Cryptocurrency Sentiment analysis

By: Crystal Han

Abstract

The goal of this project was to use data collected from APIs or extract data from already collected sentiment values of the cryptocurrency market. APIs and data are extracted from Augmento and SentiCrypt website to analyze and predict the price of cryptocurrency value. In this model, Linear Regression is used to determine the price of cryptocurrency based on the sentiments valued ranging from negative to positive. Data of the cryptocurrency sentiment values are collected every hour and every day from November 2016 to January 2022. Augmento collects cryptocurrency related talks from Reddit, Bitcointalk, and Twitter. The Augmento API allows the coder to choose what kind of coins to collect data from. Since there were already data collected for Bitcoin and Ethereum these were the two coins in consideration. In this analysis, I provide visualization of the relationship between sentiment and cryptocurrency value.

Design

The data for this project is extracted from Augmento and SentiCrypt. The amount of data is well over 100,000 data points, which is then condensed to use the mean values. Features were selected based on negative and positive talks of Bitcoin and Ethereum. Although some analysis was done for SentiCrypt in Jupyter Notebook it is not included in the Flask Web App. Clients or users interested in the rise and fall of cryptocurrency value would be interested in seeing how sentiment can affect the worth of the coins. Flask and Heroku were chosen to create a web app to allow users to interact and play around with the sentiment values and see plots of sentiment against listing close price of cryptocoins.

Data

The dataset was chosen based on negative and positive talks about Bitcoin and Ethereum. Those were the main focus and is categorized based on the websites the data was collected from. Features were then condensed to use the mean as a way to allow the dataset to be less large and allow users to see the overall relation between listing close price and sentiment value, and predict cryptocurrency price.

Data Pipeline
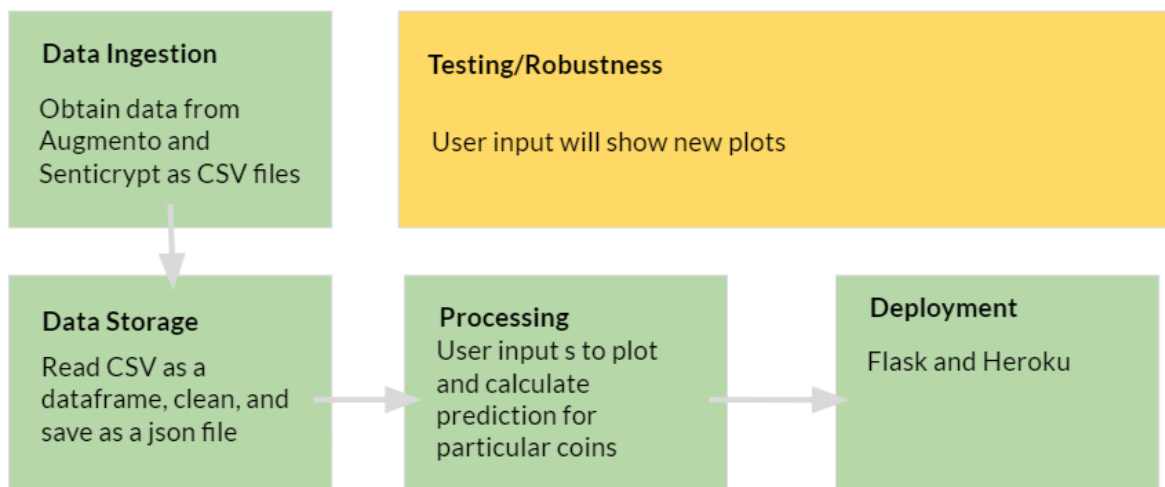
Algorithms

Feature Engineering

1. Continuous Variables: Twitter, Bitcointalk, Reddit. Negative and positive talks
2. Linear Regression

Models

Linear Regression was used to predict the price for cryptocurrency value.

Model Evaluation and Selection

The entire training dataset of data was split into 67/33 train vs. test. All scores reported were calculated with 5-fold cross validation on the testing portion to see how well we fared. Training and testing data were also scored to determine the performance of models.

Tools

- NumPy and Pandas for data manipulation
- Matplotlib and Seaborn for plotting and visualizations
- Scikit-learn for modeling
- Flask and Heroku for web app deployment

Communication

```
Linear Regression R2 train: 0.6419186244895757
Linear Regression MSE: 100548795.00038819
Linear Regression R2 test: 0.6101329512177643
Linear Regression MSE: 120925019.06479543
Mean absolute error: 7463.618134340333
0.5905028669544384
```

