

PERBANDINGAN KINERJA ALGORITMA KNN, DECISION TREE, SVM, DAN ANN UNTUK MELAKUKAN KLASIFIKASI PADA CALON PEMBELI MOBIL

Shinzi

Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara,
Jln. Letjen S. Parman No. 1, Jakarta, 11440, Indonesia
E-mail: shinzi.535220118@stu.untar.ac.id,

ABSTRAK

Pelaku bisnis jual beli mobil dihadapkan pada tantangan pasar otomotif yang semakin kompetitif, terutama dalam mengidentifikasi calon pembeli yang tepat. Untuk mengatasi masalah tersebut, penelitian ini dilakukan dengan memanfaatkan machine learning dengan menerapkan algoritma klasifikasi, berupa KNN, *Decision Tree*, SVM, dan ANN. Dalam upaya untuk mengevaluasi keefektifan masing-masing algoritma, penelitian ini melakukan perbandingan kinerja di antar algoritma. Hasil penelitian menunjukkan bahwa algoritma *Decision Tree* muncul sebagai pilihan yang paling tepat dalam proses klasifikasi pembeli potensial pada data set calon pembeli mobil. Temuan ini memberikan dasar yang kuat bagi pelaku bisnis untuk meningkatkan strategi pemasaran dan penjualan dalam bisnis jual beli mobil, serta memberikan pandangan yang lebih dalam tentang karakteristik pembeli potensial yang relevan.

Kata kunci—Klasifikasi, KNN, *Decision Tree*, SVM, calon pembeli mobil

ABSTRACT

Business operator engaged in buying and selling cars are faced within creasingly competitive challenges in the automotive market, especially in identifying the right prospective buyers. to address this issue, this research isutilizing machine learning by implementing classification algorithms, namely KNN, Decision Tree, SVM, and ANN. In aneffortto evaluate the effective nesso feach algorithm, this research compares their performance with achother. This research indicates that the Decision Tree algorithm emerges as the mostsuitable choice in the process of classifying potential buyers in the dataset of prospective car buyers. The sefindings provide a strong basis for business operators to enhancemarketing and sales strategies in the car trading business, as well as gain deeperinsights into there levant characteristics of potential buyers.

Keywords—Classification, KNN, *Decision Tree*, SVM, Prospectivebuyers

1. PENDAHULUAN

Teknologi adalah salah satu faktor pendorong utama dalam perubahankultur bisnis modern, terutama dalam pengambilan keputusan bisnis [1]. Dorongan pengambilan keputusan bisnis yang optimal dan terbaik menjadikan analisis data dengan *machinelearning* sebagai landasan yang penting dalam pengambilan keputusan bisnis[2]. Analisis data memungkinkan pelaku bisnis untuk dapat mengidentifikasi peluang bisnis baru, mengelola risiko, dan tentunya meningkatkan kinerja bisnis secara keseluruhan [3] [4], Salah satu aplikasi penting dari analisis data adalah prediksi calon pembeli atau pelanggan baru dari data historis suatu model bisnis, yang memungkinkan pelaku bisnis untuk dapat melakukan *targetedadvertising*, yang dapat memberikan keuntungan dalam bentuk efisiensi dan efektivitas dalam upaya pemasaran produk [5].

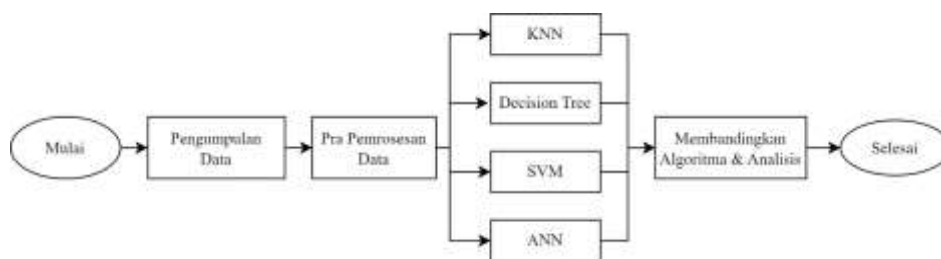
Dalam konteks industri otomotif, dalam hal ini calon pembeli mobil, penggunaan teknologi dan analisis data menjadi semakin penting dalam menangkap pasar penjualan mobil yang semakin kompetitif[6]. Oleh karena itu, tempat jual beli mobil, seperti *dealer* mobil atau pameran otomotif, menjadi titik fokus dalam upaya peningkatan penjualan dan strategi pemasaran dalam penelitian

