MULTIMEDIA UNIVERSITY

FACULTY OF COMPUTING AND INFORMATICS

BACHELOR IN COMPUTER SCIENCE

SOCIAL MEDIA COMPUTING - CDS6344

TRIMESTER 2, SESSION 2024/2025

Sentiment Analysis of Twitter's Tweet Using Transformer Models

No.	Student Name	Student ID	Tutorial
1.	Muhammad Rino Frawidya Bin Suheri	1211103423	TT2L
2.	Muhammad Afif Jazimin Bin Idris	1211103419	TT2L

Table of Contents

1. Introduction

Social media platforms have emerged as the primary channels for people to share their ideas, opinions, and feelings on a variety of subjects in the age of digital communication. Particularly, Twitter has become a rich source of unstructured textual data in real time that can provide insightful information about trends in public sentiment. The need for efficient Natural Language Processing (NLP) methods to automatically extract, evaluate, and interpret sentiments within textual content has increased due to the exponential growth of such data.

In order to process a large corpus of tweets, classify sentiments, and carry out aspect-based sentiment analysis (ABSA), this project focuses on creating an end-to-end natural language processing (NLP) pipeline. This project attempts to transform unstructured textual data into useful insight by utilizing traditional machine learning models, deep learning models, and transformer-based models NLP techniques, such as sentiment classification, text preprocessing, and visualisation, are incorporated into the project. It aims not only to classify text into categories like positive, negative, or neutral but also to provide a more detailed understanding of the data by exploring deeper into the underlying aspects that influence these sentiments. .

1.1 Project Overview

In order to categorize tweets into positive, negative, or neutral sentiments, this project focuses on sentiment analysis of Twitter data. The Kaggle dataset includes thousands of real-world tweets with sentiment labels attached. To ensure that the text data is clean and appropriate for modelling, the analysis starts with thorough data preprocessing, which includes normalization, tokenization, stopword removal, and lemmatization.

To extract meaningful features from the tweets, both traditional and modern approaches are employed. Classical feature engineering uses TF-IDF to convert text into numerical vectors, serving as a baseline for comparison. The core innovation of the project lies in leveraging transformer-based models, specifically DistilBERT, to generate dense, contextual embeddings that capture the semantic and syntactic nuances of each tweet.

Using these features, the project moves forward with training and evaluating sentiment classification models. Traditional machine learning models are compared with transformer-based approaches to assess their effectiveness. Evaluation metrics such as accuracy, precision, recall, and F1-score are used to measure model performance. The results highlight the advantages of transformer models in understanding context and improving sentiment classification accuracy, especially in the informal and diverse language found on social media platforms.

2. Problem Statement

Due to the rapid growth of social media sites like Twitter, enormous volumes of user-generated textual data are produced daily, reflecting real-time reactions to events, emotions, and public opinion. However, programmed sentiment analysis faces significant challenges due to tweets' informal, brief, and frequently ambiguous nature. The complex context and semantics present in social media language are frequently missed by traditional machine learning techniques, which rely on straightforward word frequency or bag-of-words representations.

Creating a reliable and efficient sentiment analysis pipeline that can reliably classify tweets into positive, negative, or neutral categories is the main issue this project attempts to solve. This involves overcoming challenges like misspellings, contextual ambiguity, slang, and abbreviations. This study aims to investigate whether modern transformer-based models, specifically DistilBERT, can outperform traditional methods by leveraging deep contextual embeddings to better understand the sentiment expressed in short, informal texts. The ultimate objective is to enable applications in social research, public opinion tracking, and business intelligence by offering a dependable tool for deriving actionable sentiment insights from massive Twitter data.

3. Literature Review

With each generation addressing some of the weaknesses of the earlier methods, sentiment analysis has undergone significant development, progressing from rule-based systems to modern deep learning and transformer-based models.

3.1 Early Rule-based Methods

Lexicon-based methods, which used pre-made lists of positive and negative words to determine sentiment polarity, were the first sentiment analysis systems. The study by Taboada et al. (2011) is an interesting example, as it showed how well lexicon-based approaches worked for text reviews, but it also pointed out how poorly they handled context, sarcasm, and domain-specific expressions.

In research by Hutto and Gilbert (2014), they introduced VADER (Valence Aware Dictionary and Sentiment Reasoner), a rule-based sentiment analysis tool made especially for text on social media. VADER's handling of slang, emojis, and punctuation allowed it to achieve higher accuracy on microblogging sites such as Twitter [2].

3.2 Traditional Machine Learning Models

The research community shifted to supervised machine learning models like Naïve Bayes, Logistic Regression, and Support Vector Machines (SVM) as the disadvantages of rule-based approaches became apparent. These classifiers were compared on a sentiment-labeled movie review dataset in a research by Pang et al. (2002), which showed that data-driven learning could enable machine learning models to outperform rule-based approaches [3]. Using methods like Bag of Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF), these models convert text data into feature vectors, providing flexible and scalable sentiment analysis solutions.

3.3 Deep Learning Models

By allowing models to recognise complex patterns and sequential dependencies in text, deep learning further developed the field of study. Zhang and Wallace (2015) illustrated how Convolutional Neural Networks (CNNs) can be used for sentence classification, emphasising how they can identify local text patterns without manual feature engineering [4].

