

KLASIFIKASI JUMLAH KENDARAAN DI SUMATERA UTARA DAN SUMATERA BARAT MENGGUNAKAN ALGORITMA NAIVE BAYES

Andy Tessa Saragih^{1*}, Dwi Afrida¹, R. Fazlun Dinara¹

¹Mahasiswa Prodi Sistem Informasi, STMIK Royal

**email* : andytessa10@gmail.com

Abstract : The Central Bureau of Statistics (BPS) is a non-departmental government agency established as a provider of data or information based on Law No. 6/1960 on Census and Law No. 7/1997 on Statistics. The Central Statistics Agency (BPS) recorded the number of motorized vehicles such as cars, buses, trucks, and motorcycles in the provinces of North Sumatra and West Sumatra in 2020-2021 reaching 3,043,892 million units. The purpose of this study is to classify the number of motorized vehicles in the form of cars, motorcycles, buses and trucks. This research uses quantitative research with Naive Bayes Algorithm analysis model. The data used in this study is data from several regions in North Sumatra Province and West Sumatra Province in 2020-2021. Evaluation of model performance is based on accuracy parameters, precision and total recall of the confusion-matrix. The results of testing the dataset and calculating the model performance parameters have obtained an accuracy value of 100%. With the percentage value of the description of each dataset class, namely a little 70.6%, moderate 21.6%, and a lot 7.8%.

Keywords :classification;confusion-matrix; data; motor vehicles; naive bayes

Abstrak : Badan Pusat Statistik (BPS) adalah lembaga pemerintahan non-departemen yang dibentuk sebagai penyedia data atau informasi berdasarkan UU Nomor 6 Tahun 1960 tentang Sensus dan UU Nomor 7 Tahun 1997 tentang Statistik. Badan Pusat Statistik (BPS) mencatat jumlah kendaraan bermotor seperti mobil, bus, truk, dan sepeda motor di Provinsi Sumatera Utara dan Sumatera Barat pada tahun 2020-2021 mencapai 3.043.892 juta unit. Tujuan dari penelitian ini yaitu untuk mengklasifikasikan jumlah kendaraan bermotor berupa mobil, sepeda motor, bus dan, truk. Adapun penelitian ini menggunakan jenis penelitian kuantitatif dengan model analisis Algoritma Naive Bayes. Data yang digunakan pada penelitian ini adalah data dari beberapa daerah di Provinsi Sumatera Utara dan Provinsi Sumatera Barat tahun 2020-2021. Evaluasi kinerja model didasarkan pada parameter akurasi, presisi dan recall total dari Confusion-Matrix. Hasil pengujian dataset dan perhitungan parameter performa model telah didapat nilai akurasi 100%. Dengan presentase nilai keterangan setiap kelas dataset yaitu Sedikit 70,6%, Sedang 21,6%, dan Banyak 7,8%.

Kata kunci :confusion-matrix; data; kendaraan bermotorklasifikasi; naive bayes

PENDAHULUAN

Badan Pusat Statistik (BPS) merupakan lembaga pemerintahan non-departemen yang dibentuk sebagai penyedia data atau informasi berdasarkan UU Nomor 6 Tahun 1960 tentang Sensus dan UU Nomor 7 Tahun 1960 tentang Statistik. Badan Pusat Statistik (BPS) mencatat jumlah kendaraan bermotor seperti mobil, bus, truk, dan

sepeda motor di Provinsi Sumatera Utara dan Sumatera Barat pada tahun 2020-2021 mencapai 3.043.892 juta unit. Jumlah kendaraan bermotor yang semakin meningkat di jalan mengakibatkan terjadinya kemacetan lalu lintas. Sebagian besar masyarakat Indonesia telah menjadikan kendaraan bermotor sebagai salah satu alat transportasi pertama dibandingkan dengan alat transportasi lain karena dapat menghemat waktu dan biaya menuju tempat tujuan [1].

Mengenai pengertian kendaraan bermotor tercantum dalam pasal 1 ayat 8 Undang-Undang Nomor 22 tahun 2009 tentang Lalu Lintas dan Angkutan Jalan (selanjutnya disebut UULLAJ), “Kendaraan Bermotor adalah setiap kendaraan yang digerakkan oleh perangkat mekanik berupa mesin, selain kendaraan yang berjalan diatas rel” [2].

Saat ini perhitungan jumlah kendaraan masih dilakukan secara manual dengan mencatat setiap kendaraan yang lewat. Adapun cara lain yang dapat digunakan adalah dengan memasang suatu sensor pada ruas jalan tertentu, dimana sensor akan mendeteksi setiap kendaraan yang lewat berdasarkan kuatan tekanan, inframerah, ultrasonik atau jenis sensor lainnya [3]. Maka dari itu, belum adanya sistem yang baik dalam mengklasifikasi jumlah kendaraan. Proses olah data masih dilakukan dengan cara sederhana. Hal tersebut tentunya memakan waktu yang cukup lama.

