

Prediksi Calon Pembeli Mobil Potensial menggunakan Algoritma *Logistic Regression*

Predicting Potential Car Buyers using Logistic Regression Algorithm

¹Nouval Trezandy Lapatta*, ²Abdullah Husin

¹Program Studi Teknik Informatika, Fakultas Teknik, Universitas Tadulako

²Program Studi Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Islam Indragiri

¹Jl. Soekarno Hatta Km. 9, Kel. Tondo, Kec. Mantikulore, Palu, Sulawesi Tengah, Indonesia, 94118

²Jl. Provinsi Parit 1 Tembilahan Hulu, Indragiri Hilir Riau Indonesia, 29213

*e-mail: nouval@untad.ac.id

(received: 29 Maret 2024, revised: 27 April 2024, accepted: 30 April 2024)

Abstrak

Penelitian ini diarahkan pada pengembangan model prediksi untuk mengidentifikasi individu yang memiliki potensi tinggi untuk menjadi pembeli mobil, menggunakan algoritma *logistic regression*. Tujuan utamanya adalah untuk mendukung industri otomotif dalam menyusun strategi pemasaran yang lebih efisien dan terfokus. Alasan pemilihan *logistic regression* terletak pada keunggulannya dalam menangani variabel dependen kategorikal dan kepraktisannya dalam interpretasi hasil. Data yang diolah dalam penelitian ini berasal dari informasi demografis, kebiasaan konsumsi, preferensi merek, dan berbagai faktor lain yang berpengaruh terhadap keputusan pembelian mobil. Sumber data utama adalah hasil survei online yang diikuti oleh individu yang diprediksi berpotensi membeli mobil dalam jangka waktu 12 bulan mendatang. Hasil analisis menunjukkan bahwa faktor-faktor seperti pendapatan, usia, status kepemilikan kendaraan sebelumnya, jenis kelamin dan status pernikahan memiliki peran penting dalam memprediksi kemungkinan seseorang menjadi pembeli mobil. Model yang dikembangkan berhasil mencapai akurasi dan presisi sebesar 95%, yang membuktikan kemampuannya yang signifikan dalam mengidentifikasi calon pembeli mobil dengan tingkat keberhasilan yang tinggi. Temuan ini memberikan wawasan berharga bagi industri otomotif dalam merumuskan strategi pemasaran yang lebih tepat sasaran dan efisien, serta memberikan kontribusi terhadap literatur akademik mengenai penerapan *logistic regression* dalam prediksi perilaku konsumen.

Kata kunci: prediksi pembeli mobil, *logistic regression*, model prediksi, perilaku konsumen.

Abstract

This research aims to develop a predictive model to identify individuals with a high potential to become car buyers, employing logistic regression algorithm. The primary objective is to support the automotive industry in devising more efficient and focused marketing strategies. The choice of logistic regression is based on its superiority in handling categorical dependent variables and its practicality in result interpretation. The data processed in this study derive from demographic information, consumption habits, brand preferences, and various other factors that influence car buying decisions. The main data source is the outcome of online surveys participated in by individuals predicted to have the potential to buy a car within the next 12 months. The analysis results indicate that factors such as income, age, previous vehicle ownership status, gender and marriage status play significant roles in predicting the likelihood of someone becoming a car buyer. The developed model achieved an accuracy and precision of 95%, proving its significant capability in identifying potential car buyers with a high success rate. These findings provide valuable insights for the automotive industry in formulating more targeted and efficient marketing strategies, as well as contributing to the academic literature on the application of logistic regression in consumer behavior prediction.

Keywords: car buyer prediction, logistic regression, predictive model, consumer behavior.

1 Pendahuluan

Industri otomotif global menghadapi perubahan dinamis yang signifikan, didorong oleh perkembangan teknologi, pergeseran preferensi konsumen, dan intensifikasi persaingan pasar. Dalam kondisi pasar yang semakin kompetitif ini, pemahaman tentang preferensi dan perilaku konsumen menjadi penting untuk mempertahankan dan meningkatkan pangsa pasar. Mengidentifikasi calon pembeli mobil potensial tidak hanya membantu perusahaan otomotif dalam mengoptimalkan strategi pemasaran mereka tetapi juga dalam menyesuaikan produk sesuai dengan kebutuhan dan keinginan beragam profil konsumen [1]. Namun, tantangan utama yang dihadapi adalah bagaimana mengolah dan menganalisis volume data konsumen yang besar dengan cara yang efektif untuk menghasilkan prediksi yang akurat [2].

Penelitian ini bertujuan untuk mengatasi masalah tersebut dengan mengembangkan model prediksi yang memanfaatkan algoritma *logistic regression*, sebuah metode yang terkenal dengan kemampuannya dalam mengelola variabel dependen kategorikal dan kemudahan dalam interpretasi hasilnya [3]. Algoritma ini dipilih karena potensinya dalam menangani kompleksitas data dan variabel yang berkaitan dengan perilaku pembelian konsumen, sehingga memungkinkan identifikasi calon pembeli mobil dengan lebih akurat. Selain itu, penggunaan data demografis, kebiasaan konsumsi, dan preferensi merek sebagai variabel independen diharapkan dapat meningkatkan pemahaman tentang faktor-faktor yang mempengaruhi keputusan pembelian mobil [4].

Pernyataan masalah dalam penelitian ini berfokus pada pengembangan strategi pemasaran yang efisien berdasarkan prediksi akurat tentang calon pembeli mobil. Rasionalisasi dari penelitian ini terletak pada kebutuhan untuk memperoleh wawasan yang lebih dalam mengenai faktor-faktor yang mempengaruhi keputusan pembelian konsumen, yang pada gilirannya dapat digunakan untuk merumuskan strategi pemasaran yang lebih tepat sasaran dan efektif [5].