Furthermore, Hochreiter and Schmidhuber (1997) introduced Recurrent Neural Networks (RNNs) and, in particular, Long Short-Term Memory (LSTM) networks to overcome the weaknesses of traditional RNNs in identifying long-term dependencies within sequences [5]. Because Bi-directional LSTM (BiLSTM) models could process sequences both forward and backward, they gained popularity and improved the accuracy of sentiment classification.

3.4 Transformer-Based Models

The development of transformer-based models reflected a significant advancement in NLP. By extracting bidirectional context from large data sets and fine-tuning for specific downstream tasks like sentiment analysis, Devlin et al. (2019) proposed BERT (Bidirectional Encoder Representations from Transformers), which transformed natural language processing [6]. In their study, BERT outperformed 11 NLP benchmarks, including the GLUE benchmark (General Language Understanding Evaluation) and SST-2 (Stanford Sentiment Treebank) sentiment classification task.

RoBERTa (Robustly Optimised BERT Pretraining Approach), developed by Liu et al. (2019), added more optimisations to build on the success of BERT. By eliminating the next sentence prediction objective, dynamically altering the masking pattern during training, and using ten times as much data as BERT, RoBERTa enhanced model stability and performance [7]. RoBERTa outperformed BERT in their evaluations on a number of benchmarks.

3.5 Aspect-Based Sentiment Analysis (ABSA)

Aspect-Based Sentiment Analysis (ABSA) provides more detailed sentiment assessments by combining opinions to particular features of a product or subject, whereas overall sentiment analysis provides insightful information. Through the SemEval task, Pontiki et al. (2016) developed a standardised framework for evaluating ABSA, which encouraged study of aspect-level sentiment analysis for structured data sets such as laptop and restaurant reviews [8].

Aspect tagging could be directly integrated into transformer models such as BERT using Position-Aware Tagging for ABSA tasks, as suggested by Xu et al. (2020), who achieved significant improvements over traditional BiLSTM-CRF approaches [9]. This method is ideal for short, informal texts like tweets because it allows the extraction of both aspect terms and associated sentiment polarities in a single pipeline.

4. Methodology

4.1 Data Collection and Exploration

The dataset is downloaded from Kaggle using the kagglehub library, specifically the "yasserh/twitter-tweets-sentiment-dataset" which contains thousands of tweets labeled with sentiments (positive, negative, neutral). The dataset includes columns such as textID, text, selected_text, and sentiment. Data exploration involves loading the dataset into a Pandas DataFrame and inspecting the first few rows to understand the structure of the dataset.

4.2 Data Preprocessing

Data preprocessing starts with dropping null value rows and duplicated rows. Next, the selected_text column is normalized by converting all text to lowercase, removing URLs, special characters except letters, numbers, spaces, and extra whitespace. Tokenization is performed and stopwords are removed based on the NLTK stopword list. Lemmatization is then applied using NLTK's WordNetLemmatizer to reduce words to their base forms. Before moving into the next process, the dataset is cleaned of any missing values, since some of the rows turn null after undergoing data preprocessing.

4.3 Feature Engineering

In this study, TF-IDF Vectorization is employed for future engineering. It is one of the most popular feature extraction techniques in NLP. The preprocessed selected_text are transformed from raw, unstructured text data into numerical vectors using Term Frequency-Inverse Document Frequency (TF-IDF), capturing word importance. These vectors served as input features for traditional machine learning models, which also captures both the term frequency and how unique or rare the term is.

4.4 Data Sampling and Class Balancing

A balanced dataset is created by sampling an equal number of records of 3000, with 1000 for each sentiment category (positive, neutral, negative) to prevent class imbalance, ensuring the models learn fairly from each class and reducing bias towards the majority class. The numerical vectors of TF-IDF are aligned with the sampled, balanced dataset to ensure that they correspond to the correct tweets and their sentiment labels.

4.5 Model Development

The sampled, balanced dataset is then split into a training set of 0.8 and a testing set of 0.2. Four models are then developed;

- **Logistic Regression Classifier:** This model is trained using the TF-IDF feature vectors. Hyperparameters such as the regularization strength (C), solver type (lbfgs, saga), and maximum iteration limit are optimized using grid search with three-fold cross-validation.
- **Support Vector Machine (SVM) Classifier:** This model is trained using the TF-IDF feature vectors. Hyperparameters including the regularization strength (C), kernel type, and gamma setting ARE tuned using grid search with three-fold cross-validation.
- **Random Forest Classifier:** This model is trained using the TF-IDF feature vectors. Hyperparameters including the number of estimators, maximum tree depth, minimum samples required to split a node, and minimum samples required at a leaf node are tuned using grid search with three-fold cross-validation.
- **DistilBERT Transformer Classifier:** This model is trained using raw tweet selected_text, which is first tokenized using a pre-trained DistilBERT tokenizer. The DistilBertForSequenceClassification model with three output labels is fine-tuned on this dataset using Hugging Face's Trainer API. Training hyperparameters such as batch size, number of epochs, weight decay, and logging strategy are specified through a training arguments configuration.

4.6 Model Evaluation

After conducting model development, model evaluation of four models are executed and evaluated using classification report and confusion matrix. The accuracy of all models are also compared using a bar chart.

