Social Media Predicts Future Stock Price Trends

PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING

Background

Traditionally, analytics use statistical model built on past stock prices and recent news to forecast future stock price trends.

Motivation

Social media, such as Twitter, often reflects how people think about a company and therefore can be used as an indicator of the changes of stock price in the near future.

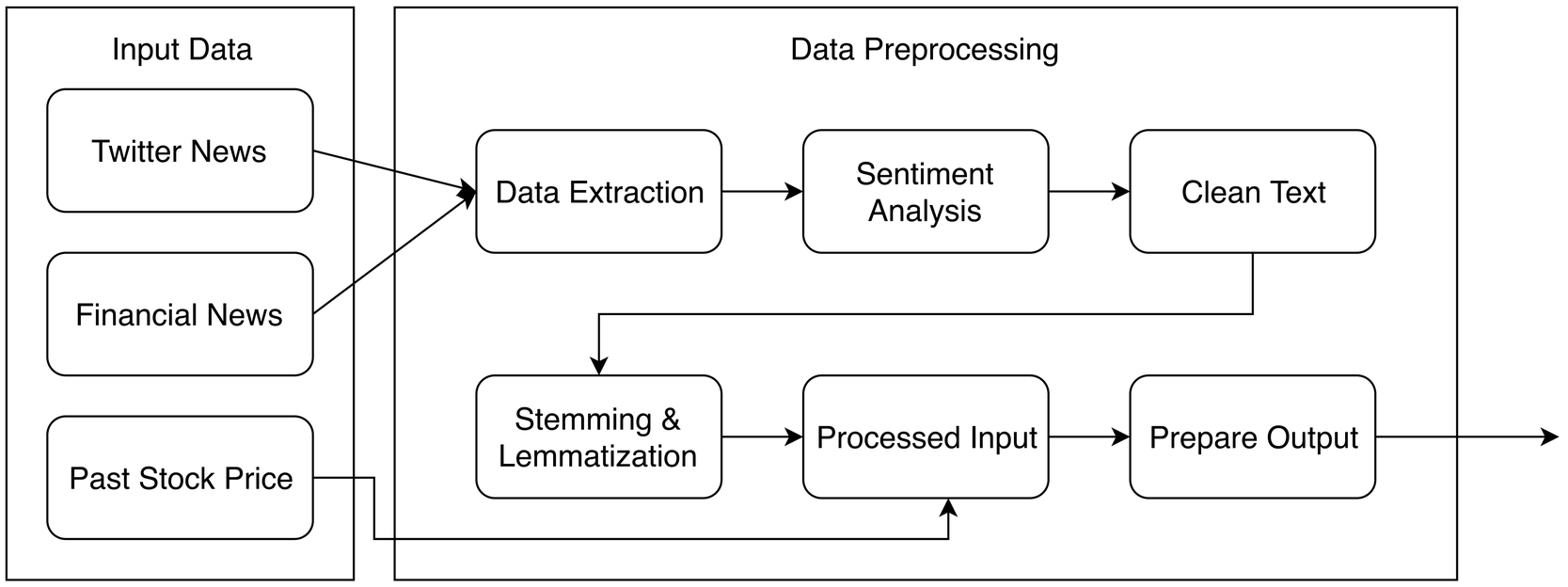
Overview

We would like to apply ML and NLP models on social media to see if it has enough information for us to make good prediction of future stock price.

Data Processing

We crawled financial news and relative news headlines from Twitter as our input corpus.

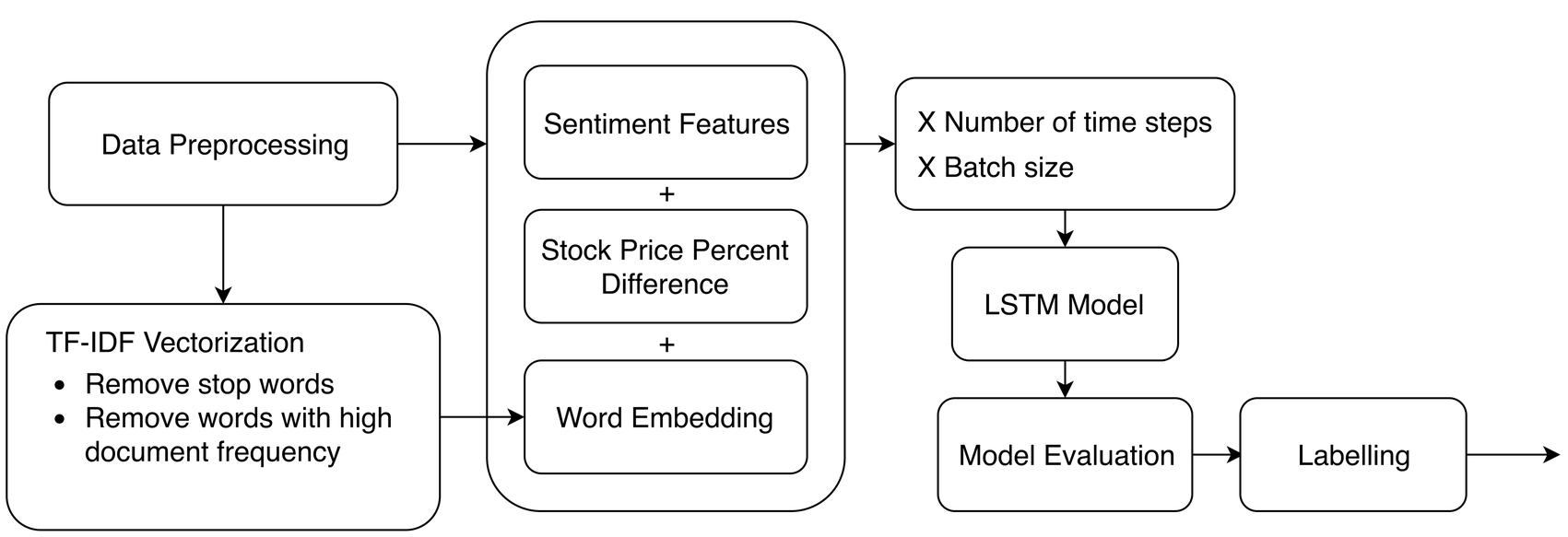
Sentiment analysis was done before cleaning text since the analyzer we used, VADER, takes emoticon and punctuation into account.



