Sentiment Analysis on Tweets Discussing ChatGPT

Chandrika Mukherjee
cmukherj@purdue.edu
Purdue University

Akhil Prasad
prasad67@purdue.edu
Purdue University

ABSTRACT

ChatGPT, being in forefront for a well-defined Large Language Model (LLM), the model has gained considerable attention and is able to find use-cases in various areas. Though, Generative AI tends to provide several applications, it comes with its flaws and the data which the model learns can tend to be biased, which can lead to misleading information/immense negative impact on millions of people. In this project, we are trying to analyse the varrious tweets on ChatGPT and analyse the general sentiment of people towards it using well-known Data Mining methods.

1 INTRODUCTION

Given the usage of ChatGPT increasing over time, there has been a significant impact of ChatGPT on the Internet community along with a significant discussion happening across the academic community. It has garnered a whopping 100 million users within two months of its release [1]. Due to this, we have found there has been significant discussion surrounding ChatGPT around the potential applications, capabilities and drawbacks that it possesses. In this project, we mainly focus our attention on the tweets related to ChatGPT. We worked on the tweets related to ChatGPT to understand the general sentiment towards ChatGPT.

The rise in the influence of ChatGPT poses both a threat and an opportunity. So, it becomes even more important to analyze the general public sentiment [10]. We have used some publicly available datasets, applied lexicon based sentiment analysis to assign ground truth label. On the prepared data, we used LSTM, Logistic Regression, Support Vector Machine to train sentiment detection models. The comparison of performance of these models are shown towards the end of this report.

Considering the scope of this project along with the time and resource constraints, we analyzed the tweets from a couple of datasets leveraging some of the attributes of the tweets which is explained further in Section 3 and 4. We have performed pre-processing on the Dataset which has been explained in section 5. Subsequently, we have discussed the chosen data mining methods in sections 6, 7, and 9 to fulfil our objectives. Next we described our detailed evaluation in section 10. We also indicated the insights out of this project work and discussed the future directions. Please refer to the released code in our github repository [8].

2 RELATED WORK

In [7], the authors have extensively studied the concepts of sentiment analysis and the various domains in which it is actively being used. They give a detailed overview of the various techniques used in sentiment analysis along with the potential harmful outcomes it tends to have. This article gives a detailed insight about sentiment analysis.

Sentiment Analysis has found a lot of applications in analysing the emotions behind the tweets. In [2], authors use sentiment analysis to detect hate speech in Urdu tweets. The authors use dynamic stop words filtering for the sparsity issue, Variable Global Feature Selection Scheme (VGFSS) for the dimensionality issue, and Synthetic Minority Optimization Technique (SMOTE) for the class imbalance issue. They SVM and Multi-nomial NBC to analyze the performance of their technique. In [11], the authors used emoticon based sentiment analysis by applying NBC on Chinese tweets to understand the abnormal events using the spatial and temporal patterns. In [3], the authors perform sentiment analysis on COVID-19 related tweets by applying various machine learning classifiers and they concluded that BERT and RoBERTa models are well suited for Twitter data. In [12], the authors propose a new text-based NLP model, TSAM to analyse the general opinion on various societal events.

There has been some work on understanding the general emotions related to ChatGPT and the applications it has in various domains [6]. Sentiment analysis of tweets surrounding usage of ChatGPT in education has been explored in [9], where they analyse different classifiers and conclude that SVM has the best accuracy and in [4], the authors use NBC to perform sentiment analysis on the tweets related to ChatGPT. In [5], the authors use AFINN, Bing and NRC sentiment dictionaries to analyse the ChatGPT related tweets. In our work, we have performed sentiment analysis using LSTM, LR and SVM, which is discussed further in the next few sections.

3 DATASET

We identified four relevant datasets, publicly available in kaggle or huggingface. Following are the selected datasets - $dataset_1$ [1], $dataset_2$ [2], $dataset_3$ [3], $dataset_4$ [4]. These datasets contain information about tweets regarding ChatGPT that covers wide range of topics, for example, artificial intelligence, openai, large language model from various sources. Every dataset contains the tweet text that can be anything from users asking questions to ChatGPT, expressing their opinions about it, or sharing their experiences using the AI model.