ini. Para pengelola tempat jual beli mobil menghadapi tantangan dalam mengidentifikasi dan membedakan calon pembeli diantarapengunjung yang hanya melihat-lihat. Penggunaan *machinelearning* dengan implementasi algoritma klasifikasi untuk menganalisis data dari pengunjung tempat jual beli mobil, dapat menjadi solusi yang efektif dalam mengatasi tantangan ini, karena menghasilkan analisis, estimasi, serta prediksi mengenai calon pembelimobil [7].

Penelitian dilakukan denganmembandingkan kinerja berbagai algoritma klasifikasi dalam memprediksi calon pembeli dari pengunjung tempat jual beli mobil, hasil dari penelitian ini dapat memberikan wawasan berharga bagi pelaku usaha jual beli mobil. Dengan pemahaman yang lebih baik tentang karakteristik pembeli potensial, pelaku bisnis jual beli dapat meningkatkan strategi pemasaran dan penjualan bisnis, meningkatkan konversi pengunjung menjadi pembeli, dan pada akhirnya, meningkatkan kinerja bisnis secara keseluruhan.

2. METODE PENELITIAN

2.1 Diagram Alir



Gambar 1 Diagram Alir

Alur jalannya penelitian dapat dilihat pada Gambar 1. Penelitian dimulai dengan pengumpulan data, data yang dikumpulkan merupakan data historis calon pembeli mobil, langkah selanjutnya adalah pra pemrosesan data, dimana data yang didapatkan dibersihkan dengan langkah-langkah pembersihan data pada umumnya yaitu, perbaikan data yang tidak biasa (*outliers*) dan penanganan data yang tidak seimbang, langkah-langkah lain seperti penanganan data yang hilang (*missing values*) tidak dilakukan pada data set, dikarenakan data set yang didapatkan tidak memiliki *missing values* [8]. Selanjutnya adalah tahapan penggunaan algoritma *KNN*, *Decision Tree*, *SVM*, dan *ANN* untuk melakukan analisis dan klasifikasi pada data calon pembeli mobil. Tahapan terakhir adalah membandingkan kinerja dari masing-masing algoritma untuk menentukan algoritma klasifikasi terbaik untuk data set yang didapatkan.

2.2 Algoritma

Klasifikasi dilakukan menggunakan empat algoritma yang berbeda, yaitu *K-NearestNeighbors*, *Decision Tree*, *SupportVectorMachine*, dan *Artificial Neural Network*. Setiap algoritma akan dilatih sebanyak 15 kali dengan tiga komposisi data latih dan data uji, yaitu 70% data latih dan 30% data uji, 80% data latih dan 30% data uji, serta 90% data latih dan 10% data uji. Keempat algoritma tersebut akan dibandingkan, Tujuan dari perbandingan sendiri adalah untuk menentukan algoritma klasifikasi mana yang paling sesuai dan efektif untuk data set yang mungkin memiliki kompleksitas yang cukup tinggi. Selain itu, perbandingan ini juga bertujuan untuk menemukan pola tersembunyi yang dapat memberikan manfaat tambahan [9][10].

