

Big data & Predictive Analytics Final project

Stock Analysis Of AAPL

Dosen pengampu:
Mulya Sulistyono S.Kom, M.A

- Anggota kelompok:
1. Muhammad Davin Al Hisyam, 23.11.5376
 2. Dionisius Nusaca Redegnosis Nolejanduma, 23.11.5377
 3. Raditya Julian Primasakti, 23.11.5391
 4. Zulafa Yahya Wibowo, 23.11.5397

Program studi Informatika
Fakultas Ilmu Komputer
Universitas Amikom Yogyakarta
2025

Daftar Isi

1. Latar belakang	3
2. Metode	4
2.1. Alur final project	4
2.2. Dataset	5
2.3. EDA	6
3. Eksperimen	7
4. Hasil dan Evaluasi	8
5. Kesimpulan	9
5.1. Kesimpulan	9
5.2. Kontribusi	9
6. Lampiran	10

1. Latar belakang

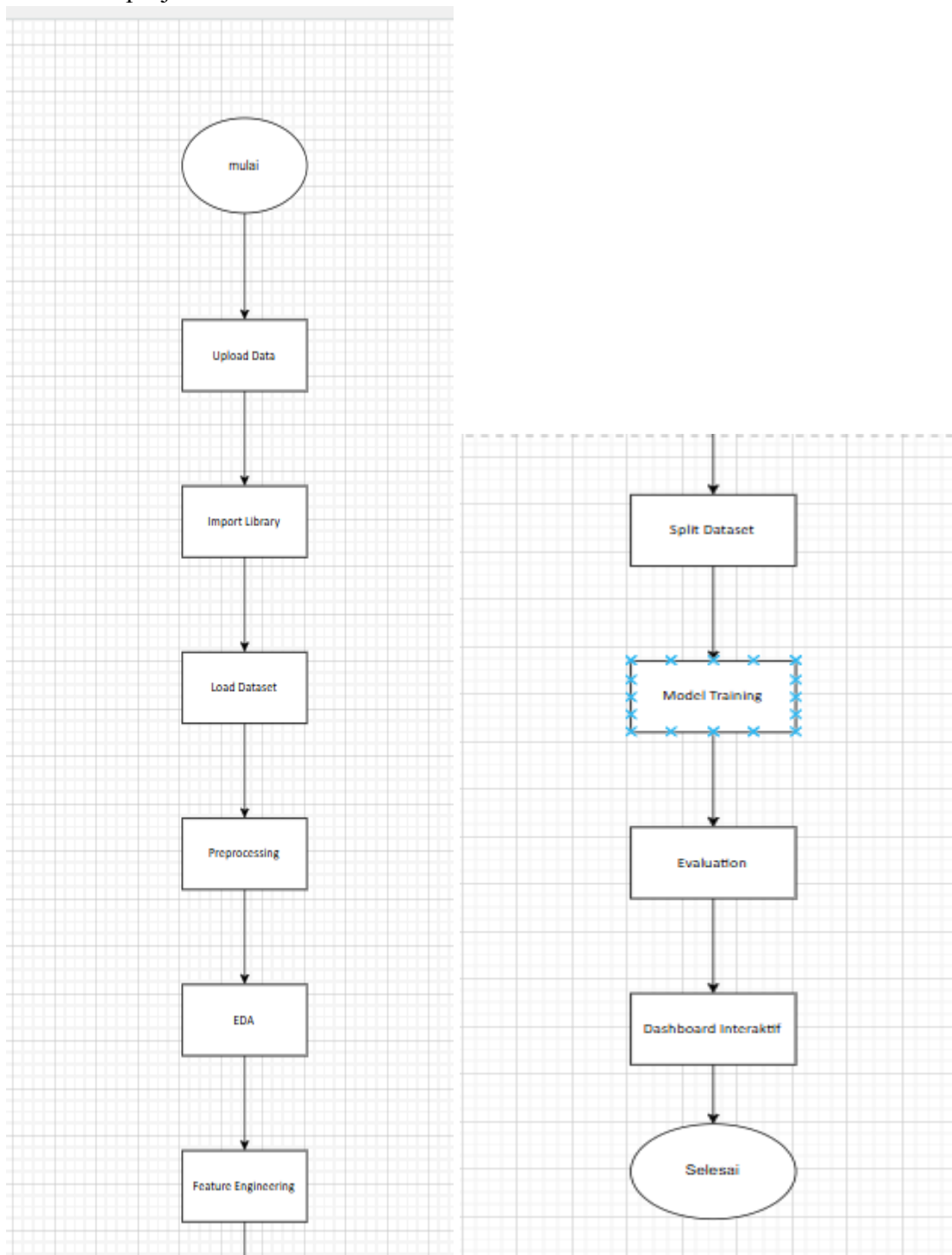
Perkembangan pasar modal dunia, termasuk saham perusahaan teknologi seperti Apple Inc. (AAPL), menunjukkan fluktuasi harga yang signifikan dari waktu ke waktu. Saham AAPL menjadi salah satu instrumen investasi paling populer di kalangan investor individu maupun institusi karena kapitalisasi pasarnya yang besar, stabilitas perusahaan, serta tingkat inovasi produk yang konsisten. Namun demikian, tingginya volatilitas harga saham seringkali menjadi tantangan bagi investor dalam menentukan waktu yang tepat untuk membeli atau menjual.

