

# Data-Driven Trading Decisions: Machine Learning on Palantir Stock

## 1. Latar Belakang

Pasar saham dikenal sebagai salah satu instrumen investasi yang menjanjikan keuntungan tinggi, namun juga memiliki risiko yang besar akibat volatilitas harga yang sulit diprediksi. Saham-saham teknologi, seperti Palantir Technologies Inc. (PLTR), menunjukkan fluktuasi harga yang signifikan dalam beberapa tahun terakhir, dipengaruhi oleh perkembangan teknologi, perubahan regulasi, serta dinamika pasar global. Kondisi ini menimbulkan tantangan bagi investor maupun analis dalam menentukan strategi investasi yang optimal.

Di sisi lain, perkembangan teknologi komputasi dan ketersediaan data historis membuka peluang untuk memanfaatkan metode analisis kuantitatif berbasis data. Machine learning dan model statistik modern dapat digunakan untuk mengidentifikasi pola, memprediksi pergerakan harga, serta mengevaluasi strategi trading dengan lebih objektif. Hal ini menjadi semakin penting mengingat banyak investor masih mengandalkan intuisi atau analisis manual, yang rentan terhadap bias dan kurang mampu mengantisipasi perubahan pasar secara cepat.

Analisis terhadap data historis saham Palantir tidak hanya memberikan gambaran mengenai performa harga dan volatilitas, tetapi juga memungkinkan pengujian berbagai strategi trading, seperti moving average crossover, relative strength index (RSI), maupun model prediksi berbasis ARIMA dan LSTM. Dengan pendekatan ini, diharapkan diperoleh insight yang lebih komprehensif terkait peluang dan risiko, sehingga mendukung pengambilan keputusan investasi yang lebih terukur dan data-driven.

## 2. Tujuan

- Menghasilkan insight dari pola historis harga.
- Menguji performa strategi sederhana (MA Crossover & RSI).
- Membandingkan hasil forecasting menggunakan ARIMA dan LSTM.

## 3. Dataset

Sumber: [Palantir Stock Data 2025](#)

Jumlah Data: 1240 baris

Variabel Data:

- Date : Date

- Open: The price market open
- High: The highest price for that day
- Low: The lowest price for that day
- Close: The price at market close, adjusted for splits
- adj\_close: The closing price after adjustments for all applicable splits and dividend distributions. Data is adjusted using appropriate split and dividend multipliers, adhering to Center for Research in Security Prices (CSRP) standards
- Volume: The number of shares traded on that day

## 4. Metodologi

### 4.1 Data Processing & Feature Engineering

Dataset yang digunakan merupakan data historis saham Palantir Technologies Inc. (PLTR) dari Kaggle, dengan periode tertentu yang memuat informasi harga harian: *open*, *high*, *low*, *close*, *adjusted close*, dan *volume*. Pada tahap ini model dan strategi memiliki input yang relevan dengan pola pergerakan harga saham. Untuk mencapainya tahap awal pemrosesan data dilakukan dengan:

1. Pembersihan data: yakni mengatasi *missing values* dan memastikan format tanggal sudah terurut
2. Transformasi variabel: yakni membuat *log return* sebagai representasi pertumbuhan harga harian
3. Feature Engineering: yakni menambahkan variabel teknikal seperti *Moving Average* (20 & 50 hari), *Relative Strength Index* (RSI) periode 14 hari, dan *Volatility Rolling Window* (20 hari)

### 4.2 Exploratory Data Analyst (EDA)

EDA dilakukan untuk memahami karakteristik data sebelum pemodelan. Dimana hasil EDA memberikan gambaran awal terkait stabilitas harga, tingkat risiko, dan pola fluktuasi yang menjadi dasar perancangan model selanjutnya. Analisisnya mencakup:

- Visualisasi tren harga guna melihat tren jangka panjang dari *closing price*
- Distribusi log return untuk mengukur distribusi tersebut mendekati normal atau menunjukkan *fat tails* yang khas pada data keuangan
- Volatilitas guna untuk menangkap dinamika risiko harga dari waktu ke waktu

### 4.3 Forecasting

#### a. ARIMA

Model ARIMA digunakan untuk melakukan prediksi harga berdasarkan pola historis deret waktu. ARIMA ini dipilih karena kemampuannya dalam menangkap pola *autocorrelation* sederhana pada data harga saham. Tahapan pemodelan ini meliputi:

- Identifikasi orde ARIMA melalui ACF dan PACF
- Estimasi parameter model dengan pendekatan *maximum likelihood*
- Validasi model melalui residual diagnostic (uji *white noise*)
- Visualisasi hasil prediksi terhadap data aktual

#### b. LSTM

Selain model statistik klasik, digunakan juga *Long Short-Term Memory* (LSTM) sebagai pendekatan *deep learning* untuk *forecasting*. LSTM dipilih karena kemampuannya mengingat pola jangka panjang dalam data sekuensial. Proses pemodelan meliputi:

- 1) Normalisasi data harga agar sesuai dengan kebutuhan jaringan saraf
- 2) Pembentukan *training* dan *test set* dengan jendela waktu (*sliding window*)
- 3) Arsitektur model:
  - a. 1 lapisan LSTM dengan 64 unit
  - b. *Dropout* untuk mengurangi *overfitting*
  - c. Lapisan *Dense* akhir untuk menghasilkan *output* harga prediksi
- 4) *Training* model dengan *loss function* MSE dan *optimizer* Adam
- 5) Evaluasi model melalui perbandingan prediksi dengan data actual

### 4.4 Backtesting Strategi

#### a. Moving Average Crossover

Strategi ini menggunakan dua *moving average* dengan periode yang berbeda yakni 20 dan 50 hari. Kinerja strategi diukur dengan metrik: *total return*, *annualized return*, *sharpe ratio*, *volatilitas tahunan*, dan *max drawdown*. Dengan aturan yang digunakan:

- *Buy signal* yakni ketika MA-20 memotong MA-50 dari bawah (*golden cross*)
- *Sell signal* yakni ketika MA-20 memotong MA-50 dari atas (*death cross*)

### b. *RSI Strategy*

Strategi berbasis momentum dengan indikator RSI (*Relative Strength Index*) periode 14 hari. Sama seperti *Moving Average*, performa akan diuji melalui *backtesting* dan dibandingkan dengan metrik risiko & return. Dengan aturan yang diterapkan:

- *Buy signal*:  $RSI < 30$  (*oversold*)
- *Sell signal*:  $RSI > 70$  (*overbought*)

## 5. Hasil dan Pembahasan

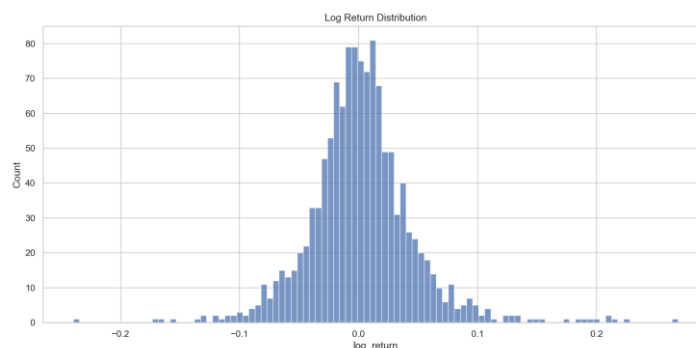
### 5.1 *Exploratory Data Analyst (EDA)*

#### a. Pergerakan Harga Saham



Grafik di atas menunjukkan harga saham Palantir dari tahun 2020 hingga 2025. Terlihat periode awal (2020–2022) cenderung stagnan dengan harga relatif rendah, kemudian mulai meningkat secara signifikan sejak 2023 hingga puncaknya pada 2025. Fluktuasi harga yang besar mengindikasikan volatilitas tinggi yang perlu diperhatikan dalam analisis risiko.

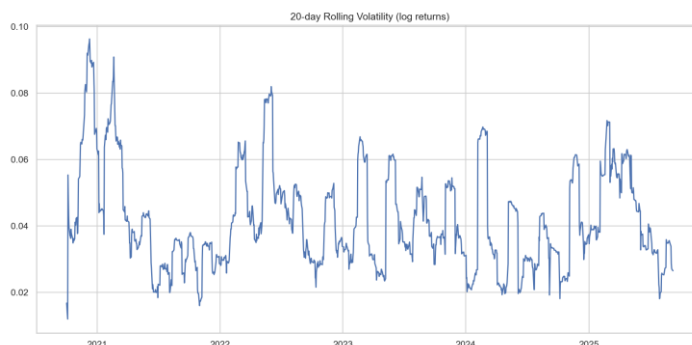
#### b. Distribusi Log Return



Histogram log return harian menunjukkan bentuk distribusi mendekati normal, namun dengan fat tails (ekor tebal). Artinya, walaupun return harian rata-rata kecil, ada kemungkinan kejadian ekstrem (spike naik/turun besar). Ini penting

untuk mempertimbangkan risiko ketika menggunakan model prediksi maupun strategi trading.

