RoboStock

Brief but technically accurate and descriptive.
Words of ‘new’, ‘improved’, ‘improvement of’, ‘improvement in’ should not be considered as part of the title.
Similarly, the articles ‘a’, ‘an’, and ‘the’ should not be included as the first words of the title.
Note: in this template, ‘invention’ and ‘web service’ that indicating the work you have done may be used interchangeably.

Dhruv Rana
dkr43@scareltmail.rutgers.edu

Table of Contents

Abstract

RoboStock presents a comprehensive approach to stock price prediction leveraging ensemble machine learning techniques and web application development. The project aims to enhance prediction accuracy and provide investors with valuable insights for informed decision-making in the stock market. Utilizing five years of historical stock data, RoboStock employs a diverse ensemble comprising Linear Regression, LSTM, and a to-be-decided model. This ensemble method amalgamates the strengths of each model to produce robust and reliable predictions. The project also involves data preprocessing, feature engineering, model development, and rigorous evaluation to ensure optimal performance.

The developed ensemble model undergoes validation using appropriate evaluation metrics to assess its effectiveness in forecasting stock prices. Furthermore, RoboStock features the implementation of a user-friendly web application using the Django framework. This application serves as a convenient platform for users to input stock data and obtain predictions effortlessly. The seamless integration of the ensemble model into the web application enhances accessibility and usability for investors, financial analysts, traders, and data scientists alike.

RoboStock's holistic approach encompasses both technical sophistication in machine learning and practical usability in web application development. By bridging the gap between cutting-edge research and real-world applications, RoboStock aims to empower users with actionable insights and facilitate informed decision-making in the dynamic landscape of the stock market.

1. Introduction

The service that has been provided

- Concentrate on making *this* assertion and *only* this assertion in a succinct set of 1 to 3 paragraphs

The values of your service

- Describe your new idea that is outperform the existing service
- This is the place to provide a succinct description of the problem context giving enough information to support the claim that a problem exists

Why our solution is worth considering and why is it effective in some way that others are not

- A succinct statement of why the reader should care enough to read the rest of the paper.
- This should include a statement about the characteristics of your solution to the problem which 1) make it a solution, and 2) make it superior to other solutions to the same problem.

How the rest of the paper is structured

- The short statement below is often all you need, but you should change it when your paper has a different structure, or when more information is *required* to describe what a given section contains. If it isn't *required* then you don't want to say it here.

The rest of this paper first discusses related work in Section 2, and then describes our implementation in Section 3. Section 4 describes how we evaluated our system and presents the results. Section 5 presents our conclusions and describes future work.

2. Cross-reference to related work

Other efforts that exist to solve this problem and why are they less effective than your method

- Statement regarding prior disclosures (existing services)
- Resist the urge to point out only flaws in other work. Do your best to point out both the strengths and weaknesses to provide as well rounded a view of how your idea relates to other work as possible

3. Background of the service

The Background of the service may (but is not required to) include the following parts:

- Field of the service: a statement of the field of art to which the service pertains, the statement should be directed to the subject matter of the claimed service.
- Description of the related art: A paragraph(s) describing to the extent practical the state of the prior art or other information disclosed known to the service provider, including references to specific prior art or other information where appropriate. Where applicable, the problems involved in the prior art or other information disclosed which are solved by the service provider's invention should be indicated.

4. Brief summary of the service

The purpose of the brief summary of your service is to apprise the public, and more especially those interested in the particular art to which the service relates, of the nature of the service, the summary should be directed to the specific service being claimed. That is, the subject matter of the service should be described in one or more clear, concise sentences or paragraphs.

5. Brief description of the several views of the drawing

When there are drawings, there shall be a brief description of the several views of the drawings and the detailed description of the invention (your web service) shall refer to the different views by specifying the numbers of the figures, and to the different parts by use of reference letters or numerals (preferably the latter).