Badan Pusat Statistik (BPS) membutuhkan suatu sistem yang dapat mengklasifikasi jumlah kendaraan secara efisien. Masalah penentuan klasifikasi tersebut dapat diatasi dengan penerapan Data Mining dengan menggunakan metode Algoritma *Naive Bayes*. Dimana metode tersebut telah banyak digunakan sebelumnya untuk melakukan klasifikasi jumlah kendaraan bermotor lainnya. Penelitian ini akan menghasilkan sebuah prediksi yang dilihat dari nilai akurasi sesuai pakar atau tidak [4].

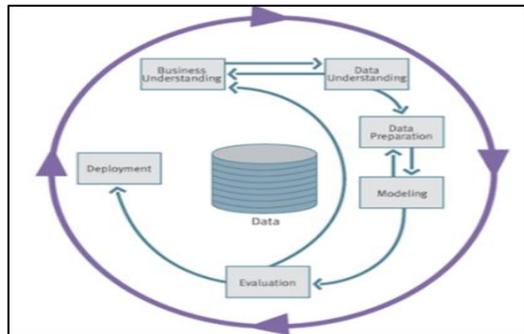
Dari rangkuman riset terdahulu yang sudah dicoba oleh dengan judul “Peningkatan Akurasi Perhitungan Jumlah Kendaraan Dengan Membangkitkan Urutan Identitas Deteksi Berbasis *Yolov4 Deep Neural Networks*”. Riset tersebut dicoba untuk melakukan peningkatan akurasi perhitungan jumlah kendaraan. Dengan tingkat akurasi pada siang hari dimana posisi kamera pada ketinggian 6 m dan sudut 50° sebesar 83%, 93% dan 94%. Sedangkan pada malam hari dimana posisi kamera pada ketinggian 1,5 m dan sudut 90° sebesar 68%, 77% dan 78% .

Penelitian yang dilakukan di Badan Pusat Statistik (BPS) di kota Medan, Sumatra Utara ini bertujuan untuk menerapkan Algoritma *Naive Bayes* untuk mengklasifikasikan jumlah kendaraan di Sumatera Utara dan Sumatera Barat.

METODE

Adapun penelitian ini menggunakan jenis penelitian kuantitatif dengan model analisis Algoritma *Naive Bayes*. Data *Warehouse* adalah suatu paradigma baru di lingkungan pengambilan keputusan strategi. Secara garis besar data *warehouse* adalah sebuah *database* penunjang keputusan yang mengandung data yang biasanya mewakili sejarah bisnis dari suatu perusahaan [5]. Data jenis ini digunakan hanya untuk proses pengklasifikasian, keputusan dan bukan untuk kegiatan operasional perusahaan lain [6].

Alur penelitian ini mengacu pada *Cross-Industry Standard Process* (CRISP-DM) untuk data *warehouse*. CRISP-DM merupakan standart proses data *warehouse* untuk strategi pemecahan masalah secara universal dari bisnis ataupun unit riset [7].



Gambar 1. Tahapan CRISP-DM

Business Understanding

Badan Pusat Statistik (BPS) adalah lembaga pemerintahan non-departemen yang dibentuk sebagai penyedia data atau informasi berdasarkan UU Nomor 6 Tahun 1960 tentang Sensus dan UU Nomor 7 Tahun 1960 tentang Statistik. Badan Pusat Statistik (BPS) mencatat jumlah kendaraan bermotor seperti mobil, bus, truk, dan sepeda motor di Provinsi Sumatera Utara dan Sumatera Barat pada tahun 2020-2021 mencapai 3.043.892 juta unit. Fokus penelitian ini adalah menganalisa bagaimana cara membuat klasifikasi untuk jumlah kendaraan bermotor.

Data Understanding

Sumber data yang didapat dari Badan Pusat Statistika (BPS) di kota Medan, Sumatera Utara. Data tersebut berisikan atribut antara lain : mobil, bus, truk, sepeda motor, jumlah dan keterangan.

```
In [3]: #Melihat Informasi Data
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51 entries, 0 to 50
Data columns (total 7 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   Nama Kota   51 non-null     object
1   Mobil       51 non-null     int64
2   Bus         51 non-null     int64
3   Truk        51 non-null     int64
4   Sepeda Motor 51 non-null     int64
5   Jumlah      51 non-null     int64
6   Keterangan  51 non-null     object
dtypes: int64(5), object(2)
memory usage: 2.9+ KB
```

