MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE GWALIOR

(A Govt. Aided UGC Autonomous Institute Affiliated to RGPV, Bhopal)

NAAC Accredited with A++ Grade

Project Report

on

"Sentiment Analysis: Amazon Fine Food Reviews"

Submitted By:

Tanishq Panwar (0901AM211066)

Pragya Bhadoriya (0901AM223D03)

Faculty Mentor:

Mr. Mir Shahnawaz Ahmad

Assistant Professor

CENTRE FOR ARTIFICIAL INTELLIGENCE

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE

GWALIOR - 474005 (MP) est. 1957

JULY-DEC. 2023

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE GWALIOR

(A Govt. Aided UGC Autonomous Institute Affiliated to RGPV, Bhopal)

NAAC Accredited with A++ Grade

CERTIFICATE

This is certified that **Tanishq Panwar (0901AM211060), Pragya Bhadoriya (0901AM223D03)** hassubmitted the project report titled **Sentiment Analysis: Amazon Fine Food Reviews** under the mentorship of **Mr. Mir Shahnawaz Ahmad**, in partial fulfilment of the requirement for the award of degree of Bachelor of Technology in **Artificial Intelligence & Machine Learning** from Madhav Institute of Technology and Science, Gwalior.

23/11/23.

Mr. Mir Shahnawaz Ahmad
Faculty Mentor
Assistant Professor
Centre for Artificial Intelligence

23/11/23

Dr. R. R. Singh
Coordinator
Centre for Artificial Intelligence

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE GWALIOR

(A Govt. Aided UGC Autonomous Institute Affiliated to RGPV, Bhopal)

NAAC Accredited with A++ Grade

DECLARATION

I hereby declare that the work being presented in this project report, for the partial fulfilment of requirement for the award of the degree of Bachelor of Technology in **AIML** at Madhav Institute of Technology & Science, Gwalior is an authenticated and original record of my work under the mentorship of **Mr. Mir Shahnawaz Ahmed**, Assistant Professor, Center for Artificial Intelligence.

I declare that I have not submitted the matter embodied in this report for the award of any degree or diploma anywhere else.

Tanishq Panwar (0901AM211060)
3rd Year 5th Sem,
Center for Artificial Intelligence

Pragya
Pragya Bhadoriya (0901AM223D03)
3rd Year 5th Sem,
Center for Artificial Intelligence

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE GWALIOR

(A Govt. Aided UGC Autonomous Institute Affiliated to RGPV, Bhopal)

NAAC Accredited with A++ Grade

ACKNOWLEDGEMENT

The full semester project has proved to be pivotal to my career. I am thankful to my institute, **Madhav Institute of Technology and Science** to allow me to continue my disciplinary/interdisciplinary project as a curriculum requirement, under the provisions of the Flexible Curriculum Scheme (based on the AICTE Model Curriculum 2018), approved by the Academic Council of the institute. I extend my gratitude to the Director of the institute, **Dr. R. K. Pandit** and Dean Academics, **Dr. Manjaree Pandit** for this.

I would sincerely like to thank my department, **Centre for Artificial Intelligence,** for allowing me to explore this project. I humbly thank **Dr. R. R. Singh**, Coordinator, Centre for Artificial Intelligence, for his continued support during the course of this engagement, which eased the process and formalities involved.

I am sincerely thankful to my faculty mentors. I am grateful to the guidance of **Mr. Mir Shahnawaz Ahmed**, Assistant Professor, and **Dr. Anshika Srivastava,** Assistant Professor Center for Artificial Intelligence for his continued support and guidance throughout the project. I am also very thankful to the faculty and staff of the department.

Tanishq Panwar (0901AM211060)
3rd Year 5th Sem,
Center for Artificial Intelligence

Pragya Bhadoriya (0901AM223D03)
3rd Year 5th Sem,
Center for Artificial Intelligence

ABSTRACT

The rapid growth of online text data has made it necessary to mine and analyze this data in order to better understand user consumption patterns and public opinion. This is especially true for user-posted online review data, which is crucial for government and business decision-making. This study uses two different models—a machine learning model and a deep learning model—to examine sentiment analysis. Comparing their performances shows that the deep learning model performs better than the ML model. The deep learning model demonstrates its better prediction capabilities with training accuracy of 0.9321 and testing accuracy of 0.8690, demonstrating its effectiveness in identifying subtle patterns within sentiment data. The purpose of this research is to enhance comprehension of the practical applications of sentiment analysis techniques and offer insightful information about the advantages and disadvantages of these models.

Keyword: Sentiment Analysis, Machine learning, Deep learning.

Table of Contents

Chapter 1: PROJECT OVERVIEW 7

1.1 Introduction 7

1.1 Motivation 7

1.2 Method based on machine learning 7

1.3 Problem Statement 8

1.4 Objectives 8

CHAPTER- 2: LITERATURE REVIEW 9

2.1 Research in Sentiment Analysis 9

2.2 Research in text classification 10

Chapter- 3: DATASET DESCRIPTION 11

2.1 Dataset collection 11

2.2 Data Pre-processing 11

Chapter – 4: METHODOLOGY 15

4.1 Exploratory Data Analysis 15

4.2 Model Training 19

4.3 Deep Learning 19

4.4 Machine Learning 20

4.5 Model Validation 20

Chapter- 5: RESULTS AND ANALYSIS 23

5.1 Result Overview 23

5.2 Result Analysis 24

5.6 Application of Model: 26

5.7 Limitations and Future Work 26

5.8 Conclusion 27

References 28

Chapter 1: PROJECT OVERVIEW

1.1 Introduction

Opinions and feelings are freely shared on a variety of online platforms in this era of digital communication, creating a wide range of public viewpoints. Online food reviews are one of the many kinds of user-generated content that can provide a wealth of information on customer attitudes. An vast and user-friendly database of reviews of dining experiences has been created by the rapid expansion of social media and specialized review sites. In this setting, sentiment analysis—a subfield of natural language processing—has emerged as a vital resource for identifying trends in the massive amount of textual data.

