

ONE-HOT ENCODING FEATURE ENGINEERING UNTUK LABEL-BASED DATA STUDI KASUS PREDIKSI HARGA MOBIL BEKAS

*Cevi Herdian¹, Ahya Kamila², Felliks Feiters Tampinongkol³, Agung Stefanus Kembau⁴,
I Gusti Agung Musa Budidarma⁵*

^{1,2}Prodi Data Science Universitas Bunda Mulia, Jl. Lodan Raya No.2 Jakarta Utara

³Prodi Informatika Universitas Bunda Mulia, Jl. Lodan Raya No.2 Jakarta Utara

^{4,5}Prodi Bisnis Digital Universitas Bunda Mulia, Jl. Lodan Raya No.2 Jakarta Utara

Email: ¹itsmecevi@gmail.com, ²akamila@bundamulia.ac.id, ³ftampinongkol@bundamulia.ac.id,
⁴akembau@bundamulia.ac.id, ⁵ibudidarma@bundamulia.ac.id

ABSTRAK

Penggunaan machine learning telah meluas di berbagai industri untuk analisis tren dan prediksi. Untuk memprediksi harga mobil bekas yang fluktuatif, penelitian ini menerapkan salah satu teknik **Feature Engineering** yaitu **One-Hot Encoding**, sebuah teknik **Feature Engineering** yang fokus kepada data-data label atau **non-numeric**. Studi ini mengeksplorasi data harga penjualan mobil bekas sebagai target variabel dan beberapa fitur seperti produsen, tahun keluaran, tipe mesin, jumlah pintu, dan popularitas. Hasil dari proses **Feature Engineering** ini sangatlah bagus, dimana **R-Squared** untuk data validasi adalah 0.85 dan untuk data testing adalah 0.86. Hasil penelitian ini memberikan informasi yang berharga bagi para peneliti dan profesional bisnis yang ingin membuat sebuah model prediksi khususnya bagaimana menangani sebuah data yang bentuknya adalah label kategori atau **non-numeric**.

Keywords: *Feature engineering, One-Hot Encoding, Machine Learning, Linear Regression.*

ABSTRACT

The use of machine learning has become widespread in various industries for trend analysis and prediction. To forecast fluctuating prices of used cars, this research employs one of the Feature Engineering techniques, namely One-Hot Encoding, which focuses on label-based or non-numeric data. The study explores data on the prices of used cars as the target variable and several features such as manufacturer, production year, engine type, number of doors, and popularity. The results of this Feature Engineering process are excellent, with an R-Squared of 0.85 for validation data and 0.86 for testing data. This research provides valuable information for researchers and business professionals interested in building a prediction model, especially in handling categorical or non-numeric data.

Keywords: *Feature engineering, One-Hot Encoding, Machine Learning, Linear Regression.*

1. PENDAHULUAN

Saat ini, dengan perkembangan teknologi yang terus berlangsung dengan cepat, teknik pembelajaran mesin (**Machine Learning**) digunakan secara lebih luas untuk mengidentifikasi dan meramalkan tren atau prediksi dalam berbagai industri [1], [2]. Teknik **Feature Engineering** merupakan bagian penting dalam pengembangan model prediktif karena esensial untuk menciptakan representasi data yang sesuai untuk model **Machine Learning** [3],[4].

Metode "**One-Hot Encoding**" menjadi fokus khusus dalam **Feature Engineering**, terutama ketika digunakan dengan data berlabel atau terkategori [5], [6]. Dalam konteks ini, tujuan utama dari penelitian ini adalah untuk menerapkan metode **One-Hot Encoding** ke dalam prediksi harga mobil bekas dengan tujuan mendapatkan model yang lebih akurat.

Prediksi harga penjualan mobil bekas merupakan hal yang tidak sederhana. Membuat prediksi harga

mobil bekas dengan akurat menjadi tugas yang menarik ketika berhadapan dengan variasi yang besar dalam karakteristik di dalam datanya seperti merek, model, tahun produksi, dan kondisi kendaraan.

