# Task 2: Research Project Plan:

# Cryptocurrency Price Prediction Using AI: A Comparative Analysis for Informed Decision-Making

**Student Name: Khizar Sabir**

**Student number: 33076684**

## 1. INTRODUCTION AND JUSTIFICATION

The use of cryptocurrency has emerged as a highly active financial instrument during the recent years (Rose, 2015). The cryptocurrency market includes Bitcoin (BTC), Ethereum (ETH) and Binance Coin (BNB) which display volatile price changes at a global level. Price predictions in the crypto market benefit both investors and analysts and decision-makers accountable for market operations. The previous research on AI-based models for crypto price prediction includes works by (Shamshad et al., 2023), (Hamayel & Owda, 2021), (Wu et al., 2018), (Fallah et al., 2024), and (Qureshi et al., 2024) as cryptocurrency markets demonstrate extreme volatility through their non-linear unpredictable price dynamics (Kakinaka & Umeno, 2021). Utilizing forecasting models becomes hard when analyzing price characteristics in markets known for substantial volatility which reduces their general predictive power. Inadequate forecasting methods will cause unstable stock markets which leads to unpredictable investor losses. This study fills the existing knowledge gap through an evaluation of ARIMA and LSTM and Prophet implemented on BTC and ETH and BNB price predictions. The primary objective of this research is developing more accurate and responsive forecasting methods for cryptocurrency prices.

## 2. RESEARCH QUESTION, AIMS & OBJECTIVES

This aim of this project is to examines various forecasting approaches. The purpose is to enhance both the accuracy rates and adaptability levels when predicting cryptocurrency price movements. The analysis enables investors analysts along with other financial systems for decision making on improved information. The project evaluates Bitcoin (BTC), Ethereum (ETH) and Binance Coin (BNB) through ARIMA, LSTM and Prophet analytical tools. The research will examine the role of real-time market sentiment and external factors on prediction quality.

### 2.1 Research Question

How effective are hybrid AI-statistical approaches in improving the accuracy of cryptocurrency price predictions for BTC, ETH, and BNB?

## 2.2 Objectives

1. A performance assessment between ARIMA, LSTM, and Prophet models for cryptocurrency price forecasting will be conducted.
2. The research will examine the performance of the ARIMA, LSTM, and Prophet models on Bitcoin (BTC), Ethereum (ETH), and Binance Coin (BNB). Because these cryptocurrencies are the most popular within digital asset markets.
3. An evaluation of prediction accuracy will use Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE).
4. A user-friendly web application will enable analysts to input a cryptocurrency through its interface alongside the display of price predictions from all three models.
5. To assess which model is more suitable under different price behavior conditions, including volatility and market shifts.

## 2.3 Deliverable

A web application will be developed to predict cryptocurrency prices using ARIMA, LSTM, and Prophet models. The system will allow users to view comparative forecasts for Bitcoin (BTC), Ethereum (ETH), and Binance Coin (BNB), incorporating both historical and real-time external data such as market sentiment. The outcome of this application can assist investors, traders, and analysts in making timely and informed decisions in fast-changing crypto markets.

# 3. LITERATURE REVIEW

This section critically examines existing literature on cryptocurrency price prediction to contextualize the proposed research. It reviews the evolution of prediction techniques, analyzes key research findings, highlights current debates, identifies existing gaps, and articulates the contribution of the current project.

### 3.1 **Evolution of Cryptocurrency Price Prediction Techniques**

Cryptocurrency markets are highly volatile. The rules around them are still developing, which makes price prediction difficult. In this section we will discuss the evolution of crypto prediction techniques from the very early days till now. In the early days, researchers used basic statistical models. Research by Vapnik laid the groundwork for some of these ideas (Vapnik, 1998). ARIMA (Box et al., 2015) and GARCH (Bollerslev, 1986) were also used to study market trends. However, these models could not handle the unpredictable and non-linear behavior of crypto prices (Qureshi et al., 2024).

As computing power and data availability improved, machine learning models became more common. Support Vector Machines (Vapnik, 1998) and Random Forests (Breiman, 2001) were better at spotting complex patterns. Later, deep learning methods like Long Short-Term Memory (LSTM) networks, introduced by (Hochreiter & Schmidhuber, 1997), helped predict prices by learning from past data over time.

More recently, researchers started using hybrid models. These combine statistical tools, machine learning, deep learning, and sentiment analysis. They use online data to improve prediction accuracy. New work is also exploring Transformer models and explainable AI (Fior et al., 2022). These helps make crypto forecasts more accurate and transparent. As the market continues to change, prediction methods keep evolving to match its volatility.

### 3.2 **Critical Review of Key Studies**

Several research papers have explored various techniques for cryptocurrency price prediction. These range from traditional statistical models to advanced AI-based approaches.

