Sentimental Analysis and Topic Modeling End to End Application

A PROJECT REPORT

submitted by

CB.EN.U4ELC20011	**Avanish Jha**
CB.EN.U4ELC20056	**Ratneshwar Kumar Bharti**
CB.EN.U4ELC20057	**Ravi Gupta**
CB.EN.U4ELC20060	**Rudraksh Singh**

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

IN

ELECTRICAL AND COMPUTER ENGINEERING

AMRITA
VISHWA VIDYAPEETHAM
DEEMED TO BE UNIVERSITY

AMRITA SCHOOL OF ENGINEERING, COIMBATORE

AMRITA VISHWA VIDYAPEETHAM

COIMBATORE - 641 112

OCTOBER 2023

AMRITA VISHWA VIDYAPEETHAM

AMRITA SCHOOL OF ENGINEERING,COIMBATORE -641 112

AMRITA
VISHWA VIDYAPEETHAM
DEEMED TO BE UNIVERSITY

BONAFIDE CERTIFICATE

This is to certify that this project entitled **“Sentimental Analysis and Topic Modeling End to End Application”** submitted by

CB.EN.U4ELC20011	**Avanish Jha**
CB.EN.U4ELC20056	**Ratneshwar Kumar Bharti**
CB.EN.U4ELC20057	**Ravi Gupta**
CB.EN.U4ELC20060	**Rudraksh Singh**

in partial fulfillment of the requirements for the award of the **Degree of Bachelor of Technology** in **ELECTRICAL & COMPUTER ENGINEERING** is a bonafide record of the work carried out under my guidance and supervision at Amrita School of Engineering.

R. Ranjith

R Ranjith
Supervisor
Assistant Professor (Sr. Gr.)
Department of Electrical and
Electronics Engineering
Amrita School of Engineering
Coimbatore- 641112

Balamurugan S
Chairperson
Professor
Department of Electrical and
Electronics Engineering
Amrita School of Engineering
Coimbatore- 641112

This project report was evaluated by us on……………

INTERNAL EXAMINER EXTERNAL EXAMINER

ABSTRACT

In our rapidly digitizing world, the exponential surge in user-generated content across various platforms demands robust and agile processing tools for extracting actionable insights. Addressing this imperative, our project showcases an avant-garde mobile application that seamlessly integrates advanced Natural Language Processing (NLP) techniques, with a primary focus on Sentiment Analysis (SA) and Topic Modeling.

For Sentiment Analysis, we harness the power of Logistic Regression, a renowned statistical and machine learning method tailored for binary outcomes. By modeling the log odds of the probability of a particular sentiment based on textual features, Logistic Regression offers both efficiency and interpretability. Its strength lies in its ability to handle non-linear relationships, its clarity in presenting the impact of individual words or phrases on sentiment, and its adaptability to various text classification challenges. In a mobile environment, these attributes translate to real-time, accurate sentiment categorizations, empowering businesses with immediate feedback on user sentiments and facilitating data-driven decision-making.

Concurrently, the application's Topic Modeling facet is powered by the Latent Dirichlet Allocation (LDA) method. LDA, a generative probabilistic model, adeptly uncovers concealed thematic structures in extensive text collections. By deciphering the blend of topics within documents and the assortment of words within topics, LDA delivers a nuanced understanding of dominant discussions and narratives.

Preliminary assessments highlight the application's prowess in rendering precise sentiment determinations and identifying cogent topic categorizations. By synthesizing methodologies and insights from leading-edge research, this mobile application stands as a beacon of the confluence of academic rigor and practical utility, setting a benchmark for future endeavors in mobile-centric NLP solutions.

CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

NLP	Natural Language Processing
SVM	Support Vector Machine
LR	Logistic Regression
NB	Naive Bayes
Max-Ent	Maximum Entropy
LDA	Latent Dirichlet Allocation
NMF	Non Negative Matrix Factorization
LSA	Latent Semantic Analysis

Chapter 1

INTRODUCTION

1.1 Introduction to Sentiment Analysis (SA)

1.1.1 Background and Context

Sentiment Analysis (SA), colloquially known as opinion mining, emerges from the confluence of Natural Language Processing (NLP) and text analytics. Its primary goal is to discern the sentiment or emotion encapsulated within textual data, whether it's positive, negative, or neutral, and in some advanced systems, even more nuanced emotions like joy or disappointment[1]. As the digital landscape burgeoned with user-generated content from various platforms, ranging from social media to e-commerce websites, the significance of SA has been accentuated.