Above mentioned $dataset_1$, has the following columns - tweet_id, created_at, like_count, quote_count, reply_count, retweet_count, tweet, country, photo_url, city, country_code. Here tweet is a text string that is the body of the tweet. The dataset includes metadata about each tweet, such as the date and time it was posted, the number of likes and retweets, replies it received. This dataset also contains some geographical information along with photo url data.

[1] https://www.kaggle.com/datasets/pcminh0505/chatgpt-twitter
[2] https://huggingface.co/datasets/deberain/ChatGPT-Tweets
[3] https://www.kaggle.com/datasets/khalidryder777/500k-chatgpt-tweets-jan-mar-2023
[4] https://www.kaggle.com/datasets/konradb/chatgpt-the-tweets

Dataset	Number of Rows	Unique Tweets
1	219294	217624
2	305424	305400
3	500002	500002
4	558405	548502

Table 1: Unique Information From Selected Datasets

Dataset	Number of Words
1	31954876
2	52395377
3	84157825
4	106602066

Table 2: Number of Words in Tweets Over Selected Datasets

Size of this dataset is 48.52 MB.

The second $dataset_2$ contains similar information along with specific information about the user. For example, this dataset contains username, user's description, count of user's followers and friends. The size of this dataset is 133 MB.

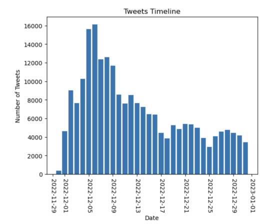
The third $dataset_3$ is of size 117 MB has similar and has less number of columns compared to $dataset_1$.

The fourth $dataset_4$ of size 248 MB is more similar to the $dataset_2$ with similar attributes.

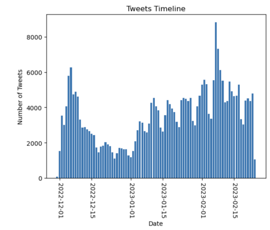
We have analyzed all four datasets to decide on the appropriate dataset for our data mining tasks. The following section shows different data analysis results.

4 EXPLORATORY DATA ANALYSIS

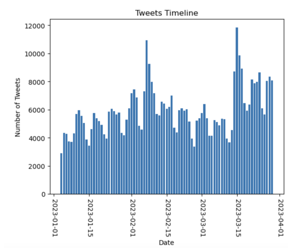
Table-1 shows the number of unique tweets from above 4 datasets. Table- 2 shows the number of words in tweets in all selected datasets. Figure-1 shows the timeline of tweets in all four datasets. Figure-2 shows the wordclouds of four datasets. The primary words used in the first three datasets are ai, chatbot, openai, gpt, future, prompt, code etc. However the content of tweets from $dataset_4$ is different. Subfigures - 2d shows the main content to be cryptocurrency, price, jump, starting small and very less content about ai. We have explored top common words in these datasets. All datasets except for $dataset_4$ have tweets with common topic revolving around chatgpt. Therefore, we will discard the $dataset_4$ from further analysis. We have visualized the $dataset_1$ and $dataset_2$ from geographical perspective (figure-3) as both of them contains information about the country/location of the user who posted the tweet. According to the analysis, top 5 countries sharing tweets in $dataset_1$ are USA, India, UK, Canada, Australia. Tweets in the $dataset_2$ are from India (New Delhi, Bangalore, Pune), USA (CA, NY), London, Canada, Germany etc. Maximum number of tweets from a location in $dataset_1$ is around 1600 and in $dataset_2$ is around 5000. The $dataset_1$ has photo meta data associated with the tweet. Figure - 4 shows that