- Shamshad et al. (2023) compared six models for short-term price prediction of ADA, BNB, and ETH. ARIMA performed best, but the study was limited by its short prediction window and lack of external factors.
- Qureshi et al. (2024) evaluated hybrid models combining machine learning and time series forecasting for Bitcoin prices. Their hybrid approach outperformed individual models but used simple averaging and focused on a limited timeframe.
- Hamayel and Owda (2021) tested recurrent neural networks including GRU, LSTM, and bidirectional LSTM for predicting BTC, LTC, and ETH prices. GRU achieved the lowest error, though the study relied on a single performance metric and did not explore feature engineering.
- Wu et al. (2018) proposed an LSTM model combined with an autoregressive component for Bitcoin price forecasting. Their model improved accuracy but was limited to one cryptocurrency and a short period.
- Fallah et al. (2024) introduced a model combining LSTM with Vector Autoregression to predict errors, improving accuracy for Bitcoin, Ethereum, and Binance Coin. However, the study lacks broader market analysis.

Table 1: Critical comparison of Cryptocurrency Prediction Models

| Study | Focus | Methods | Findings/Strengths | Limitations |
|---|---|---|---|---|
| (Shamshad et al., 2023) | Short-term price prediction (ADA, BNB, ETH) | SVR, ARIMA, Prophet, LSTMs | ARIMA best; Model comparison | Limited coins; Short-term; No external factors |
| (Hamayel & Owda, 2021) | Crypto price prediction (BTC, LTC, ETH) | GRU, LSTM, bi-LSTM | GRU best | Single metric; Limited features |
| (Wu et al., 2018) | Bitcoin price forecasting | LSTM+AR(2), LSTM | LSTM+AR(2) improves input | Single coin; Short-term; No external factors |
| (Fallah et al., 2024) | Crypto price prediction (BTC, ETH, BNB) | LSTM+VAR | Error modeling improves accuracy | Limited market analysis |
| (Qureshi et al., 2024) | Bitcoin price forecasting | ML, TS, Hybrid (avg.) | Hybrid best | Simple avg.; Model correlation |

## 3.3 Themes, Debates, and Gaps in Current Research

Most studies agree that LSTM performs well in crypto prediction. However, many only test one model or one coin. They rarely compare models across different conditions. Many ignore external factors like news or social sentiment. This makes the models limited in real-life use. Some studies also skip feature selection. Others don't test over long time periods. This research will fill these gaps by comparing models, using three major coins, and including external sentiment data.

Current cryptocurrency price prediction research highlights progress and difficulties. The successful use of deep learning models like LSTM is evident in the identification of patterns between trading times (Hochreiter & Schmidhuber, 1997) (Wu et al., 2018). However, there are recurring themes and debates in the research, and limitations exist.

Many studies concentrate on individual cryptocurrencies and short-term predictions and neglect external factors like sentiment or on-chain metrics. There is a need to improve how different data sources are used and how the models are interpreted.

## 3.4 Justification for the Proposed Research

Bitcoin (BTC), Binance Coin (BNB), and Ethereum (ETH) are among the most widely traded cryptocurrencies. Their high usage makes them important to study, especially due to their unpredictable price patterns. Many earlier studies focus on just one coin or model. They often miss the value of comparing different forecasting methods on multiple major coins. This research aims to fill that gap. It uses three well-known models and applies them side by side on BTC, ETH, and BNB. It also includes news-based sentiment data to reflect market moods, which can impact prices. By combining model comparison with external sentiment, the research offers a practical and realistic forecasting approach. This can support analysts, traders, and financial decision-makers in choosing the right method for specific market conditions.

## 3.5 Contribution to the Ongoing Debate in Cryptocurrency Prediction

There is still debate around which model works best for volatile crypto markets. This research contributes by comparing ARIMA, LSTM, and Prophet models using multiple cryptocurrencies. It also adds value by integrating sentiment data and focusing on model interpretability. The goal is to support more stable and realistic crypto price forecasts. These findings will offer useful insights for both researchers and financial analysts.

# 4. Research Design

This section presents the philosophical stance, methodology, tools, and development lifecycle used to design and execute this research project on cryptocurrency price forecasting.

### 4.1 Research Philosophy and Approach

The project adopts a **positivist philosophy** and a **deductive approach**, aiming to test forecasting models using empirical, quantifiable data. Guided by existing theories in time series analysis and deep learning, the study explores how model predictions align with real-world market trends (Qureshi et al., 2024).

### 4.2 Methodology

The project follows the **Agile SDLC model**, enabling flexibility during research sprints.

- **Sprint 1: Research Planning & Design**
  Literature review, project scope definition, research design, and dataset selection.
- **Sprint 2: Data Acquisition & Preprocessing**
  Dataset cleaning, transformation, and formatting for model training.
- **Sprint 3: Model Development, Evaluation & Optimization**
  Model implementation (ARIMA, LSTM, Prophet), performance evaluation, and tuning.
- **Sprint 4: Web App Development & Final Submission**
  Build Gradio-based interface, integrate models, testing, and final documentation.