1.1.2 Significance of SA in Mobile Applications

The proliferation of mobile apps has created new channels for users to express their opinions and feelings. User reviews, ratings, social media mentions, and other unstructured text data contain a wealth of sentiment information. Performing sentiment analysis on this data can uncover trends and patterns that would otherwise be difficult to detect manually.For example, an app developer could analyze user reviews over time to identify major peaks in negative sentiment. This could signal issues with a recent app update that have frustrated users. By drilling down on the specific features or bugs mentioned in those negative reviews, developers can pinpoint areas to fix.

Sentiment analysis can also reveal differences in how users perceive new features or design changes. A common application is comparing sentiment before and after an app redesign. A significant increase in negative sentiment could indicate the update has been received poorly by users. Analyzing the sentiment shift can provide guidance on where the update missed the mark for users.[2]

Overall, sentiment analysis provides mobile app developers an invaluable listening tool. By continuously monitoring user feedback and sentiment patterns, developers can identify problems early, understand user pain points, and gain insights to guide the app optimization process. This ultimately leads to higher user retention, engagement, and satisfaction.

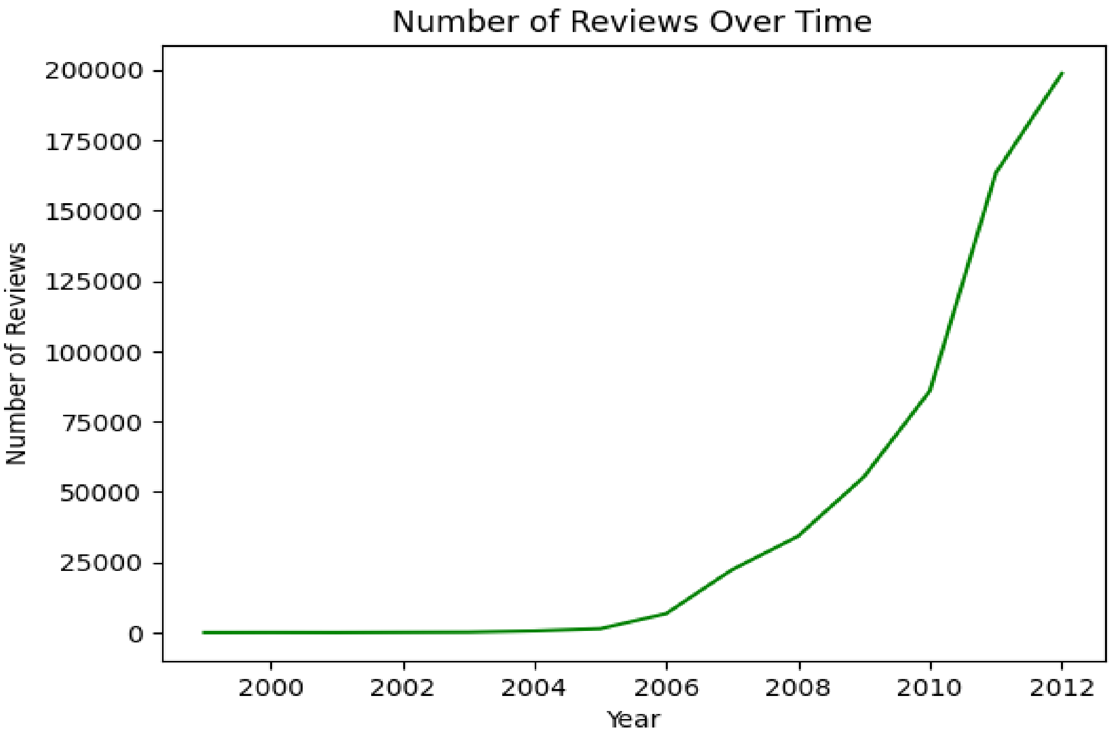


Figure 1.1: Number of Reviews Over Time for Amazon Food Reviews (1999-2012)

1.2 Introduction to Topic Modeling

1.2.1 Background and Context

Topic Modeling is an unsupervised machine learning technique that identifies topics in a large volume of text. By analyzing the co-occurrence patterns of words, it seeks to extract clusters or ’topics’ that represent semantic themes present within the documents[5]. Originating from the realms of text mining and Natural Language Processing (NLP), its foundations are rooted in the need to understand and categorize vast textual datasets, which, in the age of digital information, are growing exponentially.

