

KLASIFIKASI POLA PEMBELIAN KENDARAAN BERMOTOR UNTUK MERANCANG STRATEGI PROMOSI TERARAH MENGUNAKAN ALGORITMA LOGISTIC REGRESSION

Ryan Arya Pramudya¹, Atika Mutiarachim², Puji Setya Sunarka³
^{1,2,3}Universitas 17 Agustus 1945 Semarang, Jl. Pawiyatan Luhur I, Semarang, Indonesia
Email: ryanarya-pramudya@untagsmg.ac.id

Article History

Received: 09-11-2024

Revision: 11-11-2024

Accepted: 11-11-2024

Published: 11-11-2024

Abstract.

This research aims to determine the effect of Logistic Regression on the classification of motor vehicle consumers in 2021. Consumer classification is carried out to identify specific consumer criteria, so that targeted promotions can be directed to the right consumers. Data was obtained from several motorcycle dealerships in the Central Java area with various motorcycle brands, within the period from January to December 2024. The data has binomial labels consisting of 'cash' for cash purchases and 'credit' for credit purchases. Preprocessing is carried out by standardizing the dataset attributes from each dealer, attribute selection to choose the most appropriate and significantly influential attributes, downsampling to ensure symmetrical data, and data selection and transformation to ensure the processed data is consistent. Classification is performed using the Logistic Regression algorithm. Data is divided using the 70:30 split method. The performance results show an accuracy of 100%, a classification error of 0%, a kappa value of 1, and an AUC value of 1.00 or 100%, which are interpreted as very good, indicating that the Logistic Regression algorithm model is very accurate and suitable for use on this dataset.

Keywords: *Classification, logistic regression, motorcycle selling, performance*

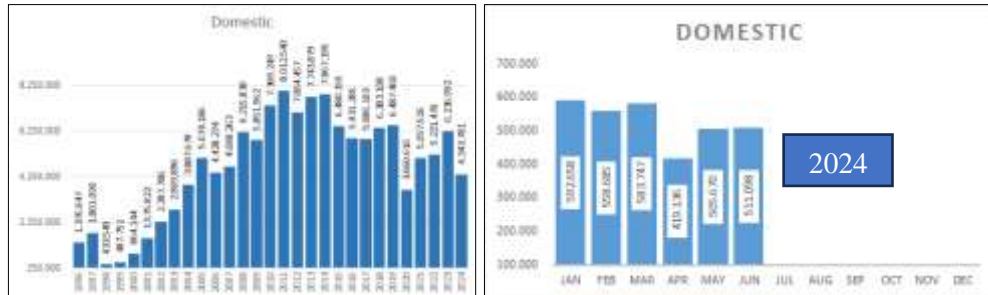
Abstrak. Penelitian ini bertujuan mengetahui pengaruh metode Logistic Regression pada klasifikasi kosumen kendaraan bermotor tahun 2021. Klasifikasi konsumen dilakukan untuk mengetahui kriteria konsumen secara spesifik, sehingga dapat melakukan promosi terarah langsung kepada konsumen yang tepat. Data diperoleh dari beberapa *dealer* penjualan motor di area Jawa Tengah dengan berbagai merk motor, pada rentang waktu Januari sampai Desember 2024. Data memiliki label binomial yang terdiri dari 'cash' untuk pembelian tunai dan 'credit' untuk pembelian kredit. Tahap *preprocessing* data dilakukan dengan menyamakan atribut dataset dari masing-masing *dealer*, *downsampling* untuk membuat jumlah data lebih simetris, kemudian seleksi data dan transformasi data untuk memastikan data yang digunakan nilainya konsisten, seleksi atribut untuk memilih atribut yang paling tepat digunakan dan memberi pengaruh signifikan. Klasifikasi dilakukan dengan algoritma Logistic Regression. Data dibagi menggunakan metode *split* dengan 70% data *training* dan 30% data *testing*, Hasil *performance* akurasi 100% nilai *classification error* 0%, nilai kappa 1, nilai AUC 1.00 atau 100% di interpretasikan sangat baik, menunjukkan model algoritma Logistic Rgression sangat akurat dan tepat digunakan pada dataset ini.

