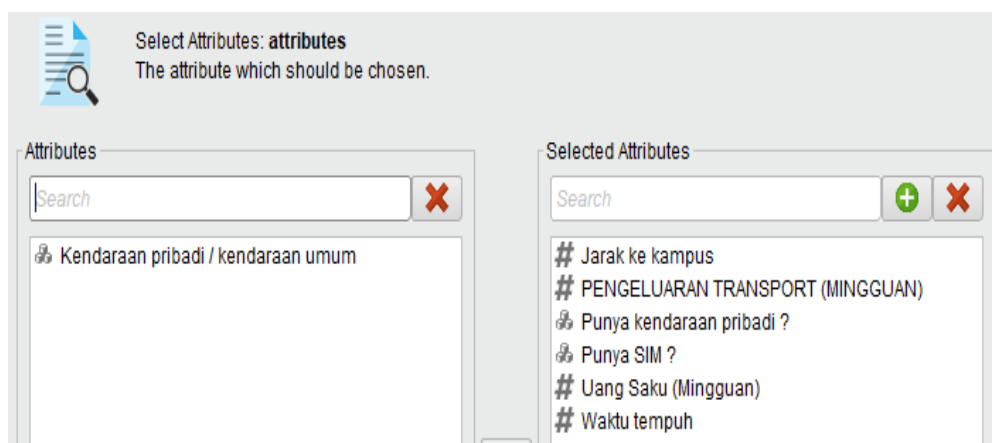
Setelah melakukan *Data Preprocessing*, selanjutnya ialah membuat pemodelan *Naïve Bayes* dengan *RapidMiner*. Penelitian ini menggunakan *RapidMiner* versi 9.2. Untuk membuat pemodelan dengan *RapidMiner* cukup melakukan *drag & drop* operator yang dibutuhkan serta menghubungkan koneksi antar operator.

Pertama Masukkan Operator *Read Excel* yang berisi file *dataset* yang telah melalui *Data Preprocessing*. Kemudian *edit dataset meta information* dimana untuk kolom *decision* (Kendaraan Umum/Kendaraan Pribadi) menjadi *label*.

column index	attribute meta data information			
0	Uang Saku (M	<input checked="" type="checkbox"/> column ...	integer	attribute
1	PENGELUAR	<input checked="" type="checkbox"/> column ...	integer	attribute
2	Jarak ke kamp	<input checked="" type="checkbox"/> column ...	integer	attribute
3	Waktu tempuh	<input checked="" type="checkbox"/> column ...	integer	attribute
4	Punya kendar	<input checked="" type="checkbox"/> column ...	binominal	attribute
5	Punya SIM ?	<input checked="" type="checkbox"/> column ...	binominal	attribute
6	daraan umum	<input checked="" type="checkbox"/> column ...	binominal	label

Gambar 2. Edit Meta Data Information

Selanjutnya, masukkan operator *Select Attributes*. Masukkan semua atribut yang digunakan dalam pemrosesan. Masukkan semua atribut kecuali atribut kendaraan umum / kendaraan pribadi karena merupakan kolom *decision*.



Gambar 3. Select Attributes

Selanjutnya masukkan operator *Split Data*. Tujuannya untuk memecah isi *dataset* kedalam dua bagian, pertama *dataset* untuk *training* dan kedua *dataset* untuk *testing*. Ubah rasio menjadi 0.7 dan 0.3 yang artinya dari 45 data yang dimiliki, sebanyak 31 data digunakan untuk *training set* dan 13 data digunakan untuk *testing set*.

Edit Parameter List: **partitions**
The partitions that should be created.

ratio

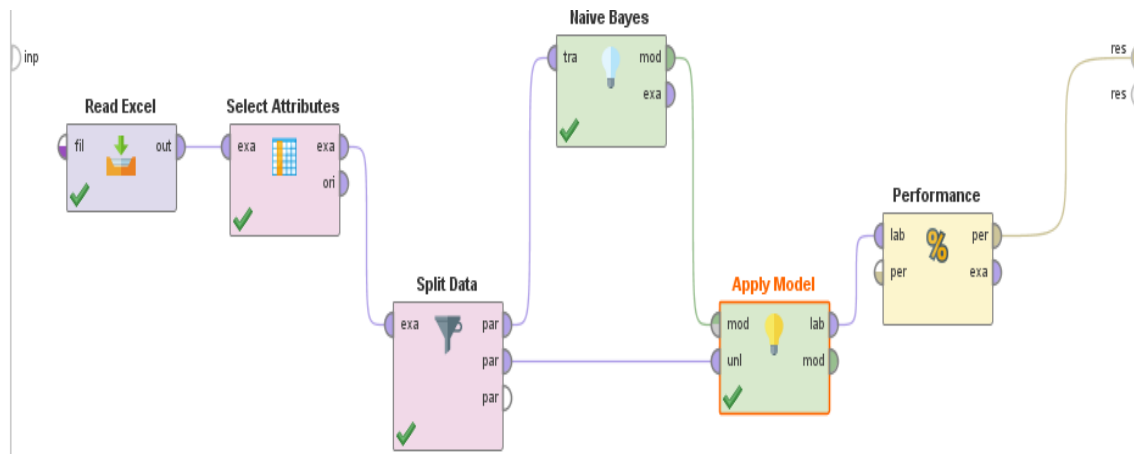
0.7

0.3

Gambar 4. Ubah Rasio Split Data

Setelah itu. Tambahkan operator *naïve bayes* dan operator *Apply Model*. Sambungkan operator *Split Data* ke operator *Naïve bayes* dan operator *Apply Model*. Kemudian sambungkan operator *Naïve bayes* ke operator *Apply Model*. Kemudian tambahkan operator *Performance* untuk mengukur akurasi hasil pemodelan yang dibuat. Ceklis pilihan *accuracy* pada operator *performance*.