1.2.2 Significance of Topic Modeling in Mobile Applications

Mobile applications, which are a nexus of user interactions and feedback, generate a plethora of textual data. This data, though rich in insights, is often unstructured and vast. Topic Modeling stands as a beacon in this context, helping categorize user feedback into coherent themes or topics, aiding developers in understanding common concerns, praises, or areas of improvement[5]. Beyond feedback, Topic Modeling can also be instrumental in content recommendation within apps, ensuring users receive information most relevant to their interests.

While Topic Modeling offers promising results, its application in the mobile domain is fraught with challenges. The concise nature of feedback, the dynamism of user interactions, and the evolving nature of app features make the extraction of stable topics challenging. Additionally, understanding the optimal number of topics or ensuring they remain interpretable and distinct can be complex[6]

1.3 Literature Review

Sentiment Analysis (SA) and Topic Modeling are pivotal techniques in the realm of Natural Language Processing (NLP), and their importance in extracting valuable insights from user-generated content in mobile applications cannot be overstated. As mobile platforms burgeon with user reviews, feedback, and discussions, effective techniques like Logistic Regression for SA and LDA for Topic Modeling are paramount. This literature survey provides a deep dive into these methodologies based on several key works.

Sentiment Analysis of Product Reviews[1]:

This comprehensive review underscores the significance of understanding consumer sentiments from product reviews. The myriad of methodologies explored in this paper, from traditional machine learning to deep learning techniques, underscores the versatility and challenges of SA. Notably, the effectiveness of logistic regression, a relatively simple yet powerful method for SA, is highlighted.

NLP in Customer Service[2]:

NLP's transformative role in enhancing customer service interactions is explored in this work. The paper elaborates on various methodologies, with a notable mention of logistic regression's efficacy in real-time SA, making it invaluable for immediate feedback systems in customer service platforms.

Comparison of Different Machine Learning Algorithms for Sentiment Analysis[3]:

In an era where a multitude of algorithms exists for SA, this paper's comparative analysis stands out. The strength of logistic regression, particularly its interpretability and efficiency, is discussed. Its performance metrics, in comparison to other algorithms, provide valuable insights for its selection in SA projects.

NLP: Current Trends[4]:

This paper offers a panoramic view of the current trends in NLP. From sentiment analysis to topic modeling and other NLP applications, the authors discuss the challenges and opportunities inherent in the field. The study highlights the rapid advancements in deep learning and their implications in NLP.

Different Topic Modeling Models[5]:

Topic modeling's essence lies in extracting structured topics from vast unstructured data. This research dives deep into various topic modeling techniques, with LDA standing out due to its effectiveness and widespread usage. The paper provides a thorough understanding of LDA's methodology, its applications, and the insights it can offer.

Sentiment Analysis Using VADER and Logistic Regression Techniques[6]:

This study's focus on logistic regression offers a detailed perspective on its application in SA. The comparison with VADER, another sentiment analysis tool, provides a comprehensive understanding of logistic regression's nuances, strengths, and potential areas of improvement.

A Survey on Sentiment Analysis Methods, Applications, and Challenges[7]:

This overarching survey provides a holistic view of SA, discussing from rule-based methods to machine learning techniques. The emphasis on logistic regression's applicability across diverse platforms, from social media to e-commerce, underscores its importance in the SA domain.

In conclusion, the combination of Logistic Regression for Sentiment Analysis and LDA for Topic Modeling promises robust and effective insights, especially in mobile applications. The literature provides a foundational understanding of these techniques, their strengths, and their challenges, guiding the successful implementation of a ”Sentiment Analysis & Topic Modeling Mobile Application” project.

1.4 Objectives

For this project, the following tasks have to be considered as objectives:

1. Real-time feedback and insights for users.
2. Enhancing business decisions through user feedback analysis.
3. Creation of a mobile application integrating sentiment analysis and topic modeling.

1.5 Report Outline

The report consists of the introductory chapter and the other chapters as follows:

Chapter 2 describes the Algorithms used for SA

Chapter 3 gives a detailed description about the dataset.

Chapter 4 gives Comparison between different algorithms. Results and Analysis is shown in this chapter.

Chapter 2

THEORETICAL FRAMEWORK

The theoretical underpinnings of any research endeavor provide the bedrock upon which practical implementations are built. In the realm of text analytics, particularly Sentiment Analysis and Topic Modeling, there exist intricate mathematical and algorithmic foundations that guide the extraction of meaningful insights from textual data.

