

## PREDIKSI KENDARAAN CUTI KADET MAHASISWA UNHAN RI

Nadiza Lediwara<sup>1</sup>, Sembada Denrineksa Bimorogo<sup>2</sup>, Aulia Khamas Heikmakhtiar<sup>3</sup>,  
Fhatur Robby Tanzil Herry<sup>4</sup>, Findi Zhafirah<sup>5</sup>

<sup>1,2,3,4,5</sup>Informatika, Fakultas Sains dan Teknologi Pertahanan,  
Universitas Pertahanan Republik Indonesia

<sup>1</sup>nadizalediwara@gmail.com; <sup>2</sup>denri5693@gmail.com; <sup>3</sup>auliakhphd@gmail.com;

<sup>4</sup>fhaturrobby763@gmail.com; <sup>5</sup>findizhafirah2004@gmail.com

**Abstrak** — Tujuan dari penelitian ini adalah untuk memprediksi jenis kendaraan yang akan digunakan kadet mahasiswa UNHAN RI pada saat cuti. Untuk memprediksi jenis kendaraan yang digunakan, peneliti menggunakan model Logistic Regression dengan menggunakan One-Hot Encoding. Hasil dari penelitian ini dapat digunakan untuk memprediksi jenis kendaraan yang digunakan. Nilai akurasi menunjukkan tingkat keakuratan sebesar 80%. Nilai ini cukup baik sebagai prediksi jenis kendaraan yang akan digunakan kadet mahasiswa cuti.

**Kata Kunci:** *Jenis Kendaraan, Cuti, Logistic Regression*

**Abstrack** — *The purpose of this study is to predict the type of vehicle that will be used by UNHAN RI students during leave. To predict the type of vehicle used, the researcher used the Logistic Regression and One-Hot Encoding. The results of this study can be used to predict the type of vehicle used. The accuracy value shows an accuracy level of 80%. This value is quite good as a prediction of the type of vehicle that will be used by student cadets on leave.*

**Keywords:** *Vehicle Type, Leave, Logistic Regression*

### 1. PENDAHULUAN

Pemanfaatan metode data mining untuk memprediksi jenis kendaraan yang digunakan oleh kadet mahasiswa selama masa cuti berdasarkan provinsi alamat tujuan mereka dapat diterapkan. Seiring dengan meningkatnya mobilitas mahasiswa antarprovinsi, terutama selama periode cuti akademik, pemahaman mengenai preferensi moda transportasi menjadi penting bagi berbagai pihak, termasuk pihak universitas, perusahaan transportasi, dan pemerintah daerah. Mengingat adanya

variasi yang signifikan dalam pilihan transportasi berdasarkan faktor geografis dan demografis, analisis yang mendalam diperlukan untuk mengidentifikasi pola yang dapat membantu dalam pengambilan keputusan yang lebih baik. Data mining telah menjadi teknik yang mampu mengolah data dalam jumlah besar untuk menemukan pola-pola tersembunyi dan menghasilkan prediksi yang akurat. Salah satu aplikasi dalam data mining adalah memprediksi jenis kendaraan berdasarkan data perjalanan. Pada penelitian ini metode Logistic Regression yang dikombinasikan

dengan One-Hot Encoding untuk melakukan prediksi tersebut. Data yang digunakan berasal dari data yang terkumpul ke keseluruhan data yang terhimpun dari keseluruhan kadet mahasiswa UNHAN RI yang berisi informasi asal, tujuan, dan jenis kendaraan. Oleh karena itu penelitian ini bertujuan untuk mengidentifikasi hubungan antara lokasi tujuan kadet mahasiswa dan pilihan moda transportasi mereka. Hasil prediksi ini kemudian dapat digunakan sebagai basis untuk pengembangan model prediktif yang memanfaatkan atribut provinsi tujuan untuk memprediksi jenis kendaraan yang kemungkinan besar akan digunakan. Salah satu model prediksi untuk klasifikasi adalah Logistic Regression (Azimah & Wardani, 2022). Penelitian ini membandingkan tiga model data mining klasifikasi. Dari hasil penelitian tersebut didapatkan untuk memprediksi gejala awal Covid-19, model Logistic Regression memiliki nilai akurasi tertinggi diantara tiga model lainnya. Berdasarkan hal tersebut, untuk memprediksi kendaraan yang digunakan kadet mahasiswa untuk cuti, maka model peneliti juga menggunakan model ini untuk prediksi kendaraan yang akan digunakan kadet mahasiswa.