This investigation explores the complex field of sentiment analysis with a particular emphasis on how it applies to the domain of restaurant reviews. Businesses and researchers alike can glean insightful information from the opinions expressed in the virtual feast of online food comments by utilizing the power of computational linguistics. We will learn about the approaches, difficulties, and revolutionary possibilities of sentiment analysis in deciphering the language of taste and emotion as we set out on this voyage through the digital culinary environment.

1.2 Motivation

Businesses can improve their products and services by gaining important insights into customer satisfaction by analyzing the attitudes expressed in online reviews, social media, and other textual sources. It acts as a guardian for managing brand reputation, providing the capacity to quickly respond to unfavorable comments and magnify favorable ones. Sentiment analysis's real-time nature helps firms stay ahead of market trends and build more individualized and robust connections with their audience, which in turn fosters brand loyalty in competitive and dynamic environments. It also makes adaptive strategies easier to implement.

1.3 Method based on machine learning

Text data is subjected to well-known machine learning technologies in order to classify sentiment using machine learning techniques. The two main categories for machine learning-based sentiment classification are supervised and unsupervised learning techniques.

Supervised learning: Training manuals with labels are used in supervised learning methods. Classifying opinions has been successfully accomplished through the use of supervised learning, an efficient classification technique. Support Vector Machine (SVM), Naïve Bayes (NB), Maximum Entropy (ME), Artificial Neural Network (NN), and Decision Tree (DT) classifiers are frequently employed supervised classification techniques in sentiment analysis. A few less popular algorithms are Bayesian Network (BN), Random Forest (RF), KNNearest Neighbor (KNN), and Logistic Regression (LR).

 Unsupervised learning: Unlike supervised learning, this method trains the classifier without the use of pre-listed data. K-means and apriori algorithms are two examples of unsupervised machine learning algorithms that are more frequently used. Another way to categorize unsupervised machine learning is into associations and clusters.

1.4 Problem Statement

The problem statement for this project revolves around sentiment analysis, a crucial aspect of Natural Language Processing. Sentiment analysis involves determining the emotional tone behind words to gain an understanding of the attitudes, opinions, and emotions expressed within an online mention. It is a vast source of information in the era of online communication. The challenge lies in accurately classifying the sentiments expressed in a dataset of reviews. The complexity of human language makes this a non-trivial task as the models need to understand the context, tone, and intent behind the words. The ultimate goal is to develop a solution that can accurately classify sentiments, providing valuable insights from the text data.

1.5 Objectives

To understand and extract the sentiments expressed in a dataset of reviews.
To develop a solution that can accurately classify these sentiments.
To provide valuable insights from the text data that can be used for further analysis or decision-making.

CHAPTER- 2: LITERATURE REVIEW

The majority of prior research on the value of reviews has focused on automating the estimation of reviews' utility in order to address issues such as the large volume and variable caliber of online customer reviews. There has been a great deal of research using text mining and statistical modeling techniques to examine the relationship between review helpfulness and text-based attributes. The study by Salehan et al. [1] looked at the effect of sentiment on the value of online user evaluations.

The study of Elzayady H. looked at how sentiment polarity affected the value of online reviews and found some interesting things. Their study methodology included both metrics related to the title and the review. The other main elements of the model, apart from the evaluations' polarity and attitude, were the reviews' duration and length. On websites such as Amazon, reviews were arranged by "most useful" instead of "most recent". Longevity was a suitable addition to the study's criteria. Review polarity was found to have a significant impact on helpfulness when compared to positive or negative evaluations.[2].

The study of Le Q. [3] shows the review's duration was another crucial factor, since a longer review would be expected for more details. They investigated how online reviews affected the quantity of helpfulness votes they received using data from CNET Download.com. The study used Lift Ratio and ALC (Akaike's Information Criterion), with the Ordinal Logistic Regression model being used to measure accuracy.

Sentiment analysis using natural language processing (NLP) can be used, for instance, to extract hedonics from food-related comments or to comprehend individual differences. Comments like "This chocolate was ok" or "This tea was absolutely delicious" are a couple examples. "This was terrible! I guess it's not too bad, but I probably won't buy it again." I would rate this with a zero! Gross! doesn't require manual tagging to be divided into positive, neutral, and negative categories, respectively [4].

A brief overview given by (D.Kawade & Oza, 2017) of the various kinds of algorithms used in sentiment analysis is provided. Sentimental analysis is the study of the subjectivity, opinions, thoughts, and sentiments expressed in a text. Sentiment analysis techniques and recently introduced algorithms are discussed, along with the significance of certain fields like resource construction, transfer learning, and feelings detection. This survey's primary goal is to categorize recent publications; 54 sentiment-analysis-based articles that were just published were categorized and summarized [5].

Sentiment analysis has several uses, such as in the consumer sector, where it uses natural language processing to automatically classify and extract the emotion from the text. Another novel machine learning technique that has surfaced is transfer education, which makes use of preexisting knowledge to solve problems and predict outcomes. It also includes the possibility of emotive analysis, such as the application of shift learning characteristics across domains that have not yet been completely studied, which makes it extremely challenging to use transfer learning to solve negative problems with text data. They come to the conclusion that aspect-level sentiment analysis for brief texts is thought to be the most promising research method going forward. (R. Liu et al., 2019) [6].

