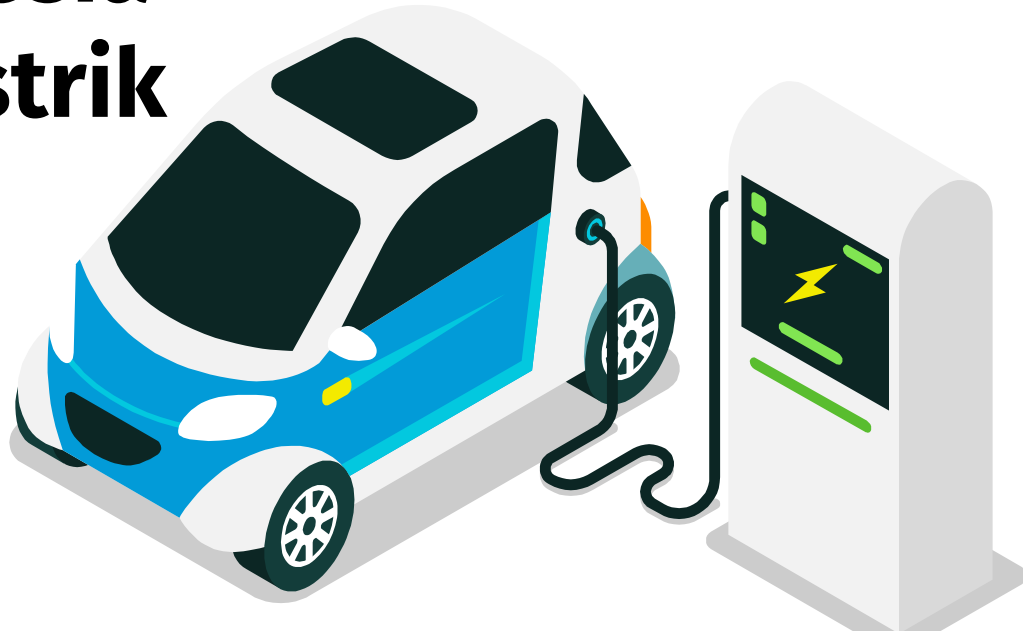


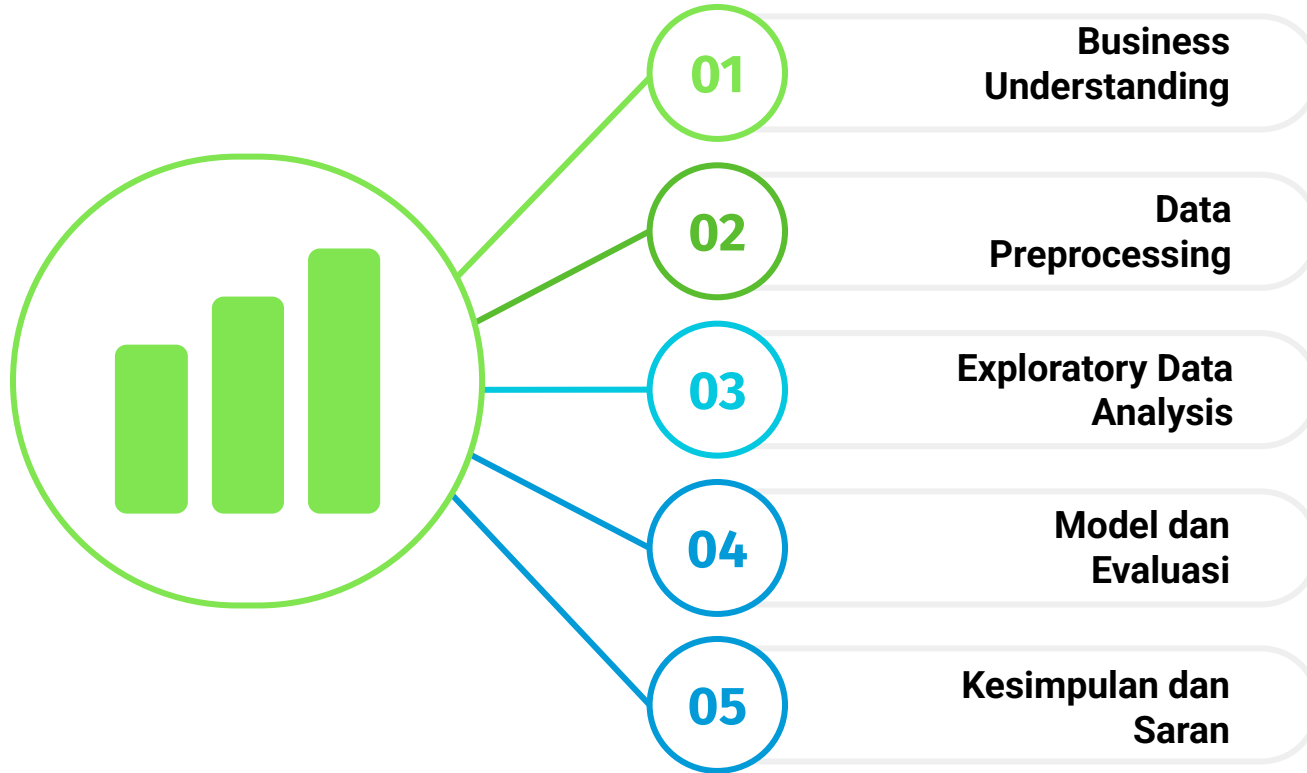
Klasifikasi Sentimen Masyarakat Indonesia terhadap Mobil Listrik

