A Digital Cryptocurrency Investment Assistant with Ratings and Customized Recommendations

Team80 - Bofei Yu, Heurys Grullon, Icaro Aguiar, Kevin Faust, Yingdan Guo
byu93, hgrullon3, iaguiar3, kfaust6, yguo448@gatech.edu

1. INTRODUCTION

Our motivation stems from a noticeable lack of investing tools that allows for user customization, friendly visualization on portfolio tracking metrics, rankings and recommendations on crypto investing strategies based on user preferences and ML results. In today's landscape, digital cryptocurrency offers unprecedented opportunities and challenges for investors. The environment is dynamic and rapidly changing given all the influx of new technologies and investment options. We've decided to build a cutting-edge assistant that may empower investors with machine learning models and easy to digest data visualizations, thereby helping investors and those interested in investing to navigate the complex world of cryptocurrency investments with more confidence and less stress.

Our project focuses on creating a program that provides ranking and customized recommendations based on user features and past performances of cryptocurrencies. Users will easily utilize various metrics that are mostly absent from portfolio management applications to aid their investment decisions.

2. PROBLEM DEFINITION

Our project focuses on creating a program that provides ranking and customized recommendations based on user features and past performances of cryptocurrencies. Users will easily utilize various metrics that are mostly absent from portfolio management applications to aid their investment decisions.

3. LITERATURE SURVEY

Paper [1] looks at using Artificial Neural Networks, SVM, random forest, and naive-Bayes using 10 technical parameters and trend based. All techniques improved with trend based, and random forests were the best predictors. Paper [2] looks at using random forest for feature selection for neural networks with LSTM to improve performance in predicting stocks, which outperformed PCA.

This paper [3] looks at using logistic regression, SVM, neural networks, and random forest, and paper [4] proposed using graph neural network (GNN) to predict cryptocurrency prices. The results show that the GNN, SVM and logistic regression have better performances overall.

Unlike the previous papers, paper [5] presents an approach for stock price prediction using fundamental analysis. The article could be useful for the project if we consider that fundamental factors prevail over technical ones for scoring stocks. We also may find it useful to combine the approaches of price prediction and stock scoring, although this paper does not specifically explore this idea. Paper [6] introduces a hybrid approach for stock selection, combining stock scoring and stock prediction. The paper should therefore be quite useful for the project, as it suggests a way to combine future market information with intrinsic stock information.

Paper [7] presents stocks scoring idea using SVM, and paper [8] proposes evolutionary algorithm. Both papers hold great similarity with what we are trying to achieve in terms of modeling and feature extraction. Paper [9] used neural network for stock selection. Benchmark uses linear model, which has been regarded as the traditional approach for predicting a stock's future return.

Paper [10] proposes using tweets sentiment and volume in combination with neural networks to predict price direction and magnitude of change. The model achieved accuracy of 63%. Paper

[11] examines volatility of cryptocurrencies and long-term memory properties. The results show the chosen cryptocurrencies have significant long memory and various performance, supporting the use of GARCH, HYGARCH, and FIGARCH as suitable modeling techniques.

Paper [12] is like Paper [10] in which it displays the important role the online world has on cryptocurrency price movements. Paper [12] employs a set of empirical approaches (VAR, Copula, and more) and suggests the frequency of google searches leads to positive returns and a surge in trading volume. Utilizing search engine data can potentially increase the accuracy of our models.

Paper [13] considered the stock ranking prediction as a multi-class classification task by pairwise combinations of stocks and used SVM for classification. Paper [14] used Deep Multi-Task Learning to solve the list-wise stock ranking problem by multi-layer modeling; the easily trained tasks provides extra gradient backpropagation to the learning of hard trained ranking task.

Paper [15] analyzed the performance of ML-based investable portfolios (based on price, fundamental, sentimental, and averaging) and attribute the performance to linear vs. non-linear effects. Results show the long-term performance by linear models is table through time.

Cryptocurrency investment applications have historically provided users with limited real-time feedback and often lack comprehensive information. Investors are required to gather metrics from multiple sources, making the process inconvenient and time-consuming. The average consumer faces many challenges when it comes to cryptocurrency. There currently exists a bunch of obstacles and hoops the average user must jump through to understand and utilize cryptocurrencies in their portfolio. [16]

In response to this challenge, our application aims to provide users with a more interactive and easily digestible visually informative interface. The application will offer a comprehensive perspective on the cryptocurrency market and provide users with ranking and customized recommendations based on user features, simplifying the process of planning, and executing optimal cryptocurrency investment strategies. For a successful cryptocurrency application to exist, the application must include these key features [17]:

- Real-time Market Data: Users can access price information, historical data, and market analysis for various cryptocurrencies.
- Investment Recommendations: Utilizing ML algorithms our application will offer personalized investment recommendations based on individual risk tolerance and goals.
- Portfolio Visualization: Users can visualize their cryptocurrency portfolios, monitor their performance, and assess their overall exposure to different digital assets.
- Security and Risk Assessment: Advanced risk assessment tools to help users evaluate the safety of their investments and take appropriate measures to protect their assets.

