

Komparasi Algoritma K-Nearest Neighbors dan Random Forest Pada Prediksi Harga Mobil Bekas

Sulaiman¹⁾, Edi Surya Negara²⁾

*^{1,2)} Jl. Jenderal Ahmad Yani, Kecamatan Seberang Ulu I, Palembang, Sumatera Selatan 30111

*^{1,2)} Teknik Informatika, Universitas Bina Darma

e-mail: *¹⁾sulaimann1704@gmail.com, *²⁾e.s.negara@binadarma.ac.id

Abstrak

Seiring dengan perkembangan zaman produksi pada kendaraan bermotor khususnya mobil dewasa ini mengalami peningkatan yang sangat signifikan, perusahaan saling berlomba-lomba dalam mengeluarkan seri terbarunya. Hal ini berdampak pada melimpahnya mobil seken atau bekas pakai di pasaran, perusahaan retail diuntut selektif dalam menentukan harga jual dan harga beli dari mobil yang akan di beli serta yang akan diperjual belikan kembali, dengan memanfaatkan teknologi khususnya Data Mining diharapkan dapat membantu proses seleksi mobil yang akan di beli dengan cepat serta akurat. Prediksi harga mobil bekas dipengaruhi beberapa faktor antara lain jenis mobil, bahan bakar, jarak tempuh, tahun produksi, serta transmisi. Dengan adanya permasalahan tersebut, penulis mencoba memkomparasi algoritma K-Nearest Neighbors dan Random Forest sebagai dasar pembuatan model machine learning yang dapat memprediksi harga mobil bekas yang sesuai dengan spesifikasi yang ada. Hasil pada penelitian ini menunjukkan Algoritma Random Forest mendapat nilai error yang lebih kecil serta akurasi yang lebih baik dibandingkan K-Nearest Neighbors, Random Forest mendapat akurasi 96.38 % dan K-Nearest Neighbors 59.17 %, nilai error yang diperoleh dihitung menggunakan metrik evaluasi MSE (Mean Squared Error) dengan menghitung selisih rata-rata nilai sebenarnya dengan nilai prediksi.

Kata kunci—Data Mining, Komparasi, Prediksi, Machine Learning

Abstract

Along with the development of the era of production of motorized vehicles, especially cars today, has experienced a very significant increase, companies are competing with each other in issuing the latest series. This has an impact on the abundance of used or used cars on the market, retail companies are demanded to be selective in determining the selling price and purchase price of the cars to be purchased and those to be resold, by utilizing technology, especially Data Mining, which is expected to assist the car selection process. which will be purchased quickly and accurately. Used car price predictions are influenced by several factors, including the type of car, fuel, distance traveled, year of production, and transmission. With these problems, the author tries to compare the K-Nearest Neighbors and Random Forest algorithms as the basis for making machine learning models that can predict the price of used cars according to existing specifications. The results of this study indicate that the Random Forest Algorithm has a smaller error value and better accuracy than K-Nearest Neighbors, Random Forest has an accuracy of 96.38% and K-Nearest Neighbors 59.17%, the error value obtained is calculated using the MSE (Mean Squared Error) evaluation metric by calculating the average difference between the actual value and the predicted value.