5. Sentiment Analysis

The sentiment analysis process in this study is designed to classify user-generated tweets into positive, negative, or neutral categories. The workflow begins with the acquisition of a labeled dataset containing thousands of tweets, each annotated with a sentiment class. To ensure the quality and consistency of the data, a comprehensive preprocessing pipeline is implemented. This pipeline included normalization of text to lowercase, removal of extraneous characters and URLs using regular expressions, and tokenization of the text into individual words. Stopwords, which removes commonly used words that do not contribute significant meaning, is also implemented to reduce noise in the data. Lemmatization is then applied to convert words to their base or dictionary forms, standardizing the vocabulary and improving the robustness of subsequent analysis.

Following preprocessing, the cleaned tweets are ready for feature extraction using TF-IDF for traditional machine learning models and modeling. The sentiment labels provide a supervised learning framework, enabling the training and evaluation of classification models. A balanced dataset is created by sampling an equal number of records of 3000, with 1000 for each sentiment category (positive, neutral, negative) to prevent class imbalance, ensuring the models learn fairly from each class and reducing bias towards the majority class. The numerical vectors of TF-IDF are aligned with the sampled, balanced dataset to ensure that they correspond to the correct tweets and their sentiment labels. This systematic approach provides a strong foundation for building effective sentiment classification models, allowing for both traditional and deep learning techniques to be applied and compared.

6. Transformer Model

The DistilBERT model is used in a transformer-based method to capture the complex conceptual and contextual information present in natural language. DistilBERT, a distilled version of BERT, offers computational efficiency and representational power, making it well-suited for large-scale text analysis tasks. The process begins by loading a pre-trained DistilBERT tokenizer and model from the Hugging Face Transformers library. Each preprocessed tweet is passed through the tokenizer, which converts the raw text into a sequence of tokens, applying truncation and padding to ensure a uniform input length compatible with the model's requirements.

Once tokenized, the text data is fed into the DistilBERT model in inference mode, which does not compute gradients and maintains the model weights. After processing each input sequence, the model generates a set of hidden state vectors for each input token. The embedding corresponding to the first token is taken out of the last hidden layer to get a fixed-size representation of every tweet.

Every tweet in the dataset is subjected to this embedding extraction procedure methodically, producing a set of high-dimensional vectors. After that, these vectors are arranged into a feature matrix, which serves as the input for classification models that come after. The model surpasses the capabilities of traditional feature extraction techniques like bag-of-words or TF-IDF by utilizing the deep contextual understanding , offered by DistilBERT to better classify sentiment and capture hidden linguistic cues.

7. Result & Visualization

For this section, findings of this project are shown and visualized, from exploratory data analysis, sentiment labeling, and model evaluation.

7.1 Findings from Model Evaluation

Classification report and actual sentiment with predicted sentiment are shown in Figure 1 for Logistic Regression model, Figure 2 for SVM model, and Figure 3 for Random Forest model. Figure 4 shows classification report for DistilBERT model.

```
Logistic Regression Classification Report
              precision    recall  f1-score   support

    Negative       0.71      0.73      0.72       200
     Neutral       0.68      0.70      0.69       200
    Positive       0.81      0.76      0.78       200

    accuracy                           0.73       600
   macro avg       0.73      0.73      0.73       600
weighted avg       0.73      0.73      0.73       600

+----+------------------------------------------------------------------------------------------------------------------+-------------------+----------------------+
|    | Tweet                                                                                                            |  Actual Sentiment |   Predicted Sentiment |
|----+------------------------------------------------------------------------------------------------------------------+-------------------+----------------------|
|  0 | LOL I love my MacBook too. Oh and my iMac. Can`t decide which I love more. OK 24' iMac trumps 13' MacBook      |                 2 |                     2 |
|  1 | thanks  i have to finish schoolwork today, no rehearsal tonight though. what ru doing?                          |                 2 |                     2 |
|  2 | _Laurie my thoughts rae with you and your family                                                                 |                 1 |                     1 |
|  3 | I Miss Daddy and Mommy                                                                                           |                 0 |                     0 |
|  4 | he so is! <3 hence, my new forum signature!                                                                      |                 2 |                     0 |
|  5 | excited to see my cousins this week.                                                                             |                 2 |                     2 |
|  6 | G*morning! Rain, rain and more rain.. ! But I don`t care so much                                                 |                 0 |                     0 |
|  7 | Hey! How u feeling? I know Charla hit me to see there were any places she can go..too young  Did she have fun? |                 1 |                     1 |
|  8 | mum`s day - ended up being happy  not that it`s my day or anything....                                           |                 2 |                     2 |
...
| 17 | A Happy Mother`s Day to all moms and soon to be mom`s out there.                                                 |                 2 |                     2 |
| 18 | PRD take a long time to review!                                                                                  |                 1 |                     1 |
| 19 | _R_Roberts I LOOK forward to seeing and sharing                                                                  |                 2 |                     0 |
+----+------------------------------------------------------------------------------------------------------------------+-------------------+----------------------+
```

