

Assignment for Senior Machine Learning Engineer Position

Data Preprocessing

1. Set the date as the index.
2. Resampled the data weekly to eliminate the weekly seasonality and make the trend more apparent.
3. Used 15 years of publicly available stock data (which includes Open, High, Low, Close, Volume) and trimmed it.
4. Created a test set with 4 weeks of averaged data.
5. Deseasonalized and detrended the train set.

Technical Indicators and Features

6. Added technical indicators:
 - **RSI**: Measures momentum.
 - **MACD**: Captures trend and momentum.
 - **Bollinger Bands**: Shows price volatility.
 - **Daily_Return**: Percent change to capture daily movement.
 - **RollingMean/RollingStd**: Reflects short-term trend and volatility.
7. Also created trend and seasonal features using ARIMA and linear regression to enhance model performance.
8. Filled any missing values (from rolling calculations) with 0.

Model Development

9. Used ARIMA to forecast baseline close values for the validation set.
10. Employed an LSTM model to predict the residual component after detrending and deseasonalizing.
11. Scaled training and validation data using MinMaxScaler to prevent data leakage.
12. Found the best hyperparameters through cross-validation with MSE as the evaluation metric.

Evaluation

13. Predicted the test set (next day's/period's closing price) using the combined ARIMA, trend, seasonal, and LSTM forecasts.
14. Evaluated the model using RMSE since it clearly shows error magnitude and penalizes larger deviations. Achieved **68.37252626876133** RMSE while predicting the the average closing values of the last 4 weeks. Since the assignment wanted me to include those I included, however using too many predicted indicators has increased the weight of the ARIMA model on the results. Result might be improved by including less

indicators (1 or 2) to the model,

15. Plotted the actual vs. forecasted closing prices to visualize performance.