Keywords— Data Mining, Comparisson, Prediction, Machine Learning

1. INTRODUCTION

Seiring dengan perkembangan zaman, produsen mobil berlomba-lomba mengeluarkan seri terbaru hingga bersaing dalam membuat teknologi yang memikat daya beli konsumen. Hal ini berimbas pada jumlah mobil bekas yang masih layak pakai, karena konsumen di kalangan atas mencari mobil yang memiliki fitur serta teknologi yang mumpuni pada masa saat ini. Pada perusahaan distributor dan retail memiliki model bisnis yang dijalankan adapun model bisnis tersebut adalah perusahaan membeli mobil bekas dari produsen kemudian menjualnya kepada konsumen. Perusahaan juga menerima penjualan kembali mobil bekas dari konsumen. Masalah yang dihadapi perusahaan adalah penetapan harga mobil bekas dengan tepat[1]. Harga adalah patokan pertama dalam pembelian serta penjualan[2]. Maka dari itu perusahaan diwajibkan bisa memperkirakan harga jual serta beli dari mobil bekas di pasaran yang sesuai dengan spesifikasinya. Dengan demikian diwajibkan perusahaan bisa mendapatkan profit yang besar. Berkaitan dengan teknologi informasi tujuan kesehatan, tujuan hukum, dan tujuan pendidikan atau pembelajaran[2]. Aplikasi-aplikasi ini sudah terbukti sangat membantu pengguna dalam mendampingi kehidupannya.

Data mining merupakan proses menemukan pola dan informasi menarik tentang data terpilih menggunakan metode tertentu[3]. *Data mining* berfungsi untuk membantu banyak pekerjaan antara lain pada kasus *classification, clustering, association, regression, forecasting, sequence analysis, deviation analysis*[4]. *Data mining* sangat berguna untuk membantu perusahaan dalam menemukan informasi yang sangat penting dari gudang data yang dimiliki setiap perusahaan itu sendiri[5]. *Data mining* bekerja dengan cara menggali informasi penting yang sebelumnya belum diketahui, atau bisa disebut memprediksi menggunakan teknik permodelan.

Tujuan penelitian ini untuk membuat model machine yang memiliki nilai error yang kecil serta dapat memprediksi harga mobil sesuai dengan spesifikasi yang ada. Dengan memanfaatkan teknologi data mining diharapkan bisa meminimalisir kesalahan pada proses prediksi yang dilakukan. Prediksi tidak harus memberikan jawaban yang pasti benar tentang peristiwa yang akan terjadi, tetapi lebih mencari jawaban yang sedekat mungkin dengan apa yang akan terjadi[5]. Penggunaan prediksi dalam bisnis merupakan alat dan faktor penting dalam keputusan konsumen[6]. Penelitian ini menggunakan 2 algoritma yaitu *K-Nearest Neighbors* dan *Random Forest* sebagai perbandingan hasil komparasi, akurasi dan nilai *error* yang di dapatkan. Algoritma ini bekerja dengan cara mencari pola di setiap kumpulan data kemudian memprediksi harga mobil bekas[7][8].

B. Kriswantara et al. melakukan penelitian terkait prediksi harga mobil bekas dengan machine learning menggunakan model *Decision Tree, Deep Neural Network* dan *Random Forest*[9]. Penelitian ini memperoleh hasil dimana akurasi pada *Decision Tree* 72%, DNN mendapat hasil R2 0.88 sedangkan *Random Forest* mendapat akurasi 84%. Dengan hasil diatas dapat disimpulkan bahwa *Random Forest* lebih cepat pada proses *training* serta mendapat akurasi paling tinggi.

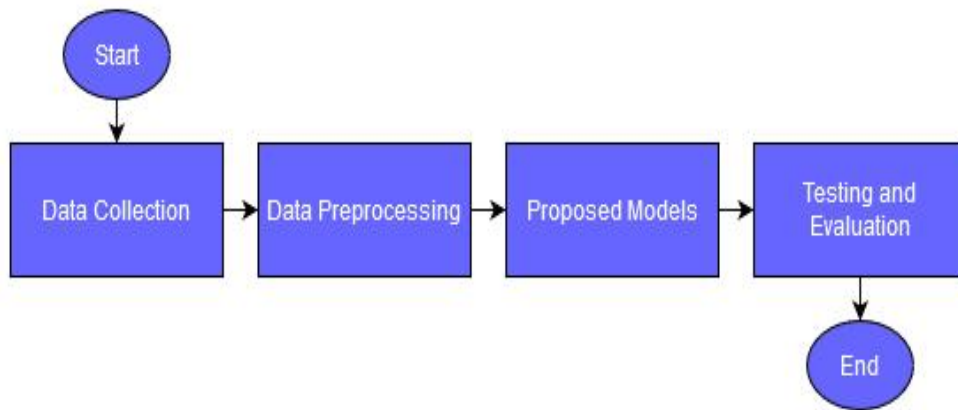
Suria Sandi Winarto et al. Melakukan penelitian tentang menentukan harga mobil bekas dengan metode *fuzzy mamdani* dan metode jaringan syaraf tiruan[10]. Penelitian ini mendapatkan hasil dimana metode jaringan syaraf tiruan dapat memprediksi harga mobil yang mendekati harga sebenarnya, dimana tingkat kebenaran hasil dari metode *fuzzy mamdani* 97.910098 % dan syaraf tiruan 97.93914 %.

