Machine Learning Engineer Assignment

Introduction

I have developed an LSTM model for predicting stock price movements using historical financial data. I have first done EDA and statistical analysis to better understand the data and get insights to model it better. I used stock Pfizer (PFE) when developing and I included last 2000 days.

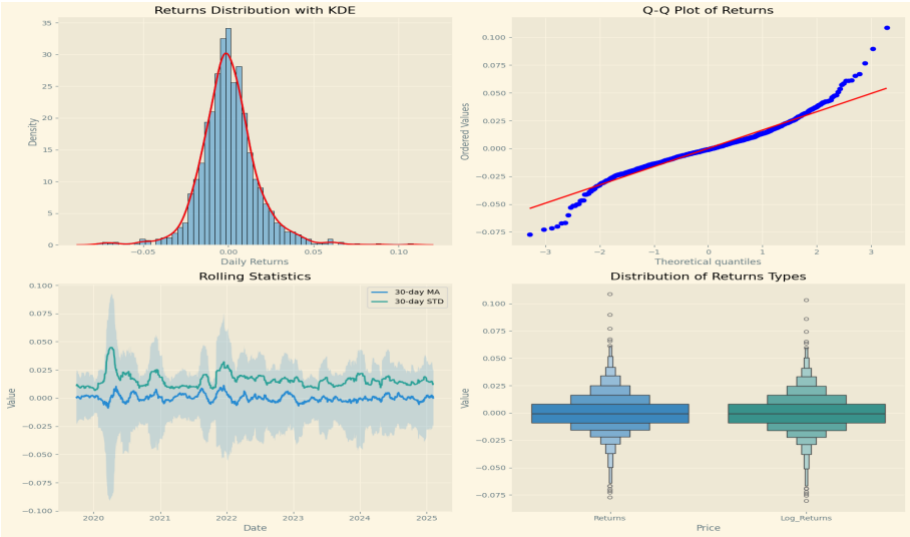
Steps Taken

Data Collection

1) Extracted stock market data using yahoo finance.
2) Selected key financial features: Open, High, Low, Close, and Volume.

Exploratory Data Analysis

1) Visualized stock volatility over time, indicating periods of high variance. Distribution plots showed almost normal distribution.

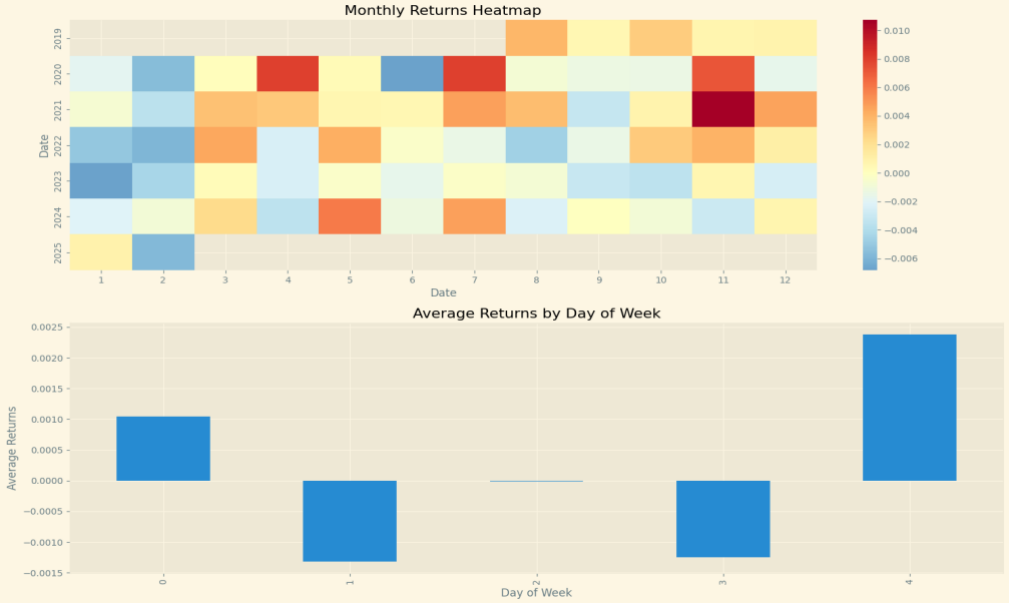


2) Confirmed stationarity which led me to use raw closing prices.

