

Prediksi Target Penjualan Mobil Menggunakan Decision Tree

Syahbagus Radithya Haryo Santoso¹, Punto Adji Bhirawa^{2*}, Dwi Fibianto³, Ghani Fauzan⁴, Basuki Rahmat⁵

^{1,2,3,4,5} Program Studi Informatika, Universitas Pembangunan Nasional “Veteran” Jawa Timur

¹22081010255@student.upnjatim.ac.id

²22081010231@student.upnjatim.ac.id

³22081010321@student.upnjatim.ac.id

⁴22081010258@student.upnjatim.ac.id

*Corresponding author email: basukirahmat.if@upnjatim.ac.id

Abstrak— Penelitian ini menyoroti pentingnya memahami target pasar dalam penjualan mobil, dengan fokus pada profil demografis konsumen potensial berdasarkan usia dan penghasilan. Menggunakan teknologi machine learning dengan metode decision tree, penelitian ini bertujuan memprediksi target pasar yang optimal. Data yang digunakan berasal dari platform pembelajaran UdeMy, mencakup berbagai kelompok usia, tingkat penghasilan, dan keinginan individu untuk membeli mobil. Hasil penelitian menunjukkan bahwa decision tree dapat membuat prediksi yang akurat, memberikan panduan praktis bagi industri otomotif dalam merumuskan strategi pemasaran dan penjualan. Implementasi analisis ini diharapkan memberikan kontribusi akademis dalam memahami perilaku konsumen terkait pembelian mobil, serta meningkatkan efektivitas kampanye pemasaran dan penjualan. Selain itu, penelitian ini membantu dalam mengembangkan produk dan layanan yang lebih tepat sasaran sesuai dengan kebutuhan konsumen. Kesimpulannya, wawasan yang diperoleh dari penelitian ini dapat membantu produsen dan pemasar menyesuaikan pendekatan mereka, meningkatkan kepuasan konsumen, dan keberhasilan bisnis di industri otomotif.

Kata Kunci— Target Penjualan, Persepsi, Pelanggan, Deep Learning, Decision Tree.

I. PENDAHULUAN

Di zaman modern ini, mobil telah menjadi salah satu elemen yang sangat vital dalam aktivitas sehari-hari banyak individu dan keluarga. Mobil tidak hanya berfungsi sebagai alat transportasi yang memudahkan mobilitas, tetapi juga sering dipandang sebagai simbol status sosial dan kenyamanan hidup. Mengingat peran penting yang dimainkan oleh mobil dalam kehidupan masyarakat, memahami target pasar untuk penjualan mobil menjadi sangat krusial bagi produsen dan pemasar. Pengetahuan ini krusial untuk menyusun strategi pemasaran yang efektif dan efisien.

Penelitian ini difokuskan pada identifikasi dan analisis profil demografis konsumen potensial yang memiliki minat untuk membeli mobil. Tujuan utama dari penelitian ini adalah mengumpulkan dan menganalisis data berdasarkan dua

variabel utama yaitu usia dan penghasilan individu. Dengan memahami korelasi antara usia, pendapatan, dan keinginan untuk membeli mobil, penelitian ini bertujuan untuk mengungkap segmen pasar yang paling prospektif bagi industri otomotif. Melalui analisis ini, kita dapat menentukan pada usia berapa individu cenderung memiliki keinginan untuk membeli mobil dan berapa penghasilan yang diperlukan agar seseorang merasa mampu membeli mobil. Selain itu, penelitian ini juga menyoroti bagaimana kombinasi antara usia dan penghasilan mempengaruhi keputusan untuk membeli mobil.

Untuk menentukan target pasar yang optimal, penelitian ini memanfaatkan teknologi machine learning dengan menggunakan metode decision tree. Metode decision tree dipilih karena kemampuannya dalam melakukan prediksi dengan akurasi yang tinggi, yang sangat mendukung perusahaan dalam mengambil keputusan yang akurat. Prediksi dalam konteks ini adalah teknik untuk memperkirakan suatu nilai di masa depan berdasarkan data historis dan data saat ini. Kemampuan untuk melakukan prediksi yang akurat adalah aset yang sangat berharga bagi perusahaan atau instansi dalam merumuskan strategi pemasaran dan penjualan yang efektif.