Dengan penelitian ini, peneliti berharap dapat meningkatkan pemahaman tentang penerapan **One-Hot Encoding** untuk meningkatkan kinerja model prediktif pada dataset. Peneliti akan menguji sejumlah statistik penjualan mobil bekas, seperti distribusi merek, volatilitas harga berdasarkan model, dan variabel lain yang memengaruhi ketidakpastian harga kendaraan.

Tujuan dari penelitian ini adalah untuk meningkatkan kualitas prediksi harga mobil bekas. Secara khusus, peneliti, praktisi industri, dan pengembang model **Machine Learning** akan menemukan wawasan berharga dari penelitian ini, yang berfokus pada **Feature Engineering** menggunakan **One-Hot Encoding**.

2. TINJAUAN PUSTAKA

Dalam Tinjauan Pustaka kali ini, peneliti menekankan dua penelitian sebelumnya, yaitu **Kuhn M, Johnson**, yang menjelaskan terkait dengan praktek dari beberapa contoh **Feature Engineering** dalam penerapan metode prediktif. Dimana dalam jurnal tersebut dijelaskan bahwa salah satu cara untuk meningkatkan akurasi adalah dengan menggunakan metode **Feature Engineering**. Metode **Feature Engineering** bisa diterapkan di dalam dataset numerik (angka) maupun sebuah label (non-angka) [7].

Selanjutnya, khusus untuk data berlabel (teks, kategori, non-numerik) dibahas oleh **Dahouda M, Joe I**, dimana pada jurnal tersebut dijelaskan bahwa terdapat peningkatan signifikan pada sebuah model pembelajaran mesin (**Machine Learning**) jika menggunakan metode ini. Metode ini tidak hanya untuk machine learning biasa, tetapi juga untuk deep learning seperti NLP (**Natural Language Processing**) dan **Image Recognition** [8].

2.1 Feature Engineering and Selection: A Practical Approach for Predictive Models

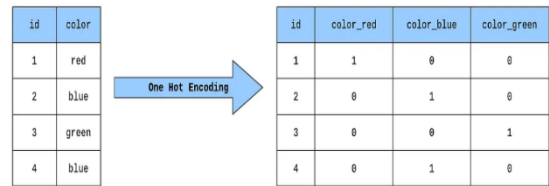
Feature engineering adalah tahap penting dalam pengembangan model prediktif, yang melibatkan transformasi variabel-variabel data menjadi bentuk yang lebih representatif untuk meningkatkan kinerja model. Salah satu teknik **Feature Engineering** yang dibahas pada penelitian ini adalah "**One-Hot Encoding**," yang digunakan untuk mengatasi data kategorikal atau berlabel. Dalam literatur, **One-Hot Encoding** telah terbukti efektif dalam meningkatkan akurasi model pada berbagai konteks, terutama ketika berhadapan dengan data yang memiliki banyak kategori atau label (**Multilabel**). Beberapa metode Feature Engineering dijelaskan di dalam tabel dibawah ini.

No	Metode Feature Engineering	Output
1	One-Hot Encoding	Vektor Biner
2	Label Encoding	Nilai Numerik
3	Imputasi	Mengisi nilai yang hilang
4	Binning	Data interval
5	Transformasi Logaritmik	Implementasi fungsi logaritmik
6	Scaling	Normalisasi nilai
7	Polynomial Features	Menciptakan fitur polynomial
8	Interaksi Fitur	Menggabungkan fitur

Tabel 1. Feature Engineering

2.2 A Deep-Learned Embedding Technique for Categorical Features Encoding

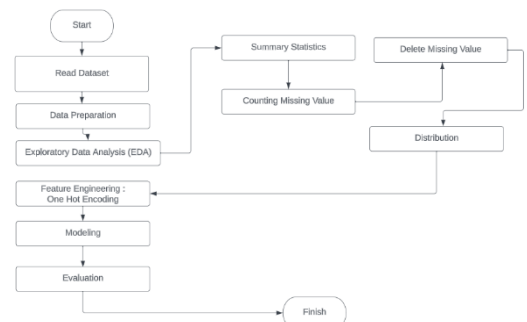
Dalam konteks prediksi harga mobil bekas, penggunaan **One-Hot Encoding** menarik perhatian karena kemampuannya untuk mengelola variabel kategori seperti merek, model, dan jenis mobil. Studi sebelumnya yang dibahas di dalam penelitian ini menunjukkan pemahaman mendalam tentang kontribusi teknik ini dalam mengatasi tantangan khusus dalam memodelkan data non-numeric.