Machine learning algorithms like Naive Bayes, Support vector machine, and Maximum entropy classifier algorithms can be used on the sentimental analysis of huge data. Using these techniques, a huge volume of data can be utilized to get optimized and strategical decisionmaking capability. sentimental analysis is also called opinion mining which gives a brilliant and humanlike brilliance which analyzes and respond emotions, the user show in social media like Facebook, Yammer, Twitter, microblogs which provide a tremendous amount of data every day in textual or numerical forms and these are classified as structured, semi-structured and non-structured and then later they are categorized as positive, negative and common based on user's attitude towards a particular topic for analysis purpose (Naiknaware et al., 2017) [7].

Sentiment analysis helps to do a review of the movies, product, and customer opinion on products. The role of sentiment analysis in natural language processing is to remove positive or negative polarities from social media messages. Digital social networks are growing increasingly, and culture focused on online media has affected young scientists in their research in the study of opinion. Organizations that are actually keen to determine their clients or the public opinion on their social media goods. Internet services should be able to test social media data on blogs, web forums, articles, tweets and user feedback (J. Singh et al., 2017) [8].

2.1 Research in text classification

This section offers some insight into the text mining and natural language processing work being done in the field in addition to the review's usefulness. The characteristics extracted from text using natural language processing (NLP) were compared and contrasted in different ways, and spam detection was classified as either "spam" or "not spam." Many lexical and semantic variables, including Bag of Words, Term Frequency, POS tagging, and others, were used in the study. The approaches of Nave Bayes, Support Vector Machine, and Logistic Regression were used to solve the classification problem [9]. Even though they weren't always better than the other tested algorithms, SVM performed better than Logistic Regression and Naive Bayes. When multiple attributes were combined, the performance also improved.

Chapter- 3: DATASET DESCRIPTION

3.1 Dataset collection

The "Amazon Fine Food Reviews" dataset, which can be found on Kaggle, is an extensive compilation of customer feedback about food items that are sold on the Amazon website. This dataset, which includes a wide variety of goods like snacks, drinks, and gourmet items, is especially concentrated on fine meals. The data offers a longitudinal view of customer sentiments because it covers a sizable time period, from October 1999 to October 2012.

A plethora of data is included in the dataset, including product ratings, helpfulness scores, user reviews, and metadata like timestamps and product IDs. Details like the reviewer's ID, the product's ID, the review summary, and the review text are all included with every review. Numerous applications, such as sentiment analysis, natural language processing, and the investigation of consumer preferences in relation to fine foods, are well-suited for this vast and comprehensive dataset. This dataset can be used by academics and data scientists to get insights into consumer attitudes, product trends, and factors that affect customer satisfaction in the context of Amazon's gourmet cuisine offerings.

3.2 Data Pre-processing

The preprocessing technique is more useful to identify and remove meaningless, noisy, and incompatible data. We attempt to comprehend the data by removing unnecessary columns and filtering all neural reviews.

1. Text preprocessing is highly essential in working with text classification problems. It helps in improving the computational efficiency and avoiding the overfitting problem by eliminating the noisy features. We discuss the various text preprocessing methods that are applied before training the model in the next sections.

```
def decontract(text):
    text = re.sub(r"won\'t", "will not", text)
    text = re.sub(r"can\'t", "can not", text)
    text = re.sub(r"n\'t", " not", text)
    text = re.sub(r"\'re", " are", text)
    text = re.sub(r"\'s", " is", text)
    text = re.sub(r"\'d", " would", text)
    text = re.sub(r"\'ll", " will", text)
    text = re.sub(r"\'t", " not", text)
    text = re.sub(r"\'ve", " have", text)
    text = re.sub(r"\'m", " am", text)
    return text
```

Fig 3.2.1 – Function For Text Preprocessing

```python
def preprocess_text(review):
    review = re.sub(r"http\S+", "", review)                # removing website links
    review = BeautifulSoup(review, 'lxml').get_text()      # removing html tags
    review = decontract(review)                            # decontracting
    review = re.sub("\S*\d\S*", "", review).strip()        # removing the words with numeric digits
    review = re.sub('[^A-Za-z]+', ' ', review)             # removing non-word characters
    review = review.lower()                                # converting to lower case
    review = [word for word in review.split(" ") if not word in stop_words] # removing stop words
    review = [lemmatizer.lemmatize(token, "v") for token in review] #Lemmatization
    review = " ".join(review)
    review.strip()
    return review
df['Text'] = df['Text'].apply(lambda x: preprocess_text(x))
```

Fig 3.2.2 – Function For Text Preprocessing

2. Tokenization Tokenization is basically the process of splitting the sentences into words. Each word is then considered as a separate token. This is the first step in text preprocessing on which all the subsequent steps are built (Hotho et al., 2005).

3. Converting uppercase tokens into lowercase Since the semantics of a word or a phrase do not depend upon which case the word is written in, the uppercase words are converted to lowercase to avoid potential duplication of the words. For example, the words ‘DOG’ and ‘dog’ convey the same meaning. This conversion helps in reducing the dimensionality of the feature set.

4. Removal of punctuation marks In this step, the punctuation marks are removed from the text as they do not provide any extra information while extracting the semantics from the text.

