Batch No – 29

BITCOIN PRICE PREDICTION

Faculty Coordinator – Mrs. Subha Meenakshi

Team Members :-

Sanjana V - 1BI20AI043

Shreyas R S - 1BI20AI046

Abstract

In the dynamic landscape of cryptocurrency markets, predicting the price movements of assets like Bitcoin has become a challenging yet crucial task for investors and traders. This project harnesses the power of machine learning to develop a robust model for forecasting Bitcoin prices.

Our approach involves collecting and preprocessing extensive historical data, encompassing various market indicators and relevant external factors. Leveraging advanced machine learning algorithms, including but not limited to regression models, neural networks, and ensemble methods, we strive to capture the intricate patterns within the cryptocurrency market.

Our results are validated through comprehensive back testing and out-of-sample testing, assessing the model's ability to generalize to unseen data. By comparing our predictions against actual market movements, we aim to demonstrate the effectiveness and reliability of our machine learning-based approach.

Additionally, we discuss the challenges faced during the project, potential limitations, and avenues for future improvements. The ultimate goal of this endeavor is to provide valuable insights for stakeholders in the cryptocurrency space, aiding them in making informed decisions in a volatile and rapidly evolving market.

Through this presentation, we share our journey in developing a predictive model for Bitcoin prices, shedding light on the potential applications of machine learning in understanding and forecasting cryptocurrency market dynamics.

Introduction

In the ever-fluctuating landscape of cryptocurrency, predicting Bitcoin prices has emerged as a critical endeavor, offering potential rewards for those adept at navigating market volatility. This project harnesses the capabilities of machine learning to address this challenge, aiming to develop a robust model that can decipher the intricate patterns governing Bitcoin's price movements. Cryptocurrency markets, epitomized by their dynamism, necessitate innovative approaches to forecasting, and our exploration into machine learning promises insights that extend beyond mere predictions.

Through this presentation, we share our journey in developing a predictive model for Bitcoin prices. From meticulous data collection and feature engineering to the optimization of machine learning algorithms, our project encapsulates the complexities of the cryptocurrency realm. Our objective is not only to enhance predictive accuracy but also to contribute to the understanding of the factors influencing Bitcoin markets. Join us as we navigate the convergence of technology and finance, unveiling the potential applications of machine learning in unraveling the mysteries of cryptocurrency price dynamics.