Agile is selected due to its iterative nature, allowing changes based on model insights or dataset adjustments. This ensures adaptability to evolving technical or experimental needs.

### 4.3 Data Acquisition

This research makes use of two publicly available and trusted datasets from **Kaggle**, covering both **price history** and **market sentiment** for major cryptocurrencies. These datasets were selected based on their completeness, real-world relevance, and usability for machine learning models.

***1. Cryptocurrency Price History Dataset***

This dataset contains **daily historical price data** for Bitcoin (BTC), Ethereum (ETH), and Binance (Sudalai Rajkumar, 2021) . It includes columns such as:

- Date: Timestamp of each record
- Open, Close: Opening and closing prices of the day
- High, Low: Highest and lowest prices of the day
- Volume: Daily trading volume
- Market cap: Market capitalization
- Name, Symbol: Asset identifier

| | SNo | Name | Symbol | Date | High | Low | Open | Close | Volume | Marketcap |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | Bitcoin | BTC | 2013-04-29 23:59:59 | 147.488007 | 134.000000 | 134.444000 | 144.539993 | 0.000000e+00 | 1.603769e+09 |
| **1** | 2 | Bitcoin | BTC | 2013-04-30 23:59:59 | 146.929993 | 134.050003 | 144.000000 | 139.000000 | 0.000000e+00 | 1.542813e+09 |
| **2** | 3 | Bitcoin | BTC | 2013-05-01 23:59:59 | 139.889999 | 107.720001 | 139.000000 | 116.989998 | 0.000000e+00 | 1.298955e+09 |
| **3** | 4 | Bitcoin | BTC | 2013-05-02 23:59:59 | 125.599998 | 92.281898 | 116.379997 | 105.209999 | 0.000000e+00 | 1.168517e+09 |
| **4** | 5 | Bitcoin | BTC | 2013-05-03 23:59:59 | 108.127998 | 79.099998 | 106.250000 | 97.750000 | 0.000000e+00 | 1.085995e+09 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **2986** | 2987 | Bitcoin | BTC | 2021-07-02 23:59:59 | 33939.588699 | 32770.680780 | 33549.600177 | 33897.048590 | 3.872897e+10 | 6.354508e+11 |
| **2987** | 2988 | Bitcoin | BTC | 2021-07-03 23:59:59 | 34909.259899 | 33402.696536 | 33854.421362 | 34668.548402 | 2.438396e+10 | 6.499397e+11 |
| **2988** | 2989 | Bitcoin | BTC | 2021-07-04 23:59:59 | 35937.567147 | 34396.477458 | 34665.564866 | 35287.779766 | 2.492431e+10 | 6.615748e+11 |
| **2989** | 2990 | Bitcoin | BTC | 2021-07-05 23:59:59 | 35284.344430 | 33213.661034 | 35284.344430 | 33746.002456 | 2.672155e+10 | 6.326962e+11 |
| **2990** | 2991 | Bitcoin | BTC | 2021-07-06 23:59:59 | 35038.536363 | 33599.916169 | 33723.509655 | 34235.193451 | 2.650126e+10 | 6.418992e+11 |

2991 rows × 10 columns

Figure 1: Sample of Bitcoin Daily Price Data

| | SNo | Name | Symbol | Date | High | Low | Open | Close | Volume | Marketcap |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | Ethereum | ETH | 2015-08-08 23:59:59 | 2.798810 | 0.714725 | 2.793760 | 0.753325 | 6.741880e+05 | 4.548689e+07 |
| **1** | 2 | Ethereum | ETH | 2015-08-09 23:59:59 | 0.879810 | 0.629191 | 0.706136 | 0.701897 | 5.321700e+05 | 4.239957e+07 |
| **2** | 3 | Ethereum | ETH | 2015-08-10 23:59:59 | 0.729854 | 0.636546 | 0.713989 | 0.708448 | 4.052830e+05 | 4.281836e+07 |
| **3** | 4 | Ethereum | ETH | 2015-08-11 23:59:59 | 1.131410 | 0.663235 | 0.708087 | 1.067860 | 1.463100e+06 | 6.456929e+07 |
| **4** | 5 | Ethereum | ETH | 2015-08-12 23:59:59 | 1.289940 | 0.883608 | 1.058750 | 1.217440 | 2.150620e+06 | 7.364501e+07 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **2155** | 2156 | Ethereum | ETH | 2021-07-02 23:59:59 | 2155.596496 | 2021.824808 | 2109.892677 | 2150.040364 | 3.179621e+10 | 2.505527e+11 |
| **2156** | 2157 | Ethereum | ETH | 2021-07-03 23:59:59 | 2237.567155 | 2117.590013 | 2150.835025 | 2226.114282 | 1.743336e+10 | 2.594475e+11 |
| **2157** | 2158 | Ethereum | ETH | 2021-07-04 23:59:59 | 2384.286857 | 2190.837703 | 2226.550382 | 2321.724112 | 1.878711e+10 | 2.706217e+11 |
| **2158** | 2159 | Ethereum | ETH | 2021-07-05 23:59:59 | 2321.922836 | 2163.041394 | 2321.922836 | 2198.582464 | 2.010379e+10 | 2.562978e+11 |
| **2159** | 2160 | Ethereum | ETH | 2021-07-06 23:59:59 | 2346.294874 | 2197.919385 | 2197.919385 | 2324.679449 | 2.089186e+10 | 2.710286e+11 |

