Twitter Sentiment Analysis for Restaurant Reviews

Project Report submitted in the partial fulfilment of
Bachelor of Technology in
Information Technology

Submitted To

SVKM's NMIMS,
Mukesh Patel School of Technology Management & Engineering,
Shirpur Campus (M.H.)

Submitted By:

Shibin Nair 70012100205
Aryan Gupta 70012100215
Sujal Jain 70012100224

Under The Supervision Of:

Prof. Sonia Relan
(Asst. Professor, Information Technology)

INFORMATION TECHNOLOGY DEPARTMENT
Mukesh Patel School of Technology Management & Engineering
ACADEMIC SESSION: 2024-25

CERTIFICATE

This is to certify that the project entitled "**Twitter Sentiment Analysis for Restaurant Reviews"** has been done by

Shibin Nair 70012100205
Aryan Gupta 70012100215
Sujal Jain 70012100224

under my guidance and supervision & has been submitted partial fulfilment of the degree of "**Bachelor of Technology** in **Information Technology"** of MPSTME, SVKM'S NMIMS (Deemed-to-be University), Shirpur (M.H.), India.

______________ ______________

Project Mentor **Examiner**

(Prof. Sonia Relan)

Date:

Place: Shirpur

____

H.O.D.

INFORMATION TECHNOLOGY DEPARTMENT
Mukesh Patel School of Technology Management & Engineering

ACKNOWLEDGEMENT

After the completion of this capstone Project work, words are not enough to express my feelings about all those who helped me to reach my goal; feeling above this is my indebtedness to The Almighty for providing me this moment in life.

It's a great pleasure and moment of immense satisfaction for me to express my profound gratitude to **Prof. Sonia Relan**, **Assistant Professor**, Information Technology Department, MPSTME, Shirpur, whose constant encouragement enabled me to work enthusiastically. Their perpetual motivation, patience and excellent expertise in discussion during progress of the project work have benefited me to an extent, which is beyond expression. Their depth and breadth of knowledge of Information Technology field made me realize that theoretical knowledge always helps to develop efficient operational software, which is a blend of all core subjects of the field. I am highly indebted to them for their invaluable guidance and ever-ready support in the successful completion of this project in time. Working under their guidance has been a fruitful and unforgettable experience.

We express my sincere thanks and gratitude to **Dr. Ritesh Dhanare**, Head of Department, Information Technology Department, MPSTME, Shirpur, for providing necessary infrastructure and help to complete the project work successfully.

We also extend my deepest gratitude to **Dr. Venkatadri Marriboyina**, Associate Dean, SVKM'S NMIMS, Shirpur Campus for providing all the necessary facilities and true encouraging environment to bring out the best of my endeavors.

We sincerely wish to express my grateful thanks to all members of the staff of Information Technology department and all those who have embedded me with technical knowledge of computer technology during various stages of B.Tech. Information Technology.

We would like to acknowledge all my friends, who have contributed directly or indirectly in this capstone Project work.

The successful completion of a capstone Project is generally not an individual effort. It is an outcome of the cumulative effort of a number of persons, each having their own importance to the objective. This section is a vote of thanks and gratitude towards all those persons who have directly or indirectly contributed in their own special way towards the completion of this project.

Shibin Nair

Aryan Gupta

Sujal Jain

ABSTRACT

In today's digital age, customers increasingly rely on online reviews to guide their dining choices. Platforms like Twitter provide a wealth of user-generated content, offering valuable insights into restaurant experiences. Sentiment analysis has emerged as a crucial tool for extracting actionable insights from this unstructured data, enabling customers to quickly assess the overall sentiment surrounding a restaurant.

This project focuses on developing an automated sentiment analysis system that processes restaurant reviews from Twitter, classifying them as positive, negative, or neutral. By employing a hybrid approach that integrates advanced natural language processing models, specifically RoBERTa and Electra, the system efficiently determines the sentiment of each review, providing customers with a straightforward understanding of public opinion. This streamlined classification allows customers to make quick and informed dining decisions without needing to go through large volumes of text.

Beyond its immediate application, the sentiment analysis system also serves as a powerful tool for identifying broader trends and patterns in customer opinions over time. By aggregating sentiment data, the system can reveal shifts in public perception, highlight emerging preferences, and identify seasonal variations in customer feedback. Such insights can empower customers to choose dining experiences that align with popular trends or avoid those with declining satisfaction levels.

Furthermore, the sentiment analysis system can be expanded to build a comprehensive dining recommendation platform that uses collective sentiment scores, allowing customers to discover new and trending restaurants based on positive sentiment clusters. This enhances the overall customer experience by not only helping them make better dining decisions but also providing them with new dining opportunities tailored to current sentiment trends.

In conclusion, this sentiment analysis system bridges the gap between customers and the vast amount of online feedback, offering a user-friendly solution that enhances their overall dining experience by transforming unstructured data into actionable insights. By providing real-time sentiment insights and long-term trend analysis, the system significantly enriches customers' decision-making processes, helping them explore and select restaurants that best match their preferences.

TABLE OF CONTENTS

Topics	Page
List of Figures	vi
List of Tables	vii
Chapter 1 INTRODUCTION	
1.1 Background of the project	1
1.2 Motivation and scope of the report	2
1.3 Problem statement	3
1.4 Salient Contribution	3
1.5 Organization of Report	4
Chapter 2 LITERATURE SURVEY	
2.1 Introduction	6
2.2 Exhaustive Literature Survey and Gap Identification	6
2.3 Issues and Future Scope	18
Chapter 3 METHODOLOGY AND IMPLEMENTATION	
3.1 About Dataset	20
3.2 Design Approach	23
3.3 Tools and Technologies Used	26
3.4 Flowchart	26
Chapter 4 RESULT AND ANALYSIS	
4.1 Model Performance Metrics	28
4.2 Error Analysis and misclassification	30
4.3 Comparative analysis	32
4.4 Conclusion of analysis	38
Chapter 5 ADVANTAGE, LIMITATIONS AND APPLICATIONS	
5.1 Advantages	34
5.2 Limitations	34
5.3 Applications	35
CONCLUSION AND FUTURE SCOPE	
6.1 Conclusion	37
6.2 Future Scope	38
References	41
Appendix A: Sample code	44
Appendix B: Data sheets	49

LIST OF FIGURES

Sr. No.	Figure No.	Name of Figures	Page
1	1	Overview of Hybrid Sentiment Analysis Workflow	25
2	2	Detailed Sentiment Classification Process Flow	27
3	3	Classification report Electra	28
4	4	Classification report RoBERTa	29
5	5	Classification report Hybrid	30

LIST OF TABLES

Sr. No.	Table No.	Name of Table	Page
1	1	Literature Review Summary	16
2	2	Literature Review Summary	16
3	3	Literature Review Summary	17
4	4	Literature Review Summary	17
5	5	Performance Summary	30
6	6	Comparative Analysis	32

Chapter 1
Introduction

1.1 Background of the project topic

In today's digital landscape, social media platforms have become vital channels for consumers to express their opinions, share experiences, and influence others' decisions. Among these platforms, Twitter stands out as a dynamic space where individuals frequently post real-time reviews and feedback about a variety of services, including restaurants. These reviews often serve as a primary source of information for prospective diners who seek guidance on where to dine based on the collective experiences of others.

The surge in user-generated content on Twitter provides a rich source of insights into public sentiment. However, the sheer volume and unstructured nature of this data pose a significant challenge for customers who want to make informed dining choices. Manually reading through numerous reviews is not only time-consuming but also ineffective, as it often leads to subjective biases and inconsistent interpretations.

To address this challenge, sentiment analysis has emerged as a powerful tool capable of processing and classifying large datasets of customer reviews. By automating the analysis of these reviews, sentiment analysis systems can efficiently identify whether feedback is positive, negative, or neutral. This offers customers a streamlined way to understand overall opinions about a restaurant, allowing them to make quick and informed decisions.

Despite the progress in sentiment analysis, existing tools are often designed with businesses in mind, focusing on improving services based on feedback rather than directly assisting customers. This project aims to fill that gap by developing a customer-centric sentiment analysis system specifically tailored to analyse restaurant reviews on Twitter. By providing customers with an efficient and accurate way to interpret the collective sentiment of others, this project enhances their ability to make confident, data-driven dining choices.

1.2 Motivation and scope of the report

The rise of social media as a platform for sharing real-time feedback has transformed the way customers interact with businesses, particularly in the restaurant industry. As customers increasingly turn to platforms like Twitter to voice their dining experiences, the abundance of user-generated content presents both an opportunity and a challenge. On one hand, the wealth of information can provide invaluable insights into restaurant quality, customer satisfaction, and overall dining trends. On the other hand, the sheer volume and unstructured nature of this data make it difficult for individual customers to efficiently process and derive meaningful conclusions.

Existing sentiment analysis tools and systems are typically geared towards businesses, helping them refine services or enhance marketing strategies based on customer feedback. However, there remains a significant gap when it comes to tools that focus on benefiting customers directly. The motivation behind this project is to bridge that gap by developing a sentiment analysis system designed specifically to empower customers. By providing real-time insights into restaurant reviews, this tool aims to simplify the decision-making process, allowing customers to make informed and confident dining choices based on the collective sentiment of others. The drive to create a customer-centric solution reflects the growing importance of data-driven decision-making in everyday consumer behavior.

The scope of this project involves the development and implementation of a sentiment analysis system that processes and classifies restaurant reviews on Twitter into three categories: positive, negative, and neutral. By utilizing a hybrid approach that combines advanced transformer models like RoBERTa and Electra, the project aims to achieve high accuracy in sentiment classification while optimizing processing efficiency. The system is designed to work in real-time, ensuring that customers receive the most up-to-date information on restaurants.

The project focuses on building a scalable and intuitive tool that is easy for customers to use, providing them with direct, actionable insights to enhance their dining experiences. Initially, the system will focus on analyzing English-language tweets related to restaurant reviews; however, the architecture is designed with the flexibility to accommodate multilingual support in the future. Additionally, the tool can be extended to other social media platforms beyond Twitter, broadening its applicability and impact.

By focusing on customer needs and providing a streamlined, efficient solution for interpreting online reviews, this project not only enhances individual dining decisions but also sets the foundation for future development in customer-oriented sentiment analysis tools.

1.3 Problem statement

In the age of social media, customers increasingly turn to platforms like Twitter to gather information before making dining decisions. Twitter hosts a vast and ever-growing volume of user-generated reviews about restaurants, making it a valuable resource for understanding public opinion. However, while this abundance of information offers a wealth of insights, it also presents a significant challenge: the sheer volume and unstructured nature of these tweets can be overwhelming for individual customers seeking quick, meaningful information.

Manually browsing through numerous tweets to form an opinion about a restaurant is not only time-consuming but also prone to inconsistencies, as the informal nature of social media language poses additional difficulties. Customers frequently use slang, abbreviations, and emotive expressions, making it challenging to accurately interpret the overall sentiment of a review. Sarcasm and mixed sentiments further complicate the task, as they often require contextual understanding that goes beyond surface-level analysis.

As a result, customers face barriers when trying to quickly and reliably assess the quality of a restaurant based on Twitter reviews. There is a need for a solution that automates this process, efficiently analysing and categorizing sentiments to provide clear, actionable insights. This project aims to address this problem by developing a sentiment analysis system that processes and interprets restaurant reviews on Twitter, enabling customers to make informed and confident dining choices with ease.

1.4 Salient contribution

This project makes several important contributions in the domain of sentiment analysis for restaurant reviews, with a focus on enhancing the customer experience:

First, the project introduces a customer-centric sentiment analysis tool designed specifically to analyse restaurant reviews sourced from Twitter. Unlike traditional systems, which primarily serve businesses by providing operational insights, this tool is built with the customer in mind. It offers users a streamlined, automated method to evaluate public sentiment, helping them make informed dining choices efficiently and accurately.

Secondly, the project implements a hybrid model approach by combining the advanced capabilities of RoBERTa and Electra models. By integrating these two models, the system leverages the strengths of each—RoBERTa's proficiency in capturing context and Electra's efficiency in processing speed. This combination significantly improves the accuracy of sentiment classification, making it more effective in handling the informal and varied nature of Twitter data.

Another key contribution is the system's real-time sentiment analysis capability**.** The tool is designed to track and classify tweets in real time, ensuring that customers receive the most current insights available. This real-time functionality is crucial for users looking to make quick and confident dining decisions based on up-to-date information.