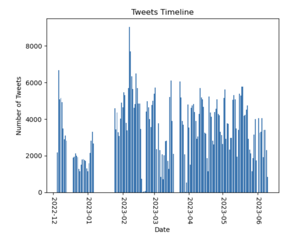
(a) $dataset_1$



(b) $dataset_2$

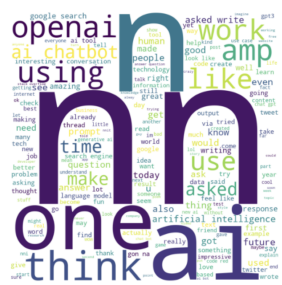


(c) $dataset_3$

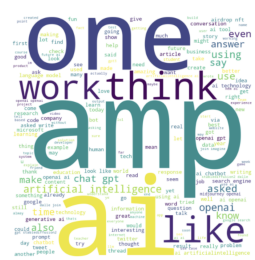


(d) $dataset_4$

Figure 1: Tweets Timeline for Four Datasets



(a) $dataset_1$



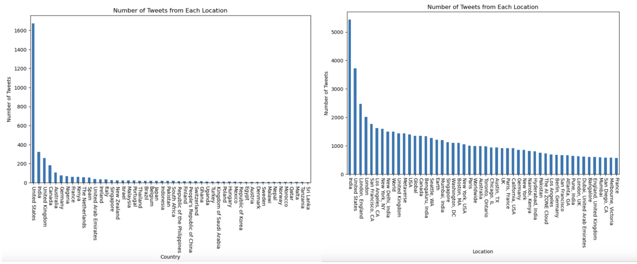
(b) $dataset_2$

(c) $dataset_3$

(d) $dataset_4$

Figure 2: Wordclouds for Four Datasets

tweets with no photo were more popular based on the like counts. Figure - 5 shows popularity of tweets over the timeline based on like, retweet, reply or quote counts wherever applicable. As expected the likes for a tweet are more than retweets. User information is available in $dataset_2$ and $dataset_3$, therefore, we have visualized the popular users from the two datasets who shared tweets. For the $dataset_2$, we have visualized popularity of the user based on user's follower and friends. For the other dataset, information of user's



(a) $dataset_1$ (b) $dataset_2$

Figure 3: Location Wise Count of Tweets

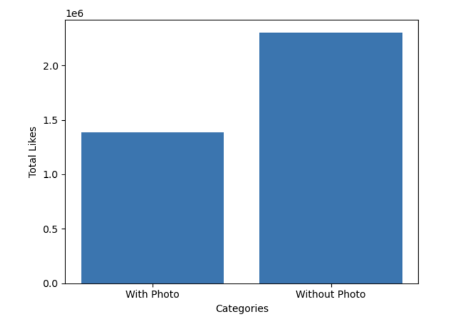
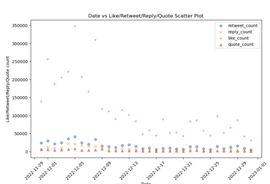
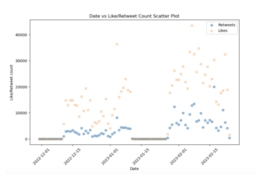


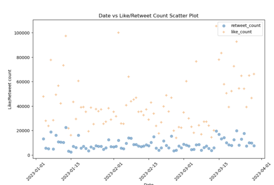
Figure 4: Count of Likes w.r.t Photo Metadata ($dataset_1$)



(a) $dataset_1$



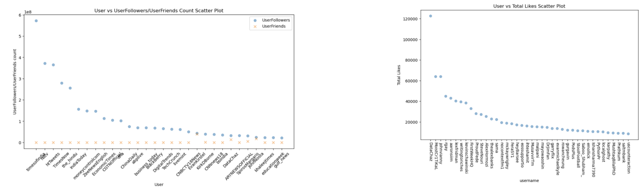
(b) $dataset_2$



(c) $dataset_3$

Figure 5: Trend of Popularity of Tweets

followers and friends were not available, so, visualized popularity based on like counts for a user's tweets. From this visualization, we found that the data from $dataset_2$ are from different news channel like TimesofIndia, NDTV, The_Hindu, IndiaToday etc. whereas the users in $dataset_3$ are mostly unknowns. Therefore to generate a better viewpoint from this data mining tasks, we will consider the $dataset_2$ instead of $dataset_3$.