Dari kedua penelitian diatas terlihat adanya perbedaan hasil prediksi pada model terutama pada metode yang digunakan. Oleh karena itu, penelitian ini dilakukan dalam rangka untuk mengetahui perbandingan kinerja pada model prediksi sehingga dapat diketahui model mana yang paling tepat berdasarkan tingkat akurasi, nilai error, hasil prediksi yang di hasilkan. Nilai *error* dari model dihitung menggunakan metrik evaluasi MSE, alasan penggunaan metrik MSE karena metrik ini cenderung minim bias dan merupakan penaksir yang *underestimate*[11]. Model yang akan dibandingkan pada penelitian ini adalah *K-Nearest Neighbors* dan *Random Forest*. Sedangkan data yang digunakan diambil dari kaggle.com yang terdiri dari 9 (sembilan)

atribut yaitu: *car_name*, *year*, *selling_price*, *kms_driven*, *fuel_type*, *seller_type*, *transmission* dan *owner*.

2. RESEARCH METHODS

Penelitian ini menggunakan metode eksperimen. Dimana metode penelitian eksperimen adalah metode penelitian yang digunakan guna mencari pengaruh *treatment* (perlakuan) tertentu[12]. Metode ini terbagi kedalam beberapa tahap antara lain: data collection, data preprocessing, model yang diajukan, *testing* dan evaluasi. Diagram alur dapat di lihat pada gambar 1.



Gambar 1. *Flow* Penelitian

2.1 Data Collection

Data yang dipakai pada penelitian ini bersumber dari kaggle.com Dataset ini merupakan kumpulan data tentang mobil bekas dan berbagai fitur yang dimiliki, Dataset ini berisi informasi tentang mobil bekas yang tercantum pada www.cardekho.com. Dataset ini dapat digunakan untuk berbagai keperluan, contohnya seperti prediksi harga.

	Car_Name	Year	Selling_Price	Present_Price	Kms_Driven	Fuel_Type	Seller_Type	Transmission	Owner
0	ritz	2014	3.35	5.59	27000	Petrol	Dealer	Manual	0
1	sx4	2013	4.75	9.54	43000	Diesel	Dealer	Manual	0
2	ciaz	2017	7.25	9.85	6900	Petrol	Dealer	Manual	0
3	wagon r	2011	2.85	4.15	5200	Petrol	Dealer	Manual	0
4	swift	2014	4.60	6.87	42450	Diesel	Dealer	Manual	0
...
296	city	2016	9.50	11.60	33988	Diesel	Dealer	Manual	0
297	brio	2015	4.00	5.90	60000	Petrol	Dealer	Manual	0
298	city	2009	3.35	11.00	87934	Petrol	Dealer	Manual	0
299	city	2017	11.50	12.50	9000	Diesel	Dealer	Manual	0
300	brio	2016	5.30	5.90	5464	Petrol	Dealer	Manual	0

301 rows x 9 columns

Pada gambar diatas, terdapat 301 baris dengan 9 kolom. Dataset ini belum bersifat final, karena masih akan dirubah sedemikian rupa untuk memudahkan proses permodelan.