Data yang digunakan dalam penelitian ini diambil dari dataset yang telah tersedia di platform pembelajaran UdeMy. Dataset ini mencakup berbagai kelompok usia dan tingkat penghasilan, yang menyediakan informasi mengenai keinginan individu untuk membeli mobil berdasarkan kondisi keuangan mereka. Analisis data ini diharapkan dapat memberikan wawasan yang jelas tentang segmen pasar yang paling prospektif untuk penjualan mobil dan membantu para pemasar dalam merancang strategi pemasaran yang lebih tepat sasaran.

Hasil dari analisis ini diharapkan tidak hanya memberikan kontribusi akademis dalam memahami perilaku konsumen terkait pembelian mobil, tetapi juga menawarkan panduan praktis bagi para pelaku industri otomotif. Dengan mengidentifikasi dan menjangkau segmen pasar yang paling relevan, produsen mobil dapat meningkatkan efektivitas kampanye pemasaran mereka dan, pada akhirnya, meningkatkan penjualan. Selain itu, hasil analisis ini dapat dimanfaatkan untuk mengembangkan produk dan layanan

yang lebih cocok dengan kebutuhan dan preferensi konsumen, sehingga menciptakan nilai tambah bagi perusahaan dan konsumennya.

Dengan memahami target pasar secara mendalam, para pemasar dan produsen mobil dapat menyesuaikan pendekatan mereka untuk memenuhi kebutuhan dan harapan konsumen. Penelitian ini diharapkan mampu memberikan wawasan yang berharga tentang faktor-faktor yang memengaruhi keputusan pembelian mobil sehingga dapat membantu dalam penyusunan strategi bisnis yang lebih efektif. Dengan demikian, diharapkan penelitian ini akan berkontribusi pada peningkatan penjualan mobil serta pengembangan produk dan layanan yang lebih unggul, yang pada akhirnya akan meningkatkan kepuasan konsumen dan keberhasilan bisnis di industri otomotif.

II. METODE

Pada penelitian kali ini kami menggunakan metode **Decision Tree** untuk memprediksi target penjualan mobil dengan beberapa proses yang kita lakukan untuk mencapai hasil akhir yang diinginkan, metodenya antara lain:

- Proses Pertama **Exploratory data**
Pada proses pertama yaitu *Exploratory data*, *Exploratory data* adalah sebuah langkah awal yang biasanya dilakukan sebelum lanjut ke metode selanjutnya, metode ini bisa dikatakan sebuah metode pendekatan terhadap sebuah dataset yang dilakukan dengan memvisualisasikan, mengidentifikasi, memahami sebuah data.
- Proses Kedua **Pre Processing data**
Pada proses yaitu Kedua *Preprocessing data* yang merupakan sebuah metode awal yang dilakukan sebelum tahap pemrosesan data atau bisa dibilang pada metode ini kita mengolah sebuah data yang masih mentah agar menjadi data yang lebih baik.
- Proses Ketiga **Split data**.
Pada proses ketiga yaitu *Split Data* merupakan sebuah proses yang dimana data yang kita miliki akan dibagi menjadi dua bagian, yaitu *data training* dan *data testing*, dengan komposisi data training 75% sedangkan data testing 25%. Fungsi *data training* yaitu untuk melatih sebuah model agar bisa mengidentifikasi pola yang relevan, sedangkan *data testing* merupakan implementasi dari pola-pola yang sudah dilatih pada data training.
- Proses Keempat **Feature scaling**
Pada proses keempat yaitu *Feature Scaling* yang merupakan sebuah proses dimana skala dari data akan diubah untuk memastikan data-data berada dalam skala dan jangkauan yang sama sehingga bisa membantu untuk mendapatkan hasil akhir.

