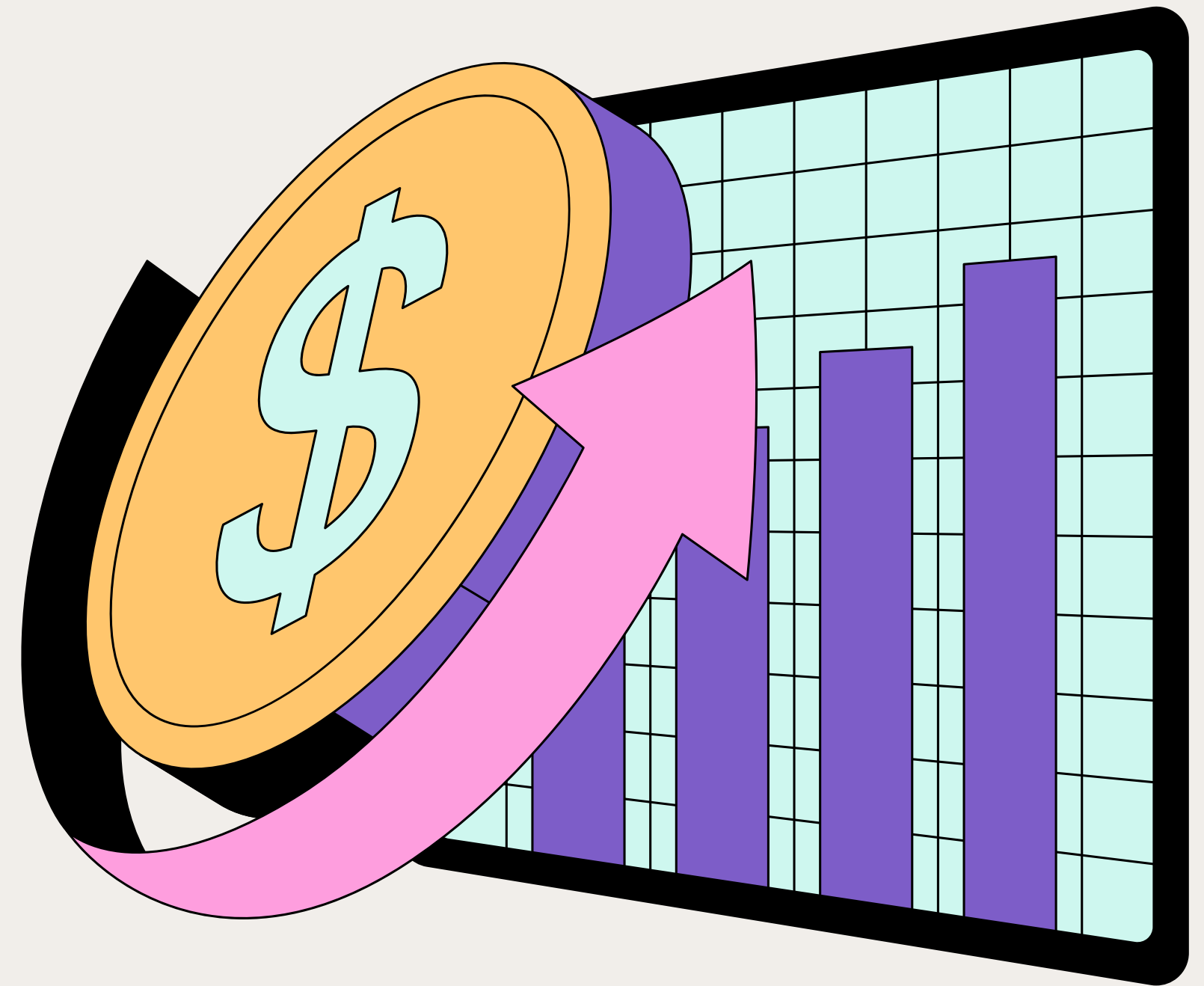


Stock Price Sentiment

2025



Dosen Pengampu:

Nisa Hanum Harani, S.T., M.T., CDSP., SFPC



Daftar Anggota

- **M.muhammad Azka Nuril**
Set env dan optimalization
- **Gaizka Wisnu Prawira**
Validation dataset dan EDA
- **Muhammad Fathir**
Pemodelan histories dan gabungan
- **Salwa Mutfia Indah Putri**
Pustaka dan preprocessing