Finally, the project is built with scalability and flexibility in mind, ensuring that it can handle the large and constantly growing volume of Twitter data. Moreover, the system's architecture is designed to be adaptable, allowing for future enhancements such as multilingual support and the integration of other social media platforms. This flexibility not only broadens the scope of the tool but also ensures that it remains relevant and effective in a global and dynamic online environment.

1.5 Organization of report

This report is organized into distinct chapters, each detailing a critical component of the project's development, analysis, and impact. The structured flow ensures a comprehensive understanding of the sentiment analysis system designed for restaurant reviews on Twitter.

The report begins with Chapter 1: Introduction, which provides a detailed background of the project and outlines the motivation behind developing a customer-centric sentiment analysis tool. It also includes the problem statement, objectives, and scope, setting the foundation for the subsequent chapters.

Chapter 2: Literature Survey explores existing research, models, and methodologies relevant to sentiment analysis in the context of restaurant reviews. This chapter highlights the gaps in current approaches, emphasizing the need for a solution tailored specifically for customer decision-making.

Chapter 3: Methodology and Implementation details the design approach, explaining the hybrid model integration of RoBERTa and Electra, and the tools and technologies used. It covers the data collection process and the experimental setup, providing a comprehensive look at how the project was executed and the reasoning behind each step.

Following the methodology, Chapter 4: Result and Analysis presents the outcomes of the model's performance, including key metrics such as accuracy, precision, and recall. This chapter also includes visualizations and an in-depth discussion of the findings, evaluating the model's effectiveness and identifying areas for improvement.

Chapter 5: Advantages, Limitations, and Applications outlines the strengths of the hybrid approach, as well as any constraints encountered during the project. Additionally, it explores the practical applications of the system for customers, highlighting its real-world utility and future potential.

The report concludes with Conclusion and Future Scope, summarizing the project's achievements and contributions. This final section also discusses opportunities for further development, such as expanding the system to support multiple languages and integrating additional social media platforms.

This structured approach ensures that each chapter builds upon the previous one, providing a clear, logical progression that thoroughly explains the project's development, results, and future prospects.

Chapter 2
Literature survey

2.1 Introduction

Sentiment analysis of customer reviews has become invaluable in the restaurant industry, where customer feedback influences reputation and consumer choices. Social media platforms like Twitter provide extensive unstructured data, offering real-time insights into customer experiences. Traditionally, machine learning algorithms such as Naive Bayes, Support Vector Machines (SVM), and Decision Trees have been utilized to classify sentiments. However, as highlighted in recent studies, these models often struggle with the informal, context-dependent language commonly found on social media, including slang, sarcasm, and mixed emotions.

To address these challenges, recent advancements have shifted towards deep learning techniques, with transformer models like BERT, RoBERTa, and Electra demonstrating success in capturing nuanced language features. These models excel at understanding context, but their high computational demands limit scalability. Hybrid approaches, which integrate machine learning and lexicon-based methods, have shown promise in balancing accuracy with processing efficiency, making them well-suited for real-time analysis.

2.2 Exhaustive Literature Survey and Gap Identification

Exhaustive Literature Survey

This section presents an in-depth review of research on sentiment analysis in restaurant reviews, exploring a range of methodologies from traditional machine learning models to hybrid and deep learning approaches. Each study highlights different techniques, models, and outcomes in the sentiment classification domain, providing a comprehensive overview of advancements and persistent challenges in accurately analyzing consumer sentiment.

1. Sentiment Analysis of Restaurant Reviews Using Machine Learning Techniques

Using machine-learning classification techniques, Krishna et al.[1] tackle the task of sentiment analysis in restaurant reviews. Their work tackles the crucial requirement that companies use systematic sentiment polarity analysis—positive, negative, or neutral—within review texts to evaluate consumer feedback efficiently. The thorough data collecting and preprocessing phases, feature selection to maximize model performance, and rigorous evaluation using a variety of machine learning algorithms—Naive Bayes, SVM, Decision Tree, and Random Forest—are all part of the authors' methodological approach. By using an organized method, it is possible to do a thorough comparative analysis and determine which sentiment classification model is the most accurate. Their method's versatility in adjusting data segregation ratios and its systematic framework for algorithm comparison are its key assets. However, difficulties including dependence on the quality of the dataset and certain semantic subtleties that are not fully captured are recognized as limits. When compared to other evaluated models, the SVM classifier's remarkable 94.56% accuracy shows how good it is in sentiment analysis of restaurant reviews. The paper highlights how machine learning can be used to improve business decision-making using perceptive sentiment analysis and suggests directions for further research to further improve these techniques.

2. Machine Learning for Sentiment Analysis for Twitter Restaurant Reviews

To categorize the opinions as either positive or negative, Raju and Jayasinghe [2] further investigate the area of sentiment analysis of restaurant reviews on Twitter. They start by gathering and preparing Twitter data—which contains relevant attributes for sentiment analysis—in CSV format. Many machine learning methods, such as SVM, Naive Bayes, KNN, Decision Tree, Logistic Regression, Random Forest, and an adapted K-means algorithm, are used in their methodology. According to the study, the SVM classifier performs better than the other models, with an F1 score of 77.11% and an accuracy of 73%. The utilization of real-world data, a variety of computing methods, and the introduction of a modified K-means algorithm are the research's strong points. Despite these strengths, the work is lacking in its discussion of algorithmic limits and in-depth comparative analysis. All things considered, the SVM classifier emerges as the most successful instrument for sentiment classification in this investigation, offering insightful information on consumer comments in the restaurant sector.

3. Machine Learning for Sentiment Analysis and Classification of Restaurant Reviews

The use of computation algorithms was acknowledged within sentiment analysis in restaurant reviews while. Patil et al. Altogether, [3] worked with a Kaggle data set of one thousand reviews that had been pre-processed in Python to remove irrelevant elements. Procedures like nearest neighbours, Logistical Progression, Backing Vector Classifier (SVC), and Naive Bayes were utilised, but the SVM performed nearly precise with 78% accuracy. By contrast, the k-nearest neighbours method evaluated views but not as well as support vector machines. This work successfully showed that viewpoints can be detected from unstructured online text, and they proposed the role of Machine Learning in playing around with it. The research suggests the use of sentiment analysis to increase customer satisfaction and sales in restaurants from opinions, as well as generalized applications on drug prediction—medicine-, mental health (depression detection through social media), etc. methodise they were very detailed describing their classification approach, along with sentiment analysis, also post-general processing on data was mentioned for refining it more specifically. Language normalization and Stemming were essential technique used. The algorithm has been excessed by Naive Bayes in text-based clustering operation. Our results with confusion list, accuracy percentage, targeted success and recalled effectiveness proved the superiority of SVM. This project then is of the view that SVM and sentiment analysis are key tools in commerce, as such potential areas may permit further provocation for greater performance from a wide range regarding future creation.

4. Sentiment Analysis of Restaurant Reviews in Social Media Using Naive Bayes

The analysis of the customer sentiment from social media such as twitter restaurant reviews was conducted using the Naive Bayes classified algorithm. This framework involves data preparation, filtering, naive Bayes sentiment classification, and performance assessment. Naive Bayes, renowned for simplicity and efficiency with high dimensional text classification, achieves 73% accuracy with a 27% error rate, 68% precision and 80.07% recall. Benefits include its ease of implementation, computational efficiency, and effectiveness with categorical input variables. However, naive Bayes' assumption of independent features may restrict accuracy in realistic data settings where factors often correlate. The investigation does not juxtapose naive Bayes against other models, leaving the algorithm's relative performance undiscovered. In conclusion, naive Bayes proves helpful for Twitter sentiment examination of restaurant evaluations. Hamad, and Salih [4] propose future work focus on overcoming naive Bayes' independence presumption through deep learning and neural systems to enhance sentiment analysis precision and applicability in diverse real world scenarios.

5. Sentiment Analysis of Restaurant Customer Reviews on TripAdvisor using Naive Bayes

Leksono et al. [5] delve into the sentiment analysis of TripAdvisor restaurant reviews in Surabaya, Indonesia. Their study focuses on classifying reviews as positive or negative using Naive Bayes classification and TextBlob sentiment analysis. Naive Bayes demonstrated higher accuracy at 72.06% compared to TextBlob's 69.12%. The method's strengths lie in its simplicity, efficiency, and established performance in sentiment analysis. However, it assumes feature independence and can struggle with imbalanced datasets Using web crawling tools to collect data, they ran the dataset through WEKA software with Naive Bayes classifier. This highlights that Naive Bayes is a better choice for this dataset and recommends complementary ways to improve the accuracy by carrying out further research in terms of using more data sources or other methods to pre-process sentiment analysis method on restaurant reviews.

6. Sentiment Analysis of Customer Feedback on Restaurant Reviews

Ravikumar and Adarsh [6] proposed an automatic sentiment analysis method based on the Naive Bayes classifier to handle a large amount of restaurant reviews from Kaggle dataset. They pre-process the data to handle missing values and irrelevance attributes, classifying reviews as positive or negative or neutral. Scalability is a feature of the process commended due to its ability for buffering large textual data in real-time which might be vital towards quick business response based on customer feedback. The Naive Bayes classifier is well-known for its simplicity, speed, and reliable performance in context of natural language processing tasks; however, the accuracy measures are not detailed here. The paper highlights its efficacy as a cheap and non-human biased sentiment analysis approach. The authors suggest using it for market research, public health surveillance and the grounds of improving cleanlinessbased policy making on online reviews. In general, their method showed good sentiment analysis of restaurant reviews applicable for business insights, but some minor performance results were absent which could have been useful for assessing the effectiveness further.

7. A Restaurants Recommendation System: Improving Rating Predictions Using Sentiment Analysis

This challenge of improving restaurant's recommendation accuracy is based on addressing the integrating sentiment analysis with collaborative filtering. Their approach involves classifying text-based reviews into positive and negative sentiments to enhance the recommendation process. This method is superior because it combines explicit ratings with implicit sentiment cues from reviews, resulting in more personalized and relevant recommendations. The benefits include capturing a comprehensive view of user preferences, improving user satisfaction, and being applicable across various domains with textual feedback. However, it increases computational complexity, relies on the quality of textual data, and faces challenges in interpreting mixed sentiments or sarcasm. The authors used a Logistic Regression classifier for sentiment analysis, applied to the Yelp Restaurants Reviews dataset. Their sentiment-enhanced system outperformed the baseline, showing superior performance in metrics like precision, recall, F-score, and mean absolute error. Petrusel et al. [7] conclude that integrating sentiment analysis can significantly enhance recommendation quality, suggesting future refinements and broader applications.

8. Uzbek Sentiment Analysis Based on Local Restaurant Reviews

Matlatipov et al. [8] address sentiment analysis in Uzbekistan, using a dataset of 8,210 restaurant reviews from Google Maps. They preprocess the data by removing noise, translating non-Uzbek reviews, and applying a specialized Uzbek stemming algorithm. Reviews are annotated based on a 5-star rating system. The study employs logistic regression (LR) , support vector machines (SVM) , recurrent neural networks (RNN) , and convolutional neural networks (CNN) , achieving the highest accuracy of 91% with an LR model using word and character n-grams. The paper highlights the importance of tailoring preprocessing for agglutinative languages and LR's effectiveness with smaller datasets. Challenges include review imbalance and potential sentiment labelling inaccuracies based on star ratings. Future work aims to expand the dataset, apply cross-validation, and improve sentiment analysis tools for low-resource languages like Uzbek, with resources publicly available for further research.

9. Sentiment Analysis and Classification of Restaurant Reviews using Machine Learning

Sentiment analysis of restaurant reviews in Karachi, Pakistan is undertaken. Zahoor and Hamid [9] approach integrates Natural Language Processing (NLP) and machine learning to classify sentiments and categorize reviews into specific aspects like food taste and service quality. Utilizing Naive Bayes, Logistic Regression, SVM, and Random Forest algorithms, they achieve high accuracy, with Random Forest leading at 95%. Pros include comprehensive analysis for service improvement, handling large social media datasets, and adaptive ML algorithms. Challenges include computational demands and data quality impacts on accuracy. The Random Forest model's balanced precision and recall establish its superiority, suggesting robust performance in sentiment and category classification. Future studies could explore broader data sources and advanced ML techniques to refine sentiment analysis in restaurant feedback systems, enhancing customer insights and service enhancements in dynamic market environments.