III. HASIL DAN PEMBAHASAN

Sumber data dalam penelitian ini diambil dari data penjualan mobil mewah suv yang terdiri dari 2 kolom sebagai variabel independen dan 1 kolom sebagai variabel dependen. Berikut Adalah link dataset nya [Social_Network_Ads.csv](#)

1. Import data menggunakan library pandas dan melakukan pengecekan info

```
1 dataset = pd.read_csv('Social_Network_Ads.csv')
2 X = dataset.iloc[:, :-1].values
3 y = dataset.iloc[:, -1].values
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 400 entries, 0 to 399
Data columns (total 3 columns):
#   Column          Non-Null Count  Dtype
---  ---
0    Age             400 non-null    int64
1  EstimatedSalary  400 non-null    int64
2   Purchased       400 non-null    int64
dtypes: int64(3)
memory usage: 9.5 KB
```

Outputnya akan menampilkan informasi tentang dataset, termasuk jumlah baris dan kolom, serta tipe data dari setiap kolom. Jika semua kolom berisi tipe data integer, maka outputnya akan mencantumkan informasi tersebut.

2. Memisahkan dataset menjadi Training set dan Test set

```
1 from sklearn.model_selection import train_test_split
2 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.25, random_state = 0)
```

Memisahkan dataset menjadi Training set dan Test Set

3. Feature Scaling

teknik yang digunakan dalam pra-pemrosesan data untuk menormalkan rentang fitur dalam data. Ini sangat penting dalam banyak algoritma machine learning karena beberapa algoritma sensitif terhadap skala data. Berikut adalah beberapa kegunaan utama dari feature scaling:

- Meningkatkan Konvergensi Algoritma
- Menghindari Dominasi Fitur Tertentu
- Meningkatkan Kinerja Model
- Mempercepat Proses Pelatihan
- Meningkatkan Stabilitas Numerik

```

1 from sklearn.preprocessing import StandardScaler
2 sc = StandardScaler()
3 X_train = sc.fit_transform(X_train)
4 X_test = sc.transform(X_test)

```

Berikut adalah data setelah melewati proses Feature Scaling

```

[[ 0.58164944 -0.88670699]
 [-0.60673761  1.46173768]
 [-0.01254409 -0.5677824 ]
 [-0.60673761  1.89663484]
 [ 1.37390747 -1.40858358]
 [ 1.47293972  0.99784738]
 [ 0.08648817 -0.79972756]
 [-0.01254409 -0.24885782]
 [-0.21060859 -0.5677824 ]
 [-0.21060859 -0.19087153]
 [-0.30964085 -1.29261101]
 [-0.30964085 -0.5677824 ]
 [ 0.38358493  0.09905991]
 [ 0.8787462  -0.59677555]
 [ 2.06713324 -1.17663843]
 [ 1.07681071 -0.13288524]
 [ 0.68068169  1.78066227]
 [-0.70576986  0.56295021]
 [ 0.77971394  0.35999821]
 [ 0.8787462  -0.53878926]
 [-1.20093113 -1.58254245]
 [ 2.1661655  0.93986109]
 [-0.01254409  1.22979253]]

```

X_train

4. Melatih model Decision Tree Classification model pada Training set

Model akan memulai dengan seluruh dataset dan mencoba membagi data menjadi dua kelompok berdasarkan kriteria tertentu dalam hal ini, menggunakan entropi untuk mengukur impurity. Proses pembagian ini terus berlanjut secara rekursif, membagi dataset menjadi subset yang lebih kecil pada setiap level hingga kondisi penghentian tertentu terpenuhi seperti jumlah maksimum kedalaman pohon atau jumlah minimum sampel dalam node. Entropi digunakan untuk mengukur ketidakmurnian dari node. Node yang lebih murni memiliki entropi yang lebih rendah. Model akan memilih split yang mengurangi entropi secara signifikan.