4. LIST OF INNOVATION

1. Providing customizable feature selections which will allow users to include or exclude various features such as coin inclusion/exclusion, investment strategies, risk tolerance, min market cap, max market cap, and more.
2. Improved visualizations for portfolio performance evaluation will be implemented, including a “stock picker” system that relies on relative rankings derived from a range of metrics specified by user-defined parameters. The parameters include critical cryptocurrency metrics like time frames, minimum returns, maximum risks, and asset prices.
3. ML algorithms will be utilized and presented to investors who have limited trading and mathematical backgrounds in a user-friendly, non-technical manner.

5. PROPOSED METHOD

Our application should be better than the current cryptocurrency investment applications due to a noticeable lack of a user-friendly cryptocurrency investment tool. The level of personalization we aim to offer goes beyond what many current applications offer, which often provide generic universal recommendations that may not suit individual needs. The new application can prioritize accessibility for casual investors. Empowering novice investors to make informed decisions through custom-tailor made recommendations, bridging the knowledge gap that currently exists in the cryptocurrency market.

5.1 Visualization Features

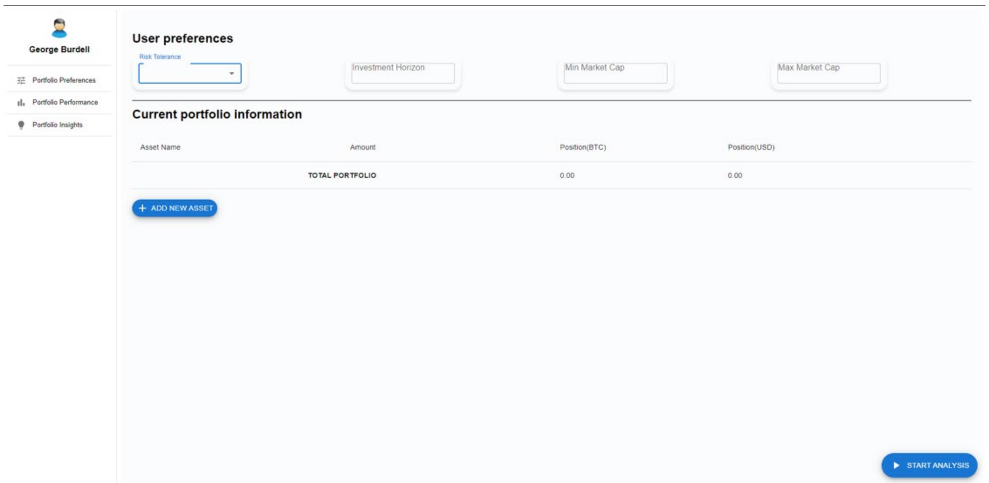


Figure 1. User preference page where user can see portfolio information and set their preferences.

We are designing an intuitive and user-friendly design to make the application accessible to users with limited financial or technical background using React. We enable users to select from various levels of risk tolerance, the amount of time they believe they will hold their portfolio, and the minimum and maximum market cap of the asset they're interested in.

Portfolio performance and insights will be demonstrated to users within this same user interface after their preferred inputs are selected, providing key portfolio performance information and insights to the user so that they may make an informed decision. Portfolio Insights can be seen in the figure below. Once the user selects their preferences and runs the analysis, ranking, customized recommendations, and insights will be generated using machine learning algorithms.

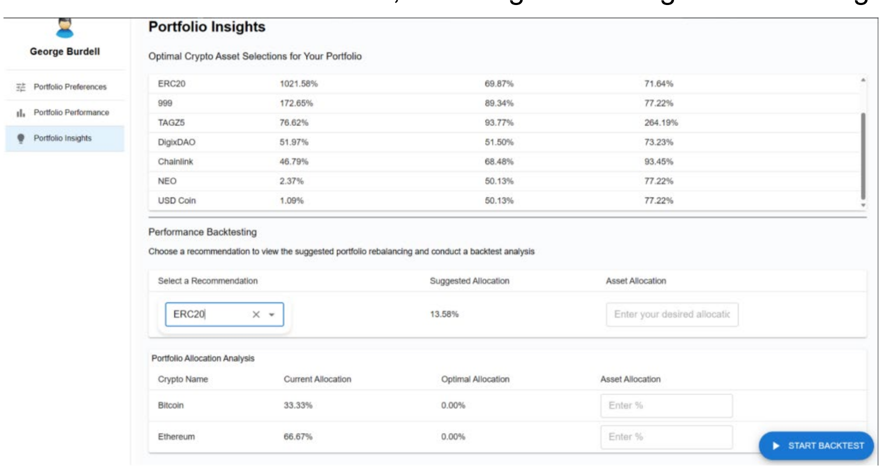


Figure 2. User insights page where user can see optimal asset rankings and start an analysis based off user-entered parameters