Gambar 1. Mekanisme One-Hot Encoding

3. METODE PENELITIAN

Dibawah ini merupakan diagram alir (**Flowchart**) metodologi penelitian yang dilakukan oleh peneliti.



Gambar 2. Flowchart Proses Penelitian

Data Preparation:

Proses mengumpulkan, membersihkan, dan mengubah data mentah menjadi format yang sesuai untuk analisis disebut sebagai persiapan data. Ini merupakan tahap penting dalam alur penelitian untuk analisis data dan **Machine Learning** karena kecocokan dan kualitas data memiliki dampak besar pada efektivitas dan akurasi dari analisis atau model yang dilakukan kemudian [9].

Data yang dipergunakan di dalam penelitian ini adalah data yang berasal dari kaggle.com, dimana Kaggle merupakan salah satu penyedia data yang berasal dari komunitas data science terbesar di dunia. Dalam persiapan data, langkah-langkah krusial dilakukan untuk memastikan analisis atau model berjalan dengan efisien dan akurat. Proses dimulai dengan mengimpor data tersebut ke dalam lingkungan python. Pembersihan data kemudian dilakukan untuk mengidentifikasi dan mengatasi

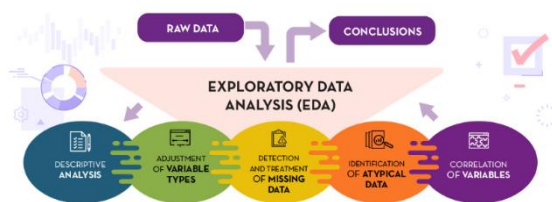
duplikat, outlier, nilai yang hilang, dan inkonsistensi dalam dataset, menjaga akurasi dan kelengkapan [9].

Selanjutnya, transformasi data melibatkan encoding variabel kategori dan feature engineering untuk meningkatkan kinerja model. Data juga diformat dengan benar dan dataset siap digunakan dalam analisis atau algoritma *Machine Learning* [10].

Exploratory Data Analysis (EDA):

Analisis data eksploratif dikenal sebagai EDA. Ini adalah metode untuk membedah dataset guna menyoroti fitur-fitur utamanya, seringkali dengan menggunakan grafik statistik dan teknik visualisasi data lainnya. Menemukan pola, hubungan, anomali, dan wawasan dalam data merupakan tujuan utama dari analisis data eksploratif (EDA), yang menjadi dasar untuk analisis dan pemodelan lebih lanjut [11].

EDA adalah langkah penting dalam proses analisis data karena memberikan pemahaman menyeluruh tentang data, membantu ilmuwan data dan analis membuat keputusan yang terinformasi tentang langkah-langkah berikutnya dalam alur kerja analisis atau pemodelan [12]. EDA sering dilakukan sebelum pemodelan statistik formal untuk membimbing arah analisis yang lebih mendalam. Beberapa metode yang paling sering dipergunakan adalah deskriptif analisis, tipe variabel, data yang hilang, dan korelasi atau hubungan antara fitur.



Gambar 3. Exploratory Data Analysis (EDA)

One-Hot Encoding:

One-Hot Encoding dalam pemrosesan data dan *Machine Learning* adalah metode yang digunakan untuk merepresentasikan variabel kategorikal sebagai vektor biner. Setiap kategori atau label dalam metode encoding ini diubah menjadi vektor biner dengan panjang yang sama dengan jumlah total kategori yang berbeda dalam variabel tersebut. Semua nilai vektor adalah nol, kecuali indeks yang sesuai dengan kategori, yang ditandai dengan angka 1 [13].