Tujuan utama dari penelitian ini adalah untuk mengembangkan model prediksi yang dapat mengidentifikasi individu dengan potensi tinggi untuk menjadi pembeli mobil menggunakan *logistic regression*. Manfaat dari penelitian ini diharapkan akan memberikan kontribusi signifikan bagi industri otomotif, khususnya dalam merancang strategi pemasaran yang lebih inovatif dan berbasis data. Selain itu, temuan penelitian ini diharapkan dapat memperkaya literatur akademik dalam bidang pemasaran dan analisis prediktif, serta memberikan wawasan baru tentang penerapan *logistic regression* dalam memprediksi perilaku pembelian konsumen.

2 Tinjauan Literatur

Penelitian terkait dengan prediksi penjualan mobil yang optimal telah dikembangkan oleh peneliti sebelumnya dimana terdapat berbagai pendekatan yang telah diterapkan dalam memprediksi penjualan atau kecenderungan pasar mobil, namun terdapat ruang untuk inovasi melalui pemanfaatan *Logistic Regression* [6] secara khusus untuk mengidentifikasi calon pembeli mobil potensial. Penelitian sebelumnya telah menunjukkan efektivitas dari berbagai metode prediktif dalam konteks penjualan mobil, termasuk *Naïve Bayes* [7], ARIMA [8], dan metode lainnya, tetapi belum banyak yang fokus pada identifikasi calon pembeli mobil.

Penggunaan algoritma *Naïve Bayes* telah ditunjukkan efektif dalam memprediksi tingkat penjualan mobil dengan akurasi yang cukup tinggi, seperti yang dijelaskan dalam penelitian oleh Hasyim et al., yang menemukan bahwa model mereka mampu mencapai akurasi sebesar 95.38% dalam prediksinya [9]. Ini menunjukkan potensi besar dalam penerapan teknik data mining dan pembelajaran mesin untuk memahami tren pasar dan preferensi konsumen dalam industri otomotif. Namun, penelitian ini lebih berfokus pada klasifikasi merk mobil berdasarkan kriteria tertentu, bukan pada identifikasi pembeli potensial.

Di sisi lain, penerapan metode ARIMA dalam prediksi penjualan mobil [10], seperti yang dijelaskan dalam penelitian oleh Walelang et al., memberikan pendekatan yang berbeda dengan fokus pada analisis deret waktu untuk memprediksi jumlah penjualan mobil dengan menggunakan model intervensi. Hasil prediksi menunjukkan bahwa penjualan mobil di PT. Hasjrat Abadi Cabang Tendeand Manado konstan sebanyak 75 unit per bulan selama tahun 2022, dengan menggunakan model intervensi ARIMA [11]. Penelitian ini menunjukkan pentingnya analisis deret waktu dalam memprediksi penjualan namun tidak mengeksplorasi potensi individu pembeli atau faktor-faktor yang mungkin mempengaruhi keputusan pembelian.

Pendekatan ini berbeda secara signifikan dari tujuan penelitian tentang prediksi calon pembeli mobil potensial menggunakan *Logistic Regression*, yang tidak hanya bertujuan untuk memprediksi jumlah penjualan tetapi juga untuk mengidentifikasi karakteristik atau faktor-faktor yang mempengaruhi individu dalam membuat keputusan pembelian. Ini mencakup analisis lebih dalam terhadap data konsumen dan perilaku pembelian, yang bisa memberikan wawasan yang lebih spesifik dan personalisasi untuk industri otomotif.

Dengan demikian, tinjauan literatur ini menunjukkan bahwa walaupun telah banyak penelitian yang menerapkan berbagai algoritma pembelajaran mesin untuk prediksi penjualan dalam industri otomotif, masih terdapat ruang yang luas untuk eksplorasi lebih lanjut, khususnya dalam penerapan *Logistic Regression* untuk identifikasi calon pembeli mobil potensial. Kebaruan dari penelitian ini terletak pada aplikasi spesifik dari *Logistic Regression* untuk mengidentifikasi faktor-faktor yang mempengaruhi keputusan pembelian potensial, yang belum banyak dijelajahi dalam literatur sebelumnya. Ini dapat membuka peluang baru untuk pemahaman yang lebih dalam tentang dinamika pasar dan perilaku konsumen dalam industri otomotif, memberikan dasar untuk strategi pemasaran yang lebih efektif dan personalisasi.

3 Metode Penelitian

Penelitian ini terdiri atas beberapa komponen, diantaranya adalah data penelitian, alat penelitian, dan teknik analisis. Teknis analisis pada penelitian ini menggunakan *Logistic Regression*.

3.1 Data Penelitian

Dataset yang digunakan pada penelitian ini merupakan data calon pembeli mobil sejumlah 1000 orang. Terdapat tujuh variabel pada dataset ini seperti yang disajikan pada Tabel 1.

Tabel 1. Variabel Dataset Calon Pembeli

No	Variabel	Deskripsi
1	ID	Nomor data
2	Usia	Usia (tahun)
3	Status	Status nikah (0 = single, 1 = menikah, 2 = menikah punya anak, 3 = duda/janda)
4	Kelamin	Jenis kelamin (0 = laki-laki, 1 = perempuan)
5	Memiliki_Mobil	Jumlah mobil yang saat ini dimiliki
6	Penghasilan	Penghasilan (x 100 juta rupiah per tahun)
7	Beli_Mobil	Keputusan membeli mobil (0 = tidak, 1 = beli)

Variabel `Beli_Mobil` menjadi kelas target pada model prediksi. Variabel lain yang dianggap dapat mempengaruhi keputusan pembelian mobil adalah, variabel `Status`, `Memiliki_Mobil`, dan `Penghasilan`.