## 2. LANDASAN TEORI

Salah satu penerapan Logistic Regression yang membahas prediksi pasien terkena diabetes atau tidak (Cahyani et al., 2022). Variabel yang digunakan untuk memprediksi seorang pasien menderita diabetes atau tidak adalah tekanan darah, kadar gula, berat badan, riwayat diabetes dari keluarga, usia, jumlah kehamilan, ketebalan lipatan kulit, dan jumlah kadar insulin. Tahapan preprocessing yang digunakan berupa normalisasi data untuk mengantisipasi rentang data yang berjauhan. Dari hasil pemodelan didapatkan nilai akurasi sebesar 75%. Penelitian kedua adalah implementasi Logistic Regression

untuk mengukur loyalitas konsumen di E-Commerce (Rahman, 2022). Model ini digunakan untuk mengetahui apakah konsumen loyal atau tidak terhadap e-commerce tersebut. Atribut yang digunakan untuk klasifikasi ini dikonversi menjadi data numerik semua seperti status bekerja atau tidak, jenis pekerjaan, tingkat keseringan menggunakan e-commerce, frekuensi berbelanja, dan lainnya. Hasil akurasi yang didapat sebesar 82%. Penelitian ketiga mengenai klasifikasi pembiayaan pinjaman koperasi juga menerapkan model Logistic Regression (P.B. & Rido, 2022). Logistic yang digunakan adalah tipe biner. Tipe biner ini pada label akhir itu hanya terdapat dua keputusan yaitu "Ya" dan "Tidak". Atribut yang digunakan untuk klasifikasi ini antara lain status keanggotaan, kelengkapan data, pekerjaan, pendapatan rata-rata, jumlah pinjaman, jangka waktu, ketersediaan jaminan, dan tanggung jawab lain.

## 3. METODE PENELITIAN

Metodologi penelitian yang digunakan pada penelitian kali ini secara garis besar menggunakan CRIPS-DM metode (Khumaidi, 2020).

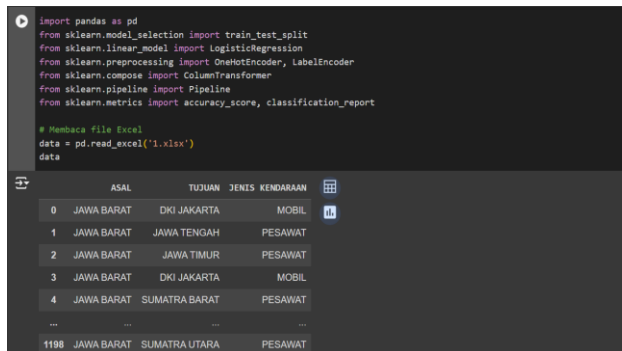
- Tahapan awal yaitu pengumpulan data (Alhapizi et al., 2020) [6]. Data dihimpun berupa sebaran cuti kadet mahasiswa berdasarkan tujuan provinsi masing-masing berdasarkan cohort.
- Tahapan kedua yaitu melakukan preprocessing terhadap data dan menyederagamkan format data agar dapat dijadikan data training pada saat pemodelan nanti. Pada tahapan ini juga ditentukan jenis kendaraan sebagai label tujuan dari pemodelan (Budiman & Widjaja, 2020).
- Tahapan ketiga yaitu melakukan pemodelan. Pemodelan yang dimaksud adalah setelah data training dan data testing dibagi maka data-data tersebut

dimasukkan ke dalam model data mining (Simanjuntak et al., 2022).