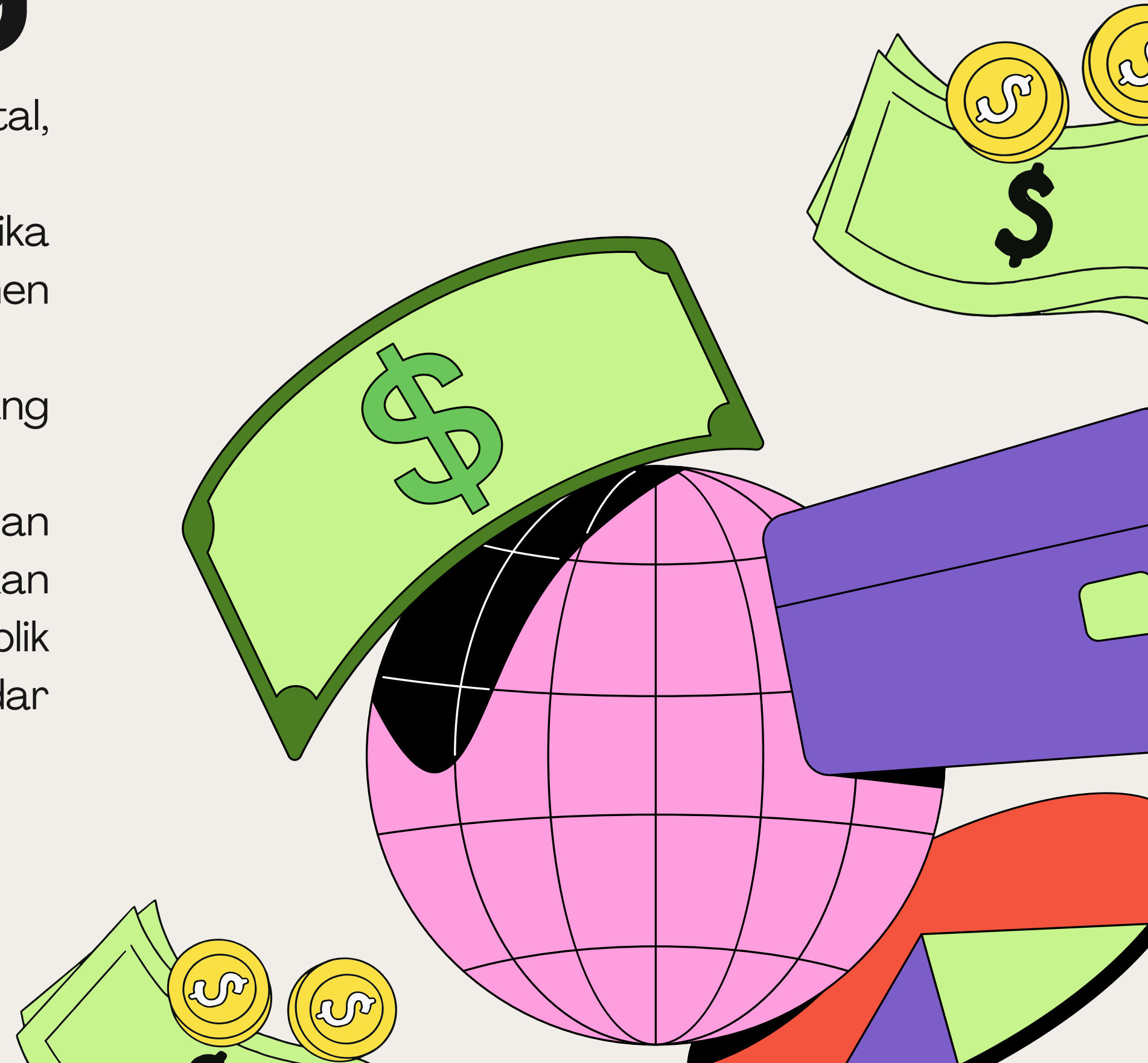
Daftar Isi



- **Latar Belakang**
- **Tinjauan Pustaka**
- **Validation Dataset**
- **Pre Processing dan Feature Engineering**
- **Pemodelan Historis dan Gabungan**
- **Optimation**
- **Visualization**
- **Kesimpulan**