5. Memprediksi hasil Test set

predicting the test set results adalah langkah penting dalam proses pengembangan model machine learning. Ini melibatkan model yang telah dilatih pada train set untuk membuat prediksi pada data yang baru atau belum pernah dilihat sebelumnya

Tujuan dari Predicting the Test Set Results:

1. Evaluasi Kinerja Model
2. Mengukur Generalisasi
3. identifikasi Kesalahan dan Perbaikan
4. Validasi Model
5. Benchmarking

6. Confusion Matrix

Confusion matrix berfungsi untuk mengevaluasi kinerja model klasifikasi. Ini adalah tabel yang memungkinkan kita untuk memvisualisasikan kinerja algoritma klasifikasi, terutama ketika kita memiliki lebih dari dua kelas. Setiap kolom dari matriks ini mewakili prediksi dari model, sedangkan setiap baris mewakili instance aktual dari kelas.

Komponen Confusion Matrix:

		Actual class	
		P	N
Predicted class	P	TP	FP
	N	FN	TN

Confusion Matrix

-**True Negative (TN)**: Jumlah instance negatif yang diprediksi dengan benar sebagai negatif.

-**False Positive (FP)**: Jumlah instance negatif yang diprediksi secara salah sebagai positif (false alarm).

-**False Negative (FN)**: Jumlah instance positif yang diprediksi secara salah sebagai negatif (miss).

-**True Positive (TP)**: Jumlah instance positif yang diprediksi dengan benar sebagai positif.

Pada artikel penelitian ini menggunakan Akurasi (Accuracy)

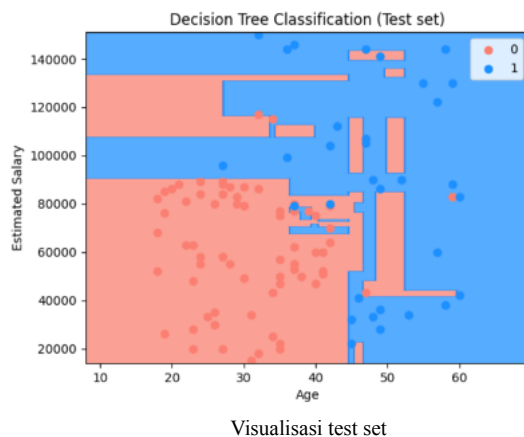
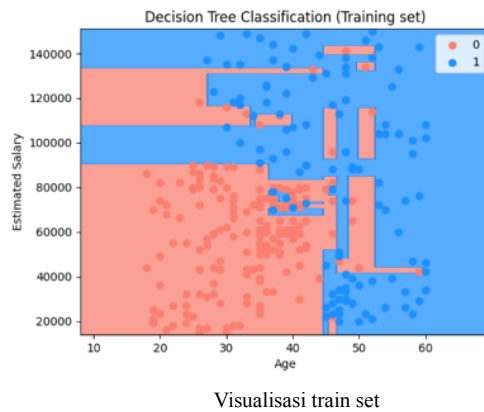
Pada kode program menunjukkan Akurasi yang cukup tinggi sebanyak 91%

```

[[62  6]
 [ 3 29]]
0.91

```

7. Visualisasi Data



Pada langkah ini, kita menampilkan visualisasi data yang diperoleh dari training set dan test set. Visualisasi ini membantu kita memahami cara model pembelajaran mesin memisahkan kelas-kelas yang berbeda dalam ruang fitur. Ini memungkinkan kita untuk melihat secara langsung bagaimana model mempelajari pola dari data pelatihan