2.2 Data Preprocessing

Pada tahap ini dilakukan proses transformasi pada data sehingga menjadi bentuk yang cocok untuk proses permodelan. Prosesnya antara lain encoding data menggunakan teknik *one-hot-encoding* serta melakukan pembaggian dataset menggunakan fungsi *train_test_split* dengan proporsi 0.2.

2.3 Proposed Models

Pada tahap ini dilakukan proses modelling menggunakan bantuan tools *google colab* dan bahasa pemrograman *python*. Model yang akan di komparasi antara lain *K-Nearest Neighbors* dan *Random Forest*. Akan dilakukan proses perhitungan nilai *error* serta akurasi pada masing-masing model sehingga dapat menjadi acuan proses komparasi.

2.3.1 K-Nearest Neighbors

Algoritma K-nearest Neighbors adalah metode klasifikasi yang mengelompokkan data baru berdasarkan jarak data terdekat atau data baru dengan tetangga[13]. KNN menerima masukan berupa *vector*[14]. KNN relatif sederhana dibandingkan dengan algoritma lainnya. Algoritma ini menggunakan kesamaan fitur untuk memprediksi nilai bagi semua data baru. KNN adalah bagian dari pembelajaran mesin terawasi yang melatih sekumpulan data yang sudah memiliki label atau bisa di katakan *supervised learning*[15]. Metrik ukuran jarak yang di pakai pada KNN antara lain *Euclidian distance*, *Hamming*, *Minkowski*, *Cosine*, *Jaccard*, dan *Manhattan distance*[16]. Rumus perhitungan jarak menggunakan *Euclidian Distance* ditunjukkan pada persamaan(1).

$$D(x, y) = \sqrt{\sum_{k=1}^n (x_i - y_i)^2} \quad (1)$$

Dimana D adalah jarak antara titik-titik pada data $x = x_1, x_2, \dots, x_i$ dan y_1, y_2, \dots, y_i dan i mewakili nilai atribut, dan n adalah dimensi atribut.

2.3.2 Random Forest

Random forest adalah algoritma *supervised learning* yang bisa digunakan pada kasus regresi dan klasifikasi. *Random forest* juga merupakan algoritma yang sering digunakan karena cukup sederhana tetapi memiliki stabilitas yang baik. *Random forest* termasuk ke pembelajaran mesin ensemble, yang merupakan model prediksi yang terdiri dari beberapa model dan bekerja secara bersama. Keputusan *random forest* dibangun berdasarkan vector acak, pembentukan pohon ini dilakukan dengan cara memilih nilai F acak. Parameter yang digunakan untuk mengatur intensitas pada *random forest* ketika nilai F dipilih dan ketika jumlah pohon yang akan dibangun[3]. Korelasi yang terbentuk berdasarkan nilai F yang dipilih apabila nilainya terlalu kecil akan menimbulkan korelasi yang kecil[17]. Perhitungan nilai F dapat dilihat pada rumus (2)

$$F = \text{Log}_2(M + 1) \quad (2)$$

Dimana M adalah total fitur, saat melakukan pemilihan fitur dilakukan juga pengacakan saat memilih set pelatihan[3].

Hyper parameter yang biasa digunakan pada Random Forest antara lain:

- *n-jobs* : jumlah job (pekerjaan) yang digunakan secara paralel.
- *n-estimator*: jumlah trees.
- *max-depth*: kedalaman pohon, merupakan ukuran banyak pohon yang membelah.
- *random_state*: control random generator yang digunakan.
-

2.4 Testing and Evaluation

Pada tahap ini dilakukan penghitungan nilai error menggunakan metrik evaluasi *MSE* (*Mean Squared Error*). Tahap ini merupakan tahap evaluasi dengan menginterpretasikan output data mining yang dihasilkan dari fase sebelumnya[18]. *MSE* bekerja dengan mengukur ketepatan nilai dugaan model yang dinyatakan dalam rata-rata kuadrat dari kesalahan dan juga dapat digunakan untuk membandingkan ketetapan ramalan antara metode peramalan yang berbeda[19]. *MSE* dihitung menggunakan rumus (3)

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - y_{pred_i})^2 \quad (3)$$

Mengurangi nilai data aktual dengan data yang diprediksi, dikuadratkan hasilnya, kemudian dijumlahkan semuanya, dan dibagi dengan jumlah data yang tersedia.

