|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | | | |
|  |  |  | |  |
| **Applications of sentiment analysis for the prediction of stock prices** | | |
| group assignment: Machine Learning  group 8  Patrick McDonnell  Karan Nandani  IMMACULATE EZEOGBO  vida rahmani  DAVID NGUYEN  Xin Tan Wang | | APRIL 2021 |

Contents

[MEMBERS 3](#_Toc68182909)

[GROUP 8 3](#_Toc68182910)

[INTRODUCTION 3](#_Toc68182911)

[OBJECTIVE 3](#_Toc68182912)

[DATA PREPARATION 3](#_Toc68182913)

[DATA SOURCES 4](#_Toc68182914)

[PREPARATION PROCESS 4](#_Toc68182915)

[MODEL DESIGN 4](#_Toc68182916)

[FLAIR 4](#_Toc68182917)

[MODEL 2 4](#_Toc68182918)

[MODEL 3 4](#_Toc68182919)

[TIME-SERIES ANALYSIS 4](#_Toc68182920)

[MODEL EVALUATION 4](#_Toc68182921)

[FLAIR 4](#_Toc68182922)

[MODEL 2 4](#_Toc68182923)

[MODEL 3 4](#_Toc68182924)

[TIME-SERIES ANALYSIS 5](#_Toc68182925)

[CONCLUSIONS 5](#_Toc68182926)

[REFERENCES 5](#_Toc68182927)

# MEMBERS

## GROUP 8

1. Patrick Mcdonnell – pmcdonne (pmcdonne@uwaterloo.ca)
2. Karan Nandani – knandani (knandani@uwaterloo.ca)
3. Immaculate Ezeogbo – iezeogbo (iezeogbo@uwaterloo.ca)
4. Vida Rahmani – v2rahman (v2rahman@uwaterloo.ca)
5. David Nguyen – d235nguy (d235nguy@uwaterloo.ca)
6. Xin Tan Wang – xtwang (xtwang@uwaterloo.ca)

# INTRODUCTION

Stock market investments play a major role in the finance sector. The volatile nature of the stock market has equal chances for earning money as well as losing money; however, profits can be maximized, and losses can be minimized by predicting the trends. Many factors affect the movement of the stock market and the sentiments of the traders are one of the significant driving factors which can change the trend of the market and affect the price of a particular stock [1].

Many AI companies use sentiment analysis to predict trends and movements of stocks. Sentiment analysis is the process of analyzing sentiments of people through various platforms where opinions and thoughts are freely expressed. Social media platforms, such as Twitter, are great networking sites to understand the sentiments of the traders and investors. Easy access to social media platforms through devices such as smartphones has made the spread of content and opinions happen at a very fast pace. News and information influence people around the world and will affect the decisions of investors on buying or selling of their stock holdings. Negative sentiments will likely pull the stock price down and positive and optimistic sentiments will likely drive the stock price up.

The tools provided by machine learning and natural language processing along with other approaches, make it possible to work with large volumes of real-time data and to extract sentiments from social media. The extracted data is then used to train AI models through sentiment analysis algorithms which is then used to predict the price of stocks.

# OBJECTIVE

This project will investigate applications of sentiment analysis for prediction of stock fluctuations and stock prices. The goal is to predict the reaction of the stock market to certain news, posts, and online conversations extracted from Twitter.

In this project, tweets related to **“electric vehicles”** were pulled, sentiment analysis was carried out and the results were used in models for predicting **“Global X Autonomous & Electric Vehicles”**(represented with **DRIV** ticker) stock values. The following objectives will be pursued in this project:

* Sentiment analysis over tweets related to “Global X Autonomous & Electric Vehicles”
* Stock price prediction for “DRIV” stock using sentiment analysis models (flair, MODEL 2, MODEL 3)
* Stock price forecasting for “DRIV” stock using time-series analysis
* Evaluating the sentiment analysis models by comparing with the time-series analysis approach

# DATA PREPARATION

## DATA SOURCES

Twitter API was set up and tweet data which were related to “electric vehicles” were pulled for the time period between ... and ... using the requests library. The bearer token given in the API setup process was passed through the authorization key in the request header and the search query was specified by adding ?q=<SEARCH QUERY> to the API address. Furthermore, in order to avoid tweet truncation parameter tweet\_mode=extended was added to the request and in order to specify the language parameter lang=en for English was included in the request.

## PREPARATION PROCESS

The Twitter API response was transformed into a Python dictionary using .json() and was iterated through to get information on all the returned tweets. The tweets returned by the API contained three fields: tweet ID (id\_str), creation date (created\_at), and untruncated text (full\_text), which were extracted using the get\_data function. Finally, the extracted tweet data were appended to the dataframe.

Cleaning the collected tweets was a necessary and fundamental step. Excessive whitespace, web addresses, and Twitter users were identified and cleaned using regular expressions (RegEx) through the re module.

# MODEL DESIGN

## FLAIR

The flair library offers a powerful pre-trained sentiment analysis model which splits the text into character-level tokens and uses the DistilBERT model to make predictions. DistilBERT is a distilled version of the powerful BERT transformer model and is considered a ‘small’ model (only 66 million parameters). However, it is still super powerful [2].

After installing flair, the model was imported and initialized and text was tokenized by passing it through flair.data.Sentence() and calling the predict method on the model. By calling the predict method, the sentiment rating was added to the data stored in sentence. The confidence/probability score and the value for the positive/negative prediction were accessed by calling score. The probability and sentiment were then added to the dataframe.

## MODEL 2

## MODEL 3

## TIME-SERIES ANALYSIS

# MODEL EVALUATION

## FLAIR

## MODEL 2

## MODEL 3

## TIME-SERIES ANALYSIS

# CONCLUSIONS

# REFERENCES

[1] “Psychology influences markets,” *California Institute of Technology*, 2013. https://www.sciencedaily.com/releases/2013/07/130701151608.htm (accessed Mar. 31, 2021).

[2] V. Sanh, “Smaller, faster, cheaper, lighter: Introducing DistilBERT, a distilled version of BERT,” *2019*. https://medium.com/huggingface/distilbert-8cf3380435b5 (accessed Apr. 01, 2021).