

Integrating Sentiment Analysis into AI-
Driven Budgeting:
Advancing Personal Financial Management
through
Emotion-Aware Financial Insights

Natural Language Processing - Sentiment Analysis

MSc Research Project

Artificial Intelligence

Ravi Kumar Sabbavarapu

Student ID: 23301953

School of Computing

National College of Ireland

Supervisor: Lavish Thomas

National College of Ireland
MSc Project Submission Sheet
School of Computing

Student Name: Ravi Kumar Sabbavarapu

Student ID: 23301953

Program me: Artificial Intelligence **Year:** 2025

Module: Msc. Research Practicum

Lecturer: Lavish Thomas
Submission Due Date: 15-Sep-2025

Project Title: Integrating Sentiment Analysis into AI-Driven Budgeting:
Advancing Personal Financial Management through
Emotion-Aware Financial Insights
Natural Language Processing - Sentiment Analysis

Word Count: **Page Count:** 07

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: S Ravi Kumar

Date: 11-Aug-2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	☐
---	--------------------------

Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only		
Signature:		
Date:		
Penalty Applied (if applicable):		

Integrating Sentiment Analysis into AI-Driven Budgeting: Advancing Personal Financial Management through Emotion-Aware Financial Insights

Natural Language Processing - Sentiment Analysis

Ravi Kumar Sabbavarapu
23301953

1 Introduction

The study aims to incorporate sentiment analysis models into AI-powered personal budgeting systems using innovative ML, DL, and LSTM methods to track real-time investor opinions from finance texts and social media platforms.

2 System Specifications

The system specification contains all the resources used in order to complete the research study. Figure 1 shows the local system used to run this project, while Figure 2 (A) shows all the run time utilized for research. Figure 2 (B) shows detailed information about memory with detailed specification limits. All hardware accelerators have been utilized for research and need to be used with caution due to usage limits.

Hardware:

Hardware Overview:

Model Name: MacBook Air
Model Identifier: MacBookAir10,1
Model Number: MGN63HN/A
Chip: Apple M1
Total Number of Cores: 8 (4 performance and 4 efficiency)
Memory: 8 GB
System Firmware Version: 11881.140.96

OS Loader Version: 11881.140.96
Serial Number (system): HXJMNW021WFV
Hardware UUID: 574142E7-E874-52F3-80F5-75F134B9437D
Provisioning UDID: 00008103-000C58210E83A01E
Activation Lock Status: Disabled

(base) ravikumarsabbavarapu@Ravis-MacBook-Air ~ %

Figure 1. System specifications

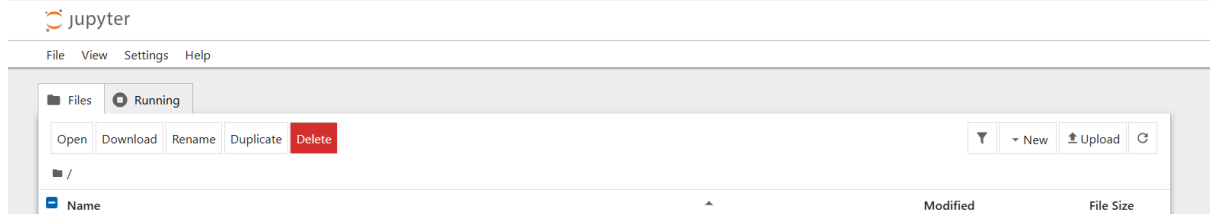


Figure 2 (A): Jupyter Notebook RunTime Types

Software Used:

- Microsoft excel: Used for custom dataset description.
- Jupyter Notebook: Used for all processing and as code runtime environment
- Google Drive: Used to store models, datasets and handle while runtime

3 Dataset specification

The dataset is derived from a publicly available authentic platform, named Kaggle, which is consistent with the research aim of incorporating sentiment analysis into AI-budgeting. The dataset consists of 11,932 finance tweets, which are classified into three sentiment classes such as Bearish (LABEL_0), Bullish (LABEL_1), and Neutral (LABEL_2) (Kaggle, 2024). These affective labels serve as the target variable in this study, reflecting the user's emotional reaction towards financial news, hence leading to appropriate investment and budgetary decisions.

4. Project Development

After collecting all data and storing it on drive or storage place, Colab notebook can be launched.