2160 rows × 10 columns

Figure 2: Sample of Ethereum Daily Price Data

### *2. Crypto News Sentiment Dataset*

This dataset captures **market sentiment data** derived from news headlines (Olivier Van Hauwaert, 2021). It helps integrate external factors such as public opinion and market mood into the forecasting models. Key features include:

- Date: Timestamp of each news item
- flair_sentiment: Sentiment score assigned by the Flair model

- tb_polarity, tb_subjectivity: Polarity and subjectivity from TextBlob
- vader_pos, vader_neg, vader_neu, vader_compound: Sentiment metrics from the VADER sentiment analyzer

| | date | flair_sentiment | tb_polarity | tb_subjectivity | vader_pos | vader_neg | vader_neu | vader_compound |
|---|---|---|---|---|---|---|---|---|
| **0** | 2021-11-05 04:42:00 | -0.9959 | 0.050000 | 0.250000 | 0.187 | 0.081 | 0.733 | 0.4404 |
| **1** | 2021-11-05 08:15:00 | -0.8736 | 0.000000 | 1.000000 | 0.000 | 0.213 | 0.787 | -0.4019 |
| **2** | 2021-11-05 10:24:00 | -0.5914 | 0.316667 | 0.483333 | 0.144 | 0.000 | 0.856 | 0.4019 |
| **3** | 2021-11-05 16:58:00 | -0.9848 | 0.250000 | 0.750000 | 0.137 | 0.000 | 0.863 | 0.3612 |
| **4** | 2021-11-05 21:00:00 | 0.5093 | 0.068182 | 0.227273 | 0.110 | 0.000 | 0.890 | 0.2732 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... |
| **9894** | 2023-12-18 18:33:34 | -0.9962 | -0.266667 | 0.333333 | 0.052 | 0.193 | 0.754 | -0.6486 |
| **9895** | 2023-12-19 02:59:59 | -0.9852 | -0.080000 | 0.640000 | 0.124 | 0.000 | 0.876 | 0.4767 |
| **9896** | 2023-12-19 04:10:00 | -0.9995 | 0.000000 | 0.000000 | 0.116 | 0.000 | 0.884 | 0.3612 |
| **9897** | 2023-12-19 04:50:11 | -0.9979 | 0.300000 | 0.100000 | 0.303 | 0.000 | 0.697 | 0.7506 |
| **9898** | 2023-12-19 05:25:00 | -0.7180 | -0.011111 | 0.383333 | 0.119 | 0.000 | 0.881 | 0.4019 |

9899 rows × 8 columns

Figure 3: Sample Record from Crypto Sentiment Dataset

These datasets together allow for a comprehensive modeling framework. That combines **technical price indicators** with **external sentiment signals.** Enabling the development of more robust and accurate crypto price forecasting models.

### *4.4 Data Preparation*

- **Missing values** handled using linear interpolation
- **Outlier treatment** using IQR method with winsorization
- **Normalization** applied using Min-Max scaling for LSTM compatibility

### *4.5 Model Development*

To evaluate and compare forecasting performance, this study implements three distinct models: **ARIMA**, **LSTM**, and **Prophet**. These models were selected due to their unique strengths and complementary forecasting capabilities:

- ARIMA (AutoRegressive Integrated Moving Average) is a basic model used for time series forecasting. It works best when data follows a steady pattern and doesn't change too much over time. In this project, I will use ARIMA as a starting point to compare it with other models. It helps to set a baseline for prediction accuracy. To make ARIMA work well, I will tune its settings (p, d, q) using the ADF test and ACF/PACF plots.
- LSTM (Long Short-Term Memory) is a type of deep learning model. It's useful for learning from time-based data. Since crypto prices change fast and often don't follow a regular pattern, LSTM can help catch these changes. I will train this model using the Adam optimizer. The accuracy will be checked using Mean Squared Error (MSE). This model may take more time to train, but it could offer better results for crypto price prediction.