By White Fans Team

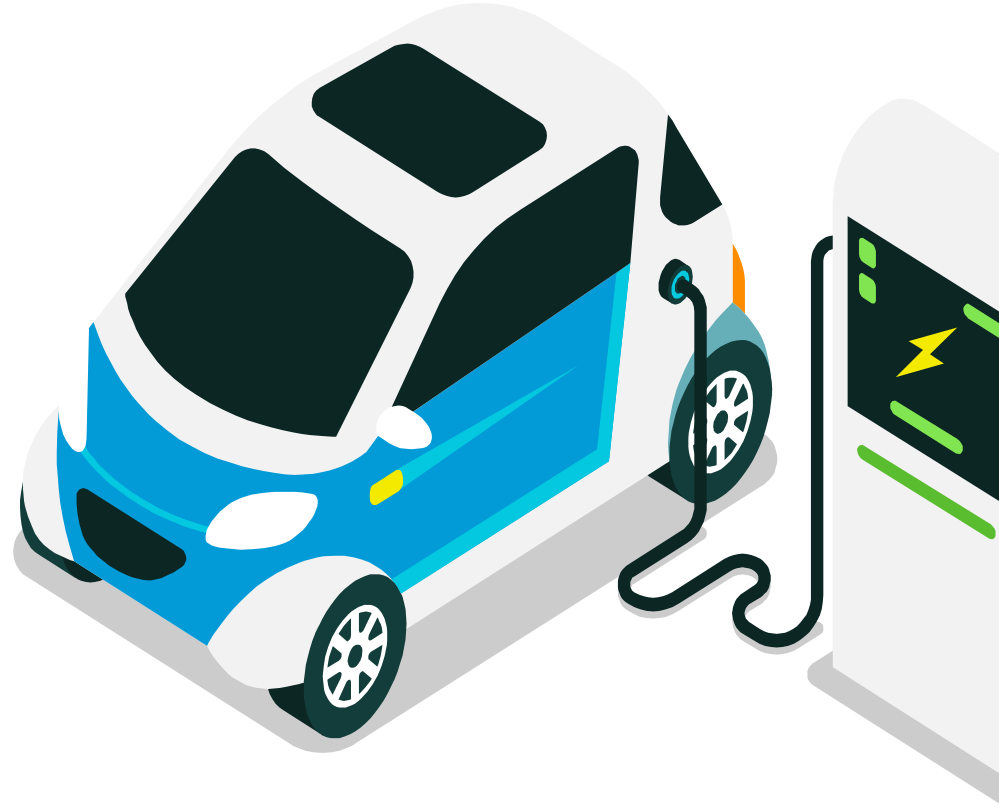
Final Data Competition ISFEST 2024



Daftar Isi

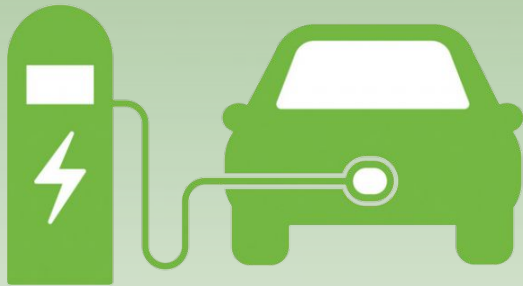


Business Understanding



Melalui Peraturan Presiden (Perpres) Nomor 55 Tahun 2019, Indonesia didorong untuk mempercepat program kendaraan bermotor listrik berbasis baterai

TARGET :
2 JUTA UNIT TERJUAL
pada tahun 2030

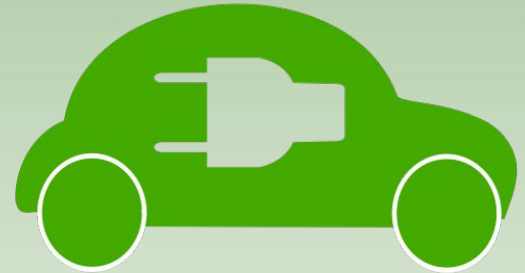


Kementrian ESDM (2024)


**PENJUALAN MOBIL
LISTRIK SEBANYAK**

20.414 UNIT


Mentri ESDM, Arifin Tasrif (2023)



Penyebab:




CNBC Indonesia > Research > Berita Research




Subsidi Kendaraan Listrik Hanya Dinikmati Orang Kaya, Ini Hitungannya

Muhammad Reza Ilham Taufani, CNBC Indonesia
22 February 2024 19:20




PEMERINTAH BAKAL "PAKSA" WARGA PAKAI KENDARAAN LISTRIK

Persulit Aturan Kendaraan BBM, Cara Pemerintah "Paksa" Warga Beralih ke Kenda...




Studi: Pengguna Mobil Listrik Ternyata Memiliki Lebih Banyak Keluhan Dibanding Mobil Bensin


8 Jan 2024




Alasan 80 Persen Masyarakat Tolak Subsidi Mobil Listrik




Perlu Reformasi Kebijakan untuk Dorong Keberhasilan Insentif Kendaraan Listrik




Mobil Listrik yang Ramah Lingkungan Ternyata Bisa Merusak Lingkungan - Kompasiana.com



RINCIAN ANGGARAN SUBSIDI MOTOR DAN MOBIL LISTRIK



APA KABAR INFRASTRUKTUR KENDARAAN LISTRIK DI INDONESIA



Warganet Banyak yang Tak Setuju Mobil Listrik Disubsidi, Ini Alasannya

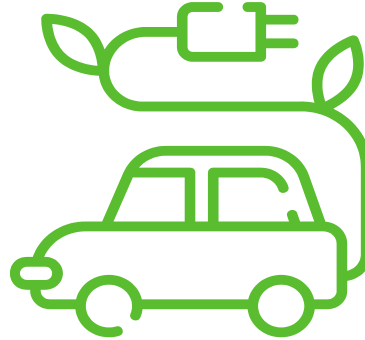
Tim detikcom - detikOto
Kamis, 25 Mei 2023 14:10 WIB