4.1 Importing Library:

It depicts the bottom-line configuration of introducing sentiment analysis to artificial intelligence-based budgeting systems. The figure shows how crucial libraries (TensorFlow, sklearn, pandas, and SMOTE) need to be imported to accomplish relevant preprocessing data and model formulation, as well as text analysis.

```

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from wordcloud import WordCloud
from collections import Counter
from imblearn.over_sampling import SMOTE
import re
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier, AdaBoostClassifier
from sklearn.metrics import classification_report, accuracy_score, precision_score, recall_score, f1_score, confusion_matrix
from tensorflow.keras.layers import Dense, Embedding, SpatialDropout1D, GlobalAveragePooling1D
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, LSTM, Embedding, SpatialDropout1D
from tensorflow.keras.layers import Embedding, SpatialDropout1D, GlobalAveragePooling1D, Dense
from tensorflow.keras.preprocessing.text import Tokenizer
from tensorflow.keras.preprocessing.sequence import pad_sequences

```

Figure 3: Packages used in Project

4.2 Importing Files

This step makes machine learning outputs dependable by checking the data structure, shape, and content. This data, labelled with such sentiment, is critical to constructing the emotion-sensitive financial model to have a more natural and responsive personal budgeting tool according to the sentiment.

```
train = pd.read_csv('sent_train.csv')
```

```
valid = pd.read_csv('sent_valid.csv')
```

Figure 4: Importing Files

4.3 Preprocessing and cleaning

Checking and managing null and duplicate values helps to ensure data consistency and enhance predictive accuracy (Ali, Emran and Asmai, 2021). It also verifies that there is no Null and duplication in the training and validating datasets, making the data dependable and complete.

```
## Checking Null Values in train data
```

```
train.isnull().sum()
```

```
text      0  
label     0  
dtype: int64
```

```
## Checking Null Values in valid data
```

```
valid.isnull().sum()
```

```
text      0  
label     0  
dtype: int64
```

```
## Checking Duplicated Values in train data
```

```
train.duplicated().sum()
```

```
np.int64(0)
```

```
## Checking Duplicated Values in valid data
```

```
valid.duplicated().sum()
```

```
np.int64(0)
```

```
## Cleaning Text in train data
```

```
### Enhanced Text Cleaning Function
```

```
def clean_text(text):  
    text = str(text)  
    text = text.lower()  
    text = re.sub(r'http\S+|www.\S+', '', text)  
    text = re.sub(r'^a-z0-9\s', '', text)  
    text = re.sub(r'\s+', ' ', text)  
    text = text.strip()  
    return text
```

```
# Apply cleaning
```

```
train['clean_text'] = train['text'].apply(clean_text)
```

```
# Preview cleaned data
```

```
train[['text', 'clean_text']].head()
```

	text	clean_text
0	\$BYND - JPMorgan reels in expectations on Beyo...	bynd jpmorgan reels in expectations on beyond ...
1	CCLRCL - Nomura points to bookings weakness...	ccl rcl nomura points to bookings weakness at ...
2	\$CX - Cemex cut at Credit Suisse, J.P. Morgan ...	cx cemex cut at credit suisse jp morgan on wea...
3	\$ESS: BTIG Research cuts to Neutral https://t...	ess btig research cuts to neutral
4	\$FNKO - Funko slides after Piper Jaffray PT cu...	fnko funko slides after piper jaffray pt cut