Figure 1: Classification report for Logistic Regression Model

```
SVM Classification Report
              precision    recall  f1-score   support

    Negative       0.72      0.73      0.72       200
     Neutral       0.69      0.70      0.70       200
    Positive       0.81      0.78      0.79       200

    accuracy                           0.74       600
   macro avg       0.74      0.74      0.74       600
weighted avg       0.74      0.74      0.74       600

+----+------------------------------------------------------------------------------------------------------------------+-------------------+----------------------+
|    | Tweet                                                                                                            |  Actual Sentiment |   Predicted Sentiment |
|----+------------------------------------------------------------------------------------------------------------------+-------------------+----------------------|
|  0 | LOL I love my MacBook too. Oh and my iMac. Can`t decide which I love more. OK 24' iMac trumps 13' MacBook      |                 2 |                     2 |
|  1 | thanks  i have to finish schoolwork today, no rehearsal tonight though. what ru doing?                          |                 2 |                     2 |
|  2 | _Laurie my thoughts rae with you and your family                                                                 |                 1 |                     1 |
|  3 | I Miss Daddy and Mommy                                                                                           |                 0 |                     0 |
|  4 | he so is! <3 hence, my new forum signature!                                                                      |                 2 |                     0 |
|  5 | excited to see my cousins this week.                                                                             |                 2 |                     2 |
|  6 | G*morning! Rain, rain and more rain.. ! But I don`t care so much                                                 |                 0 |                     1 |
|  7 | Hey! How u feeling? I know Charla hit me to see there were any places she can go..too young  Did she have fun? |                 1 |                     1 |
|  8 | mum`s day - ended up being happy  not that it`s my day or anything....                                           |                 2 |                     2 |
...
| 17 | A Happy Mother`s Day to all moms and soon to be mom`s out there.                                                 |                 2 |                     2 |
| 18 | PRD take a long time to review!                                                                                  |                 1 |                     1 |
| 19 | _R_Roberts I LOOK forward to seeing and sharing                                                                  |                 2 |                     2 |
+----+------------------------------------------------------------------------------------------------------------------+-------------------+----------------------+
```

Figure 2: Classification report for SVM Model

```
Random Forest Classification Report
              precision    recall  f1-score   support

    Negative       0.69      0.79      0.74       200
     Neutral       0.73      0.70      0.72       200
    Positive       0.83      0.75      0.79       200

    accuracy                           0.75       600
   macro avg       0.75      0.75      0.75       600
weighted avg       0.75      0.75      0.75       600

+----+-------------------------------------------------------------------------------------------------------+------------------+---------------------+
|    | Tweet                                                                                                 |  Actual Sentiment |  Predicted Sentiment |
|----+-------------------------------------------------------------------------------------------------------+------------------+---------------------|
|  0 | LOL I love my MacBook too. Oh and my iMac. Can`t decide which I love more. OK 24' iMac trumps 13' MacBook     |                2 |                   2 |
|  1 | thanks  i have to finish schoolwork today, no rehearsal tonight though. what ru doing?              |                2 |                   2 |
|  2 | _Laurie my thoughts rae with you and your family                                                      |                1 |                   0 |
|  3 | I Miss Daddy and Mommy                                                                                |                0 |                   0 |
|  4 | he so is! <3 hence, my new forum signature!                                                           |                2 |                   0 |
|  5 | excited to see my cousins this week.                                                                  |                2 |                   2 |
|  6 | G*morning! Rain, rain and more rain.. ! But I don`t care so much                                      |                0 |                   0 |
|  7 | Hey! How u feeling? I know Charla hit me to see there were any places she can go..too young  Did she have fun? |                1 |                   1 |
|  8 | mum`s day - ended up being happy  not that it`s my day or anything....                                |                2 |                   2 |
...
| 17 | A Happy Mother`s Day to all moms and soon to be mom`s out there.                                      |                2 |                   2 |
| 18 | PRD take a long time to review!                                                                       |                1 |                   1 |
| 19 | _R_Roberts I LOOK forward to seeing and sharing                                                       |                2 |                   2 |
+----+-------------------------------------------------------------------------------------------------------+------------------+---------------------+
```

Figure 3: Classification report for Random Forest Model

```
DistilBERT Classification Report
              precision    recall  f1-score   support

    Negative       0.89      0.91      0.90       169
     Neutral       0.87      0.86      0.87       237
    Positive       0.90      0.89      0.90       194

    accuracy                           0.89       600
   macro avg       0.89      0.89      0.89       600
weighted avg       0.89      0.89      0.88       600
```

Figure 4: Classification report for DistilBERT Model

Based on the classification report of all models developed, their accuracy is compared using a bar chart. Figure 5 shows model accuracy comparison.