Business Understanding

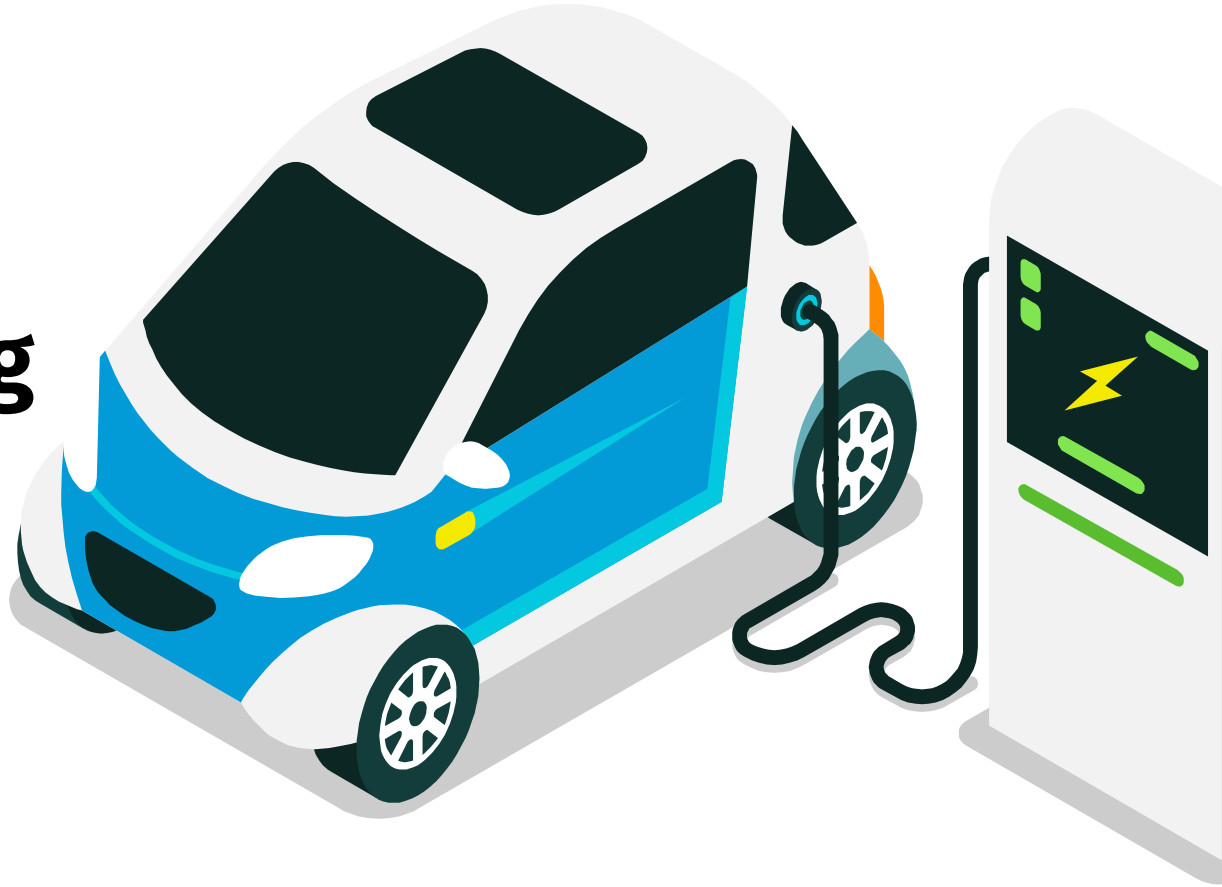
**Determine Business
Objectives**

**Determine Data
Mining Goals**

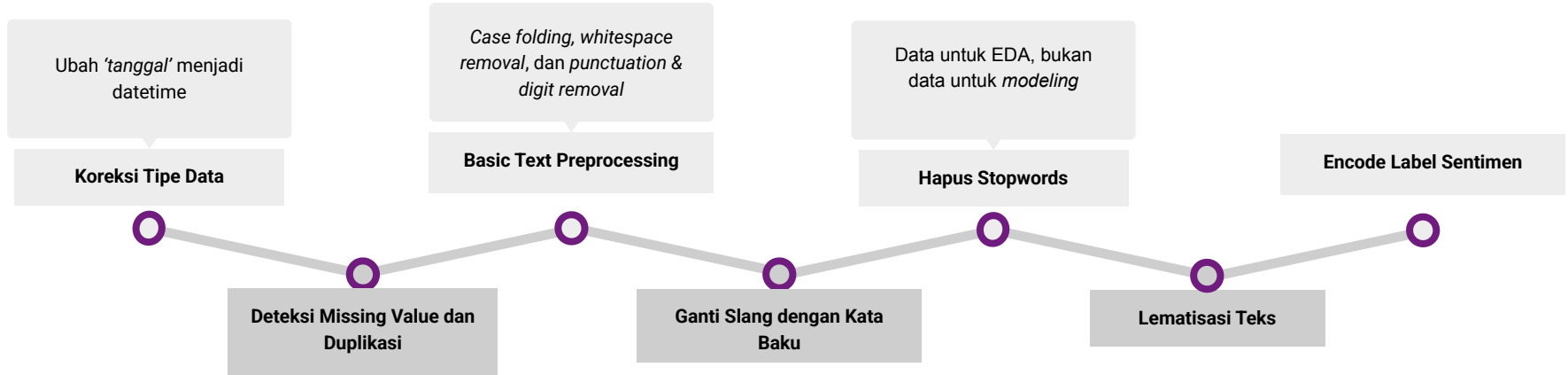
Produce Project Plan



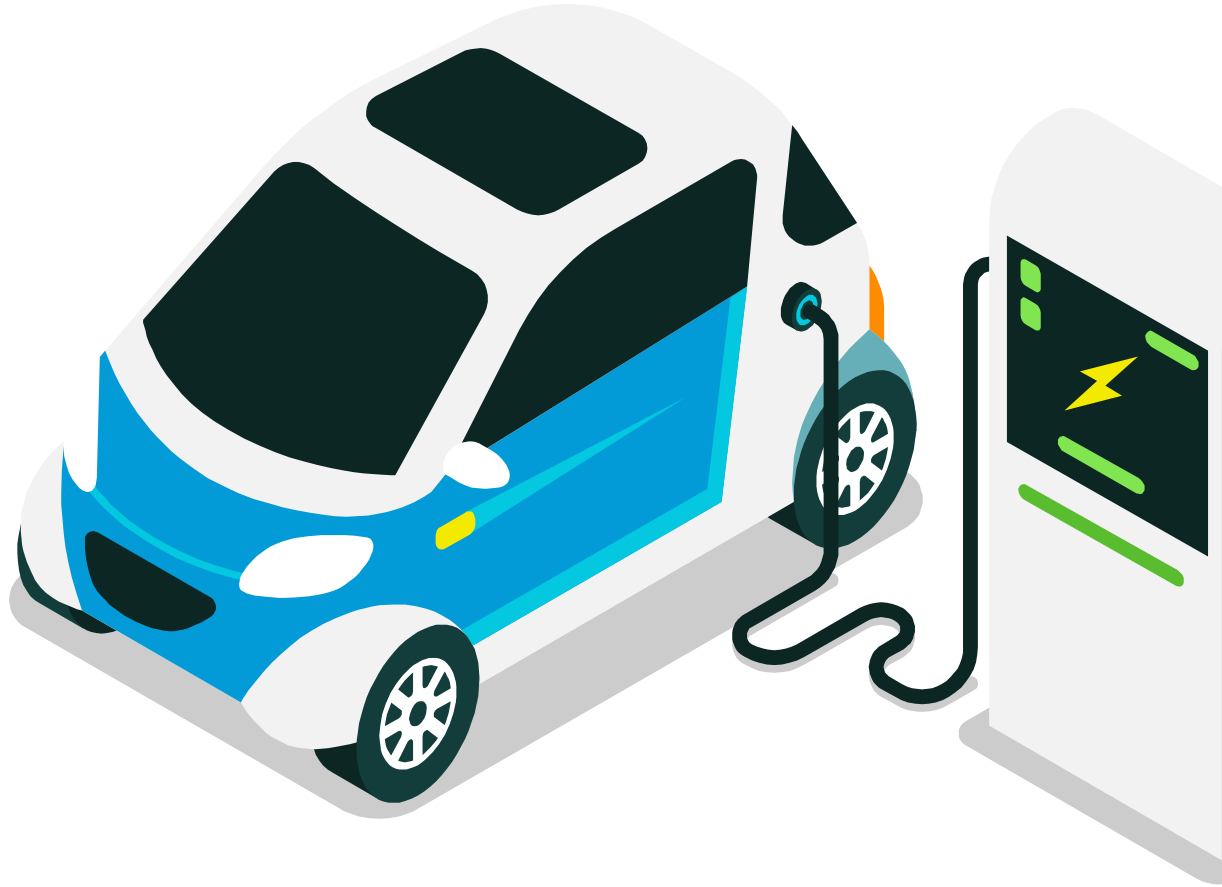
Data Preprocessing



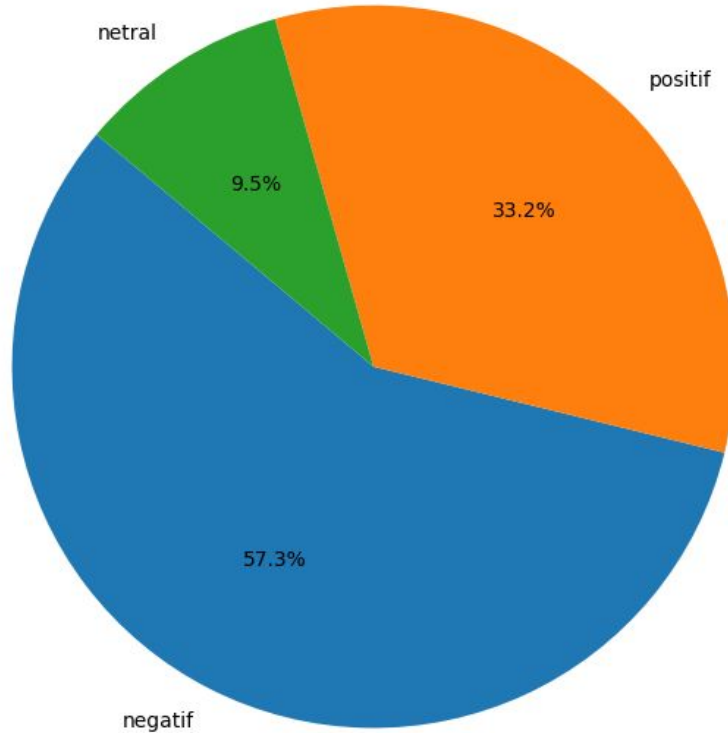
Data Preprocessing



Exploratory Data Analysis



Distribusi Tipe Sentimen



Positif

504

Neutral

114

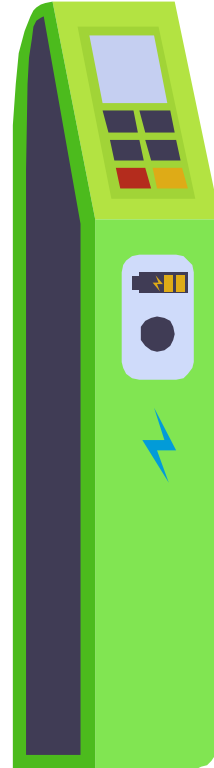
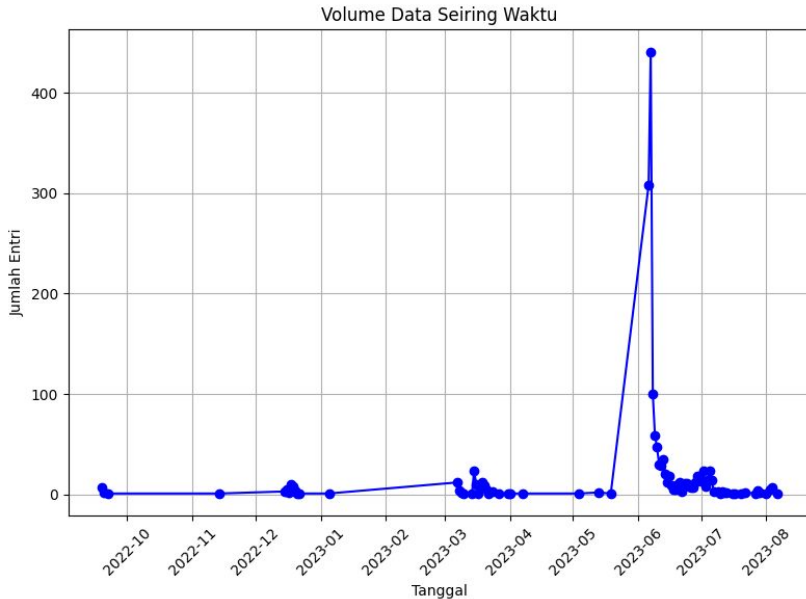
Negatif

869

Data menunjukkan ketidakseimbangan sentimen, dengan sentimen negatif jauh lebih dominan dibandingkan netral atau positif. Hal ini akan dibahas lebih lanjut dalam segmen **Model and Evaluation**.



Distribusi Komentar terhadap waktu



Top
Date

Date	Count