Bayangkan, sebagai contoh, bahwa ada tiga kategori dalam variabel kategorikal "Warna": Merah, Hijau, dan Biru. Setiap kategori direpresentasikan seperti berikut dalam *One-Hot Encoding*:

Red: [1, 0, 0]
 Green: [0, 1, 0]
 Blue: [0, 0, 1]

Meskipun memiliki kelebihan, one-hot encoding dapat menghasilkan ruang fitur berdimensi tinggi, terutama pada dataset dengan banyak kategori. Dalam situasi tersebut, perlu pertimbangan hati-hati untuk menyeimbangkan keuntungan representasi dengan kemungkinan peningkatan kompleksitas komputasi [14].

Machine Learning Modeling: Linear Regression

Machine Learning Modeling Linear Regression adalah sebuah algoritma dasar pembelajaran mesin yang untuk meramalkan variabel hasil yang kontinu berdasarkan satu atau lebih variabel prediktor. Secara sederhana, algoritma ini memodelkan hubungan antara variabel independen (fitur atau prediktor) dan variabel dependen (juga dikenal sebagai variabel target atau respons) dengan menyesuaikan persamaan linear pada data yang diamati [15].

Persamaan untuk Linear Regression bisa dilihat dibawah ini:

$$Y = mx + b \quad (1)$$

Dimana,
 y adalah variabel dependen,
 x adalah variabel independen,
 m adalah kemiringan garis (koefisien yang mewakili hubungan antara x dan y),
 b adalah y -intercept (titik di mana garis memotong sumbu y).

Dalam kasus regresi linear berganda, di mana terdapat beberapa variabel prediktor, persamaannya menjadi:

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n \quad (2)$$

Dimana,
 b_0 adalah y -intercept,
 b_1, b_2, \dots, b_n adalah koefisien untuk masing-masing variabel prediktor yang sesuai, yaitu x_1, x_2, \dots, x_n .

Untuk tugas-tugas seperti prediksi penjualan, memprediksi harga rumah, dan meneliti hubungan antar variabel, regresi linear sering digunakan. Ini adalah metode yang berguna dalam bidang pemodelan statistik dan analisis prediktif karena algoritmanya yang sederhana namun efektif, serta asumsi dan interpretasinya sangat bagus [16].

Evaluation:

Setiap machine learning model membutuhkan sebuah parameter terkait bagus atau tidaknya model tersebut [17]. Dalam penelitian ini, penelitian menggunakan beberapa parameter seperti dijelaskan dibawah ini.

-**Mean Absolut Error (MAE)** mengukur rata-rata perbedaan absolut antara nilai yang diprediksi dan nilai sebenarnya. Ini memberikan bobot yang sama pada semua kesalahan [18].

- **Mean Squared Error (MSE)** mengukur rata-rata perbedaan kuadrat antara nilai yang diprediksi dan nilai sebenarnya. Ini memberikan hukuman yang lebih berat pada kesalahan yang lebih besar dibandingkan MAE [19].

-**Root Mean Squared Error (RMSE)** adalah akar kuadrat dari MSE dan memberikan ukuran rata-rata besaran kesalahan dalam satuan yang sama dengan variabel target [20].

- **R-squared (R2)** mengukur proporsi varians pada variabel target yang dijelaskan oleh model. Rentang nilai R2 adalah 0 hingga 1, di mana nilai yang lebih tinggi menunjukkan kesesuaian yang lebih baik [21].

4. PEMBAHASAN

Pada bagian pembahasan ini dijelaskan hasil analisis, temuan penelitian, dan pembahasannya secara ilmiah. Di dalamnya juga terdapat bagian-bagian grafik yang menunjang hasil tersebut.

