Machine Learning Project Proposal

USMAN AHMAD BSDSF22M033

AREEBA CHAUDHRY BSDSF22M018

ZAINAB NAVEED BSDSF22M013

Project Title

Bitcoin Price Prediction Using Sentiment Analysis and Historical Data Modeling

Problem Statement

Cryptocurrency markets are known for their volatility, influenced not only by technical indicators and historical data but also by real-time public sentiment, especially from platforms like Twitter. This project proposes a hybrid approach to predict the next day's Bitcoin price using two independent sources: historical price trends analyzed with machine learning models such as LSTM or regression, and sentiment scores derived from tweets using Natural Language Processing (NLP).

Objectives

- Collect real-time tweets and historical price data for Bitcoin and selected cryptocurrencies.
- Perform sentiment analysis on tweets to classify public opinion as positive, negative, or neutral.
- Build a predictive model using historical price data to forecast the next day's price.
- Combine insights from sentiment analysis and price prediction to enhance decision-making.
- Develop an intuitive frontend with Streamlit and optionally integrate FASTAPI for fast predictions.

Proposed Methodology

- Tweet Collection & Preprocessing

Use Twitter API to fetch tweets and clean them using NLP techniques (removal of URLs, emojis, etc.).

- Sentiment Analysis

Apply pre-trained NLP models or train a classifier to categorize tweet sentiments.

- Historical Data Modeling

Use historical Bitcoin prices to train models like LSTM, Linear Regression, or XGBoost.

- Hybrid Decision Logic

Combine sentiment trend and price forecast to generate a final predictive insight.

- Frontend Development

Build an interactive user interface using Streamlit and optionally enable quick API responses using FASTAPI.

Machine Learning Project Proposal

Dataset Description

1. Tweets Dataset: Live tweets collected using Twitter API filtered with keywords such as #Bitcoin, #BTC.

2. Price Data: Historical prices of Bitcoin and 2-3 other coins obtained from sources like CoinGecko or Yahoo Finance.

The datasets will be synchronized by timestamp to analyze the relationship between public sentiment and market performance.

Expected Outcomes

- A dual-model prediction system combining historical price modeling and tweet sentiment analysis.
- An interactive platform where users can input a date or see the next day's forecast based on real-time data.
- Insightful visualizations showing the correlation between tweet sentiments and price movements.
- A foundational framework for expanding to more cryptocurrencies or longer-term forecasting.

Timeline

Day 1-3: Data collection (tweets and price data) and preprocessing.

Day 4-6: Sentiment model development and analysis.

Day 7-9: Time series modeling with historical data.

Day 10-12: Integration and development of hybrid prediction logic.

Day 13-14: Frontend creation with Streamlit and API setup.

Day 15: Testing, polishing UI, documentation, and submission.