2.1 Theoretical Foundations of Sentiment Analysis

Sentiment Analysis, at its core, is a classification problem where pieces of text are categorized into predefined sentiment classes[1]. Depending on the granularity required, this can range from binary classification (positive/negative) to multi-class (e.g., positive/neutral/negative) or even multi-label scenarios.

2.1.1 Logistic Regression

One of the most widely-used algorithms for such classification tasks is Logistic Regression. Unlike its name suggests, Logistic Regression is used for binary classification problems, and its extension, Multinomial Logistic Regression, tackles multi-class problems.

Nature of the Algorithm:

- Logistic Regression operates by estimating probabilities using a logistic (sigmoid) function. This ensures that the estimated probabilities are between 0 and 1, making them interpretable as class probabilities.
- The decision boundary in Logistic Regression is linear, which means it works best when the data points of different classes can be separated by a straight line (in 2D), a plane (in 3D), or a hyperplane (in higher dimensions).

Mathematical Formulation:

Given a feature vector x, the probability $P(Y = 1|x)$

$$P(Y = 1|x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

Here e is the base of natural logarithms, and B_0 and B_1 are the parameters of the model.

Applications and Insights:

- Beyond its vast applications, Logistic Regression's strength lies in its interpretability. Each feature's weight provides insights into its impact on the sentiment classification[4].

2.1.2 Gradient Boosting

Gradient Boosting is an ensemble learning technique that combines the predictions of multiple weak models, usually decision trees, to create a strong predictive model. It operates in an iterative manner, with each new model correcting the errors of the previous ones, gradually improving overall predictive performance.

Nature of the Algorithm:

- Gradient Boosting is an ensemble learning method that builds a strong predictive model by combining the predictions of multiple weak models, usually decision trees.
- It works in an iterative manner, where each new tree corrects the errors of the previous ones, gradually improving the model's performance.

Mathematical Formulation:

1. **Initialization**:
   Start with an initial estimate, which can be the average of the target values (for regression problems) or the log odds ratio (for classification problems).

   $$F_0(x) = \arg\min_{\gamma} \sum_{i=1}^{N} L(y_i, \gamma)$$

2. **Iterative Updates**:
   For each stage $m = 1, 2, \ldots, M$, where M is the number of boosting iterations:

(a) **Compute Pseudo-Residuals**:

Calculate the negative gradient (pseudo-residuals) of the loss function $L(y, F)$ with respect to the model predictions F at the previous step.

$$r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$$

(b) **Fit a Weak Learner**:

Fit a weak learner (usually a decision tree) to the pseudo-residuals.

$$h_m(x) = fit(x, r_{im})$$

(c) **Compute Multiplier**:

Compute a multiplier γ_m that minimizes the loss when added to the current model.

$$\gamma_m = \arg\min_{\gamma} \sum_{i=1}^{N} L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i))$$

(d) **Update the Model**:

Update the model with the weak learner scaled by the multiplier.

$$F_m(x) = F_{m-1}(x) + \nu\gamma_m h_m(x)$$

Here, ν is the learning rate, a parameter that scales the contribution of each tree.

3. **Final Model**:

The final model is a sum of the initial estimate and the contributions from all the weak learners.

$$F_M(x) = F_0(x) + \nu \sum_{m=1}^{M} \gamma_m h_m(x)$$

Applications and Insights:

- Gradient Boosting is widely used for both regression and classification problems. It can be applied to detect anomalies in data. It is robust to outliers and can handle complex relationships in data[3].

2.1.3 Support Vector Machines

A supervised learning algorithm that is used for classification and regression. SVM finds the optimal hyperplane that separates data points of different classes in feature space. It aims to maximize the margin between the classes, and it can handle both linear and non-linear decision boundaries through the use of kernel functions.

Nature of the Algorithm:

- SVM is a supervised learning algorithm used for classification and regression.
- It works by finding the hyperplane that best separates the data into different classes while maximizing the margin between the classes.

Mathematical Formulation:

1. **Data Representation**:
   Consider a set of training examples $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ where each x_i is a feature vector representing a data point (e.g., a text document) and y_i is the class label (e.g., +1 for positive and -1 for negative sentiment).

2. **Objective Function**:
   The goal is to find a hyperplane defined by the weight vector w and bias b that separates the classes with the maximum margin. The hyperplane can be represented by the equation $w \cdot x + b = 0$.