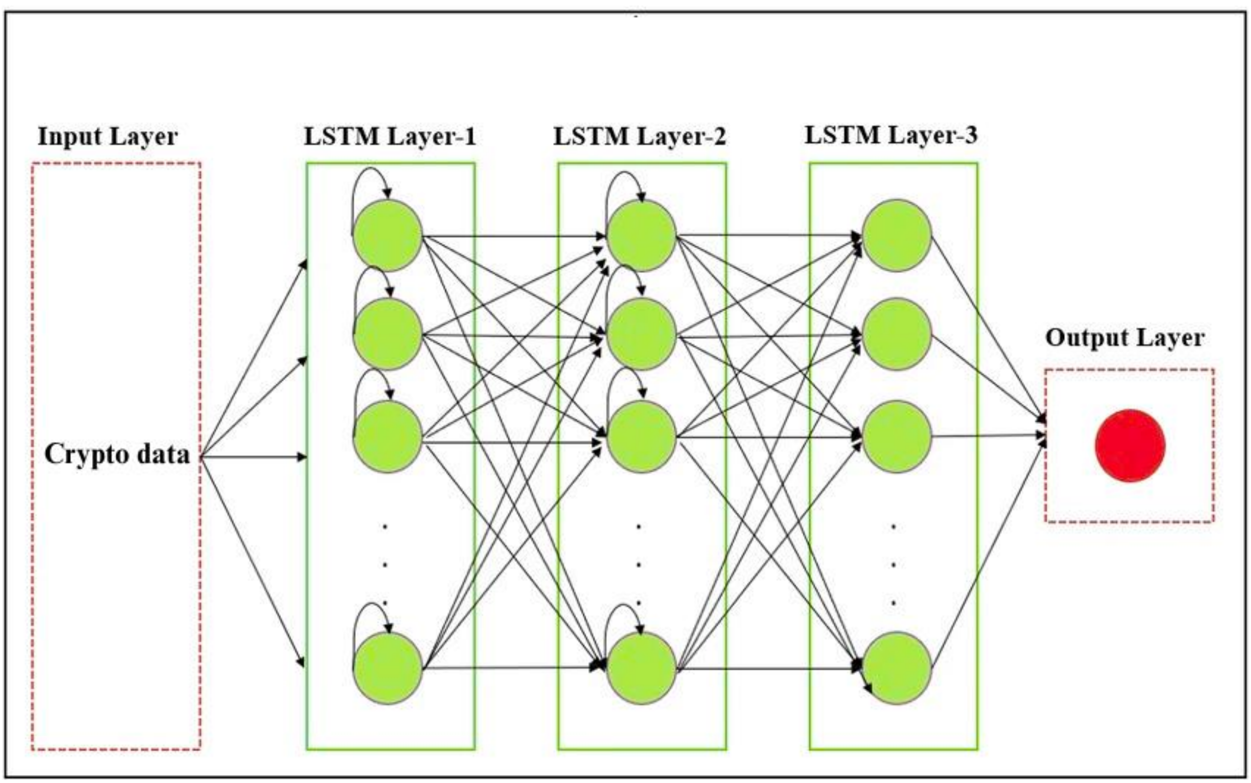



Figure 4: Architecture of the LSTM Network used in this study.

- **Prophet**, developed by Facebook, is a decomposable time series model that handles **seasonality, trend changes, and holiday effects** effectively. It is robust to missing data and outliers, making it a practical choice for financial time series with irregular behavior.

By combining these three models each leveraging different assumptions and strengths this project aims to identify which method or combination provides the most accurate and reliable predictions for Bitcoin (BTC), Ethereum (ETH), and Binance Coin (BNB) under different market conditions.

### *4.6 Model Evaluation*

Performance will be measured using:

- Mean Absolute Error (MAE)
- Root Mean Square Error (RMSE)
- Mean Absolute Percentage Error (MAPE)
- $R^2$ (coefficient of determination)

## 4.3 Tools and Technologies

To ensure effective model development, data processing, sentiment analysis, and deployment, this project integrates a range of industry-standard tools and libraries:

- **Programming Language**:
  Python – chosen for its versatility, robust machine learning ecosystem, and rich libraries for time series and NLP tasks.
- **Data Handling and Preprocessing**:
  - pandas and numpy – used for data manipulation, cleaning, and efficient numerical computations.
  - os and datetime – employed for managing file paths and handling time-based data.
- **Sentiment Analysis Tools**:
  - Flair – used to extract advanced sentiment scores from crypto news headlines.
  - TextBlob – for calculating polarity and subjectivity.
  - VADER from nltk.sentiment – for rule-based sentiment scoring on financial text.
- **Modeling Libraries**:
  - **TensorFlow/Keras** – for building and training LSTM-based deep learning models.
  - **Prophet** – for decomposable trend and seasonality-based forecasting.

    - **Statsmodels** – to implement and optimize ARIMA models for baseline comparison.
- **Evaluation and ML Utilities**:
    - scikit-learn – for preprocessing utilities, metrics calculation (e.g., MAE, RMSE, $R^2$), and scaling.
- **Development and Execution Environment**:
    - **Jupyter Notebooks** – for modular experimentation and step-by-step visualization.
    - **Google Colab** – provides GPU-based training capabilities for LSTM, enhancing performance and accessibility without local hardware constraints.
- **Deployment Tool**:
    - **Gradio** – utilized to build a lightweight and interactive web application that demonstrates real-time cryptocurrency price predictions using the trained models.