2023-06-07	441
2023-06-06	308
2023-06-08	100

Top
Month

Month	Count
2023-06	1235
2023-07	123
2023-03	93

Terdapat lonjakan signifikan dalam jumlah komentar antara tanggal 6 hingga 8 Juni 2023. Selain itu, secara keseluruhan, bulan Juni mencatatkan jumlah komentar tertinggi dalam dataset ini.

Kata yang paling banyak muncul

Word Cloud: positif



5 Kata yang Paling Sering Muncul untuk Sentimen 'positif':

- 1) harga: 240
- 2) mahal: 117
- 3) laku: 80
- 4) ev: 77
- 5) juta: 77

Word Cloud: negatif



5 Kata yang Paling Sering Muncul untuk Sentimen 'negatif':

- 1) harga: 195
- 2) subsidi: 171
- 3) rakyat: 150
- 4) indonesia: 146
- 5) ev: 129

Word Cloud: netral



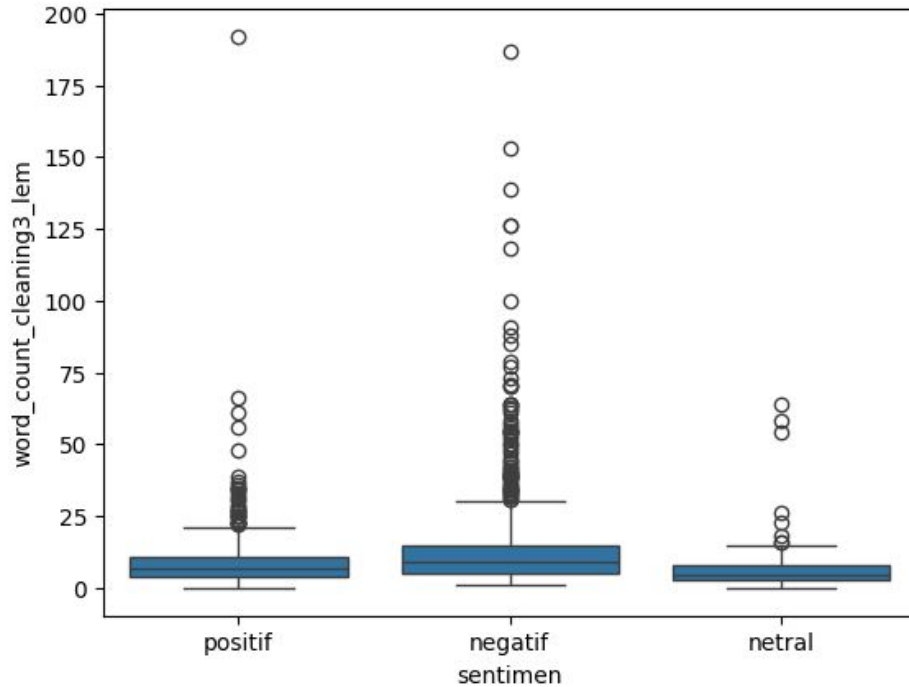
5 Kata yang Paling Sering Muncul untuk Sentimen 'netral':