6. Detailed description of the web service

A detailed description of the invention and drawings follows the general statement of invention and brief description of the drawings. This detailed description must be in such particularity as to enable any person skilled in the pertinent art or science to develop the same service without involving extensive experimentation.

Enablement: enable your paper reader to do the same without a sweat
Best mode: how your service is used to bring up the best results
Recommendation: things learned from your development, that including mistakes you made, solutions to the issues encountered, etc.

To prepare input data for our stock price prediction algorithms, we need to conduct thorough analysis to select relevant features that have a high correlation with the target output, which is the stock price. Here's how we can proceed:

Feature Selection:

- We will consider various features commonly used in stock price prediction, such as opening price, closing price, highest price, lowest price, and trading volume.
- Additionally, we can incorporate technical indicators like moving averages, Relative Strength Index (RSI), Moving Average Convergence Divergence (MACD), and Bollinger Bands.

Correlation Analysis:

- We will calculate the correlation coefficients between each feature and the target output (stock price).
- Features with high correlation coefficients indicate a strong linear relationship with the target variable and are more likely to be valuable for prediction.

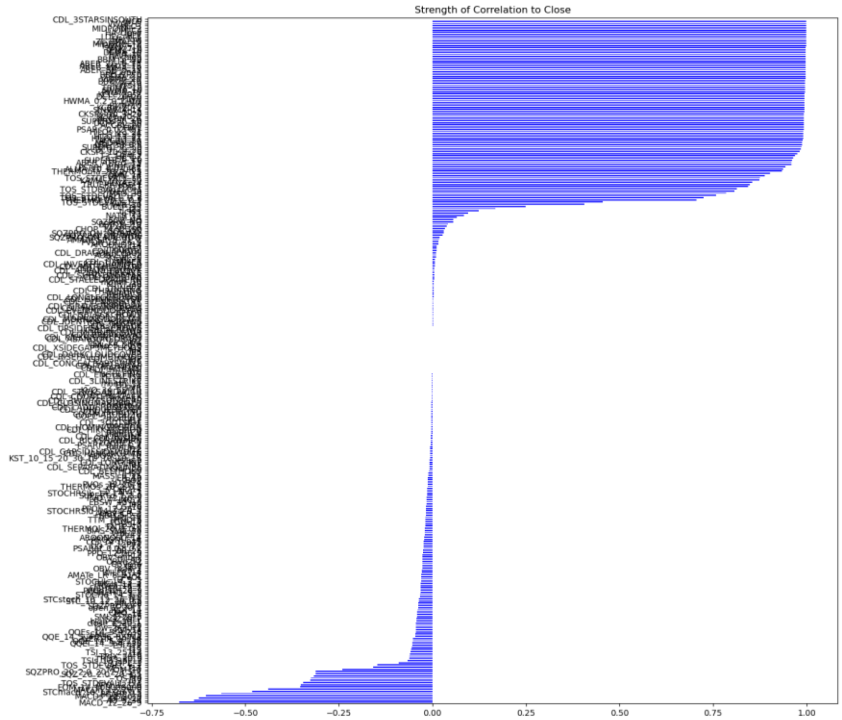


Figure x : Correlation Strength of Several features to next day closing price

- A correlation coefficient close to 1 indicates a strong positive correlation, while a coefficient close to -1 indicates a strong negative correlation.

Feature Engineering:

- Next, we analysis where all the features lie in the feature space.

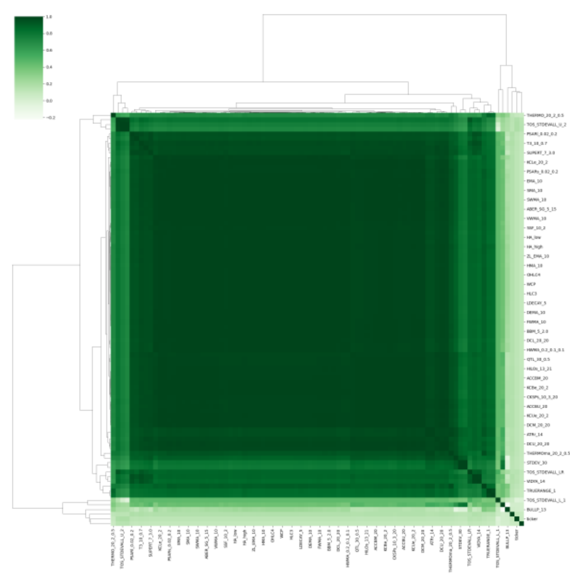


Figure x : Correlation between highly correlated features (sorted based on distance)