5. Stop words removal: In natural language processing, stop words are the most commonly used words which do not convey much meaning. Some of the examples of the stop words include short 26 function words such as ‘a’, ‘an’, ‘the’, ‘is’, ‘are’, ‘which’, ‘at’ and, ‘on’. ‘NLTK’ library of python is used in this study to remove stop words.

```python
stop_words= set(['br', 'the', 'i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you're", "you've",\
            "you'll", "you'd", 'your', 'yours', 'yourself', 'yourselves', 'he', 'him', 'his', 'himself', \
            'she', "she's", 'her', 'hers', 'herself', 'it', "it's", 'its', 'itself', 'they', 'them', 'their',\
            'theirs', 'themselves', 'what', 'which', 'who', 'whom', 'this', 'that', "that'll", 'these', 'those', \
            'am', 'is', 'are', 'was', 'were', 'be', 'been', 'being', 'have', 'has', 'had', 'having', 'do', 'does', \
            'did', 'doing', 'a', 'an', 'the', 'and', 'but', 'if', 'or', 'because', 'as', 'until', 'while', 'of', \
            'at', 'by', 'for', 'with', 'about', 'against', 'between', 'into', 'through', 'during', 'before', 'after',\
            'above', 'below', 'to', 'from', 'up', 'down', 'in', 'out', 'on', 'off', 'over', 'under', 'again', 'further',\
            'then', 'once', 'here', 'there', 'when', 'where', 'why', 'how', 'all', 'any', 'both', 'each', 'few', 'more',\
            'most', 'other', 'some', 'such', 'only', 'own', 'same', 'so', 'than', 'too', 'very', \
            's', 't', 'can', 'will', 'just', 'don', "don't", 'should', "should've", 'now', 'd', 'll', 'm', 'o', 're', \
            've', 'y', 'ain', 'aren', "aren't", 'couldn', "couldn't", 'didn', "didn't", 'doesn', "doesn't", 'hadn',\
            "hadn't", 'hasn', "hasn't", 'haven', "haven't", 'isn', "isn't", 'ma', 'mightn', "mightn't", 'mustn',\
            "mustn't", 'needn', "needn't", 'shan', "shan't", 'shouldn', "shouldn't", 'wasn', "wasn't", 'weren', "weren't", \
            'won', "won't", 'wouldn', "wouldn't"])
```

Fig 3.2.3– Stopwords

5 Lemmatization : Lemmatization is a text normalization process of reducing the inflectional form of words into base or dictionary form which is called 'lemma'. For example, lemmatization converts the word 'ran' into its base form 'run'. In lemmatization, complete morphological analysis of words is done to ensure that the base word belongs to the dictionary (Manning et al., 2008). In this way, lemmatization has a slight edge over its counterpart 'stemming' method which usually removes the prefix or suffix associated with the word. The root word in stemming is not required to be a valid word from the language. Keeping this in mind, lemmatization will be used in this study to reduce the inflectional form of words. This study uses 'TextBlob' library of python programming language for performing lemmatization.

Chapter – 4: METHODOLOGY

The methodology for this project involves preparing the data, selecting and training the models for sentiment analysis, and then evaluating their performance to ensure they can accurately classify sentiments in the dataset.

Fig. 4.0.1 - Flow Chart of Machine learning & Deep Learning based sentiment analysis technique

4.1 Exploratory Data Analysis

This section discusses the data manipulation performed before training the model and also describes the goal of this study is to build a model which predicts whether the review is helpful or not, the reviews which did not get any vote - either helpful or not helpful, will not be considered in the study. This reduces the number of observations from the original 568,454 to 298,402. Now, the next step is to examine the distribution of the attributes 'positive' and 'Score' by looking at their descriptive statistics. After resampling :