3. RESULT AND DISCUSSION

3.1 Proprocessing

Tahap ini dilakukan dengan beberapa tahap, seperti *One Hot Encoding*, serta membagi dataset menggunakan fungsi *train_test_split*. Disamping itu dilakukan penambahan fitur age untuk merepresentasikan usia pada mobil, merubah fitur kategorik ke bentuk numerik secara manual. Proses ini dilakukan untuk merubah setiap nilai di dalam kolom menjadi kolom baru dan mengisinya dengan nilai biner dari variabel kategori ke variabel numerik menjadi 0 dan 1. *One-hot-encoding* berfungsi untuk mendapatkan fitur baru yang sesuai sehingga dapat mewakili variabel kategori. Pentingnya proses pembagian dataset menggunakan fungsi *train_test_split* ini karena dalam *machine learning*, hal yang pedulikan adalah seberapa baik kinerja model ketika dihadapkan pada data yang belum pernah dilihat sebelumnya. Oleh karena itulah, harus melatih dan menguji model menggunakan dua data yang berbeda. Korelasi fitur pada data dapat dilihat pada gambar dibawah ini.



Gambar 3. *Correlation matrix* pada data

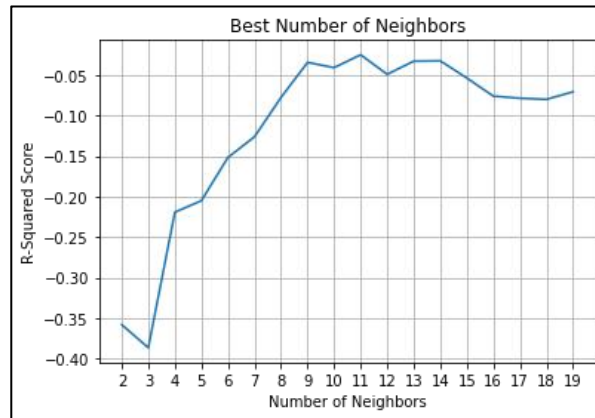
Koefisien korelasi berkisar antara -1 dan +1. mengukur kekuatan hubungan antara dua variabel serta arahnya (positif atau negatif). Mengenai kekuatan hubungan antar variabel, semakin dekat nilainya ke 1 atau -1, korelasinya semakin kuat. Sedangkan, semakin dekat nilainya ke 0, korelasinya semakin lemah.

3.2 Prediction

Setelah selesai pada proses preparasi data selanjutnya masuk ke tahap modelling menggunakan 2 algoritma yaitu *K-Nearest Neighbors* dan *Random Forest*. Hasil dari proses *train* ini selanjutnya akan di hitung nilai *error* nya serta melakukan uji coba terhadap data uji. Percobaan pertama adalah membandingkan akurasi dari masing-masing model, percobaan kedua adalah menghitung nilai *error* menggunakan metrik MSE dan membandingkan hasilnya dan percobaan terakhir adalah melakukan perbandingan dari hasil yang di dapatkan kedua model menggunakan fungsi *predict*.

3.2.1 *K-Nearest Neighbors*

Melakukan proses modelling dengan set hyper parameter tuning neighbors yang memiliki performa paling bagus pada data. Hasilnya dapat dilihat pada gambar dibawah ini.



Gambar 4. Best neighbors

Hasil yang di dapat nilai $n = 2$, jika mengacu pada gambar diatas harusnya 3 adalah nilai terbaiknya. Tetapi saat melakukan pengujian nilai 3 sebagai parameter n mendapat akurasi yang lebih rendah di dibandingkan nilai 2, maka kami memutuskan memakai nilai 2. Akurasi serta nilai *error* yang di dapat pada model dapat dilihat pada tabel 1.