- 1) subsidi: 43
- 2) juta: 14
- 3) mahal: 12
- 4) laku: 12
- 5) indonesia: 12

Analisis dari kata-kata yang paling sering muncul menunjukkan bahwa diskusi berfokus pada aspek harga dan subsidi.



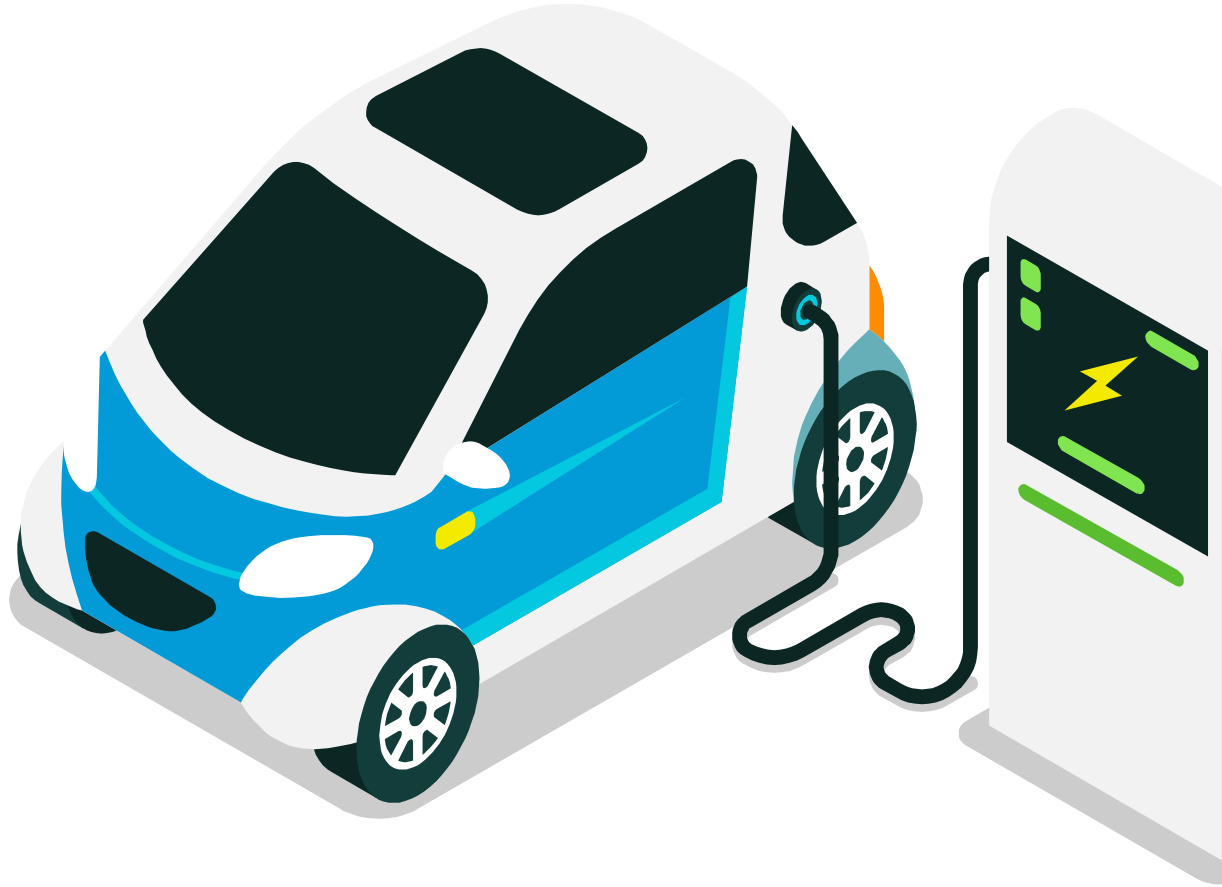
Frekuensi Banyak Kata Pada Setiap Komentar



Dari plot, kita dapat melihat bahwa terdapat minimnya perbedaan pada tipe sentimen positif dan negatif. Sentimen netral memiliki kata-kata yang jauh lebih dikit.



Model dan Evaluasi



Model Klasifikasi

Random
Forest

01

Multinomial Naive
Bayes Classifier

02

Gaussian Naive
Bayes classifier

03

Support Vector
Classifier

04

XGBoost (Extreme
Gradient Boosting)

05



06

Logistic Regression

07

Decision Tree Classifier

08

Bagging Classifier

09

Adaptive Boosting
Classifier

10

Gradient Boosting
Classifier

Problem: Imbalance Data

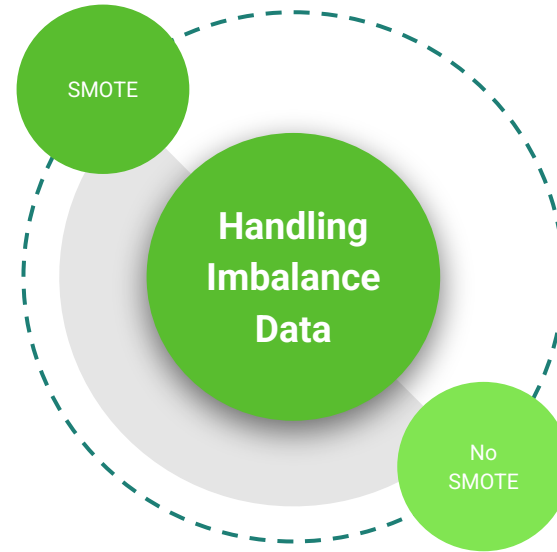
Sentimen	Banyak Data
Positif	504
Netral	144
Negatif	869

Tabel di atas adalah distribusi dari jumlah data tiap sentimen.

Terdapat kondisi **data yang tidak seimbang** (*imbalance data*).

