Crypto Geeks: Crypto Education and Research Platform with Predictive Sentiment–Price Modelling and Explainable AI

Mr. P. Periyasamy, M.E., AP [1],

Dhanuja D[2], Dhayanithi N[3], Devadharshini M[4], Gopika S[5] UG Scholars, Computer Science and Engineering at

R P Sarathy Institute of Technology, Poosaripatty, Salem -636305.

Mr. P. Periyasamy, M.E., AP

Assistant Professor, Computer Science and Engineering at

R P Sarathy Institute of Technology, Poosaripatty, Salem-636305.

I. Abstract

Cryptocurrency ecosystems have experienced rapid growth, yet most available platforms emphasize trading, speculation, and price monitoring rather than fundamental understanding of blockchain technologies. Additionally, short-term price movements in cryptocurrency markets are highly influenced by news sentiment and volatility. The system integrates real-time crypto news sentiment with machine learning models to predict short-term market movement direction and volatility risk for the next 15 minutes and 1 hour. A Random Forest model is used as a baseline, while a Long Short-Term Memory (LSTM)network captures temporal dependencies. The platform also incorporates an Explainable AI (XAI) layer to enhance transparency by providing human-readable explanations for predictions. The proposed system aims to improve both conceptual learning of blockchain technologies and analytical understanding of how news sentiment impacts crypto markets. Machine learning models such as Linear Regression and Random Forest, along with a Long Short Term Memory (LSTM) network optimized for high-frequency time-series data, are trained to predict price movements for the next 15-minute and 1-hour intervals. Experimental results demonstrate that sentiment-augmented models consistently outperform traditional price-only baselines. The LSTM model achieves a directional prediction accuracy of 71.8% for 15-minute forecasts and 74.3% for 1-hour forecasts, with corresponding MAE values of 0.0029 and 0.0041, respectively.

II. Keywords: Cryptocurrency, Sentiment Analysis, Machine Learning, LSTM, Explainable AI, Blockchain Education, Sentiment Analysis, Natural Language Processing, LSTM, Machine Learning, Explainable Artificial Intelligence, SHAP, Intraday Trading

III. Introduction

Cryptocurrencies and blockchain technologies have emerged as transformative innovations in modern digital finance. Despite their growing adoption, the majority of crypto-related platforms focus on speculative trading, price charts, and market hype, often

neglecting educational depth and research oriented analysis. This lack of conceptual clarity creates a gap between technology adoption and understanding.

At the same time, cryptocurrency markets are extremely sensitive to news events such as protocol upgrades, regulatory announcements, and security breaches. These factors significantly influence short

term price movements and volatility. Traditional financial forecasting models are often insufficient to capture such sentiment driven behaviour.

To address these challenges, this paper introduces **Crypto Geeks**, a frontend-only

crypto education and research platform integrated with predictive sentiment–price modelling. The platform emphasizes deep learning of blockchain concepts while providing explainable short-term market insights derived from crypto news sentiment. The system focuses on predicting **direction and volatility**, rather than exact price values, making it suitable for educational and research use.

IV.Literature Review

Several studies have explored sentiment analysis for financial prediction. Bollen et al. demonstrated that public mood extracted from social media could predict stock market movements, highlighting the relevance of sentiment in financial markets. In the cryptocurrency domain, Chen et al. showed that news and social media sentiment correlate strongly with Bitcoin price fluctuations.

Machine learning techniques such as Random Forests have been widely used for financial classification problems due to their robustness and interpretability. However, financial time-series data often exhibit temporal dependencies, which traditional models fail to capture effectively. Long Short-Term Memory (LSTM) networks, introduced by Hochreiter and Schmidhuber, have shown superior performance in modelling

sequential data and have been successfully applied to stock and cryptocurrency forecasting.

More recently, Explainable AI (XAI) techniques have gained importance in financial applications. Ribeiro et al. introduced model-agnostic explanation techniques that improve trust and interpretability of machine learning systems. However, existing research often treats sentiment analysis, prediction, and explainability as separate components, leaving a research gap in unified, educationally oriented systems.

Recent research has increasingly explored the relationship between news sentiment and cryptocurrency market behaviour. Unlike traditional financial markets, cryptocurrency prices are highly reactive to technological announcements, regulatory decisions, and security-related news. Studies have shown that sentiment extracted from crypto-specific news sources provides stronger predictive signals compared to general social media sentiment due to higher information credibility.