## 4.5 Web Application Requirement Analysis

This project will culminate in a **simple forecasting web app** where users can select a model and cryptocurrency, and view visual price predictions.

***Functional Requirements***

The web application will allow users to forecast cryptocurrency prices using different machine learning models. It is designed to be user-friendly, with intuitive navigation and interactive features. The main functional components are:

**User Interface (UI):**
A clean and minimal interface that enables users to:

- Select a cryptocurrency (BTC, ETH, or BNB).
- Choose a forecasting model (ARIMA, LSTM, or Prophet) via dropdown menus.
- View prediction results displayed as interactive line graphs.

**Prediction Output:**

Once the user selects options and clicks “Predict,” the system will:

- Run the selected model on the dataset.
- Display predicted prices.

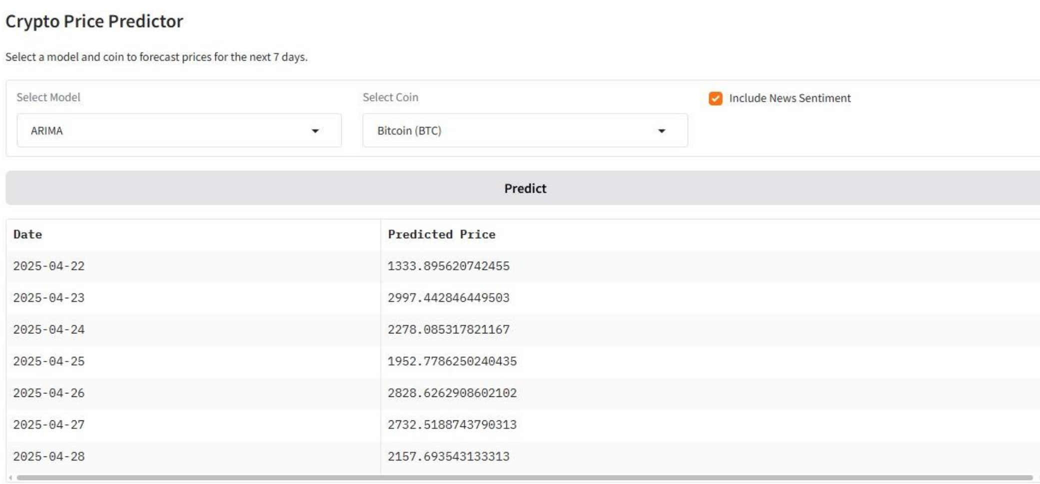



Figure 5: UI of Cryptocurrency Price Forecasting Web App

### *Non-Functional Requirements*

- **Performance**: Lightweight UI, fast load times
- **Scalability**: Can add more coins/models in future
- **Security**: Safe from data injections or unauthorized access
- **Usability**: Responsive and easy-to-use interface

### 4.6 Data Science Workflow

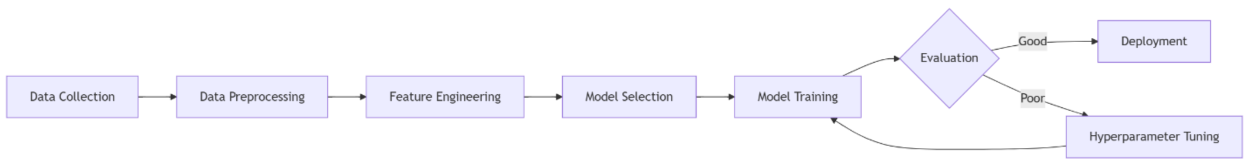



Figure 6: End-to-End Data Science Workflow Diagram

The workflow begins with data collection and ends with either deployment or hyperparameter tuning based on evaluation results. This cycle ensures model improvement in case performance is below expectation.

# 5. ETHICS, RISKS AND ISSUES

### *5.1 Ethics and Legal Considerations*

This project uses only public datasets. No personal or sensitive data is involved. That reduces the legal and ethical risks. Still, ethical care is taken. Sentiment data is from published news, not individuals. The system will not give financial advice. It is for educational use only. All sources are credited. A research ethics form is filled and attached in Appendix A.

This project does not involve human participants or private data, minimizing ethical concerns.

All datasets are sourced from publicly available Kaggle repositories and used solely for academic purposes.

Sources are properly cited, and no personal or sensitive data is collected or processed.

No commercial or real-time trading use is associated with the study.

A completed **Ethics Checklist is included in Appendix A** to confirm full compliance with university ethical standards.

***6.2 Risks and Mitigation Measures***

Table 2: Risks and mitigation measures of this project.

| Risk | Description | Mitigation |
|---|---|---|
| **Data Volume** | Dataset may require heavy preprocessing. | Use staged pipelines and EDA checkpoints. |
| **Model Load** | LSTM requires high computational resources. | Use Google Colab's free GPU environment. |
| **Accuracy Variability** | Forecasting performance may vary across assets. | Apply hyperparameter tuning and model benchmarking. |
| **Evaluation Complexity** | Multiple models and metrics increase complexity. | Use a standardized evaluation framework. |
| **Time Constraints** | Multi-model training may delay progress. | Follow Agile methodology and detailed Gantt chart. |

A detailed **Risk Log Table is included in Appendix B** for transparency and planning.

