Sentiment Analysis for Financial News

Harrison DiStefano, Jaehee Jeong, Lisa Kaunitz, Junghwan Park, David Sun

Table of Contents

01 **Problem Statement**
Literature Review

02 **Methods**
Tools Utilized & Our Data

03 **EDA**
Data Exploration

04 **Analysis**
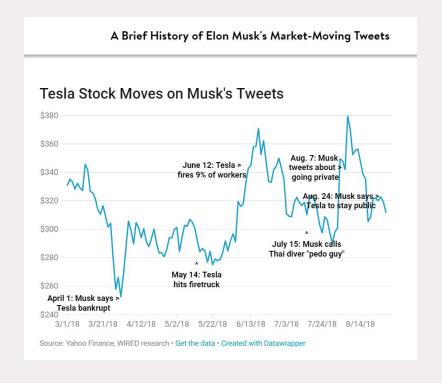Models & Demo

05 **Conclusion**
Results & Future Recommendations

Problem Statement

- There is a statistically significant association between strong media sentiment and abnormal market returns and volatility[1]

- While we can not directly change or control the headlines, our goal is to

  1. **Build a model that predicts news sentiments which may affect changes in the market.**

  2. **Implement an app that can take in a potential news headline and output the predicted sentiment.**

A Brief History of Elon Musk's Market-Moving Tweets

Tesla Stock Moves on Musk's Tweets

June 12: Tesla fires 9% of workers

Aug. 7: Musk tweets about going private

Aug. 24: Musk says Tesla to stay public

July 15: Musk calls Thai diver "pedo guy"

May 14: Tesla hits firetruck

April 1: Musk says Tesla bankrupt

Source: Yahoo Finance, WIRED research • Get the data • Created with Datawrapper

[1]Wan, Xingchen, et al. "Sentiment Correlation in Financial News Networks and Associated Market Movements." *Nature News*, Nature Publishing Group, 4 Feb. 2021, https://www.nature.com/articles/s41598-021-82338-6.
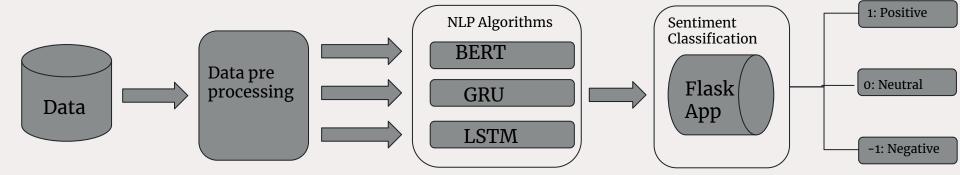
Methods

Our Data

- 4,845 financial news headlines with predetermined sentiment labels
    - Neutral
    - Positive
    - Negative

- Imported as .csv via Kaggle

Tools Utilized

- Github

- Python (numpy, pandas, TensorFlow, Keras, matplotlib)

- R (ggplot, nlp, tm, caTools, topicmodels, fpc, RTextTools)

- Flask

Data → Data pre processing → NLP Algorithms (BERT, GRU, LSTM) → Sentiment Classification (Flask App) → 1: Positive, 0: Neutral, -1: Negative

Exploratory Data Analysis

Our Data at a Glance

Positive
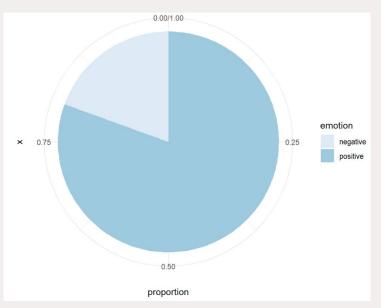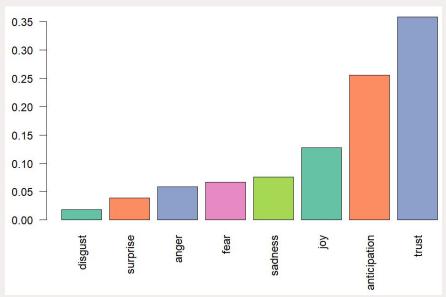

Neutral

Negative

Exploratory Data Analysis

Composition of 10 sentiments in all headlines after tokenization looks like :

Exploratory Data Analysis
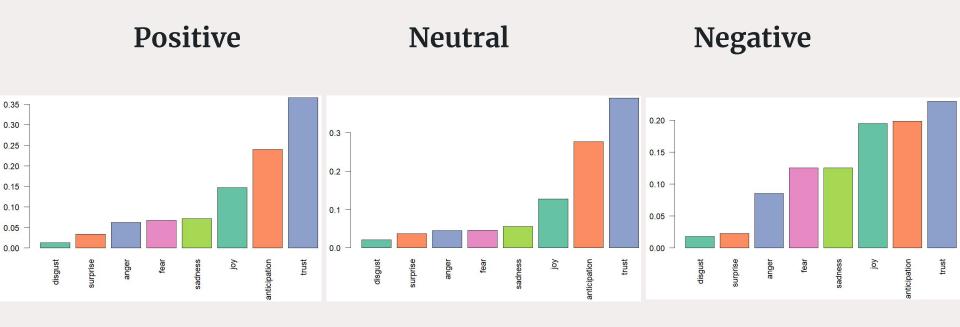
8 sentiments for the positive, negative, neutral after tokenization looks like :

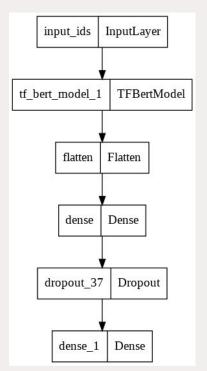## Positive

Neutral

Negative

Models

- Text Cleaning
  - Remove Punctuations
  - English Only
- Train & Test Datasets: 80/20
- Models:
  - LSTM (Long Short Term Memory)
  - GRU (Gated Recurrent Unit)
  - BERT (Bidirectional Encoder Representations from Transformers)
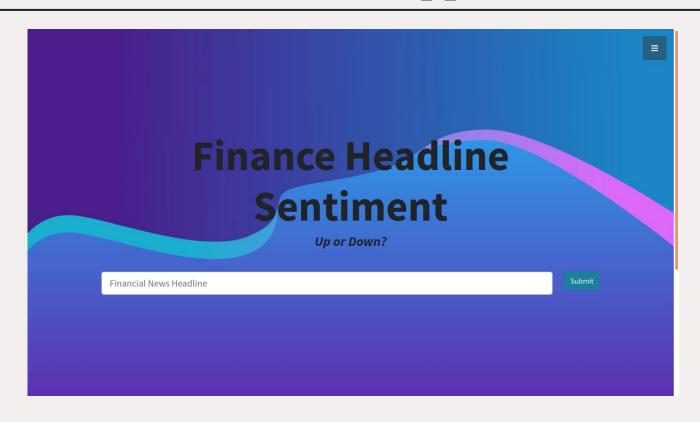   - Sentimental Analysis
    - Fine-Tuning

Results

BERT Arch.

Models	Epoch	Loss	Test Accuracy
LSTM	18	0.845	63.9%
GRU	23	0.838	63.4%
BERT Sentimental Analysis	-	-	23.1%
BERT Fine-Tuning	8	0.43	84.6%

Further: Flask App Demo

Further: Flask App Demo

Shortcomings

- No price data and date data
  - We can't check how the new sentiments affect stock prices

Future Recommendations

- Building a stock market price forecasting model
  - Collecting data
    - Date, Stock price per day for a specific company
    - Specific company's news headline
    - Other features like dividend yield, price to book, etc.
  - Using sentiment analysis result as a predictor feature
- Adding more labels
  - Now:
    - Negative/ Neutral/ Positive
  - Future:
    - Strongly Negative/ Weakly Negative/ Neutral/ Weakly Positive/ Strongly Positive

Thank you
&
Questions?