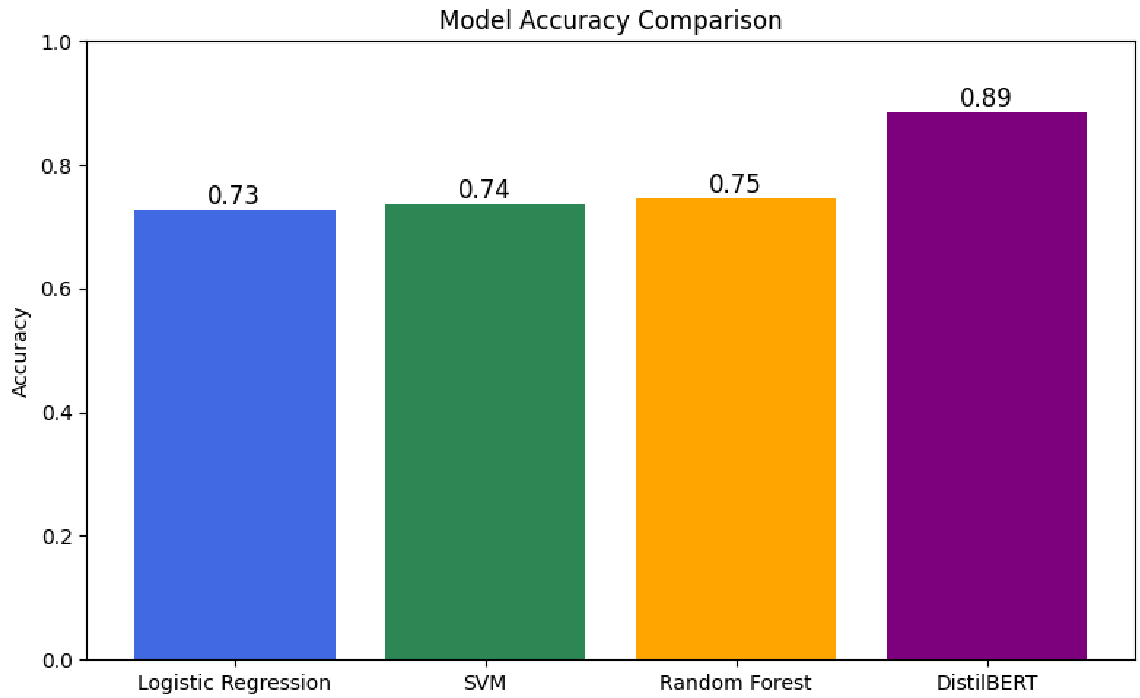

Figure 5: Model Accuracy Comparison

Based on Figure 5, the DistilBert model has outperformed all models developed in terms of accuracy with 0.89 accuracy. Random Forest has 0.75 accuracy, following SVM model with 0.74 accuracy, and Logistic Regression model with 0.73 accuracy. The confusion matrix of all models are also shown in Figure 6, Figure 7, Figure 8, and Figure 9.