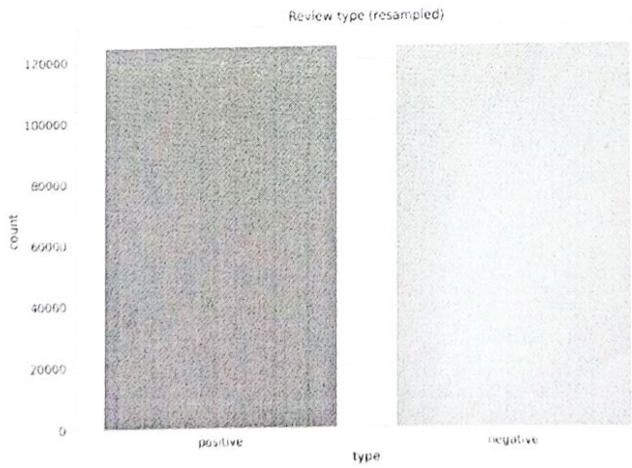

Fig 4.1.1 – Review Type After Resampling

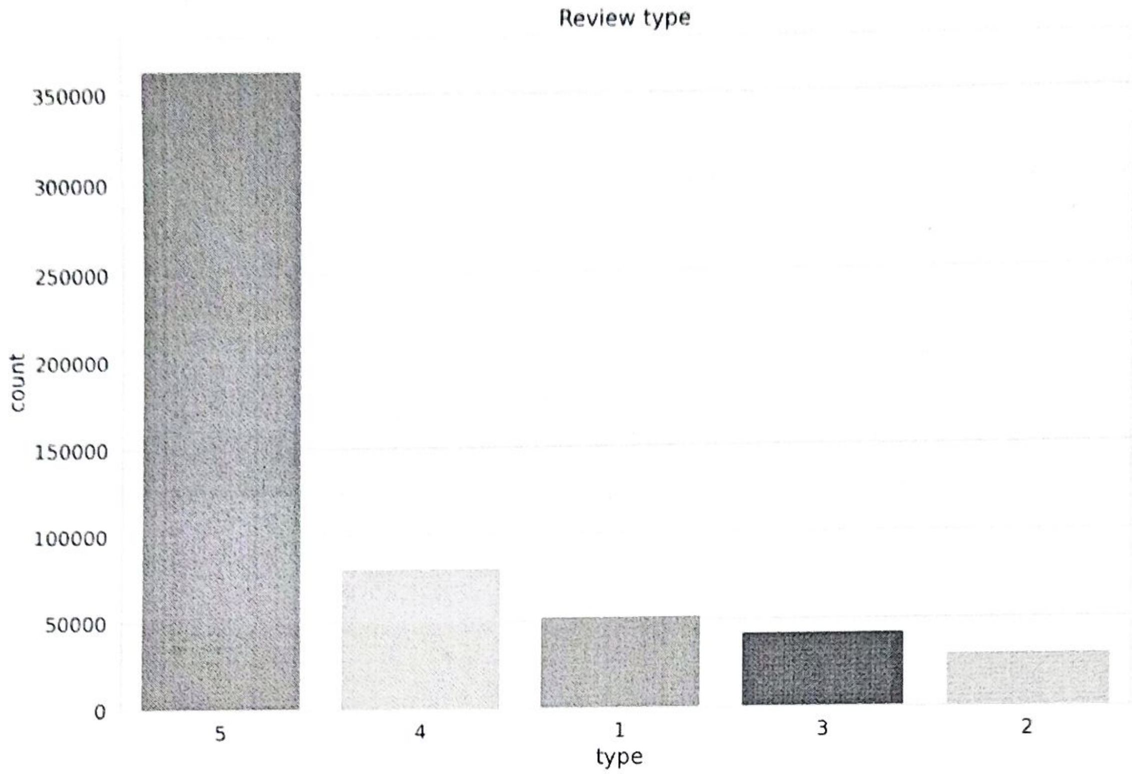

Fig 4.1.2 - Count Plot for Review types from 1 to 5

Now we will be visualizing the most frequent words for positive and negative sentiments using word clouds. Word clouds provide a simple and effective way to understand the main themes or concepts within a body of text. By creating separate word clouds for positive and negative sentiments, we can gain insights into the specific words that are commonly associated with each sentiment in our dataset. This can help us better understand the nature of the sentiments expressed in the reviews. The images of the word clouds will be presented next.

Fig 4.1.3 - Word Cloud for the Positive Sentiments

Fig 4.1.4 - Word Cloud for the Negative Sentiments

Next, we will be visualizing the distribution of review scores in our dataset. This will be done using a pie chart, which is an effective way to represent the proportion of each score in the total number of reviews. The pie chart will show the percentage of reviews that have received a particular star rating, providing a clear picture of the overall sentiment of the reviews in our dataset. The image of the pie chart will be presented next.

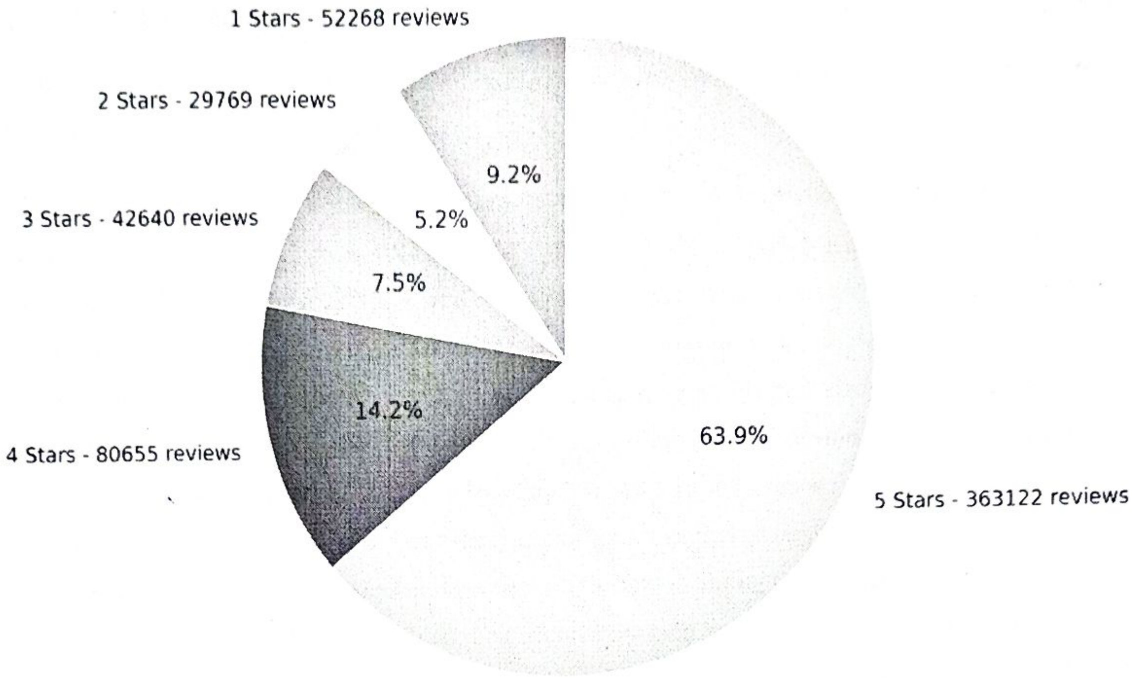

Fig 4.1.5 – Pie Chart for Reviews

4.2 Model Training

Two models are employed in this study: A Deep Learning model and an XGBoost model, which is a Machine Learning model.