Imbalance data menyebabkan **performa** dan **akurasi** model yang **menurun** (Hawari, 2024)

Handling Imbalance Data: SMOTE (Synthetic Minority Over-sampling Technique)



Akan dibandingkan 10 model sebelumnya
1) **dengan** dan 2) **tanpa SMOTE** melalui beberapa metrik penilaian



Model



2. Vektorisasi Teks: TF-IDF

Vektorisasi untuk mengubah data teks komentar menjadi vektor numerik.

4. Training Model

Data akan di-*train* dengan 10 model sebelumnya



1.Split Data

Data komentar dibagi menjadi data *train* dan data *test*. Digunakan **test size** sebesar 30%



3. Imbalance Data

Akan dicoba 1) **dengan SMOTE** dan 2) **tanpa SMOTE**



Next Slide

Electric Car Infographics

Previous Slide

6. SMOTE vs No SMOTE

Memilih model 1) **dengan** atau 2) **tanpa SMOTE** berdasarkan metrik evaluasi

8. Hyperparameter Tuning

Hyperparameter Tuning dengan GridSearchCV pada model yang dipilih sebelumnya

5. Evaluasi

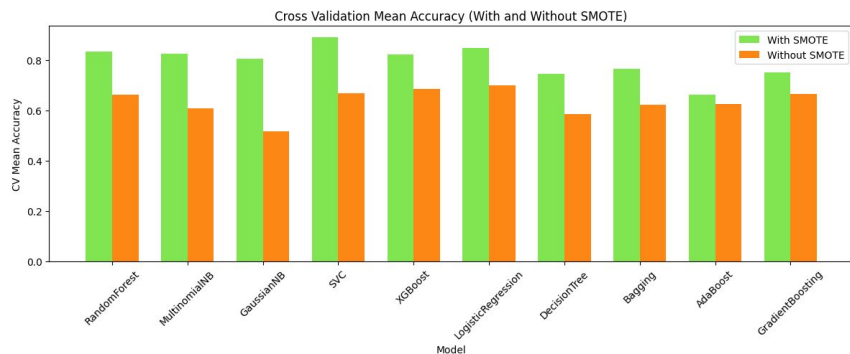
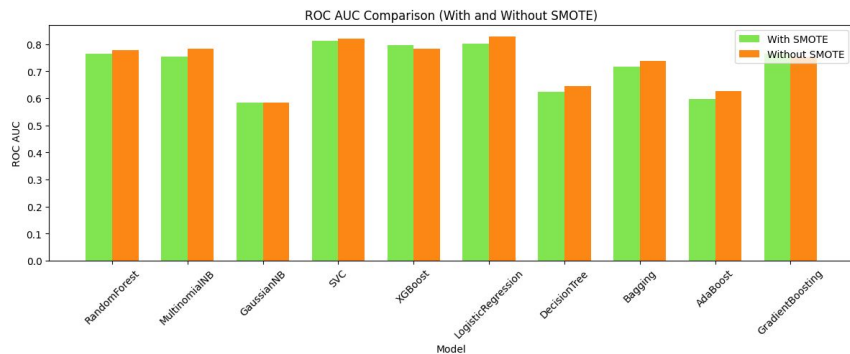
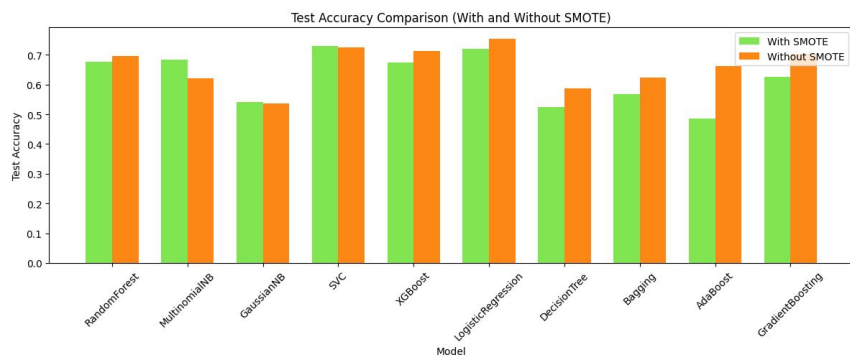
Akan digunakan beberapa metrik evaluasi: *accuracy, precision, recall, f1-score, cross-validation, ROC AUC*

7. Pilih Model Klasifikasi

Akan dipilih 1 dari 10 model klasifikasi berdasarkan metrik evaluasi

9. Training dengan Parameter Hasil GridSearch CV

Akan dipilih 1 dari 10 model klasifikasi berdasarkan metrik evaluasi



SMOTE vs No SMOTE

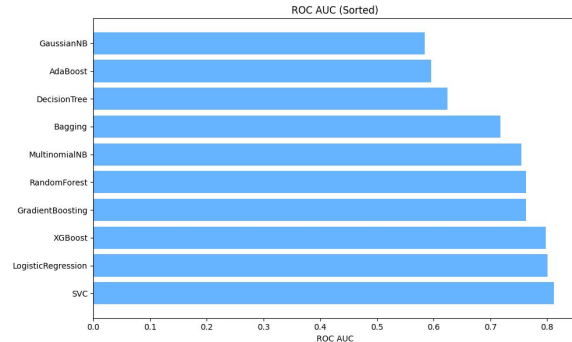
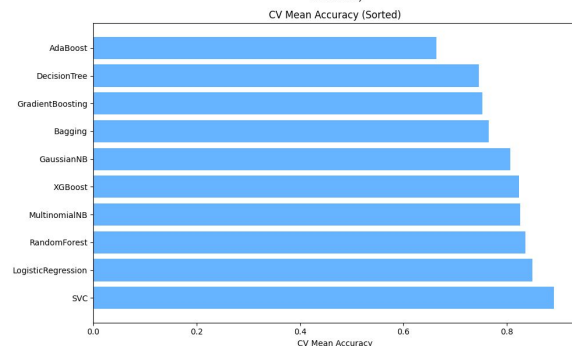
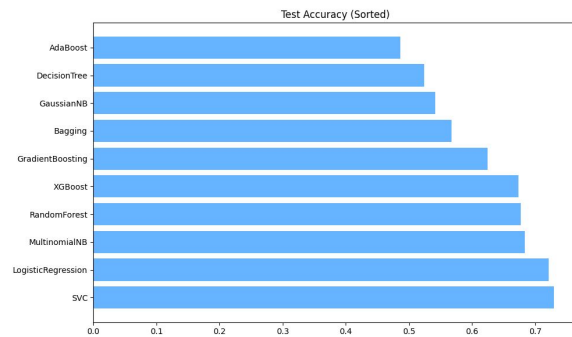
Tanpa SMOTE:

1. **Akurasi lebih tinggi pada beberapa algoritma**
2. **CV Mean Accuracy lebih rendah pada semua algoritma**

Keputusan:

1. SMOTE diperlukan untuk meningkatkan kinerja model secara umum.
2. **Pilih model dengan SMOTE** untuk performa yang lebih stabil.

