Jade Bell
Professor Zhou
CSC 425 001
7 December 2025

CSC 425 AI Project Report

Topic

Sentiment Analysis - ChatGPT User Reviews from Google Play Store

Project Team Members

Jade Bell

Responsibilities

My responsibilities include all aspects of this project, including dataset acquisition, preprocessing and cleaning, feature engineering, exploratory analysis, model development, training/test split, performance evaluation, visualisation, and sentiment analysis.

1. Introductory

Sentiment analysis, a part of Natural Language Processing (NLP), identifies and categorises opinions in textual data to determine whether sentiments are positive, neutral, or negative. This provides insights into user perceptions and experiences with products and services. In the context of AI products like ChatGPT, user-generated reviews provide an ongoing stream of feedback that reflects satisfaction, usability issues, and public perception. By analysing thousands of real-world user reviews, sentiment analysis can reveal trends, identify themes in user concerns, and help developers prioritise improvements.

This project will focus on sentiment analysis of ChatGPT reviews from the Google Play Store, which is a frequently updated dataset that reflects users' experiences with the most used AI applications. Given the evolution of AI applications, user sentiment shifts in response to new updates, feature changes, and improvements in app performance. Understanding these variations is valuable for users to gain insight into experiences in AI, while developers receive actionable information summarising user feedback at a scale. By building machine learning models to classify these entiments, this project seeks to evaluate how well the methods capture sentiment patterns through their feedback.

2. Problem Statement

How can sentiment analysis of an AI Product, ChatGPT reviews, be used to better understand user satisfaction and evolving public perception of AI systems over time?

2.1 Main Goal

To develop and evaluate sentiment classification for ChatGPT app reviews that not only achieve high accuracy but also provide deeper insight into user sentiment patterns, feature importance, and trends in feedback towards AI applications.

- Will introduce feature engineering techniques and comparative modelling to improve its performance.
- Real World Impact: ChatGPT represents one of the widely used AI systems. Understanding user sentiment toward this product benefits both consumers and producers. For consumers (users), it ensures their feedback is effectively analysed, enabling greater improvement and greater trust in AI technology. For producers and developers, sentiment analysis provides insights into which features users value and the issues they face, helping with updates, bug fixes, and AI decisions. On a bigger scale, monitoring sentiments around AI tools is critical for innovation and user satisfaction.

3. Dataset

The dataset contains reviews on ChatGPT from the Google Play Store, which is updated daily.
Dataset Size: ~ 60,000 reviews (updated October 2025)
Source:
https://www.kaggle.com/datasets/ashishkumarak/chatgpt-reviews-daily-updated/data

Features of the dataset include:

Feature	Type	Description
Username	string	Identifieer of user
Review Id	string	Reviewer Identifier
Review Content	string	Content of review
Rating	integer	User rating(1-5 stars)
Likes	integer	Number of thumbs-up likes

App Version	string	ChatGPT app version
Review Timestamp	datetime	When user review was posted on Google Play Review
Created date	datetime	When review was created

For the rating score, the following sentiment labels were created:
Positive -> 4-5 stars
Neutral -> 3 stars
Negative -> 1-2 stars

4. Preprocessing/Feature Engineering

Before modelling, preprocessing and cleaning were necessary to normalise the text.

4.1 Dataset Cleaning

a. Removal of missing or duplicate reviews
b. Lowercase Text
c. Removal of punctuation, HTML, URLs, special characters, emojis, and numbers
d. Removed extra whitespace

4.2 Textual Processing

a. Tokenization
b. Stpword Removal using NLTK
c. Lemmatization

4.3 Feature Engineering

Feature engineering is the process of creating or modifying features that machine models can understand. This additional step is crucial for preparing the dataset to enhance its performance for machine learning models. Numerical features added to strengthen the model predictions include:

a. Review Length (number of chars)
b. Word count
c. Sentiment intensity
d. Reviews datetime(date features)

Project Timeline and Plan

Week	Milestone Task

Week 1	Choose a dataset and create an objective/plan
Week 2	Data cleaning, preprocessing, and data analysis; TD-IDF, WordCloud for separating sentiment review words
Week 3	Implement Naive Bayes AI Model, DT, Gradient Booster, confusion matrix, and Feature Engineering
Week 4	Improve any metrics and charts.

5. Methodology

The following machine learning(Supervised ML) algorithms were implemented:

1. Multinomial Naive Bayes - used as required and widely used for classification tasks
2. Decision Tree Classifier - provides insight into patterns within the data
3. Gradient Booster Classifier - used to see whether boosting weaker learning models could achieve higher accuracy

These models were trained with TF-IDF vectorised text.