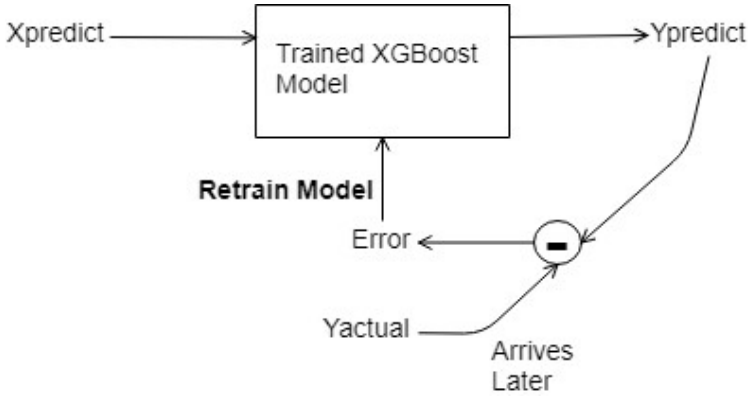
Literature Survey

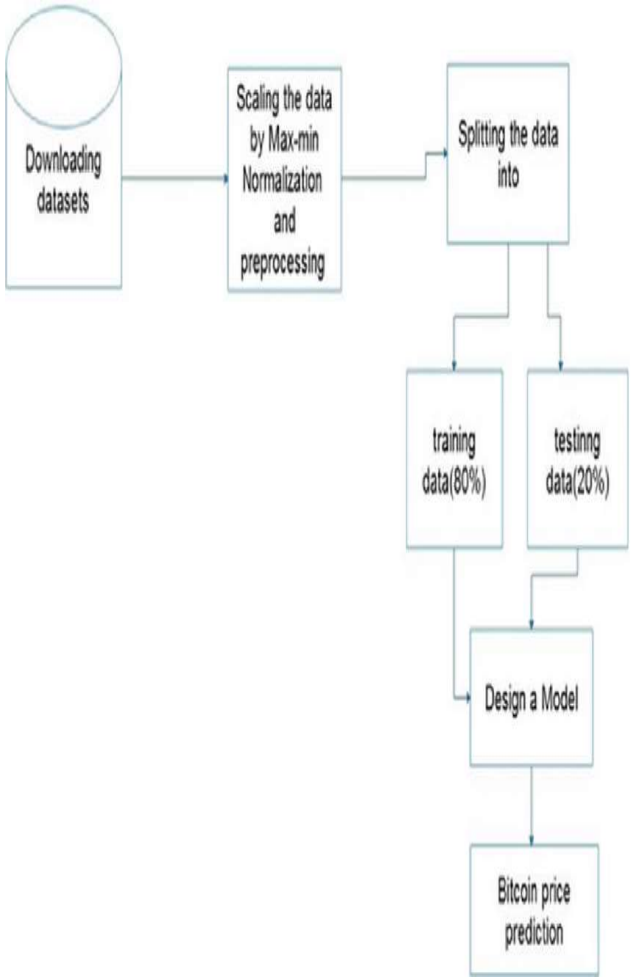
S.no	Author name and year of Publication	Title name and Journal Name	Abstract or objectives	Techniques used	Limitations
1	Antonis Polemitis and Antonis C. Simotas 2015	Forecasting Financial Time Series with Machine Learning Algorithms European Journal of Operational Research	This paper explores the application of machine learning algorithms for financial time series forecasting, including Bitcoin prices. The objective is to evaluate the predictive performance of various algorithms in the context of cryptocurrency markets.	Support Vector Machines, Neural Networks, Random Forests	Limited consideration of external factors influencing cryptocurrency markets.
2	Yaya O.S. and Ogbonna A.O. 2016	Modelling and Forecasting Bitcoin Volatility Index Financial Innovation	The paper focuses on modeling and forecasting the volatility of Bitcoin prices. Objectives include understanding the key determinants of Bitcoin volatility and assessing the efficacy of different models in capturing and predicting volatility patterns.	Generalized Autoregressive Conditional Heteroskedasticity (GARCH), Autoregressive Integrated Moving Average (ARIMA)	Primarily centered on volatility rather than price prediction.
3	Moazeni, S., & Bhowmik, T. 2018	Forecasting daily exchange rate using hybrid model of ARIMA and machine learning techniques Journal of Ambient Intelligence and Humanized Computing	The study proposes a hybrid model combining ARIMA and machine learning techniques for forecasting exchange rates, including Bitcoin. Objectives involve evaluating the synergy of these models in enhancing prediction accuracy.	ARIMA, Support Vector Machines, Neural Networks	Limited focus on specific features relevant to cryptocurrency markets.
4	Kristoufek, L. 2015	What are the main drivers of the Bitcoin price? Evidence from wavelet coherence analysis PLOS ONE	This paper employs wavelet coherence analysis to identify the main drivers influencing Bitcoin prices. The objective is to uncover the relationships between Bitcoin and various macroeconomic and financial indicators.	Wavelet Coherence Analysis	Limited application of machine learning techniques.