Analisis prediksi harga saham sangat penting untuk membantu investor mengambil keputusan berbasis data. Dengan memanfaatkan metode analitik modern, terutama pendekatan *big data* dan *machine learning*, kita dapat mencoba memodelkan pergerakan harga saham historis dan memprediksi tren harga di masa depan. Hal ini memberikan nilai tambah berupa *insight* prediktif yang sebelumnya sulit diperoleh hanya dengan analisis konvensional.

Selain itu, ketersediaan data harga saham yang melimpah dan bersifat time-series membuat domain ini sangat relevan dijadikan studi kasus dalam penerapan algoritma prediksi, *feature engineering*, dan evaluasi model prediksi.

2. Metode

2.1. Alur final project



2.2. Dataset

Dataset yang digunakan dalam proyek ini berasal dari file 'apple_stock.csv' yang diperoleh melalui platform Kaggle dan sudah kami download dan pindahkan ke dalam drive(https://drive.google.com/drive/folders/1ukmAVGeOP3fWSycjdORQK4hiX_otsdqz?usp=sharing). Dataset mencakup data historis saham Apple (AAPL) dengan periode waktu dari tahun 2017 hingga 2024, dan memiliki frekuensi harian sesuai dengan hari perdagangan bursa. Beberapa kolom penting yang terdapat dalam dataset ini antara lain *Date* (tanggal), *Open* (harga pembukaan), *High* (harga tertinggi), *Low* (harga terendah), *Close* (harga penutupan), *Adj Close* (harga penutupan yang telah disesuaikan), dan *Volume* (volume transaksi). Ukuran dataset ini mencerminkan jumlah hari perdagangan selama sekitar tujuh tahun, sehingga cukup kaya untuk digunakan dalam analisis dan pemodelan prediksi harga saham.

2.3. EDA

Proses EDA yang dilakukan:

- Mengecek informasi data (info())
- Mengecek missing values
- Deskripsi statistik (describe())
- Visualisasi tren harga penutupan
- Heatmap korelasi harga (Open, High, Low, Close)
- Histogram distribusi harga penutupan
- Boxplot harga
- Scatter plot Open vs Close
- Membuat kolom Moving Average 30 hari
- Visualisasi harga penutupan dan Moving Average

3. Eksperimen

Eksperimen dilakukan dalam beberapa tahap utama. Pada tahap preprocessing dan feature engineering, data diolah dengan cara mengkonversi kolom tanggal menjadi format datetime, mengurutkan data berdasarkan tanggal agar urut secara kronologis, serta membuat fitur moving average (MA) selama 30 hari untuk membantu analisis trend. Selanjutnya, dilakukan pemilihan dan pelatihan model menggunakan algoritma Linear Regression dari pustaka `sklearn.linear_model`. Fitur yang digunakan sebagai prediktor meliputi kolom Open, High, dan Low, sedangkan target prediksi adalah kolom Close. Dataset kemudian dibagi menjadi dua bagian, yaitu 80% data training dan 20% data testing, sehingga model dapat diuji performanya pada data yang belum pernah dilihat. Proses evaluasi model dilakukan dengan menghitung tiga metrik utama, yaitu Mean Absolute Error (MAE), Mean Squared Error (MSE), dan R-squared (R^2), guna menilai ketepatan prediksi. Untuk memudahkan interpretasi hasil, dibuat visualisasi berupa scatter plot yang membandingkan harga penutupan aktual dengan prediksi model, serta dashboard interaktif menggunakan Plotly. Dalam eksperimen ini digunakan berbagai library dan tools, antara lain pandas untuk manipulasi data, matplotlib untuk visualisasi statis, seaborn untuk eksplorasi korelasi, sklearn sebagai pipeline machine learning, plotly untuk visualisasi interaktif, dan Google Collab sebagai platform eksekusi notebook.

4. Hasil dan Evaluasi

Hasil dari model Linear Regression menunjukkan bahwa intercept yang dihasilkan adalah sebesar 5.823627154001742. Koefisien dari masing-masing fitur adalah Open sebesar -0.611787, High sebesar 0.801635, dan Low sebesar 0.810430. Berdasarkan perhitungan matrik evaluasi, diperoleh nilai Mean Absolute Error (MAE) sebesar 0.4789, Mean Squared Error (MSE) sebesar 0.4839, dan R-squared (R^2) sebesar 0.9999, yang menunjukkan performa model sangat baik dan mendekati prediksi sempurna. Visualisasi hasil prediksi dalam bentuk scatter plot memperlihatkan bahwa nilai prediksi berada sangat dekat dengan garis diagonal, menandakan akurasi yang tinggi. Sementara itu, dashboard interaktif yang dibuat dengan Plotly memungkinkan pengguna melakukan eksplorasi lebih lanjut terhadap tren harga penutupan aktual, pergerakan moving average, serta hasil prediksi model secara interaktif dan mendetail.