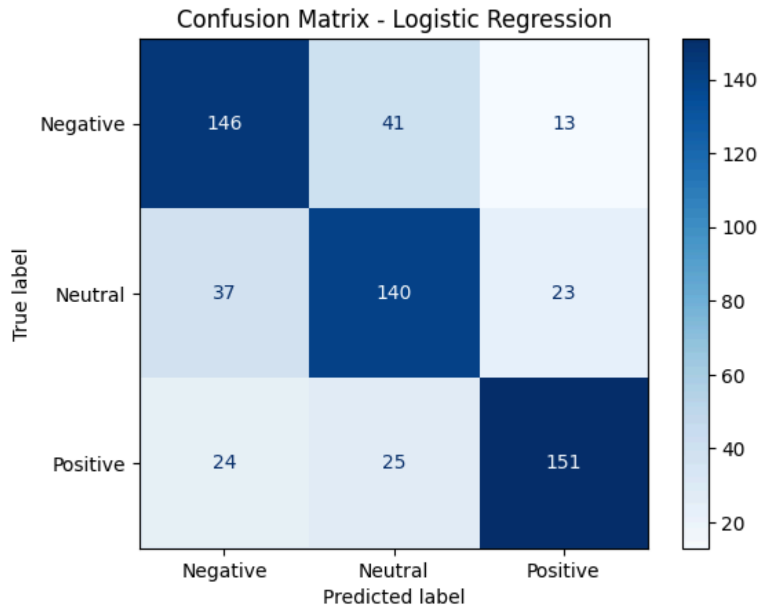

Figure 6: Confusion Matrix for Logistic Regression Model

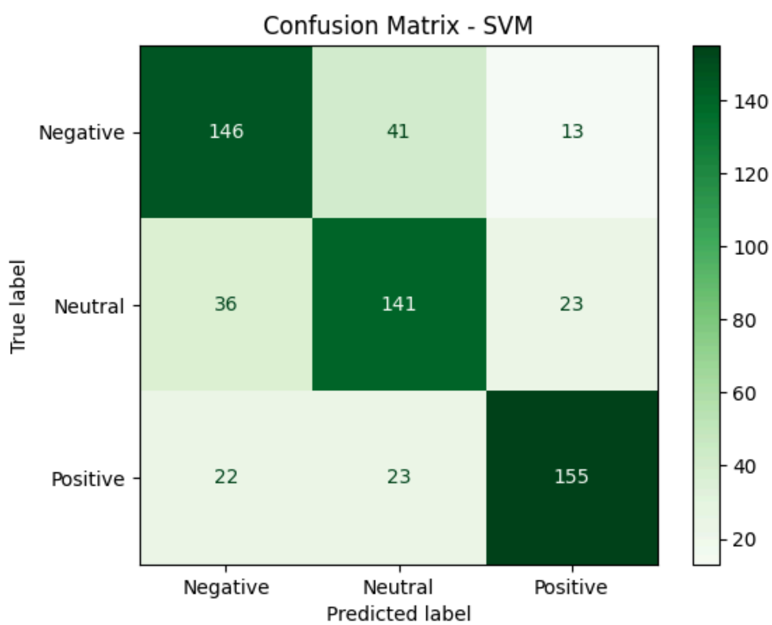

Figure 7: Confusion Matrix for SVM Model

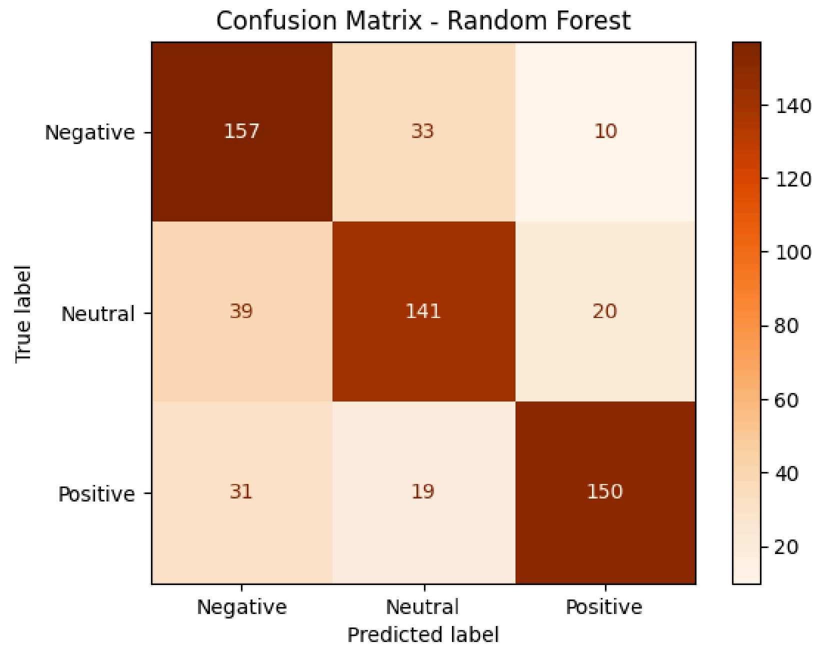

Figure 8: Confusion Matrix for Random Forest Model

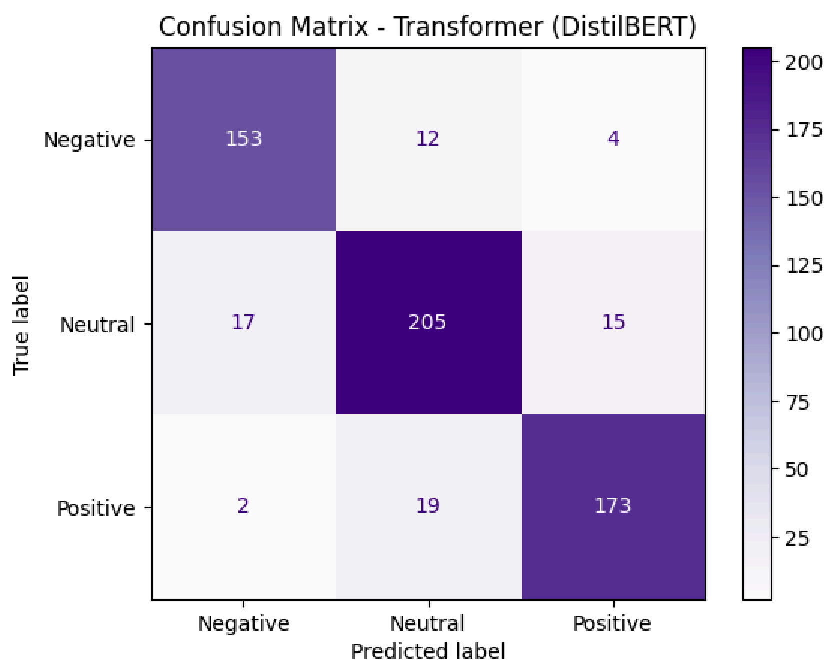

Figure 9: Confusion Matrix for DistilBERT Model

The Confusion matrix of DistilBERT model shows the highest true positive and true neutral values compared to other models, which indicates strong overall accuracy. Although the true negative value of DistilBERT model is slightly lower than Random Forest model, the superior performance on the other two categories contributes to DistilBERT's higher overall accuracy in sentiment classification.

7.2 Opinion Mining

7.2.1 Aspect-Opinion Pair Extraction

This process is used to find pairs of words in tweets where a specific aspect (e.g. camera, service, food) is linked to an opinion word (e.g. good, bad, hate, love). Using spaCy's dependency parser, the system checked two types of grammar patterns, which are when an adjective describes a noun (e.g. "great phone"), and when a verb expresses a feeling about a noun (e.g. "hate service").