Tabel 1 Hasil pengujian model KNN

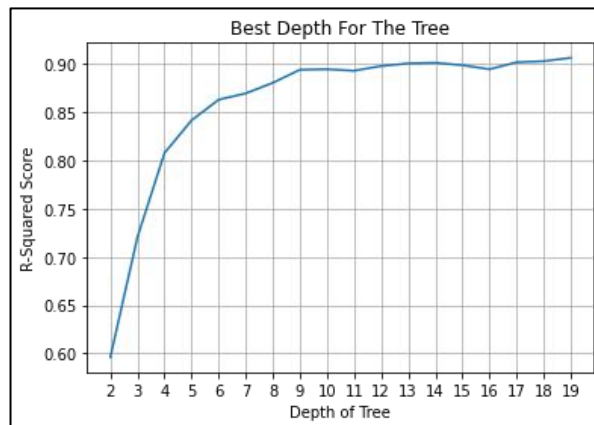
Model	Accuracy	Nilai Error	
		Train	Test
<i>K-Nearest Neighbors</i>	59.17 %	0.009817	0.044094

Berdasarkan hasil modelling pada Tabel 1. *Accuracy* sebesar 59.7 %, sedangkan pada nilai *error train* sebesar 0.009817 dan *test nya* sebesar 0.044094.

3.2.2 Random Forest

Melakukan proses modelling menggunakan hyper parameter tuning *max_depth* yang paling bagus pada data. Hasilnya dapat dilihat pada gambar dibawah ini.

Pada proses ini kami menggunakan *max_depth* pada model sebesar 12, nilai inilah yang akan dipakai pada proses permodelan. Akurasi serta nilai *error* yang di dapat pada model dapat dilihat pada tabel 2.



Gambar 5. Best max_depth

Tabel 2 Hasil pengujian model *Random Forest*

Model	Accuracy	Nilai Error	
		Train	Test
<i>Random Forest</i>	96.38 %	0.001003	0.003206

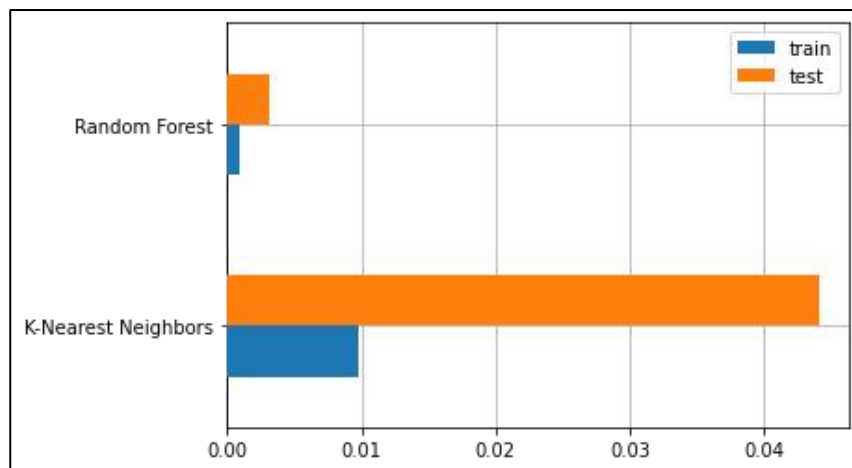
Berdasarkan hasil modelling pada Tabel 2. *Accuracy* sebesar 96.38 %, sedangkan pada nilai error train sebesar 0.001003 dan test nya sebesar 0.003206.

3.3 Comparative Evaluation

Pada tahap ini akan dilakukan komparasi dari kedua model, terkait *accuracy*, nilai *error* serta hasil prediksinya. Hasil komparasi dapat dilihat pada tabel dibawah ini.