3.2 Alat Penelitian

Analisis pada penelitian ini menggunakan Google Collabatory dengan menggunakan *runtime* Python 3 Google Compute Engine Backend pada mode CPU dan RAM sebesar 2GB.

3.3 Logistic Regression

Logistic Regression merupakan teknik analisis statistik yang digunakan untuk memprediksi peluang suatu peristiwa berdasarkan satu atau lebih variabel independen [12]. Teknik ini sangat bermanfaat dalam kasus-kasus di mana variabel dependen yang ingin diprediksi bersifat biner, seperti ya/tidak, sukses/gagal, atau hadir/tidak hadir. *Logistic Regression* beroperasi dengan mengestimasi probabilitas bahwa variabel dependen berada dalam kategori tertentu, dengan memanfaatkan fungsi logit sebagai *link function* [13]. Formula dari *Logistic Regression* seperti pada Persamaan 1 [14].

$$P(Y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)}} \quad (1)$$

<http://sistemasi.ftik.unisi.ac.id>

Keterangan :

($Y=1$) adalah probabilitas bahwa variabel dependen (Y) sama dengan 1, diberikan nilai-nilai dari variabel independen (X).

$\beta_0, \beta_1, \beta_2, \dots, \beta_p$ adalah koefisien regresi yang akan diestimasi.

X_1, X_2, \dots, X_p adalah variabel independen.

Fungsi e adalah fungsi eksponensial, yaitu konstanta Euler (sekitar 2.71828).

Logistic Regression sering digunakan dalam berbagai bidang seperti kedokteran untuk memprediksi kemungkinan penyakit [15], dalam keuangan untuk penilaian risiko kredit [16], dan dalam pemasaran untuk memprediksi perilaku pembelian konsumen [17].

Salah satu kelebihan utama dari *Logistic Regression* adalah mampu menangani variabel prediktor baik yang bersifat numerik maupun kategorikal, dan cocok untuk situasi di mana hubungan antara variabel independen dan peluang kejadian tidak bersifat linear [18].

4 Hasil dan Pembahasan

Terdapat beberapa proses dalam penelitian ini yang akan dibahas, diantaranya praproses data, *Exploratory Data Analysis* (EDA), serta pemodelan dan evaluasi.

4.1 Praproses Data

Sebelum masuk ke tahap pemodelan, maka dilakukan beberapa persiapan data berikut ini. Pada Tabel 2 kita dapat melihat deskripsi dataset dan sebaran data dari daftar 1000 calon pembeli mobil.

Tabel 2. Deskripsi dan Sebaran Data Calon Pembeli

	Usia	Status	Kelamin	Memiliki Mobil	Penghasilan	Beli Mobil
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	43.532000	1.469000	0.481000	0.952000	270.090000	0.633000
std	12.672078	1.073402	0.499889	0.80146	95.23681	0.482228
min	24.000000	0.000000	0.000000	0.000000	95.000000	0.000000
25%	33.000000	1.000000	0.000000	0.000000	187.000000	0.000000
50%	43.000000	1.000000	0.000000	1.000000	258.500000	1.000000
75%	53.000000	2.000000	1.000000	2.000000	352.250000	1.000000
max	164.000000	3.000000	1.000000	4.000000	490.000000	1.000000

Terlihat pada kolom usia terdapat calon pembeli dengan usia tertinggi, yakni 164 tahun. Untuk rasionalisasi data, maka pada penelitian ini akan menghapus data calon pembeli dengan usia lebih dari 100 tahun, dengan pertimbangan sangat jarang sekali pembeli mobil dengan usia lebih dari 100 tahun. Setelah dilakukan penghapusan, jumlah data berkurang 2 menjadi 998 calon pembeli. Hal ini tidak memberikan dampak berarti terhadap model prediksi, karena hanya sebesar 0,2% dari keseluruhan data calon pembeli yang diobservasi. Tabel 3 menunjukkan hasil normalisasi usia calon pembeli, dimana usia terendah adalah 24 tahun dan tertinggi adalah 65 tahun. Selain itu dilakukan pengecekan data *null* (kosong), dengan hasil bahwa seluruh kolom memiliki 998 baris *record* dimana seluruhnya memiliki tipe numerik (bilangan kontinyu) dan tidak terdapat data *null*.

Tabel 3. Hasil Normalisasi Usia Calon Pembeli

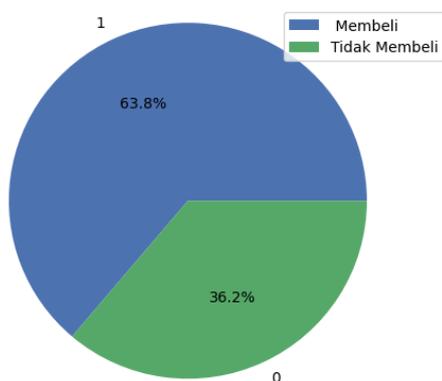
	Usia	Status	Kelamin	Memiliki Mobil	Penghasilan	Beli Mobil
count	998.000000	998.000000	998.000000	998.000000	998.000000	998.000000
mean	43.323647	1.469940	0.480962	0.953908	270.430862	0.634269
std	11.774139	1.073338	0.499888	0.801127	95.026802	0.481876
min	24.000000	0.000000	0.000000	0.000000	95.000000	0.000000
25%	33.000000	1.000000	0.000000	0.000000	187.250000	0.000000
50%	43.000000	1.000000	0.000000	1.000000	259.000000	1.000000
75%	53.000000	2.000000	1.000000	2.000000	352.750000	1.000000
max	65.000000	3.000000	1.000000	4.000000	490.000000	1.000000