Dipilih Model dengan SMOTE



Memilih 1 dari 10 Algoritma dengan SMOTE

Evaluasi dilakukan berdasarkan **Accuracy**, **CV Mean Accuracy**, **Precision**, **Recall**, dan **F1-Score**.

SVM Classifier dipilih sebagai **model terbaik** dalam mengklasifikasikan sentimen pada dataset ini.

Evaluasi

Evaluasi	Nilai
Test Accuracy	73.03%
CV Mean Accuracy	89.17%
CV Std Dev	0.0834
ROC AUC	0.8126

Classification Report

Class	Precision	Recall	F1-Score	Support
Class 0	0.71	0.96	0.82	269
Class 1	1.00	0.05	0.09	42
Class 2	0.79	0.50	0.61	145

Overall Metrics	Nilai
Accuracy	73.03%
Macro Avg F1-Score	0.51
Weighted Avg F1-Score	0.69

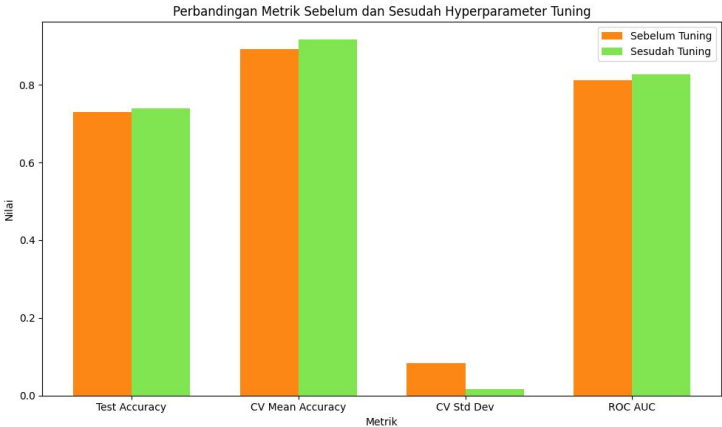
Dipilih Model SVM Classifier dengan SMOTE

Perbandingan Model SVC dengan SMOTE Sebelum dan Sesudah Hyperparameter Tuning

Setelah tuning, model SVC menghasilkan performa yang lebih baik pada semua metrik, khususnya pada CV Mean Accuracy dan ROC AUC.

Stabilitas model meningkat dengan CV Std Dev yang lebih rendah.

Perbandingan Evaluasi



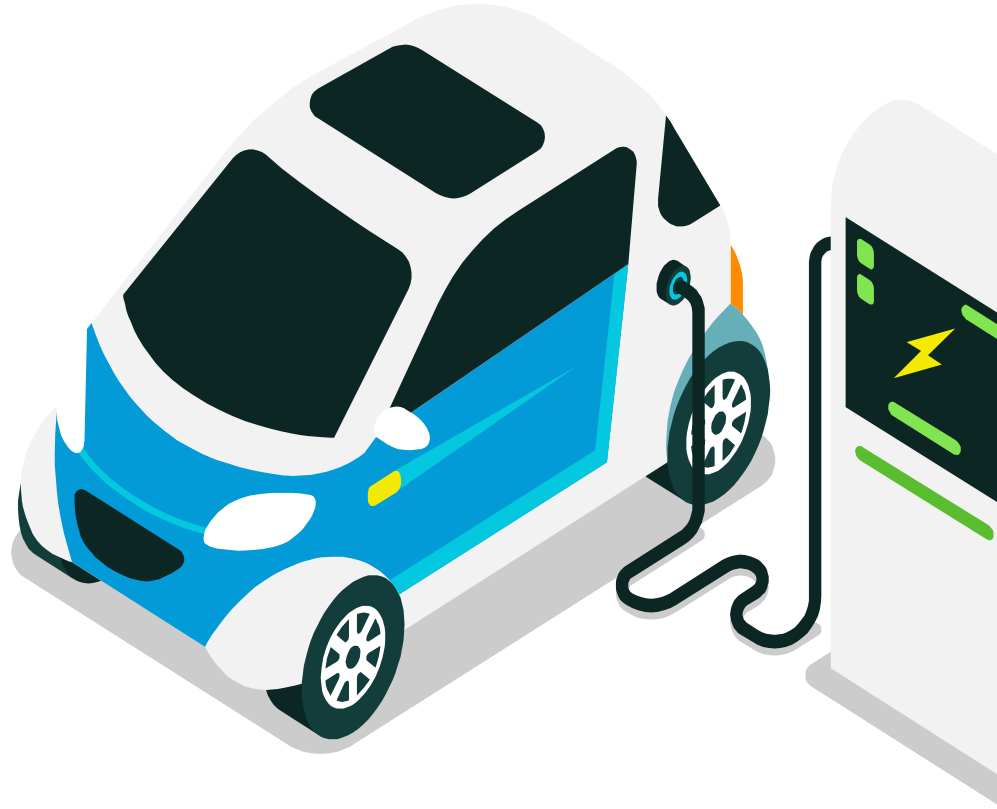
Model SVC Setelah Tuning Menunjukkan Performa Lebih Baik

Evaluasi	Nilai
Test Accuracy	73.03%
CV Mean Accuracy	89.17%
CV Std Dev	0.0834
ROC AUC	0.8126