# 6. TIME PLAN

This research project will run from **1st February 2025 to 1st August 2025**, divided into structured objectives and tasks following the Agile methodology. Each sprint includes focused activities to ensure the timely completion of the project. Key milestones are highlighted for tracking progress.

***Objective 1: Literature Review and Research Design Finalization***

- **Task 1.1:** Conduct detailed literature review → *Feb 1 – Feb 14, 2025*
- **Task 1.2:** Define research problem, gap, and finalize methodology → *Feb 15 – Feb 22, 2025*
  *Milestone 1: Research design finalized – Feb 22, 2025*

***Objective 2: Data Acquisition and Preprocessing***

- **Task 2.1:** Acquire datasets from Kaggle (price + sentiment) → *Feb 23 – Feb 26, 2025*
- **Task 2.2:** Clean and preprocess data (handle nulls, scale, format) → *Feb 27 – Mar 7, 2025*
  *Milestone 2: Cleaned datasets ready – Mar 7, 2025*

***Objective 3: Model Development & Evaluation***

- Implement ARIMA, LSTM, and Prophet models → *Mar 8 – Mar 25, 2025*
- Hyperparameter tuning & model optimization → *Mar 26 – Apr 5, 2025*
- Evaluate models using MAE, RMSE, $R^2$, MAPE → *Apr 6 – Apr 14, 2025*
- Analyze comparative performance across BTC, ETH, BNB → *Apr 15 – Apr 20, 2025*
  *Milestone 3: Evaluation and insights completed – Apr 20, 2025*

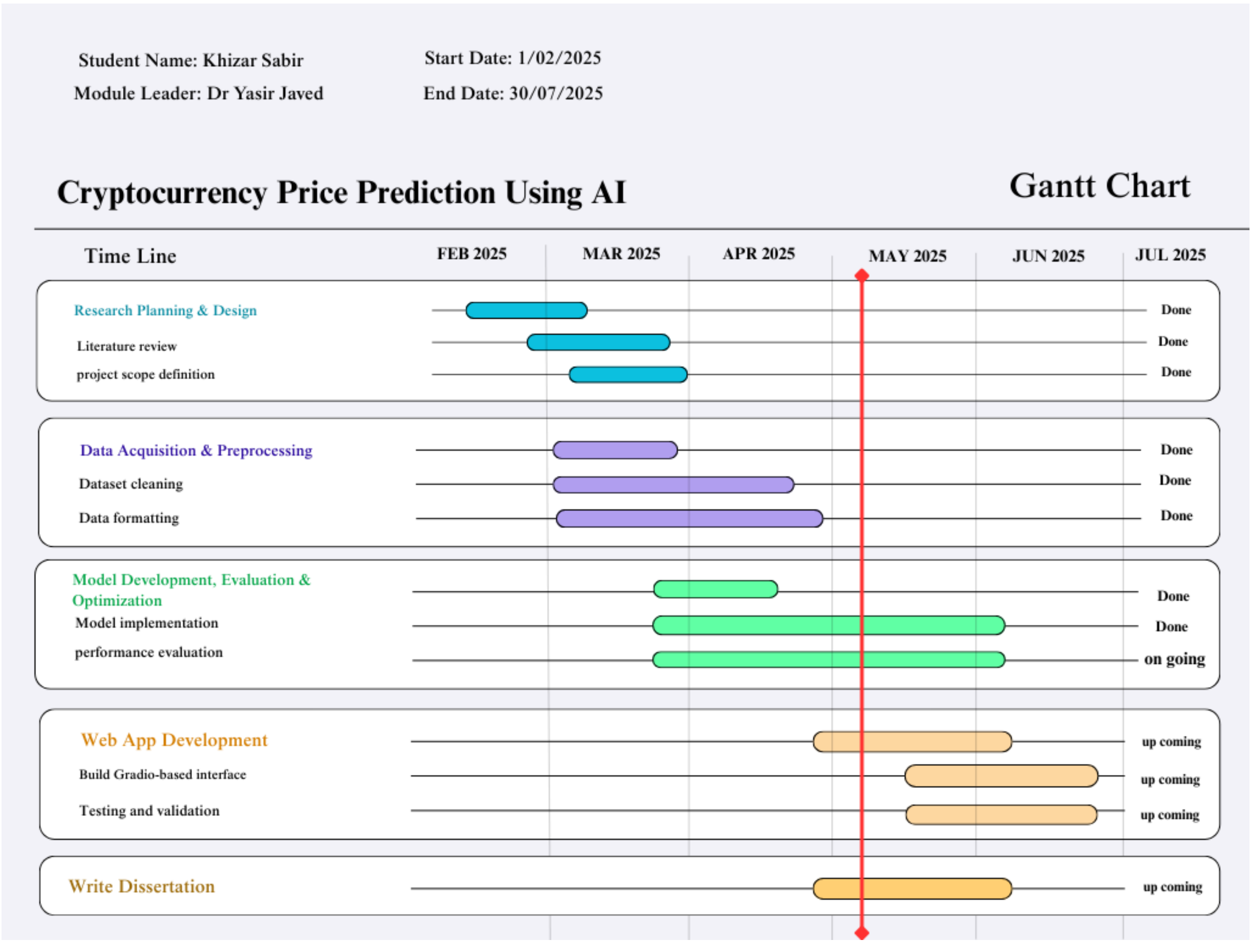
***Objective 4: Web App Development and Deployment***