4.2 Exploratory Data Analysis

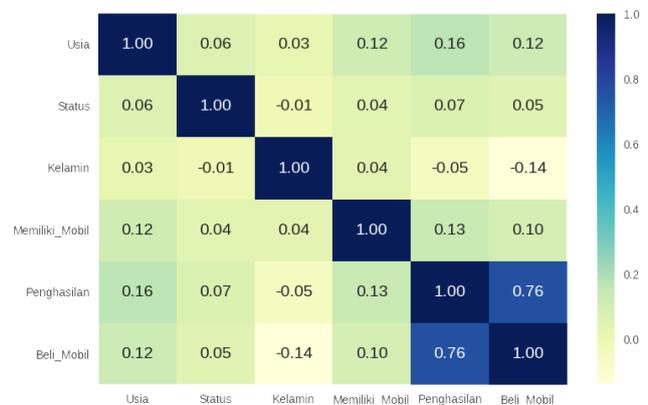
Hasil praproses data calon pembeli mobil disajikan pada Tabel 4 dengan sampel 10 data. Adapun Gambar 1 menunjukkan persentase pembelian mobil berdasarkan data calon pembeli. Dari 998 calon pembeli, sebanyak 631 (63,8%) telah membeli mobil dan 358 (36,2%) tidak membeli mobil. Adapun Gambar 2 menampilkan korelasi antar variabel, dimana terlihat bahwa variabel `Penghasilan` memiliki pengaruh yang paling kuat terhadap keputusan pembelian mobil.

Tabel 4. Hasil Praproses Data Calon Pembeli Mobil

ID	Usia	Status	Kelamin	Memiliki_Mobil	Penghasilan	Beli_Mobil
1	32	1	0	0	240	1
2	49	2	1	1	100	0
3	52	1	0	2	250	1
4	26	2	1	1	130	0
5	45	3	0	2	237	1
6	39	2	0	1	280	1
7	38	2	1	0	150	0
8	29	1	1	0	143	0
9	30	2	0	0	200	1
10	51	1	1	0	174	0

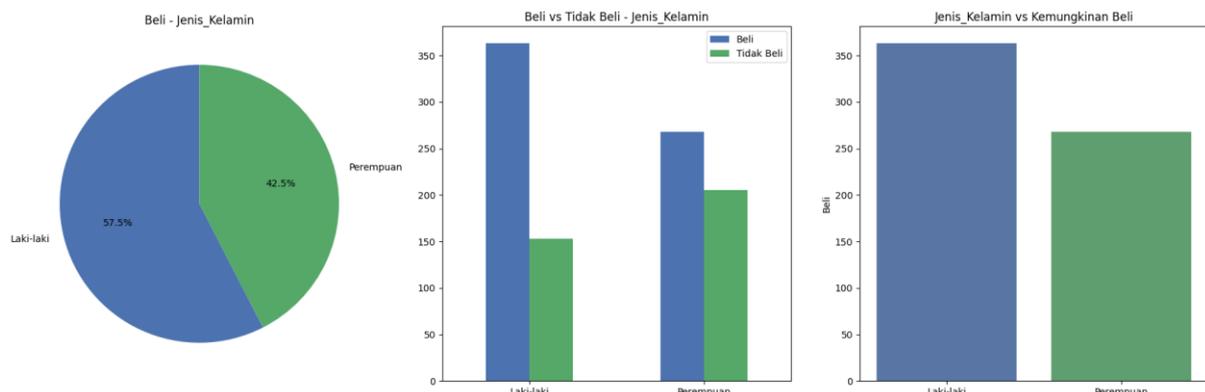


Gambar 1. Persentase Pembelian Mobil



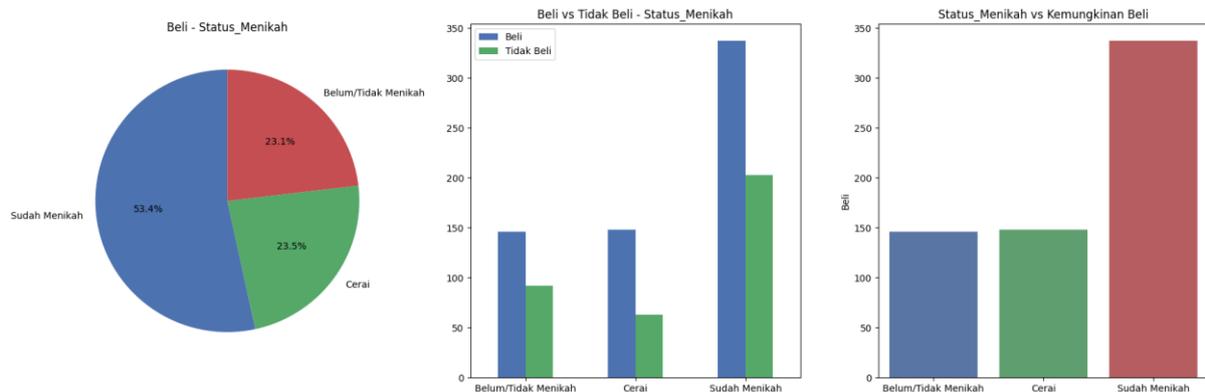
Gambar 2. Korelasi Antar Variabel

Gambar 3 menampilkan hubungan pembelian mobil dan jenis kelamin dari calon pembeli. Mayoritas calon pembeli yang memutuskan untuk membeli mobil adalah laki-laki sebanyak 57,5%, dan perempuan sebanyak 42,5%.