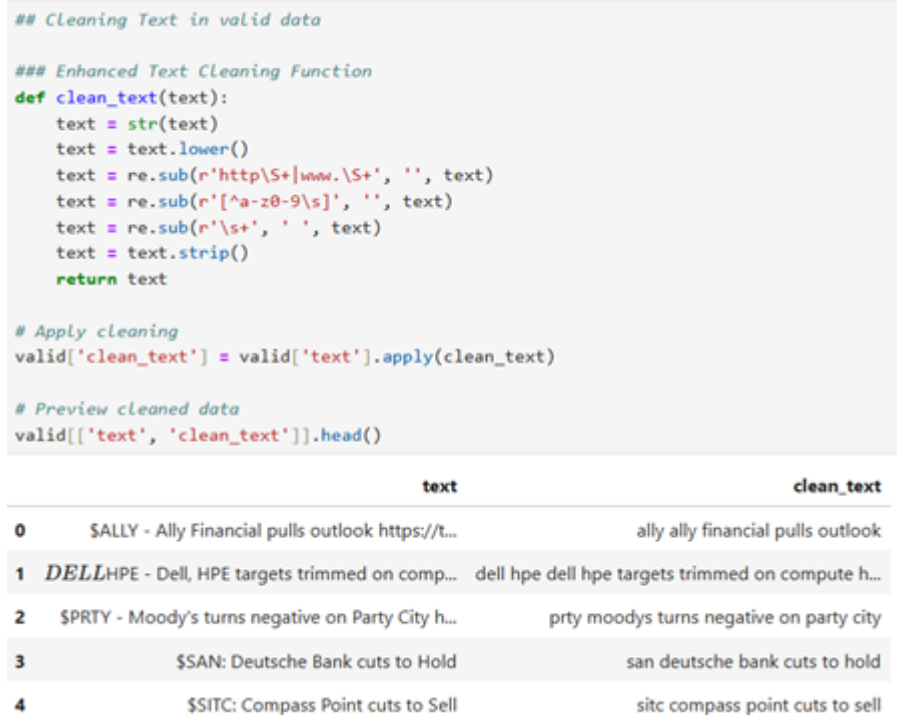
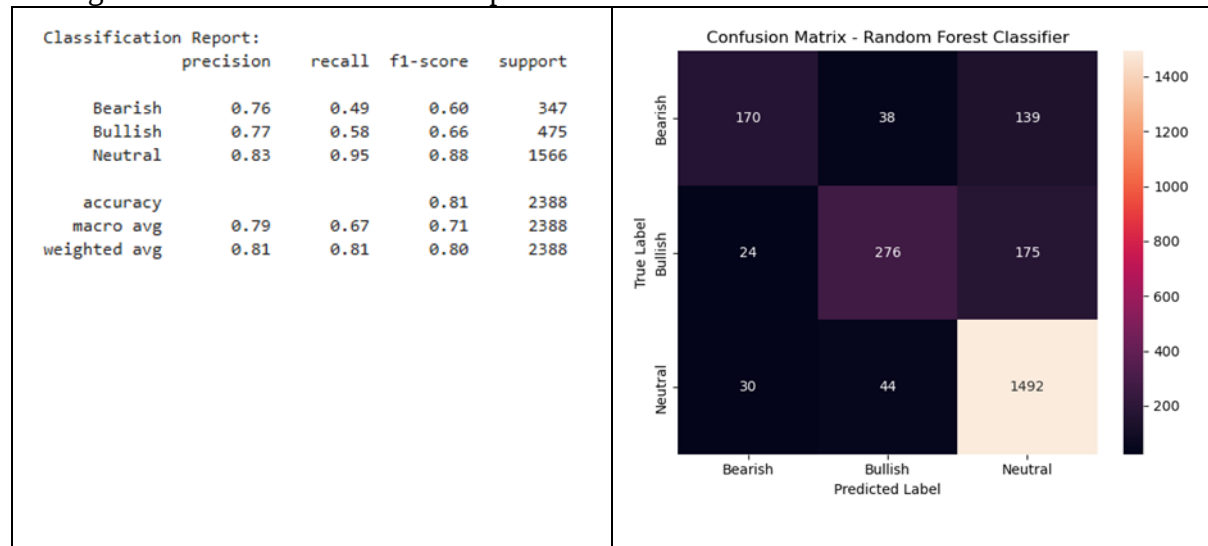


Figure 5: Data Preprocessing

This preprocessing plays a key role in the accurate extraction of the sentiment behind the financial news headlines. Clean text builds a solid background for ML models, which enhances the accuracy of sentiment-based budgeting tools in personal fiscal management systems.

4.4 Modelling:

The figure shows the classification report and confusion of all four trained models.