Figure 1 Linear Regression Model and Scoring

Pairwise Relationship Negative sentiment vs Listing price (Augmento BTC)
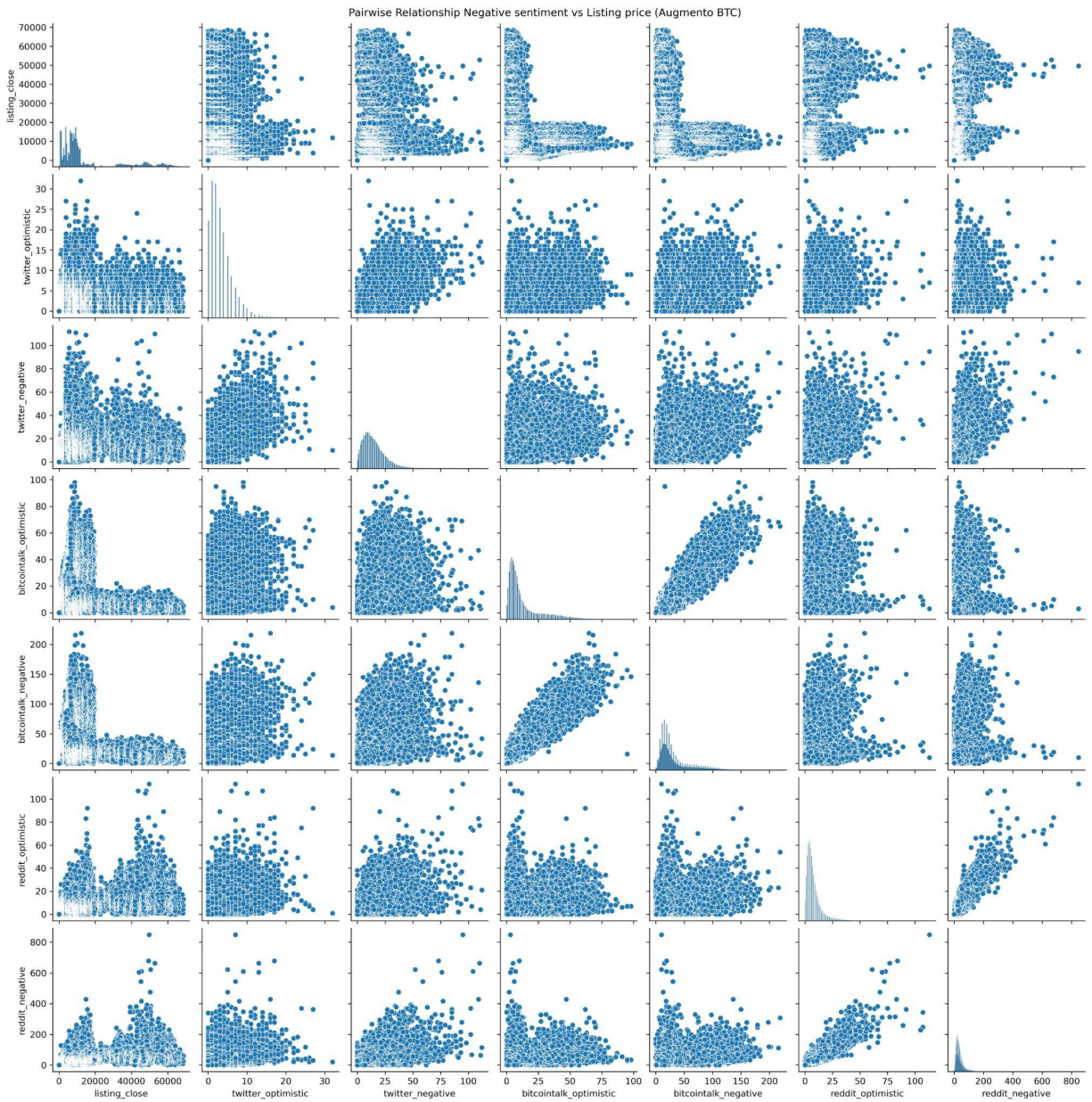