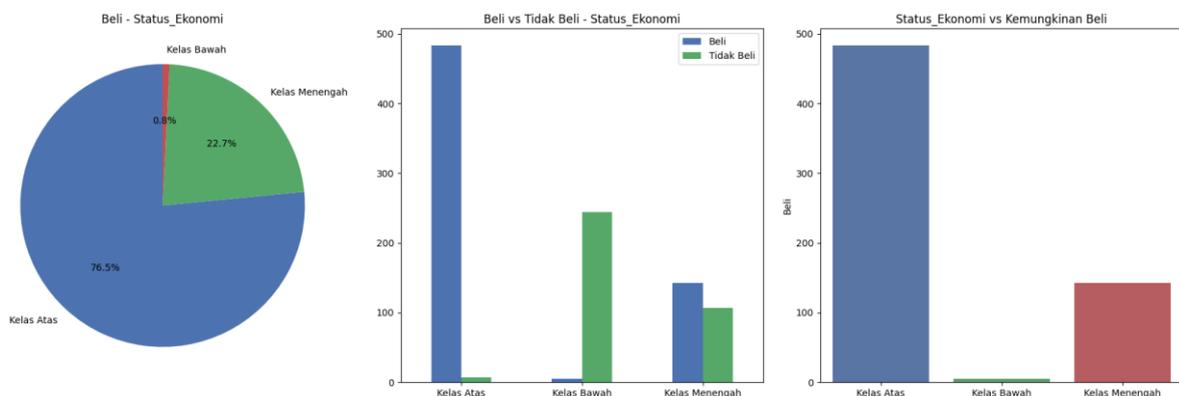
Gambar 3. Hubungan Pembelian Mobil dan Jenis Kelamin

Lebih dari setengah calon pembeli mobil yang akhirnya memutuskan membeli adalah berstatus sudah menikah sebanyak 53,4%. Kemudian diikuti oleh calon pembeli yang berstatus cerai sebanyak 23,5% dan belum atau tidak menikah sebanyak 23,1%, seperti yang ditampilkan pada Gambar 4.



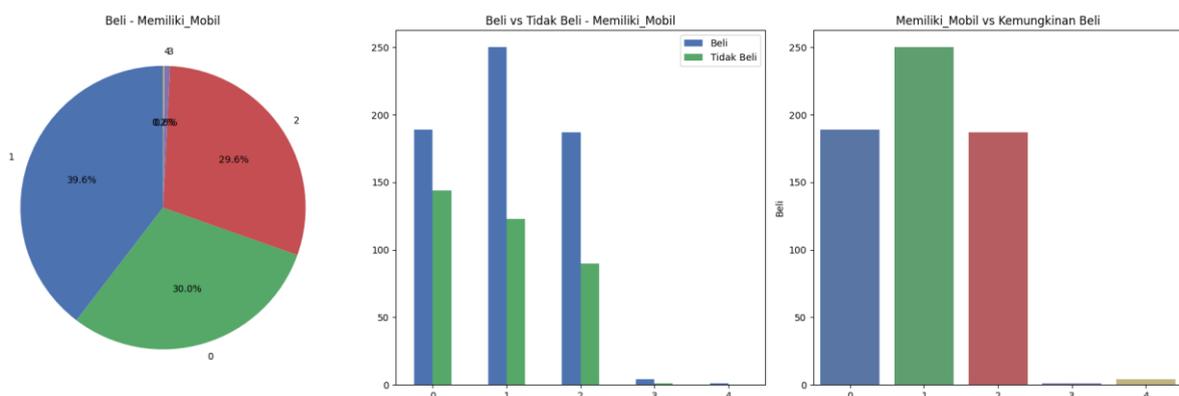
Gambar 4. Hubungan Pembelian Mobil dan Status Menikah

Lebih dari setengah calon pembeli yang akhirnya memutuskan membeli mobil adalah berstatus sudah menikah sebanyak 53,4%. Kemudian diikuti oleh calon pembeli yang berstatus cerai sebanyak 23,5% dan belum atau tidak menikah sebanyak 23,1%, seperti yang ditampilkan pada Gambar 5.