Tabel 3 Hasil komparasi

Model	Accuracy	Nilai Error	
		Train	Test
<i>K-Nearest Neighbors</i>	59.17 %	0.009817	0.044094
<i>Random Forest</i>	96.38 %	0.001003	0.003206

Gambar 6. Visualisasi komparasi nilai *error*

Pada gambar 6. Serta Tabel 3. Terlihat dimana model *Random Forest* menghasilkan nilai yang error yang lebih kecil serta akurasiya juga menunjukkan unggulnya model *Random Forest*. Tahap selanjutnya adalah komparasi hasil prediksi pada masing-masing model. Prediksi dilakukan dengan cara membuat fungsi predict pada model, kami melakukan test pada mobil wagon r dengan harga jual sebenarnya adalah 2.85. Hasil prediksi dapat dilihat pada tabel di bawah ini.

Tabel 4 Komparasi hasil prediksi

Model	Nama Mobil	Hasil Prediksi
<i>K-Nearest Neighbors</i>	Wagon r	1.9500000000000002
<i>Random Forest</i>	Wagon r	2.82227752941503

Berdasarkan hasil komparasi pada Tabel 4. Terlihat bahwa model *Random Forest* mendapatkan hasil yang sangat mendekati harga asli dari mobil bekas wagon r, sedangkan pada model *K-Nearest Neighbors* hasilnya masih terlampaui jauh dari yang diharapkan.

4. CONCLUSION

Berdasarkan beberapa skenario pengujian dan analisis yang dilakukan, dapat diambil kesimpulan sebagai berikut, perbandingan antara Algoritma *K-Nearest Neighbors* dan *Random Forest* mendapatkan hasil nilai *error* yang cukup jauh. Hal ini dibuktikan dengan proses pengujian dimana KNN mendapatkan nilai 0.009817, *Random Forest* 0.001003. Algoritma yang di sarankan penulis dalam pembuatan model *machine learning* pada studi kasus prediksi harga mobil bekas ini adalah *Random Forest*. Serta akurasi pada model menunjukkan keunggulan model *Random Forest* yang mendapatkan nilai 96.38 % sedangkan model KNN mendapatkan nilai 59.17 %. Dapat ditarik kesimpulan bahwa model menggunakan algoritma *Random Forest* lebih unggul dibandingkan *K-Nearest Neighbors*.

5. SUGGESTION

Saran bagi penelitian selanjutnya agar menerapkan *EDA(Exploratory Data Analysis)* lebih lanjut sebelum menghapus fitur pada data, serta menggunakan teknik *feature selection*, karena bisa jadi fitur tersebut dibutuhkan dan dapat meningkatkan performa pada model yang dibuat.