- Tahapan terakhir yaitu evaluasi untuk melihat seberapa baik nilai akurasi yang dihasilkan sekaligus menguji model yang dibuat dengan data yang baru (Zai, 2020).

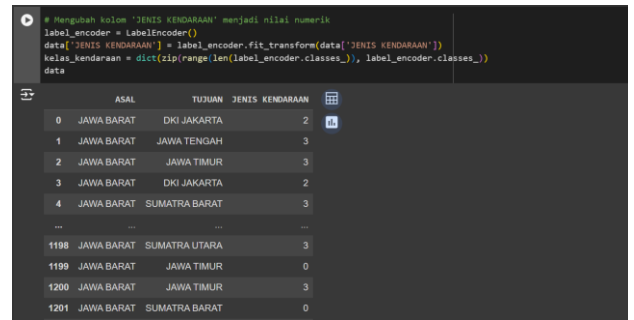
#### 4. HASIL PENELITIAN DAN PEMBAHASAN

Pada tahapan pengumpulan data, data dikumpulkan berdasarkan angkatan (cohort) tiap angkatan. Data ini dihimpun menjadi satu-kesatuan yang nantinya akan menjadi suatu data training yang akan menjadi pembelajaran mesin. Pengumpulan data ini sangat beragam, oleh karena itu perlu adanya penyeragaman format agar data dapat diolah lebih lanjut lagi. Dataset berasal dari sejumlah kadet mahasiswa UNHAN RI sebanyak 1198 orang. Tampilan dataset dapat dilihat pada Gambar 1.



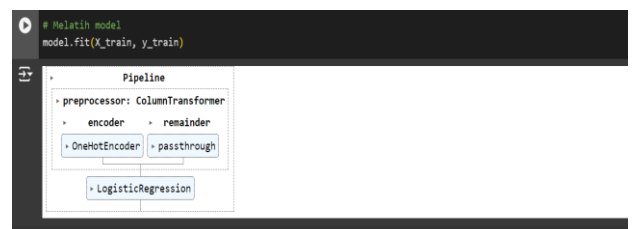
Gambar 1. Dataset Sebaran Cuti

Setelah semua dataset terhimpun maka ditentukan variabel label tujuan yaitu “Jenis Kendaraan”. Pada proses ini label dijadikan sebagai LabelEncoder dari library scikit-learn. Guna LabelEncoder disini adalah untuk mengubah data kategorikal menjadi nilai numerik (Herdian et al., 2024) Tampilan LabelEncoder dapat dilihat pada Gambar 2.