S.no	Author name and year of Publication	Title name and Journal Name	Abstract or objectives	Techniques used	Limitations
5	Wang, Y., Ma, Y., Wang, J., & Liu, W. 2015	Forecasting the Price of Bitcoin Using Social Media Sentiment Analysis IEEE International Conference on Data Mining Workshop (ICDMW)	The paper explores the integration of sentiment analysis from social media into Bitcoin price forecasting models. Objectives include understanding the impact of social sentiment on cryptocurrency markets.	Sentiment Analysis, Machine Learning Models	Reliance on the assumption that social media sentiment accurately reflects market sentiment.
6	Garcia, D., & Schweitzer, F 2015	Social signals and algorithmic trading of Bitcoin Royal Society Open Science	This paper investigates the influence of social signals on Bitcoin trading strategies. Objectives involve understanding how social media and online communities impact trading decisions in cryptocurrency markets.	Algorithmic Trading Models	Limited emphasis on price prediction models.
7	Kim, Y.B., Lee, J.H., & Ki, E.J. 2016	Predicting the price of Bitcoin using Bayesian regression International Journal of Business and Management	The study employs Bayesian regression to predict Bitcoin prices. Objectives include assessing the Bayesian approach's effectiveness in capturing the probabilistic nature of cryptocurrency price movements.	Bayesian Regression	Potential sensitivity to choice of priors in Bayesian analysis.
8	Shah, D., & Zhang, Q. 2019	Bitcoin Price Prediction and Trading Strategy using Machine Learning with Python International Journal of Advanced Research in Computer Science	This paper presents a comprehensive approach to Bitcoin price prediction and trading strategy development using machine learning. Objectives involve building a model capable of generating profitable trading signals.	Machine Learning Models, Trading Strategy Development	Limited discussion on potential overfitting and generalization issues.
9	Mollah, M.A.H., & Tariq, U. 2018	Predicting the price of Bitcoin using Artificial Neural Network Procedia Computer Science	The study focuses on predicting Bitcoin prices using Artificial Neural Networks (ANN). Objectives include evaluating the performance of ANN in capturing the complex patterns inherent in cryptocurrency markets.	Artificial Neural Networks	Potential challenges in training neural networks with limited data.

Proposed System

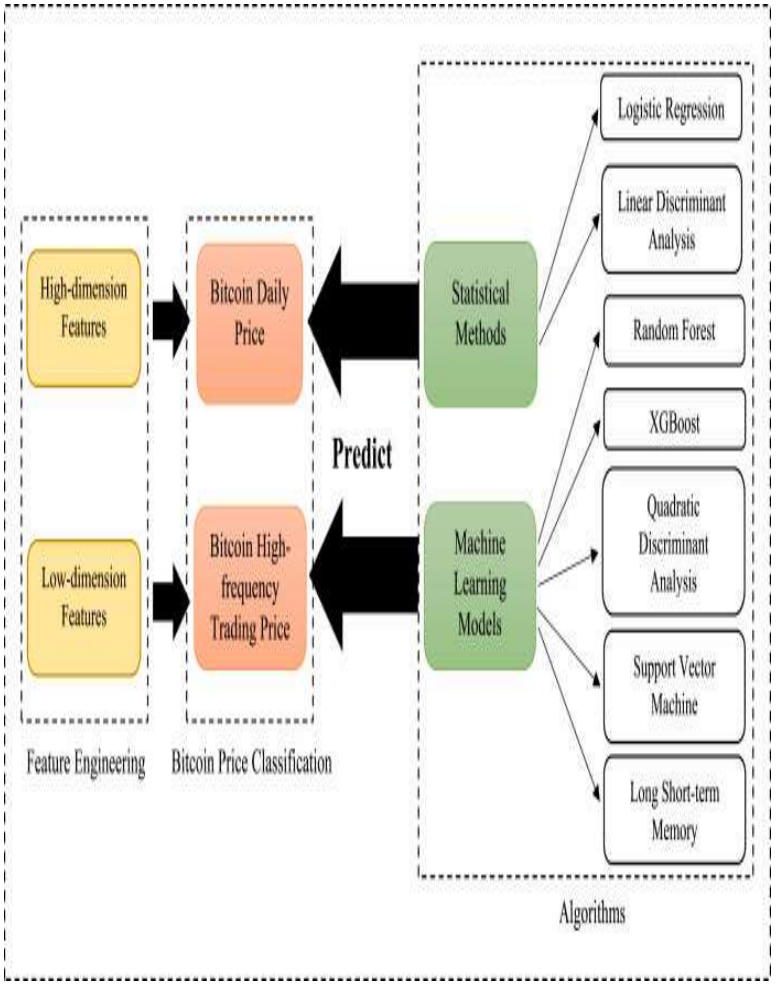
The proposed system entails a comprehensive Jupyter Notebook designed for Bitcoin price prediction using the XGBoost algorithm. The notebook guides through a meticulous process, from data preprocessing to model creation and evaluation. It begins by ensuring data integrity with an absence of missing or null values, followed by an exploration of key trends through exploratory data analysis (EDA) for specific years, offering valuable insights into Bitcoin's price dynamics. The dataset spans from September 17, 2014, to February 19, 2022, covering 2712 days and includes essential features such as Date, Open, High, Low, and Close. The predictive modeling phase involves splitting the data into training and testing sets, followed by the implementation of the XGBoost algorithm from the sklearn library. Post-training, the model's performance is assessed using Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). The proposed system culminates in visualizations that juxtapose predicted 10-day price charts against actual closed prices, providing a holistic view of the model's effectiveness in forecasting Bitcoin prices.