5.2 Data Sources

The cryptocurrency price dataset we are using is provided and downloaded from Kaggle. The dataset contains information such as trading volume, price in USD, market cap of many cryptocurrency assets. This data ranges from January 1st, 2016, to January 8th, 2022. The dataset consists of 2,382,643 rows

5.3 Machine Learning Algorithms and Insight Generation

Our price forecasting model consists of a hybrid time series model featuring both additive and multiplicative trends. This model is adept at capturing seasonal trends whether yearly, weekly, and monthly, plus holiday effects. The implementation is carried out using the Prophet library in Python. The Prophet library is specifically designed to handle time series data with strong seasonal patterns, as well as multiple seasonality and holiday effects. It's renowned for its ability to accurately forecast in various domains, including finance and economics. Prophet is designed to handle missing data gracefully, which allows us to forecast accurately even in the presence of incomplete information. Since Prophet is great at handling missing data, shifts in trends, and typically handles outliers well, we believe using it as our chosen model is a good choice, as it has excellent forecasting ability in an environment full of volatility, strong seasonal effects, and irregular events. This is key to our project, as we aim to provide substantial analysis and insights of a volatile market to users.

To dive into more details regarding our algorithm, let's begin with the training process. We train two separate models using Prophet on historical cryptocurrency ranging from Jan 1st, 2020 to Jan 8th, 2022. One model is trained to capture multiplicative seasonality effects, while the other is trained to capture additive seasonality effects.

Before training, the algorithm must specify 25 potential changepoints which are uniformly placed along the time series. We chose values for our changepoint scale that avoided both too much flexibility and not enough flexibility in capturing and projecting trends. After specifying changepoints, we must choose whether to capture additive and multiplicative trends. Instead of choosing between one, we believed combining the two would give the most accurate results in this case, which is what led us to creating two separate Prophet models. Next, we chose Fourier orders for the yearly and weekly seasonality effects. Since seasonality is estimated using a partial Fourier sum in Prophet models, we took care in determining this value, as higher orders allows for seasonality to fit to faster-changing cycles but it comes with the cost of potential overfitting.

With parameters selected for both our additive and multiplicative models, we were able to train them on historical data and receive two separate forecasts. We then combined the forecasts from the two models by taking the mean, which is an approach known as ensemble forecasting. By combining and taking the mean of our two forecasting models, we improve overall predictive accuracy and robustness.

Using our model, we can generate insights for our users. Whenever a user enters, we can initiate training on up to approximately 450 cryptocurrencies up to a 1 billion market cap threshold. Once the training is completed, the user inputs their current portfolio information. Afterwards, to generate insights, we perform a 6-month price prediction by selecting the top 20 most profitable assets that yielded positive returns. From those top ~20 selected assets, we combine each one with the user's existing portfolio. For each combination, we perform portfolio optimization and generate new allocation recommendations for both the user's current portfolio and a new recommended portfolio. The user can then choose from the recommendations and perform a backtest analysis, which checks how well the strategic allocation would have worked on historical data. Based on the analysis, a user can then select what they deem best for their portfolio and goals.