Data Preparation:

Hasil dari data yang sudah dipersiapkan menjadi sebuah dataframe (**rows** dan **columns**) yang bisa dilihat di bawah ini.

	make	model	year	engine_fuel_type	engine_hp	engine_cylinders	transmission_type	driven_w
0	bmw	1_series_m	2011	premium_unleaded_(required)	335.0	6.0	manual	rear_wheel_
1	bmw	1_series	2011	premium_unleaded_(required)	300.0	6.0	manual	rear_wheel_
2	bmw	1_series	2011	premium_unleaded_(required)	300.0	6.0	manual	rear_wheel_

Gambar 4. Dataframe hasil Data Preparation

Di dalam dataframe tersebut terdapat elemen data fitur dan juga target, dimana memiliki fitur sebagai berikut:

Make, Model, Year, Engine Fuel Type, Engine Cylinders, Transmission Type, Driven Wheels, Number of Doors, Market Category, Vehicle Size, Vehicle Style, Highway Mpg, City MPG, dan Popularity.

Dan sebagai targetnya yaitu MSRP atau **Price**.

Exploratory Data Analysis:

Beberapa EDA yang telah dibuat diantaranya adalah:

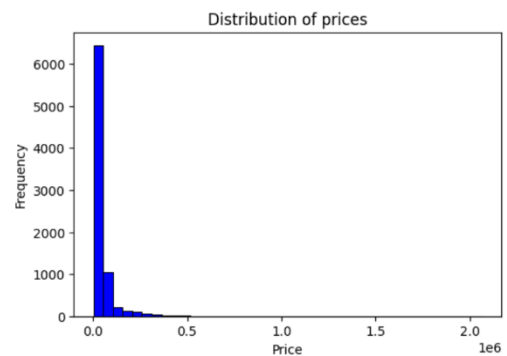
1. Summary Statistics
2. Counting Missing Value
3. Distribution
4. Outliers

	year	engine_hp	engine_cylinders	number_of_doors	highway_mpg	city_mpg	popularity	msrp
count	11914.000000	11845.000000	11884.000000	11908.000000	11914.000000	11914.000000	11914.000000	1.191400e+04
mean	2010.384338	249.38607	5.628629	3.438093	26.637485	19.732255	1554.911197	4.099474e+04
std	7.579740	109.19187	1.780559	0.881315	8.863001	8.987798	1441.855347	6.010910e+04
min	1990.000000	55.000000	0.000000	2.000000	12.000000	7.000000	2.000000	2.000000e+03
25%	2007.000000	170.000000	4.000000	2.000000	22.000000	16.000000	549.000000	2.100000e+04
50%	2015.000000	227.000000	6.000000	4.000000	26.000000	18.000000	1385.000000	2.999500e+04
75%	2016.000000	300.000000	6.000000	4.000000	30.000000	22.000000	2009.000000	4.223125e+04
max	2017.000000	1001.000000	16.000000	4.000000	354.000000	137.000000	5857.000000	2.085902e+06

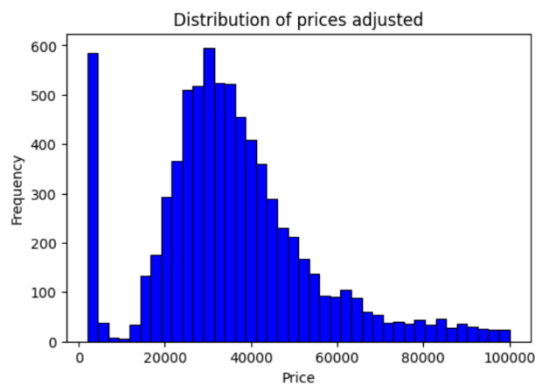
Gambar 5. Summary Statistics

```
make          0
model         0
year          0
engine_fuel_type  3
engine_hp      69
engine_cylinders 30
transmission_type 0
driven_wheels  0
number_of_doors 6
market_category 3742
vehicle_size   0
vehicle_style  0
highway_mpg    0
city_mpg       0
popularity     0
msrp           0
dtype: int64
```