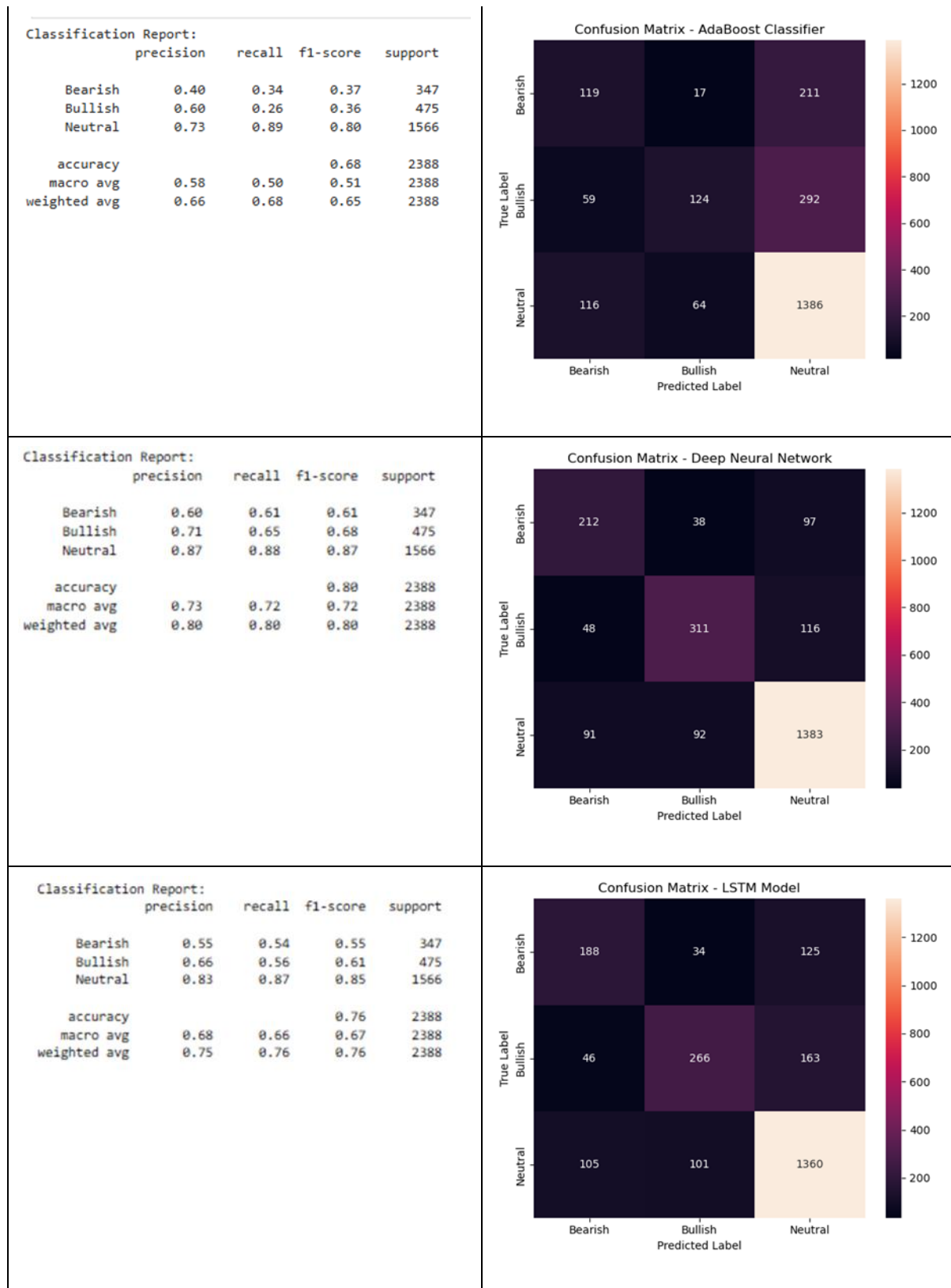


Figure 6: Model Evaluation

Across all four classifiers, the neutral class is the easiest to detect based on the highest recall and precision, while bearish and bullish sentiments prove more challenging to identify

accurately. False positives occur when the model predicts a sentiment that is not present, for example, calling a neutral tweet bearish. False negatives occur when the model fails to detect a sentiment that is there, for example, missing a bullish signal.

References

Kaggle (2024). *Twitter Financial News Sentiment Dataset*. [online] Kaggle.com. Available at: https://www.kaggle.com/datasets/borhanitrash/twitter-financial-news-sentiment-dataset/data?select=sent_dataset_meta.txt [Accessed 6 Jun. 2025].

Ali, A., Emran, N.A. and Asmai, S.A. (2021). Missing values compensation in duplicates detection using hot deck method. *Journal of Big Data*, [online] 8(1), pp.1–19. doi:<https://doi.org/10.1186/s40537-021-00502-1>.