(a) $dataset_1$ (b) $dataset_2$

Figure 6: User's Popularity

Finally as the $dataset_1$ and $dataset_2$ both contain popular tweets about chatgpt and have some unique features - $dataset_2$ with more information about user, $dataset_1$ with more information about the tweet. We will primarily apply data mining algorithms on $dataset_1$. However, we will also apply the selected algorithms on $dataset_2$ and comapre the results.

	tweet	final_tweet
0	ChatGPT: Optimizing Language Models for Dialog...	optimizing language model dialogue
1	Try talking with ChatGPT, our new AI system wh...	try talking new ai system optimized dialogue f...
2	ChatGPT: Optimizing Language Models for Dialog...	optimizing language model dialogue ai machinel...
3	THRILLED to share that ChatGPT, our new model ...	thrilled share new model optimized dialog publ...
4	As of 2 minutes ago, @OpenAI released their ne...	2 minute ago released new nnand use right
5	Just launched ChatGPT, our new AI system which...	launched new ai system optimized dialogue
6	As of 2 minutes ago, @OpenAI released their ne...	2 minute ago released new nnand use right n n
7	ChatGPT coming out strong refusing to help me ...	coming strong refusing help stalk someone agre...
8	#0penAI just deployed a thing I've been helpin...	0penal deployed thing ive helping build last c...
9	Research preview of our newest model: ChatGPT\...	research preview newest model nnwere trying so...
10	GOD DAMN IT @OpenAI STOP ANNOUNCING THINGS I A...	god damn stop announcing thing busy
11	OpenAI ChatGPT: Optimizing Language Models for...	openai optimizing language model dialoguenl
12	ChatGPT: Optimizing Language Models for Dialog...	optimizing language model dialogue ai machinel...
13	OpenAI announced ChatGPT a model optimized for...	openai announced model optimized dialogue
14	OpenAI ChatGPT: Optimizing Language Models for...	openai optimizing language model dialogue n2
15	#AI #techforgood ChatGPT: Optimizing Language ...	ai techforgood optimizing language model dialo...
16	#Technical ChatGPT: Optimizing Language Models...	technical optimizing language model dialogue a...
17	#ai Models are set to become the search engine...	ai model set become search engine future atm s...
18	OpenAI ChatGPT: Optimizing Language Models for...	openai optimizing language model dialogue n2
19	I asked ChatGPT (a new AI system that is optim...	asked new ai system optimized dialogue teach s...

(a)

Figure 7: Example - Main Tweet and Preprocessed Tweet

5 PREPROCESSING PHASE

After conducting exploratory data analysis on selected datasets, we have applied various preprocessing techniques on the data so that the available data can provide us with more and important information. Tweet texts are first converted to lower case. Then stopwords, hyperlinks, emojis were removed from the texts. We applied lemmatization on the tweets to group together the inflected forms of a word so they can be analyzed as a single item, identified by the word's lemma, or dictionary form. Figure-7 shows an example of available tweet text and corresponding preprocessed text.

6 SENTIMENT ANALYSIS

There are multiple approaches to determine the emotional tone or sentiment expressed in a text. Lexicon-based approach is simple and involves using a set of predefined rules and heuristics to determine the ground truth label. These rules are typically based on lexical and syntactic features of the text, such as the presence of positive or negative words and phrases. However, the accuracy may not be good when dealing with complex texts or emotion. To label the sentiment of each tweet using ML algorithms, we need to manually label the available training data as ground truth which is time consuming. Therefore, we have not used machine learning techniques to assign ground truth labels. We could also use pretrained transformer based deep-learning models for assigning sentiment label for each tweet. However, this will require significant computational resources. Therefore, for this proposed task, we relied on simple lexicon-based approach to determine the ground truth label.

We have incorporated Valence Aware Dictionary and sEntiment Reasoner (VADER Sentiment Analysis) labelling the data. VADER uses a combination of A sentiment lexicon is a list of lexical features (e.g., words) which are generally labeled according to their semantic orientation as either positive or negative. VADER not only tells about the Positivity and Negativity score but also tells us about how positive or negative a sentiment is.

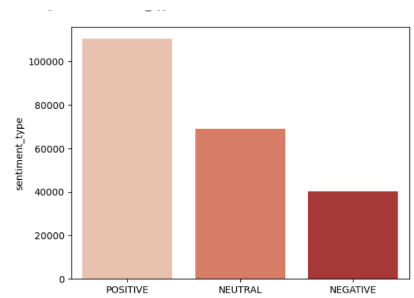
Figure - 8, 9 shows the number of labelled positive, negative and neutral data in the $dataset_1$ and $dataset_2$. Both the datasets have comparable amount of positive, negative and neutral data.

Figure -10a shows the number of positive, negative and neutral tweets from November 30th to December 31st 2022 in case of $dataset_1$. We observe that the overall count of tweets discussing ChatGPT is less towards the end of the above timeline. From figure-10b, we can observe that for the $dataset_2$, the tweets are from November 30th 2022 to February 24th 2023. Here the number of tweets are more towards the end of the timeline. For both the datasets, number of positive tweets are relatively more in count than negative and neutral tweets.

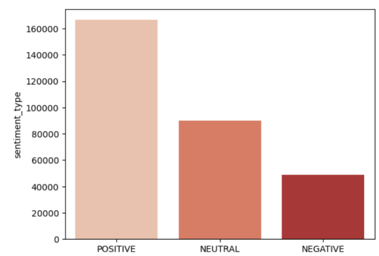
7 FEATURE EXTRACTION

Until now, we have explored and selected our primary and secondary datasets. We have labelled the ground truth for each of the tweets using VADER sentiment analysis. So, after this we initialized a separate dataset with preprocessed tweets and corresponding ground truth sentiment labels (positive, negative, neutral). Therefore, after this, we worked on converting the text of the tweets into feature vectors. The feature vector represents a numerical representation of the input text data. In the context of natural language processing (NLP) and machine learning, these feature vectors are used to represent the characteristics of the text in a format that can be easily understood and processed by machine learning algorithms.

To extract feature vectors, we have incorporated the Word2vec technique. Word2vec is a group of related models that are used to produce word embeddings. These models are shallow, two-layer neural networks that are trained to reconstruct linguistic contexts of words. Word2vec takes as its input a large corpus of text and produces a vector space, typically of several hundred dimensions, with each unique word in the corpus being assigned a corresponding vector in the space. Word2vec can utilize either of two model architectures to produce these distributed representations of words: continuously sliding bag-of-words (CBOW) or continuously sliding skip-gram. In both architectures, word2vec considers both individual words and a sliding context window as it iterates over the corpus.



(a) $dataset_1$



(b) $dataset_2$

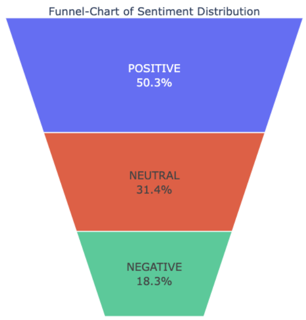
Figure 8: Sentiment Type Labelled

8 DATA SPLITTING -

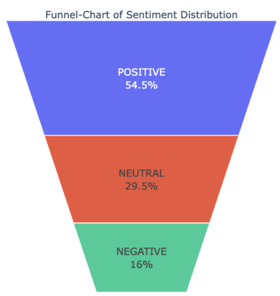
We have divided the datasets into training and testing sets by creating a 80 : 20 split with a specific random state (42 in our case).

9 APPLYING DATA MINING ALGORITHMS -

According to our literature review, we observed that Logistic Regression (LR), Support Vector Machine (SVM) and Long Short-Term Memory (LSTM) are three common data mining algorithms used for sentiment analysis. Logistic Regression (LR) works well when the relationship between features and sentiment is approximately linear. In this case, we have used multinomial Logistic Regression as we have three different class labels in the dataset. SVM is effective in handling non-linear relationships between features and sentiment. Long Short-Term Memory (LSTM) networks are a type of recurrent neural network (RNN) that are particularly well-suited for natural language processing (NLP) tasks - handles sequential data, captures long-range dependencies, handles vanishing gradient problem and variable-length sequences. LSTMs are well-suited for sequential data like text. They can capture dependencies and relationships in the sequence of words, making them effective for sentiment analysis tasks where the context of words matters. We have used same training data to generate three different models using LR, SVM and LSTM algorithms. Later we used these models on test dataset to evaluate the accuracy.