2.2.1 *K-NearestNeighbors (KNN)*

KNN (K-NearestNeighbors) merupakan algoritma *machinelearning* yang dapat digunakan untuk klasifikasi ataupun regresi [11]. Algoritma ini melakukan prediksi kelas berdasarkan kemiripan fitur data set uji dengan tetangga data set, kemiripan fitur data set uji dicari menggunakan *euclidean distance*, rumus untuk mencari *euclidean distance* dapat dilihat pada rumus

5 [11]. Pada penelitian ini analisis dengan algoritma KNN dengan mempertimbangkan dua skenario berbeda, yaitu menggunakan 5 tetangga dan 11 tetangga.

$$d = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \quad (5)$$

Dimana :

d : *euclidean distance*

a_i : Fitur ke- i pada dataset uji

b_i : Fitur ke- i pada dataset tetangga

Euclidean distance dihitung sesuai dengan jumlah tetangga yang telah ditentukan dalam model *machine learning*. Tetangga yang memiliki kedekatan terbesar dengan data uji akan dipilih, kemudian data uji akan diklasifikasikan berdasarkan mayoritas kelas dari tetangga tersebut.

2.2.2 Decision Tree

Decision Tree merupakan algoritma umum yang dapat digunakan untuk melakukan klasifikasi. *Decision Tree* bekerja dengan cara membuat pohon keputusan dari kelas fitur yang diambil dari data set. Proses pembentukan pohon keputusan ini dilakukan dengan memilih atribut terbaik yang memisahkan data menjadi sub set yang paling *homogen*, sehingga memungkinkan untuk menentukan prediksi kelas dari data baru dengan mengikuti jalur pohon yang sesuai [12].

2.2.3 Support Vector Machine

Support Vector Machine (SVM) merupakan salah satu algoritma *machine learning* yang efektif dan populer untuk tugas klasifikasi dan regresi [13][14], SVM bekerja dengan cara membuat garis pemisah (*hyperplane*) terbaik yang memisahkan dua kelas data dengan *margin* maksimum [13][14]. Garis pemisah ini berfungsi sebagai pembatas antara data set. SVM mencari garis pemisah tersebut dengan memaksimalkan *margin*, yaitu jarak antara garis pemisah dan titik-titik data terdekat dari kedua kelas [13].

SVM juga memiliki kemampuan untuk menangani data yang tidak linier dengan memanfaatkan fungsi transformasi yang disebut *kernel* [15]. *Kernel SVM* mengubah ruang fitur asli menjadi ruang fitur yang lebih kompleks, di mana data menjadi dapat dipisahkan secara linear. Beberapa jenis *kernel* yang umum digunakan dalam SVM berupa *kernel linear*, *radial basis function* (RBF), *polynomial*, dan *sigmoid* [15]. Pada penelitian ini, *kernel* yang digunakan untuk dibandingkan kinerjanya adalah *kernel linear*, *RBF*, dan *polynomial*.

2.2.4 Artificial Neural Network

Artificial Neural Network (ANN) merupakan algoritma analisis data yang terinspirasi dari jaringan saraf manusia [16], ANN bekerja dengan cara mengolah data dengan serangkaian *neuron* [17]. Setiap *neuron* menerima input dari *neuron* sebelumnya lalu menghasilkan *output* dengan menggunakan fungsi aktivasi. Pada saat masa pelatihan, fitur dari data yang masuk akan diproses dan diingat untuk dapat membedakan satu kelas dengan yang lain [17]. ANN dapat melakukan klasifikasi pada data set dengan memanfaatkan proses pembelajaran yang dilakukan selama masa pelatihan [16][17].

2.3 Metode Evaluasi

Kinerja dari masing-masing algoritma pada metode penelitian akan dievaluasi menggunakan metrik akurasi yaitu, presisi, *recall*, dan F1-score. Metrik-metrik tersebut memberikan gambaran komprehensif tentang kinerja model klasifikasi secara keseluruhan [18].

