MACHINE LEARNING NANO DEGREE

Capstone Project :

Stock market prediction based on twitter posts by influential personalities

VAMSHI KRISHNA VENKATASWAMY

MAY 19TH ,2020

Description

Investment firms, hedge funds and even individuals have been using financial models to better understand market behavior and make profitable investments and trades. A wealth of information is available in the form of historical stock prices and company performance data, suitable for machine learning algorithms to process.

The project I am going to implement is to do with finance and stock predictions, one of the main reasons for choosing the stock market is the vast amount of data available and secondly we would like to see how we could use machine learning to predict the fluctuation of the entire market or a particular stocks based on tweets made by influential personalities. Not too long ago during the rise of cryptocurrencies Charlie Lee creator of Litecoin tweeted to all Litecoin holders that these were good times and one should be prepared for and to expect the coin value to fall to as low as $20, this tweet caused a stir in the market causing the coin to drop hundreds of dollars in in matter of no time, however when he tweeted about some of the new security features and robustness of the software we saw a spike in the coin price. Another example is President Trump tweeting on increasing import duty on Chinese products, we saw a frenzy in the markets. In today's world where information is readily available and at your fingertips, we think and believe that the paradigm of how stocks rise and fall based on information has changed in a way where we can use machine learning to use available data to decide if the outcome will cause a positive or negative impact.

PROBLEM STATEMENT:

The main aim of this project is to predict how stocks rise and fall based on information has changed in a way where we can use machine learning to use available data to decide if the outcome will cause a positive or negative impact of a certain stock using historic data

--Obtain data from yahoo finance and tweets from twitter

--Libraries: Sentiment Analyzer, numpy, pandas, NaiveBayes

-- Algorithms: Naïve Bayes, Neural Networks, Logistic Regression

--Training the model

--Data Pre-processing

--API: Twitter API, StocksAPI from rapidAPI

Datasets:

The datasets used for this project are obtained from Yahoo Finance. The tweets are taken from twitter and tesla tweets using Twitter API, StocksAPI from rapidAPI. For a stock there are 7 values particularly features namely Date, Open, High, Low, Close , Volume and Adj closing price. And the tweets contain the date and tweets.

Solution Statement:

I used sentiment Analyzer to calculate the polarity and confidence of the tweets and applied Naïve Bayes, Neural Networks, Logistic Regression algorithms to predict the best model for prediction and train the neural network using dropouts method for preventing overfitting.

Benchmark Model:

The benchmark model will utilize the prediction from the sentiment analyzer model and actual stock price predicted by the model.

Project Design:

Data pre-processing using Normalizing, Data transformation, Data Cleaning. For predicting the stock price and applying train and test sets to various models and selecting the best model for prediction.