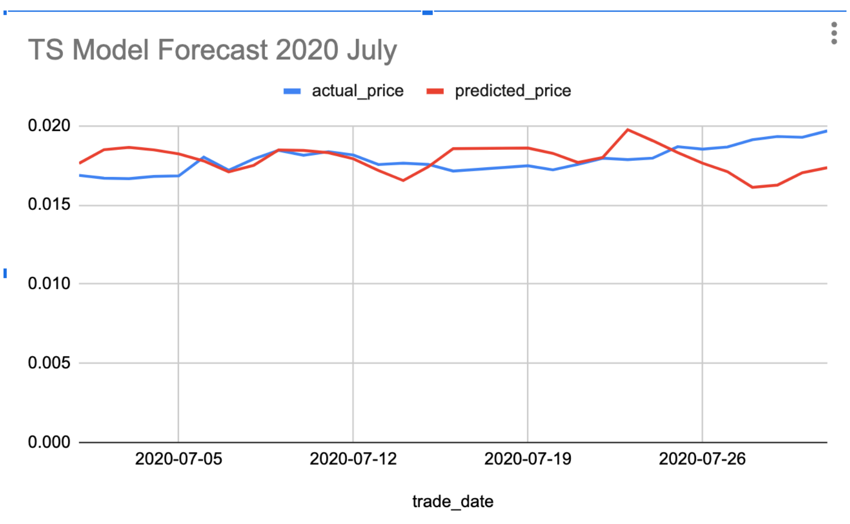
5.4 Testing and Validation

We thoroughly back tested our Prophet model to assess the performance against historical data. We tested a multiplicative model and an additive model and calculated the mean absolute percentage error of our ensemble forecast. By providing historical time series data to our two models with different configuration and parameters, we were able to choose the most reliable models that best captured trends within cryptocurrency historical data.

6. EXPERIMENTS/EVALUATION

To measure the success of our application, we employed model user studies and accuracy of trend forecasting. By utilizing user studies, we learned how this tool may help investors/analysts to achieve their goal. We evaluated the results of our model by comparing our model forecast to ground truth data. Although accuracy is not the goal for our investment tool, the models must give a reasonable prediction to direct the ranking.

Additionally, to determine how easily digestible and usable our cryptocurrency investment recommendation tool works, we believe it would be helpful to show the interface to users and capture their feedback using Qualtrics Survey. We presented a 5-minute survey to respondents that asked them questions on ease of use, interpretation, and the opportunity to provide feedback.



Prediction Error Rate

To evaluate the model performance, we used MAPE (Mean Absolute Percentage Error) rate to compare the actual market prices vs. predicted price on daily basis. The smaller the MAPE is, the better the model is tracking the market. The projected performance was estimated by the simulation process, which means, for each cryptocurrency, we trained the hybrid time series model on price history until T-1, and generate the T forecast by averaging time series models' results.

As a result, the estimated MAPE for TRON in July 2020 was about 7.6%. The above chart displayed how the predicted price (red) tracked the actual market price (blue) of TRON. Ideally the predicted (red) should overlap with actual (blue) if we achieved 100% accuracy.

Survey results

We gathered feedback from 10 users, the majority of whom identified as experienced investors. The average satisfaction score was 3.9 out of 5. Key areas highlighted for enhancement included customization options, user interface improvements, and more accurate price forecasting.

7. CONCLUSION AND DISCUSSION

A significant gap exists in the realm of cryptocurrency investment aid software. Given this evident absence, our user-friendly investment tool will empower investors ranging from casual to veterans with better decision-making opportunities. With our tool, users will have the opportunity to input metrics pertaining to their situation and quickly receive easily digestible insights and rankings derived from mathematical models to users, thereby alleviating the complexities and uncertainties around the cryptocurrency investing world and increasing their wealth.

Effort Distribution Statement:

In the project’s final phase, the team’s contributions were as follows:

Bofei Yu selected the dataset, created the prediction models, ran experiments, drafted the final poster, and managed questionnaires, as decided in agreement with the team.

Heurys Grullon drafted this report, as decided in agreement with the team.

Icaro Aguiar developed most of the code, enhanced the machine learning models, deployed the application for user testing, and complemented this report.

Kevin Faust partially completed the Portfolio Performance module and created the readme file.

Yingdan Guo did not contribute, appearing to have left the class.

REFERENCE

[1] Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques. Expert Systems with Applications, 42(1), 259-268.

[2] Khaidem, L., Saha, S., & Dey, S. R. (2016). Predicting the direction of stock market prices using random forest. arXiv preprint arXiv:1605.00003.

[3] Akyildirim, E., Goncu, A., & Sensoy, A. (2021). Prediction of cryptocurrency returns using machine learning. Annals of Operations Research, 297, 3–36. https://doi.org/10.1007/s10479-020-03575-y

[4] Wei Yin, Ziling Chen, Xinxin Luo & Berna Kirkulak-Uludag. “Forecasting cryptocurrencies’ price with the financial stress index: a graph neural network prediction strategy.” Applied Economics Letters (2022). Link: https://www.tandfonline.com/doi/full/10.1080/13504851.2022.2141436'

[5] S. D. Shende, A. S. Singh, S. S. Shah, M. M. Shinde, S. R. More and B. Ainapure, "Stocks Price Prediction by Fundamental Analysis Using Machine Learning Algorithms," 2022 5th

International Conference on Contemporary Computing and Informatics (IC3I), Uttar Pradesh, India, 2022, pp. 1515-1522, doi: 10.1109/IC3I56241.2022.10072563.

