Time Series Prediction on Social Media Data

R S Subbulakshmi | 20171069
Atirek Kumar | 20171060
Pulkit Sharma | 2019201006
Himanshu Kumar | 2019201094

Overview

The objective of the project is to implement **Sentiment Analysis** on a database scrapped from Social media and thereby do **Time series prediction .** Here we are specifically taking the case of Cryptocurrencies and its popularity in social media sites, thereby affecting the stock prices in the market, especially Bitcoin from a Twitter database.

Goals

1. Scrape and maintain a considerable sized database from a Social Media site(ie. Twitter)
2. Apply Sentiment analysis and gain how Bitcoin is favoured or not amongst the global citizens
3. Design and implement models for training and testing and thus achieve Time Series prediction on Bitcoin and its stock prices in the market
    - Create a Baseline model using SVM
    - Create secondary model using RNN/LSTM
    - Create time series model CNN

Data Scrapping

The data is specifically scrapped from 2 prominent sources

1. Twitter - top tweets of a specific day which are relevant for Bitcoin
2. Coinmarketcap.com - official prices of the stocks at given time stamps

Sentiment Analysis

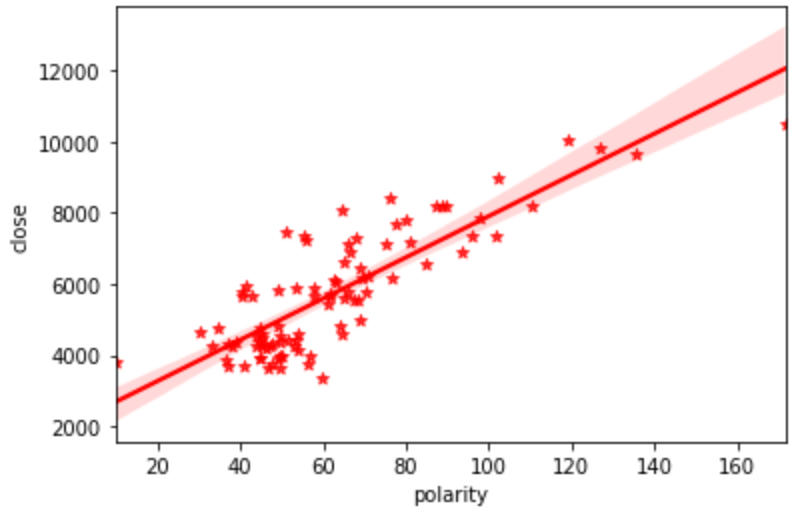
Current sentiment analysis is done using Text Blob to gain basic polarity of the tweets and thus according to a self ascertained range, we deem the tweet to be positive, negative or neutral.

We have applied the Multinomial Naive Bayes classifier for text classification with Regex Tokenizer and CountVectorizer used for relevant word extraction (keeping the language english and removing stop words).

Accuracy:0.69

Data Analysis

Plotting the polarity data vs bitcoin price over 3 months we found a correlation between the two with a `Pearson Score of 0.83.`



Time Series Prediction

Model 1:

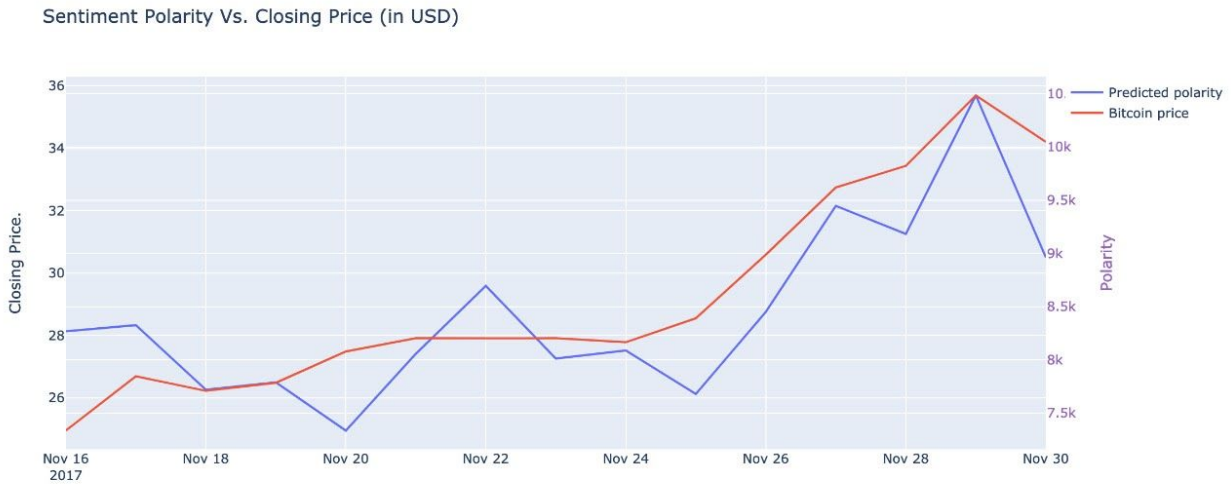
Basic linear regression with only past stock prices :

Model 2:

Basic RNN with 4 layers, adam optimizer and MinMaxScaler normalization. This also takes only old stock prices for training and prediction

Model 3:

Polarity prediction based on sentiment analysis of previous tweets to see trends in bitcoin price.



Model 4:

Price prediction of next hour based on last hour prices and last day polarity using CNN with **r2 score=0.98**.

Price prediction by CNN

Relevant papers and sites referred

http://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf

Goals for future

1. Compare and test with ARMA, Prophet models for TIme series prediction
2. Manipulate different model criterias for higher efficiency in Time series prediction
3. Implement and compare different techniques and models for higher efficiency in Sentiment analysis