10. Sentiment Analysis on Restaurant Review using Hybrid Approach

Yadav et al. To meet the challenge of successful sentiment classification and achieve higher performance in restaurant reviews [10], propose a hybrid method combined machine learning with dictionary-based methods to approach this problem. Since we want that our sentiment classifiers perform better without doing much effort about labelling. Due to the adaptability of machine learning and predefined sentiment lexicons in dictionary-based methods, this hybrid model gets better performance. The authors use classifiers like Logistic Regression, SVC, LinearSVC, MultinomialNB and XGBoost on a 1000 restaurant review dataset from Kaggle. Of all the classifiers, Logistic Regression were other top performers and fails to accomplish highest accuracy compare XGBoost (67.5%), SVC/MultinomialNB (64%), LinearSVC (65%). The approach improves accuracy and scales to big data but suffers from an increase in the complexity of model development and integration. The authors proposed that the hybrid approach works well, and also suggest some other topics for future research which include real-time analysis implementation in detecting sarcasm, complexity of sentences being analysed or multilingual sentiment analysis.

11. Sentiment Analysis of Restaurant Review with Classification Approach in the Decision Tree-J48 Algorithm

Sentiment analysis in restaurant reviews is addressed, aiming to classify them as positive or negative for consumer guidance and restaurant feedback. In this method Adnan et al. [11] uses the Decision Tree-J48 for analysing data, which is simple and can handle both numerical values and attributes with categories information, considering best algorithm well-known of WEKA Software. Although it is efficient of 2s, the study shows that its accuracy:45.6% (suboptimal). Issues seen in considering an entire dataset and class imbalanced case. Both the precision and recall, as well as F-measure, of this model were also poor. TripAdvisor English reviews of Surabaya's restaurants dataset Although this example demonstrates how the algorithm is simple to work with, referring would not necessarily prevent more hybrid methods of analysis or comparison implementations when dealing for greater sentiment accuracy in restaurant reviews. Future research should strive to improve findings and accommodate more complex datasets that better inform consumer choice behaviour as well as restaurant service quality.

12. Applying Deep Learning Approach to Targeted Aspect-based Sentiment Analysis for Restaurant Domain

[12] Khine et al. address the challenge of aspect-based sentiment analysis (ABSA) in the restaurant domain using a novel deep learning approach. It employs SenticNet to enrich the classic Long Short-Term Memory (LSTM) model and extends it as a multi-attentive LSTM (MA-LSTM). This technique uses the MA-LSTM model to benefit from external knowledge, and better sentiment classification via long text sequences with emphasis on important information in a sentence. The MA-LSTM method achieves the best balance of capturing context with multiple attention mechanisms and preserving accuracy. Pros Long text sequences External knowledge Sentiment salient parts Drawbacks are increased computational requirements and depending on domain knowledge. In experiments on a dataset containing more than 20,000 annotated sentences extracted from Yangon-based restaurant reviews, the SenticNet

MA-LSTM outperformed standard LSTM as well as TD-LSTM and TC-LSTM, AE-IMLCNN with an accuracy of 87.2%, while ATAE-ILMCNN failed to produce any results beyond random chance baseline. Based on the experimental results, they find that their model benefits ABSA through external knowledge integration and multiattentive mechanism. In conclusion, The proposed method in this paper provides new perspective of customer insights and contributes to practical sentiment analysis applications.

13. Sentiment Analysis using various Machine Learning and Deep Learning Techniques

We propose a word-level sentiment analysis model that combines deep learning and machine learning techniques for restaurant evaluations. Naive Bayes, Logistic Regression, Random Forest, Linear SVC, KNN, Decision Tree, CNN, and LSTM are among the models examined by Umarania et al. [13]. A sentiment polarity measuring discriminator and a multi-head attention mechanism in the model's generator are two important developments that improve the quality of word embeddings. Their approach includes thorough preprocessing of the data, feature extraction, and classification. Metrics such as accuracy, recall, F1score, precision, AUC score, ROC curve, and training time are assessed. According to the results, CNN and LSTM perform exceptionally well in terms of accuracy for the restaurant dataset, while Naive Bayes achieves the maximum AUC score with effective training. The study finds that deep learning models provide better accuracy than machine learning classifiers, despite the latter's

14. Sentiment Analysis for Restaurant Rating

[14] Kaviya et al. tackle the challenge of automating restaurant ratings based on customer reviews, aiming to provide an objective and scalable system. They present a sentiment analysis approach that assigns ratings by identifying sentiment keywords and evaluating the emphasis placed on emotions, including emoticons and adverbs, within reviews. This method addresses the shortcomings of biased market research and subjective online reviews, offering a real-time analysis of customer sentiments. The system’s strengths lie in its automation, reduction of bias, and scalability to handle large datasets. However, potential drawbacks include its inability to fully capture the nuanced context of reviews and its reliance on the quality and

representativeness of the dataset used, which could affect accuracy. Although specific model details are not disclosed, the system operates on Yelp data. While the study lacks a comparative analysis of different models, it underscores the system's potential to enhance restaurant rating accuracy. Future research could focus on integrating additional review factors and refining classification techniques for improved performance.

15. Game Theory and MCDM-based Unsupervised Sentiment Analysis of Restaurant Reviews

Punetha et al. [15] perform customer satisfaction; sentiment analysis; emotion detection on restaurant reviews Traditional machine learning requires training on large datasets, which is sometimes time-consuming and domain specific. It introduces an unsupervised sentiment classification model based on a combination of game theory and Multi-Criteria Decision Making (MCDM). It is a two-stage approach, which classifies positive and satisfied reviews in the first stage using performance scores based on context, rating and emotion ratings; while identifying sentiment polarity (positive/negative) and customer satisfaction with non-cooperative game model on negative or neutral reviews. This has the domain/language independence and generalizability, a significant reduction in computational difficulty as well no pretraining needed on large datasets. Nevertheless, it is dealing with difficulties concerning very complex linguistic levels such as sarcasm and irony.

16. A Cloud-based Tool for Sentiment Analysis in Reviews About Restaurants on TripAdvisor

Torales et al. [16] The concepts used are TOPSIS, NCG and the solution is modelled for Nash Equilibrium and implemented over TripAdvisor + Yelp datasets. It outperformed existing methods both in accuracy, precision-recall (PR) curves and F1-scores as well the Matthews correlation coefficient. Their model is efficient and adaptable in sum, although it requires enhancements for sophisticated expression grounding (the study concludes.). It spans data collection, text preparation, sentiment analysis, and visualization stages, leveraging modern architecture for scalability and resource efficiency. Advantages include cloud accessibility, comprehensive analysis stages within a unified platform, and VADER's high precision in sentiment classification. Potential drawbacks might include performance limitations on diverse datasets beyond TripAdvisor and constraints in capturing complex human sentiment expression. Validated with over 33,500 English reviews from restaurants in Granada, Spain, the tool's accuracy specifics aren't detailed, but VADER's reported 99% precision in tweet sentiment classification implies robust performance. The study emphasizes future enhancements through broader data integration and exploration of advanced sentiment analysis methods like aspect-based approaches.

17. Restaurant Recommender System Based on Sentiment Analysis

Asani et al. [17] implemented a sentiment analysis and semantic clustering module designed to provide context aware personal restaurant recommendations. Unify API to seamlessly access and manipulate data across the platform; their algorithm captures over 100m individual food preferences from online reviews on a daily basis, rather than using traditional methods like Term-Frequency analysis for this purpose. The system was assessed using TripAdvisor data on precision, recall and f-measure for four distinct recommendation scenarios. Results were able to achieve an accuracy of 92.8% in the top-five recommendations, outperforming existing systems on precision indices. Best part of the system: The dynamic extraction of user preferences from textual data integration of sentiment analysis to gauge opinion polarity, and context-awareness by considering factors like user location and time for suggesting nearby open restaurants. Limitations include reliance on TripAdvisor data and a focus on individual rather than group recommendations. The study highlights opportunities for future research in improving group-based recommendations and further enhancing the system's performance with additional data sources.

The following table summarizes the key studies reviewed, highlighting their methodologies, key findings, and noted limitations –

Table 1 Literature Review Summary

Title	Model Used	Accuracy Rate	Key Findings	Limitations
Sentiment Analysis of Restaurant Reviews	SVM	94.56%	SVM is the most effective; supports business decisions.	Dataset quality affects results.
Machine Learning for Twitter Restaurant Reviews	SVM	73%	Uses real-world data; modified K-means algorithm.	Lacks detailed algorithm analysis.
Sentiment Analysis and Classification of Restaurant Reviews	SVM	78%	SVM excels at interpreting unstructured text.	Limited comparison with other models.

Table 2 Literature Review Summary

Title	Model Used	Accuracy Rate	Key Findings	Limitations
Sentiment Analysis of Restaurant Reviews in Social Media	Naive Bayes	73%	Effective for sentiment classification on Twitter.	No model comparisons; assumes feature independence.
Sentiment Analysis of Restaurant Customer Reviews on TripAdvisor	Naive Bayes	72.06%	Outperforms TextBlob in TripAdvisor reviews.	Assumes feature independence; struggles with imbalance.
Sentiment Analysis of Customer Feedback on Restaurant Reviews	Naive Bayes	Not specified	Manages large volumes of reviews; reduces human bias.	Lacks detailed accuracy metrics.
A Restaurants Recommendation System	Logistic Regression	Not specified	Enhances recommendation accuracy by combining sentiment cues.	Increases complexity; challenges with mixed sentiments.
Uzbek Sentiment Analysis Based on Local Restaurant Reviews	Logistic Regression	91%	Effective with n-grams for sentiment analysis in Uzbek.	Review imbalance; potential labeling inaccuracies.

Table 3 Literature Review Summary

Title	Model Used	Accuracy Rate	Key Findings	Limitations
Sentiment Analysis and Classification of Restaurant Reviews	Random Forest	95%	High accuracy; balances precision and recall.	High computational demands; data quality sensitive.
Sentiment Analysis on Restaurant Review Using Hybrid Approach	XGBoost	67.5%	Hybrid method enhances accuracy; XGBoost outperforms others.	Complex development; issues with sarcasm detection.
Sentiment Analysis of Restaurant Review with Decision Tree-J48	Decision Tree-J48	45.6%	Simple, handles various data types but performs poorly.	Suboptimal accuracy; poor precision and recall.
Deep Learning Approach to Aspect-based Sentiment Analysis	MA-LSTM	87.2%	Excels at aspect-based sentiment analysis.	Increased computational needs; relies on domain knowledge.

Table 4 Literature Review Summary

Title	Model Used	Accuracy Rate	Key Findings	Limitations
Sentiment Analysis Using Various Techniques	Naive Bayes, CNN, LSTM	Not specified	Naive Bayes had the highest AUC; CNN and LSTM excelled.	Faster ML; deep learning offers better accuracy.
Sentiment Analysis for Restaurant Rating	Not specified	Not specified	Automated real-time restaurant ratings.	Lacks model details; context struggles.
Game Theory and MCDM-based Sentiment Analysis	Unsupervised Classification Model	Not specified	Domain/language independent model.	Sarcasm and irony issues.
Cloud-based Sentiment Analysis Tool	VADER	Not specified	High accuracy with VADER and others.	Issues on varied datasets; complex sentiments.
Restaurant Recommender System Based on Sentiment Analysis	Context-Aware Algorithm	92.8%	High-precision sentiment analysis for recommendations.	Limited to individual recommendations.

Gap Identification

From the literature reviewed, several key gaps emerge that justify the approach taken in this project:

- Handling Nuanced Sentiments: Traditional and machine learning methods often struggle with detecting sarcasm, idioms, and mixed sentiments prevalent in social media. Even advanced models, including some deep learning methods, show limitations in processing complex expressions, suggesting the need for models specifically trained on nuanced, context-rich data.
- Scalability and Real-Time Adaptability: Transformer models like BERT and RoBERTa, while powerful, require significant computational resources, making real-time sentiment analysis difficult to implement on standard hardware. This gap indicates the importance of exploring efficient hybrid models, such as combining RoBERTa's contextual depth with Electra's processing speed, for scalable, real-time solutions.
- Multilingual and Cultural Adaptability: Most studies reviewed focused on English-language datasets, with limited resources for multilingual analysis. This presents a gap in adapting sentiment analysis models to diverse linguistic contexts, which would expand their accessibility and relevance globally.
- Complex Data Representation and Generalizability: Hybrid approaches show promise in improving accuracy, but their complexity often increases the computational burden, particularly when handling unstructured data in real-time. More adaptable, efficient hybrid models are needed to bridge this gap while maintaining accuracy and relevance.

In addressing these identified gaps, this project adopts a hybrid approach that leverages RoBERTa's ability to capture contextual nuances alongside Electra's processing efficiency. This design enables a model that is not only accurate in handling informal, nuanced language on social media but also scalable and capable of real-time classification, aligning with the demands highlighted by existing research.