Gambar 2. Deskripsi Data

Data Preparation

Sebelum masuk kedalam tahap *modelling*, hal yang harus dilakukan terlebih dahulu ialah melakukan persiapan data. Kegunaan persiapan data pada tahap ini adalah untuk memperbaiki masalah yang terdapat pada data sebelumnya, yaitu memilih data, membangun data, mengintegrasikan data, dan membersihkan data untuk menghasilkan data *modelling* yang bagus.

```
In [24]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51 entries, 0 to 50
Data columns (total 5 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   Mobil           51 non-null     int64
1   Bus              51 non-null     int64
2   Truk             51 non-null     int64
3   Sepeda Motor    51 non-null     int64
4   Keterangan       51 non-null     int32
dtypes: int32(1), int64(4)
memory usage: 1.9 KB
```

Gambar 3. Deskripsi Data Yang Dipakai

Data Modelling

Metode yang digunakan dalam penelitian ini yaitu metode Algoritma *Naive Bayes*. Algoritma *Naive Bayes* merupakan pengklasifikasian data dalam jumlah besar untuk memprediksi probabilitas di masa depan yang memiliki kelebihan seperti mudah digunakan, satu kali *scan* data *training*, penanganan nilai atribut yang hilang dan data kontinu. Langkah selanjutnya yaitu menggunakan metode algoritma *Naive Bayes* sebagai alat kepastian dengan menghitung fakta yang keluar. Setelah semua tahapan dilakukan dengan baik menggunakan metode algoritma *Naive Bayes*, hasil akhir yang akan didapatkan akan menghasilkan sebuah prediksi yang dilihat dari nilai akurasi sesuai pakar atau tidak [8]. Dengan penulisan rumus [9] :

$$P(C_i \setminus X) = \frac{P(X \setminus C_i) P(C_i)}{P(X)} \quad (1)$$

Keterangan :

$P(C_i \setminus X)$: Probabilitas hipotesa C_i jika diberikan fakta atau record X (*posterior probability*),

$P(X \setminus C_i)$: Mencari nilai parameter yang memberi kemungkinan yang paling besar (*Likelihood*), dan

$P(X)$: Jumlah probability tuple yang muncul.

Evaluation

Evaluasi model adalah mengukur sejauh mana *performance* dari model *machine learning* yang kita digunakan untuk mengolah *dataset*. Pada tahap ini, digunakan untuk menganalisa kualitas klasifikasi data yang evaluasinya dengan *confusion-matrix* yang menampilkan akurasi, presisi dan *recall*. *Confusion-Matrix* merupakan metode evaluasi yang dapat digunakan untuk menghitung kinerja atau tingkat kebenaran dari proses klasifikasi. *Confusion-Matrix* disini adalah tabel matrix yang terdiri menjadi tiga kelas, yaitu : kelas sedikit, kelas sedang dan kelas banyak.

Metode *confusion-matrix* memberikan informasi aktual dan dapat diprediksi, pada penelitian ini menggunakan *confusion-matrix* agar didapat nilai akurasi perhitungan yang sudah dilakukan sebelumnya menggunakan metode yang dipilih dimana menggunakan metode Algoritma *Naive Bayes*, *confusion-matrix* juga digunakan untuk mendapatkan hasil jumlah kendaraan bermotor dari aspek yang didapat oleh sistem [10]. Adapun rumus untuk mencari Akurasi sebagai berikut :

$$a = \frac{t}{n} \times 100\% \quad (2)$$

Keterangan :

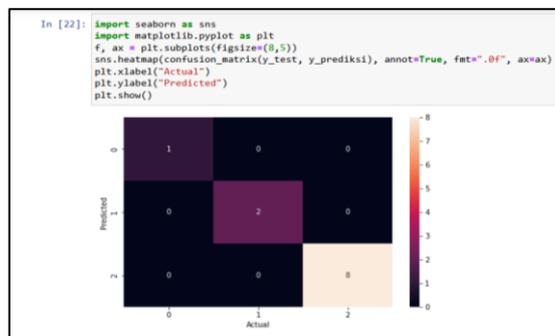
a : Akurasi dalam persen,

t : Jumlah percobaan dengan prediksi valid, dan

n : Jumlah percobaan.