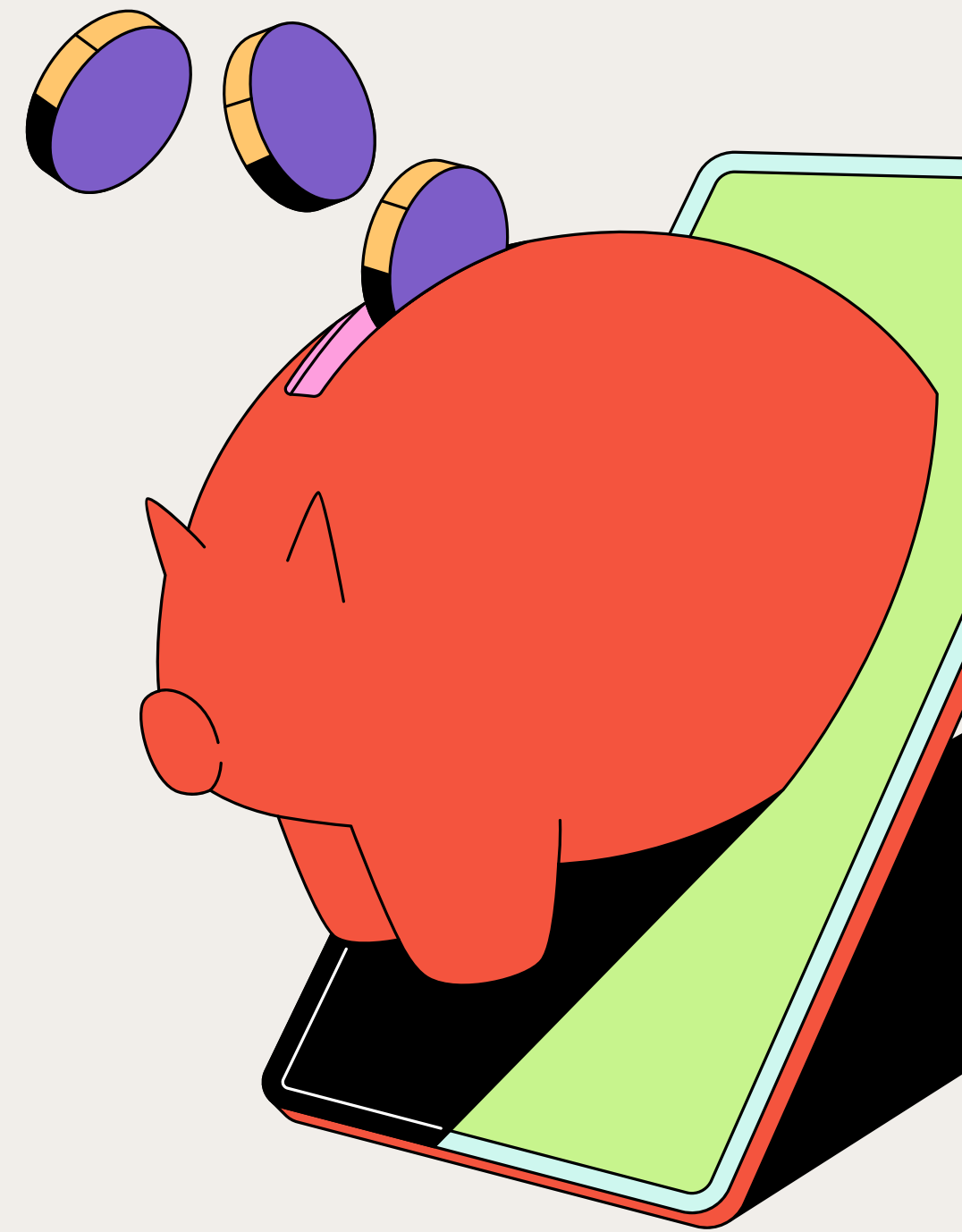
Latar Belakang

- Harga saham dipengaruhi oleh faktor fundamental, teknikal, dan sentimen publik.[2]
- Data historis saja tidak cukup menangkap dinamika psikologis pasar sehingga dibutuhkan sentimen sosial untuk menganalisis harga saham.[5][7][8]
- Twitter menjadi sumber informasi real-time yang mencerminkan opini pasar.[4][5][6]
- Penelitian ini berfokus pada saham BBRI dengan pendekatan machine learning untuk menghasilkan jawaban, “Seberapa pengaruh sentimen publik terhadap prediksi harga saham dibanding sekedar data historis?”.



Tinjauan Pustaka

- Skor sentimen didasarkan menggunakan RobeRTA
- Algoritma yang digunakan::
 - Random Forest
 - XGBoost
 - LogReg
 - SVC
 - MLP
- Evaluasi: dalam bentuk rata-rata dan standar deviasi: *Accuracy, Precision, Recall, F1-Score, ROC AUC Score, Confusion Matrix, Directional Accuracy,*





Validation Dataset:

- Data sentimen: 2232 tweet dari Twitter
- Data historis: 37 data dari Investing.com berdasarkan periode data sentimen
- Sinkronisasi berdasarkan tanggal

Exploratory Data Analysis (EDA):

- Analisis distribusi sentimen: positif, negatif, netral
- Korelasi antara sentimen publik dan perubahan harga

Validation Dataset

Pre Processing dan Feature Engineering

Preprocessing Data:

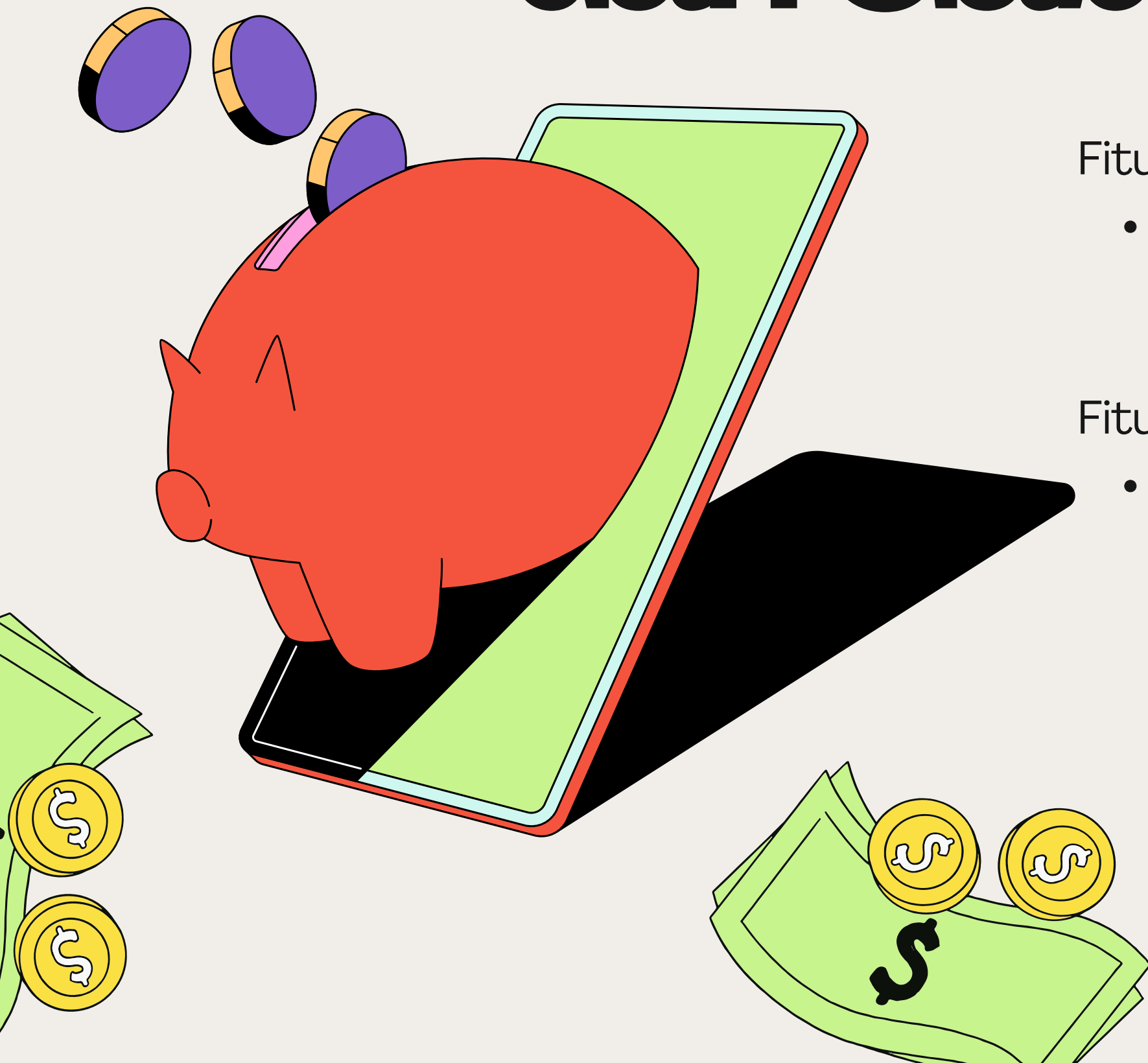
- Konversi format tanggal dan numerik
- Hilangkan missing value

Feature Engineering:

- Sentimen Ratio, Day Return, Lag1, Lag2, dan lain lain
- Target sebagai y



Pemodelan Historis dan Gabungan



Fitur Model Data Historis:

- Terakhir, Pembukaan, Tertinggi, Terendah, Vol., Perubahan%, Range, Day Return, Lag 1, Lag 2

Fitur Model Gabungan (Sentimen + Historis) :

- Terakhir, Pembukaan, Tertinggi, Terendah, Vol., Perubahan%, Avg Signed Sentiment, Count Positive, Count Negative, Count Neutral, Total Tweets, Range, Day Return, Sentiment Ratio, Tweet Intensity, Lag 1, Lag 2

Hasil Pemodelan

Historis

Rangkuman Evaluasi Model (Tahap Historis):												
	accuracy_mean	precision_mean	recall_mean	f1_mean	roc_auc_mean	directional_acc_mean	accuracy_std	precision_std	recall_std	f1_std	roc_auc_std	directional_acc_std
RandomForest	0.500000	0.333333	0.666667	0.440476	0.652778	0.500000	0.136083	0.249444	0.471405	0.319882	0.199343	0.136083
XGBoost	0.388889	0.250000	0.277778	0.261905	0.310185	0.555556	0.207870	0.204124	0.207870	0.204817	0.187142	0.078567
LogReg	0.500000	0.166667	0.222222	0.190476	0.717593	0.500000	0.136083	0.235702	0.314270	0.269374	0.066448	0.000000
SVM	0.500000	0.333333	0.666667	0.444444	0.634259	0.500000	0.136083	0.235702	0.471405	0.314270	0.371353	0.136083
MLP	0.555556	0.383333	0.666667	0.476190	0.555556	0.500000	0.207870	0.306413	0.471405	0.356348	0.218722	0.136083

Gabungan

Rangkuman Evaluasi Model (Tahap Gabungan):												
	accuracy_mean	precision_mean	recall_mean	f1_mean	roc_auc_mean	directional_acc_mean	accuracy_std	precision_std	recall_std	f1_std	roc_auc_std	directional_acc_std
RandomForest	0.611111	0.416667	0.666667	0.507937	0.833333	0.500000	0.207870	0.311805	0.471405	0.367487	0.117851	0.136083
XGBoost	0.666667	0.472222	0.666667	0.552381	0.675926	0.611111	0.235702	0.335640	0.471405	0.391288	0.124914	0.078567
LogReg	0.611111	0.416667	0.500000	0.452381	0.796296	0.555556	0.207870	0.311805	0.408248	0.351543	0.214367	0.078567
SVM	0.444444	0.166667	0.333333	0.222222	0.583333	0.555556	0.078567	0.235702	0.471405	0.314270	0.424918	0.078567
MLP	0.555556	0.361111	0.500000	0.419048	0.750000	0.500000	0.207870	0.306816	0.408248	0.350186	0.270031	0.136083

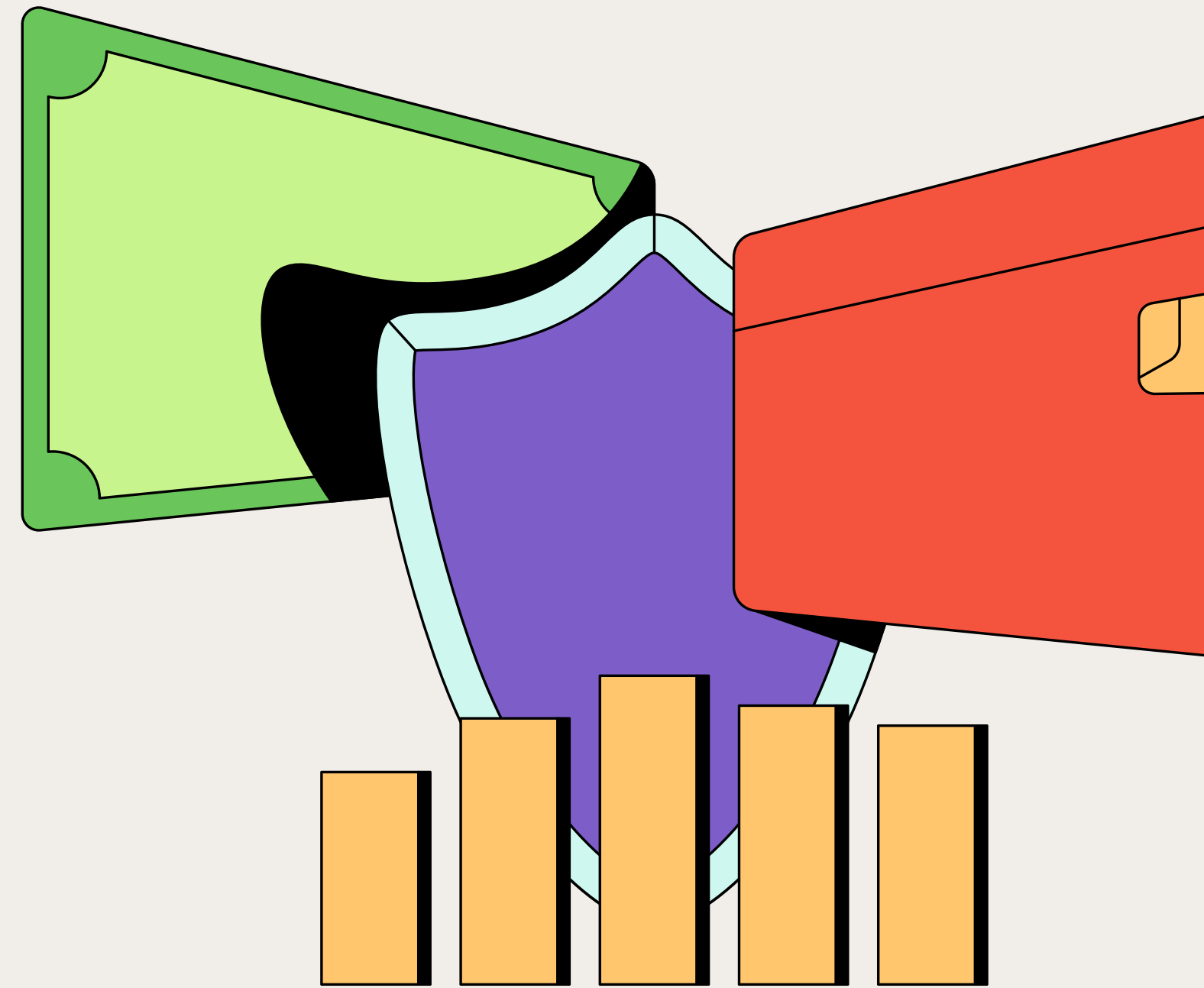
Optimasi Model:

- Tujuan: Meningkatkan hasil & stabilitas prediksi
- Teknik:
 - GridSearchCV untuk hyperparameter tuning
 - Normalisasi fitur (StandardScaler)

Model yang Dioptimasi:

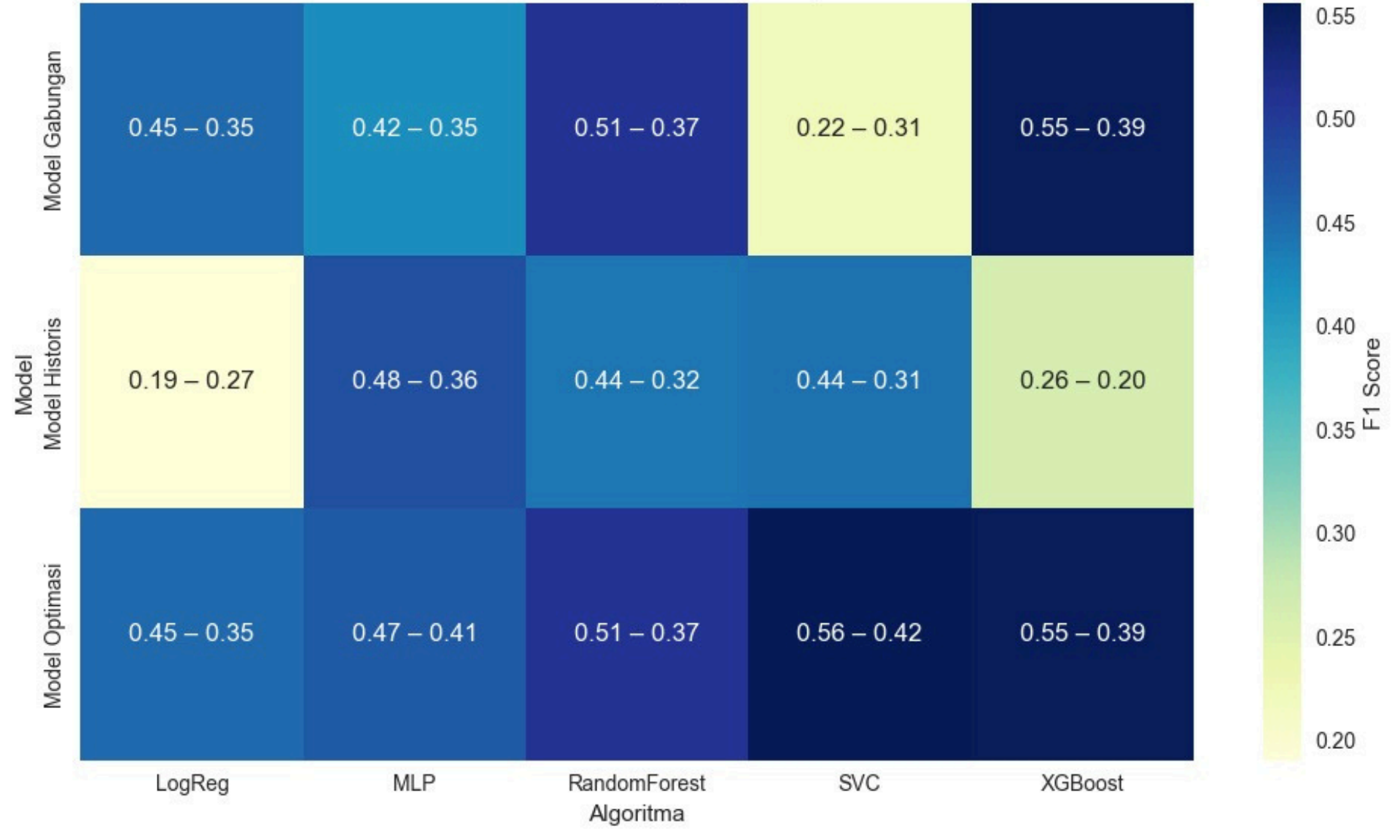