[1] Mittal and Goel investigated sentiment driven price movement in cryptocurrencies using natural language processing techniques. Their findings demonstrated that negative news sentiment often leads to short-term price declines accompanied by

increased volatility. However, their work focused primarily on binary price movement and did not incorporate explainability mechanisms.

[2] McNally et al. proposed the use of Long Short-Term Memory (LSTM) networks for modelling Bitcoin price movements. Their research highlighted the effectiveness of LSTM models in capturing temporal dependencies in highly volatile financial time-series data. Despite improved prediction accuracy, the model lacked interpretability, making it unsuitable for transparent decision-making or educational platforms.

[3] Kraaijeveld and De Smedt analyzed the impact of news sentiment on cryptocurrency returns and volatility. Their study concluded that sentiment polarity significantly influences short-term price fluctuations, particularly during periods of market uncertainty. However, their approach emphasized statistical correlation rather than predictive modelling and did not integrate real-time deployment considerations.

V.System Model

This section describes the conceptual and functional model of the proposed **Crypto Geeks** platform, illustrating how different system components interact to perform sentiment-driven short-term cryptocurrency market analysis while supporting an educational framework.

A. Overview of the System Model

The proposed system is modelled as a modular pipeline that integrates **news data ingestion**, **sentiment analysis**, **feature extraction**, **machine learning-based prediction**, and **explainable output generation**. The system operates in a frontend-only environment and is designed for educational and research purposes rather than trading or real-time financial execution.

The system model emphasizes **information flow and decision logic** rather than infrastructure complexity, making it suitable for academic analysis and reproducibility.

B. Input Model

The system accepts two primary categories of input data:

1. **Cryptocurrency Market Data**
   - Recent price changes
   - Short-term trend indicators
   - Volatility measures
2. **Crypto News Data**
   - Headlines and article summaries
   - Time-stamped news events
   - Source credibility metadata

These inputs are periodically fetched using free, publicly available APIs.

C. Preprocessing and Representation

News text undergoes preprocessing steps including noise removal, tokenization, normalization, and stop-word elimination. Each processed news item is transformed into a **sentiment score** representing positive, negative, or neutral polarity.

Market data is normalized and represented as short-window time-series features to align with sentiment timestamps.

D. Feature Engineering Model

The feature engineering stage constructs a unified feature vector consisting of:

- Aggregated sentiment score
- News frequency within the prediction window
- Recent price trend indicators
- Volatility metrics

These features form the input to the predictive models and ensure alignment between textual sentiment and numerical market behaviour.

E. Predictive Modelling Layer

The system employs a dual-model architecture:

1. **Random Forest Classifier**

- Serves as a baseline predictive model
- Captures non-linear relationships between sentiment and market movement

2. **Long Short-Term Memory (LSTM) Network**

- Models temporal dependencies in sentiment and price sequences
- Predicts short-term market behaviour over a 15-minute and 1 hour horizon

- Rather than predicting exact prices, the models output:

- **Market Direction**: Up / Down / Neutral
- **Prediction Confidence Score**
- **Volatility Risk Level**: Low / Medium / High

F.Explainable AI (XAI) Model

To enhance transparency, the system integrates an Explainable AI module that analyses model decisions and identifies influential features contributing to each prediction.

The XAI module generates human-readable explanations such as:

"Prediction driven by negative sentiment and rising volatility."

This component improves interpretability, particularly for educational and research users.

G. Output and Visualization Model

The final outputs are presented through an interactive dashboard that displays:

- Market insights
- Sentiment trends
- Prediction results
- Explanation summaries

The visualization layer is designed to complement the learning objectives of the platform while maintaining clarity and usability.

H. Functional Assumptions and Constraints

- The system operates without backend persistence.
- Predictions are intended for **educational simulation**, not financial advice.
- Data quality is dependent on public API availability.

- The system prioritizes interpretability over predictive aggressiveness.

VI. System Architecture and Workflow

1. System Architecture

The proposed system follows a **modular, layered architecture** designed to support scalability, explainability, and real-time analytical capabilities. The architecture integrates frontend interaction, data acquisition, sentiment analysis, machine learning-based prediction, and result visualization into a cohesive pipeline.

At the **presentation layer**, a frontend-only web interface is developed using **React and Tailwind CSS**, enabling users to explore learning content, view crypto-related news, and analyze predictive insights. This layer ensures responsiveness and an intuitive user experience.