HASIL DAN PEMBAHASAN

Tahap awal dalam analisis, yaitu meng-*import* modul yang diperlukan dan juga *dataset*. Jumlah Kendaraan Bermotor di Daerah Provinsi Sumatra Utara dan Sumatra Barat. Setelah membangun model dengan data *training*, langkah selanjutnya adalah menggunakan data *testing* untuk melakukan pengujian. Visualisasi *Confusion-Matrix* untuk metode *Naive Bayes* adalah :



Gambar 4. Visualisasi *Confusion Matrix*

```
In [23]: from sklearn.metrics import classification_report
print (classification_report(y_test, y_prediksi))
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	1
1	1.00	1.00	1.00	2
2	1.00	1.00	1.00	8
accuracy			1.00	11
macro avg	1.00	1.00	1.00	11
weighted avg	1.00	1.00	1.00	11

Gambar 5. Tingkat Akurasi

Berdasarkan gambar 5, maka dapatlah hasil Algoritma *Naive Bayes* menghasilkan nilai *precision* untuk kelas banyak 100%, untuk kelas sedang 100%, dan untuk kelas sedikit 100%. Nilai *recall* untuk kelas banyak 100%, kelas sedang 100%, dan kelas sedikit 100%. Nilai *f1-score* untuk kelas banyak 100%, kelas sedang 100% dan kelas sedikit 100%. Dan nilai akurasi 100% untuk metode *Naive Bayes*.

SIMPULAN

Data yang digunakan dalam penelitian ini berisi 51 data tabel dengan 5 *field name* yang sudah disortir sebelumnya yaitu mobil, bus, truk, sepeda motor, dan

keterangan. Berdasarkan pembahasan penggunaan Algoritma *Naive Bayes* dan Model *Confusion-Matrix*, mendapatkan nilai *precision* 100%, *recall* 100%, *f1-score* 100% dan nilai akurasi 100% yang masuk dalam kategori *Good Classification*. Pengujian model belum dilakukan secara *real time*. Metode *Naive Bayes* adalah model yang dikategorikan baik dan diimplementasikan untuk memprediksi jumlah kendaraan bermotor. Berdasarkan pengalaman historis sebelumnya untuk mempermudah proses pengklasifikasian jumlah kendaraan bermotor di daerah Sumatra Utara dan Sumatra Barat.

DAFTAR PUSTAKA

- [1] A. M. Mulyadi, B. Perencanaan, P. Daerah, W. Santosa, F. Teknik, and U. K. Parahyangan, "Walkability Pada Kawasan Transit Oriented," vol. 8, no. 1, pp. 27–38, 2022.
- [2] Sjaifurrachman, "Keberadaan Kendaraan Bermotor (Mobil) Pribadi Sebagai Angkutan Umum Dalam Perspektif Undang-Undang Nomor 22," vol. I, no. April, pp. 1–15, 2014.
- [3] F. Rofii, G. Priyandoko, M. I. Fanani, and A. Suraji, "Vehicle Counting Accuracy Improvement By Identity Sequences Detection Based on Yolov4 Deep Neural Networks," *Teknik*, vol. 42, no. 2, pp. 169–177, 2021, doi: 10.14710/teknik.v42i2.37019.
- [4] D. Muriyatmoko, T. Taufiqurrahman, and D. B. Pratama, "Aplikasi Diagnosa Kerusakan Sepeda Motor Menggunakan Metode Naive Bayes dan Rekomendasi Kesesuaian Bengkel," *Semin. Nas. Inform. dan Apl.*, pp. 1–6, 2021.
- [5] Y. Indarta, D. Irfan, M. Muksir, W. Simatupang, and F. Ranuharja, "Analisis dan Perancangan Database Menggunakan Model Konseptual Data Warehouse Sistem Manajemen Transaksi Toko Online Haransaf," *Edukatif J. Ilmu Pendidik.*, vol. 3, no. 6, pp. 4448–4455, 2021, doi: 10.31004/edukatif.v3i6.1477.
- [6] K. Khotimah, "Perancangan Dan Implementasi Data Warehouse Untuk Mendukung Sistem Akademik (Studi Kasus Pada STKIP Muhammadiyah Kotabumi)," *J. Teknol. Inf. Magister Darmajaya*, vol. 2, no. 01, pp. 94–107, 2016.
- [7] Y. Suhandi, I. Kurniati, and S. Norma, "Penerapan Metode Crisp-DM Dengan Algoritma K-Means Clustering Untuk Segmentasi Mahasiswa Berdasarkan Kualitas Akademik," *J. Teknol. Inform. dan Komput.*, vol. 6, no. 2, pp. 12–20, 2020, doi: 10.37012/jtik.v6i2.299.
- [8] A. A. A. Arifin, W. Handoko, and Z. Efendi, "Implementasi Metode Naive Bayes Untuk Klasifikasi Penerima Program Keluarga Harapan," *J-Com (Journal Comput.*, vol. 2, no. 1, pp. 21–26, 2022, doi: 10.33330/j-com.v2i1.1577.
- [9] D. Ayuningsih and N. A. Hasibuan, "Sistem Pakar Mendiagnosa Kerusakan Pada Mesin Penggilingan Padi Menggunakan Metode Naive Bayes," *J. JURIKOM (Jurnal Ris. Komputer)*, vol. 5, no. 4, pp. 371–376, 2018.