TD-IDF (Term Frequency-Inverse Document Frequency) was used to convert textual data into numerical features. This emphasises meaningful words while reducing the impact of uninformative words. A maximum of 5,000 features was selected with both uni/bigrams.

5.1 Model Training

a. Transform preprocessed text into TD-IDF
b. Train/Test split (80% training, 20& testing)
c. Train models and evaluate using
    i. Accuracy, precision, confusion matrix, cross-validation 5 5-fold

6. Results:

A summary of results based on my code output and outcomes of models on a large TD-IDF dataset includes:

6.1 Sentiment Distribution Percentages

The analysis shows that neutral reviews are most prevalent in the dataset, indicating that users do not show strong emotions.

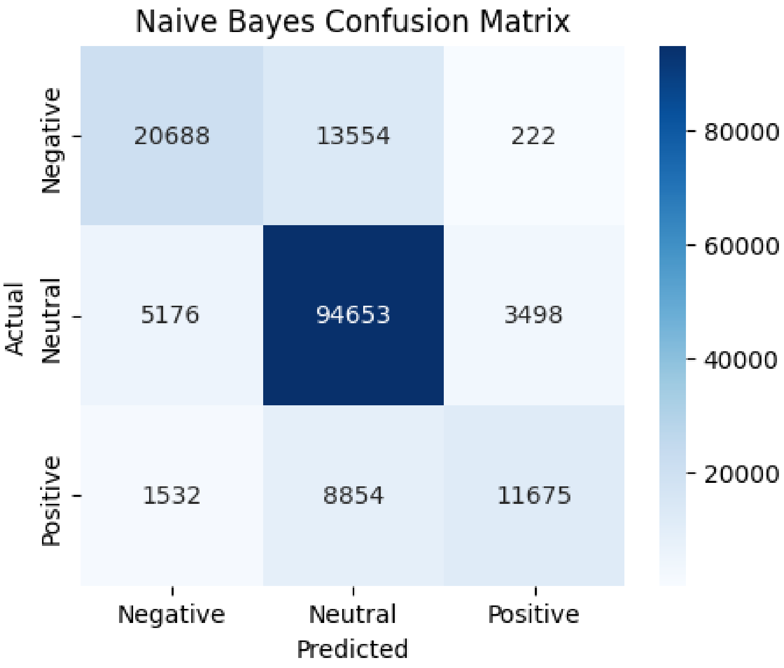
Sentiment	Count of Reviews	Percentage
Neutral	515,414	64.5%
Negative	172,792	21.6%
Positive	111,054	13.9%

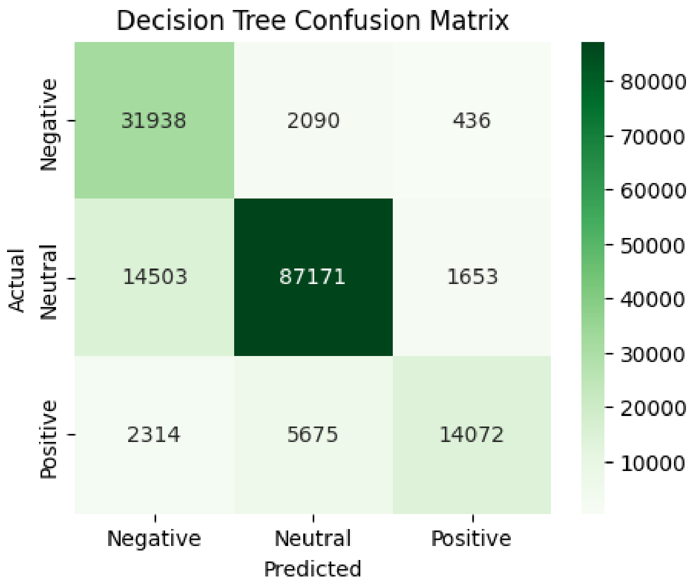
Model	Test Accuracy	Positive Prec/Recall	Neutral Prec/Recall	Negative Prec/Recall
Naive Bayes	79%	Precision:81% Recall:92%	Precision:76% Recall:53%	Precision:76% Recall:60%
DT	83%	Precision:87% Recall:64%	Precision:92% Recall:84%	Precision:66% Recall:93%
Gradient Booster	85%	Precision:83% Recall:60%	Precision:90% Recall:90%	Precision:77% Recall:92%