The Deep Learning model is built using Keras, a high-level neural networks API. It is a Sequential model, which is a linear stack of layers. The model consists of Dense layers with ReLU activation functions, Dropout layers to prevent overfitting, and a final Dense layer with a softmax activation function for multi-class classification. The model is compiled with the Adam optimizer and the categorical cross-entropy loss function. It is trained for a specified number of epochs with a certain batch size, and a fraction of the training data is used for validation.On the other hand, the XGBoost model is a decision-tree-based ensemble Machine Learning algorithm that uses a gradient boosting framework. It is renowned for its execution speed and model performance. This model is particularly useful for structured or tabular data. XGBoost implements the gradient boosting decision tree algorithm, which works by combining the predictions from multiple decision trees.Both models are evaluated based on their accuracy, which is the proportion of correct predictions made by the models. The choice of these two models allows for a comprehensive understanding of the data and ensures robust predictions.

4.2.1 Deep Learning

The model is built using Keras, a high-level neural networks API. It is a Sequential model, which is a linear stack of layers that you can easily create by passing a list of layer instances to the constructor.

The first layer is a Dense layer with 256 units and a Rectified Linear Unit (ReLU) activation function. The input shape corresponds to the number of features in the training data. This layer is followed by a Dropout layer with a rate of 0.2, which randomly sets a fraction of the input units to 0 at each update during training time, helping to prevent overfitting.

The next layer is another Dense layer with 128 units and a ReLU activation function, followed by another Dropout layer with a rate of 0.2.

The final layer is a Dense layer with 2 units and a softmax activation function, which is often used for multi-classification problems. The model is compiled with the Adam optimizer with a learning rate of 0.001, and the categorical cross-entropy loss function, which is suitable for multi-class classification.

The model is trained for 15 epochs with a batch size of 16. The validation split is set to 0.1, meaning 10% of the training data is used for validation. The 'shuffle' parameter is set to True, which means the training data is shuffled at each epoch.The model's performance is evaluated based on its accuracy, which is the proportion of correct predictions made by the model.

Model: "sequential"

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 256)	131328
dropout (Dropout)	(None, 256)	0
dense_1 (Dense)	(None, 128)	32896
dropout_1 (Dropout)	(None, 128)	0
dense_2 (Dense)	(None, 2)	258

Total params: 164482 (642.51 KB)
Trainable params: 164482 (642.51 KB)
Non-trainable params: 0 (0.00 Byte)

Fig 4.2.1 – Model Summary of Deep Learning

4.2.2 Machine Learning

The XGBoost model is implemented as follows:

First, the features are extracted from the training and testing data using the previously trained Deep Learning model. The predict method of the model is used to generate these features.

Next, the XGBoost classifier is initialized. XGBoost, short for eXtreme Gradient Boosting, is an optimized distributed gradient boosting library designed to be highly efficient, flexible, and portable. It implements machine learning algorithms under the Gradient Boosting framework.

The XGBoost model is then trained on the extracted features from the training data and the corresponding labels. The fit method of the XGBoost model is used for this purpose.

Finally, the trained XGBoost model is used to predict the labels for the test data. The predict method of the XGBoost model is used to generate these predictions.

This methodology allows the model to leverage the feature extraction capabilities of the Deep Learning model and the predictive power of the XGBoost model, providing a robust solution for the task at hand. The model's performance can be evaluated by comparing the predicted labels with the actual labels.

4.3 Model Validation

This section discusses model validation using the train and test sets and sheds light on the model performance measurement in order to perform model comparison. This study uses the hold-out method to split the original dataset into train and test sets. The dataset is shuffled and divided into train and test sets with the ratio 9:1 for all the approaches used in the study. The train_test_split function is used to split the review data into training and test sets. The embed_reviews and embed_reviews_test functions are used to convert the reviews into numerical representations, or embeddings, which can be used as input for machine learning models. These embeddings are generated in batches for efficiency. The resulting X_train and X_test variables contain the embeddings for the training and test reviews, respectively, and are used for training and evaluating the models.

This study deals with the binary classification of the review texts. Hence, the potential performance measures for this study include accuracy, precision, recall, F1 Score, classification report and area under ROC (Receiver Operating Characteristics) curve. In a binary classification setting, correction of classification can be evaluated using the confusion matrix which is shown in Figure 4.3.1

Actual class	**Predicted as positive**	**Predicted as negative**
Positive	True positive (tp)	False negative (fn)
Negative	False positive (fp)	True negative (tn)

Fig 4.3.1 – Confusion Matrix

Confusion matrix Actual class Predicted as positive Predicted as negative Positive True positive (tp) False negative (fn) Negative False positive (fp) True negative (tn) The parameters of the confusion matrix can be described as below:

1) True positive – sample belonging to positive class predicted as positive

2) True negative – sample belonging to negative class predicted as negative

3) False positive – sample belonging to negative class predicted as positive

4) False negative- sample belonging to positive class predicted as negative

This explanation of the parameters provides a basis for understanding the evaluation purposes of different performance measures used for performing binary classification. Since the goal of this study is to extract and label reviews based on their usefulness, it is not affordable to misclassify the review of either class into its opposite class. Moreover, it is also imperative to evaluate the overall performance of the classification models. Hence, the performance measures used in this study are overall accuracy and AUC (area under the curve).

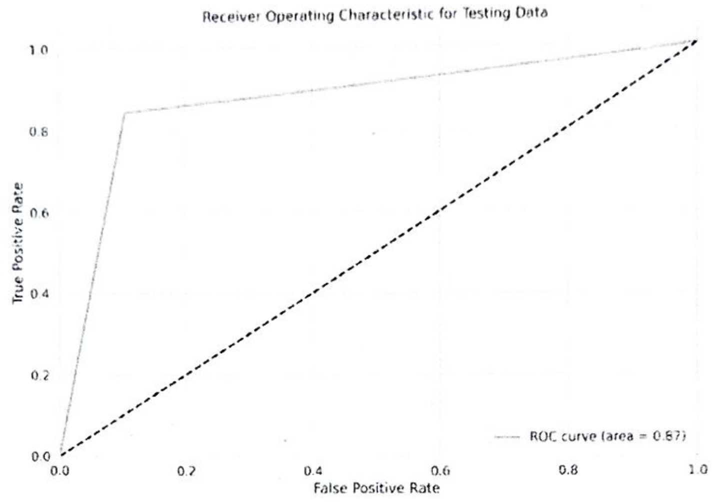

Fig 4.3.2 - ROC curve for Deep Learning

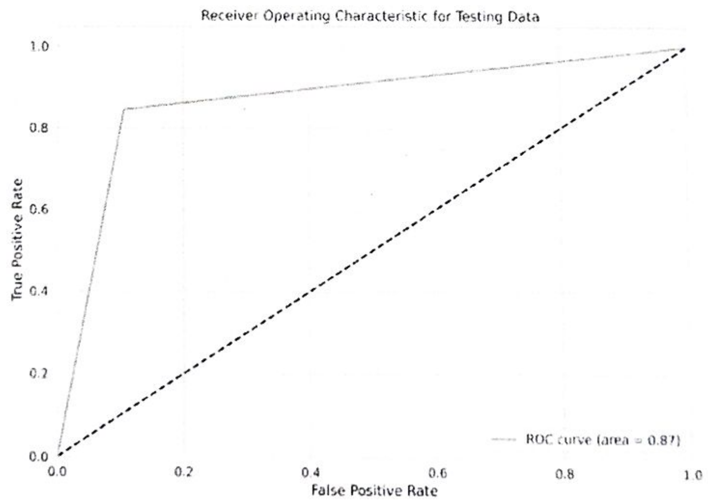

Fig 4.3.3 -ROC curve for Machine Learning

Chapter- 5: RESULTS AND ANALYSIS

5.1 Result Overview

Based on accuracy testing, the deep learning model outperformed the XGBoost model, which had an accuracy of about with an accuracy of 87%. The slight variation in accuracy between the two models indicates that the deep learning model performs marginally better in categorizing the sentiments expressed in food reviews.

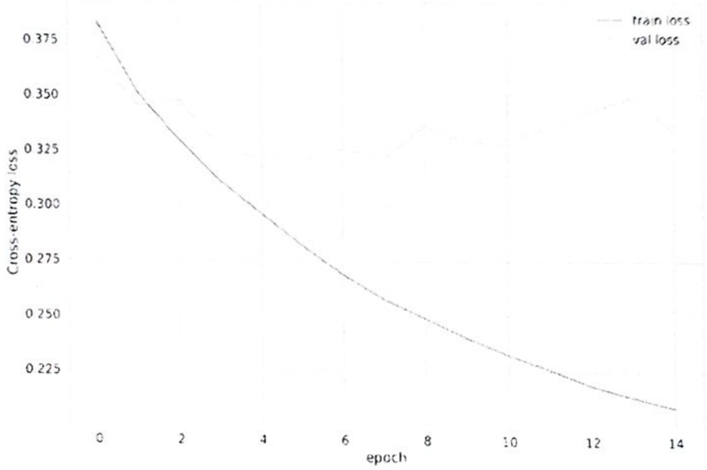

Fig 5.1.1 – Validation Loss Graph

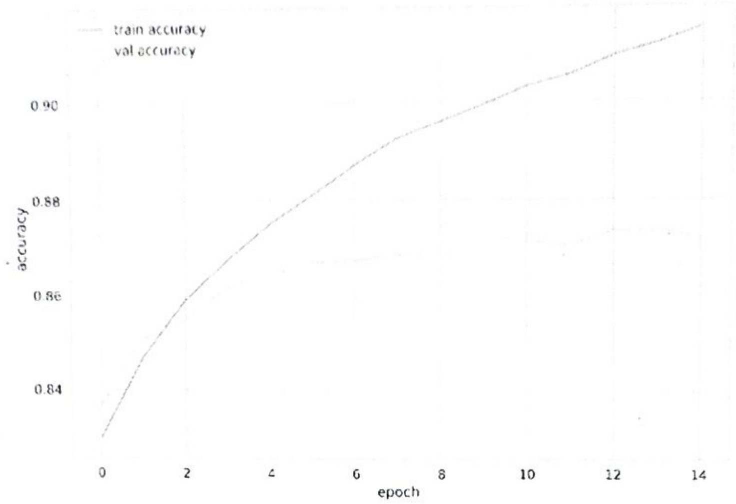

Fig 5.1.1 – Accuracy Graph

5.2 Result Analysis

Classification Report for testing data:

	precision	recall	f1-score	support
0	0.85	0.89	0.87	12473
1	0.89	0.84	0.87	12463
accuracy			0.87	24936
macro avg	0.87	0.87	0.87	24936
weighted avg	0.87	0.87	0.87	24936