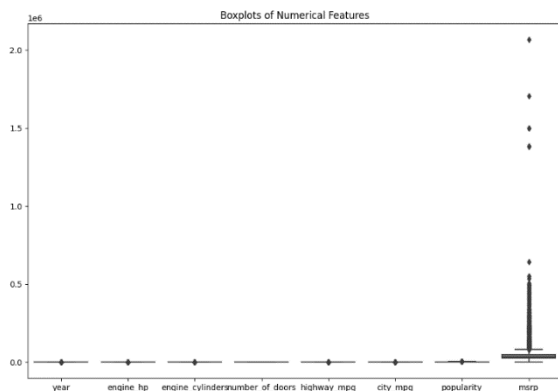
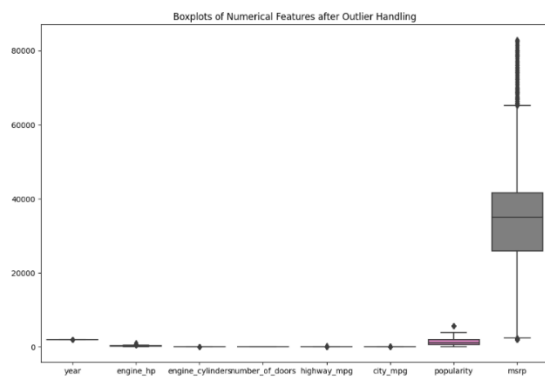
Gambar 7. Missing Value



Gambar 8. Distribution of Price



Gambar 9. Distribution of Price <10.000

Gambar 10. Ourliers Sebelum *Data Cleaning*Gambar 11. Outliers Setelah *Data Cleaning*

One-Hot Encoding:

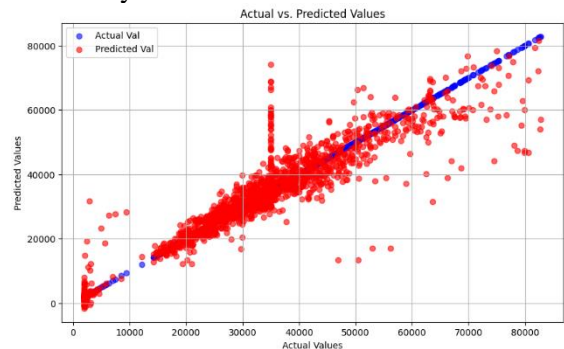
Pada tahapan ini, semua data yang berjenis teks atau *non-numeric* dibuat fitur baru seperti yang telah dijelaskan sebelumnya. Sehingga fitur yang ada di dalam dataframe meningkat dengan sangat banyak.

popularity	msrp	make_acura	make_alfa_romeo	...	vehicle_style_convertible	vehicle_style_convertible_suv
3916	46135	0	0	...	0	0
3916	40650	0	0	...	1	0
3916	36350	0	0	...	0	0
3916	29450	0	0	...	0	0
3916	34500	0	0	...	1	0

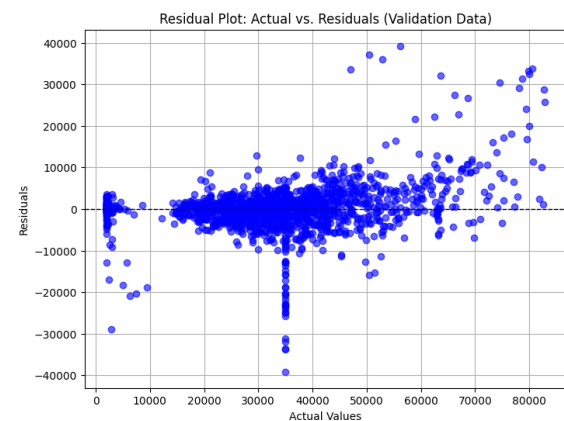
Gambar 12. Hasil *One-Hot Encoding*

Machine Learning Modeling: Linear Regression:

Proses pemodelan Machine Learning Linear Regression diimplementasikan melalui **Library Python scikit-learn**. Model tersebut menghasilkan prediksi pada dataset yang ada seperti gambar di bawah berikut juga dengan residualnya.