Kata Kunci: *Klasifikasi, logistic regression, penjualan motor, performance*

How to Cite: Pramudya, A. et. al. (2024). Klasifikasi Pola Pembelian Kendaraan Bermotor Untuk Merancang Strategi Promosi Terarah Menggunakan Algoritma Logistic Regression. *Indo-Fintech Intellectuals: Journal of Economics and Business*, 4 (5), 2549-2557. [10.54373/ifiijeb.v4i5.2121](https://doi.org/10.54373/ifiijeb.v4i5.2121)

PENDAHULUAN

Motor merupakan alat transportasi utama masyarakat Indonesia. Peningkatan penjualan sepeda motor tidak hanya terjadi setiap tahun, bahkan setiap bulan. Penjualan sepeda motor di Indonesia pada bulan Juni 2024 mencapai 511.098 unit, naik dari bulan-bulan sebelumnya sepanjang tahun 2024.



Gambar 1. Penjualan Sepeda Motor di Indonesia (*Indonesia, 2024*)

Motor memiliki harga yang lebih terjangkau dibandingkan mobil, mengendarai sepeda motor dinilai praktis dan lebih cepat. Tingginya persaingan industri motor membuat produsen bersaing dalam berbagai hal untuk meningkatkan penjualan, mulai dari desain yang lebih bagus, canggih, nyaman, sampai dengan bagaimana cara mengoptimalkan penjualan kepada konsumen. Berbagai metode promosi dilakukan guna meningkatkan penjualan produk kendaraan. Melakukan klasifikasi konsumen sangat perlu dilakukan untuk mengetahui kriteria konsumen yang tepat secara spesifik, sehingga dapat melakukan promosi terarah langsung kepada konsumen yang tepat.

Penelitian (Astuti, 2016) menggunakan metode C45, KNN dan Neural Network untuk menentukan kelayakan konsumen memperoleh kredit kendaraan bermotor. Populasi penelitian merupakan konsumen Bess Finance tahun 2012/2013, data sebanyak 486 dengan 14 atribut prediktor dan 1 atribut kelas Good 309 data dan Bad 177 data. Hasil akurasi terbaik diperoleh C45 92,89%. Atribut yang digunakan pada penelitian beserta pembobotannya adalah, bobot tertinggi yaitu 1 untuk atribut Jumlah Tanggungan, Lama Tinggal, Masa Kerja, Penghasilan Perbulan dan Pembayaran Pertama. Bobot 0.8 pada atribut Kepemilikan Rumah, Jenis Pekerjaan, Status Kepegawaian, Bobot 0.6 pada Kondisi Rumah. Bobot 0.5 pada atribut Status Perkawinan, Pendidikan Terakhir, Usia, dan Status Perusahaan.

Penelitian (Fernando et al, 2023) klasifikasi 1620 data kreditur motor Honda dari Showroom Astra Motor Sako Palembang menggunakan metode Naive Bayes, dengan atribut kelas Layak 1399 data dan Tidak Layak 221 data, atribut prediktor yang digunakan nama, kecamatan, kelurahan, jumlah tenor, pengeluaran per bulan, jenis kelamin, pendidikan terakhir,

pemakai kendaraan, dp awal, dan cicilan. Uji klasifikasi dilakukan lima kali dengan split data 90:10, 80:20, 70:30, 60:40, 50:50, akurasi tertinggi pada split data 90:10 yaitu 74,83%.