Design and build UI with Gradio → *Apr 21 – Apr 27, 2025*
Integrate models and deploy app → *Apr 28 – May 6, 2025*
*Milestone 4: Web app deployed – May 6, 2025*

***Objective 5: Dissertation***

Write Dissertation → *May 6 – Jully 30, 2025*

## Gantt Chart:



Student Name: Khizar Sabir
Module Leader: Dr Yasir Javed
Start Date: 1/02/2025
End Date: 30/07/2025
Cryptocurrency Price Prediction Using AI
Gantt Chart
Time Line
FEB 2025
MAR 2025
APR 2025
MAY 2025
JUN 2025
JUL 2025
Research Planning & Design
Done
Literature review
Done
project scope definition
Done
Data Acquisition & Preprocessing
Done
Dataset cleaning
Done
Data formatting
Done
Model Development, Evaluation & Optimization
Done
Model implementation
Done
performance evaluation
on going
Web App Development
up coming
Build Gradio-based interface
up coming
Testing and validation
up coming
Write Dissertation
up coming

## References:


1. Fallah, M. F., Pourmansouri, R., & Ahmadpour, B. (2024). Presenting a New Deep Learning-Based Method with the Incorporation of Error Effects to Predict Certain Cryptocurrencies. *International Review of Financial Analysis*, *95*, 103466. https://doi.org/10.1016/j.irfa.2024.103466
2. Hamayel, M. J., & Owda, A. Y. (2021). A Novel Cryptocurrency Price Prediction Model Using GRU, LSTM and bi-LSTM Machine Learning Algorithms. *AI*, *2*(4), 477–496. https://doi.org/10.3390/ai2040030
3. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.
4. Olivier Van Hauwaert. (2021). Crypto News Sentiment Dataset. In *Kaggle*. https://www.kaggle.com/datasets/oliviervha/crypto-news
5. Qureshi, M., Iftikhar, H., Rodrigues, P. C., Rehman, M. Z., & Salar, S. A. A. (2024). Statistical modeling to improve time series forecasting using machine learning, time series, and hybrid models: a case study of bitcoin price forecasting. *Mathematics*, *12*(23), 3666.
6. Rose, C. (2015). The Evolution Of Digital Currencies: Bitcoin, A Cryptocurrency Causing A Monetary Revolution. *International Business & Economics Research Journal (IBER)*, *14*(617).
7. Shamshad, H., Ullah, F., Ullah, A., Kebande, V. R., Ullah, S., & Al-Dhaqm, A. (2023). Forecasting and Trading of the Stable Cryptocurrencies With Machine Learning and Deep Learning Algorithms for Market Conditions. *IEEE Access*, *11*, 122205–122220. https://doi.org/10.1109/ACCESS.2023.3327440
8. Sudalai Rajkumar. (2021). Cryptocurrency Historical Price Data. In *kaggle*. kaggle. https://www.kaggle.com/datasets/sudalairajkumar/cryptocurrencypricehistory
9. Vapnik, V. (1998). *Statistical learning theory*. Wiley.
10. Wu, C.-H., Lu, C.-C., Ma, Y.-F., & Lu, R.-S. (2018). A New Forecasting Framework for Bitcoin Price with LSTM. *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, 168–175. https://doi.org/10.1109/ICDMW.2018.00032

## Appendix A – Ethics Checklist

| Ethical Consideration | Response |
|---|---|
| Human subjects involved? | No |
| Private or sensitive data used? | No |
| Use of publicly available datasets? | Yes |
| Data cited properly? | Yes |
| Commercial use? | No |
| Personal data processed? | No |
| Research purpose? | Educational / Academic |
| Ethics form submitted? | Yes |

## Appendix B – Risk Log Table

| Risk ID | Risk Description | Impact | Likelihood | Mitigation Strategy |
|---|---|---|---|---|
| R1 | Data preprocessing delays | Medium | High | Staged pipelines and early EDA |
| R2 | LSTM model slow training | Medium | High | Use Google Colab GPU + batch processing |
| R3 | Inconsistent model accuracy | Medium | Medium | Tune hyperparameters and compare results |
| R4 | Evaluation complexity | Low | Medium | Standardized metric framework |
| R5 | Time management | High | Medium | Follow Agile sprints and weekly targets |