- The dark green square on top right of figure indicates that several features are correlated and provide the same information as other features in relation to closing price for the next day. Thus, we can ignore features that are close in feature space to each other to reduce our number of features.

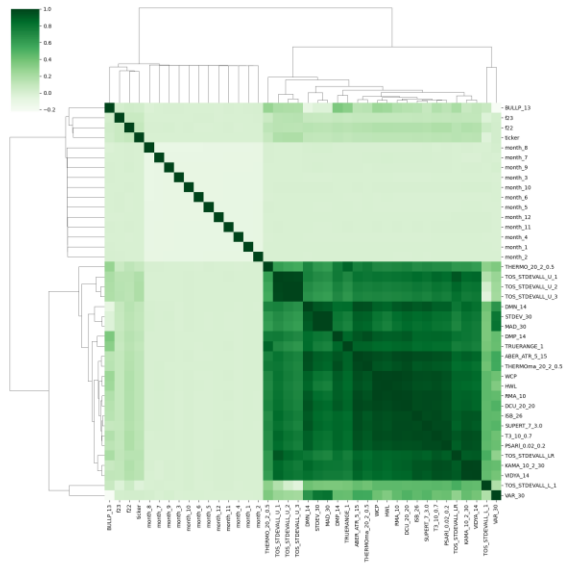


Figure x : Correlation between highly correlated features (ignoring features that provide some information)

- Now the graph looks a lot better, and we don't have a lot of dark green squares as before. Next, we want to make sure the relationship we found is not stationary and will hold over time. Thus, we check the correlation of the features over quarterly change.

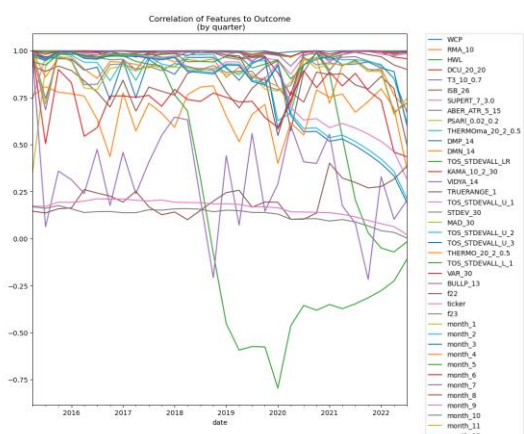


Figure x: Correlation of Selected Features to Outcome (sampled quaterly)

Based on the correlation analysis, we may perform feature engineering to create new features or transform existing ones to enhance predictive power. By conducting thorough analysis and selecting features with high correlation coefficients, we ensure that our prediction algorithms receive meaningful input data that are strongly associated with the target output. This approach enhances the predictive accuracy and reliability of our stock price prediction models.

Model Choice:
Choice of LSTM (Long Short-Term Memory):

Reasoning: LSTM networks are well-suited for sequential data modeling, making them a natural choice for time series forecasting tasks like stock price prediction. They are capable of capturing long-term dependencies in the data while mitigating the vanishing gradient problem often encountered in traditional recurrent neural networks (RNNs).

Advantages: LSTMs have memory cells that can maintain information over long periods, allowing them to capture complex temporal patterns present in financial data. This architecture enables the model to learn from historical price movements and technical indicators, leading to more accurate predictions.

Model Architecture:

Input Layer: The input layer accepts sequences of historical data and technical indicators.

