PROJECT PROPOSAL

Title of the Project: Crypto market capitalization change based on S&P 500 movements.
Student: Khalil Garaali - 40226310
Date submitted: 2/16/2025

1. **Abstract**

   This project aims to predict daily changes in the cryptocurrency market capitalization, which refers to the total value of all coins in circulation, based on movements in the S&P 500 index. By analyzing historical correlations between traditional equity markets and digital assets, I seek to determine whether stock market trends can serve as predictive indicators for cryptocurrency market fluctuations.

   The dataset will include historical crypto market cap data from Kaggle, S&P 500 price movements from Kaggle and Yahoo Finance if possible, and additional macroeconomic indicators such as Bitcoin dominance or the fear and greed index if needed. Data preprocessing will involve handling missing values, feature engineering, and merging time-series datasets.

   Performance will be evaluated using regression models, with Root Mean Square Error (RMSE), R-squared (R^2), and Directional Accuracy (MDA) as primary metrics. I will also explore time-series models such as LSTMs to capture market trends. It would also be interesting to test out transformer-based models and graph neural networks (GNNs).

   Key challenges include market volatility, finding datasets that align well together in terms of date-time, potential nonlinear relationships, and external macroeconomic factors influencing both markets. Addressing these will require careful feature selection and robust modeling techniques. If successful, this study could enhance understanding of cross-market influences and provide insights for traders and analysts navigating the crypto market.

2. **Problem Statement**

   Many crypto traders and investors overlook the fact that cryptocurrency markets are influenced by stock market movements, particularly the S&P 500. Ignoring this correlation can lead to incomplete market analysis and poor investment decisions resulting in loss of money. Solving this problem is important because recognizing the influence of stock market movements on crypto can help traders make more informed decisions, reduce risks, and develop better trading strategies based on a more complete market analysis,

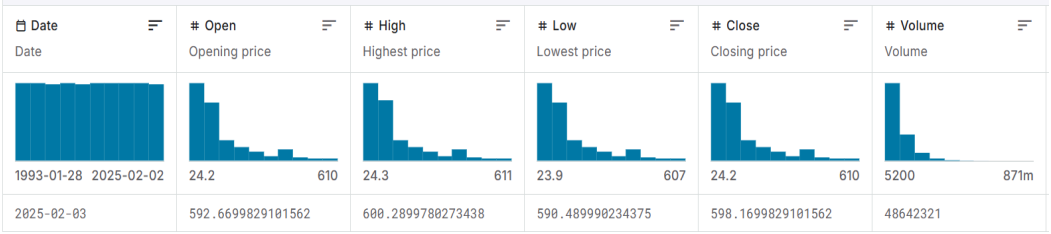
thus reducing people's stress in terms of money, especially in these economic times.

3. **Solution Statement**
   The solution is to predict daily changes in the crypto market capitalization based on S&P 500 movements. The solution is feasible since recent studies show a strong correlation between crypto and the S&P 500 movement. Other macroeconomic indicators can be added to reinforce the accuracy of the model. I will explore regression models such as XGBoost and Random Forest for their ability to capture nonlinear relationships, as well as LSTMs to model time-series dependencies. These methods are well-suited for financial forecasting due to their capacity to handle complex patterns and temporal trends.

4. **Literature Review**
   Previous research has applied machine learning to cryptocurrency price prediction and portfolio management. One study used a convolutional neural network (CNN) trained with historical price data in a reinforcement learning framework, achieving significant returns in backtesting. Another approach, KryptoOracle, integrated real-time Twitter sentiment analysis with an adaptive learning model to improve crypto price predictions. Additionally, Long Short-Term Memory (LSTM) networks have been used to forecast cryptocurrency closing prices by capturing time-series dependencies. These studies mainly focus on price prediction.

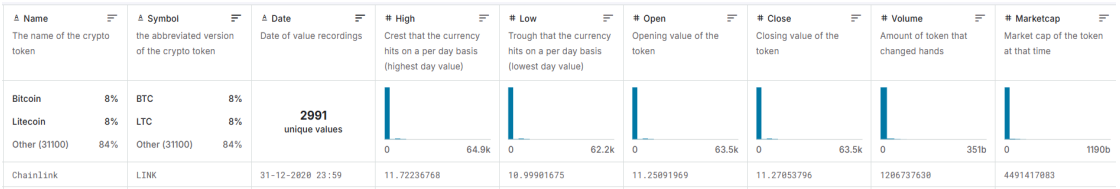
5. **Originality**
   Unlike previous studies that mainly focus on predicting individual cryptocurrency prices using deep learning for sentiment analysis, my approach examines the broader relationship between S&P 500 movements and crypto market capitalization. This provides a macroeconomic perspective, helping investors understand how traditional financial markets influence crypto trends. My approach is novel as this relationship has not been extensively studied before.

6. **Datasets**
   The data consists of tabular, time-series financial data from two sources gathered from Kaggle: one for S&P 500 price and volume data over a long time frame (1993 - 2025) and another for cryptocurrency market cap, price, and volume over a shorter period of time (2018-2021). Each input will include features such as date, S&P 500 volume and open/close prices, and the market cap of the top 50 to 100 cryptocurrencies from the same time frame.



Date	# Open	# High	# Low	# Close	# Volume
Date	Opening price	Highest price	Lowest price	Closing price	Volume
1993-01-28 2025-02-02	24.2 610	24.3 611	23.9 607	24.2 610	5200 871m
2025-02-03	592.6699829101562	600.2899780273438	590.489990234375	598.1699829101562	48642321

S&P 500 input data



Name	Symbol	Date	# High	# Low	# Open	# Close	# Volume	# Marketcap
The name of the crypto token	the abbreviated version of the crypto token	Date of value recordings	Crest that the currency hits on a per day basis (highest day value)	Trough that the currency hits on a per day basis (lowest day value)	Opening value of the token	Closing value of the token	Amount of token that changed hands	Market cap of the token at that time
Bitcoin 8% Litecoin 8% Other (31100) 84%	BTC 8% LTC 8% Other (31100) 84%	2991 unique values	0 64.9k	0 62.2k	0 63.5k	0 63.5k	0 351b	0 1190b
Chainlink	LINK	31-12-2020 23:59	11.72236768	10.99981675	11.25091969	11.27053796	1206737630	4491417083

Crypto market input data

The expected output will be the daily percentage change in total crypto market cap or altcoin market cap. The goal of the model is to find patterns that emerge between the S&P 500 movement and the crypto market cap.

The dataset will require preprocessing, including handling missing values and aligning time frames between the stock and crypto data. Training and testing splits will follow a chronological order, ensuring that past data is used to predict future trends to avoid data leakage. The dataset size will determine whether training can be done locally or requires cloud resources like Google Colab or Gradient with better GPU acceleration. Proper feature scaling and normalization may also be necessary to improve model performance.

7. **Describe how you will measure “success.”**

Success will be determined by whether the model can accurately predict upcoming cryptocurrency market cap movements based on S&P 500 trends, demonstrating a strong correlation between the financial markets. Key performance metrics will include Root Mean Squared Error (RMSE), R^2 score to assess prediction accuracy and Directional Accuracy (MDA). Additionally, backtesting on historical data will help confirm whether incorporating stock market trends improves crypto market predictions. A successful outcome would show a consistent performance over different test inputs and provide evidence that crypto markets react to stock market movements. Failure would mean high prediction errors, weak correlation, or a model that does not generalize well. Even if the model fails to establish a strong link, the project / research remains valuable if it provides conclusive insights into the relationship between these markets.

8. **Describe the main challenges**
   One major challenge is ensuring high-quality and reliable data, as financial markets are influenced by numerous unpredictable factors. Finding data about the S&P 500 and the crypto market in the same time frame will also prove to be difficult. Another challenge is capturing the complex, non-linear relationships between S&P 500 movements and crypto market capitalization, which may require advanced feature engineering and hyperparameter tuning. Computational constraints could also arise when training deep learning models, but I will mitigate this by using Google Colab or Gradient for GPU acceleration.

9. **Describe the computational resources you will use:**
   I will primarily use a personal computer with a multi-core CPU (Ryzen 5) and GPU (RTX 4050) for model training and evaluation. For more computationally intensive tasks, such as training deep learning models, I may leverage Google Colab or Gradient, which provide better access to GPUs and TPUs.

10. **List the main Python packages you expect to use**
    I expect to use Scikit-learn for machine learning models like Random Forest and XGBoost, and TensorFlow/Keras for deep learning approaches such as LSTMs. Pandas and NumPy will be used for data manipulation. Additionally, Matplotlib and Seaborn will be used for data visualization and analysis.