Setelah itu sambungkan operator *Performance* ke *res*. Berikut adalah bentuk pemodelan *Naïve Bayes* yang telah dibuat



Gambar 5. Pemodelan *Naïve Bayes* dengan *Rapidminer*

Jalankan model yang dibuat dengan menekan tombol *Play*. Setelah itu *RapidMiner* akan menampilkan hasil *performance* berupa *confusion matrix*. Berikut adalah tabel *confusion matrix* yang didapat

Tabel 2. Model *confusion matrix* hasil pemodelan *naive bayes*

<i>Accuracy : 92.86%</i>			
	<i>True pribadi</i>	<i>True umum</i>	<i>Class precision</i>
<i>Pred. pribadi</i>	6	0	100%
<i>Pred. umum</i>	1	7	87.50%
<i>Class recall</i>	85.71%	100%	

3.4 Analisa Hasil

Setelah diproses, didapatkan hasil dari *PerformanceVector* berupa model *confusion matrix* yang mengandung nilai *accuracy*, *class precision*, dan *class recall* berdasarkan pemodelan yang dibuat. *Precision* didefinisikan sebagai derajat akurasi antara permintaan akurasi yang pengguna inginkan dibandingkan dengan jawaban yang sistem berikan kepada pengguna. Sesuai pemodelan yang telah dibuat, didapatkan hasil *class precision* untuk prediksi kendaraan pribadi sebesar 100% sedangkan *class precision* yang didapat prediksi kendaraan umum sebesar 87.50%. *Class recall* pengertian dari derajat keberhasilan yang dilakukan sistem dalam mendapatkan informasi kembali. Nilai *class recall* yang didapat untuk kendaraan pribadi sebesar 85.71% dan untuk kendaraan umum mendapat nilai *class recall* sebesar 100%. *Accuracy* adalah presentase kedekatan antara nilai prediksi dengan nilai aktual. Nilai *accuracy* yang didapat dari pemodelan ini sebesar 92.86%.

4. KESIMPULAN

Untuk menuju UIN Syarif Hidayatullah Jakarta, mahasiswa FST memerlukan moda transportasi. Mahasiswa dapat menggunakan moda transportasi umum maupun pribadi untuk mencapai kampus. Untuk menentukan moda transportasi yang digunakan, ada banyak faktor yang dapat dijadikan bahan pertimbangan oleh mahasiswa. Faktor tersebut antara lain seperti jumlah uang saku mingguan, alokasi untuk transportasi per minggu, jarak menuju kampus, waktu tempuh, kepemilikan kendaraan pribadi, serta kepemilikan SIM. Dengan *data mining* algoritma *Naïve Bayes* dihasilkan nilai *Accuracy* sebesar 92.86%, yang dengan demikian dapat memberikan rekomendasi moda transportasi apa yang efektif dan efisien untuk digunakan oleh mahasiswa untuk menuju kampus berdasarkan faktor-faktor yang tercantum dalam penelitian ini.

DAFTAR PUSTAKA

- [1] D. Andriansyah, “Manajemen Transportasi Dalam Kajian Dan Teori,” *Jakarta Pus. Fak. Ilmu Sos. dan Ilmu Polit. Univ. Prof. Dr. Moestopo Beragama*, 2015.
- [2] B. Santosa, “Data mining teknik pemanfaatan data untuk keperluan bisnis,” *Yogyakarta Graha Ilmu*, vol. 978, no. 979, p. 756, 2007.

- [3] Y. Lukito, "Deteksi Komentar Spam Bahasa Indonesia Pada Instagram Menggunakan Naive Bayes," *Ultimatics*, vol. 9, no. 1, pp. 50–58, 2017.
- [4] M. S. Mustafa, M. R. Ramadhan, and A. P. Thenata, "Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," *Creat. Inf. Technol. J.*, vol. 4, no. 2, pp. 151–162, 2018.
- [5] M. H. Rifqo and A. Wijaya, "Implementasi Algoritma Naive Bayes dalam Penentuan Pemberian Kredit," *Pseudocode*, vol. 4, no. 2, pp. 120–128, 2017.
- [6] D. T. Larose, "Discovering Knowledge in Data. New Jersey: John Willey & Sons." Inc, 2005.
- [7] D. Ariadi and K. Fithriasari, "Klasifikasi Berita Indonesia Menggunakan Metode Naive Bayesian Classification dan Support Vector Machine dengan Confix Stripping Stemmer," *J. Sains dan Seni ITS*, vol. 4, no. 2, 2016.
- [8] B. R. CTI, A. A. Gafar, N. Fajriani, U. Ramdani, F. R. Uyun, and N. Ransi, "Implementasi k-means clustering pada rapidminer untuk analisis daerah rawan kecelakaan," in *Prosiding Seminar Nasional Riset Kuantitatif Terapan 2017*, 2017, vol. 1, no. 1.