University of Hertfordshire

School of Engineering and Computer Science

MSc Computer Science (Data Science and Analytics)

Advanced Computer Science Masters Project (7COM1039)

2023

Detailed Project Proposal

**Crypto currency Price Prediction using Machine Learning**

Student Name: Ezugwu Thankgod Chukwuebuka

Student ID: 21084271

Supervisor: Liang Chen

1. **Project Introduction**



The study of crypto currency price dynamics has grown in popularity as a result of their explosive rise. Digital or virtual currencies that utilise cryptography to safeguard transactions and regulate the generation of new units are known as crypto currencies. They run on decentralised platforms like blockchain technology and provide special benefits including anonymity, accessibility on a global scale, and large potential earnings. Considering that there are hundreds of crypto currencies, it is essential for traders, investors, and fans to comprehend and anticipate their price alterations.

With this project, I want to use machine learning to solve the issue of anticipating bitcoin values. Since crypto currency markets are so erratic, it is difficult to predict their prices with any degree of certainty. It is a difficult and dynamic problem to address since a variety of factors, including market mood, news events, regulatory changes, and investor behaviour, may affect bitcoin pricing. I will use machine learning algorithms to create predictive models for changes in crypto currency prices because they have shown promise in identifying patterns and trends in financial data.

This research was motivated by the rising popularity of crypto currencies and the demand for accurate tools to predict price fluctuations. Although crypto currency markets present huge opportunity for investors, they also carry a number of hazards. I can help investors make wise decisions, reduce risks, and maybe boost their profits by creating precise price prediction models. The initiative also fits with the overarching objective of investigating machine learning's potential in financial analysis and forecasting.

Creating machine learning models to forecast bitcoin values is the goal of this project. To train and assess various algorithms, I will use key market indicators and historical price data. Data preparation, feature engineering, model selection, and performance assessment will all be part of the project. I will work to develop precise and reliable prediction models by investigating various machine learning approaches, including regression models, time series analysis, and deep learning architectures. The outcomes of this study will not only advance the area of crypto currency analysis but also give traders and investors important information that they can use to make well-informed choices.

**2.0 Project Goal or Aim**

**Aim:**

The aim of this project is to create and use machine learning models that can precisely forecast crypto currency values. The plan is to develop reliable prediction models that will help investors, traders, and fans make educated decisions in the extremely volatile crypto currency market by using previous price data and pertinent market indicators.

**Objectives:**

1) Create machine learning models that are precise and trustworthy for forecasting bitcoin values.

2) To improve the forecasting power of the models, incorporate pertinent market indicators and past pricing data.

3) Improve the models to attain high accuracy in forecasting bitcoin prices.

4) Assess the models' performance using the proper metrics and contrast it to baseline methods or benchmark models.

**2.0– Project requirements (or features, or functionalities)**

The project requirements for Crypto currency Price Prediction using Machine Learning include access to reliable and comprehensive historical price data, computing resources, software and libraries, pre-processing and feature engineering, and model development and evaluation. Data requirements include reliable and comprehensive historical price data, market indicators, computing resources, software and libraries, pre-processing and feature engineering, and model development and evaluation. Computing resources include sufficient computing power and resources, software and libraries, pre-processing and feature engineering, and model development and evaluation. Software and libraries include Python, TensorFlow, scikit-learn, Keras, Pandas, NumPy, or statsmodels. Pre-processing and feature engineering involves developing data pre-processing pipelines to handle missing values, outliers, and inconsistencies in the crypto currency data, and feature engineering involves extracting meaningful features from the available data.

Model development and evaluation involves developing and evaluating models to predict crypto currency prices. The most important details in this text are the exploration and implementation of machine learning algorithms suitable for price prediction, design and implementation of appropriate training and testing protocols, utilization of hyper parameter tuning techniques, user interface and deployment, documentation and collaboration, and communication of the project's findings, insights, and limitations to stakeholders and the wider community. Machine learning algorithms such as regression models, time series analysis techniques, and deep learning architectures are used to predict crypto currency prices. Cross-validation strategies and evaluation metrics are used to assess the models' performance. Hyper parameter tuning techniques are used to optimize and fine-tune the models for improved prediction accuracy.

2.1Core project requirements

• Gather historical price data for various crypto currencies from reliable sources, including Bitcoin, Ethereum, Litecoin, etc.