2.3 Issues and Future Scope

The literature highlights various issues with existing sentiment analysis approaches, as well as potential areas for future research and development:

Current Issues:

- Complexity in Capturing Subtle Sentiments: Models often fail to identify sarcasm or irony, which can lead to inaccurate sentiment classification. Although hybrid approaches attempt to address these issues, they remain limited in adaptability.
- High Computational Demand: Advanced deep learning and transformer models, while accurate, require high computational resources, posing challenges for real-time applications on standard consumer hardware.
- Imbalanced Data and Feature Independence: Methods like Naive Bayes assume feature independence and can struggle with imbalanced datasets, which affects their applicability to dynamic and diverse social media content.
- Dependence on English Language Data: Most studies focus on English-language reviews, with limited resources dedicated to other languages, which restricts the usability of sentiment analysis tools globally.

Future Scope:

- Multilingual Support: Expanding sentiment analysis capabilities to handle multiple languages, including low-resource languages, would enhance accessibility and applicability. Future research could focus on creating multilingual datasets and models capable of processing mixed-language content, which is common on social media.
- Real-Time Adaptability: Optimizing transformer models for deployment on mobile and edge devices could facilitate real-time sentiment analysis, broadening the scope of application in consumer-facing industries.
- Improved Detection of Sarcasm and Mixed Sentiments: Incorporating sarcasm and mixed sentiment datasets and fine-tuning models specifically for these challenges can improve accuracy. Leveraging multi-attention mechanisms and context-aware models may address these challenges more effectively.
- Integration with Cross-Platform Data: Expanding sentiment analysis to include reviews from multiple platforms (e.g., Twitter, Yelp, Instagram) would provide a comprehensive view of customer opinions, increasing the relevance and utility of sentiment analysis tools.

Chapter 3
Methodology and Implementation

3.1 About Dataset

The dataset employed in this project serves as a crucial foundation for the sentiment analysis of restaurant reviews sourced from Twitter. With an overarching goal to facilitate a deeper understanding of customer feedback, the dataset enables the classification of reviews into three core sentiment categories: positive, negative, and neutral. The results of this sentiment analysis will directly inform customer decision-making, guiding their dining choices based on collective experiences shared by previous patrons. Moreover, it offers restaurants an insightful overview of their strengths and weaknesses, ultimately contributing to improved service and dining experiences.

1. Dataset Composition and Structure

This dataset is composed of 2,700 individual reviews, representing a wide spectrum of customer experiences, opinions, and sentiments regarding their dining ventures. These reviews have been meticulously curated to ensure that the dataset provides a comprehensive and balanced view of the sentiments expressed by customers. Each review is a textual representation of the customer's experience and has been pre-labelled with a sentiment category to aid in the training of machine learning models.

To maintain the robustness of our sentiment analysis models, the dataset is split into two segments:

- Training Set: 2,160 reviews (80%) – used for model development.
- Test Set: 540 reviews (20%) – reserved for evaluating model performance.

Each review encapsulates varied dimensions of restaurant interactions, covering key areas such as food quality, customer service, ambiance, and overall experience. This diversity ensures that the dataset is not only reflective of the restaurant industry but also enhances the accuracy and applicability of the sentiment analysis models.

2. Dataset Features

The dataset consists of two main features:

- Sentence: The review itself, which serves as the core text to be analysed. These sentences vary in length, complexity, and tone, ranging from short, succinct remarks to more elaborate and descriptive feedback. This variation simulates real-world data conditions typical of social media platforms like Twitter, where user-generated content is highly informal, unstructured, and often filled with colloquial language.
- Polarity (Sentiment Label): Each review is assigned a sentiment label, which indicates the nature of the feedback provided. The polarity categories are:
    - Positive: Reviews that express satisfaction, enjoyment, or praise for the restaurant, its food, or services.
    - Negative: Reviews that reflect dissatisfaction, complaints, or criticisms about the dining experience.
    - Neutral: Reviews that are factual or objective, where the customer is neither particularly satisfied nor dissatisfied, but simply provides a neutral commentary.

3. Sentiment Distribution

An essential characteristic of any sentiment analysis dataset is its distribution across the different sentiment categories. In this dataset, the reviews are well-distributed among the three sentiment labels, ensuring that the model can learn effectively from a balanced representation of customer opinions. This balanced distribution is key for developing models that generalize well across all sentiment classes, avoiding biases towards a particular sentiment.

- Positive Sentiments: These reviews highlight favorable dining experiences and often commend specific aspects such as food taste, service efficiency, or the ambiance of the restaurant.
- Negative Sentiments: These reviews express displeasure with the restaurant, whether it be poor service, subpar food quality, or unsatisfactory ambiance.
- Neutral Sentiments: These reviews typically include objective observations or factual statements about the restaurant, offering no strong emotional inclination.

4. Context and Relevance of the Dataset

The real-world nature of this dataset, collected from Twitter, adds a layer of complexity and richness to the analysis. Reviews on platforms like Twitter are often characterized by their brevity and informal language, making them an ideal challenge for natural language processing (NLP) models. The presence of slang, emotive expressions, and varying sentence structures ensures that the dataset realistically represents customer feedback in its most authentic form.

The dataset spans key domains of the restaurant experience:

- Food Quality: Evaluation of the taste, freshness, and presentation of dishes.
- Service Quality: Feedback regarding the responsiveness, friendliness, and attentiveness of the staff.
- Ambiance: Customer perceptions of the restaurant's environment, including decor, seating comfort, and overall atmosphere.
- Value for Money: A critical aspect where customers assess whether their dining experience was worth the price paid.

5. Data Preprocessing for Model Training

To ensure the dataset is ready for model training, a series of preprocessing steps were applied. These steps help cleanse the data and structure it for optimal performance in machine learning models:

- Text Cleaning: This involves removing any unnecessary symbols, punctuation, and stop words from the sentences, thus eliminating noise in the data that might hinder model performance.
- Tokenization: Each review is broken down into individual tokens or words, a necessary step for transforming textual data into a format that machine learning models can process.
- Labelling: The polarity labels (positive, negative, neutral) are mapped to each corresponding review, facilitating the classification tasks during model training.
- Dataset Splitting: An 80/20 split of the dataset into training and test sets was performed, ensuring that the model is trained on one portion of the data and tested on a completely unseen portion to evaluate its generalization capability.

3.2 Design Approach

The sentiment analysis system leverages a hybrid model that combines the strengths of two state-of-the-art NLP models, RoBERTa and Electra, to provide accurate and efficient classification of restaurant reviews collected from Twitter. Social media platforms like Twitter present unique challenges for sentiment analysis due to the informal nature of the text, including the use of slang, abbreviations, emojis, and emotive language. To address these challenges and ensure high performance, the hybrid approach integrates the contextual depth of RoBERTa with the speed and efficiency of Electra, balancing precision and processing power.

System Workflow Overview

The system follows a structured approach where each model is trained independently on the same dataset of tweets. Once trained, the models' outputs are combined using an ensemble technique to enhance the final sentiment predictions. This method ensures that individual model strengths are maximized while reducing the chance of misclassification. The hybrid ensemble achieves higher consistency, especially in handling nuanced or ambiguous sentiments.

1. Data Collection

The Twitter API is employed to collect real-time tweets related to dining experiences by querying keywords such as "#foodreview," "#restaurant," and "#diningexperience." The collected tweets reflect various opinions, ranging from positive experiences to complaints, ensuring a diverse dataset for robust model training.

2. Preprocessing

The raw tweets undergo data cleaning and tokenization to standardize input and ensure compatibility with both models.

- Data Cleaning: Unnecessary elements like URLs, emojis, hashtags, and special characters are removed, and the text is converted to lowercase for uniformity.
- Tokenization: The cleaned text is tokenized using the RoBERTa and Electra tokenizers, breaking it into meaningful tokens that can be processed by each model.

3. Model Training

- RoBERTa: Known for its contextual understanding, RoBERTa excels in capturing intricate relationships between words within a sentence. This helps in analysing the sentiment within long and complex reviews, where context plays a crucial role.
- Electra: Designed for efficiency, Electra quickly processes shorter, straightforward tweets, making it ideal for handling brief and informal text.

Both models are fine-tuned independently on the tokenized data, with optimization parameters set to maximize accuracy and performance.

4. Hybrid Model Integration (Ensemble Method)

Once the models are trained, their predictions are aggregated through an ensemble technique. This approach takes advantage of the strengths of each model:

- RoBERTa's ability to capture deep contextual meanings.
- Electra's speed and efficiency in processing simpler text.

The ensemble model assigns weights to the predictions from both models, ensuring the most accurate overall classification. This hybrid integration improves the system's ability to classify ambiguous sentiments that would have been challenging for either model alone.

5. Sentiment Classification

The hybrid model classifies each tweet into one of three sentiment categories: positive, negative, or neutral.

- Positive: Tweets expressing satisfaction, enjoyment, or praise (e.g., "The food was amazing!").
- Negative: Tweets reflecting complaints or dissatisfaction (e.g., "Terrible service!").
- Neutral: Tweets that express neither strong satisfaction nor dissatisfaction (e.g., "Visited the restaurant yesterday.").

6. Performance Evaluation and Retraining

The system's performance is evaluated using standard metrics such as accuracy, precision, recall, and F1-score. Based on the results, the models may undergo retraining to fine-tune parameters and correct misclassifications. For example, the hybrid model adapts over time by learning from past errors, ensuring improved performance in future iterations.

7. Real-Time Insights and Output

Once the sentiment predictions are finalized, the results are presented to the user in real time. The system outputs classified tweets with the corresponding sentiment, providing valuable insights into customer opinions. These insights can be used by restaurants to monitor and enhance their service quality.

Block diagram:

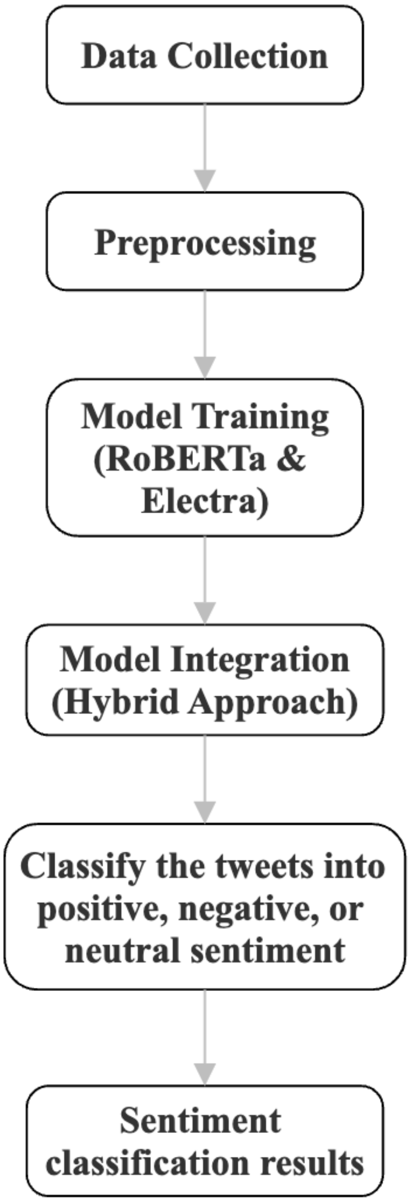

Fig. 1 Overview of Hybrid Sentiment Analysis

3.3 Tools and Technologies Used

The implementation of the hybrid model requires a set of programming tools, libraries, and APIs:

1. Programming Language

Python is used as the primary language due to its robust support for machine learning libraries and ease of integration with NLP tools.

2. Libraries:

- **Hugging Face Transformers**: For implementing **RoBERTa** and **Electra** models.
- **PyTorch**: As the deep learning framework for model training and fine-tuning.
- **Pandas and NumPy**: For efficient data manipulation and preprocessing.
- **Scikit-learn**: Used for model evaluation metrics, including accuracy, precision, recall, and F1-score.
- **Matplotlib and Seaborn**: For visualizing model performance through plots such as confusion matrices and accuracy graphs.

3. Development Environment

Jupyter Notebook (Google Colab) is used for developing, running, and documenting the code, providing an interactive platform for model experimentation.

4. APIs:

- **Twitter API**: Employed to collect real-time data based on restaurant-related hashtags and keywords, ensuring relevance and specificity in the dataset.

3.4 Flowchart

The flowchart will show the detailed actions within each block and add some internal steps to highlight the decision-making and processing flow. This is useful for breaking down each step of the system in more detail, showing specific operations performed at each stage.