Gambar 13. Hasil Prediksi



Gambar 14. Hasil Residual Prediksi

Evaluasi:

Setelah dataset dipisah menjadi *training*, *validation*, dan *testing*, dimana fitur dan target juga dipisahkan. Didapatkan hasil untuk parameternya di bawah ini.

MSE Validation: 35849710.15965941

R-squared (R2) Validation: 0.8505181445876561

MAE Validation: 3462.1756800781664

RMSE Validation: 5987.462748081144

MSE Test: 35510421.5103415

R2 Test: 0.8587799213777815

MAE Test: 3549.5871152778022

RMSE: 5959.062133452

5. KESIMPULAN

Pengukuran ini memberikan gambaran tentang seberapa baik model *Machine Learning* bekerja. Kinerja model yang lebih baik ditunjukkan oleh nilai yang lebih rendah untuk *MAE*, *RMSE*, dan *MSE*, serta nilai yang lebih tinggi untuk *R2*.

Dengan jumlah kesalahan yang relatif sedikit dan skor R-squared yang tinggi dalam kasus ini, model ini tampak berfungsi dengan baik.

Meskipun hasil pemodelan yang dilakukan mendapatkan hasil yang sangat baik, tetapi pemodelan *Machine Learning* selalu menyisakan ruang untuk melakukan perbaikan. Berikut adalah beberapa teknik dan strategi lanjutan yang bisa dijadikan pertimbangan untuk meningkatkan model yang telah ada.

1. **Regularization Techniques:** Regresi Ridge, Regresi Lasso, dan kombinasi keduanya (L1 dan L2)
2. **Polynomial Regression Degree:** Eksperimen dengan derajat polinomial yang berbeda. Meskipun Anda menggunakan derajat 2, Anda dapat mencoba derajat yang lebih tinggi atau lebih rendah untuk melihat apakah itu meningkatkan kinerja model tanpa overfitting.
3. **Feature Engineering:** Dengan cermat merancang, membuat, dan memilih fitur berdasarkan pengetahuan dalam bidang tertentu serta analisis data. Pemilihan fitur menggunakan metode seperti Recursive Feature Elimination (RFE) dapat memberikan manfaat.

Khusus untuk Feature Engineering, berikut adalah beberapa ide untuk fitur yang ditambahkan:

Age of the Car, Engine Displacement, Fuel Efficiency, Popularity Interaction, Categorical Feature Interactions, Log Transformation, Binning, Year Grouping, dan Feature Scaling.