2.3.1 Akurasi

Akurasi adalah ukuran yang menggambarkan sejauh mana model klasifikasi dapat memprediksi kelas dengan benar, dibandingkan dengan total jumlah sampel yang dievaluasi [19]. Rentang nilai akurasi selalu antara 0 hingga 1, di mana nilai yang mendekati 0 menunjukkan bahwa prediksi yang tepat hampir tidak terjadi, sedangkan nilai mendekati 1 menunjukkan bahwa model memiliki tingkat keakuratan yang tinggi, di mana hampir semua prediksi kelasnya benar [20]. Dengan kata lain, semakin tinggi nilai akurasi, semakin dapat diandalkan model klasifikasi dalam memprediksi kelas dengan benar, sehingga menunjukkan keakuratan *model* klasifikasi [20]. Secara matematis akurasi dapat dituliskan seperti pada rumus 1.

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

Dimana:

ACC : Akurasi

TP : *TruePositive*

TN : *TrueNegative*

FP : *FalsePositive*

FN : *FalseNegative*

TruePositive adalah hasil prediksi model yang berhasil mengidentifikasi kelas positif pada kelas yang benar positif, sementara *TrueNegative* adalah hasil prediksi yang mengonfirmasi kelas negatif dengan benar pada kelas yang benar negatif. Di sisi lain, *FalsePositive* terjadi ketika model salah mengklasifikasikan kelas positif pada kelas yang sebenarnya negatif, sedangkan *FalseNegative* terjadi ketika model salah mengklasifikasikan kelas negatif pada kelas yang sebenarnya positif [21].

2.3.2 Presisi dan Recall

Presisi dan *recall*, Presisi adalah nilai perbandingan antara prediksi kelas positif yang benar-benar positif dibandingkan dengan jumlah keseluruhan prediksi positif yang dilakukan, tanpa memperhatikan berapa banyak prediksi kelas negatif yang salah diklasifikasikan sebagai kelas positif [22], secara matematis presisi dapat dituliskan seperti rumus 2.

$$PREC = \frac{TP}{TP + FP} \quad (2)$$

Dimana:

PREC : Presisi

TP : *TruePositive*

FP : *FalsePositive*

Sedangkan, *recall* adalah nilai perbandingan antara prediksi kelas positif yang benar-benar positif dibandingkan dengan jumlah keseluruhan instansi yang sebenarnya positif dalam data set [22], secara matematis *recall* dapat dituliskan seperti rumus 3.

$$REC = \frac{TP}{TP + FN} \quad (3)$$

Dimana:

REC : *Recall*

TP : *TruePositive*

FP : *FalsePositive*

2.3.3 F1-Score

F1-score adalah metrik evaluasi yang menggabungkan presisi dan *recall* menjadi satu nilai tunggal. Hal ini dilakukan dengan menghitung *harmonicmean*[23] dari presisi dan *recall*. F1-score memberikan gambaran yang seimbang tentang kinerja model dalam mengklasifikasikan kelas positif, dengan mempertimbangkan baik jumlah prediksi yang benar positif maupun kemampuan model untuk menemukan semua kelas yang sebenarnya positif[18]. Secara matematis F1-score dapat dituliskan seperti rumus 4.

$$F1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad (4)$$

Dimana:

F1 : F1-score

FP : *FalsePositive*

TP : *TruePositive*

FN : *FalseNegative*

3. HASIL DAN PEMBAHASAN

Data diperoleh dari situs Kaggle yang diakses pada tanggal 26 Maret 2024 melalui tautan berikut: <https://www.kaggle.com/datasets>. Data set ini merupakan kumpulan data historis mengenai calon pembeli mobil. Terdapat lima fitur yang diobservasi dalam data set ini, yaitu Usia (dalam rentang 24 hingga 60 tahun), Status Pernikahan (dinyatakan dalam skala numerik di mana 0 menunjukkan belum menikah, 1 menikah, 2 menikah dengan memiliki anak, dan 3 sebagai duda atau janda), Jenis Kelamin (dinyatakan sebagai 0 untuk pria dan 1 untuk perempuan), Jumlah Mobil, dan Penghasilan per Tahun. Selain itu, terdapat satu variabel target yang diamati, yaitu Keterangan Membeli Mobil, di mana 0 mengindikasikan tidak membeli dan 1 mengindikasikan membeli. Data set ini terdiri dari 1000 baris data yang diamati untuk setiap fitur dan variabel target.