BIBLIOGRAPHY

- [1] A. Amalia, M. Radhi, S. H. Sinurat, D. R. H. Sitompul, E. Indra, et al., "Prediksi harga mobil menggunakan algoritma regresi dengan hyper-parameter tuning," *Jurnal Sistem Informasi dan Ilmu Komputer Prima (JUSIKOM PRIMA)*, vol. 4, no. 2, pp. 28–32, 2021.
- [2] V. W. Siburian and I. E. Mulyana, "Prediksi harga ponsel menggunakan metode random forest," in *Annual Research Seminar (ARS)*, vol. 4, pp. 144–147, 2019.
- [3] R. Amanda and E. S. Negara, "Analysis and implementation machine learning for youtube data classification by comparing the performance of classification algorithms," *Jurnal Online Informatika*, vol. 5, no. 1, pp. 61–72, 2020.
- [4] A. Saleh, F. Nasari, et al., "Penerapan equal-width interval discretization dalam metode naive bayes untuk meningkatkan akurasi prediksi pemilihan jurusan siswa," *Masyarakat Telematika Dan Informasi: Jurnal Penelitian Teknologi Informasi dan Komunikasi*, vol. 9, no. 1, pp. 1–12, 2018.
- [5] B. Kriswantara and R. Sadikin, "Machine learning used car price prediction with random forest regressor model," *JISICOM (Journal of Information System, Informatics and Computing)*, vol. 6, no. 1, pp. 40–49, 2022.
- [6] G. N. Ayuni and D. Fitriana, "Penerapan metode regresi linear untuk prediksi penjualan properti pada pt xyz," *Jurnal telematika*, vol. 14, no. 2, pp. 79–86, 2019.
- [7] N. Nafiyah, "Penerapan regresi linear dalam memprediksi harga jual mobil bekas," in *Prosiding Seminar Nasional Pengembangan Aktual Teknologi Informasi (SENA BAKTI)*, vol. 2, pp. 2–7, 2015.
- [8] F. Rahmawati and N. Merlina, "Metode data mining terhadap data penjualan sparepart mesin fotocopy menggunakan algoritma apriori," *PIKSEL: Penelitian Ilmu Komputer Sistem Embedded and Logic*, vol. 6, no. 1, pp. 9–20, 2018.
- [9] B. Kriswantara, K. Kurniawati, and H. F. Pardede, "Prediksi harga mobil bekas dengan machine learning," *Syntax Literate; Jurnal Ilmiah Indonesia*, vol. 6, no. 5, pp. 2100–2110, 2021.
- [10] S. S. Winarto and T. Sutojo, "Menentukan harga mobil bekas dengan menggunakan metode fuzzy mamdani dan metode jaringan syaraf tiruan," *Techno. com*, vol. 11, no. 3, pp. 134–141, 2012.

-
- [11] Y. Suparman, "S-26 perlukah cross validation dilakukan? perbandingan antara mean square prediction error dan mean square error sebagai penaksir harapan kuadrat kekeliruan model yusep suparman,"
- [12] Z. Arifin, "Metodologi penelitian pendidikan," *Jurnal Al-Hikmah*, vol. 1, no. 1, 2020.
- [13] B. Santosa, "Data mining teknik pemanfaatan data untuk keperluan bisnis," *Yogyakarta: Graha Ilmu*, vol. 978, no. 979, p. 756, 2007.
- [14] A. Wanto, M. N. H. Siregar, A. P. Windarto, D. Hartama, N. L. W. S. R. Ginantra, D. Napitupulu, E. S. Negara, M. R. Lubis, S. V. Dewi, and C. Prianto, *Data Mining: Algoritma dan Implementasi*. Yayasan kita menulis, 2020.
- [15] N. M. Putry, "Komparasi algoritma knn dan naïve bayes untuk klasifikasi diagnosis penyakit diabetes mellitus," *EVOLUSI: Jurnal Sains dan Manajemen*, vol. 10, no. 1, 2022.
- [16] K. M. A. Pasaribu, R. E. Saputra, and C. Setianingsih, "Sistem informasi monitoring bencana alam dari data media sosial menggunakan metode k-nearest neighbor," *eProceedings of Engineering*, vol. 8, no. 5, 2021.
- [17] M. Mambang and A. Byna, "Analisis perbandingan algoritma c. 45, random forest dengan chaid decision tree untuk klasifikasi tingkat kecemasan ibu hamil," *SEMNASSTEKNOMEDIA ONLINE*, vol. 5, no. 1, pp. 2–1, 2017.
- [18] N. L. W. S. R. Ginantra, F. N. Arifah, A. H. Wijaya, R. S. Septarini, N. Ahmad, D. P. Y. Ardiana, F. Effendy, A. Iskandar, H. Hazriani, I. Y. Sari, *et al.*, *Data mining dan penerapan algoritma*. Yayasan Kita Menulis, 2021.
- [19] S. M. Robial, "Perbandingan model statistik pada analisis metode peramalan time series:(studi kasus: Pt. telekomunikasi indonesia, tbk kandatel sukabumi)," *SANTIKA is a scientific journal of science and technology*, vol. 8, no. 2, pp. 823–838, 2018.