5. Kesimpulan

5.1. Kesimpulan

Pada bagian yang saya kerjakan, yaitu pengumpulan data dan evaluasi model regresi linear, saya menyimpulkan bahwa proses pengambilan data berjalan dengan baik menggunakan dataset *apple_stock.csv* dari Kaggle yang mencakup data historis saham AAPL periode 2017–2024. Setelah model dilatih, saya melakukan evaluasi menggunakan metrik MAE, MSE, dan R-squared (R^2). Hasil evaluasi ini menunjukkan bahwa model regresi linear cukup mampu memprediksi harga penutupan berdasarkan fitur *Open*, *High*, dan *Low*. Evaluasi ini penting untuk mengetahui seberapa baik performa model dan apakah masih perlu perbaikan atau pengembangan lebih lanjut.

5.2. Kontribusi

NIM	Nama	Kontribusi
23.11.5376	Muhammad Davin Al Hisyam	Membuat laporan eksperimen bagian Latar Belakang dan kesimpulan, juga membuat code bagian EDA dan Visualisasi Data di google collab
23.11.5377	Dionisius Nusaca Redegnosis N.	Membuat laporan eksperimen bagian Metode, Mencari Dataset AAPL Stocks, Membuat code bagian Data Collection dan Evaluasi Model Linear di google collab
23.11.5391	Raditya Julian Primasakti	Membuat laporan eksperimen bagian Eksperimen, dan membuat code bagian Analisis Korelasi di google collab
23.11.5397	Zulafa Yahya Wibowo	Membuat laporan eksperimen bagian Hasil dan Evaluasi, dan membuat code bagian Model Regresi Linear di google collab

6. Lampiran

Google Collab:

<https://colab.research.google.com/drive/19pZbxyzahxDPIzUDGfRIuvKxKJYwpbyN?usp=sharing>

Poster :



LATAR BELAKANG

Perkembangan pasar modal dunia, termasuk saham perusahaan teknologi seperti Apple Inc. (AAPL), menunjukkan fluktuasi harga yang signifikan dari waktu ke waktu. Saham AAPL menjadi salah satu instrumen investasi paling populer di kalangan investor individu maupun institusi karena kapitalisasi pasarnya yang besar, stabilitas perusahaan, serta tingkat inovasi produk yang konsisten. Namun demikian, tingginya volatilitas harga saham seringkali menjadi tantangan bagi investor dalam menentukan waktu yang tepat untuk membeli atau menjual.

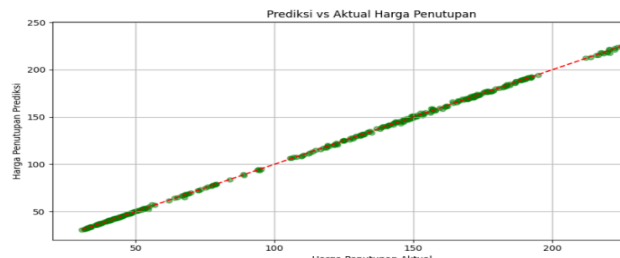


EKSPERIMEN

Eksperimen dilakukan melalui beberapa tahap. Pertama, data diproses dengan mengonversi kolom tanggal ke format datetime, mengurutkannya secara kronologis, dan menambahkan fitur moving average (MA) 30 hari. Model yang digunakan adalah Linear Regression dari sklearn, dengan fitur Open, High, dan Low untuk memprediksi Close. Data dibagi 80% untuk pelatihan dan 20% untuk pengujian. Evaluasi model dilakukan menggunakan MAE, MSE, dan R^2 . Hasil divisualisasikan dengan scatter plot dan dashboard interaktif menggunakan Plotly. Tools yang digunakan antara lain pandas, matplotlib, seaborn, sklearn, plotly, dan Google Colab.

HASIL DAN EVALUASI

Hasil dari model Linear Regression menunjukkan bahwa intercept yang dihasilkan adalah sebesar 5.823627154001742. Koefisien dari masing-masing fitur adalah Open sebesar -0.611787, High sebesar 0.801635, dan Low sebesar 0.810430. Berdasarkan perhitungan matrik evaluasi, diperoleh nilai Mean Absolute Error (MAE) sebesar 0.4789, Mean Squared Error (MSE) sebesar 0.4839, dan R^2 sebesar 0.9999, yang menunjukkan performa model sangat baik dan mendekati prediksi sempurna. Visualisasi hasil prediksi dalam bentuk scatter plot memperlihatkan bahwa nilai prediksi berada sangat dekat dengan garis diagonal, menandakan akurasi yang tinggi.



https://www.canva.com/design/DAGtD8w4aQk/1JYVtctkP0WgqhMdel8p8w/edit?utm_content=DAGtD8w4aQk&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton