

Cryptocurrency Daily Price Predictor

Using machine learning and sentiment analysis to predict cryptocurrency
gains and losses

Matthew Viafora

Overview:

With the recent boom and hype behind cryptocurrencies, there seems to be new cryptocurrencies and NFTs popping up every single day. The main driving force behind these new currencies is hype and sentiment in the community. As was seen with cryptocurrencies Bitcoin, Ethereum, Dogecoin, etc. as well as some stocks involved in the January 2021 investing drama including GME (Gamestop), AMC, NOK (Nokia), it is clear that community drive heavily influences currency or stock price. Utilizing this community sentiment along with previous price trends, we can create a model that outputs short-term gains and losses, while at the same time indexing user sentiment over time to predict long term gains and losses.

Technologies:

In order to perform sentiment analysis on what the community is saying about any given cryptocurrency, we need to pull tweets from the Twitter API under “cryptocurrency” and related tags to the currency we are analyzing. After this data is gathered and formatted, it will be bounded between a value of 0 and 1 representing the mean sentiment for a given day where 0 is negative sentiment and 1 and positive sentiment. This data is then merged with our dataset that will be trained on a Microsoft Azure virtual machine to streamline training in order to form a hypothesis representation of our data to make future predictions. Technologies we are using include:

- Python 3
- Tensorflow 2
- Keras
- Twitter API
- Pandas
- Sci-Kit Learn
- Numpy
- Matplotlib
- Microsoft Azure Virtual Machine

Automated High Level Timeline of Algorithm:

D_i = Given day

X_j = Input features

$H(X_j)$ = Hypothesis representation

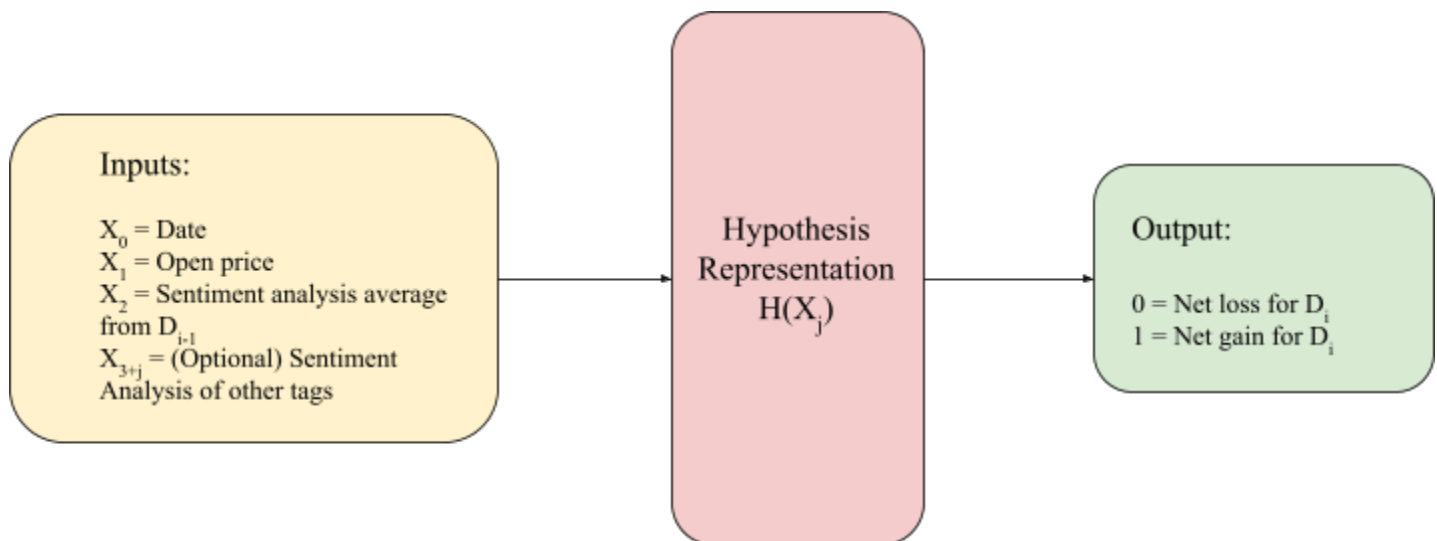
0 = Net loss for given day

1 = Net gain for given day

FOR D_i :