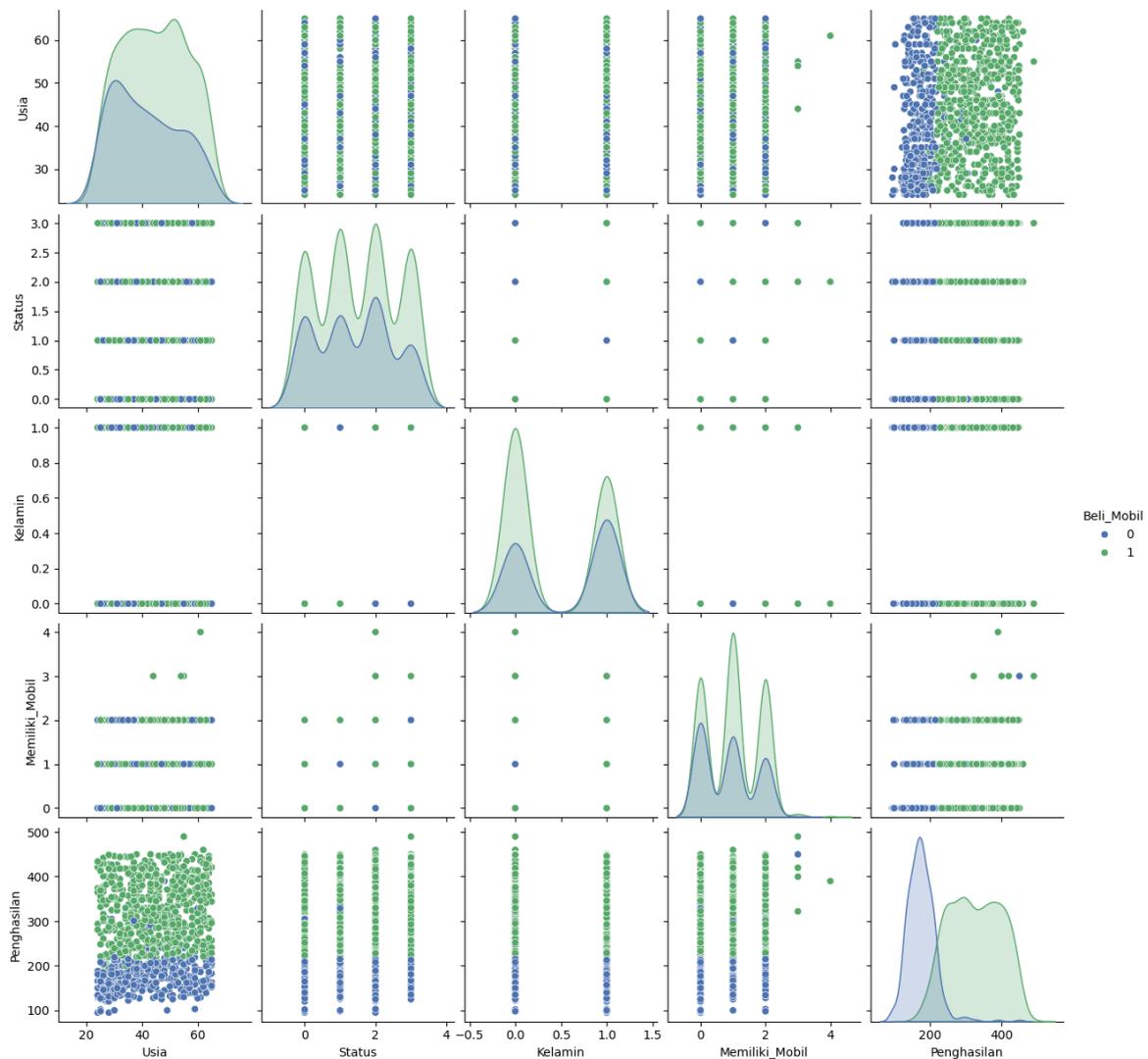
Gambar 5. Hubungan Pembelian Mobil dan Status Ekonomi

Sebanyak 70% dari calon pembeli telah memiliki mobil sebelumnya, dimana dua peringkat terbesar adalah telah memiliki 1 mobil sebanyak 39,6% kemudian disusul yang telah memiliki 2 mobil sebanyak 29,6%. Sementara 30% calon pembeli mobil belum pernah memiliki mobil sebelumnya, seperti yang ditampilkan pada Gambar 6.



Gambar 6. Hubungan Pembelian Mobil dan Kepemilikan Mobil

Hubungan antar variabel disajikan dalam bagan *pairplot* [19] pada Gambar 7 yang menunjukkan bahwa usia dan penghasilan memiliki pengaruh yang paling besar dalam keputusan calon pembeli untuk akhirnya memutuskan membeli mobil.



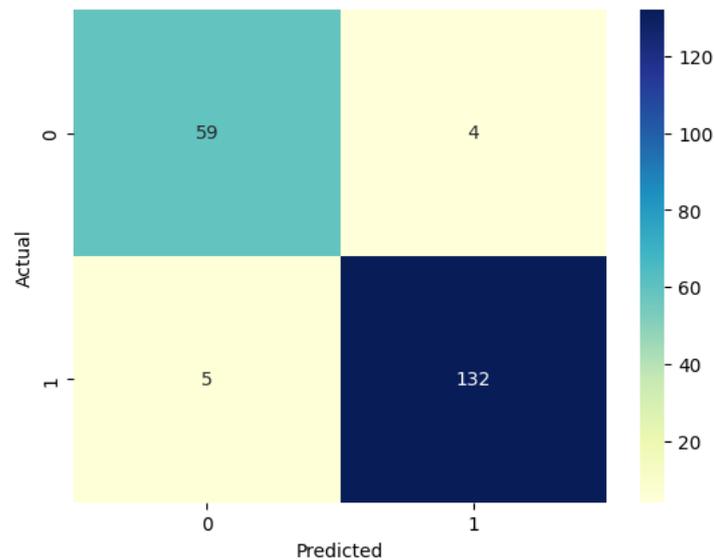
Gambar 7. Hubungan Berpasangan (Pairplot) Antar Variabel

4.3 Pemodelan dan Evaluasi

Proses pelatihan model menggunakan algoritma *Logistic Regression*. Dengan parameter *solver* `lbfgs` yang merujuk pada *Limited-memory Broyden–Fletcher–Goldfarb–Shanno* dimana merupakan algoritma optimasi yang bekerja dengan memperkirakan gradien fungsi tujuan menggunakan sejumlah titik sebelumnya. Parameter *solver* ini sangat cocok untuk dataset yang mengandung data ordinal [20], sehingga pada penelitian ini menggunakan *solver* tersebut.

Gambar 8 merupakan tabel *confusion matrix* yang menampilkan hasil prediksi model dengan penjelasan sebagai berikut [21] :

- True Positive (TP) = 59, merupakan jumlah yang diprediksi membeli mobil dan memang benar membeli mobil.
- True Negative (TN) = 132, merupakan jumlah yang diprediksi tidak membeli mobil dan memang benar tidak membeli mobil.
- False Positive (FP) = 5, merupakan jumlah yang diprediksi membeli mobil namun sebenarnya tidak membeli mobil.
- False Negative (FN) = 4, merupakan jumlah yang diprediksi tidak membeli mobil namun sebenarnya membeli mobil.