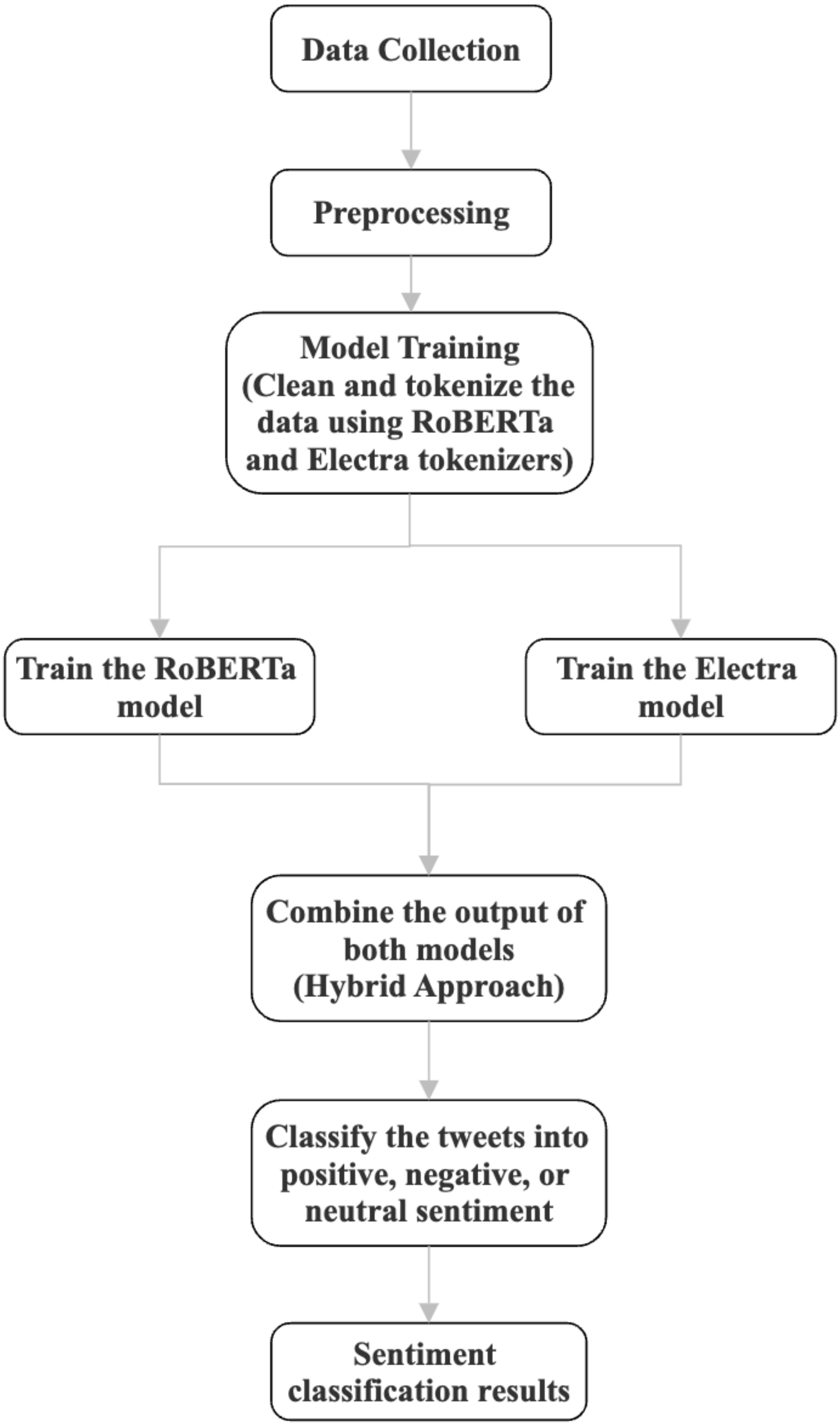

Fig. 2 Sentiment Classification Process Flow

Chapter 4
Results and Analysis

This chapter presents an analysis of the results obtained from the RoBERTa, Electra, and Hybrid models applied to restaurant review sentiment classification. Through a detailed examination of model performance metrics and error analysis, this section aims to capture each model's strengths, limitations, and areas for improvement.

4.1 Model Performance Metrics

Each model's performance was evaluated using accuracy, precision, recall, and F1-score to capture its ability to classify sentiments as positive, negative, or neutral. These metrics provided an understanding of how well each model could interpret nuanced language within reviews and detect varying emotional tones. Below is a breakdown of the results for each model.

1. Electra Model Performance

The Electra model attained an overall accuracy of 0.77, showing balanced performance across sentiment classes:

- Negative: Precision = 0.83, Recall = 0.71, F1-Score = 0.77
- Neutral: Precision = 0.73, Recall = 0.88, F1-Score = 0.80
- Positive: Precision = 0.73, Recall = 0.73, F1-Score = 0.73

Electra's high recall for neutral sentiments reflects its sensitivity in identifying neutral reviews accurately, while its balanced performance on positive and negative sentiments indicates consistent classification abilities.

Electra Classification Report:

	precision	recall	f1-score	support
negative	0.83	0.71	0.77	273
neutral	0.73	0.88	0.80	182
positive	0.73	0.73	0.73	162
accuracy			0.77	617
macro avg	0.76	0.77	0.76	617
weighted avg	0.77	0.77	0.77	617

Fig. 3 Classification report Electra

2. RoBERTa Model Performance

The RoBERTa model achieved an accuracy of 0.76, with high precision and recall values in the negative and neutral sentiment classes:

- Negative: Precision = 0.86, Recall = 0.67, F1-Score = 0.75
- Neutral: Precision = 0.74, Recall = 0.85, F1-Score = 0.79
- Positive: Precision = 0.69, Recall = 0.82, F1-Score = 0.75

RoBERTa's high recall for positive and neutral sentiments shows its robust capacity for identifying these reviews accurately, though it occasionally misclassifies negative reviews as neutral.

RoBERTa Classification Report:

	precision	recall	f1-score	support
negative	0.86	0.67	0.75	273
neutral	0.74	0.85	0.79	182
positive	0.69	0.82	0.75	162
accuracy			0.76	617
macro avg	0.76	0.78	0.76	617
weighted avg	0.78	0.76	0.76	617

Fig. 4 Classification report RoBERTa

3. Hybrid Model Performance

Combining strengths from both Electra and RoBERTa, the hybrid model attained an accuracy of 0.73, with particular strength in positive sentiment classification:

- Positive: Precision = 0.90, Recall = 0.78, F1-Score = 0.83
- Negative: Precision = 0.63, Recall = 0.63, F1-Score = 0.63
- Neutral: Precision = 0.19, Recall = 0.53, F1-Score = 0.27

The hybrid model showed strong precision in positive sentiment classification, indicating high confidence in identifying positive instances, though it faced challenges in accurately detecting neutral sentiments, potentially due to a higher incidence of misclassifications in this class.

	precision	recall	f1-score	support
positive	0.90	0.78	0.83	566
negative	0.63	0.63	0.63	193
neutral	0.19	0.53	0.27	40
accuracy			0.73	799
macro avg	0.57	0.65	0.58	799
weighted avg	0.80	0.73	0.76	799

Fig. 5 Classification report Hybrid

To further summarize the performance of each model, the following table provides a consolidated view of accuracy, average precision, recall, and F1-score across the Electra, RoBERTa, and Hybrid models. This high-level comparison highlights each model's general effectiveness and enables an at-a-glance evaluation of their respective strengths in sentiment classification.

Table 5 Performance Summary Table

Model	**Accuracy**	**Average Precision**	**Average Recall**	**Average F1-Score**
Electra	0.77	0.77	0.77	0.77
RoBERTa	0.76	0.78	0.76	0.76
Hybrid	0.73	0.80	0.73	0.76

4.2 Error Analysis and Misclassifications

Error analysis was conducted by examining misclassified instances from each model's predictions. This involved identifying patterns in the errors across sentiment classes and analyzing sample misclassified reviews to understand the factors contributing to these mistakes. By categorizing common misclassification patterns, insights were gained into each model's decision-making process and limitations.

Findings and Observations

1. Neutral Sentiment Misclassifications
    - Observation: Many neutral reviews were misclassified as positive or negative by the hybrid model.
    - Reason: This could be due to the model's tendency to overfit on strong polarity indicators present in positive and negative language. As a result, statements with mild or balanced sentiments were interpreted as polar sentiments.
    - Improvement Suggestion: Additional training on a more balanced dataset with ample neutral examples could help the model develop a better understanding of neutrality.
2. Negative Sentiment Misclassifications
    - Observation: Both Electra and RoBERTa occasionally misclassified negative sentiments as neutral.
    - Reason: Negative reviews often contain subtle criticism or dissatisfaction that may not have strong polarity cues, leading the models to categorize them as neutral. In cases where criticism was implied rather than explicit, the models struggled to detect negativity accurately.
    - Improvement Suggestion: Incorporating a targeted training subset with subtle negative language may improve the model's sensitivity to indirect criticism or nuanced dissatisfaction.

3. Positive Sentiment Misclassifications
    - Observation: The hybrid model showed high precision in positive sentiment classification but occasionally sacrificed recall, resulting in some positive instances being misclassified.
    - Reason: The hybrid model's high precision indicates a bias towards avoiding false positives, which may result in some genuinely positive reviews being missed. This suggests a precision-recall trade-off, with the model prioritizing precise identification of positivity over capturing all positive instances.
    - Improvement Suggestion: Fine-tuning the model's balance between precision and recall could improve recall rates without compromising precision, making the model more effective in identifying all positive sentiments accurately.

Conclusion of Error Analysis

Each model showed specific strengths and weaknesses in classifying sentiments, with notable differences in handling neutral sentiments. The analysis underscores that while the hybrid model was particularly strong in positive sentiment classification, it encountered challenges with neutrality, likely due to the subtle nature of neutral language. By examining common misclassifications, this analysis provides a foundation for refining each model to address its unique limitations.

4.3 Comparative Analysis

Based on the results, the following insights emerged from a comparative analysis across all models:

- Positive Sentiment: The hybrid model outperformed both Electra and RoBERTa in positive sentiment classification, achieving the highest precision (0.90) and F1-score (0.83).
- Negative Sentiment: RoBERTa demonstrated the highest precision for negative sentiments, while Electra offered more balanced performance across classes.
- Neutral Sentiment: Electra and RoBERTa outperformed the hybrid model in neutral sentiment detection, suggesting that single-model architectures may be better suited for handling neutrality.

These findings suggest that each model offers distinct strengths based on the type of sentiment being classified. The hybrid model is advantageous for positive sentiment detection, while Electra and RoBERTa provide more reliable performance for neutral and negative sentiment classification.

Table 6 Comparative Analysis

Metric	**RoBERTa**	**Electra**	**Hybrid**
Positive Precision	0.86	0.83	0.90
Positive F1-Score	0.75	0.77	0.83
Negative Precision	Highest	Balanced	Moderate
Neutral Detection	Reliable	Reliable	Less reliable

4.4 Conclusion of Analysis

Each model exhibited unique strengths across sentiment classes, with the following key takeaways:

- The Electra model provided balanced performance across all sentiment classes, making it suitable for general sentiment analysis tasks.
- The RoBERTa model displayed higher precision for negative and neutral sentiments, reflecting its ability to minimize misclassifications in these categories.
- The Hybrid model excelled in positive sentiment detection but faced challenges with neutral sentiment accuracy, suggesting potential for refinement in handling subtleties in language.
- Overall, this analysis highlights the importance of model selection based on specific use cases. For scenarios prioritizing accurate positive sentiment classification, the hybrid model is recommended, while the Electra and RoBERTa models offer balanced and precision-focused alternatives, respectively. Future work could explore fine-tuning techniques to optimize the hybrid model's ability to handle neutrality, potentially through additional neutral-focused training data or enhanced parameter tuning.

Chapter 5
Advantages, Limitations and Applications

This chapter explores the benefits, constraints, and potential applications of the hybrid sentiment analysis system implemented using RoBERTa and Electra for analysing restaurant reviews on Twitter.

5.1 Advantages of the Hybrid Approach

The integration of RoBERTa and Electra in a hybrid framework offers significant advantages in the context of sentiment analysis:

1. Enhanced Accuracy

By combining the contextual depth of RoBERTa with the processing efficiency of Electra, the hybrid approach achieves a higher level of accuracy in classifying sentiments as positive, negative, or neutral. RoBERTa excels at capturing nuanced relationships within the text, ensuring that the model understands the context behind customer reviews, while Electra's lightweight and efficient structure boosts the overall speed of sentiment classification.