Tabel 1 Perbandingan Kinerja Algoritma *K-NearestNeighbors*(KNN) Berdasarkan Jumlah Tetangga dan Komposisi Data Latih dan Data Uji

Jumlah Tetangga	Data Latih	Data Uji	Nilai Rata-Rata			
			Akurasi	Presisi	Recall	F1-Score
5	70%	30%	0.94	0.936	0.936	0.936
	80%	20%	0.9406	0.9373	0.9346	0.936
	90%	10%	0.942	0.9373	0.9386	0.939
11	70%	30%	0.938	0.936	0.932	0.933
	80%	20%	0.9393	0.9373	0.932	0.9346
	90%	10%	0.9406	0.9373	0.9353	0.9366

Tabel 1 menampilkan perbandingan kinerja algoritma *K-NearestNeighbors* (KNN) berdasarkan jumlah tetangga dan komposisi data latih dan data uji. Terdapat enam *model*KNN yang terbentuk dengan memvariasikan jumlah tetangga serta komposisi data latih dan uji. *Model*KNN dengan jumlah tetangga 5, cenderung menunjukkan nilai rata-rata metrik evaluasi di atas *model*KNN dengan tetangga 11.

Penggunaan data latih 70%, 80%, dan 90% pada *model* KNN tetangga 5 juga memberikan nilai presisi, *recall*, dan F1-Score yang relatif serupa, dengan nilai tertinggi terdapat pada *model* KNN dengan komposisi 90% data latih dan 10% data uji. Secara keseluruhan, hasil eksperimen menunjukkan bahwa penggunaan *model* KNN dengan tetangga 5 cenderung memberikan kinerja yang sedikit lebih baik model KNN dengan jumlah tetangga 11.

Tabel 2 Perbandingan Kinerja Algoritma *Decision Tree* Berdasarkan Komposisi Data Latih dan Data Uji

Data	Data	Nilai Rata-Rata
------	------	-----------------

Latih	Uji	Akurasi	Presisi	Recall	F1-Score
70%	30%	0.9453	0.9406	0.9433	0.9426
80%	20%	0.9433	0.942	0.94	0.9393
90%	10%	0.9433	0.9386	0.9393	0.9393

Berdasarkan Tabel 2, dapat diamati bahwa kinerja model memiliki perbedaan yang cukup kecil antara ketigakomposisi data latih dan data uji. Secara konsisten, model mencapai akurasi yang cukup tinggi, dengan nilai rata-rata metrik evaluasi sekitar 94%. Hal ini menunjukkan bahwa model memiliki kemampuan yang baik dalam mengklasifikasikan data pembeli mobil.

Akan tetapi, terdapat sedikit perbedaan dalam nilai presisi, *recall*, dan F1-score diantara ketigamodelDecision Tree. Pada modeldengan komposisi data latih 70% dengan data uji 30%, model memiliki nilai presisi, *recall*, dan F1-Score yang sedikit lebih tinggi dibandingkan dengan modeldengan komposisi data latih 80% dengan data uji20%, juga denganmodeldengan komposisi data latih 90% dengan data uji10%. Hal ini menunjukkan bahwa proporsi data latih dan uji dapat mempengaruhi kinerja darimodelDecision Tree.

Tabel 3 Perbandingan Kinerja Algoritma *SupportVectorMachine* (SVM) Berdasarkan Jenis *Kernel* dan Komposisi Data Latih dan Data Uji

Jenis <i>Kernel</i>	Data Latih	Data Uji	Nilai Rata-Rata			
			Akurasi	Presisi	Recall	F1-Score
Linear	70%	30%	0.9313	0.9246	0.9306	0.9273
	80%	20%	0.9333	0.928	0.934	0.9286
	90%	10%	0.9306	0.9166	0.9313	0.9273
RBF	70%	30%	0.9393	0.9366	0.9346	0.9373
	80%	20%	0.936	0.9346	0.93	0.932
	90%	10%	0.9413	0.9373	0.9373	0.938
Polynomial	70%	30%	0.9406	0.9326	0.9386	0.934
	80%	20%	0.938	0.934	0.9393	0.9353
	90%	10%	0.9373	0.9293	0.938	0.934