Fig 5.2.1 – Classification Report for Testing Data for Deep Learning Model

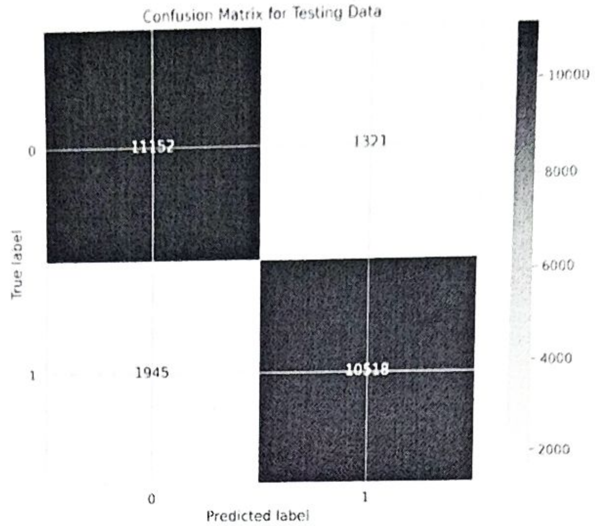

Fig 5.2.2 – Confusion Matrix for Testing Data for Deep Learning Model

Classification Report for testing data:

	precision	recall	f1-score	support
0	0.86	0.88	0.87	12473
1	0.88	0.86	0.87	12463
accuracy			0.87	24936
macro avg	0.87	0.87	0.87	24936
weighted avg	0.87	0.87	0.87	24936

Fig 5.2.3 – Classification Report for Testing Data for Machine Learning Model

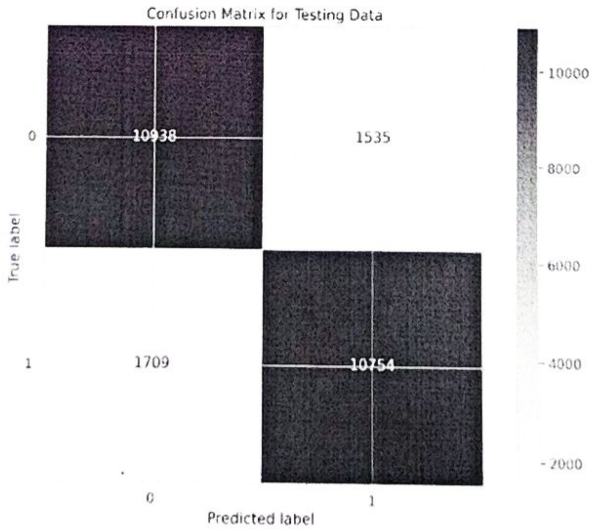

Fig 5.2.4 – Confusion Matrix for Testing Data for Machine Learning Model

5.3 Application of Model:

The model can be applied in various domains such as sentiment analysis, spam detection, and topic classification. It can help businesses understand customer feedback, filter out irrelevant information, and categorize data for easy analysis. Additionally, it can be used in social media monitoring, customer service, and market research to gain insights from large volumes of text data.

5.4 Limitations and Future Work

Limitations: The model, while effective, has certain limitations. It assumes that the input data is clean and well-structured, which may not always be the case in real-world scenarios. The performance of the model can be affected by the quality of the data. Additionally, the model might not perform as well on data that is significantly different from the training data.

Future Work: In the future, the model could be improved by incorporating more sophisticated preprocessing techniques to handle unstructured and noisy data. Additionally, exploring other architectures or models could potentially improve the performance. Incorporating more features or using ensemble methods could also be beneficial. Lastly, the model could be made more robust by training it on a more diverse dataset.

5.5 Conclusion

In conclusion, the study successfully applied Deep Learning and XGBoost models for sentiment analysis. The models were trained and tested on a dataset of reviews, with the Deep Learning model achieving the highest accuracy. This demonstrates the effectiveness of Deep Learning models in understanding and classifying sentiments in text data. However, it's important to note that the performance of these models depends on the quality and diversity of the training data. Future work could explore more sophisticated preprocessing techniques, different model architectures, and ensemble methods to further improve the performance. Despite the challenges, the results of this study are promising and provide a strong foundation for future research in this area.

References

1. Malik M. S. I, Hussain, A. Helpfulness of product reviews as a function ofdiscrete positive and negative emotions. Computers in Human Behavior, 2017, 73, pp. 290-302.
2. Elzayady H, Badran K. M, Salama G. I.). Sentiment analysis on twitter datausing apache spark framework. In 2018 13th International Conference on Computer Engineering and Systems (ICCES). 2018, December. IEEE. pp. 171-176.
3. Le Q, and Mikolov, T. Distributed representations of sentences and documents. In International conference on machine learning. 2014, June,pp. 1188-1196.
4. Reed, D.R.; Mainland, J.D.; Arayata, C.J. Sensory Nutrition: The Role ofTaste in the Reviews of Commercial Food Products. *Physiol. Behav.* **2019**, *209*, 112579. [Google Scholar] [CrossRef]
5. Kawade, D., & Oza, K. (2017). Sentiment Analysis: Machine LearningApproach. International Journal of Engineering and Technology, 9(1), 2183–2186. DOI: https://doi.org/10.21817/ijet/2017/v9i3/1709030151
6. Liu, R., SHI, Y., JI, C., & JIA, M. (2019). A Survey of Sentiment AnalysisBased on Transfer Learning. IEEE Access, 7(1), 85401-85412. DOI: https://doi.org/10.1109/ACCESS.2019.2925059
7. Naiknaware, B., Kushwaha, B., & Kawathekar, S. (2017). Social MediaSentiment Analysis using Machine Learning Classifiers. International Journal of Computer Science and Mobile Computing, 6(6), 465-472.
8. Singh, J., Singh, G., & Singh, R. (2017). Optimization of sentiment analysisusing machine learning classifiers. Human-Centric Computing and Information Sciences, 7(32), 1-7. DOI: https://doi.org/10.1186/s13673-017-0116-3
9. Yu X, Liu Y, Huang X, An, A. Mining online reviews for predicting salesperformance: A case study in the movie domain. IEEE Transactions on Knowledge and Data Engineering, 2010, 24(4), 720-734.