Crypto Forecast Analyzer in Machine Learning

Team Members
Patel Nency (A20550076)
Patel Saloni (A20558369)
Patel Darshan (A20527204)

1. Introduction

In this study, we look at the dynamic and diverse realm of cryptocurrencies, which has attracted global attention due to their high volatility and emerging importance in the financial ecosystem. The inclusion of a diversified range of digital currencies—Bitcoin,Cardano,BinanceCoin, Ethereum,Dogecoin,Litecoin,Monero,Stellar,Tether, and XRP—extends the scope of our research, recognizing the varying attractiveness different cryptocurrencies have for a wide range of investors and financial institutions. This research digs into the critical role that advanced computing technologies like Python and Jupyter Notebooks play in processing and analyzing financial data. Python's rich libraries and broad use in financial analytics provide unrivaled capabilities in data manipulation, statistical computations, and machine learning methods.

The financial industry has witnessed substantial transformations since the introduction of cryptocurrencies, digital assets designed to function as methods of exchange that use encryption to protect transactions. Bitcoin and other cryptocurrencies have emerged as highly volatile but potentially valuable assets, attracting significant interest from investors, traders, and researchers alike. This project focuses on using machine learning techniques, notably Long Short-Term Memory (LSTM) networks and the Prophet forecasting tool, to estimate bitcoin prices.

The objectives of projects are:

- Develop and compare LSTM and Prophet models for their effectiveness in forecasting cryptocurrency prices.

- Analyze the performance of these models based on historical data, evaluating their accuracy and reliability in predicting future market movements.

- Explore the applicability of machine learning in providing actionable insights for cryptocurrency trading and investment strategies.

2. Problem Statement

The cryptocurrency market is highly volatile, with prices evolving substantially over short periods of time. While volatility provides enormous profit potential, it also carries major hazards for investors and traders. The challenge is precisely anticipating these price swings so that traders may make informed selections. Traditional financial forecasting approaches frequently fail to capture the complex dynamics and inherent volatility of bitcoin values.

This project addresses the requirement for improved machine learning models that can accurately and timely estimate bitcoin prices. Using Long Short-Term Memory (LSTM) networks and the Prophet forecasting tool, we hope to create models that can understand and anticipate the subtle patterns of bitcoin price changes. The comparison of LSTM and Prophet will shed light on the advantages and disadvantages of employing deep learning vs a more typical time series forecasting approach in the context of cryptocurrency price prediction.

The effective development and deployment of these models has the potential to greatly improve trading methods, reduce risks, and provide new profit opportunities in the volatile cryptocurrency market. This research aims to increase the accuracy of cryptocurrency forecasts while also contributing to the larger field of financial technology by investigating novel machine learning applications.

3. Data Collection and Preprocessing

The initial phase entails thoroughly cleaning the dataset to correct missing numbers, eliminate inconsistencies, and remove outliers. Given the diversity of cryptocurrencies, normalization or standardization processes are used to align disparate scales, allowing for comparison studies. Furthermore, feature engineering plays an important role in improving the dataset by introducing new variables that may reveal hidden trends among different cryptocurrencies.

3.1 Data Sources

The dataset used in this bitcoin forecasting project came from Kaggle, a popular website that contains a wide range of datasets supplied by a diverse community of data scientists, researchers, and hobbyists.The chosen dataset contains historical price data for a variety of cryptocurrencies, providing an excellent foundation for creating forecasting models.
Here is link to download:
https://www.kaggle.com/datasets/sudalairajkumar/cryptocurrencypricehistory

3.2 Data Description

The dataset includes a full collection of metrics for each cryptocurrency, such as daily opening prices, high and low prices of the day, closing prices, and trading volumes. This information is critical for analyzing market dynamics and forecasting future price fluctuations. Each cryptocurrency's data contains

Date: Timestamps identify each entry, enabling for chronological analysis and modeling of time-dependent patterns.
Open: The price at which a cryptocurrency is first traded when the exchange opens on a given day.
High: The highest price at which a cryptocurrency traded during the day.
Low: The lowest price at which a cryptocurrency traded during the day.
Close: The price at which a cryptocurrency was last traded when the exchange closed on any particular day.
Volume: The total volume of cryptocurrency exchanged in a 24-hour period.
Market Cap: The overall market value of cryptocurrencies circulating supply.

3.3 Preprocessing Steps

Given the project's expanding scope, which now includes cryptocurrencies such as Bitcoin, Cardano, Binance Coin, Ethereum, Dogecoin, Litecoin, Monero, Stellar, Tether, and XRP, data preparation and visualization are even more critical. This stage is critical to assuring the integrity and usability of the obtained data across many cryptocurrencies.

3.3.1 Data Cleaning

In the cryptocurrency project, the data cleaning procedure began with the combining of distinct files for each coin into a single dataset. This consolidation was essential for conducting a thorough investigation of the bitcoin industry. The dataset, normally in CSV format, was then put into a Python environment and manipulated with packages such as Pandas and NumPy.
A key step involved assigning unique asset IDs to each cryptocurrency, facilitating analysis and application of machine learning models. The data cleaning method addressed issues such as missing values, outliers, and standardizing timestamps. Consistent date and time formats throughout the collection were required for accurate temporal analysis.

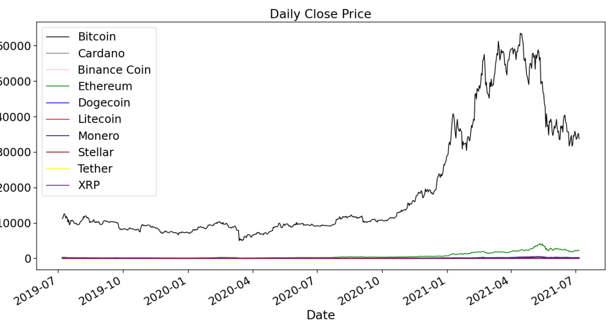
Date	id	Name	High	Low	Open	Close	Volume	Marketcap	Target	assetid
2017-07-26 23:59:59	0	Binance Coin	0.109013	0.099266	0.105893	0.105138	2.003950e+05	1.051380e+07	0	1
2017-07-27 23:59:59	1	Binance Coin	0.108479	0.100888	0.105108	0.107737	3.444990e+05	1.077370e+07	1	1
2017-07-28 23:59:59	2	Binance Coin	0.109019	0.101473	0.107632	0.104067	3.425680e+05	1.040670e+07	0	1
2017-07-29 23:59:59	3	Binance Coin	0.111264	0.101108	0.104782	0.107811	3.402180e+05	1.078110e+07	1	1

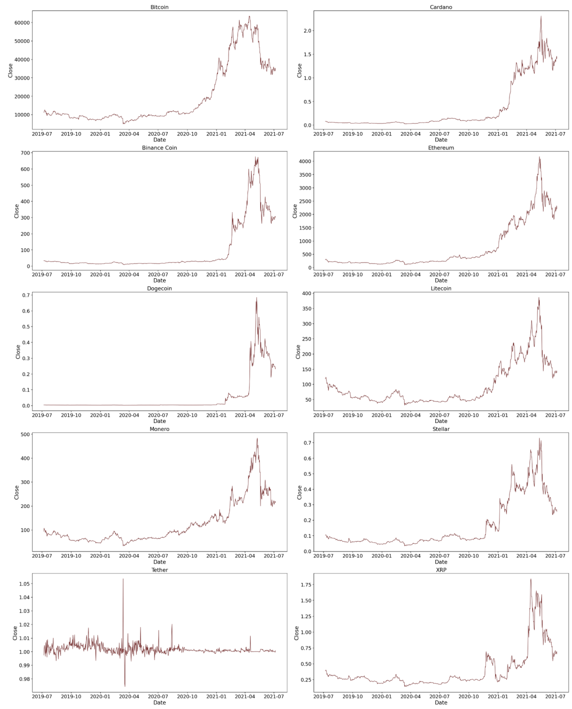
3.3.2 Exploratory Data Analysis

The EDA phase of our cryptocurrency forecasting research used a number of analytical and visualization tools to identify underlying patterns, trends, and anomalies in bitcoin market data. A big part of our investigation entailed creating line graphs for the closing values of ten major cryptocurrencies: Bitcoin, Cardano, Binance Coin, Ethereum, Dogecoin, Litecoin, Monero, Stellar, Tether, and XRP.

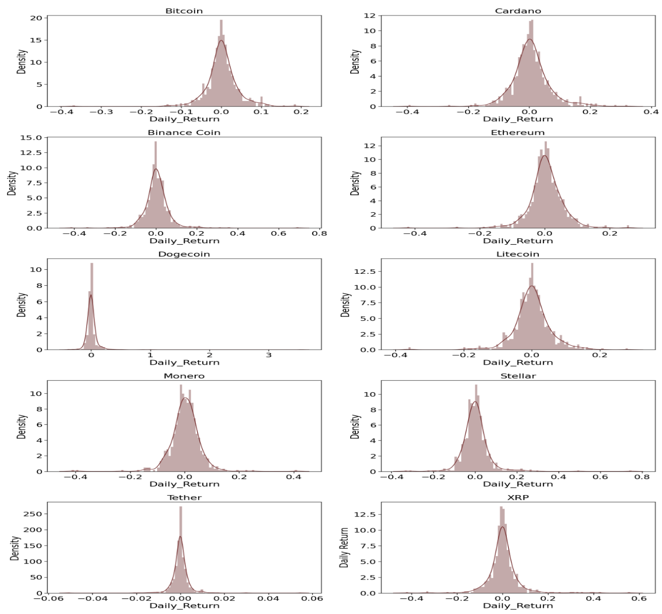
Visualization of Closing Prices:

We used the seaborn and matplotlib packages to build line plots of each cryptocurrency's closing price over time.



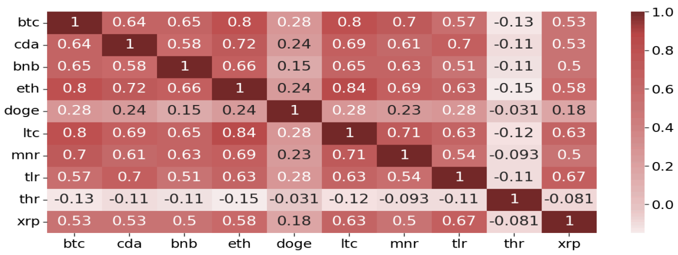


We intended to get insight into each cryptocurrency's risk profiles and return characteristics by visualizing its daily return distributions. We used Seaborn's Displot function to construct distribution plots for each cryptocurrency's daily results.



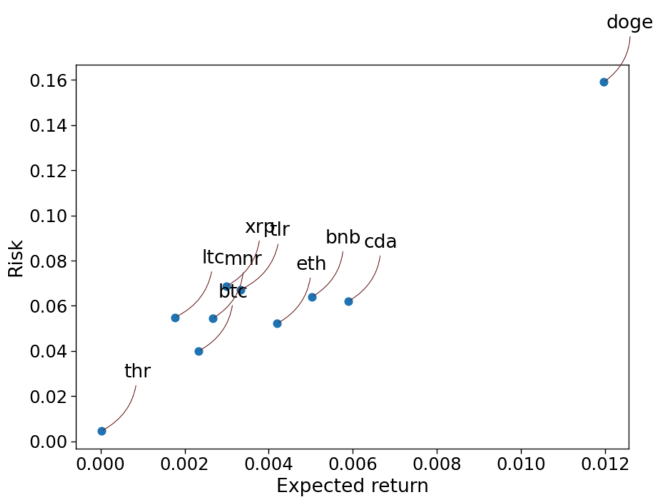
3.3.3 Correlation and Risk

Correlation Among Ten Cryptocurrencies' Closing Prices:



We calculated the Pearson correlation coefficients between the cryptocurrency' closing prices during the observation period. This was illustrated using a correlation heatmap, which provides a simple graphical picture of how one cryptocurrency's closing price movement affects others.

Risk and Expected Returns for Ten Cryptocurrencies:



Understanding cryptocurrencies' investment potential requires evaluating their risk and expected return. This analysis calculates the volatility (a proxy for risk) and average returns of each cryptocurrency, laying the groundwork for portfolio optimization and risk management measures. Risk (Volatility): We estimated the standard deviation of each cryptocurrency's daily returns and used it to assess risk. Expected Return: The average daily return for each cryptocurrency throughout the recorded period was calculated to determine its expected return.

4. Implementation

4.1 LSTM Model Implementation:

Long short-term memory is a type of recurrent neural network architecture used in deep learning. Unlike normal feedforward neural networks, LSTMs have feedback connections. It can process lengthy sequences of data as well as individual data points.

Model Architecture:

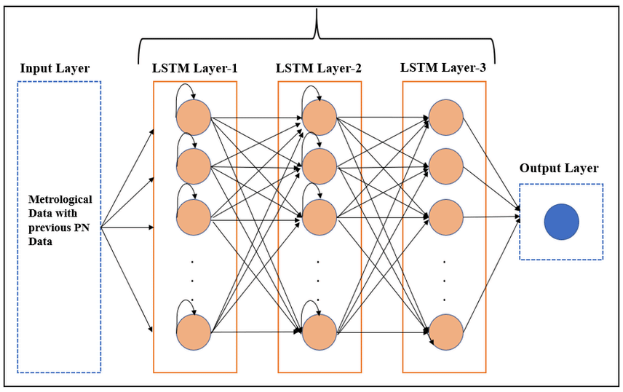
The LSTM model architecture for predicting Bitcoin prices could be as follows:

The Input Layer is the first layer of the model that accepts sequence data.

LSTM Layers: One or more LSTM layers with a fixed number of neurons each. These layers can learn from time-series data.

The Dense Layer is a fully linked layer that interprets the information learned by the LSTMs.

The output layer is a single neuron with a linear activation function that predicts the continuous value of the Bitcoin price.



Data Integration and Preprocessing: The notebook combines Bitcoin price data with attributes from a Wikipedia dataset. It prepares the data by defining a 'tomorrow' column to retain future closing prices, which are then utilized to create a binary 'target' column that indicates if the price increased. The preprocessing phase involves

The normalization of the 'close' price with the MinMaxScaler ensures that the LSTM model receives data on a scale that improves its performance.

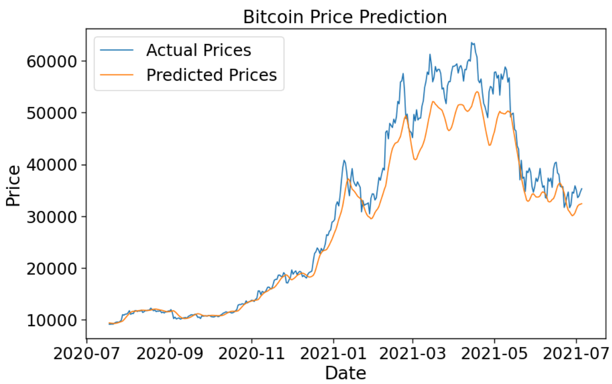
The notebook generates sequences from pricing data to estimate future points. A sequence of 60 time steps is chosen to provide the model enough historical context to learn and predict.

Model Architecture and Training:

To minimize overfitting, the LSTM model consists of two LSTM layers separated by dropout layers. After Adam optimization and mean squared error loss, the model is trained on the prepared sequences for 20 epochs.

Post-training, the model forecasts Bitcoin values based on the test set. The forecasts are rescaled back to their original price range, and the root mean squared error (RMSE) measure is used to assess the model's performance.

Visualization of the model's predictions compared to actual Bitcoin values, providing a visual assessment of how effectively the LSTM model captures cryptocurrency price trends and volatility.



4.2 Prophet Model Implementation

Data Collection and Preparation:

The procedure starts with the collection of Bitcoin price data, which usually includes timestamps and price information like opening, closing, high, and low values. Concurrently, relevant Wikipedia statistics such as page visits and edit counts are acquired. The two datasets are integrated using timestamps to ensure that each pricing entry corresponds to Wikipedia data.

Data Preprocessing:

This stage requires preparing the data to match the Prophet's specifications. The combined dataset is altered such that it has two columns: 'ds', the datetime column, and 'y', the column indicating the variable we want to forecast, which in this case is likely to be Bitcoin's closing price. Any null values or outliers that may bias the model are addressed during preprocessing.

Model Configuration and Training:

With the data prepared, a Prophet forecasting model instance is created. This stage may include adjusting the model to account for weekly and annual seasonality, as well as including Wikipedia data as extra regressors if there is a chance they will influence Bitcoin prices. The model is then trained on historical data with its fit approach.

Real-Time Data Fetching:

An initial step is taken to obtain live Bitcoin price data. Using the yfinance package, which provides a handy API for downloading financial data from Yahoo Finance, the script retrieves the most recent Bitcoin price data. Typically, this data retrieval is done by supplying the Bitcoin ticker symbol, as well as the required time period and frequency.

Data Preprocessing for Live Prediction:

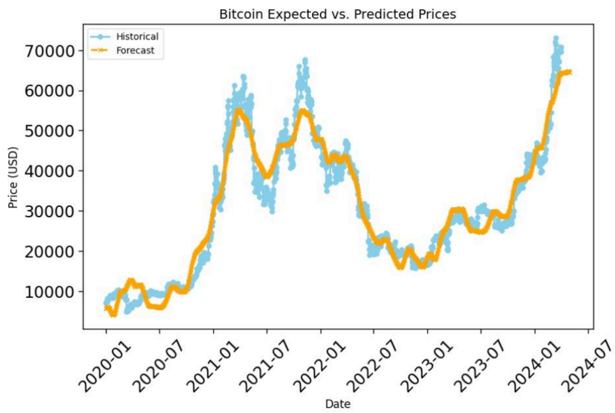
The live data collected from yfinance is then preprocessed to meet the Prophet model's format requirements. This involves renaming the columns 'ds' for the datetime index and 'y' for the closing price, as well as scaling the data as necessary during the initial model training.

Updating the Model with Live Data:

In this stage, the Prophet model, already trained on historical data, can be updated with the latest live data to refine its predictions. This step is crucial for maintaining the model's relevance and accuracy over time.

Forecasting with Updated Model:
The revised model can now generate fresh projections. A future DataFrame is constructed, which may include dates beyond the current date in order to estimate future prices. The model makes predictions based on the entire dataset, including live data.

Visualization of Live Predictions:
To visually evaluate the model's performance using live data, the predictions are shown against the historical data. This image demonstrates the model's real-time forecasting power and how the freshly added live data interacts with historical trends.



The forecast model appears to closely track the historical trend, indicating that it may have captured the data's underlying trends. However, real future values may differ due to the volatility nature of cryptocurrency markets and unanticipated external factors not factored into the model.
This visualization helps investors and analysts comprehend probable future price movements based on past trends, but it should be emphasized that it is not a guarantee of future performance and should be used in conjunction with other market analyses.

Our team is hard working on estimating future Bitcoin prices using advanced forecasting models presented in the latest research, as well as on model train, in which future prediction dates are constructed using Prophet's make_future_dataframe function. If more regressors were included, the model would require future values. The model forecasts future Bitcoin prices based on the parameters it has learned for these dates.

As we keep perfecting the model with new data and approaches, our goal remains to produce a well-informed forecast that captures the possible directions that the cryptocurrency could go. While acknowledging the inherent uncertainties in such unpredictable markets, we are committed to leveraging this model to aid in strategic decision-making.

5. References

[1] S. McNally, J. Roche and S. Caton, "Predicting the Price of Bitcoin Using Machine Learning,"

[2] I. Yenidoğan, A. Çayir, O. Kozan, T. Dağ and Ç. Arslan, "Bitcoin Forecasting Using ARIMA and PROPHET

[3] Jigar Patel, Sahil Shah, Priyank Thakkar, Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques

[4] “A. Ng, "Linear Regression With Multi Variable", Stanford, CA” 3.

[5] “D. Nelson, A. Pereira, and R. de Oliveira, "Stock Market’s Price Movement Prediction With LSTM Neural Networks", in Neural Networks (IJCNN).