• Pre-process the collected data by handling missing values, outliers, and inconsistencies to ensure data integrity.

• Engineer relevant features from the historical price data, such as moving averages, relative strength index (RSI), and volume indicators.

• Incorporate additional market indicators that can impact crypto currency prices, such as news sentiment, social media activity, or regulatory announcements.

• Select and implement suitable machine learning algorithms for price prediction, such as regression models (e.g., linear regression, random forest), time series analysis techniques (e.g., ARIMA, LSTM), or deep learning architectures (e.g., neural networks).

• Split the data into training and testing sets, utilizing appropriate cross-validation strategies (e.g., k-fold cross-validation, time-series cross-validation) to evaluate model performance.

• Train the models using the training dataset and tune hyper parameters to optimize their performance.

• Evaluate the models' performance using various evaluation metrics, including mean squared error (MSE), mean absolute error (MAE), or accuracy, to assess their effectiveness in predicting crypto currency prices.

• Continuously iterate and refine the models by adjusting the algorithm configurations, feature selection, or data pre-processing techniques to improve prediction accuracy.

• Conduct thorough analysis and interpretation of the model's predictions, providing insights into factors influencing crypto currency prices.

**2.2Advanced Project aims**

• To achieve the highest possible levels of accuracy and performance for the selected machine learning algorithms, surpassing existing benchmark models or state-of-the-art approaches in crypto currency price prediction.

• To conduct extensive comparative analysis of multiple machine learning algorithms, evaluating their strengths, weaknesses, and suitability for different crypto currency markets and time periods.

• To explore and implement advanced feature engineering techniques, such as sentiment analysis using natural language processing (NLP) on news articles or social media data, to capture more nuanced market dynamics and improve prediction accuracy.

• To investigate ensemble learning techniques, combining multiple machine learning models or algorithms, to further enhance prediction accuracy and robustness.

**3.0 – Secondary Research Aims (Literature Review)**

* Identify and review existing literature and research papers related to crypto currency price prediction using machine learning techniques. Evaluate various machine learning algorithms to assess their strengths, weaknesses, and performance. Select the most promising algorithms that align with the project's objectives and have demonstrated success in predicting crypto currency prices.
* Identify key features and indicators used in previous research for crypto currency price prediction, such as historical price data, trading volume, social media sentiment, news sentiment, or macroeconomic indicators, to determine the most effective and informative features to be incorporated in the project.
* Conduct an extensive literature search to gather results from state-of-the-art research in the field of crypto currency price prediction. Analyze performance metrics such as accuracy, precision, recall, or F1 score to establish a baseline for performance evaluation. Use the collected results as a benchmark to evaluate and compare the performance of developed models in terms of accuracy, efficiency, or novelty.

The goal of the literature review is to provide readers a thorough overview of the state of the field, assist in choosing the best algorithms and features, and set a standard for the project's performance. The project may build on prior work, make wise judgements, and advance knowledge in the area of machine learning-based crypto currency price prediction by completing a complete literature analysis.

**4.0 – Primary Research**

* **How you intend to train your implemented algorithm with the dataset**

1. Using the dataset to train the implemented algorithm:
2. The dataset, which consists of key market indicators and historical price data, will be split into training and testing sets.
3. The machine learning algorithm will be trained using the training set and the chosen features and methods.
4. To create predictions on data that has not yet been seen, the algorithm will learn the correlations and patterns found in the training data.
5. Depending on the algorithm and dataset size, several training techniques, such as batch training or online training, may be used.
6. The performance of the algorithm will be optimised by applying hyper parameter tuning methods like grid search or random search.

* **How you intend to test the algorithm, again with the dataset that you have obtained from somewhere. Think about how much time you will allow for these phases, because planning your project is extremely important**

Analysing the dataset and the algorithm:

1. The performance and generalizability of the algorithm will be assessed using the testing set, which is distinct from the training set.
2. The trained model and the supplied features will be used by the algorithm to generate predictions on the testing set.
3. To evaluate the accuracy and dependability of the method, the projected values will be contrasted with the actual values from the testing set.
4. The performance of the algorithm will be measured using several evaluation metrics, such as mean squared error (MSE), mean absolute error (MAE), or correctness.
5. To get more solid and trustworthy assessment findings, cross-validation approaches like time-series cross-validation and k-fold cross-validation may be used.