(a) $dataset_1$

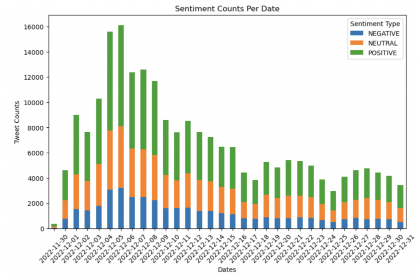


(b) $dataset_2$

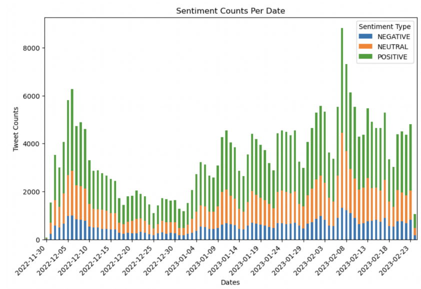
Figure 9: Distribution of Positive, Negative and Neutral Sentiments

Model	Accuracy (Dataset1)	Accuracy (Dataset2)
Logistic Regression	64%	65%
SVM	74%	75%
LSTM	90%	89%

Table 3: Model Type vs Accuracy



(a) $dataset_1$



(b) $dataset_2$

Figure 10: Trend of Different Sentiment Types

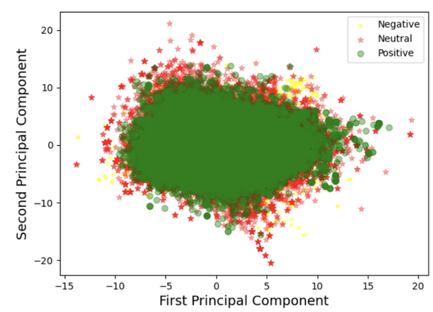


Figure 11: PCA on $dataset_1$

10 EVALUATION

Table 3 shows the test accuracy obtained by using different models on $dataset_1$ and $dataset_2$. According to the results we observe that the Logistic Regression performs poorly on both of the datasets with only 64% and 65% test accuracy respectively. SVM performs better than LR with accuracy of 74% and 75% respectively. Reducing the dimension of the feature vectors, we observe that the data is non-linear in figure 11 . As already mentioned in section 9, SVM is more effective in handling non-linearity than LR. Therefore, the outcome validates the previously mentioned fact. Also LSTM performed better than SVM in case of both the datasets. LSTMs can automatically learn relevant features from the input data, which is advantageous in tasks like sentiment analysis where the identification of sentiment-bearing words or phrases can be complex and context-dependent. SVMs, on the other hand, may require explicit feature engineering. The result shows that the test accuracy for LSTM was highest with 90% and 89% for $dataset_1$ and $dataset_2$ respectively. Even though performance wise LSTM, SVM were better than LR, both of them required significant amount of time to train the models.

Along with accuracy, we report the precision, recall, F1-score for performance of the above selected algorithms on selected datasets.

A high accuracy score indicates that the model is making a large proportion of correct predictions and is performing well overall.

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Test\ Data\ Points}$$

	precision	recall	f1-score	support
negative	0.53	0.22	0.31	8038
neutral	0.61	0.60	0.60	13637
positive	0.67	0.82	0.74	22184
accuracy			0.64	43859
macro avg	0.60	0.55	0.55	43859
weighted avg	0.63	0.64	0.62	43859

(a) $dataset_1$

	precision	recall	f1-score	support
negative	0.52	0.20	0.29	9778
neutral	0.59	0.53	0.56	18023
positive	0.69	0.85	0.76	33284
accuracy			0.65	61085
macro avg	0.60	0.53	0.54	61085
weighted avg	0.64	0.65	0.63	61085

(b) $dataset_2$

Figure 12: Classification Report using Logistic Regression

	precision	recall	f1-score	support
negative	0.75	0.29	0.41	8038
neutral	0.76	0.75	0.76	13637
positive	0.73	0.90	0.80	22184
accuracy			0.74	43859
macro avg	0.75	0.65	0.66	43859
weighted avg	0.74	0.74	0.72	43859

(a) $dataset_1$

	precision	recall	f1-score	support
negative	0.68	0.26	0.37	9778
neutral	0.75	0.73	0.74	18023
positive	0.75	0.90	0.82	33284
accuracy			0.75	61085
macro avg	0.73	0.63	0.65	61085
weighted avg	0.74	0.75	0.73	61085