[6] Fengmei Yang, Zhiwen Chen, Jingjing Li, Ling Tang, A novel hybrid stock selection method with stock prediction, Applied Soft Computing, Volume 80, 2019, Pages 820-831, ISSN 1568-4946, https://doi.org/10.1016/j.asoc.2019.03.028.

[7] Zhang, Ru, et al. "Multi-factor stock selection model based on kernel support vector machine." J. Math. Res 10.9 (2018).

[8] Yu, Lean, Lunchao Hu, and Ling Tang. "Stock selection with a novel sigmoid-based mixed discrete-continuous differential evolution algorithm." IEEE Transactions on Knowledge and Data Engineering 28.7 (2016): 1891-1904.

[9] Levin, Asriel. "Stock selection via nonlinear multi-factor models." Advances in Neural Information Processing Systems 8 (1995).

[10] Wu, Y., Yang, Y., Wang, S., & Lai, K. K. (2022). Bitcoin price change and trend prediction through twitter sentiment and data volume Financial Innovation, 8(1), 37. https://doi.org/10.1186/s40854-022-00352-7

[11] Elngar, A. A., Youssef, A. M., & Loey, M. (2020). Long Memory in the Volatility of Selected Cryptocurrencies: Bitcoin, Ethereum and Ripple. Journal of Risk and Financial Management, 13(6), 107. https://doi.org/10.3390/jrfm13060107

[12] Ke, Z., Xu, P., & Xu, X. (2018. Forecasting cryptocurrency returns and volume using search engines. Financial Innovation, 4(1), 12. https://doi.org/10.1186/s40854-018-0119-8

[13] Engin Tas, Ayca Hatice Atli. “Stock Price Ranking by Learning Pairwise Preferences.” Computational Economics (2022). Link: https://link.springer.com/article/10.1007/s10614-022-10350-7

[14] Tao Ma, Ying Tan. “Stock Ranking with Multi-Task Learning.” Expert Systems with Applications: An international Journal, Volume 199, Issue C (Aug 2022). Link: https://doi.org/10.1016/j.eswa.2022.116886

[15] Stephane Daul, Thibault Jaisson, Alexandra Nagy. “Performance attribution of machine learning methods for stock returns prediction.” Pictet Asset Management, Acacias 60, 1211 Geneva, Switzerland (Apr 2022). Link: https://doi.org/10.1016/j.jfds.2022.04.002

[16] Chen, Y., Li, M., Li, X., Wang, L., & Wang, X. (2019). Research on Factors Affecting the Price of Bitcoin and Investment Strategy. In 2019 2nd International Conference on Management, Education Technology and Economics (METE 2019) (pp. 171-175). Atlantis Press.

[17] Li, X., Li, M., Chen, Y., & Wang, H. (2021). Cryptocurrency Investment Recommendations Based on User Behavior Analysis. IEEE Transactions on Network and Service Management, 18(2), 1350-1359.

Dataset Candidates:
https://www.kaggle.com/datasets/georgezakharov/historical-data-on-the-trading-of-cryptocurrencies/

New York Stock Exchange (kaggle.com)

S&P 500 Companies with Financial Information - Dataset - DataHub - Frictionless Data

Non-Financial Corporate Data| RAW (rutgers.edu)

Investing.com (scraping, [GitHub - alvarobartt/investpy: Financial Data Extraction from Investing.com with Python])

Questionnaire:

Survey Link: https://qfreeaccountssjc1.az1.qualtrics.com/jfe/form/SV_3CoTkVNQQI7TYGy

Section 1 - About yourself:

Q1. How would you describe your level of investment experience?

A. Beginner
B. Intermediate
C. Advanced

Q2. Investment Goals - What is your primary goal for investing?

A. Wealth accumulation
B. Retirement planning / Education funding
C. Other (please specify)

Q3. What is your anticipated investment time horizon?

A. Short-term (weeks)
B. Medium-term (months)
C. Long-term (years)

Q4. What investment strategy aligns best with your preferences?

A. Passive investing
B. Active trading
C. Balanced approach
D. Socially responsible investing
E. Other (please specify)

Q5. How often would you like to monitor and adjust your investment portfolio?

A. Daily
B. Weekly
C. Monthly
D. Quarterly
E. Annually

Section 2 - Ratings of the App:

Q6. Satisfaction: On a scale of 1 to 5, with 1 being very bad and 5 being very good, how would you rate your experience?

Q7. Anything that you recommend to further improve? (Select all that applies)

A. User interface
B. Price forecasting accuracy
C. Customization
D. Recommendation mechanism
E. Other (please specify)

Q8. Any other comments or suggestions?