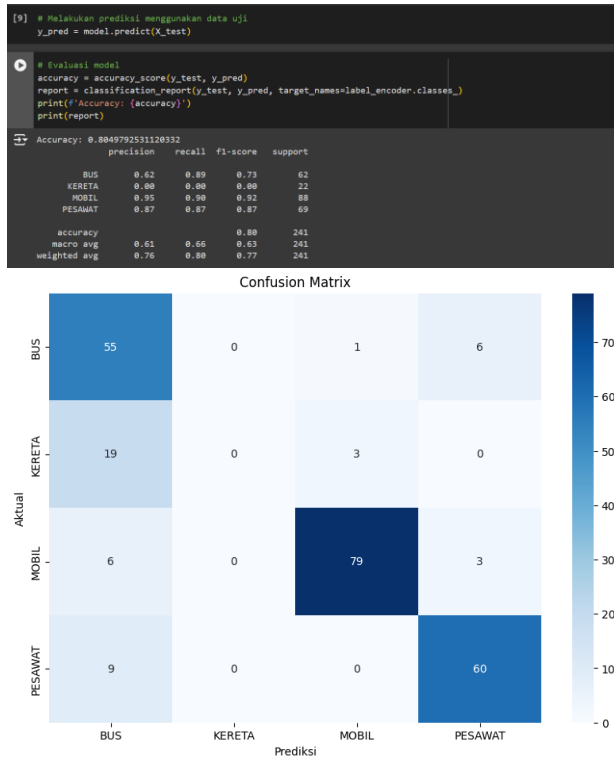
Gambar 2. Tahapan LabelEncoder

Data yang telah dihimpun diberi split data dengan proporsi 80:20. Pembagian ini bertujuan untuk memisahkan dataset menjadi dua bagian: satu untuk melatih model (80%) dan satu untuk menguji model (20%). Ini dilakukan untuk mengevaluasi kinerja model pada data yang tidak terlihat selama pelatihan. Setelah split data tahapan berikutnya adalah transformer untuk one-hot encoding pada kolom 'ASAL' dan 'TUJUAN', serta membuat pipeline yang menggabungkan preprocessing dan pemodelan. One-hot encoding digunakan untuk mengubah data kategorikal menjadi format yang dapat digunakan oleh model machine learning dengan membuat kolom biner untuk setiap kategori unik. Sedangkan pipeline memungkinkan kita untuk menggabungkan beberapa langkah preprocessing dan model machine learning menjadi satu alur kerja yang efisien. Setelah itu adalah melatih model menggunakan data latih. Pada tahap ini, pipeline yang telah dibuat dilatih menggunakan data latih yang telah dibagi sebelumnya, hal tersebut dapat dilihat pada Gambar 3.



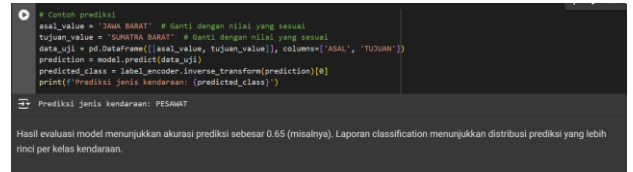
Gambar 3. Melatih Model

Setelah melakukan pelatihan model, berikutnya adalah memprediksi menggunakan data uji dan mengevaluasi model menggunakan metrik akurasi dan *classification report*. Tampilan tersebut dapat dilihat pada Gambar 4.



Gambar 4. Matriks Evaluasi

Dari hasil evaluasi didapat nilai akurasi yaitu 80%. Nilai akurasi ini cukup baik untuk mengklasifikasikan kendaraan yang digunakan kadet mahasiswa untuk cuti. Tahapan pengujian dengan data baru juga dilakukan untuk mengetes apakah model yang telah dibuat dapat bekerja dengan baik. Untuk daerah asal sudah pasti diisi dengan “Jawa Barat” sesuai dengan lokasi kampus UNHAN RI. Tampilan pengujian dengan data baru dapat dilihat pada Gambar 5. Peneliti mengisi tujuan dengan provinsi “Sumatra Barat”, maka kendaraan yang akan digunakan kadet mahasiswa adalah “Pesawat”. Tampilan pengujian dapat dilihat pada Gambar 5.



Gambar 5. Tampilan Pengujian

Dari hasil penelitian ini dapat membangun model prediksi jenis kendaraan berdasarkan data asal dan tujuan perjalanan, menggunakan metode Logistic Regression dan teknik *preprocessing One-Hot Encoding*. Sehingga dapat membantu pihak kampus untuk pengambilan keputusan jika ingin melacak perjalanan cuti kadet mahasiswa.

## 5. KESIMPULAN DAN REKOMENDASI

Penelitian ini telah berhasil membangun model prediksi jenis kendaraan yang digunakan kadet mahasiswa untuk cuti menggunakan metode Logistic Regression dan teknik *One-Hot Encoding*. Model menunjukkan akurasi yang cukup baik, namun masih dapat ditingkatkan. Evaluasi hasil menunjukkan bahwa model dapat digunakan untuk memprediksi jenis kendaraan dengan cukup akurat.