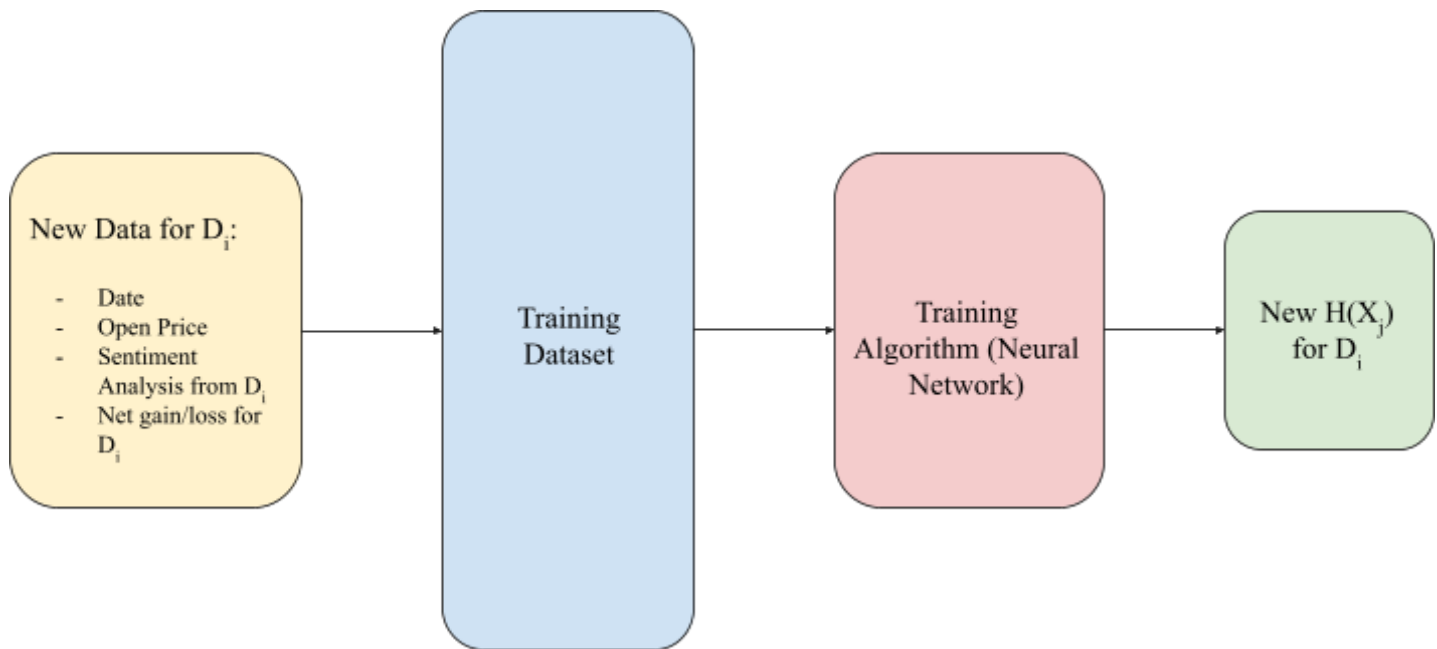
TIME: 12:00 AM

- Inputs are gathered from sources and run through the model to evaluate output for that given day.



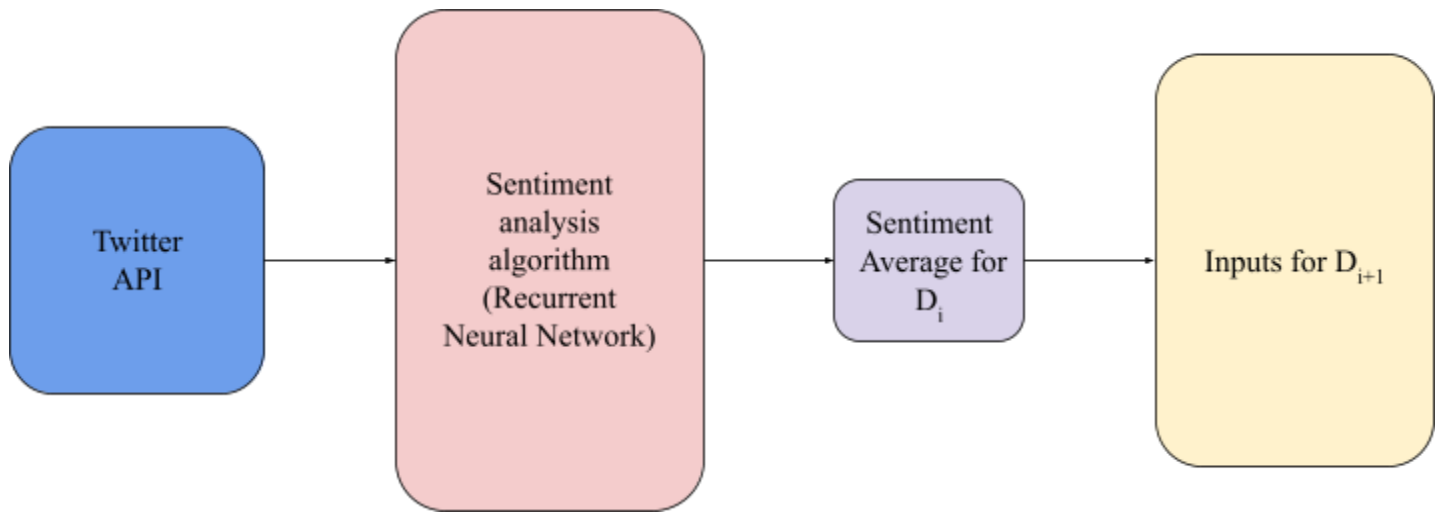
TIME: 11:55 PM

- Data from D_i , including the actual net loss or gain is compiled and added to the cumulative dataset in which the model will be trained from.
- New model is trained with updated dataset obtaining a new $H(X_j)$



TIME: 11:59 PM

- Request Twitter API for ≥ 500 tweets from D_i relating to specific cryptocurrency tag.
- Perform sentiment analysis on tweets to obtain an average sentiment score bounded between 0 and 1 for that day.



Process Overview