REFERENSI

- [1] Al-jabery, K. K., Obafemi-Ajayi, T., Olbricht, G. R., & Wunsch II, D. C. (2020). Data analysis and machine learning tools in MATLAB and Python. *Computational Learning Approaches to Data Analytics in Biomedical Applications*, 231–290. <https://doi.org/10.1016/B978-0-12-814482-4.00009-7>
- [2] Awujoola, O. J., Ogwueleka, F. N., Odion, P. O., Awujoola, A. E., & Adelegan, O. R. (2022). Genomic data science systems of Prediction and prevention of pneumonia from chest X-ray images using a two-channel dual-stream convolutional neural network. *Data Science for Genomics*, 217–228. <https://doi.org/10.1016/B978-0-323-98352-5.00013-6>
- [3] *Confusion Matrix - an overview | ScienceDirect Topics*. (n.d.). Retrieved June 4, 2024, from <https://www.sciencedirect.com/topics/engineering/confusion-matrix>
- [4] Demir, F. (2022). Deep autoencoder-based automated brain tumor detection from MRI data. *Artificial Intelligence-Based Brain-Computer Interface*, 317–351. <https://doi.org/10.1016/B978-0-323-91197-9.00013-8>
- [5] Dutta, P., Paul, S., Cengiz, K., Anand, R., & Majumder, M. (2022). A predictive method for emotional sentiment analysis by machine learning from electroencephalography of brainwave data. *Implementation of Smart Healthcare Systems Using AI, IoT, and Blockchain*, 109–130. <https://doi.org/10.1016/B978-0-323-91916-6.00008-4>
- [6] Kulkarni, A., Chong, D., & Batarseh, F. A. (2020). Foundations of data imbalance and solutions for a data democracy. *Data Democracy: At the Nexus of Artificial Intelligence, Software Development, and Knowledge Engineering*, 83–106. <https://doi.org/10.1016/B978-0-12-818366-3.00005-8>
- [7] Nailil Amani, N., & Hayati, U. (2024). PENGGUNAAN ALGORITMA DECISION TREE UNTUK PREDIKSI PRESTASI SISWA DI SEKOLAH DASAR NEGERI 3 BAYALANGU KIDUL. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Issue 1).
- [8] Radhi, M., Ryan Hamonangan Sitompul, D., Hamonangan Sinurat, S., & Indra, E. (2021). ANALISIS BIG DATA DENGAN METODE EXPLORATORY DATA ANALYSIS (EDA) DAN METODE VISUALISASI MENGGUNAKAN JUPYTER NOTEBOOK. *Jurnal Sistem Informasi Dan Ilmu Komputer Prima*, 4(2).
- [9] Reddi, S., & Eswar, G. V. (2020). Fake news in social media recognition using Modified Long Short-Term Memory network. *Security in IoT Social Networks*, 205–227. <https://doi.org/10.1016/B978-0-12-821599-9.00009-1>
- [10] Satapathy, S., Loganathan, D., Bhoi, A. K., & Barsocchi, P. (2022). Effects of EEG-sleep irregularities and its behavioral aspects: review and analysis. *Cognitive and Soft Computing Techniques for the Analysis of Healthcare Data*, 239–267. <https://doi.org/10.1016/B978-0-323-85751-2.00009-8>
- [11] Sharma, D. K., Chatterjee, M., Kaur, G., & Vavilala, S. (2022). Deep learning applications for disease diagnosis. *Deep Learning for Medical Applications with Unique Data*, 31–51. <https://doi.org/10.1016/B978-0-12-824145-5.00005-8>
- [12] Sinambela, D. P., Naparin, H., Zulfadhilah, M., & Hidayah, N. (2023). Implementasi Algoritma Decision Tree dan Random Forest dalam Prediksi Perdarahan Pascasalin. *Jurnal Informasi Dan Teknologi*, 58–64.
- [13] Singh, P., Singh, N., Singh, K. K., & Singh, A. (2021). Diagnosing of disease using machine learning. *Machine Learning and the Internet of Medical Things in Healthcare*, 89–111. <https://doi.org/10.1016/B978-0-12-821229-5.00003-3>
- [14] Tiwari, A. (2022). Supervised learning: From theory to applications. *Artificial Intelligence and Machine Learning for EDGE Computing*, 23–32. <https://doi.org/10.1016/B978-0-12-824054-0.00026-5>