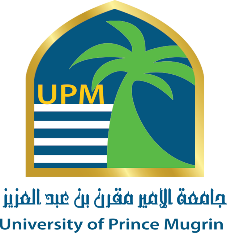
****

**Project Report**

**Bitcoin Price Predictor**

**Data Mining – AI306**

**Team Members**

**Abubakar Waziri – 4220056 – 4220056@upm.edu.sa**

**Mohammed Sattar – 4310129 - 4310129@upm.edu.sa**

**Youssef ElNahas – 4311779 - 4311779@upm.edu.sa**

**Supervisor: Dr. Mohammed Temraz**

# Abstract

* Brief overview of the study’s objective: predicting Bitcoin prices
* Summary of methodology: data sources, models used
* Key findings and implications

# Introduction

Bitcoin, the world’s leading cryptocurrency since around 2013, is known for its high volatility and unpredictable price movements. Therefore, accurate price prediction is crucial for traders and investors aiming to navigate this dynamic market and actually be successful in it. Traditional statistical methods often fall short in capturing Bitcoin’s complex behaviour, while machine learning and deep learning models offer new potential for better and more accurate forecasts. This study compares several predictive approaches using historical Bitcoin data to identify effective strategies for price prediction and support informed decision-making in the cryptocurrency market.

# Literature Review (Related Work)

Recent studies on Bitcoin price prediction have explored a variety of machine learning and deep learning approaches to enhance forecasting accuracy. Mohammad jafari (2024) conducted a comparative analysis of two recurrent neural network architectures—Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU)—using historical Bitcoin data from 2015 to 2023. The study found that GRUs outperformed LSTMs, achieving a lower mean squared error (MSE) of 4.67 compared to 6.25, while also training approximately 30% faster. This highlights GRU's effectiveness in handling long-term dependencies and its computational efficiency in financial time series forecasting.

Complementing this, Swetha (2022) evaluated traditional machine learning and deep learning models—including Linear Regression, Facebook Prophet, and LSTM—for predicting the prices of Bitcoin, Ethereum, and Litecoin. While Linear Regression demonstrated high R² values, LSTM models exhibited lower Root Mean Squared Error (RMSE), indicating better generalization and predictive performance. The study emphasized that LSTM is better suited for capturing the non-linear and volatile nature of cryptocurrency markets, though it also recommended expanding the input feature set beyond just closing prices.

Building on the limitations of traditional deep learning models, Khaniki and Manthouri (2024) proposed a hybrid Transformer-based framework combining technical indicators, the Performer neural network, and Bidirectional LSTM (BiLSTM). The Performer, employing the FAVOR+ attention mechanism, provided a scalable and efficient alternative to standard Multi-head Attention Transformers. When coupled with BiLSTM, the model effectively captured both long-range dependencies and bidirectional temporal dynamics. Across multiple benchmarks, including LSTM, GRU, and standard Transformers, their model consistently achieved the highest accuracy, with the lowest RMSE and highest R² scores on both daily and hourly Bitcoin price data.

Collectively, these studies highlight the evolving landscape of cryptocurrency forecasting. From conventional time series models to cutting-edge attention-based neural networks, the integration of temporal modeling, attention mechanisms, and technical indicators plays a pivotal role in enhancing prediction accuracy for Bitcoin and other digital assets.

# Experiments

## a- Data Description

The dataset used in this study consists of historical Bitcoin price and trading data spanning from January 1, 2012, to January 1, 2025, totalling around 7 million entries. The data is organized in a time series format with a one-minute frequency, providing a comprehensive view of Bitcoin’s market activity over more than a decade.

Each record in the dataset includes the following features:

* **Timestamp**: Unix timestamp representing the precise minute of the record.
* **Open**: The price of Bitcoin at the start of the minute.
* **High**: The highest price reached within the minute.
* **Low**: The lowest price reached within the minute.
* **Close**: The price of Bitcoin at the end of the minute.
* **Volume**: The amount of Bitcoin traded during the minute.
* **Datetime**: Date and time corresponding to the timestamp (in YYYY-MM-DD Time:TImezone format)

A sample from the dataset is shown below:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Timestamp | Open | High | Low | Close | Volume | Datetime |
| 1360887300.0 | 26.63 | 26.63 | 26.63 | 26.63 | 17.98955015 | 2013-02-15 00:15:00+00:00 |
| 1360887360.0 | 26.63 | 26.63 | 26.63 | 26.63 | 0.0 | 2013-02-15 00:16:00+00:00 |
| 1360887420.0 | 26.60 | 26.60 | 26.60 | 26.60 | 20.0 | 2013-02-15 00:17:00+00:00 |

This granular dataset enables detailed analysis and modelling of Bitcoin’s price dynamics, capturing both short-term fluctuations and long-term trends. The inclusion of open, high, low, close, and volume data supports the extraction of technical indicators and the development of robust predictive models.

## b- Data Preprocessing

* Handling missing values and outliers
* Feature engineering (e.g., lag features, moving averages)
* Data normalization or scaling
* Splitting data into training and test sets

## c- Models Built

* Description of baseline models (e.g., Linear Regression, ARIMA)
* Advanced models (e.g., Random Forest, LSTM, GRU)
* Hyperparameter tuning strategies

## d- Results

* Evaluation metrics (e.g., RMSE, MAE, R²)
* Comparative performance of models
* Visualization of predictions vs. actual prices

# Interpretation & Discussion

* Analysis of model performance
* Discussion of factors influencing prediction accuracy
* Limitations of the study
* Implications for traders and researchers

# Conclusion

* Recap of main findings
* Contributions to the field
* Suggestions for future work

# References

* List of all academic papers, datasets, and tools cited in the report