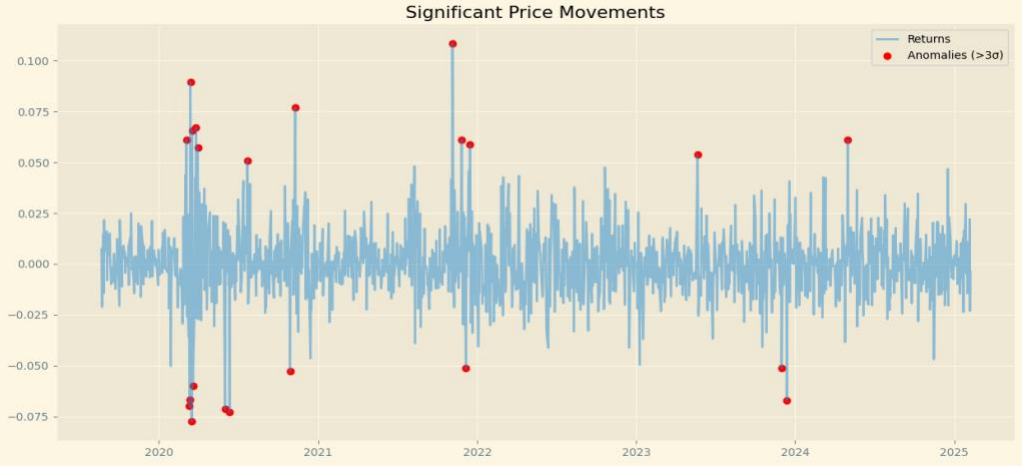
Stationarity Test Results

Test	Statistic	P-Value	Result
ADF	-8.1840	0.0000	✓ Stationary
KPSS	0.2556	0.1000	✓ Stationary

3) Analyzed seasonal trends in stock returns to consider using time-based features.



4) Identified significant market movements and anomalies



Feature Engineering

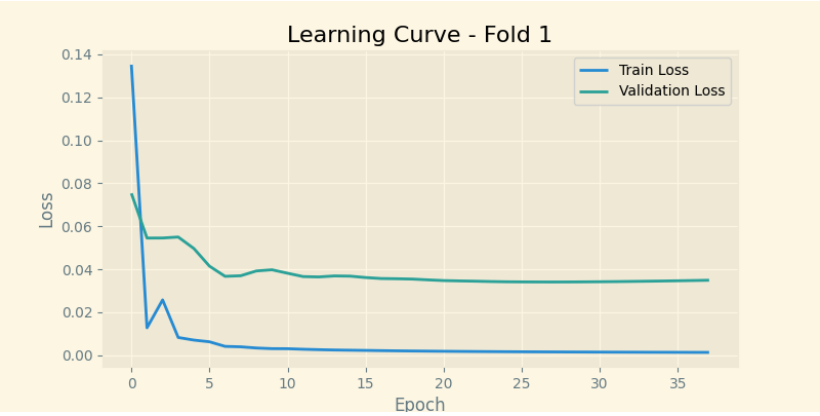
1) Implemented technical indicators (RSI, MACD, Bollinger Bands, ATR, ADX, OBV, Stochastic Oscillator)
2) Engineered additional features like moving averages and momentum indicators.
3) Applied Recursive Feature Elimination to select the most relevant features.

Data Preparation

1) Scaled features using Robust Scaler to handle outliers and maintain distribution.
2) Created sequences for LSTM input.
3) Split data into training, validation, and test sets while maintaining time-series nature to prevent data leakage.

Modeling Approach

1) Created an LSTM model and applied time series cross validation.
2) Implemented dropout layers and L2 regularization to prevent overfitting on the training set.
3) Used early stopping and learning rate reduction strategies to optimize convergence.
4) Utilized the Adam optimizer, because of hardware restrictions and convergence speed.



Metric Selection

RMSE was prioritized as it penalizes large errors more than MAE, and it is often the choice in financial forecasting especially since they have a lot of anomalies.

Results & Insights

We have achieved a low RMSE but a low R^2 because the model minimizes prediction errors, but stock price exhibit high volatility not explained by the model. This suggests that while our model captures trends well, it lacks power over macro price changes.

Test metrics:

1) **MAE:** 0.99
2) **MSE:** 1.33
3) **RMSE:** 1.15
4) **R^2 Score:** 0.28