(b) $dataset_2$

Figure 13: Classification Report using SVM

	precision	recall	f1-score	support
negative	0.91	0.71	0.80	8038
neutral	0.84	0.99	0.91	13637
positive	0.95	0.92	0.93	22184
micro avg	0.90	0.90	0.90	43859
macro avg	0.90	0.87	0.88	43859
weighted avg	0.91	0.90	0.90	43859
samples avg	0.90	0.90	0.90	43859

(a) $dataset_1$

	precision	recall	f1-score	support
negative	0.88	0.63	0.74	9778
neutral	0.81	0.99	0.89	18023
positive	0.95	0.91	0.93	33284
micro avg	0.89	0.89	0.89	61085
macro avg	0.88	0.84	0.85	61085
weighted avg	0.90	0.89	0.89	61085
samples avg	0.89	0.89	0.89	61085

(b) $dataset_2$

Figure 14: Classification Report using LSTM

Task	Planned Completion Date
Literature Review	Sep 29
Analysis and Shortlisting of Dataset	Oct 5
Analysing the various available Data Mining Methods	Oct 10
Implementation of Various Models	Nov 5
Training the Various Models with the shortlisted Dataset	Nov 10
Testing and Verification	Nov 24
Presentation Prep	Dec 1
Report Completion	Dec 10

Table 4: Initial Plan to Achieve Our Objectives

A high precision score indicates that the model is good at making positive predictions and is less likely to make false positive error.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

A high recall score indicates that the model is good at capturing most of the positive instances and has a low rate of false negatives.

$$Recall = \frac{True\ Positive}{True\ Positives + False\ Negative}$$

A high F1 score indicates that the model is good at both minimizing false positives and capturing most of the positive instances. It is generally a desirable characteristic in situations where you want a balance between precision and recall.

$$F1\ Score = \frac{2\ *\ Precision\ *\ Recall}{Precision + Recall}$$

We included classification report for all the models and datasets. Refer to the figures - 12, 13, 14. Based on these results, we find that the performance of LSTM is better than that of SVM and LR models.

According to our initial plan (table - 4), we have completed all our tasks and reported all the outcomes.

11 INSIGHTS AND LIMITATIONS

Based on the VADER sentiment analysis, the selected datasets had more positive data than neutral and negative instances. Therefore, overall, the public has a positive outlook towards the use of ChatGPT in their daily lives. However as suspected the opinion is not fully positive, contains some negative concerns as well.

In this work, to label the ground truth data, we have utilized VADER sentiment analysis. However, this is a simple lexicon based approach and it has some shortcomings like VADER often treats each sentence independently. It doesn't consider the sentiment flow across multiple sentences or the overall sentiment of a paragraph, which may lead to misinterpretations.

As the ground truth has been assigned by VADER sentiment analysis, the results obtained using different data mining algorithms provides a direct relation with VADER assigned labels. This can be improved by either manually label subset of the datasets and then correct the ground truth labels for the rest of the dataset. Or we can make use of pretrained transformer based deep-learning models like BERT, RoBERTa, GPT etc.

12 FUTURE DIRECTIONS

We have relied on social media - twitter to gather the dataset discussing about ChatGPT. However, we can plan to conduct offline or online interviews of diverse set of people with curated questions targeted towards use of ChatGPT in our daily lives. Datasets obtained in this process will be specific and have more reliability. Then our trained model can be directly used on the obtained dataset to summarize general public opinion about ChatGPT.

Other than relying solely on text-based data, we can plan for multimodal sentiment analysis. Integrating information from multiple modalities, such as text, images, audio, and video, for a more comprehensive understanding of sentiment and therefore will provide better summary of general public opinion.

In this project, we focus our work mainly using the generalized Machine Learning models and we have not looked at other learning models such as Transformers. Intuitively, we would be able to get better understanding in case we used Transformers which would help us in identifying the word embedding from the tweets and can help us get better insights other than a simple Positive/Negative output label. Applying Transformer based models would be an interesting direction to proceed for our project.

This project limits our work to sentiment analysis, as an extension of this, we want to work on the causality of the tweet data. Would be interesting to perform an extensive causal analysis of the tweets. With causal analysis on the tweets, would be able to understand the performance of ChatGPT in a better way and we can help in making the GenAI models better.