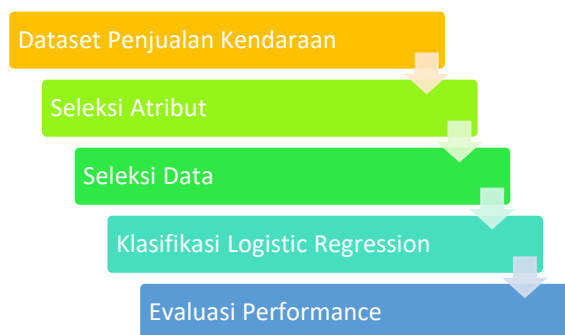
Penelitian (Sriani et al, 2024) klasifikasi Naive Bayes 89 data nasabah tahun 2021-2022 PT Pegadaian Area Medan 1, dengan atribut kelas Layak dan Tidak Layak. Atribut prediktor terdiri dari Nama, Umur, Jenis Kelamin, Karakter, Pekerjaan, Pendapatan, Tanggungan, Pengeluaran, Rumah, Pinjaman. Data dibagi menjadi 84 training dan 21 testing, menghasilkan akurasi sebesar 100%.

Penelitian (Mutiarachim & Tyoso, 2024) menggunakan klasifikasi Logistic Regression pada dataset Bank Marketing untuk menemukan pola deposito. Klasifikasi pada 9.999 data nasabah, akurasi yang dihasilkan sebesar 88,53%, AUC 93,4% dan error value yang rendah yaitu 11,4%. Menemukan pola yang tepat, membantu pihak bank mengoptimalkan promosi terarah pada nasabah yang tepat, sehingga kemungkinan untuk closing produk deposito lebih besar.

Memahami pola pembelian yang tepat, sangat penting untuk mendukung promosi yang lebih terarah, dalam penelitian ini pola membantu dealer memahami apakah konsumen akan membeli produk secara cash atau credit. Promosi hanya pada konsumen spesifik dengan pola tertentu, yang paling tepat dan akurat, meningkatkan kemungkinan produk terjual dengan pasti, sehingga lebih efisien dan efektif. Hal tersebut sekaligus meminimalisir kebutuhan tenaga dan biaya jika dibandingkan dengan promosi secara global.

METODE

Penelitian ini merupakan penelitian kuantitatif yang dilakukan dengan mengolah dataset agar konsisten, memilih atribut yang tepat, dan kemudian memprosesnya menggunakan metode data mining dengan algoritma klasifikasi Logistic Regression. Proses preprocessing terdiri dari seleksi atribut dan seleksi data. Proses penelitian digambarkan pada alur penelitian..



Gambar 2. Alur Penelitian

Penelitian ini menggunakan dataset privat penjualan kendaraan bermotor, dengan jumlah total 9.952 data, terbagi menjadi 6036 *cash* dan 3916 *credit*. Total atribut predictor sebanyak 102, dan 1 atribut kelas. Tahap *preprocessing* bertujuan untuk mengolah data agar lebih lengkap dan konsisten. Preprocessing dilakukan dengan seleksi atribut untuk menghilangkan atribut redundan dan atribut-atribut yang tidak perlu, seratan seleksi data untuk memastikan dataset yang digunakan konsisten dan tepat.

Seleksi Atribut

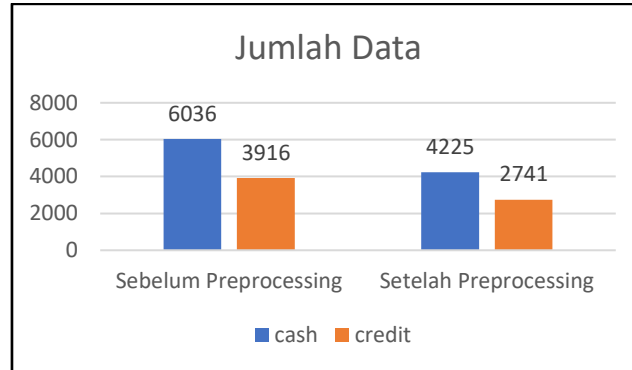
Seleksi atribut dilakukan dengan memilih atribut yang tepat untuk memaksimalkan akurasi dan mengoptimalkan kinerja model klasifikasi. Terdapat 103 atribut dengan banyak atribut *redundan*. Seleksi dilakukan dengan menghapus atribut *redundan*, memilih atribut berdasarkan studi literature pada penelitian yang telah dilakukan mengenai klasifikasi penjualan kendaraan dan dengan melakukan weighting pada atribut, sehingga dihasilkan 19 atribut terpilih dan 1 *class* terdiri dari label *cash* didefinisikan dengan “1” dan *credit* didefinisikan dengan “2”. Atribut yang banyak mengandung *missing value* dihapus, anatra lain atribut Flag dan atribut Facebook, Twitter, Instragam, dan lainnya. memiliki hampir 90% data yang kosong, tidak memiliki nilai, sehingga dihapus.

