Dan Licht & Marc Ojalvo

Milestone 1

Our project aims to determine what role cryptocurrencies will play in the post-pandemic environment and if the sector-specific coin model is sustainable. We will use machine learning models to create an industry-based comparative analysis of different cryptocurrencies, particularly during the COVID-19 pandemic and current economic recession. We will be looking at cryptocurrencies across five sector categories: Currency, Infrastructure, Financial, Services, and Media and Entertainment. We will be using several datasets to analyze and predict pricing, the average volume traded, and adoption rates of particular cryptocurrencies. These include historical pricing data and datasets specific to the pandemic, such as different government responses, economic data, and COVID-19 case numbers. Using the Google Dataset search engine, we have obtained these relevant datasets from trusted sources such as Statista, Johns Hopkins, and ESRI.

Since 2017, predictive analysis of cryptocurrencies (cryptos) has increased in popularity due to their volatility and appreciation. With a market cap of 1.03 billion dollars and a projected 1.4 billion dollar market by 2024, financial analysts worldwide are attempting to profit from the crypto market. Previous researchers have built crypto price prediction models applying standard currency pricing models such as fundamental analysis, technical analysis, and sentiment analysis using Natural Language Processing (NLP). More generally, asset price analysis and prediction is a common research field for economists. Although predicting prices based on financial theory has been done, adequate comparative analysis based on cryptos' intended use has not been done, especially during an economic recession. Researchers have yet to explore this topic because the pandemic period is the first economic recession in the U.S. since the crypto craze of 2017.

Cryptocurrencies have the potential to play a large role in our future economy. Given this possibility, it would be beneficial to gain insight into how cryptocurrency will affect the future's financial landscape. Understanding how people will use cryptos, which people will be using them, and what purposes people will use them can be useful for their financial well-being. The idea of using different coins for different purposes is unique to the cryptocurrency market and, therefore, will be a critical feature of our analysis.

We will be using existing models to derive correlations between our pricing data, economic data, and COVID-19 data. We will be using Python for this project because of its robust data science and machine learning tools. Python has packages such as Pandas and NumPy for data analysis and visualization data. Additionally, libraries such as scikit-learn, Tensorflow, and Keras provide tools to build machine learning models such as artificial neural networks and random forests. Mainly, recurrent neural networks (RNNs) and Long Short Term Memory neural networks (LSTMs) work well for time series data. Using these libraries, we will aim to find statistically significant information to help answer our questions.

We intend to use the CCXT library to gather data from the ten most reputable exchanges to obtain historical crypto pricing data. CCXT enables users to pull cryptocurrency data from popular exchanges with different programming languages, including Python. According to the SEC, only ten exchanges show actual trade volume, and we will pull data from these exchanges. We are using the Messari Sector Category breakdowns in order to categorize top cryptos by their industry. By analyzing one or two of the top-performing and most popular coins from each sector, we will attempt to find trends across our time series data.

Machine learning models like neural nets and random forests can yield insight into cryptocurrency prices' best predictors. Additionally, we will use traditional statistical and exploratory data analysis methods to identify trends and correlations between our pricing and pandemic related data. With the information found, we will include visualizations, such as plots and charts. Our project will contain dimensionality reduction techniques, like multidimensional scaling and principal component analysis (PCA), to relate multiple datasets to each other. We believe combining these techniques could yield unique insight regarding cryptocurrencies' performance during an economic recession. With this information, we will attempt to make predictions about cryptos' role in future society.

More specifically, we will be answering the following questions using specific methods and techniques:

1. *Which sector-based cryptocurrencies will maintain, gain, or lose value during a recession?*

Of the five sectors we are examining, we believe Currency and Media and Entertainment coins would withstand economic recession the best, while Financial coins would take the largest hit. To test this hypothesis, we will use the CCXT library to aggregate historical cryptocurrency pricing and Pandas and NumPy to analyze and visualize this data.

1. *Is it possible to predict cryptocurrency prices or average daily volume traded based on COVID and recession data?*

We hypothesize that with certain datasets, we can determine which cryptocurrencies, based on sector, will increase or decrease in value. To test this hypothesis, we will build a neural network with our pandemic-related time series data as input features and cryptocurrency pricing or volume traded over time as output.