Model & Featurization

We tried out logistic regression, vanilla RNN, LSTM and different featurization strategies.

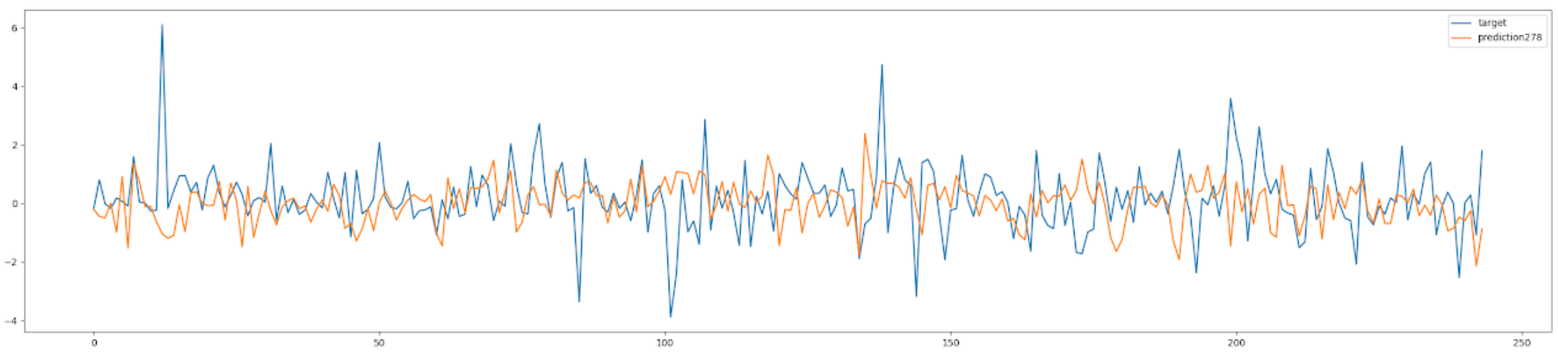
Our model takes word embedding vectors and history stock prices as input and predicts the rate of change of future stock price. We then further label the predictions as DOWN, STAY, and UP.

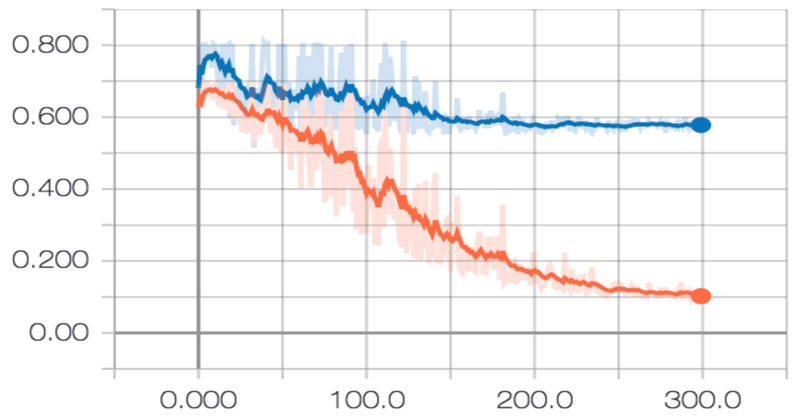


Result and Comments

The bottleneck for this project is the input data noise, which makes it hard to extract meaningful features for our model training.

One possible way to improve the performance is using labelled Twitter and financial news data (attitudes toward the company or its products; noise data or not).





- Accuracy on Tesla without sentiment features: 34%
- Accuracy on Tesla with sentiment features: 41%
- Accuracy on Apple with sentiment features: 44.3%

Dawei Shen Eric Wang Antony Liu