Tabel 3 menampilkan perbandingan kinerja algoritma SVM berdasarkan jenis *kernel* dan komposisi data latih dan uji, terdapat 9 model SVM yang terbentuk dari memvariasikan jenis *kernel* dengan dengan komposisi data latih dan data uji. Dari tabel tersebut dapat dilihat bahwa SVM dengan *kernelRBF* memiliki nilai rata-rata metrik evaluasi paling baik diantara ketiga jenis *kernel* untuk data set calon pembeli mobil.

Dari tabel tersebut juga dapat dilihat tren yang cukup menarik, dimana nilai rata-rata dari metrik evaluasi cenderung memiliki nilai yang sama pada model SVM dengan komposisi data latih 70% dan data uji 30% dengan model dengan data latih 90% dan data uji 10%. Hal ini menunjukkan bahwa dalam data set ini, komposisi data latih dan data uji tidak memberikan pengaruh yang signifikan terhadap kinerja model SVM.

Tabel 4 Perbandingan Kinerja Algoritma *Artificial Neural Network* (ANN) Berdasarkan Komposisi Data Latih dan Data Uji

Data Latih	Data Uji	Nilai Rata-Rata			
		Akurasi	Presisi	Recall	F1-Score
70%	30%	0.8546	0.8533	0.8326	0.8406
80%	20%	0.8453	0.8453	0.8246	0.832
90%	10%	0.86	0.852	0.8453	0.8453

Tabel 4 memberikan perbandingan kinerja algoritma *Artificial Neural Network* (ANN) berdasarkan komposisi data latih dan data uji. Pada tabel tersebut dapat dilihat bahwa *model* ANN dengan komposisi data latih 90% dan data uji 10% memiliki yang rata-rata metrik evaluasi yang paling tinggi jika dibandingkan dengan dua komposisi data latih dan data uji lainnya.

Tabel 5 Perbandingan Kinerja Model Terbaik dari Setiap Algoritma

Algoritma	Nilai Rata-Rata			
	Akurasi	Presisi	Recall	F1-Score
<i>K-NearestNeighbours (KNN)</i>	0.942	0.9373	0.9386	0.939
<i>Decision Tree</i>	0.9453	0.9406	0.9433	0.9426
<i>SupportVectorMachine(SVM)</i>	0.9413	0.9373	0.9373	0.938
<i>Artificial Neural Network (ANN)</i>	0.86	0.852	0.8453	0.8453

Tabel 5 menggambarkan perbandingan kinerja *model* terbaik dari masing-masing algoritma, *model* terbaik diambil dari Tabel 1 – Tabel 4, pengambilan *model* didasarkan pada nilai rata-rata metrik evaluasi tertinggi. Algoritma yang dibandingkan pada tabel ini adalah *K-NearestNeighbour*(KNN), *Decision Tree*, *SupportVectorMachine* (SVM), dan *Artificial Neural Network* (ANN). Berdasarkan tabel tersebut dapat dilihat bahwa *model* klasifikasi dengan algoritma *Decision Tree* memiliki nilai akurasi, presisi, *recall*, dan F1-score terbaik diantara 4 algoritma klasifikasi lainnya, disusul oleh algoritma *K-NearestNeighbour* (KNN) dan *SupportVectorMachine* (SVM) yang memiliki nilai rata-rata metrik evaluasi yang hampir sama. Algoritma *Artificial Neural Network* (ANN) memiliki nilai rata-rata metrik evaluasi yang sangat buruk jika dibandingkan dengan 3 algoritma lainnya, sehingga *model* dengan algoritma ANN tidak disarankan dalam melakukan klasifikasi pada data set ini. Oleh karena itu, *model* dengan algoritma *Decision Tre* merupakan *model* klasifikasi terbaik untuk melakukan klasifikasi pada data set calon pembeli mobil.