Tabel 1. Atribut Terpilih

No	Atribut	Tipe Data	Keterangan
1	Kelurahan	Polynominal	Alamat kelurahan konsumen
2	Kecamatan	Polynominal	Alamat kecamatan konsumen
3	Kode Dealer	Polynominal	Kode dealer
4	Jenis Sales	Polynominal	Jenis sales
5	Jenis Kelamin	Binominal	Jenis kelamin konsumen
6	Agama	Polynominal	Agama konsumen
7	Pekerjaan	Polynominal	Pekerjaan konsumen
8	Pengeluaran	Integer	Pengeluaran konsumen
9	Pendidikan	Polynominal	Pendidikan terakhir konsumen
10	diHubungi	Binominal	Apakah konsumen di hubungi? Ya-Tidak
11	Merk	Polynominal	Merk sepeda motor
12	Jenis	Polynominal	Jenis sepeda motor
13	Fungsi	Polynominal	Utilitas / kepentingan konsumen sebagai pengguna kendaraan
14	Pemakai	Polynominal	Segmentasi pengguna kendaraan
15	Umur	Integer	Umur konsumen
16	Tipe	Polynominal	Tipe sepeda motor
17	DP Aktual	Integer	DP Aktual dari konsumen
18	Warna	Polynominal	Warna kendaraan
19	Hobi	Polynominal	Hobi konsumen
20	Cash/Credit	Binominal	Kelas/Label apakah pembeli melakukan pembayaran secara tunai/ <i>cash</i> atau kredit/ <i>credit</i> Cash = 1 , Credit = 2

Seleksi Data

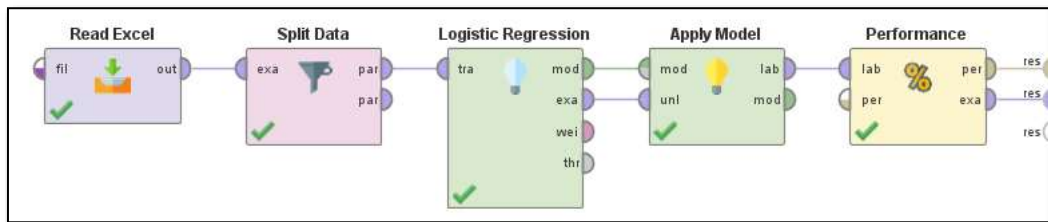
Seleksi dilakukan dengan *cleansing data* yaitu menghapus data dengan nilai negatif. Menghilangkan data yang mengandung *missing value*. Hasil seleksi dari 9.952 data dengan 6036 *cash* dan 3916 *credit*, menjadi 6966 data dengan 4225 *cash* dan 2741 *credit*.



Gambar 3. Jumlah Data Sebelum dan Setelah *Preprocessing*

Klasifikasi Logistic Regression

Penelitian menggunakan *tool* Altair Rapid Miner. Penilaian klasifikasi Logistic Regression dinilai dari *performance* yang akan dihasilkan. *Performance* terdiri dari *accuracy*, *classification error*, *kappa*, *absolute error*, *RMSE*, *correlation*, *AUC*, *precision*, *recall*, *F measure*, *senitivity* dan *specificity*. Data dibagi dengan metode *split* 70% *training* dan 30% *testing*.



Gambar 4. Proses pada Aplikasi RapidMiner

HASIL

Hasil klasifikasi berupa *performance* disampaikan pada tabel Hasil. *Confussion matrix* merupakan matrik dua dimensi yang menggambarkan perbandingan hasil prediksi dengan kenyataan. Terdapat empat kondisi dalam *confussion matrix* yaitu *true positive* (TP), *true negative* (TN), *false positive* (FP) dan *false negative* (FN).