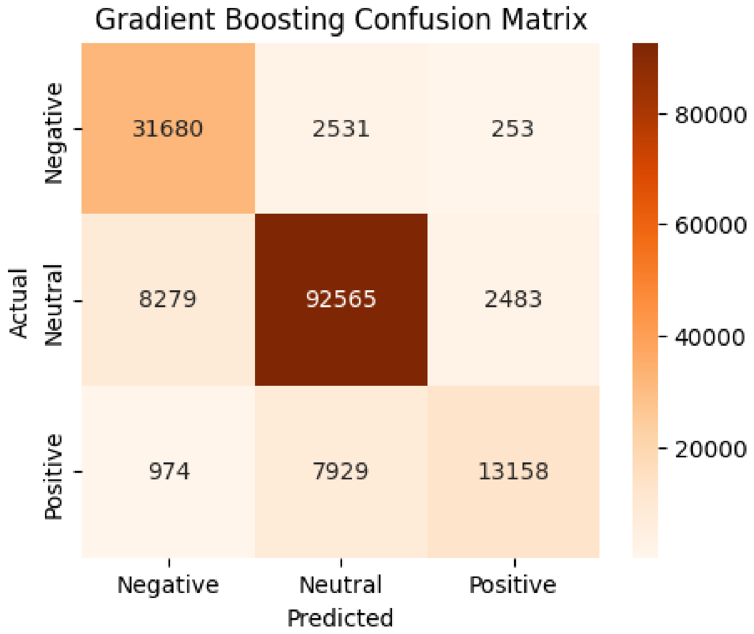
6.2 Model Comparisons

```
Naive Bayes Model Evaluation
Accuracy: 0.794584991116783

Classification Report:
               precision    recall  f1-score   support

           0       0.76      0.60      0.67     34464
           1       0.81      0.92      0.86    103327
           2       0.76      0.53      0.62     22061

    accuracy                           0.79    159852
   macro avg       0.77      0.68      0.72    159852
weighted avg       0.79      0.79      0.79    159852

Decision Tree Model Evaluation
Accuracy: 0.8331519155218577

Classification Report:
               precision    recall  f1-score   support
    Negative       0.66      0.93      0.77     34464
     Neutral       0.92      0.84      0.88    103327
    Positive       0.87      0.64      0.74     22061

    accuracy                           0.83    159852
   macro avg       0.81      0.80      0.79    159852
weighted avg       0.85      0.83      0.84    159852

Gradient Boosting Model Evaluation
Accuracy: 0.8595638465580662

Classification Report:
               precision    recall  f1-score   support

    Negative       0.77      0.92      0.84     34464
     Neutral       0.90      0.90      0.90    103327
    Positive       0.83      0.60      0.69     22061

    accuracy                           0.86    159852
   macro avg       0.83      0.80      0.81    159852
weighted avg       0.86      0.86      0.86    159852
```

6.3 Confusion Matrices


Naive Bayes Confusion Matrix





7. Conclusions:

Through examination of my results, the models revealed several patterns. Most reviews were classified as neutral. This is because a large number of users write very short or low-emotion comments such as “works fine” or “okay,” which do not provide strong sentiment clues. Since the models relied mainly on TF-IDF features, short reviews produced very weak signals, causing the classifier to default to the neutral category. Additionally, the positive class had a lower recall because many words used in positive reviews also appear in neutral ones. Terms like “good,” “fine,” and “works” overlap across both classes, making it harder for the model to separate them and leading to some of the misclassifications.

In conclusion, this project demonstrates that sentiment analysis is a powerful tool for understanding how people feel about AI tools like ChatGPT. By analysing thousands of reviews at once, we can quickly see overall satisfaction levels, how people react to updates over time, and what kinds of concerns or frustrations users commonly mention. Even though my models had challenges like neutral bias and short reviews, the patterns still reveal that user perception of AI is mixed. Overall, sentiment analysis enables us to track user satisfaction, identify trends in public opinion, and understand how people respond to AI as it becomes an increasingly integral part of everyday life.

8. Ethics Consideration:

In this project, there was no private user information accessed, since the reviews come from the public app review 'Google Play Store'. All reviews were analysed in groups and not individually. These sentiment models may contain some bias due to how the users wrote their reviews, so my results should not be viewed as perfect or as how every user feels. Understanding user sentiment toward AI applications like ChatGPT can help improve these tools, which may benefit users by making the technology more reliable.

References

https://www.geeksforgeeks.org/machine-learning/decision-tree-vs-naive-bayes-classifier/
https://www.geeksforgeeks.org/machine-learning/how-to-compute-entropy-using-scipy/
https://www.geeksforgeeks.org/machine-learning/ml-gradient-boosting/
https://www.ibm.com/think/insights/how-can-sentiment-analysis-be-used-to-improve-customer-experience
https://aws.amazon.com/what-is/nlp/