4. KESIMPULAN

Kesimpulan dari hasil penelitian ini adalah analisis data menggunakan berbagai algoritma klasifikasi memiliki kinerja yang cukup beragam dalam memprediksi calon pembeli mobil. Penggunaan *machinelearning* dengan algoritma klasifikasi seperti *K-NearestNeighbors* (KNN), *Decision Tree*, *SupportVectorMachine* (SVM), dan *Artificial Neural Network* (ANN) memberikan pemahaman yang mendalam tentang efektivitas masing-masing algoritma dalam konteks ini.

Perbandingan kinerja algoritma menunjukkan bahwa *model Decision Tree* memiliki kinerja terbaik dalam hal akurasi, presisi, *recall*, dan F1-Score. *Model* ini menunjukkan kemampuan yang baik dalam mengklasifikasikan data pembeli mobil, dengan nilai rata-rata metrik evaluasi sekitar 94%. Di sisi lain, algoritma KNN dan SVM dengan *kernel RBF* juga menunjukkan hasil yang kompetitif dengan kinerja yang hampir sama.

Meskipun demikian, perlu diperhatikan bahwa penggunaan algoritma ANN tidak disarankan dalam kasus ini karena memberikan kinerja yang buruk dibandingkan dengan algoritma lainnya. Hal ini menunjukkan bahwa pemilihan algoritma klasifikasi sangat penting dalam menghasilkan *model* yang optimal untuk memprediksi calon pembeli mobil.

Untuk pengembangan selanjutnya, penelitian ini dapat melibatkan lebih banyak variabel atau fitur yang relevan untuk meningkatkan akurasi prediksi. Selain itu, eksperimen dapat diperluas dengan menggunakan teknik *ensemblelearning* atau mempertimbangkan algoritma klasifikasi lainnya untuk mendapatkan pemahaman yang lebih mendalam tentang data calon pembeli mobil dan meningkatkan kinerja *model*. Dengan demikian, penelitian ini memberikan dasar yang kuat

untuk pengembangan strategi pemasaran dan penjualan yang lebih efektif dalam industri jual beli mobil.