* **How you intend to evaluate your results. Have you established the criteria for evaluation in the Literature Review? Have you planned to collect results from other projects as well?**

Analysing the outcomes:

1. During the phase of collecting current research findings and benchmarks known as the literature review, the assessment criteria for the project will be defined.
2. The findings gathered from other projects will be used as a benchmark to assess the effectiveness of the proposed algorithm.
3. The efficacy and contribution of the project will be evaluated based on the outcomes using the predetermined criteria, such as accuracy, efficiency, or uniqueness.
4. Additionally, the findings will be subjected to qualitative analysis and interpretation in order to gather knowledge of the variables affecting crypto currency prices as well as the model's propensity for prognostication.

5) The project's findings will be compared to current research and literature to identify strengths, weaknesses, and potential areas for further advancement.

5.0 **Project resources**

I am going to use software mostly for this project. The software used for execution of artefact (Python Code) will be Jupyter notebook or Google Colab. And for writing report, we will use Microsoft Office and Ms-Paint for creation of diagrams.

**6.0 Project risks and their mitigation**

A table should be created here with project risk information:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **Risk description** | **Probability** | **Possible effects** | **Mitigation methods** |
| 1 | Insufficient or low-quality data | Medium | Inaccurate or biased predictions | Collect data from multiple reliable sources |
| 2 | Overfitting of the model | High | Poor generalization to unseen data | Regularize the model, use cross-validation |
| 3 | Lack of interpretability of the model | Low | Difficulty in understanding model decisions | Use interpretable models or interpretability techniques (e.g., SHAP values) |
| 4 | Rapidly changing market dynamics | High | Inaccurate predictions due to changing patterns | Implement a mechanism to update and retrain models frequently |
| 5 | Hardware or software failures | Low | Disruption of model training or prediction | Implement backup systems, regular maintenance, and monitoring |

**7.0 Project Plan**

Objective 1 – To select a suitable dataset to be used with the selected algorithm and pre-process it by <date>

Objective 2 – To conduct Background Study in order to identify the problem, draft the research questions, write motivation and rationale for the project. Finalise the aim and objectives by <date>.

Objective 3 -To undertake literature review into ML algorithms to select the most suitable ones for this project, understand the parameters used for detection, and to obtain results from state-of-the-art research in this field, by <date>

Objective 4 - To perform Research in order to understand the methodology to be followed which will include survey of the collected data, visualisation of data using Exploratory Data Analysis by <date>

Objective 5 - To implement the Machine Learning models the steps followed will be Feature Engineering on the chosen dataset, Evaluation and Fine Tuning of the parameters chosen via feature engineering, Model Evaluation based on several factors by <date>

**7.0Project outcomes and lessons to be learned**

1. Gain in-depth knowledge and practical experience in applying machine learning algorithms and techniques to the domain of crypto currency price prediction.
2. Develop proficiency in data pre-processing, feature engineering, model training, and evaluation in the context of financial time series data.
3. Acquire expertise in selecting and implementing appropriate machine learning algorithms, considering their strengths, weaknesses, and performance metrics.
4. Enhance analytical and problem-solving skills by addressing challenges related to crypto currency market dynamics, data quality, and model interpretation.
5. Improve programming skills through hands-on implementation of machine learning models, leveraging popular libraries and frameworks.
6. Validate the effectiveness of machine learning models in predicting crypto currency prices, contributing to the broader field of financial forecasting and trading strategies.
7. Gain insights into the complex relationships between various market indicators, news sentiment, and crypto currency price movements.
8. Develop the ability to critically analyze and compare research papers, identifying gaps and areas for further improvement in the field.
9. Demonstrate the capability to independently plan, execute, and evaluate a complex machine learning project, showcasing project management and research skills.
10. Establish a solid foundation for further exploration of advanced machine learning techniques, deep learning, and ensemble methods for financial forecasting.
11. Apply the acquired knowledge and skills in real-world scenarios, such as investment strategies, risk management, or portfolio optimization.

Overall, the project aims to provide a comprehensive learning experience, combining technical skills, domain knowledge, and research capabilities, which will contribute to personal growth and professional development in the field of Data science, Machine learning and finance.