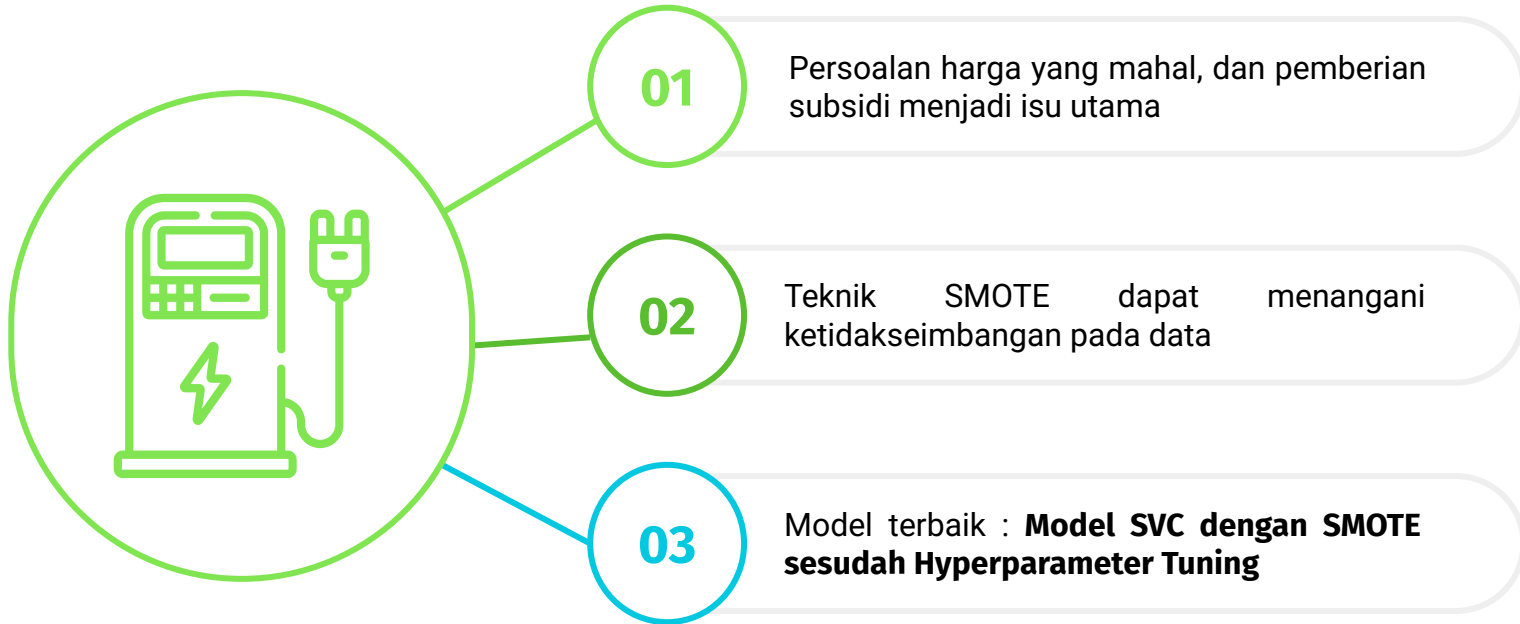
Perbandingan Classification Report

Class	Metric	Sebelum Tuning	Setelah Tuning
Class 0	Precision	0.71	0.79
	Recall	0.96	0.88
	F1-Score	0.82	0.83
Class 1	Precision	1.00	0.23
	Recall	0.05	0.12
	F1-Score	0.09	0.16
Class 2	Precision	0.79	0.72
	Recall	0.50	0.66
	F1-Score	0.61	0.69
Overall	Accuracy	0.73	0.74
Macro Avg	Precision	0.83	0.58
	Recall	0.50	0.55
	F1-Score	0.51	0.56
Weighted Avg	Precision	0.76	0.71
	Recall	0.73	0.74
	F1-Score	0.69	0.72

CONCLUSION & SUGGESTION



KESIMPULAN



SARAN

IMPLEMENTASI TEKNIK SMOTE

Pertimbangkan untuk menggunakan teknik SMOTE dalam menangani ketidakseimbangan kelas pada data.

EKSPLORASI MODEL SVM

Eksplorasi penggunaan model SVM yang lebih kompleks dapat meningkatkan daya prediktif dan keakuratan analisis sentimen

MODEL DEEP LEARNING

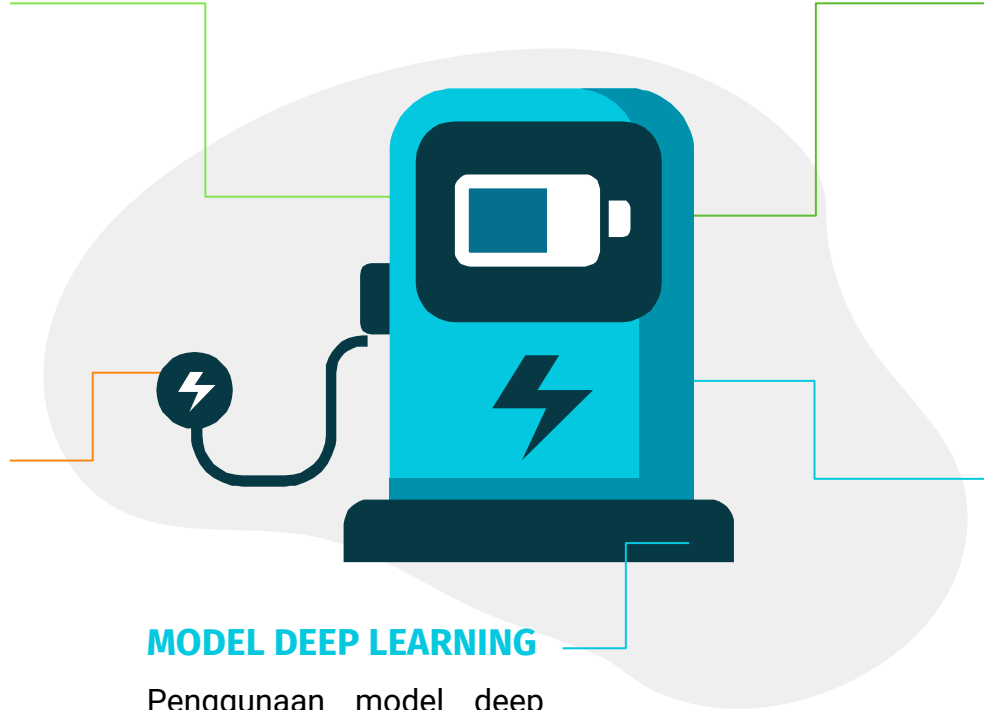
Penggunaan model deep learning untuk analisis sentimen yang lebih akurat

PERLUAS PENGUNAAN MODEL SVM

Perluas penggunaan model SVM dalam analisis sentimen kendaraan listrik dengan menguji data baru secara berkala

PEMANTAUAN SENTIMEN

Pantau sentimen secara berkelanjutan untuk menangkap perubahan tren dan masalah baru.



Daftar Pustaka

- [1] Hawari, I. F., Najib, M. K., Nurdianti, S., Marpaung, Y. F. Y., Kusumawati, N., Nurfadila, M., ... & Hernawan, B. F. (2024). PENGARUH TEKNIK OVERSAMPLING PADA ALGORITMA MACHINE LEARNING DALAM KLASIFIKASI BODY MASS INDEX (BMI). Jurnal Riset dan Aplikasi Matematika (JRAM), 8(1), 51-68.
- [2] Suryani, S., Fayyad, M. F., Savra, D. T., Kurniawan, V., & Estanto, B. H. (2023). Sentiment Analysis of Towards Electric Cars using Naive Bayes Classifier and Support Vector Machine Algorithm. Public Research Journal of Engineering, Data Technology and Computer Science, 1(1), 1-9.
- [3] Suresha, H. P., & Kumar Tiwari, K. (2021). Topic Modeling and Sentiment Analysis of Electric Vehicles of Twitter Data. Asian Journal of Research in Computer Science, 12(2), 13-29.
- [4] Jiang, X., & Everts, J. (2021). Making sense of electrical vehicle discussions using sentiment analysis on closely related news and user comments. arXiv preprint arXiv:2112.12327.





**Terima
Kasih.**