Figure 2. Correlation between sentiment and cryptocurrency listing close price (Augmento BTC)
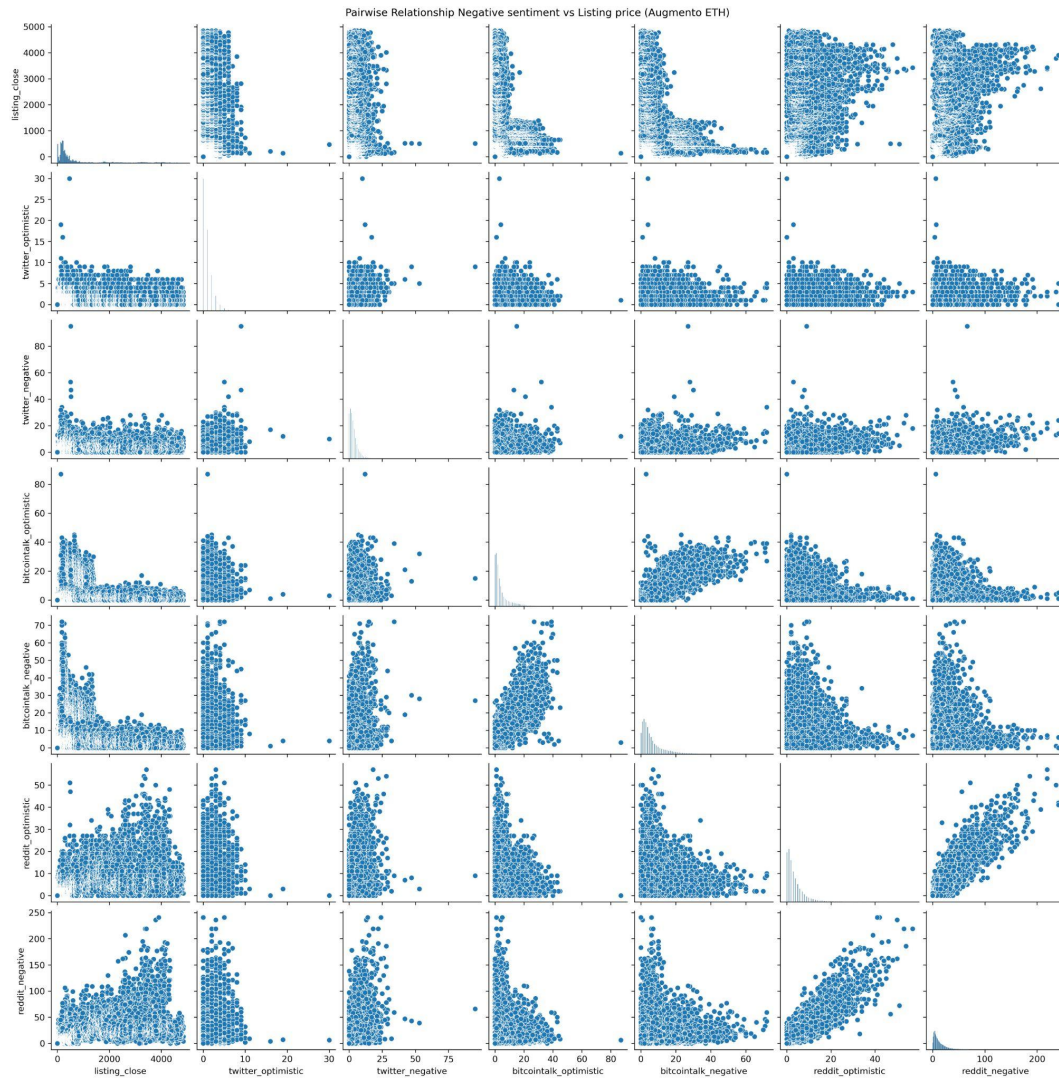
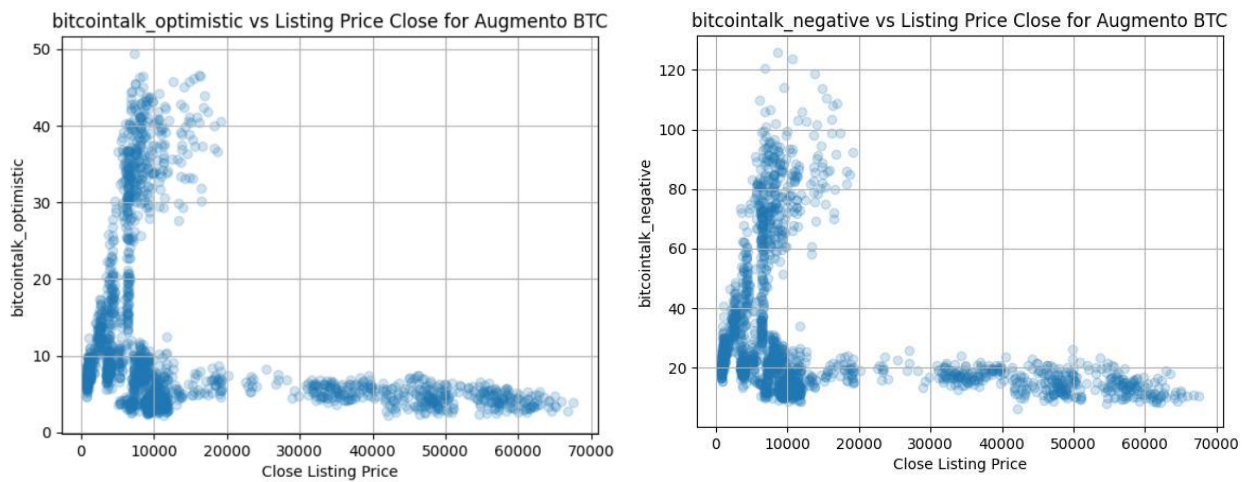Figure 3 Correlation between sentiment and cryptocurrency listing close price (Augmento ETH)

Figure 4. Comparison between Bitcointalk Optimistic and Bitcointalk Negative

Select a Sentiment Value:
○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○
listing_close twitter_negative twitter_optimistic bitcointalk_negative bitcointalk_optimistic reddit_optimistic reddit_negative bitcointalk_positive reddit_positive twitter_positive twitter_prediction bitcointalk_prediction reddit_prediction

Select a Dataframe:
○ btc_aug ○ eth_aug

twitter_negative	twitter_optimistic	bitcointalk_negative	bitcointalk_optimistic	reddit_optimistic	reddit_negative	bitcointalk_positive	reddit_positive	twitter_positive
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Submit

Model Prediction:
listing_close : 1459616.49

Model Graphs:

Figure 5. Flask Web App deployment