DAFTAR PUSTAKA

- [1] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Comput Sci*, vol. 2, no. 3, p. 160, May 2021, doi: 10.1007/s42979-021-00592-x.
- [2] E. Mosqueira-Rey, E. Hernández-Pereira, D. Alonso-Ríos, J. Bobes-Bascarán, and Á. Fernández-Leal, "Human-in-the-loop machine learning: a state of the art," *Artif Intell Rev*, vol. 56, no. 4, pp. 3005–3054, Apr. 2023, doi: 10.1007/s10462-022-10246-w.
- [3] E. S. Gualberto, R. T. De Sousa, T. P. De B. Vieira, J. P. C. L. Da Costa, and C. G. Duque, "From Feature Engineering and Topics Models to Enhanced Prediction Rates in Phishing Detection," *IEEE Access*, vol. 8, pp. 76368–76385, 2020, doi: 10.1109/ACCESS.2020.2989126.
- [4] E. Esenogho, I. D. Mienye, T. G. Swart, K. Aruleba, and G. Obaido, "A Neural Network Ensemble With Feature Engineering for Improved Credit Card Fraud Detection," *IEEE Access*, vol. 10, pp. 16400–16407, 2022, doi: 10.1109/ACCESS.2022.3148298.
- [5] T. Al-Shehari and R. A. Alsowail, "An Insider Data Leakage Detection Using One-Hot Encoding, Synthetic Minority Oversampling and Machine Learning Techniques," *Entropy*, vol. 23, no. 10, p. 1258, Sep. 2021, doi: 10.3390/e23101258.
- [6] S. Okada, M. Ohzeki, and S. Taguchi, "Efficient partition of integer optimization problems with one-hot encoding," *Sci Rep*, vol. 9, no. 1, p. 13036, Sep. 2019, doi: 10.1038/s41598-019-49539-6.
- [7] M. Kuhn and K. Johnson, *Feature Engineering and Selection*. Chapman and Hall/CRC, 2019. doi: 10.1201/9781315108230.
- [8] M. K. Dahouda and I. Joe, "A Deep-Learned Embedding Technique for Categorical Features Encoding," *IEEE Access*, vol. 9, pp. 114381–114391, 2021, doi: 10.1109/ACCESS.2021.3104357.
- [9] R. Croft, Y. Xie, and M. A. Babar, "Data Preparation for Software Vulnerability Prediction: A Systematic Literature Review," *IEEE Transactions on Software Engineering*, vol. 49, no. 3, pp. 1044–1063, Mar. 2023, doi: 10.1109/TSE.2022.3171202.
- [10] C. Starbuck, "Data Preparation," in *The Fundamentals of People Analytics*, Cham: Springer International Publishing, 2023, pp. 79–95. doi: 10.1007/978-3-031-28674-2_6.
- [11] A. Páez and G. Boisjoly, "Exploratory Data Analysis," 2022, pp. 25–64. doi: 10.1007/978-3-031-20719-8_2.
- [12] M. Moscarelli, "Exploratory Data Analysis in 'R,'" in *Biostatistics With 'R': A Guide for Medical Doctors*, Cham: Springer International Publishing, 2023, pp. 23–40. doi: 10.1007/978-3-031-33073-5_3.
- [13] K. Zhang et al., "Description-Enhanced Label Embedding Contrastive Learning for Text Classification," *IEEE Trans Neural Netw Learn Syst*, pp. 1–14, 2023, doi: 10.1109/TNNLS.2023.3282020.
- [14] D. Faronbi, I. Roman, and J. P. Bello, "Exploring Approaches to Multi-Task Automatic Synthesizer Programming," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, Jun. 2023, pp.

- 1–5. doi: 10.1109/ICASSP49357.2023.10095540.
- [15] G. James, D. Witten, T. Hastie, R. Tibshirani, and J. Taylor, “Linear Regression,” 2023, pp. 69–134. doi: 10.1007/978-3-031-38747-0_3.
- [16] H. Zhao, Y. Ding, L. Meng, Z. Qin, F. Yang, and A. Li, “Bayesian Multiple Linear Regression and New Modeling Paradigm for Structural Deflection Robust to Data Time Lag and Abnormal Signal,” *IEEE Sens J*, vol. 23, no. 17, pp. 19635–19647, Sep. 2023, doi: 10.1109/JSEN.2023.3294912.
- [17] G. Varoquaux and O. Colliot, “Evaluating Machine Learning Models and Their Diagnostic Value,” 2023, pp. 601–630. doi: 10.1007/978-1-0716-3195-9_20.
- [18] D. S. K. Karunasingha, “Root mean square error or mean absolute error? Use their ratio as well,” *Inf Sci (N Y)*, vol. 585, pp. 609–629, Mar. 2022, doi: 10.1016/j.ins.2021.11.036.
- [19] J. A. Ujong, E. M. Mbadike, and G. U. Alaneme, “Prediction of cost and duration of building construction using artificial neural network,” *Asian Journal of Civil Engineering*, vol. 23, no. 7, pp. 1117–1139, Nov. 2022, doi: 10.1007/s42107-022-00474-4.
- [20] G. Zenkner and S. Navarro-Martinez, “A flexible and lightweight deep learning weather forecasting model,” *Applied Intelligence*, vol. 53, no. 21, pp. 24991–25002, Nov. 2023, doi: 10.1007/s10489-023-04824-w.
- [21] R. Di Mari, S. Ingrassia, and A. Punzo, “Local and Overall Deviance R-Squared Measures for Mixtures of Generalized Linear Models,” *J Classif*, vol. 40, no. 2, pp. 233–266, Jul. 2023, doi: 10.1007/s00357-023-09432-4.