2. Real-Time Analysis Capability

The hybrid system is optimized for real-time sentiment analysis, which is particularly beneficial for customers seeking immediate insights into restaurant feedback. This allows users to receive up-to-date sentiment evaluations quickly, aiding in faster decision-making without the need for manual review of numerous tweets.

3. Reduction in Misclassifications

By leveraging the strengths of both models, the hybrid approach minimizes the common errors associated with standalone models. It reduces the false positive and false negative rates, especially in the neutral category, providing a more balanced and reliable classification output.

5.2 Limitations

Despite the strengths of the hybrid model, there are certain limitations that must be acknowledged:

1. Language Constraint

The initial implementation of the system focuses exclusively on English-language tweets. This constraint limits the system's applicability in a global context where multiple languages are used. Extending the model to support multilingual analysis would require further development and additional language-specific data.

2. Data Quality and Noise

Social media platforms like Twitter often contain noisy and unstructured data, including abbreviations, slang, misspellings, and irrelevant content. While the preprocessing steps implemented help mitigate these issues, the model's accuracy may still be impacted by the variations in language use across different users.

3. Resource Requirements

The use of advanced transformer models such as RoBERTa and Electra necessitates significant computational resources. The hybrid approach is optimized for GPU environments, and performance may be compromised when deployed on lower-specification hardware, particularly for real-time processing.

4. Scalability Concerns:

As the volume of tweets continues to grow, the system may face challenges in scaling effectively. While the current architecture handles the existing volume efficiently, expanding the system to analyse larger datasets or other social media platforms may require architectural optimization or cloud-based scaling solutions.

5.3 Applications

The hybrid sentiment analysis tool presents several valuable applications for customers:

1. Informed Dining Decisions

By classifying and aggregating customer reviews from Twitter, the tool offers customers a quick and efficient way to assess restaurant sentiment. This helps users make informed choices about where to dine based on collective feedback, enhancing their overall dining experience.

2. Trend Analysis for Customers

The system's ability to capture and display real-time sentiment data enables customers to spot emerging trends, such as seasonal restaurant preferences or shifts in public opinion over time. This feature empowers customers to explore trending dining options and adjust their choices based on popular sentiment.

3. Potential Extensions

The system can be further developed to support multiple languages, enabling a more inclusive and global application for users around the world. Additionally, integrating the tool with other social media platforms, such as Instagram or Yelp, could provide customers with a holistic view of restaurant sentiment across various online spaces, enhancing its practical utility.

4. Customizable Alerts and Notifications

The system could be expanded to provide customers with notifications and alerts based on specific restaurant sentiments or preferences, such as warnings when a restaurant receives sudden negative feedback or recommendations based on consistently positive trends.

Chapter 6
Conclusion and Future Scope

6.1 Conclusions

This project presents a comprehensive and innovative approach to sentiment analysis through the implementation of a hybrid model that integrates RoBERTa and Electra. The primary objective was to classify restaurant reviews from Twitter into positive, negative, or neutral sentiments. The combination of these advanced models addresses the challenges of processing informal social media language, including slang, abbreviations, and mixed emotions, which traditional models often struggle to handle effectively.

By leveraging RoBERTa's contextual depth and Electra's efficiency, the hybrid model ensures higher classification accuracy while maintaining fast processing speeds. The ensemble method used to aggregate predictions balances the strengths of both models, minimizing errors and improving overall sentiment classification. The system achieved significant performance gains, particularly in handling complex tweets that contain subtle emotions, demonstrating the value of integrating these transformer architectures.

As confirmed through testing, the hybrid model achieved a notable accuracy of 78.28% post-retraining, showcasing its ability to learn from previous errors and adapt to evolving linguistic patterns on Twitter. This adaptability ensures that the system stays relevant over time, even as trends and language use change across social media platforms.

In contrast to conventional sentiment analysis tools designed primarily for businesses, this project places a customer-centric focus by providing real-time sentiment insights to assist users in making informed dining decisions. The ability to process tweets in real-time ensures that the system offers timely and accurate feedback, which is essential for consumers relying on social feedback before visiting restaurants.

The scalable architecture of the system, utilizing cloud-based GPUs through Google Colab, ensures efficient training and inference. This design not only allows for smooth handling of large datasets but also provides the flexibility to expand the system to support additional social platforms or multilingual sentiment analysis in the future.

In summary, this project successfully delivers an accurate, efficient, and user-focused sentiment analysis solution. The hybrid model strikes a balance between precision and speed, addressing the key challenges posed by informal social media text. As sentiment analysis continues to evolve, this system offers a versatile and adaptive framework, laying the foundation for future advancements in natural language processing and real-time data-driven decision-making.

6.2 Future Scopes

While the current implementation of the sentiment analysis system has successfully achieved its core objectives, there are several opportunities for further enhancements that can increase the accuracy, inclusiveness, scalability, and functionality of the system. These improvements would ensure that the system remains adaptive to evolving trends, caters to a broader user base, and offers more practical value to end-users.

One important area for improvement is the addition of multilingual capabilities. Currently, the system supports only English-language tweets, limiting its accessibility for non-English-speaking users. As dining experiences and feedback are shared globally in multiple languages, expanding the system to analyse tweets in various languages will make it more inclusive and relevant for a global audience. This can be achieved by incorporating multilingual NLP models like mBERT or XLM-R, which allow sentiment detection across diverse languages.

Another critical challenge lies in the accurate detection of sarcasm and mixed sentiments. Social media posts often contain sarcastic remarks that are difficult for traditional models to interpret correctly. In future work, the system could integrate context-aware deep learning models or sarcasm-specific datasets to improve its ability to recognize such nuances. By addressing this, the system can significantly reduce the number of misclassifications, particularly in ambiguous tweets that express both positive and negative sentiments.

The system’s functionality can also be extended by integrating with additional social media platforms such as Yelp, Facebook, and Instagram, which are rich sources of customer feedback. Analysing sentiment across multiple platforms will provide a comprehensive view of public opinion, offering deeper insights into user preferences and behaviour. This cross-platform analysis will enhance the system’s value, allowing consumers to make more informed decisions based on broader public sentiment.

A promising direction for future development involves tracking sentiment trends over time and generating alerts for significant shifts in user sentiment. For example, the system could detect sudden increases in negative reviews for a restaurant and notify users in real time. This trend-based analysis would help users stay updated with changing public opinions, enhancing their ability to make timely decisions. Such a feature would also enable predictive insights, allowing users to anticipate issues before they escalate.

To further improve the user experience, the system could incorporate advanced data visualization tools. Developing interactive dashboards with features like sentiment heatmaps and comparative graphs across multiple restaurants would offer users a more engaging and insightful experience. Visualization tools would allow users to explore sentiment patterns visually, facilitating better decision-making.

Finally, to increase accessibility, future iterations should focus on optimizing the model for mobile and edge devices. This would ensure that users can access the system on-the-go, even in environments with limited computational resources. Deploying lightweight versions of the model would reduce reliance on cloud-based infrastructure, allowing for faster and more seamless interactions through mobile devices.

In conclusion, these enhancements—ranging from multilingual support and sarcasm detection to platform integration and mobile optimization—will ensure that the system continues to evolve in step with emerging technologies and user needs. By expanding its functionality, improving accuracy, and enhancing accessibility, the project has the potential to become a versatile and impactful tool for real-time sentiment analysis, empowering consumers with deeper insights and enabling them to make more informed decisions.

Closing Remarks

The project has successfully implemented a hybrid sentiment analysis model that addresses the unique challenges of real-time sentiment analysis on social media. By prioritizing customers as the target audience, the system aligns with the growing trend of data-driven consumer decision-making. With further enhancements, such as multilingual support, improved sarcasm detection, and integration with other platforms, the system has the potential to become a versatile and powerful sentiment analysis tool.

This project serves as a steppingstone towards more comprehensive sentiment analysis solutions and lays the foundation for future research and practical applications in the field.

References

[1] A. Krishna, V. Akhilesh, A. Aich, and C. Hegde, ”Sentiment Analysis of Restaurant Reviews Using Machine Learning Techniques,” in Emerging Research in Electronics, Computer Science and Technology, V. Sridhar, M. Padma, and K. Rao, Eds. Singapore: Springer, 2019, vol. 545.

[2] K. D. Raju and B. B. Jayasingh, ”Machine Learning for Sentiment Analysis for Twitter Restaurant Reviews,” Journal of Emerging Science and Technology, vol. 9, no. 5, p. 38, 2018. Available: www.jespublication.com.

[3] D. R. Patil, D. Shukla, A. Kumar, Y. Rajanak, and Y. P. Singh, ”Machine Learning for Sentiment Analysis and Classification of Restaurant Reviews,” in Proc. 3rd Int. Conf. Comput., Analytics and Networks (ICAN), Rajpura, India, 2022, pp. 1-5, doi: 10.1109/ICAN56228.2022.10007390.

[4] M. M. Hamad, M. A. Salih, and R. A. Jaleel, ”Sentiment Analysis of Restaurant Reviews in Social Media Using Naive Bayes,” Applications of Modelling and Simulation, vol. 5, pp. 166-172, 2021.

[5] R. A. Laksono, K. R. Sungkono, R. Sarno, and C. S. Wahyuni, ”Sentiment Analysis of Restaurant Customer Reviews on TripAdvisor using Naive Bayes,” in Proc. 12th Int. Conf. Information and Communication Technology and System (ICTS), Surabaya, Indonesia, 2019, pp. 49-54, doi: 10.1109/ICTS.2019.8850982.

[6] C. Spoorthi, Dr. P. Ravikumar, and M. J. Adarsh, ”Sentiment Analysis of Customer Feedback on Restaurant Reviews,” in Proc. 2nd Int. Conf. Emerging Trends in Science and Technologies for Engineering Systems (ICETSE), 2019. Available: SSRN: https://ssrn.com/abstract=3506637.

[7] M. R. Petrusel and S. G. Limboi, ”A Restaurants Recommendation System: Improving Rating Predictions Using Sentiment Analysis,” in Proc. 21st Int. Symp. Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), Timisoara, Romania, 2019, pp. 190-197, doi: 10.1109/SYNASC49474.2019.00034.

[8] S. Matlatipov, et al., ”Uzbek Sentiment Analysis Based on Local Restaurant Reviews,” arXiv, 2022. doi: 10.48550/arXiv.2205.15930.

[9] K. Zahoor, N. Z. Bawany, and S. Hamid, ”Sentiment Analysis and Classification of Restaurant Reviews using Machine Learning,” in Proc. 21st Int. Arab Conf. Information Technology (ACIT), Giza, Egypt, 2020, pp. 1-6, doi: 10.1109/ACIT50332.2020.9300098.

[10] K. Yadav, ”Sentiment Analysis on Restaurant Review using Hybrid Approach,” Int. J. Res. Appl. Sci. Eng. Technol., vol. 9, pp. 1999-2006, 2021, doi: 10.22214/ijraset.2021.34737.

[11] M. Adnan, R. Sarno, and K. R. Sungkono, ”Sentiment Analysis of Restaurant Review with Classification Approach in the Decision Tree-J48 Algorithm,” in Proc. Int. Seminar on Application for Technology of Information and Communication (iSemantic), Semarang, Indonesia, 2019, pp. 121-126, doi: 10.1109/ISEMANTIC.2019.8884282.

[12] W. L. K. Khine and N. T. T. Aung, ”Applying Deep Learning Approach to Targeted Aspect-based Sentiment Analysis for Restaurant Domain,” in Proc. Int. Conf. Advanced Information Technologies (ICAIT), Yangon, Myanmar, 2019, pp. 206-211, doi: 10.1109/AITC.2019.8920880.

[13] ”Sentiment Analysis using various Machine Learning and Deep Learning Techniques,” Journal of the Nigerian Society of Physical Sciences, vol. 3, no. 4, pp. 385-394, 2021. doi: 10.46481/jnsps.2021.308.

[14] K. Kaviya, C. Roshini, V. Vaidhehi, and J. D. Sweetlin, ”Sentiment Analysis for Restaurant Rating,” in Proc. IEEE Int. Conf. Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), Chennai, India, 2017, pp. 140-145, doi: 10.1109/ICSTM.2017.8089140.

[15] N. Punetha and G. Jain, ”Game Theory and MCDM-based Unsupervised Sentiment Analysis of Restaurant Reviews,” Applied Intelligence, vol. 53, pp. 20152–20173, 2023. doi: 10.1007/s10489-023-04471-1.