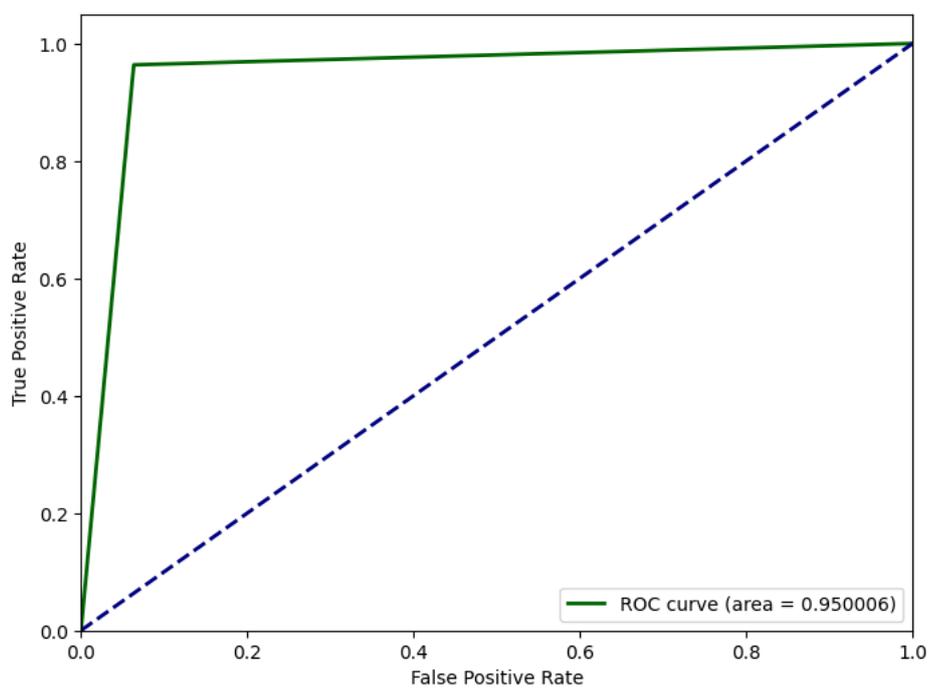


Gambar 8. Confusion Matrix

Tabel 3 menyajikan hasil pengukuran kinerja model, dimana model telah mampu untuk melakukan prediksi dengan benar yang ditunjukkan oleh perolehan skor *accuracy*, *precision*, *recall*, dan *f1-score* sebesar 95%. Hasil ini tervalidasi melalui representasi kurva ROC [22] [23] pada Gambar 9 dimana area kesesuaian prediksi sebesar 95%.

Tabel 3. Hasil Pengukuran Kinerja Model

	Precision	Recall	F1-score	Support
0	0.92	0.94	0.93	63
1	0.97	0.96	0.97	137
accuracy			0.95	200
macro avg	0.95	0.95	0.95	200
weighted avg	0.96	0.95	0.96	200



Gambar 9. Kurva ROC

5 Kesimpulan

Penelitian ini telah berhasil mengembangkan sebuah model prediksi efektif dengan memanfaatkan algoritma regresi logistik, yang secara signifikan dapat mengidentifikasi individu yang berpotensi menjadi pembeli mobil. Dengan mencapai akurasi dan presisi yang tinggi, yakni sebesar 95%, penelitian ini menunjukkan bahwa variabel-variabel seperti pendapatan, usia, status kepemilikan kendaraan sebelumnya, jenis kelamin, dan status pernikahan memegang peranan krusial dalam mempengaruhi keputusan pembelian mobil. Temuan ini memberikan wawasan yang sangat berharga bagi industri otomotif dalam menyusun strategi pemasaran yang lebih efisien dan terfokus, sekaligus memberikan sumbangan penting pada literatur akademik tentang penggunaan regresi logistik dalam analisis perilaku konsumen. Melalui penelitian ini, terbukti bahwa penggunaan teknik analisis data yang canggih seperti regresi logistik tidak hanya memungkinkan perusahaan otomotif untuk mengidentifikasi segmen pasar yang paling mungkin melakukan pembelian tetapi juga membantu dalam merancang kampanye pemasaran yang lebih efektif dan efisien, yang pada akhirnya akan meningkatkan penjualan dan kepuasan pelanggan. Ini menegaskan kembali pentingnya penerapan metodologi analitik dalam memahami dan merespons dinamika pasar untuk mencapai keunggulan signifikan dalam industri yang kompetitif. Penelitian lebih lanjut dapat menerapkan *feature selection* (seleksi fitur) untuk mengidentifikasi variabel yang paling berpengaruh dalam prediksi calon pembeli mobil potensial dengan tetap mempertahankan akurasi dan presisi yang tinggi. Selain itu juga dapat diperluas dengan teknik analisis data untuk menangani kemungkinan *imbalanced data* (data yang tidak seimbang) saat membangun model prediksi.