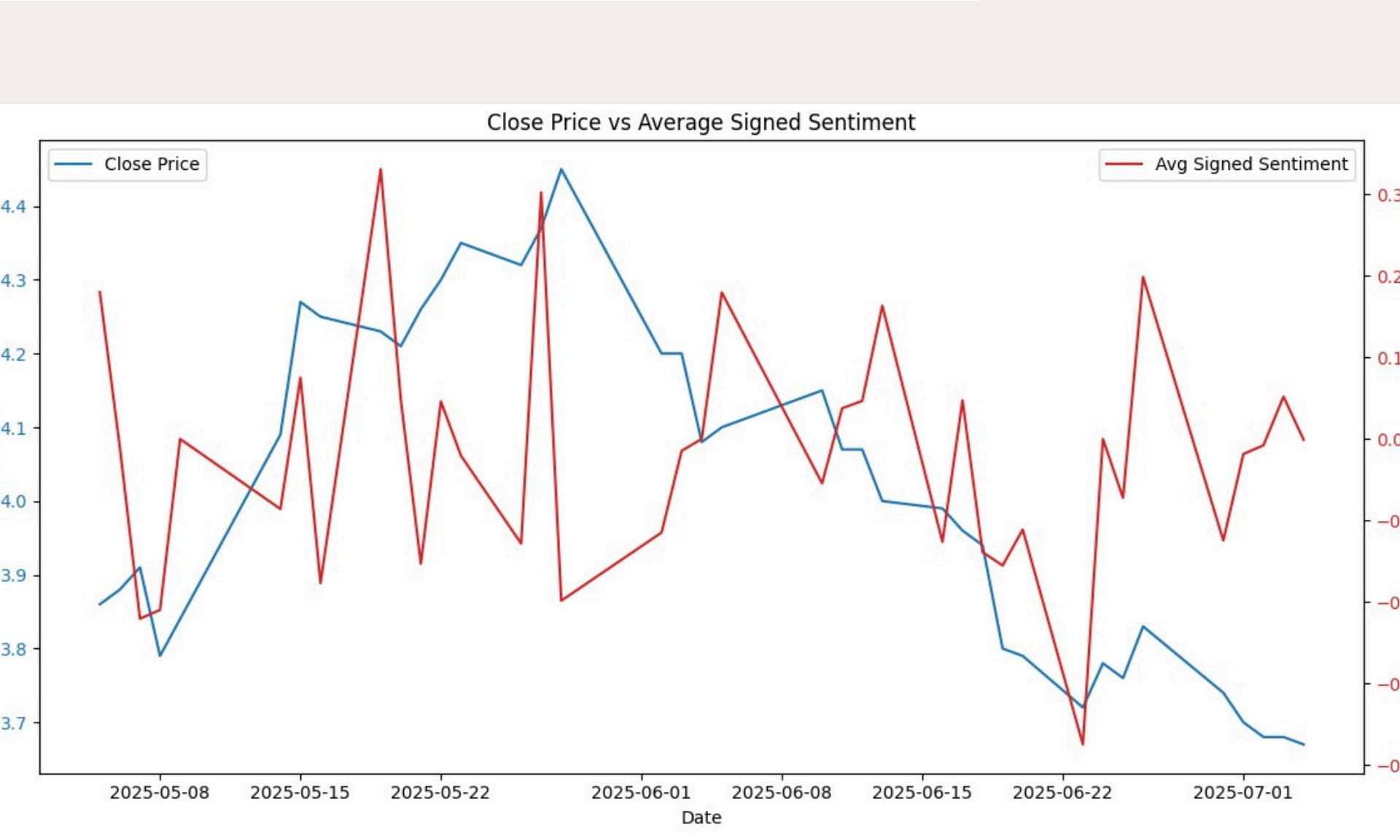
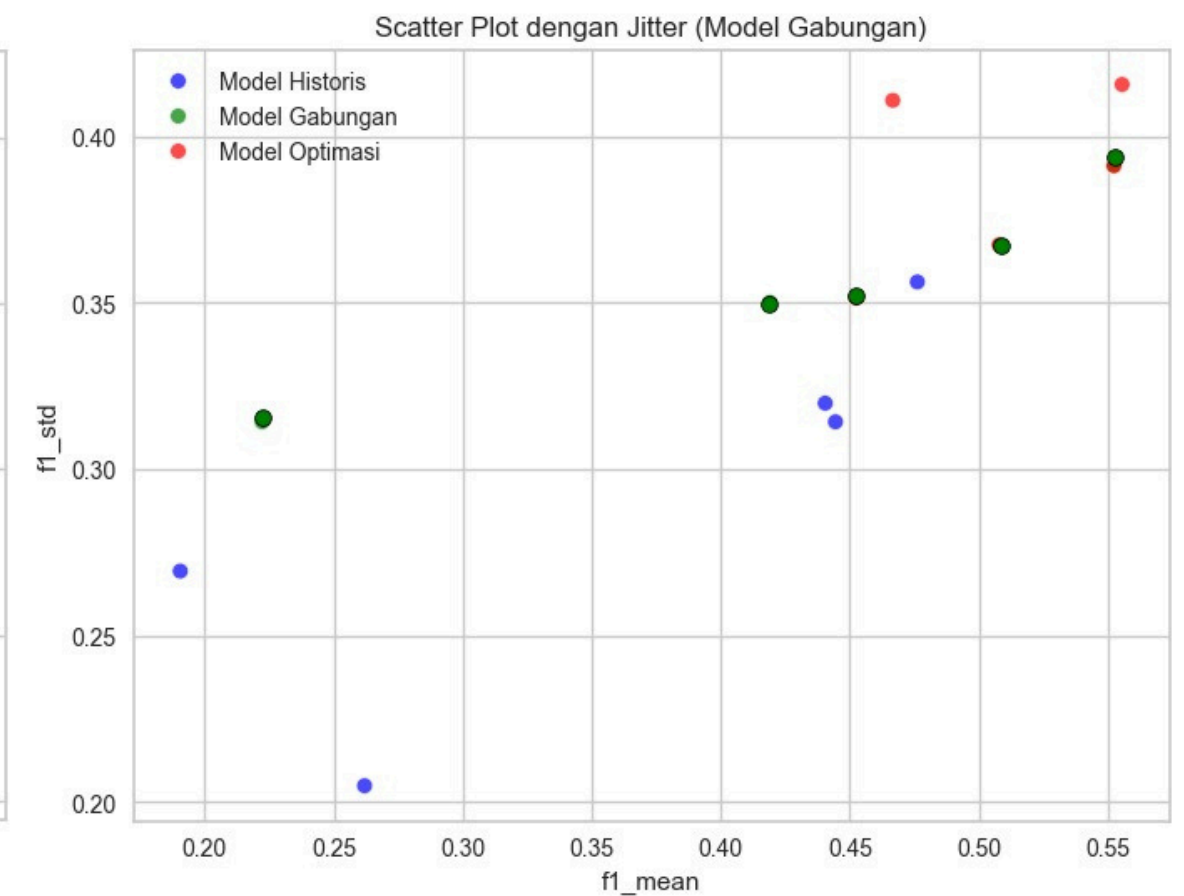
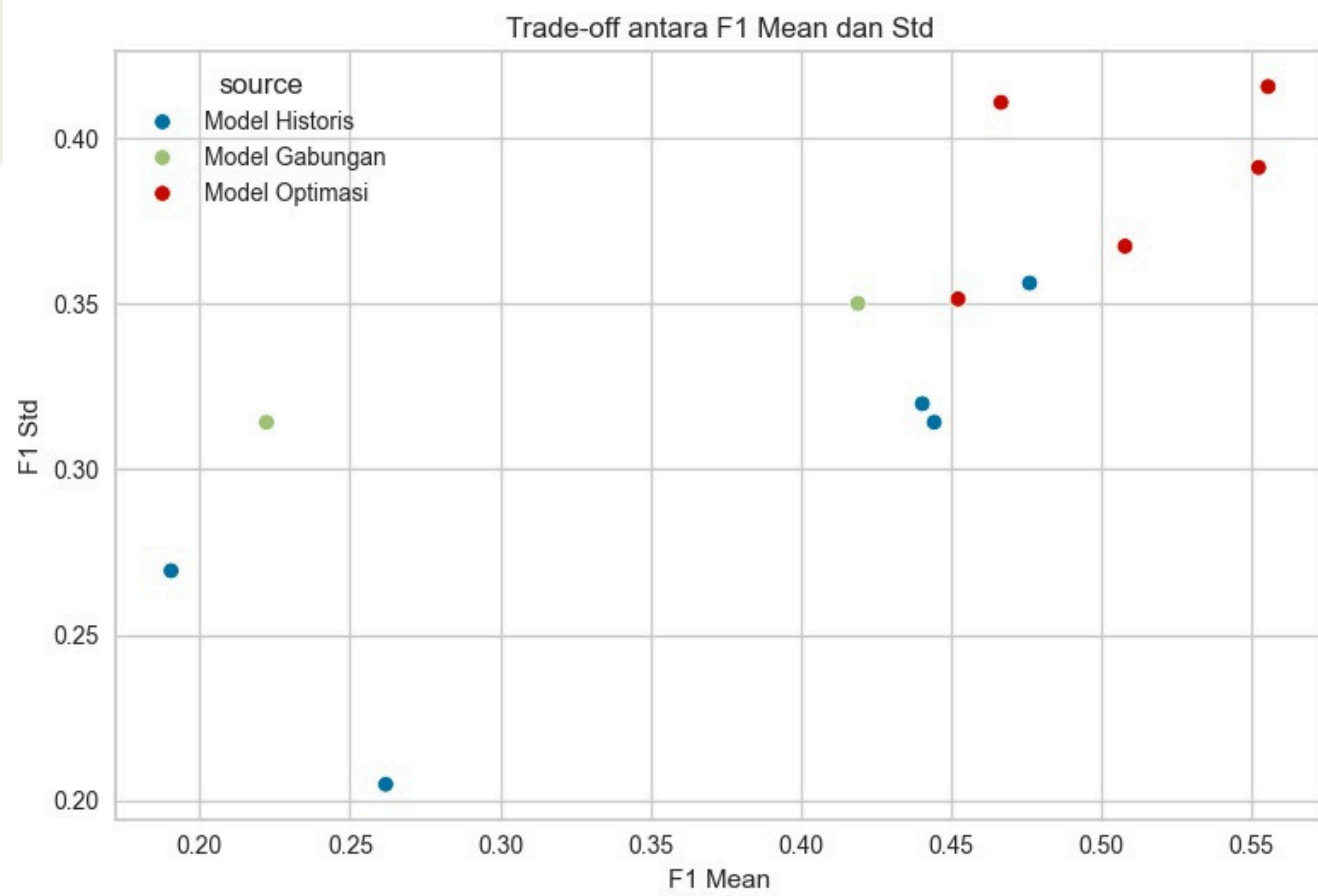
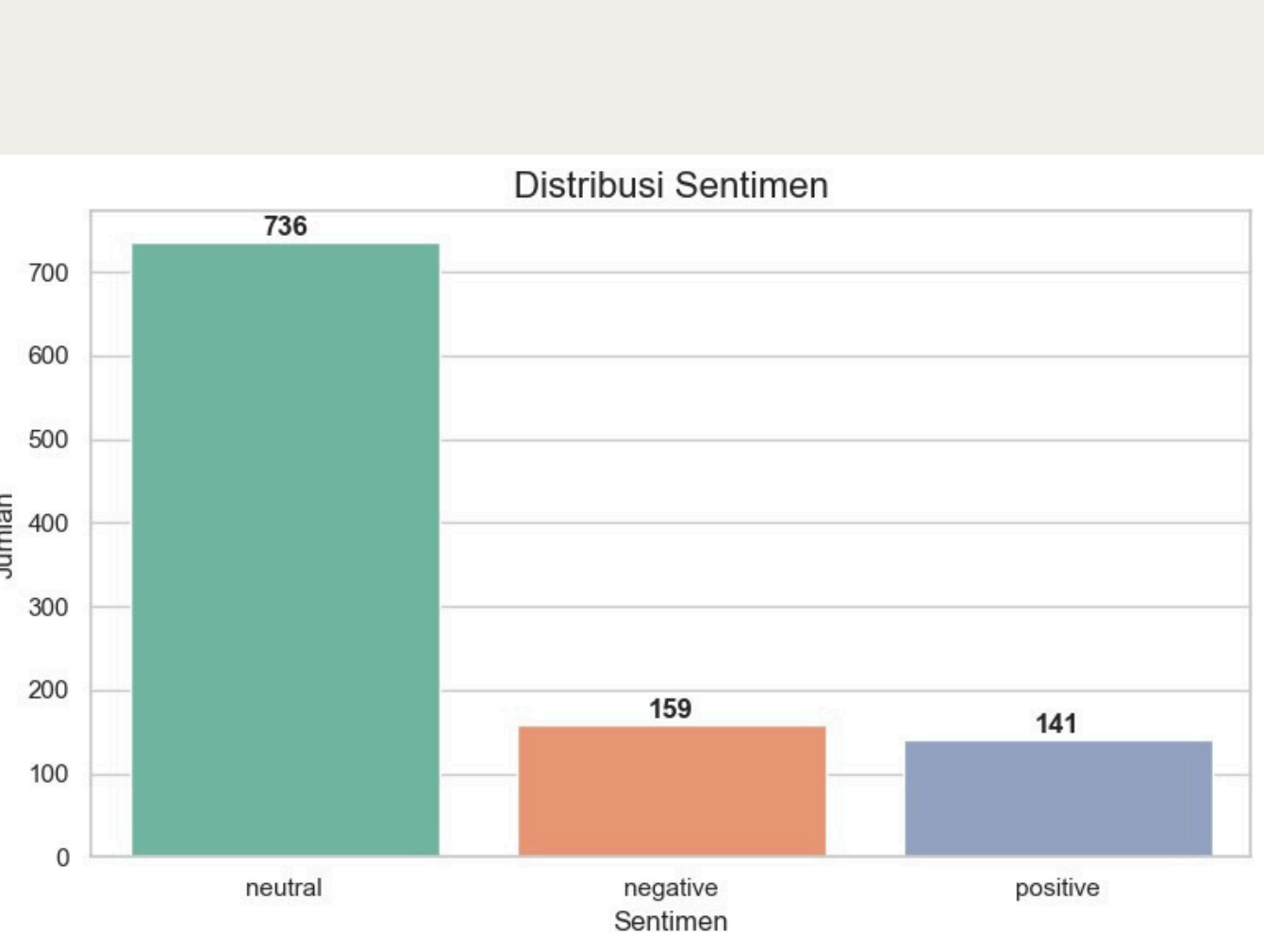
- Random Forest: `n_estimators`, `max_depth`
- XGBoost: `max_depth`, `learning_rate`, dan `n_estimators`
- LogRes: `C`, `penalty`, `solver`
- SVC: `C`, `gamma`, dan `kernel`
- MLP: `hidden_layer_sizes`, `activation`, dan `alpha`