A total of 200 tweet texts are sampled, and the detected aspect-opinion pairs are stored in a table along with their relation type and original tweet sentence. This step helps to understand not just overall sentiment, but what people feel about specific topics mentioned in the tweets. Figure 10 shows the output of aspect-opinion pair extraction.

	aspect	opinion	relation	sentence
0	hobby	fabulous	amod	well well...so sorry! let`s dance it`s fabulo...
1	tickets	has	dobj	We want to trade with someone who has Houston ...
2	hole	leave	dobj	i know! i just got off the phone with them. it...
3	hole	Giant	amod	i know! i just got off the phone with them. it...
4	Regions	closed	nsubj	i know! i just got off the phone with them. it...
5	mothers	Happy	amod	Happy mothers day everybody
6	pals	have	dobj	hey, but you have gal pals here in Phoenix!!!
7	pals	gal	amod	hey, but you have gal pals here in Phoenix!!!
8	star	replies	nsubj	No star replies to me
9	Birthday	bowling	nsubj	GRANADA BOWL WITH THE CREW! Birthday bowling f...
10	bottle	broke	dobj	I just broke a bottle, I should probably focus...
11	songgoeswrongs	more	amod	I just broke a bottle, I should probably focus...
12	day	big	amod	big hair rock day today me thinks here I go a...
13	lol	wrong	amod	. Probably spelt it wrong lol.
14	soup	get	dobj	oh dear, hope you are feeling better soon, get...

Figure 10: Output of Aspect-Opinion Pair Extraction

7.2.2 Top 10 Most Common Opinion Words

The purpose of this insight is to identify and display the most frequently occurring opinion words extracted from the tweets. The system counts how often each opinion word appears in the previously extracted aspect-opinion pairs and selects the top 10 most common words. Figure 11 shows a horizontal bar chart of top 10 most common opinion words.

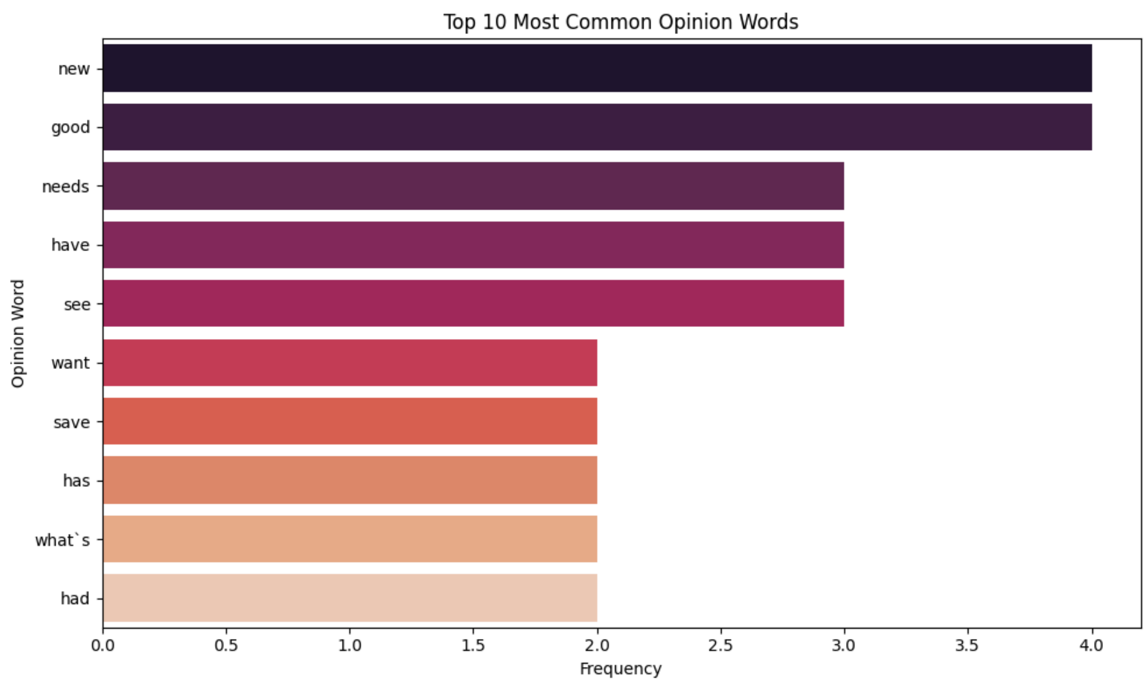

Figure 11: Top 10 Most Common Opinion Words

7.2.3 Opinion WordCloud

This process is used to create a word cloud of the opinion words extracted from the tweets. All the opinion words are combined into a single string, generating a word cloud to visually display the most common opinion words based on their frequency. In the word cloud, words that appear more frequently are shown in larger and bolder text. Figure 12 shows the opinion wordcloud.

Figure 12: Opinion WordCloud

7.3 Aspect-Based Sentiment Analysis (ABSA)

This process is used to identify specific aspects mentioned in tweets and determine the sentiment expressed towards them. A sample of 200 tweet texts is analyzed using spaCy's dependency parser to detect pairs where an adjective describes a noun or a verb relates to a noun. At the same time, the overall sentiment of each tweet is measured using the VADER sentiment analyzer. The extracted aspect, associated opinion word, their grammatical relation, and the overall sentiment label are stored in a table for review. Figure 13 shows the output of ABSA.