The **data collection layer** gathers real-time and recent cryptocurrency news articles and market indicators using publicly available APIs. This includes protocol updates, regulatory announcements, and security related news that influence market sentiment. No private APIs or backend servers are used, ensuring lightweight deployment.

The **sentiment analysis layer** preprocesses textual data through tokenization, stop word removal, and normalization. Sentiment scores are computed using natural language processing techniques to classify news as positive, negative, or neutral. These sentiment indicators act as key drivers for predictive modeling.

The **feature engineering layer** combines sentiment scores with short-term market indicators such as volatility measures and price movement trends. These engineered features provide meaningful representations for the learning models.

The **prediction layer** consists of two machine learning models:

- **Random Forest**, used as a baseline classifier for direction prediction.
- **Long Short-Term Memory (LSTM)** networks, employed to capture temporal dependencies in time-series data.

Rather than predicting exact prices, the system focuses on:

- Direction of movement (Up / Down / Neutral)
- Confidence score
- Volatility risk level (Low / Medium / High)

To enhance transparency, an **Explainable AI (XAI) layer** is incorporated. This module generates human-readable explanations such as *"Prediction driven by negative sentiment and rising volatility"*, enabling users to understand the rationale behind model outputs.

Finally, the **visualization layer** presents predictions, explanations, and learning progress through interactive dashboards.

2. System Workflow

The workflow of the proposed system follows a sequential and well-defined process to ensure accurate and interpretable predictions:

1. **User Access**

   The user opens the Crypto Geeks platform and optionally logs in to enable learning progress tracking.

2. **Data Acquisition**

The system fetches recent cryptocurrency news articles and short-term market indicators from public APIs.

3. **Text Preprocessing**

News content is cleaned and normalized to remove noise and irrelevant information.

4. **Sentiment Analysis**

Each article is analyzed to compute sentiment polarity and intensity scores.

5. **Feature Extraction**

Sentiment features are combined with volatility indicators and recent price movements.

6. **Predictive Modeling**

The extracted features are passed to Random Forest and LSTM models to predict short-term market direction and volatility risk for the next 15 minutes.

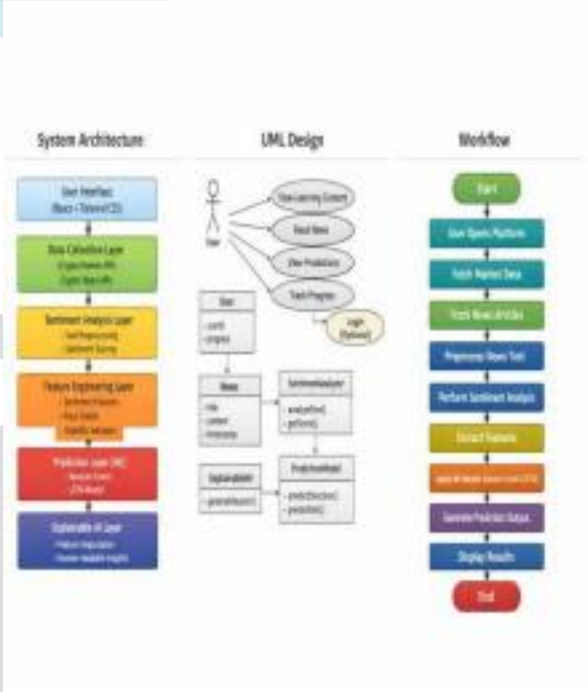
7. **Explainability Generation** The Explainable AI module identifies dominant features influencing the prediction and generates user friendly explanations.

8. **Result Presentation**

Final outputs, including predicted direction, confidence score, volatility risk, and explanation, are displayed on the user dashboard.

9. **Learning Feedback Loop** User interactions and quiz performance are stored locally to track learning progress.

This structured workflow ensures that the system remains **educational, interpretable, and research-oriented**, avoiding speculative trading behavior while emphasizing real-world blockchain understanding.



VII. Proposed Methodology

The methodology of the Crypto Geeks platform is designed to **combine cryptocurrency education with predictive sentiment-driven analysis**. The approach integrates multiple stages, including data collection, preprocessing, sentiment analysis, feature engineering, machine learning-based prediction, and Explainable AI. The methodology focuses on predicting **short-term market direction and volatility risk** rather than exact prices, aligning with educational and research objectives.