LSTM Layers: Multiple LSTM layers are stacked to capture different levels of temporal dependencies in the data. Each LSTM layer processes the input sequences and passes its output to the next layer.

Output Layer: The final LSTM layer is followed by a dense output layer that predicts the closing prices for multiple future time steps.

Activation Functions: ReLU (Rectified Linear Unit) activation functions are commonly used in the LSTM layers to introduce non-linearity, while linear activation is used in the output layer for regression tasks.

```
Model: "Ensemble_Model"
__________________________________________________________________________________________________
 Layer (type)                   Output Shape         Param #     Connected to
==================================================================================================
 input_5 (InputLayer)           [(None, 1825, 308)]  0           []

 LSTM_Model (Functional)        (None, 182)          4507215     ['input_5[0][0]']

 GRU_Model (Functional)         (None, 182)          30035534    ['input_5[0][0]']

 CNN_Model (Functional)         (None, 182)          7702484     ['input_5[0][0]']

 NN_Model (Functional)          (None, 182)          59440874    ['input_5[0][0]']

 average (Average)              (None, 182)          0           ['LSTM_Model[0][0]',
                                                                  'GRU_Model[0][0]',
                                                                  'CNN_Model[0][0]',
                                                                  'NN_Model[0][0]']

 dense_15 (Dense)               (None, 182)          33306       ['average[0][0]']

==================================================================================================
Total params: 101,719,413
Trainable params: 101,718,333
Non-trainable params: 1,080
__________________________________________________________________________________________________
None
```

Training Objective:

The primary objective during the training phase is to minimize the mean squared error (MSE) loss function, which measures the discrepancy between predicted and actual closing prices.
The model learns to adjust its parameters (weights and biases) iteratively through backpropagation, optimizing them to make accurate predictions on unseen data.
Training Process:

Dataset Splitting: The dataset is split into training and validation sets. The training set is used to update the model's parameters during training, while the validation set helps monitor the model's performance and prevent overfitting.
Batch Processing: The training data is divided into batches to facilitate efficient computation and optimization. Each batch undergoes forward propagation to compute predictions and backward propagation to update the model's parameters.
Gradient Descent Optimization: The Adam optimizer or another suitable optimization algorithm is employed to minimize the MSE loss function. Adaptive learning rates provided by the optimizer help accelerate convergence towards the optimal solution.
Early Stopping: To prevent overfitting, early stopping is applied based on the validation loss. Training is halted if the validation loss fails to improve over a specified number of epochs.

Validation Phase (Using Test Data):

The test dataset, separate from the training and validation sets, is used to evaluate the model's performance on unseen data. This phase assesses the model's generalization ability and its ability to make accurate predictions on new data.

Evaluation Metrics:

Metrics such as mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE) are computed to quantify the disparity between predicted and actual closing prices.
These metrics provide insights into the model's predictive accuracy and are crucial for comparing the performance of different models or configurations.

Model Evaluation:

The trained LSTM model is applied to the test dataset to generate predictions for future closing prices.
The predicted prices are compared against the actual prices in the test dataset to assess the model's accuracy and effectiveness.

Visualizations, such as line plots or candlestick charts, can be used to visualize predicted versus actual prices, helping to identify patterns and discrepancies.

Hyperparameter Tuning:
Model performance can be further enhanced through hyperparameter tuning. Parameters such as the number of LSTM units, learning rate, batch size, and dropout rate can be adjusted to optimize the model's performance.

Model Iteration:
Based on the evaluation results, adjustments may be made to the model architecture or training process to address any shortcomings or improve performance.