## 6. REFERENSI

- [1] A. Rahman, “Mengukur Loyalitas Konsumen terhadap Suatu *E-Commerce* untuk Meningkatkan Penjualan dengan Metode Logistic Regression,” *Jurnal Ilmu Data*, vol. 2, no. 10, pp. 1–11, 2022, Accessed: Jun. 24, 2024. [Online]. Available: <http://ilmudata.org/index.php/ilmudata/article/view/241/233>
- [2] A. Khumaidi, “Data Mining for Predicting The Amount of Coffee Production Using CRIPS-DM Method,” *Jurnal Techno*

- Nusa Mandiri*, vol. 17, no. 1, pp. 1–8, Feb. 2020,  
doi: 10.33480/TECHNO.V17I1.1240.
- [3] A. E. Budiman and A. Widjaja, “Analisis Pengaruh Teks Preprocessing Terhadap Deteksi Plagiarisme Pada Dokumen Tugas Akhir,” *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 6, no. 3, Dec. 2020,  
doi: 10.28932/JUTISI.V6I3.2892.
- [4] A. Y. Simanjuntak, I. S. S. Simatupang, and A. Anita, “Implementasi Data Mining Menggunakan Metode *Naïve Bayes Classifier* untuk Data Kenaikan Pangkat Dinas Ketenagakerjaan Kota Medan,” *JOURNAL OF SCIENCE AND SOCIAL RESEARCH*, vol. 5, no. 1, pp. 85–91, Feb. 2022,  
doi: 10.54314/JSSR.V5I1.804.
- [5] C. Zai, “Implementasi Data Mining sebagai Pengolahan Data,” *Portal data.org*, vol. 2, no. 3, pp. 1–12, 2020, Accessed: Jun. 24, 2024. [Online]. Available:  
<http://portaldata.org/index.php/portaldat/article/view/107/109>
- [6] C. Herdian, A. Kamila, and I. G. A. M. Budidarma, “Studi Kasus *Feature Engineering* Untuk Data Teks: Perbandingan *Label Encoding* dan *One-Hot Encoding* Pada Metode Linear Regresi,” *Technologia: Jurnal Ilmiah*, vol. 15, no. 1, pp. 93–108, Jan. 2024, doi: 10.31602/TJI.V15I1.13457.
- [7] F. Azimah and K. R. N. Wardani, “Klasifikasi Deteksi Gejala Awal COVID-19 dengan Metode *Logistic Regression*, *Random Forest Classifier* dan *Support Vector Machine*,” *Jurnal Locus Penelitian dan Pengabdian*, vol. 1, no. 6, pp. 405–418, Sep. 2022,  
doi: 10.58344/LOCUS.V1I6.135.
- [8] G.P.P.B. and R. Febryansyah, “Klasifikasi Persetujuan Permohonan Pinjaman pada Koperasi Simpan Pinjam Menggunakan Algoritma *Logistic Regression*,” *Ilmudata.org*, vol. 2, no. 12, pp. 1–12, 2022, Accessed: Jun. 24, 2024. [Online]. Available:  
<http://ilmudata.org/index.php/ilmudata/article/view/281/270>
- [9] M. R. Alhapizi, M. Nasir, and I. Effendy, “Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru Universitas Bina Darma Palembang,” *Journal of Software Engineering Ampera*, vol. 1, no. 1, pp. 1–14, Mar. 2020,  
doi: 10.51519/JOURNALSEA.V1I1.10.
- [10] Q. R. Cahyani, M. J. Finandi, J. Rianti, D. L. Arianti, and A. D. P. Putra, “Diabetes Risk Prediction using *Logistic Regression Algorithm*,” *JOMLAI: Journal of Machine Learning and Artificial Intelligence*, vol. 1, no. 2, pp. 107–114, Jul. 2022,  
doi: 10.55123/JOMLAI.V1I2.598.