	sentence	aspect	opinion_word	relation	sentiment
0	well well...so sorry! let`s dance it`s fabulo...	hobby	fabulous	amod	positive
1	We want to trade with someone who has Houston ...	tickets	has	dobj	negative
2	i know! i just got off the phone with them. it...	hole	leave	dobj	negative
3	i know! i just got off the phone with them. it...	hole	giant	amod	negative
4	i know! i just got off the phone with them. it...	regions	closed	nsubj	negative
5	Happy mothers day everybody	mothers	happy	amod	positive
6	hey, but you have gal pals here in Phoenix!!!	pals	have	dobj	neutral
7	hey, but you have gal pals here in Phoenix!!!	pals	gal	amod	neutral
8	No star replies to me	star	replies	nsubj	negative
9	GRANADA BOWL WITH THE CREW! Birthday bowling f...	birthday	bowling	nsubj	neutral
10	I just broke a bottle, I should probably focus...	bottle	broke	dobj	negative
11	I just broke a bottle, I should probably focus...	songgoeswrongs	more	amod	negative
12	big hair rock day today me thinks here I go a...	day	big	amod	neutral
13	. Probably spelt it wrong lol.	lol	wrong	amod	negative
14	oh dear, hope you are feeling better soon, get...	soup	get	dobj	positive

Figure 13: Output of ABSA

8. Discussion

The ability of DistilBERT's to deeply understand the context and semantics of language, which traditional machine learning models inherently lack, makes this transformer-based model outperforms the other models in this study. DistilBERT uses the transformer architecture to process entire sentences, capturing the relationships between words and their surrounding context, whereas TF-IDF models treat words as independent features and ignore word order. This is particularly crucial in sentiment analysis since context, negation, or subtle linguistic cues can completely alter a sentence's sentiment.

Additionally, tweets and other social media texts are frequently informal and contain misspellings, slang, and abbreviations. Due to their heavy reliance on precise word matches and frequency counts, traditional models find it difficult to generalize in such noisy environments. DistilBERT, on the other hand, can produce meaningful embeddings even for uncommon or out-of-vocabulary words because it has been pre-trained on large and varied text corpora. The model can differentiate between tweets that may seem similar at first glance but convey distinct sentiments because DistilBERT generates dense, low-dimensional vectors that capture the text's syntactic structure and underlying meaning.

DistilBERT's foundation in transfer learning is another advantage of using it. The model has a general understanding of language by default, which can be adjusted for the particular sentiment analysis task. With less task-specific data needed to attain high performance, DistilBERT can swiftly adjust to the nuances of sentiment in tweets, thanks to this pre-training. Because of this, DistilBERT can generalise more effectively and produce predictions that are more accurate, even in the presence of sparse labelled data or subtle sentiment differences.

9. Conclusion & Future Work

9.1 Conclusion

In this project, sentiment analysis on Twitter data is studied using both transformer-based model and traditional machine learning models. The input data is made sure to be clean and consistent by implementing thorough preprocessing, which includes normalization, tokenization, stopword removal, and lemmatization. Although traditional methods provide good sentiment analysis, their limitations in capturing context and semantic meaning are evident in the evaluation results. The transformer-based DistilBERT model, on the other hand, shows a definite performance advantage. DistilBERT achieves greater accuracy and more reliable sentiment classification by utilising deep contextual embeddings to capture specific details and relationships in the text. The findings demonstrate the great efficacy of transformer-based models for sentiment analysis, especially when applied to informal and context-dependent social media data.

9.2 Future Work

Even though the current approach with DistilBERT generates excellent results, there are still a number of opportunities for development. First, experimenting with larger or more specialised transformer models, like domain-adapted models, BERT, or RoBERTa, may improve performance, particularly for more subtle sentiment distinctions. Sentiment classification may benefit from the addition of extra features like user metadata, tweet timestamps, or emoji analysis. Additionally, class imbalance may be addressed and generalization enhanced by utilising semi-supervised learning. Lastly, expanding the analysis to multilingual data or implementing the trained models in a real-time sentiment monitoring system would increase the work's usefulness and relevance.

References

1. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, *37*(2), 267–307. https://doi.org/10.1162/coli_a_00049
2. Hutto, C., & Gilbert, E. (2014). VADER: a Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1). https://doi.org/10.1609/icwsm.v8i1.14550
3. Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment Classification using Machine Learning Techniques. *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - EMNLP '02*, *10*. https://doi.org/10.3115/1118693.1118704
4. Zhang, Y., & Wallace, B. (2015). *A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification*. ArXiv.org. https://arxiv.org/abs/1510.0382
5. Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
6. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North*, *1*. https://doi.org/10.18653/v1/n19-1423
7. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019, July 26). *RoBERTa: A Robustly Optimized BERT Pretraining Approach*. ArXiv.org. https://arxiv.org/abs/1907.11692
8. Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., AL-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., De Clercq, O., Hoste, V., Apidianaki, M., Tannier, X., Loukachevitch, N., Kotelnikov, E., Bel, N., Jiménez-Zafra, S. M., & Eryiğit, G. (2016, January 1). *SemEval-2016 Task 5: Aspect Based Sentiment Analysis*. HAL Archives Ouvertes. https://doi.org/10.18653/v1/S16-1002

9. Xu, L., Li, H., Lu, W., & Bing, L. (2020). Position-Aware Tagging for Aspect Sentiment Triplet Extraction. *ArXiv (Cornell University)*. https://doi.org/10.18653/v1/2020.emnlp-main.183