1. *Which economic indicators best predict fluctuations in crypto prices?*

We believe consumer spending and restaurant and bar restrictions may best predict crypto pricing, while features such as real estate spending and travel restrictions will not be as telling. Using techniques like gradient boosting feature importance and permutation importance on our neural network, we will derive the relative importance of each of our features in predicting crypto pricing.

1. *Does a change in volatility affect the rate at which businesses adopt cryptocurrencies?*

We think that as cryptocurrencies become less volatile, more businesses will begin accepting coins in exchange for goods and services. To test this hypothesis, we will compare fluctuations in pricing over time to cryptocurrency adoption rates among global corporations.

Regarding feasibility, we think our project is accomplishable within the given time frame. One of our main obstacles will be aggregating historical time series pricing data across different exchanges over an extended period. Additionally, we must find a way to quantify and categorize our pandemic-related data like government restrictions on bars and restaurants and travel restrictions to use this information in machine learning models. If we are to build a neural net, we will also need to determine the best architecture, likely through trial and error. We will need to find whether an LSTM, RNN, or another architecture will be most powerful. We must also determine what activation functions to use, whether to use techniques like batch normalization and dropout, how many hidden layers to have, and other neural net specifications. Though neural nets and other machine learning models and tactics can get quite complicated, we are confident in our ability to harness these tools within our deadlines successfully.

We plan on doing most of the work together. We both have a background in working with datasets and data manipulation; the work we intend to complete is feasible and within our range of knowledge and ability. Nevertheless, Dan has a stronger background in neural nets, so he will do a larger portion of the programming. On the other hand, Marc is a finance major and understands how to use statistics to find prices. Our unique backgrounds will prove to be beneficial to the completion of our project.

References

Aungiers, Jakob. “Time Series Prediction Using LSTM Deep Neural Networks.” *Altum Intelligence*, 1 Sept. 2018, www.altumintelligence.com/articles/a/Time-Series-Prediction-Using-LSTM-Deep-Neural-Networks.

Ccxt. “Ccxt/Ccxt.” *GitHub*, 2019, github.com/ccxt/ccxt.

Cerliani, Marco. “Feature Importance with Neural Network.” *Medium*, Towards Data Science, 21 Jan. 2020, towardsdatascience.com/feature-importance-with-neural-network-346eb6205743.

“COVID-19 Global Travel Restrictions and Airline Information.” *Humanitarian Data Exchange*, 2020, data.humdata.org/dataset/covid-19-global-travel-restrictions-and-airline-information.

“COVID-19: U.S. Housing Policies by State 2020.” *Statista*, Jennifer Rudden, 27 Aug. 2020, www.statista.com/statistics/1111306/policy-housing-states-covid19-usa/.

“Cryptocurrency Market.” *Market Research Firm*, <https://www.marketsandmarkets.com/Market-Reports/cryptocurrency-market-158061641.html#:~:text=Cryptocurrency%20is%20a%20disruptive%20concept,6.18%25%20during%20the%20forecast%20period.>

“Crypto News, Pricing, and Research.” *Messari Crypto News*, messari.io/classifications.

“Exploratory Data Analysis.” *Wikipedia*, Wikimedia Foundation, 31 Aug. 2020, en.wikipedia.org/wiki/Exploratory\_data\_analysis.

Fusaro, Teddy, and Matt Hougan. “Bitwise Asset Management.” *Sec.gov*, 19 Mar. 2019, www.sec.gov/comments/sr-nysearca-2019-01/srnysearca201901-5164833-183434.pdf.

*Johns Hopkins Coronavirus Resource Center*, 2020, coronavirus.jhu.edu/.

Palaniappan, Vivek. “Neural Networks to Predict the Market.” *Medium*, Towards Data Science, 23 Sept. 2018, towardsdatascience.com/neural-networks-to-predict-the-market-c4861b649371.

“Restaurants and Bars Restrictions.” *COVID-19 Resources*, 2020, coronavirus-resources.esri.com/datasets/b5409d33573344e18d1fa634eb41507e\_0/data?geometry=149.558,-30.835,-148.918,66.754.

“U.S. Unemployment Rate: Adjusted, August 2020.” *Statista*, Erin Duffin, 8 Sept. 2020, www.statista.com/statistics/273909/seasonally-adjusted-monthly-unemployment-rate-in-the-us/.