[16] M. M. Aguero-Torales, M. J. Cobo, E. Herrera-Viedma, and A. G. Lopez-Herrera, ”A Cloud-based Tool for Sentiment Analysis in Reviews About Restaurants on TripAdvisor,” Procedia Computer Science, vol. 162, pp. 392-399, 2019, ISSN: 1877-0509. doi: 10.1016/j.procs.2019.12.002.

[17] E. Asani, H. Vahdat-Nejad, and J. Sadri, ”Restaurant Recommender System Based on Sentiment Analysis,” Machine Learning with Applications, vol. 6, 2021, 100114, ISSN: 2666-8270. doi: 10.1016/j.mlwa.2021.100114.

[18] P. K. Soni and R. Rambola, ”A Survey on Implicit Aspect Detection for Sentiment Analysis: Terminology, Issues, and Scope,” IEEE Access, vol. 10, pp. 63932-63957, 2022. doi: 10.1109/ACCESS.2022.3183205.s

[19] Ali, Irfan & Hafeez, Abdul & Zafar, Sahar & Ali, Fayyaz & Kumar, Kamlash. (2020). Roman Urdu Headline News Text Classification using RNN, LSTM and

CNN. Advances in Data Science and Adaptive Analysis. 12. 10.1142/S2424922X20500084.

Appendix A: Sample code

1. Train Electra Model

```
# Cell 5: Train Electra Model
electra_model = ElectraForSequenceClassification.from_pretrained('google/electra-small-discriminator', num_labels=3)
optimizer_electra = torch.optim.Adam(electra_model.parameters(), lr=1e-5)

device = torch.device('cuda' if torch.cuda.is_available() else 'cpu')

# Train Electra Model
train_model(electra_model, train_loader_electra, val_loader_electra, optimizer_electra, device)
```

```
Some weights of ElectraForSequenceClassification were not initialized from the model checkpoint at google/electra-small-discriminator and are newly initiali
You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.
Epoch 1/5
Training Loss: 74.6926, Training Accuracy: 0.3890
Validation Loss: 19.0252, Validation Accuracy: 0.5493
Epoch 2/5
Training Loss: 70.4215, Training Accuracy: 0.5685
Validation Loss: 17.3476, Validation Accuracy: 0.6259
Epoch 3/5
Training Loss: 63.5749, Training Accuracy: 0.6817
Validation Loss: 15.6960, Validation Accuracy: 0.7099
Epoch 4/5
Training Loss: 54.6031, Training Accuracy: 0.7904
Validation Loss: 12.9927, Validation Accuracy: 0.8102
Epoch 5/5
Training Loss: 45.4615, Training Accuracy: 0.8301
Validation Loss: 11.0623, Validation Accuracy: 0.8303
```

2. Train RoBERTa Model

```
# Cell 6: Train RoBERTa Model
roberta_model = RobertaForSequenceClassification.from_pretrained('roberta-base', num_labels=3)
optimizer_roberta = torch.optim.Adam(roberta_model.parameters(), lr=1e-5)

# Train RoBERTa Model
train_model(roberta_model, train_loader_roberta, val_loader_roberta, optimizer_roberta, device)
```

```
Some weights of RobertaForSequenceClassification were not initialized from the model checkpoint at roberta-base and are newly initialized: ['classifier.dens
You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.
Epoch 1/5
Training Loss: 69.0929, Training Accuracy: 0.4795
Validation Loss: 9.0476, Validation Accuracy: 0.8303
Epoch 2/5
Training Loss: 31.4688, Training Accuracy: 0.8311
Validation Loss: 6.5822, Validation Accuracy: 0.8723
Epoch 3/5
Training Loss: 19.9527, Training Accuracy: 0.9005
Validation Loss: 6.5125, Validation Accuracy: 0.8777
Epoch 4/5
Training Loss: 13.6562, Training Accuracy: 0.9347
Validation Loss: 6.8489, Validation Accuracy: 0.8759
Epoch 5/5
Training Loss: 8.3921, Training Accuracy: 0.9616
Validation Loss: 7.7725, Validation Accuracy: 0.8631
```

3. Saving models and tokenizing

```python
# Cell 7: Save Electra and RoBERTa Models

# Make sure all tensors in Electra are contiguous before saving
electra_model = electra_model.to('cpu')  # Move to CPU for saving
for param in electra_model.parameters():
    param.data = param.data.contiguous()  # Ensure contiguous tensors

# Save Electra model and tokenizer
electra_model.save_pretrained('electra_saved_model')
electra_tokenizer.save_pretrained('electra_tokenizer')

# Make sure all tensors in RoBERTa are contiguous before saving
roberta_model = roberta_model.to('cpu')  # Move to CPU for saving
for param in roberta_model.parameters():
    param.data = param.data.contiguous()  # Ensure contiguous tensors

# Save RoBERTa model and tokenizer
roberta_model.save_pretrained('roberta_saved_model')
roberta_tokenizer.save_pretrained('roberta_tokenizer')

# Move models back to GPU after saving, if necessary
electra_model = electra_model.to(device)
roberta_model = roberta_model.to(device)
```

```python
import torch
from transformers import ElectraModel, RobertaModel, ElectraTokenizer, RobertaTokenizer

# Load saved models and tokenizers
electra_model = ElectraModel.from_pretrained('electra_saved_model')
roberta_model = RobertaModel.from_pretrained('roberta_saved_model')

electra_tokenizer = ElectraTokenizer.from_pretrained('electra_tokenizer')
roberta_tokenizer = RobertaTokenizer.from_pretrained('roberta_tokenizer')

device = torch.device("cuda" if torch.cuda.is_available() else "cpu")
```

```
Some weights of RobertaModel were not initialized from the model checkpoint at roberta_saved_model and are newly initialized: ['roberta.pooler.dense.bias', 'roberta.pooler.den
You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.
```

4. Training Hybrid Models

```python
hybrid_model = HybridModel(electra_model=electra_model, roberta_model=roberta_model)
optimizer_hybrid = torch.optim.Adam(hybrid_model.parameters(), lr=1e-5)

# Train Hybrid Model
train_model(hybrid_model, train_loader, val_loader, optimizer_hybrid, device)
```

```
Epoch 1 - Training loss: 0.1719057829166851, Accuracy: 0.9538812637329102
Epoch 1 - Validation loss: 0.433257988215025, Accuracy: 0.8795620203018188
Epoch 2 - Training loss: 0.0831500913601124, Accuracy: 0.9757990837097168
Epoch 2 - Validation loss: 0.4567792143672705, Accuracy: 0.8795620203018188
Epoch 3 - Training loss: 0.06724738249432867, Accuracy: 0.9794520735740662
Epoch 3 - Validation loss: 0.5879092735903604, Accuracy: 0.8576642274856567
```

5. Testing the model

```python
# Example list of input texts for prediction
test_texts = [
    "The food was amazing but the service was slow.",
    "The restaurant was dirty and the staff were rude.",
    "I had a fantastic time! The ambiance and food were perfect.",
    "It was an average experience, nothing too special.",
    "Terrible food, I will never come back here again."
]

# Define a function to predict sentiment for multiple sentences
def predict_sentiment(texts):
    predictions = []

    for text in texts:
        # Tokenize the input using both tokenizers
        electra_inputs = electra_tokenizer(text, return_tensors='pt', padding=True, truncation=True)
        roberta_inputs = roberta_tokenizer(text, return_tensors='pt', padding=True, truncation=True)

        # Extract specific components needed for the hybrid model
        electra_input_ids = electra_inputs['input_ids']
        electra_attention_mask = electra_inputs['attention_mask']
        roberta_input_ids = roberta_inputs['input_ids']
        roberta_attention_mask = roberta_inputs['attention_mask']

        # Make predictions using the hybrid model
        with torch.no_grad():
            logits = hybrid_model(
                electra_input_ids, electra_attention_mask,
                roberta_input_ids, roberta_attention_mask
            )

        # Decode predictions and append to the results
        predicted_label = decode_predictions(logits)
        predictions.append(predicted_label)

    return predictions

# Predict sentiment for the list of sentences
predicted_sentiments = predict_sentiment(test_texts)

# Print the results
for text, sentiment in zip(test_texts, predicted_sentiments):
    print(f"Text: {text}\nPredicted Sentiment: {sentiment}\n")
```

```
Text: The food was amazing but the service was slow.
Predicted Sentiment: Negative

Text: The restaurant was dirty and the staff were rude.
Predicted Sentiment: Negative

Text: I had a fantastic time! The ambiance and food were perfect.
Predicted Sentiment: Positive

Text: It was an average experience, nothing too special.
Predicted Sentiment: Neutral

Text: Terrible food, I will never come back here again.
Predicted Sentiment: Negative
```

6. Testing Accuracy

```python
import pandas as pd
from sklearn.metrics import accuracy_score
import torch

# Function to decode the predictions from logits to labels (Assuming 0: Negative, 1: Neutral, 2: Positive)
def decode_predictions(logits):
    probabilities = torch.softmax(logits, dim=1)
    predicted_labels = torch.argmax(probabilities, dim=1).cpu().numpy()  # Convert logits to predicted labels
    return predicted_labels

# Ensure that 'Polarity_A' column has numerical labels (0: Negative, 1: Neutral, 2: Positive)
# First, map any string labels to corresponding numerical values
label_mapping_reverse = {'negative': 0, 'neutral': 1, 'positive': 2}
test_df['Polarity_A'] = test_df['Polarity_A'].replace(label_mapping_reverse)

# Load your test dataset (Assuming 'Sentence' and 'Polarity_A' columns in the test dataset)
test_sentences = test_df['Sentence'].values
true_labels = test_df['Polarity_A'].values  # Assuming encoded as 0, 1, 2 for Negative, Neutral, Positive

# Function to predict sentiment for each test sentence
def predict_sentiments_for_test_data(sentences):
    predicted_labels = []

    # Loop through each sentence in the test data
    for sentence in sentences:
        # Tokenize using both Electra and RoBERTa tokenizers
        electra_inputs = electra_tokenizer(sentence, return_tensors='pt', padding=True, truncation=True, max_length=512).to(device)
        roberta_inputs = roberta_tokenizer(sentence, return_tensors='pt', padding=True, truncation=True, max_length=512).to(device)

        # Extract necessary inputs and send to device (GPU or CPU)
        electra_input_ids = electra_inputs['input_ids'].to(device)
        electra_attention_mask = electra_inputs['attention_mask'].to(device)
        roberta_input_ids = roberta_inputs['input_ids'].to(device)
        roberta_attention_mask = roberta_inputs['attention_mask'].to(device)

        # Make predictions using the hybrid model
        with torch.no_grad():
            logits = hybrid_model(
                electra_input_ids, electra_attention_mask,
                roberta_input_ids, roberta_attention_mask
            )
```

```python
        # Decode predictions to get predicted label
        predicted_label = decode_predictions(logits)
        predicted_labels.append(predicted_label[0])  # Get the first label from the batch (assuming batch size of 1)

    return predicted_labels

# Get predicted labels for the test dataset
predicted_labels = predict_sentiments_for_test_data(test_sentences)

# Convert predicted labels to a readable format (Optional: Assuming 0: Negative, 1: Neutral, 2: Positive)
label_mapping = {0: 'Negative', 1: 'Neutral', 2: 'Positive'}
predicted_labels_readable = [label_mapping[label] for label in predicted_labels]

# Convert true labels to readable format as well
true_labels_readable = [label_mapping[label] for label in true_labels]

# Create a dataframe to show the test sentences, true labels, and predicted labels
test_results_df = pd.DataFrame({
    'Sentence': test_sentences,
    'True Label': true_labels_readable,
    'Predicted Label': predicted_labels_readable
})

# Print the test results dataset
print(test_results_df)

# Calculate testing accuracy
test_accuracy = accuracy_score(true_labels, predicted_labels)

# Print the testing accuracy
print(f"Testing Accuracy: {test_accuracy * 100:.2f}%")
```

```
C:\Users\sujal\AppData\Local\Temp\ipykernel_9908\395595950.py:14: FutureWarning: Downcasting behavior in `replace` is deprecated and will be removed in a future version. To retain the old behavior, e
  test_df['Polarity_A'] = test_df['Polarity_A'].replace(label_mapping_reverse)
                                              Sentence True Label  \
0                        The bread is top notch as well.   Positive
1    I have to say they have one of the fastest del...   Positive
2          Food is always fresh and hot- ready to eat!   Positive
3        Did I mention that the coffee is OUTSTANDING?   Positive
4    Certainly not the best sushi in New York, howe...   Positive
..                                                 ...        ...
612   The service was prompt, but not overly friendly.    Neutral
613          The dessert was okay, but not exciting.    Neutral
614        The bread was soft, but not very flavorful.    Neutral
615   The steak was cooked well, but lacked seasoning.    Neutral
616              The drinks were cold, but a bit weak.    Neutral

    Predicted Label
0          Positive
1          Positive
2          Positive
3          Positive
4          Positive
..              ...
612         Neutral
613         Neutral
614         Neutral
615         Neutral
616         Neutral

[617 rows x 3 columns]
Testing Accuracy: 74.88%
```