Referensi

- [1] B. P. Siahaan and E. A. Prasetio, "Understanding Customer Insights Through Big Data: Innovations in Brand Evaluation in the Automotive Industry," *Asian J. Technol. Manag. AJTM*, vol. 15, no. 1, pp. 49–66, 2022, doi: 10.12695/ajtm.2022.15.1.4.
- [2] M. Vanhala, C. Lu, J. Peltonen, S. Sundqvist, J. Nummenmaa, and K. Järvelin, "The usage of large data sets in online consumer behaviour: A bibliometric and computational text-mining-driven analysis of previous research," *J. Bus. Res.*, vol. 106, pp. 46–59, Jan. 2020, doi: 10.1016/j.jbusres.2019.09.009.
- [3] S. Nusinovici *et al.*, "Logistic regression was as good as machine learning for predicting major chronic diseases," *J. Clin. Epidemiol.*, vol. 122, pp. 56–69, Jun. 2020, doi: 10.1016/j.jclinepi.2020.03.002.
- [4] A. D. Nurhalim, "Analisis Faktor-Faktor yang Mempengaruhi Perilaku Konsumen dalam Keputusan Pembelian Mobil Toyota Avanza di Kota Tangerang," *Jambura Econ. Educ. J.*, vol. 5, no. 1, pp. 51–59, Dec. 2022, doi: 10.37479/jeej.v5i1.15263.
- [5] S. N. Untari, S. Djaja, and J. Widodo, "Strategi Pemasaran Mobil Merek Daihatsu pada Dealer Daihatsu Jember," *J. Pendidik. Ekon. J. Ilm. Ilmu Pendidik. Ilmu Ekon. Dan Ilmu Sos.*, vol. 11, no. 2, p. 82, Jan. 2018, doi: 10.19184/jpe.v11i2.6451.
- [6] M. P. LaValley, "Logistic Regression," *Circulation*, vol. 117, no. 18, pp. 2395–2399, May 2008, doi: 10.1161/CIRCULATIONAHA.106.682658.
- [7] F. Harahap, A. Y. N. Harahap, E. Ekadiansyah, R. N. Sari, R. Adawiyah, and C. B. Harahap, "Implementation of Naïve Bayes Classification Method for Predicting Purchase," in *2018 6th International Conference on Cyber and IT Service Management (CITSM)*, Parapat, Indonesia: IEEE, Aug. 2018, pp. 1–5. doi: 10.1109/CITSM.2018.8674324.
- [8] S. P. Shakti, M. K. Hassan, Y. Zhenning, R. D. Caytiles, and I. N.Ch.S.N, "Annual Automobile Sales Prediction Using ARIMA Model," *Int. J. Hybrid Inf. Technol.*, vol. 10, no. 6, pp. 13–22, Jun. 2017, doi: 10.14257/ijhit.2017.10.6.02.
- [9] A. Hasyim, M. Fatchan, and W. Hadikristanto, "Penerapan Algoritma Naïve Bayes Dalam Memprediksi Tingkat Penjualan Mobil Tahun 2022," *J. Ilm. Intech Inf. Technol. J. UMUS*, vol. 4, no. 02, pp. 207–215, Nov. 2022, doi: 10.46772/intech.v4i02.872.
- [10] S. A. Sinaga, "Implementasi Metode Arima (Autoregressive Moving Average) Untuk Prediksi Penjualan Mobil," *J. Glob. Technol. Comput.*, vol. 2, no. 3, pp. 102–109, Aug. 2023, doi: 10.47065/jogtc.v2i3.4013.

- [11] F. G. Walelang, C. E. J. C. Montolalu, and D. Hatidja, "Penerapan Metode Autoregressive Integrated Moving Average (ARIMA) dalam Model Intervensi Fungsi Step untuk Memprediksi Penjualan Mobil di PT. Hasjrat Abadi Cabang Tendeand Manado." *Indonesian Journal of Intelligence Data Science*, May 15, 2023.
- [12] I. Hadjar, "Regresi Logistik: Menaksir Probabilitas Peristiwa Variabel Binari," *Phenom. J. Pendidik. MIPA*, vol. 7, no. 2, pp. 137–163, Jan. 2018, doi: 10.21580/phen.2017.7.2.1385.
- [13] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied logistic regression*, Third edition. in Wiley series in probability and statistics. Hoboken, NJ: Wiley, 2013.
- [14] D. G. Kleinbaum and M. Klein, *Logistic Regression*. in *Statistics for Biology and Health*. New York, NY: Springer New York, 2010. doi: 10.1007/978-1-4419-1742-3.
- [15] M. Nourelahi, A. Zamani, A. Talei, and S. Tahmasebi, "A Model to Predict Breast Cancer Survivability Using Logistic Regression".
- [16] Z. Khemais, D. Nesrine, and M. Mohamed, "Credit Scoring and Default Risk Prediction: A Comparative Study between Discriminant Analysis & Logistic Regression," *Int. J. Econ. Finance*, vol. 8, no. 4, p. 39, Mar. 2016, doi: 10.5539/ijef.v8n4p39.
- [17] G. M. Allenby and P. J. Lenk, "Modeling Household Purchase Behavior with Logistic Normal Regression," *J. Am. Stat. Assoc.*, vol. 89, no. 428, pp. 1218–1231, Dec. 1994, doi: 10.1080/01621459.1994.10476863.
- [18] C. Mood, "Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It," *Eur. Sociol. Rev.*, vol. 26, no. 1, pp. 67–82, Feb. 2010, doi: 10.1093/esr/jcp006.
- [19] G. Rajagopalan, *A Python Data Analyst's Toolkit: Learn Python and Python-based Libraries with Applications in Data Analysis and Statistics*. New York: Apress, 2021.
- [20] D. R. S. Saputro and P. Widyaningsih, "Limited memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) method for the parameter estimation on geographically weighted ordinal logistic regression model (GWOLR)," Yogyakarta, Indonesia, 2017, p. 040009. doi: 10.1063/1.4995124.
- [21] J. T. Townsend, "Theoretical analysis of an alphabetic confusion matrix," *Percept. Psychophys.*, vol. 9, no. 1, pp. 40–50, Jan. 1971, doi: 10.3758/BF03213026.
- [22] N. A. Obuchowski, "ROC Analysis," *Am. J. Roentgenol.*, vol. 184, no. 2, pp. 364–372, Feb. 2005, doi: 10.2214/ajr.184.2.01840364.
- [23] D. Phung, G. I. Webb, and C. Sammut, *Encyclopedia of Machine Learning and Data Science*. New York, NY: Springer US : Imprint: Springer, 2020.