**DAB 401 (FINANACIAL ANALYTICS)**

**FINAL GROUP PROJECT REPORT.**

**GROUP MEMBERS:**

**BENCY ELDHO (0731915)**

**DARIN AMICHANDWALA (0734860)**

**DONGMEI MO (0733239)**

**MEHAK RAJDEV (0732638)**

**NANBAL AKILA FOMPUN (0714107)**

**SOURAV THAKUR (0733904)**

**TOPIC: PREDICTION OF STOCK PRICE MOVEMENT FROM NEWS HEADLINES USING NATURAL LANGUAGE PROCESSING.**

**INTRODUCTION**

In recent times we have discovered that media outlets such as news headlines and social media outlets have played a vital role on the movement of stock prices in the market. However, there were challenges as these were done manually. Some of these challenges with the manual systems of prediction include poor accuracy, high time consumption and poor productivity on the part of the staff. The need to automate this process became increasingly apparent and in this project, we will be applying a machine learning technique known as Natural Language Processing (NLP) to predict stock price movements for the next day using a Recurrent Neural Network (RNN) particularly Long-Short Term Memory (LSTM) as the model for prediction.

**DATA SOURCES**

We obtained our data from two sources. They are www.seekingalpha.com and Yahoo Finance where we obtained headline data and stock price data. The two datasets were merged into one dataset which had 1185 rows and 3 columns.

**DATA CLEANING**

The procedures below were followed when cleaning our data;

* Removal of special characters from text data using the Regular expression library in python.
* Creating a new column for date using the datetime library in python.
* Imputation using the mean method where there were missing values in the Price column.

**EXPLORATORY DATA ANALYSIS**

In exploring our data, we carried out some manipulation using Natural Language Processing libraries and basic arithmetic operations. These analyses include;

* Calculating the polarity scores of each sentiment with the Sentiment Intensity Analyzer (SIA) in python which assigns a score to each sentiment and calculates a compound.
* Calculating daily returns and merge with the closing rice column in a new data frame.
* Creating a new column for score by shifting by one since we want to predict stock price movements for the next day.
* Merging the new Date, Score and the Returns column.
* We carried out a correlation analysis between the Returns and the Score column and discovered we had a weak negative score of 0.1154 which suggests that we are not likely to get high accuracy if we build a model based on the data and the correlation is inverse.

**DATA PREPROCESSING**

Data Preprocessing is basically the use of statistical techniques to prepare data for modelling. Some of the statistical techniques we applied include;

* Shifting the price and compound columns by one and storing them as new columns in the data frame as we are trying to predict for the next day.
* Separating our data into target and feature with price\_shifted column being the target. (y) and the price, comp and comp\_shifted the feature variables (X).
* Scale our feature data using the MinMax scaler.
* We split our data into training and test sets.

**MODEL BUILDING AND EVALUATION**

In building the model, we selected the Long Short Term memory due to its high predictive power for time series problems. It was applied to both the training and the test sets. The Mean Squared Error was our chosen metric for model evaluation and our results are as shown below;

**MSE on Training set: 0.0092**

**MSE on Test set: 0.0137**

**A picture containing clock

Description automatically generated**

**CONCLUSION**

In this project we have been able to establish that Natural Language Processing is a useful and powerful tool in the prediction of stock price movements and eases the burden of wasting man hours.

Recurrent Neural Networks particularly LSTM provide high accuracy as shown by the evaluation metric we used (MSE) and the visualization of our results.

**REFERENCES**

[**https://ca.finance.yahoo.com/**](https://ca.finance.yahoo.com/)

[**https://seekingalpha.com/symbol/GM?s=gm**](https://seekingalpha.com/symbol/GM?s=gm)

[**https://algotrading101.com/learn/sentiment-analysis-with-python-finance/**](https://algotrading101.com/learn/sentiment-analysis-with-python-finance/)

[**http://jonathankinlay.com/2019/10/trading-market-sentiment/**](http://jonathankinlay.com/2019/10/trading-market-sentiment/)

[**https://www.quantnews.com/exploring-trader-sentiment-data-python-pandas/**](https://www.quantnews.com/exploring-trader-sentiment-data-python-pandas/)

**.**