Block Diagram

- **Data Collection and Downloading**: Obtain historical Bitcoin price data from reliable sources such as financial data APIs, cryptocurrency exchanges, or financial websites.
- **Data Preprocessing:** Clean the data by handling missing values, removing outliers, and addressing any data quality issues. Convert the data into a format suitable for analysis, with date-time indexing if applicable.
- **Scaling the Data by Max-Min Normalization:** Use Max-Min normalization to scale the Bitcoin price data to a specific range, typically between 0 and 1. This will ensure that all features have the same scale and prevent biases in the model. Splitting the Data into Training and
- **Testing Sets:** Split the preprocessed data into training data (approximately 80% of the data) and testing data (approximately 20% of the data). This will allow you to train the model on a subset of the data and then evaluate its performance on unseen data.
- **Feature Engineering:** Identify relevant features that could impact the price of Bitcoin, such as volume, previous price trends, market sentiment, etc. It's crucial to take into account various factors that could influence the price, and create features based on these factors.
- **Model Design and Selection**: Choose an appropriate model for time series prediction, such as Long Short-Term Memory (LSTM) networks, Autoregressive Integrated Moving Average (ARIMA) models, or other neural network architectures. Consider incorporating additional external factors such as macroeconomic indicators, news sentiment, or social media activity that could influence Bitcoin prices.
- **Training and Validation**: Train the model using the training data and validate its performance using the testing data. This involves optimizing model hyperparameters and evaluating its predictive accuracy.
- **Evaluation and Validation**: Use appropriate metrics to evaluate the model's performance, such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), or Mean Absolute Error (MAE). Ensure the model doesn't exhibit overfitting to the training data and generalizes well to new, unseen data.
- **Prediction and Future Forecasting**: Utilize the trained model to make predictions on future Bitcoin prices, and implement strategies to assess the model's predictive capabilities in real-time.



System Architecture

XGBoost Algorithm

XGBoost is a decision-tree-based ensemble Machine Learning algorithm that uses a gradient boosting framework. In prediction problems involving unstructured data (images, text, etc.) artificial neural networks tend to outperform all other algorithms or frameworks. However, when it comes to small-to-medium structured/tabular data, decision tree based algorithms are considered best-in-class right now. Please see the chart below for the evolution of tree-based algorithms over the years. XGBoost algorithm was developed as a research project at the University of Washington. Tianqi Chen and Carlos Guestrin presented their paper at SIGKDD Conference in 2016 and caught the Machine Learning world by fire. Since its introduction, this algorithm has not only been credited with winning numerous Kaggle competitions but also for being the driving force under the hood for several cutting-edge industry applications. As a result, there is a strong community of data scientists contributing to the XGBoost open source projects with ~350 contributors and ~3,600 commits on GitHub

System Requirements

Hardware Requirements:

-CPU : Processor i5 or more, GPU

-RAM : 8GB

-Operating System

Software Requirements:

-Jupyter Notebook

-Libraries and Xgboost library

-Python 3.8

Non Functional Requirements

Performance:

The system should exhibit high performance in terms of model training and prediction speed, ensuring timely processing of data.

Scalability:

The system should be scalable to handle an increase in the size of the dataset or additional features without a significant decrease in performance.