Tabel 2. *Confussion Matrix*

	true 1	true 2	class precision
pred 1	TP 4225	FP 0	100%
pred 2	FN 0	TN 2741	100%
class recall	100%	100%	

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{4225+2741}{4225+2741+0+0} = 1 = 100\%$$

$$Precision = \frac{TP}{TP+FP} = \frac{4225}{4225+0} = 1 = 100\%$$

$$Recall \ \& \ Sensitivity = \frac{TP}{TP+FN} = \frac{4225}{4225+0} = 1 = 100\%$$

$$Specificity = \frac{TN}{TN+FP} = \frac{2741}{2741+0} = 1 = 100\%$$

Accuracy merupakan tingkat kedekatan antara nilai yang dapat terhadap nilai sebenarnya. *Precision* mengukur ketepatan data yang relevan. *Recall* mengukur proporsi positif sebenarnya yang diklasifikasi dengan benar. Nilai *recall* sama dengan nilai *sensitivity*. Nilai *recall* tinggi berarti *false positive* lebih banyak terjadi dibandingkan *false negative*. Nilai *precision* tinggi berarti *true positive* lebih banyak dibandingkan terjadinya *false negative* (Martin & Powers, 2015)(Setiawan, 2020)(Setiawan, 2020). Hasil perhitungan menunjukkan *precision* lebih tinggi dibandingkan *recall*, hal ini berarti baik karena penelitian bertujuan untuk mengoptimalkan hasil klasifikasi data berlabel *yes*, yang berarti nasabah bersedia berlangganan deposito. *Specificity* merupakan kebenaran memprediksi negatif dibandingkan dengan keseluruhan data negatif.

$$F1\text{-Measure} = 2 \times \frac{Recall \times Precision}{Recall + Precision} = 2 \times \frac{100\% \times 100\%}{100\% + 100\%} = 100\%$$

F-Measure, atau skor F, atau F1-score, adalah perbandingan rata-rata dari ketepatan dan recall. Nilai F-Measure seratus persen menunjukkan bahwa algoritma regresi logistik melakukan prediksi klasifikasi dengan baik jika jumlah data *false negative* dan *false positive* dalam dataset mendekati atau simetris, sebaliknya jika jumlah data tidak simetris, nilai akurasi sangat baik digunakan sebagai acuan kinerja algoritma (Christen et al., 2023).

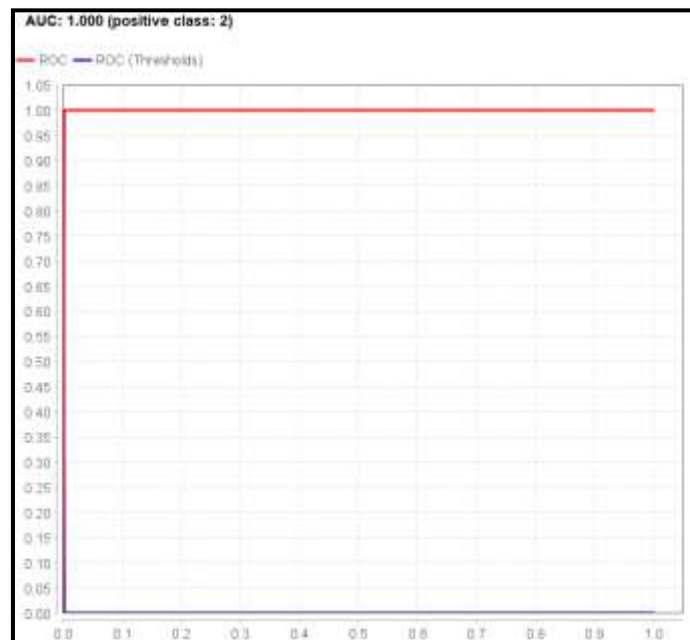
DISKUSI

Hasil menunjukkan bahwa nilai akurasi yang dihasilkan adalah 100%. *Classification error* merupakan performa kesalahan dari algoritma klasifikasi yang digunakan (Silangen & Matdoan, 2022). *Precision, Recall, Sensitivity, Specificity, F-Measure* 100%.