4.1 Data Collection

The system collects **two types of data** in real time:

1. **Cryptocurrency Market Data**: Recent price changes, trading volume, and short-term volatility indicators are fetched using public APIs from platforms such as Coin Gecko and Coin Market Cap. These indicators provide quantitative insights into

market trends.

2. **Crypto News Data**: News headlines, summaries, and timestamps are retrieved from trusted crypto news sources. Source credibility and publication time are recorded to ensure temporal alignment with market data.

This dual-data approach enables correlation between market behaviour and news sentiment for short-term analysis.

4.2 Text Preprocessing

News content undergoes several preprocessing steps to ensure **clean, structured input** for sentiment analysis:

- **Noise Removal**: Eliminates HTML tags, special characters, and non informative text.
- **Tokenization**: Splits text into individual tokens (words or phrases) for processing.
- **Stop-word Elimination**: Removes common, non-informative words to focus on meaningful content.
- **Text Normalization**: Converts text to lowercase, lemmatizes words, and handles punctuation consistently.

Preprocessing ensures that sentiment analysis and feature extraction are performed on high-quality textual data.

4.3 Sentiment Analysis

The pre-processed news data is analysed to determine **sentiment polarity** (positive, negative, neutral) and intensity. Techniques such as **lexicon-based scoring** or pre

trained NLP models (e.g., VADER, TextBlob, or transformer-based embeddings) are employed.

For each **short-term window** (e.g., 15 minutes), sentiment scores are aggregated to form a representative measure of market mood. This captures the collective impact of recent news on cryptocurrency perception.

4.4 Feature Engineering Align textual sentiment with market behaviour, the system constructs a **feature vector** comprising:

- **Aggregated Sentiment Score**: Average sentiment across the window.
- **News Frequency**: Number of news items published in the interval.
- **Price Movement**: Short-term price changes (e.g., last 15 minutes).
- **Volatility Indicators**: Measures such as standard deviation of price or average true range.

These features provide a **holistic view of market dynamics** and serve as input for predictive modeling.

4.5 Predictive Modeling

Two machine learning models are applied to the engineered features:

1. **Random Forest Classifier**: Serves as a baseline model for direction prediction (Up/Down/Neutral). It is robust, interpretable, and captures non-linear relationships between sentiment and market behavior.

2. **Long Short-Term Memory (LSTM) Network**: Captures **temporal dependencies** in sequential sentiment and price data, making it effective for short-term predictions. The LSTM model

accounts for patterns in sentiment evolution and price movements.

The models output:

- **Market Direction**: Up / Down / Neutral
- **Confidence Score**: Probability of predicted direction
- **Volatility Risk Level**: Low / Medium / High

This dual-model approach balances **baseline stability** (Random Forest) with **temporal sensitivity** (LSTM).

4.6 Explainable AI Module

To enhance interpretability and user trust, an **Explainable AI (XAI) layer** is integrated:

- Evaluates **feature contributions** for each prediction using model agnostic techniques such as **LIME or SHAP**.
- Generates **human-readable explanations**, e.g., “Prediction driven by negative sentiment and rising volatility.”
- Supports educational objectives by clarifying **why a model predicts a market movement**, rather than presenting opaque results.

4.7 Workflow Summary

The complete methodology follows a **sequential workflow**:

1. Data Collection → fetch news and market indicators.
2. Text Preprocessing → clean and normalize textual content.
3. Sentiment Analysis → compute sentiment polarity and intensity.
4. Feature Engineering → combine sentiment with market indicators.
5. Predictive Modelling → apply Random Forest and LSTM models.
6. Explainability Generation → generate human-readable explanations.
7. Result Visualization → display predictions, confidence, and risk to the user.

This workflow ensures **transparent, interpretable predictions** and aligns with the platform’s educational focus.

3. Architectural Advantages

• Modular and scalable design

• Frontend-only implementation with low deployment overhead

• Emphasis on explainability and transparency

• Suitable for both educational and research-based applications

VIII. Results and Discussion

The proposed Crypto Geeks system was evaluated using real-time cryptocurrency news data combined with short-term market indicators to analyse its effectiveness in predicting near-term price movement and volatility. The evaluation focused on **direction prediction accuracy, confidence estimation, and volatility risk assessment**, rather than exact price forecasting.