Optimization



Hasil Optimasi

📊 Rangkuman Evaluasi Model (Tahap Optimasi):												
	accuracy_mean	precision_mean	recall_mean	f1_mean	roc_auc_mean	directional_acc_mean	accuracy_std	precision_std	recall_std	f1_std	roc_auc_std	directional_acc_std
RandomForest_3_fold	0.611111	0.416667	0.666667	0.507937	0.895833	0.500000	0.207870	0.311805	0.471405	0.367487	0.106230	0.136083
RandomForest_5_fold	0.611111	0.416667	0.666667	0.507937	0.833333	0.500000	0.207870	0.311805	0.471405	0.367487	0.117851	0.136083
RandomForest_10_fold	0.611111	0.416667	0.666667	0.507937	0.895833	0.500000	0.207870	0.311805	0.471405	0.367487	0.106230	0.136083
XGBoost_3_fold	0.555556	0.383333	0.666667	0.476190	0.833333	0.500000	0.207870	0.306413	0.471405	0.356348	0.235702	0.136083
XGBoost_5_fold	0.666667	0.472222	0.666667	0.552381	0.675926	0.611111	0.235702	0.335640	0.471405	0.391288	0.124914	0.078567
XGBoost_10_fold	0.555556	0.383333	0.666667	0.476190	0.833333	0.500000	0.207870	0.306413	0.471405	0.356348	0.235702	0.136083
LogReg_3_fold	0.555556	0.250000	0.333333	0.285714	0.800926	0.555556	0.207870	0.353553	0.471405	0.404061	0.153966	0.078567
LogReg_5_fold	0.611111	0.416667	0.500000	0.452381	0.796296	0.555556	0.207870	0.311805	0.408248	0.351543	0.214367	0.078567
LogReg_10_fold	0.555556	0.250000	0.333333	0.285714	0.800926	0.555556	0.207870	0.353553	0.471405	0.404061	0.153966	0.078567
SVC_3_fold	0.666667	0.500000	0.666667	0.555556	0.583333	0.611111	0.272166	0.408248	0.471405	0.415740	0.424918	0.283279
SVC_5_fold	0.666667	0.500000	0.666667	0.555556	0.583333	0.611111	0.272166	0.408248	0.471405	0.415740	0.424918	0.283279
SVC_10_fold	0.666667	0.500000	0.666667	0.555556	0.583333	0.611111	0.272166	0.408248	0.471405	0.415740	0.424918	0.283279
MLP_3_fold	0.611111	0.416667	0.500000	0.452381	0.791667	0.555556	0.207870	0.311805	0.408248	0.351543	0.155902	0.078567
MLP_5_fold	0.611111	0.444444	0.500000	0.466667	0.791667	0.611111	0.283279	0.415740	0.408248	0.410961	0.294628	0.283279
MLP_10_fold	0.611111	0.444444	0.500000	0.466667	0.791667	0.611111	0.283279	0.415740	0.408248	0.410961	0.294628	0.283279



Kesimpulan

- Distribusi sentimen publik terbukti berkorelasi dengan arah harga saham.
- Model gabungan (historis + sentimen) memberikan hasil prediksi yang lebih baik dibandingkan model historis.
- Berdasarkan evaluasi F1 Score, model XGBoost menjadi model terbaik dengan skor 0.5524, sedangkan SVC menjadi model yang paling perlu ditingkatkan dengan skor 0.2222.
- Namun setelah setiap model dioptimasi, SVC 3/5/10 Fold menjadi model terbaik dengan skor 0.5556, sedangkan LogReg 3/10 Fold menjadi model yang paling perlu ditingkatkan, dengan skor 0.2857



Work Together

Thank You For Your Attention