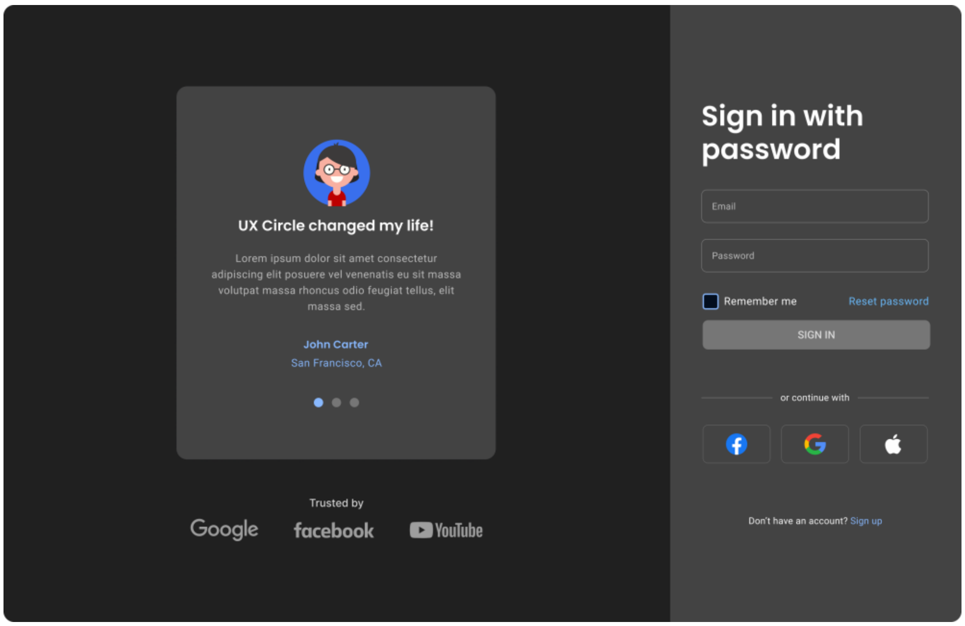
Iterative experimentation and refinement help iteratively improve the model's predictive accuracy and robustness.

```
Epoch 1/40
1542/1542 [==============================] - 1714s 1s/step - loss: 1407.501
Epoch 2/40
1542/1542 [==============================] - 1761s 1s/step - loss: 523.0694
Epoch 3/40
1542/1542 [==============================] - 1739s 1s/step - loss: 433.5826
Epoch 4/40
1542/1542 [==============================] - 1707s 1s/step - loss: 335.8069
Epoch 5/40
1542/1542 [==============================] - 1736s 1s/step - loss: 289.7407
Epoch 6/40
1542/1542 [==============================] - 1728s 1s/step - loss: 260.5909
Epoch 7/40
1542/1542 [==============================] - 1714s 1s/step - loss: 246.4886
Epoch 8/40
1542/1542 [==============================] - 1732s 1s/step - loss: 212.6625
Epoch 9/40
1542/1542 [==============================] - 1738s 1s/step - loss: 247.5602
Epoch 10/40
1542/1542 [==============================] - 1766s 1s/step - loss: 190.4338
Epoch 11/40
1542/1542 [==============================] - 1667s 1s/step - loss: 172.4104
Epoch 12/40
1542/1542 [==============================] - 1730s 1s/step - loss: 173.1429
Epoch 13/40
1542/1542 [==============================] - 1754s 1s/step - loss: 159.1414
Epoch 14/40
1542/1542 [==============================] - 1777s 1s/step - loss: 110.2315
Epoch 15/40
1542/1542 [==============================] - 1732s 1s/step - loss: 105.4090
Epoch 16/40
1542/1542 [==============================] - 1758s 1s/step - loss: 101.8306
Epoch 17/40
1542/1542 [==============================] - 1699s 1s/step - loss: 83.2076 -
Epoch 18/40
1542/1542 [==============================] - 1674s 1s/step - loss: 79.6317 -
Epoch 19/40
1542/1542 [==============================] - 1696s 1s/step - loss: 77.8040 -
Epoch 20/40
1542/1542 [==============================] - 1723s 1s/step - loss: 69.8421 -
```

Web Application Design:
Our web application, tentatively named "RoboStock Dashboard," is designed using Figma, a collaborative interface design tool. The prototype provides users with an intuitive interface to interact with stock price predictions and analysis tools.

