

NLP Project proposal

Cryptocurrency Trend Prediction With Natural Language Processing

Collecting Data Phase

Mostafa Ahooei (std num. 97521117)
Sahand Nazarzadeh (std num. 97522265)

Abstract

This project combines natural language processing, deep learning, cryptocurrency market prediction, and sentiment analysis. Its goal is to develop a predictive model that can analyze news articles and social media posts related to the cryptocurrency market, and forecast the future value and trend of cryptocurrencies.

Introduction

The project will make use of natural language processing (NLP) techniques to extract relevant information from news articles and social media posts. NLP algorithms will be used to identify key phrases and sentiment in the text. Deep learning models will be developed to analyze the extracted information and make predictions about the future value and trend of crypto currencies.

The project will focus on sentiment analysis as a key factor in predicting the stock market. Sentiment analysis will be used to analyze the tone and emotion expressed in the news articles and social media posts related to the crypto currency market. This analysis will help to identify whether the sentiment is positive, negative, or neutral, and how it may impact the crypto currency market.

The project will be divided into several phases. The first phase will involve data collection and preprocessing. The second phase will focus on developing the NLP and deep learning models. The third phase will involve evaluating the models and refining them based on their performance.

The final outcome of the project will be a predictive model that can analyze news articles and social media posts related to the crypto currency market and predict the future value and trend of

crypto currencies. The model will be evaluated based on its accuracy and its ability to generate profitable trading strategies.

First Phase (Data collecting)

This project aims to collect two types of data to create a dataset that includes natural language data as features and price trends as labels. The data is related to financial markets and will primarily be in English, encompassing both formal and informal language styles. The natural language data will consist of characters, and the price trends will be represented as numerical values. The data will be classified into three main classes: positive emotions, neutral emotions, and negative emotions. The interesting thing about the data is that it can be used to develop trading strategies. Our hypothesis is that the data has a significant impact on price trends in the short and long term. However, the main limitation of the project is that the cryptocurrency market is relatively new, with the oldest available data dating back to 2000. These types are:

1. Language Based Data

Data collection is a crucial step in any machine learning project, and it is particularly important in natural language processing (NLP) projects. In the context of the proposed project, data collection involves gathering relevant news articles and social media posts related to the crypto currency market, which will be used to train and test the predictive model.

There are several sources of data that can be used for this project. News articles can be obtained from various financial news websites, such as Bloomberg, Reuters, and Yahoo Finance. Social media posts can be collected from platforms such as Twitter, TradingView, and CoinMarketCap. These sources provide a wealth of information that can be used to

analyze market sentiment and make predictions about future crypto currency prices.

Once the data sources have been identified, the next step is to collect the data. This can be done using web scraping tools or APIs provided by the data sources. Web scraping involves extracting data from websites using automated tools, while APIs provide a more structured and standardized way of accessing data.

2. Price Based Data

Collecting data from the Binance API involves using the API to retrieve trading data such as price, volume, and order book information. The Binance API provides several endpoints that allow you to retrieve different types of data, including historical and real-time data.

Resources

- Bloomberg: Has API for news data
- Reuters: Has API for news data
- Yahoo Finance: Has API for news data
- Twitter: Has not API social media posts data
- TradingView: Has API for social media posts data
- CoinMarketCap: Has API for social media posts and numerical data
- Binance: Has API for numerical data