DAFTAR PUSTAKA

- [1] Y. Al-Anqoudi, "Using Machine Learning in Business Process Re-Engineering," *Big Data and Cognitive Computing*, vol. 5, 2021.
- [2] N. S. S. Sugiana dan B. Musty, "Analisis Data Sistem Informasi Monitoring Marketing," *Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, vol. 12, pp. 696-708, 2023.
- [3] V. Chandgude dan B. Kawade, "Role of Artificial Intelligence and Machine Learning in Decision Making for Business Growth," *International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)*, vol. 3, no. 1, pp. 54-57, 2023.
- [4] M. A. Tripathi, K. Madhavi, P. Kandi, V. K. Nassa, B. Mallik dan K. Chakravarthi, "Machine learning models for evaluating the benefits of business intelligence systems," *The Journal of High Technology Management Research*, vol. 34, no. 2, 2023.
- [5] H. Ji, X. Xianon, G. Su, J. Wang dan Y. Wang, "Utilizing Machine Learning for Precise Audience Targeting in Data Science and Targeted Advertising," *Academic Journal of Science and Technology*, vol. 9, pp. 216-220, 2024.
- [6] J. Nagy dan Z. Jambor, "Competitiveness in global trade: The case of the automobile industry," *ECONOMIC ANNALS*, vol. LXIII, pp. 61-84, 2018.
- [7] I. Lewaa, "Customer Segmentation Using Machine Learning Model: An Application of RFM Analysis," *Journal of Data Science and Intelligent Systems*, vol. XX, pp. 1-12, 2023.
- [8] F. Ridzuan, W. M. Nazmee dan W. Zainon, "A Review on Data Cleansing Methods for Big Data," dalam *Procedia Computer Science* 161, 2019.
- [9] M. B. Antor, "A Comparative Analysis of Machine Learning Algorithms to Predict Alzheimer's Disease," *Journal of Healthcare Engineering*, 2021.
- [10] S. Uddin, A. Khan, E. M. Hossain dan . M. . A. Moni, "Comparing different supervised machine," *BMC Medical Informatics and Decision Making*, pp. 1-16, 2019.
- [11] K. Taunk, S. De, S. Verm dan A. Swetapadma, "A Brief Review of Nearest Neighbor Algorithm for Learning and Classification," dalam *Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS 2019)*, 2019.
- [12] E. Oktavianti, M. Agustin dan R. Sari, "MPLEMENTASI ALGORITMA DECISION TREE DENGAN FITUR SELEKSI WEIGHT BY INFORMATION GAIN," *JURNAL MULTINETICS*, vol. 9, pp. 118-126, 2023.
- [13] C. M. Sitorus, A. Rizal dan M. Jajuli, "Prediksi Risiko Perjalanan Transportasi Online Dari Data Telematik Menggunakan Algoritma Support Vector Machine," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 6, pp. 254-265, 2020.
- [14] J. Cervantes, F. Garcia-Lamon, L. Rodríguez-Mazahua dan A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, vol. 408, pp. 189-205, 2020.
- [15] A. Hosameldin dan K. N. Asoke, "Condition Monitoring with Vibration Signals: Compressive Sampling and Learning Algorithms for Rotating Machines," *IEEE*, 2019, pp. 259-277.
- [16] D. Kurniasari, R. V. Mahyunis, Warsono dan A. Nurayaman, "IMPLEMENTATION OF ARTIFICIAL NEURAL USING BACKPROPAGATION ALGORITHM BY COMPARING FOUR ACTIVATION FUNCTIONS IN PREDICTING GOLD PRICES," *Kumpulan jurnal Ilmu Komputer (KLIK)*, vol. 10, pp. 93-105, 2023.
- [17] M. Lendo-Siwicka, K. Zabłocka, E. Soból, A. Markiewicz dan G. Wrzesiński, "Application of an Artificial Neural Network (ANN) Model to Determine the Value of the Damping Ratio

- (D) of Clay Soils,” *Applied Sciences*, vol. 13, no. 10, pp. 1-15, 2023.
- [18] S. Orozco-Arias, J. S. Piña dan R. Tabares-Soto, “Measuring Performance Metrics of Machine Learning,” *MDPI*, pp. 1-18, 2020.
- [19] W. I. Rahayu, C. Prianto dan E. A. Novia, “PERBANDINGAN ALGORITMA K-MEANS DAN NAÏVE BAYES UNTUK MEMREDIKSI PRIORITAS PEMBAYARAN TAGIHAN RUMAH SAKIT BERDASARKAN TINGKAT KEPENTINGAN PADA PT. PERTAMINA (PERSERO),” *Jurnal Teknik Informatika*, vol. 13, no. 2, pp. 1-8, 2021.
- [20] S. A. Hicks, . I. Strümke dan V. Thambawita1, “Evaluation metrics for medical applications of artificial intelligence,” *natureportofolio*, 2022.
- [21] T. F. Monaghan, S. N. Rahman,, C. W. Agudelo, A. J. Wein dan J. . M. Laz, “Foundational Statistical Principles in Medical Research: Sensitivity, Specificity, Positive Predictive Value, and Negative Predictive Value,” *MDPI*, pp. 1-7, 2021.
- [22] P. Fränti dan R. Mariescu-Istodor, “Soft precision and recall,” *Pattern Recognition Letters*, vol. 167, pp. 115-121, 2023.
- [23] D. Chakrabarty, “Arithmetic-Harmonic Mean: A Measure of Central Tendency of Ratio-Type Data,” *International Journal of Advanced Research in Science, Engineering and Technology* , vol. 8, no. 5, pp. 1-9, 2021.