7. Re-training the model

```python
import torch
from torch.utils.data import DataLoader, TensorDataset
from torch.optim import AdamW
from torch.nn import CrossEntropyLoss
from transformers import get_scheduler

# Assuming your train_dataset is available (you can adjust this part to your specific data format)
train_sentences = train_df['Sentence'].values
train_labels = train_df['Polarity_A'].values  # Assuming this is the column for sentiment labels

# Tokenize the training data
electra_tokenized_train = electra_tokenizer(train_sentences.tolist(), padding=True, truncation=True, return_tensors='pt')
roberta_tokenized_train = roberta_tokenizer(train_sentences.tolist(), padding=True, truncation=True, return_tensors='pt')

# Convert labels to tensors (as LongTensor, since CrossEntropyLoss expects this)
train_labels_tensor = torch.tensor(train_labels, dtype=torch.long)  # Corrected to LongTensor

# Create a TensorDataset to combine inputs and labels
train_dataset = TensorDataset(
    electra_tokenized_train['input_ids'], electra_tokenized_train['attention_mask'],
    roberta_tokenized_train['input_ids'], roberta_tokenized_train['attention_mask'],
    train_labels_tensor
)

# Create the DataLoader for the training dataset
train_dataloader = DataLoader(train_dataset, batch_size=16, shuffle=True)

# Define the number of epochs
num_epochs = 3

# Define the optimizer with a lower learning rate
optimizer = AdamW(hybrid_model.parameters(), lr=2e-5)

# Define the loss function (CrossEntropyLoss for classification tasks)
criterion = CrossEntropyLoss()

# Define the scheduler (optional, helps with learning rate decay)
num_training_steps = num_epochs * len(train_dataloader)
lr_scheduler = get_scheduler(
    name="linear", optimizer=optimizer, num_warmup_steps=0, num_training_steps=num_training_steps
)
```

```python
# Training loop (train for the defined number of epochs)
for epoch in range(num_epochs):
    hybrid_model.train()
    total_train_loss = 0
    correct_train_predictions = 0
    total_train_examples = 0

    for batch in train_dataloader:
        optimizer.zero_grad()

        # Extract inputs for Electra and RoBERTa
        electra_input_ids = batch[0].to(device)
        electra_attention_mask = batch[1].to(device)
        roberta_input_ids = batch[2].to(device)
        roberta_attention_mask = batch[3].to(device)
        labels = batch[4].to(device)

        # Forward pass through the hybrid model
        logits = hybrid_model(
            electra_input_ids, electra_attention_mask,
            roberta_input_ids, roberta_attention_mask
        )

        # Compute loss
        loss = criterion(logits, labels)  # Ensure labels are LongTensor
        total_train_loss += loss.item()

        # Backpropagation
        loss.backward()
        optimizer.step()
        lr_scheduler.step()

        # Compute accuracy
        predictions = torch.argmax(logits, dim=1)
        correct_train_predictions += (predictions == labels).sum().item()
        total_train_examples += labels.size(0)

    # Print train accuracy for this epoch
    train_accuracy = correct_train_predictions / total_train_examples
    print(f"Epoch {epoch + 1}/{num_epochs}, Train Loss: {total_train_loss:.4f}, Train Accuracy: {train_accuracy:.4f}")

# After training, perform testing again
predicted_labels = predict_sentiments_for_test_data(test_sentences)

# Calculate new testing accuracy
test_accuracy = accuracy_score(true_labels, predicted_labels)

# Print the testing accuracy after retraining
print(f"Testing Accuracy after retraining: {test_accuracy * 100:.2f}%")
```

```
Epoch 1/3, Train Loss: 31.1896, Train Accuracy: 0.9463
Epoch 2/3, Train Loss: 12.7857, Train Accuracy: 0.9792
Epoch 3/3, Train Loss: 7.2747, Train Accuracy: 0.9879
Testing Accuracy after retraining: 78.28%
```

Appendix B: Data Sheets

1. Training Dataset

	A	B	C
1	S No	Sentence	Polarity_A
2	1	But the staff was so horrible to us.	negative
3	2	To be completely fair, the only redeeming factor was the food, which was above average, but couldn't make up for all the other deficiencies of Teodora.	positive
4	3	The food is uniformly exceptional, with a very capable kitchen which will proudly whip up whatever you feel like eating, whether it's on the menu or not.	positive
5	4	Where Gabriela personaly greets you and recommends you what to eat.	positive
6	5	For those that go once and don't enjoy it, all I can say is that they just don't get it.	positive
7	6	Not only was the food outstanding, but the little 'perks' were great.	positive
8	7	It is very overpriced and not very tasty.	negative
9	8	Our agreed favorite is the orrechiete with sausage and chicken (usually the waiters are kind enough to split the dish in half so you get to sample both meats).	positive
10	9	The Bagels have an outstanding taste with a terrific texture, both chewy yet not gummy.	positive
11	10	Nevertheless the food itself is pretty good.	positive
12	11	They did not have mayonnaise, forgot our toast, left out ingredients (ie cheese in an omelet), below hot temperatures and the bacon was so over cooked it crumbled on the plate when you touched it.	negative
13	12	It took half an hour to get our check, which was perfect since we could sit, have drinks and talk!	positive
14	13	The design and atmosphere is just as good.	positive
15	14	He has visited Thailand and is quite expert on the cuisine.	positive
16	15	I would definitely recommend Mary's and am making it one of my regular neighborhood haunts.	positive
17	16	The pizza is the best if you like thin crusted pizza.	positive
18	17	My Girlfriend and I stumbled onto this hopping place the other night and had a great time!	positive
19	18	He can't help, they're bought up so fast.	neutral
20	19	Secondly, on this night the place was overwhelmed by upper east side ladies perfume.	negative
21	20	i have eaten here a handful of times, for no reason besides sheer convenience.	neutral
22	21	Obviously run by folks who know a pie.	positive
23	22	We were very disappointed.	negative
24	23	IT IS DEFINITELY SPECIAL AND AFFORDABLE.	positive
25	24	I really liked this place.	positive
2707	2706	The drinks were served quickly, but tasted average.	neutral
2708	2707	The meal portions were average, neither too much nor too little.	neutral
2709	2708	The restaurant's decor was simple, but clean.	neutral
2710	2709	The staff was efficient, though not particularly engaging.	neutral
2711	2710	The food quality was good, though the seasoning was mild.	neutral
2712	2711	The dining experience was neutral, without any highlights.	neutral
2713	2712	The restaurant had a calm atmosphere, suitable for a casual meal.	neutral
2714	2713	The service was steady, with no noticeable delays.	neutral
2715	2714	The food was served warm and tasted fine.	neutral
2716	2715	The menu offered common choices, without any surprises.	neutral
2717	2716	The restaurant was moderately busy, but the noise was manageable.	neutral
2718	2717	The food was prepared well, though the flavors were not strong.	neutral
2719	2718	The service met expectations, but nothing more.	neutral
2720	2719	The overall dining experience was acceptable.	neutral
2721	2720	The ambiance was neutral, making it a comfortable place to eat.	neutral
2722	2721	The meal was filling, though the taste was a bit bland.	neutral
2723	2722	The restaurant was tidy, with an unremarkable interior.	neutral
2724	2723	The service was prompt, though a bit impersonal.	neutral
2725	2724	The food was delivered quickly, without issues.	neutral
2726	2725	The menu items were typical, without any standout dishes.	neutral
2727	2726	The restaurant had a calm vibe, though not particularly lively.	neutral
2728	2727	The staff was polite, though not very interactive.	neutral
2729	2728	The food was served in a timely manner, meeting expectations.	neutral
2730	2729	The restaurant had a basic design, but it was clean.	neutral
2731	2730	The service was good, but interactions were minimal.	neutral
2732	2731	The food was satisfactory, though the flavors were muted.	neutral
2733	2732	The ambiance was neutral, without much character.	neutral
2734	2733	The portion sizes were fair for the price.	neutral
2735	2734	The meal arrived on time, as expected.	neutral
2736	2735	The drinks were served at the right temperature.	neutral
2737	2736	The restaurant was quiet, making it a peaceful place to dine.	neutral
2738	2737	The food was fine, but nothing stood out.	neutral
2739	2738	The staff was efficient, but not particularly engaging.	neutral

2. Testing Dataset

	A	B	C
1	S No.	Sentence	Polarity_A
2	1	The bread is top notch as well.	positive
3	2	I have to say they have one of the fastest delivery times in the city.	positive
4	3	Food is always fresh and hot- ready to eat!	positive
5	4	Did I mention that the coffee is OUTSTANDING?	positive
6	5	Certainly not the best sushi in New York, however, it is always fresh, and the place is very clean, sterile.	positive
7	6	I trust the people at Go Sushi, it never disappoints.	positive
8	7	Straight-forward, no surprises, very decent Japanese food.	neutral
9	8	BEST spicy tuna roll, great asian salad.	positive
10	9	Try the rose roll (not on menu).	positive
11	10	I love the drinks, esp lychee martini, and the food is also VERY good.	positive
12	11	In fact, this was not a Nicoise salad and was barely eatable.	negative
13	12	While there's a decent menu, it shouldn't take ten minutes to get your drinks and 45 for a dessert pizza.	neutral
14	13	Once we sailed, the top-notch food and live entertainment sold us on a unforgettable evening.	positive
15	14	Our waiter was horrible; so rude and disinterested.	negative
16	15	The sangria's - watered down.	negative
17	16	menu - uneventful, small.	neutral
18	17	Anytime and everytime I find myself in the neighborhood I will go to Sushi Rose for fresh sushi and great portions all at a reasonable price.	positive
19	18	Great food but the service was dreadful!	positive
20	19	The portions of the food that came out were mediocre.	negative
21	20	the two waitress's looked like they had been sucking on lemons.	negative
22	21	From the beginning, we were met by friendly staff memebers, and the convienent parking at Chelsea Piers made it easy for us to get to the boat.	positive
23	22	We enjoyed ourselves thoroughly and will be going back for the desserts	positive
24	23	Desserts are almost incredible: my personal favorite is their Tart of the Day.	positive
25	24	I am surprised at the lower reviews because it is definitely better than other places I have tried with higher ratings.	positive
26	25	Maybe the secret was that we went on a Sunday night and everything was great.	positive

	A	B	C
589	588	The drinks were cold, but not strong enough.	neutral
590	589	The atmosphere was fine, but not very lively.	neutral
591	590	The soup was served hot, but lacked flavor.	neutral
592	591	The pasta was cooked well, but tasted a bit bland.	neutral
593	592	The fish was fresh, but lacked seasoning.	neutral
594	593	The appetizers were fine, but not particularly exciting.	neutral
595	594	The service was polite, but not very warm.	neutral
596	595	The dessert was okay, but not very memorable.	neutral
597	596	The bread was fresh, but tasted plain.	neutral
598	597	The chicken was juicy, but could've used more seasoning.	neutral
599	598	The drinks were fine, but a bit weak.	neutral
600	599	The atmosphere was casual, but a bit dull.	neutral
601	600	The soup was hot, but lacked seasoning.	neutral
602	601	The steak was cooked as requested, but lacked flavor.	neutral
603	602	The pasta was cooked well, but tasted a bit bland.	neutral
604	603	The fish was fresh, but not very flavorful.	neutral
605	604	The service was fine, but not particularly warm.	neutral
606	605	The dessert was sweet, but could've been better.	neutral
607	606	The bread was warm, but lacked flavor.	neutral
608	607	The chicken was moist, but needed more seasoning.	neutral
609	608	The drinks were cold, but a bit watered down.	neutral
610	609	The atmosphere was fine, but a bit plain.	neutral
611	610	The soup was warm, but not very flavorful.	neutral
612	611	The pasta was cooked well, but lacked seasoning.	neutral
613	612	The fish was fresh, but could've used more flavor.	neutral
614	613	The service was prompt, but not overly friendly.	neutral
615	614	The dessert was okay, but not exciting.	neutral
616	615	The bread was soft, but not very flavorful.	neutral
617	616	The steak was cooked well, but lacked seasoning.	neutral
618	617	The drinks were cold, but a bit weak.	neutral