### c. Volatilitas 20 Hari



Grafik ini menampilkan volatilitas rolling 20 hari. Terlihat beberapa periode lonjakan volatilitas, terutama di awal 2021 serta lonjakan signifikan pada 2024–2025. Lonjakan volatilitas ini biasanya berhubungan dengan fase bullish atau bearish tajam di harga saham.

## 5.2 Forecasting

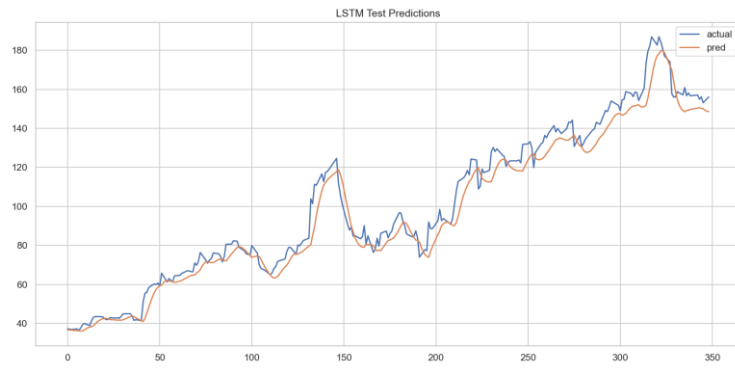
### a. ARIMA



Hasil forecasting ARIMA menunjukkan prediksi harga yang cenderung kembali ke rata-rata (mean-reverting). ARIMA dapat menangkap pola musiman dan tren linear, tetapi terbatas untuk data saham yang cenderung non-linear dan sangat dipengaruhi faktor eksternal.

### b. LSTM

Model LSTM dilatih dengan 20 epoch menghasilkan training loss  $4.89e-05$  dan validation loss 0.0015. Hal ini menunjukkan model cukup baik dalam mempelajari pola harga saham.



Grafik hasil prediksi memperlihatkan bahwa LSTM mampu mengikuti tren harga lebih baik daripada ARIMA, khususnya dalam menangkap pola jangka pendek.

## 5.3 Backtesting Strategi

### a. MA Crossover



Grafik memperlihatkan sinyal beli (panah hijau) dan jual (panah merah) berdasarkan crossover MA20 dan MA50. Terlihat strategi ini mampu menangkap tren naik panjang, tetapi juga memberikan sinyal palsu di periode volatil tinggi.

- **Kinerja Backtest:**
  - Total Return: **465%**
  - Annualized Return: **42.18%**
  - Annualized Volatilitas: **56.81%**
  - Sharpe Ratio: **0.74**
  - Max Drawdown: **-78.55%**

Interpretasi: strategi ini sangat menguntungkan dalam periode tren naik, namun memiliki risiko kerugian besar jika terjadi tren berbalik.

## b. RSI Strategy



Grafik menunjukkan sinyal beli ketika  $RSI < 30$  (oversold) dan jual ketika  $RSI > 70$  (overbought). Strategi ini lebih konservatif dibanding crossover MA karena hanya memberikan sinyal pada kondisi ekstrem.

### Kinerja Backtest:

- Total Return: **142%**
- Annualized Return: **19.74%**
- Annualized Volatilitas: **48.95%**
- Sharpe Ratio: **0.40**
- Max Drawdown: **-66.88%**

Interpretasi: return lebih rendah dibanding MA Crossover, tetapi strategi ini lebih stabil karena tidak terlalu sering memberi sinyal beli/jual.

## 6. Kesimpulan

- 1) LSTM lebih unggul dibanding ARIMA dalam forecasting, karena mampu menangkap pola non-linear dan tren jangka pendek.
- 2) MA Crossover menghasilkan keuntungan paling besar, namun risikonya juga paling tinggi (high risk – high return).
- 3) RSI Strategy lebih konservatif, cocok untuk investor yang ingin mengurangi frekuensi transaksi meskipun return lebih kecil.
- 4) Kombinasi antara forecasting model (LSTM) dan teknikal indikator (MA/RSI) dapat memberikan pendekatan yang lebih kuat dalam pengambilan keputusan trading berbasis data.