Reliability and Availability:

The system should operate reliably, minimizing downtime or disruptions during model training and prediction tasks.

Module Design:

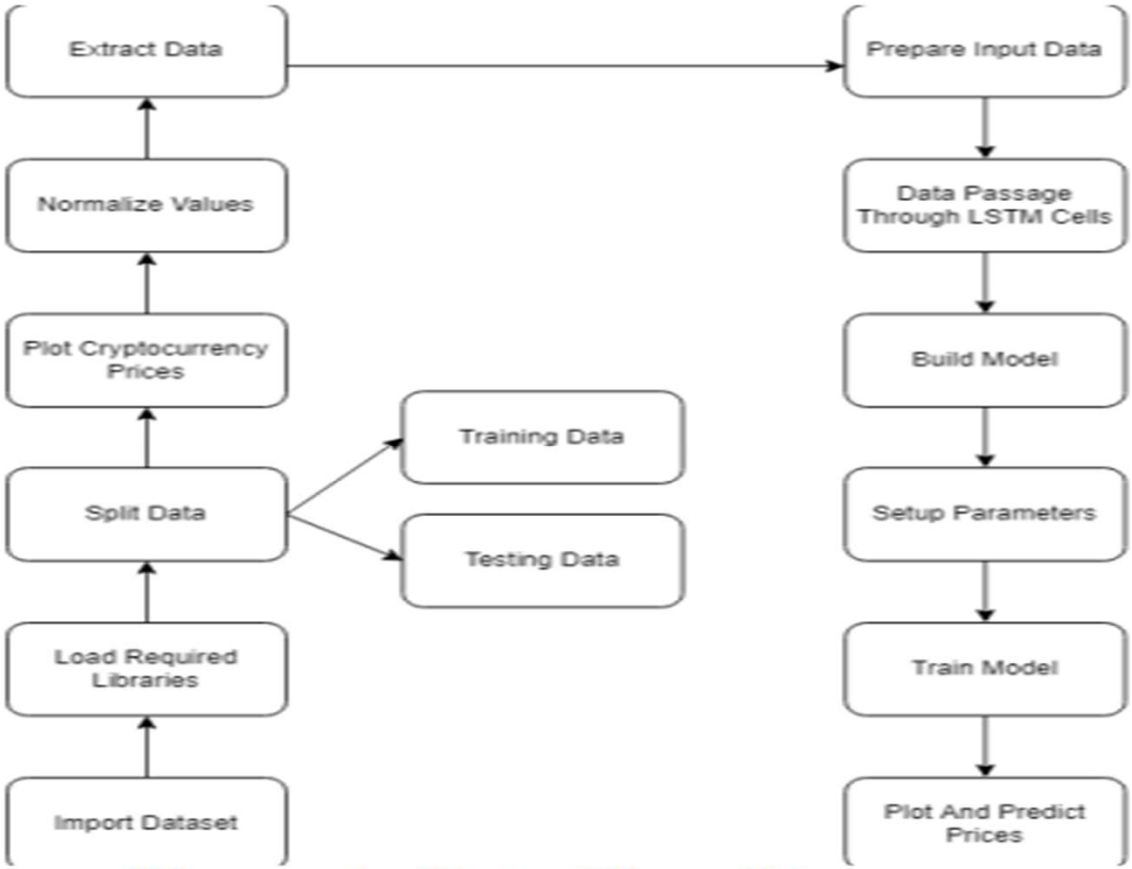
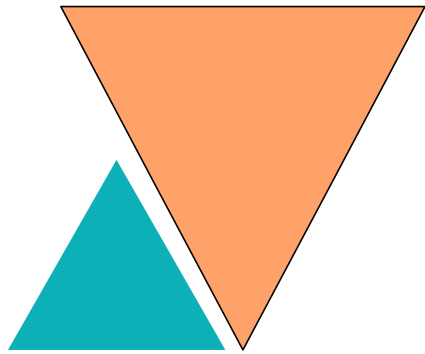


Figure 1: Data Flow Diagram



Thank You