Tabel 3. *Performance* Klasifikasi

<i>Criterion</i>	<i>Performance</i>
Accuracy	100%
Precision	100%
Recall	100%
F-Measure	100%
Sensitivity	100%
Specificity	100%
Classification Error	0%
Kappa	1.00
Youden	1.00
AUC	1.00

Nilai kappa yang dihasilkan adalah 1 sehingga model dikatakan sempurna. Kappa merupakan normalisasi dari *accuracy*. *Classification error* 0% artinya tidak terjadi kesalahan. Nilai AUC yang dihasilkan 1.00, jika di prosentase, nilai AUC 100%, menunjukkan model klasifikasi semakin baik dan tepat digunakan pada data tersebut.



Gambar 5. Grafik AUC

KESIMPULAN

Performance klasifikasi menghasilkan nilai yang baik, akurasi 100%, *classification error* 0%, nilai kappa 1, nilai AUC 1 artinya 100%, artinya baik dari segala sisi. Hasil menunjukkan bahwa algoritma Logistic Regression tepat digunakan pada dataset ini, karena menghasilkan nilai *performance* yang baik.

REKOMENDASI

Penelitian menggunakan dataset dari berbagai *dealer* sehingga memerlukan waktu untuk melakukan *preprocessing* penyesuaian atribut. Jumlah data yang diperoleh terbatas, dan skala wilayah hanya Jawa Tengah. Pemilihan *dealer* dilakukan secara acak. Penelitian selanjutnya bisa dilakukan menambah jumlah data juga dapat memberi pengaruh terhadap hasil klasifikasi.

REFERENSI

- Astuti, P. (2016). Komparasi Penerapan Algoritma C45 KNN dan Neural Network dalam Proses Kelayakan Penerimaan Kredit Kendaraan Bermotor. *Faktor Exacta*, 87-101.
- Fernando, S. D., Purnamasari, S. D., Syaputra, H., & Oktarini Saputri, N. A. (2023). Prediksi Kredit Sepeda Motor pada Showroom Astra Motor Sako Palembang dengan Metode Naive Bayes. *Journal of Computer and Information Technology*, 17-23.
- Hill, R. C., Griffiths, W. E., & Lim, G. C. (2018). *Principles of Econometrics 5th Edition*. Wiley.
- Indonesia, A. I. (2024). *AISI / Asosiasi Industri Sepeda Motor Indonesia*. Retrieved from Asosiasi Industri Sepeda Motor Indonesia: <https://www.aisi.or.id/>
- Kumar, K. N. (2020). *Econometrics 1st Edition*. USA: CRS Press LLC.
- Kusrini, & Luthfi, E. T. (2009). *Algoritma Data Mining*. Yogyakarta: Andi Offset.
- Meirisah, F., & Sutabri, T. (2023). Analisa Kasus Kejahatan Siber dengan Menggunakan Visualisasi Data. *JINTEKS (Jurnal Informatika Teknologi dan Sains)*, 32-37.
- Mutiarachim, A., & Tyoso, J. S. (2024). Optimasi Prediksi Pemasaran Nasabah Deposito Bank dengan Metode Klasifikasi. *Jurnal Cakrawala Informasi*, 20-28.
- North, M. A. (2012). *Data Mining for The Masses*.
- Sarosa, S. (2023). *Eksplorasi dan Analisis Data Bisnis*. Daerah Istimewa Yogyakarta: Penerbit PT Kanisius.
- Sriani, Nasution, R. H., & Humaira, S. (2024). Implementasi Algoritma Naive Bayes untuk Memprediksi Kelayakan Pemberian Kredit Sepeda Motor. *Journal of Science and Social Research*, 546 – 554.
- Stock, J. H., & Watson, M. W. (2020). *Introduction to Econometrics 4th Edition*. Pearson.
- Wooldridge, J. M. (2020). *Introductory Econometrics A Modern Approach 7th Edition*. Boston: Cengage.