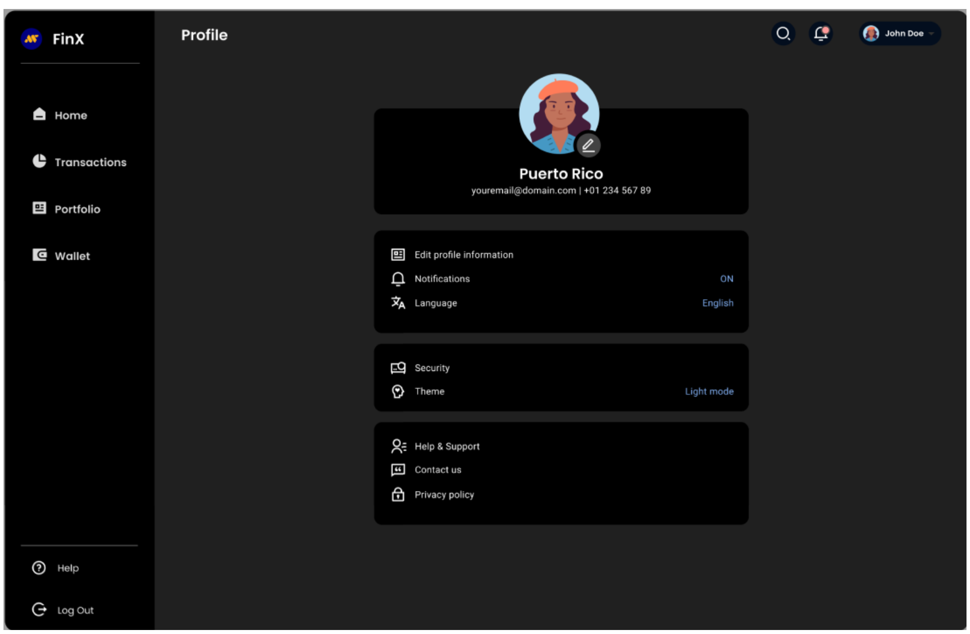
The dashboard design encompasses key components such as a navigation bar, prediction input form, prediction results display, and user profile management, all meticulously crafted for optimal user experience.

Prototype Development:
Leveraging Figma, we've developed a comprehensive prototype design for the RoboStock Dashboard, focusing on layout, styling, and interactive elements.



The prototype serves as a visual representation of the final product, allowing stakeholders to provide feedback and iterate on the design before implementation.

The design is tailored to meet user requirements and preferences, ensuring ease of navigation, clarity of information, and aesthetic appeal.



7. Evaluation

How your web service is tested

- Performance metrics
- Performance parameters
- Experimental design

How your solution performed, how its performance compared to that of other solutions mentioned in related work, and how these results show that your solution is effective

- Presentation and Interpretation
- Why, how, and to what degree our solution is better
- Why the reader should be impressed with our solution
- Comments

Context and limitations of your solution as required for summation

- What the results *do* and *do not* say

8. Claims

A claim or claims particularly pointing out and distinctly claiming the subject matter which you regard as your invention or discovery.

References

[1] Anderson, J., Ramamurthy, S., Jeffay, K., "Real-Time Computing with Lock-Free Shared Objects," *Proceedings of the 16th IEEE Real-Time Systems Symposium*, IEEE Computer Society Press, December 1995, pp. 28-37.

[2] Baruah, S., Howell, R., Rosier, L., "Algorithms and Complexity Concerning the Preemptively Scheduling of Periodic, Real-Time Tasks on One Processor," *Real-Time Systems Journal*, Vol. 2, 1990, pp. 301-324.

[3] Goddard, S., Jeffay, K., "Analyzing the Real-Time Properties of a Dataflow Execution Paradigm using a Synthetic Aperture Radar Application," *Proc. IEEE Real-Time Technology and Applications Symposium*, June 1997, pp. 60-71.