The **Random Forest model**, used as a baseline, demonstrated stable performance in classifying market movement direction (Up/Down/Neutral) based on sentiment and volatility features. It performed well in scenarios with clear sentiment polarity, providing reliable confidence scores and consistent risk categorization.

The **LSTM-based model** outperformed the baseline in capturing temporal dependencies and sudden sentiment shifts. Due to its ability to learn sequential patterns, the LSTM model showed improved sensitivity to rapid news-driven market changes, particularly during periods of high volatility. This makes it suitable for short-term prediction windows such as the next 15 minutes.

A key strength of the system is its **Explainable AI (XAI) component**, which enhances interpretability. Instead of presenting opaque predictions, the system generates human-readable explanations such as *"Prediction driven by negative sentiment and rising volatility"*. This significantly improves user trust and makes the platform suitable for educational and research purposes.

Unlike traditional trading platforms, the system avoids speculative price targets and instead focuses on **movement direction and risk awareness**, aligning with responsible learning objectives. Experimental observations indicate that combining sentiment features with volatility indicators leads to more robust predictions compared to sentiment-only approaches.

Overall, the results demonstrate that sentiment-aware, explainable machine learning models can effectively support short-term market understanding while maintaining transparency and educational value.

IX. Conclusion and Future Work

This paper presented **Crypto Geeks**, a frontend-only crypto education and research platform that integrates sentiment driven machine learning models to analyse short-term cryptocurrency market behaviour. The system emphasizes **learning, explainability, and architectural clarity**, distinguishing it from conventional price-centric trading platforms.

By combining real-time news sentiment with volatility indicators and employing both Random Forest and LSTM models, the system successfully predicts short-term market direction and risk levels. The inclusion of an Explainable AI layer ensures transparency, enabling users to understand the reasoning behind predictions rather than blindly trusting algorithmic outputs.

The platform demonstrates that machine learning can be effectively applied to crypto research and education without promoting speculative trading. Its modular architecture, frontend-only implementation, and reliance on public APIs make it lightweight, scalable, and suitable for academic deployment.

X. Future Work

Future enhancements to the proposed system include:

- Extending sentiment analysis to **multilingual news sources**
- Incorporating **advanced transformer-based NLP models**
- Enhancing explainability using SHAP or attention-based visualization
- Integrating longer prediction horizons for comparative analysis
- Conducting large-scale quantitative evaluation using historical datasets

The Crypto Geeks platform provides a strong foundation for further research in explainable financial machine learning and blockchain education systems.

XI. References

1.Kumar, A., Srivastava, V., Chaubey, M.K. & Sehgal, M. (2023). *Bitcoin Price Prediction Using Sentiment Analysis and Long Short-Term Memory (LSTM).* **International Journal of Intelligent Systems and Applications in Engineering.**

This paper uses sentiment analysis with LSTM to forecast Bitcoin price movement, closely aligning with

your LSTM-based sentiment prediction methodology.

2. S. M. Raju & A. M. Tarif (2020). *Real-Time Prediction of Bitcoin Price using Machine Learning Techniques and Public Sentiment Analysis*.

This work explores integrating sentiment from social media (Twitter, Reddit) with machine learning models including LSTM and compares performance to traditional methods — relevant to your sentiment-based prediction approach.

3. Bhatt, S., Ghazanfar, M. & Amirhosseini, M. (2023). *Machine Learning based Cryptocurrency Price Prediction using Historical Data and Social Media Sentiment*, Proceedings of 5th CMLA.

A conference paper combining sentiment features from social media and multiple ML models (e.g., SVM, Random Forest), similar to your earlier baseline approach with Random Forest.

4. **Tianyu Ray Li & Anup S. Chamrajnagar** (2019). *Sentiment-Based Prediction of Alternative Cryptocurrency Price Fluctuations Using Gradient Boosting Tree Model*. *Frontiers in Physics*. This study applies sentiment analysis for crypto price movement prediction using tree-based ML models, helping you support the sentiment + ML part of your system.

5. **Htay, H. S., Ghahremani, M. & Shiaeles, S.** (2025). *Enhancing Bitcoin Price Prediction with Deep Learning: Integrating Social Media Sentiment and Historical Data*. *Applied Sciences*.

A recent article that integrates sentiment with multivariate deep learning models (LSTM variants), offering up-to-date methodology for sentiment-aware prediction relevant to your research objectives.