13 CONTRIBUTION OF EACH TEAMMATE -

- Both of us worked together in standardizing the preprocessing and data exploration methods.
- Chandrika reported the data exploration results of dataset1 and dataset3. Akhil reported the data exploration results of dataset2 and dataset4.
- Both of us worked together to understand the related work.
- Both of us worked together in selecting and standardizing the data mining processes involved in our work.
- Chandrika worked on obtaining results using Dataset1 utilizing the selected data mining methods. Akhil reported the results for Dataset2.
- Chandrika worked on Insights and Limitations.
- Akhil worked on Future Directions.
- We both worked on determining the references and update final results to GitHub [8].
- We equally contributed to the writing of this report and the presentation.

REFERENCES

[1] 2023. ChatGPT Tops 25 Million Daily Visits. https://www.similarweb.com/blog/insights/ai-news/chatgpt-25-million/.

[2] Muhammad Z. Ali, Ehsan-Ul-Haq, Sahar Rauf, Kashif Javed, and Sarmad Hussain. 2021. Improving Hate Speech Detection of Urdu Tweets Using Sentiment Analysis. *IEEE Access* 9 (2021), 84296–84305. https://doi.org/10.1109/ACCESS.2021.3087827

[3] Niklas Braig, Alina Benz, Soeren Voth, Johannes Breitenbach, and Ricardo Buettner. 2023. Machine Learning Techniques for Sentiment Analysis of COVID-19-Related Twitter Data. *IEEE Access* 11 (2023), 14778–14803. https://doi.org/10.1109/ACCESS.2023.3242234

[4] Adhitia Erfina and Muhamad Nurul. 2023. Implementation of Naive Bayes classification algorithm for Twitter user sentiment analysis on ChatGPT using Python programming language. *Data Metadata* 2 (06 2023), 45. https://doi.org/10.56294/dm202345

[5] Adem Korkmaz, Cemal Aktürk, and Tarik Talan. 2023. Analyzing the User's Sentiments of ChatGPT Using Twitter Data. *Iraqi Journal for Computer Science and Mathematics* 4 (05 2023), 202–214. https://doi.org/10.52866/ijcsm.2023.02.02.018

[6] Reza Hadi Mogavi, Chao Deng, Justin Juho Kim, Pengyuan Zhou, Young D. Kwon, Ahmed Hosny Saleh Metwally, Ahmed Tlili, Simone Bassanelli, Antonio Bucchiarone, Sujit Gujar, Lennart E. Nacke, and Pan Hui. 2023. Exploring User Perspectives on ChatGPT: Applications, Perceptions, and Implications for AI-Integrated Education. arXiv:2305.13114 [cs.CY]

[7] Saif M. Mohammad. 2021. Sentiment Analysis: Automatically Detecting Valence, Emotions, and Other Affectual States from Text. arXiv:2005.11882 [cs.CL]

[8] Chandrika Mukherjee and Akhil Prasad. 2023. *ChatGPT Sentiment Analysis*. https://github.com/chandms/ChatGPT-Tweets-Analysis

[9] Mohammad Tubishat, Feras Al-Obeidat, and Ahmed Shuhaiber. 2023. Sentiment Analysis of Using ChatGPT in Education. *WSEAS TRANSACTIONS ON ADVANCES in ENGINEERING EDUCATION* 20 (08 2023), 60–66. https://doi.org/10.37394/232010.2023.20.9

[10] Rui Wang, Xing Liu, Yanan Wang, and Haiping Huang. 2023. What are Public Concerns about ChatGPT? A Novel Self-Supervised Neural Topic Model Tells You. arXiv:2309.01522 [cs.CL]

[11] Jichang Zhao, Li Dong, Junjie Wu, and Ke Xu. 2012. MoodLens: An emoticon-based sentiment analysis system for chinese tweets. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (08 2012). https://doi.org/10.1145/2339530.2339772

[12] Xujuan Zhou, Xiaohui Tao, Jianming Yong, and Zhenyu Yang. 2013. Sentiment analysis on tweets for social events. In *Proceedings of the 2013 IEEE 17th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*. 557–562. https://doi.org/10.1109/CSCWD.2013.6581022