3. **Optimization Problem**:
   To find the optimal w and b, solve the following optimization problem:

   $$\min_{w,b} \frac{1}{2} \|w\|^2$$

   subject to the constraints:

   $$y_i(w \cdot x_i + b) \geq 1, \quad for alli = 1, \ldots, n$$

4. **Support Vectors**:
   Data points that lie on the margins (defined by $w \cdot x + b = \pm 1$) are called support vectors.

These are the critical elements of the training set as they are the closest to the hyperplane and determine its position and orientation.

5. **Sentiment Classification**:
   Once the model is trained, a new text document x can be classified by evaluating the sign of $w \cdot x + b$. A positive sign indicates a positive sentiment, and a negative sign indicates a negative sentiment.

Applications and Insights:

- SVM is used in image classification tasks. It is employed in text classification problems. It performs well even in high-dimensional spaces. The regularization parameter in SVM helps prevent overfitting.[1].

2.1.4 Random Forest

An ensemble learning method that constructs a collection of decision trees during training. Each tree is trained on a random subset of the data, and a random subset of features is considered at each split. The final prediction is made by aggregating the predictions of all individual trees, providing a robust and accurate model.

Nature of the Algorithm:

- Random Forest is an ensemble learning method that constructs a multitude of decision trees during training.
- It outputs the mode of the classes for classification problems or the mean prediction for regression problems.

Mathematical Formulation:

Random Forest is an ensemble learning method that operates by constructing multiple decision trees during training and outputting the class that is the mode of the classes (classification) of the individual trees.

1. **Training Phase**:

   - Given a training set $X = \{x_1, x_2, ..., x_n\}$ with corresponding target values $Y = \{y_1, y_2, ..., y_n\}$, where each x_i is a feature vector representing a document and y_i is its sentiment label.
   - A number of decision trees are constructed. For each tree:

     (a) A random sample of the training set is selected with replacement (bootstrap sample).

     (b) At each node of the tree, a random subset of features is chosen, and the best split on these features is used to split the node. The process is repeated recursively.

2. **Prediction Phase**:

   - For a new document represented by a feature vector x, each tree in the forest makes a prediction about the sentiment.
   - The final prediction is made based on the majority vote of all the trees in the forest.

$$\hat{y} = mode\{tree_1(x), tree_2(x), ..., tree_k(x)\}$$

   where $tree_i(x)$ is the prediction of the i-th tree.

Note: Random Forest is particularly effective for sentiment analysis because it can handle high-dimensional feature spaces and complex data structures often found in text data.

Applications and Insights:

- It is used in image recognition tasks. It is applied in credit scoring models. It is used for tasks like gene expression analysis. It is less prone to overfitting compared to individual decision trees. It provides a measure of the importance of each feature in the prediction.[2].

Chapter 3

DATA DESCRIPTION

Amazon Fine Food Reviews:

Context

This dataset consists of reviews of fine foods from amazon. The data span a period of more than 10 years, including all approximating 500,000 reviews up to October 2012. Reviews include product and user information, ratings, and a plain text review. It also includes reviews from all other Amazon categories.

Data Includes:

- Reviews from Oct 1999 - Oct 2012
- 568,454 reviews
- 256,059 users
- 74,258 products
- 260 users with > 50 reviews

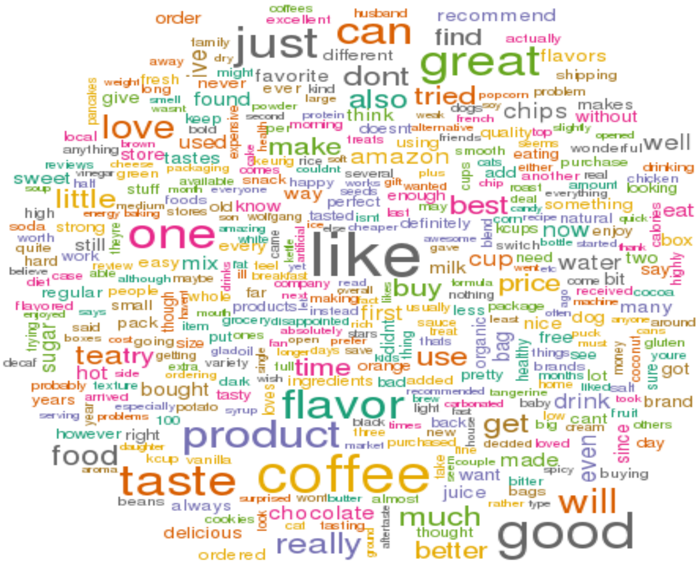
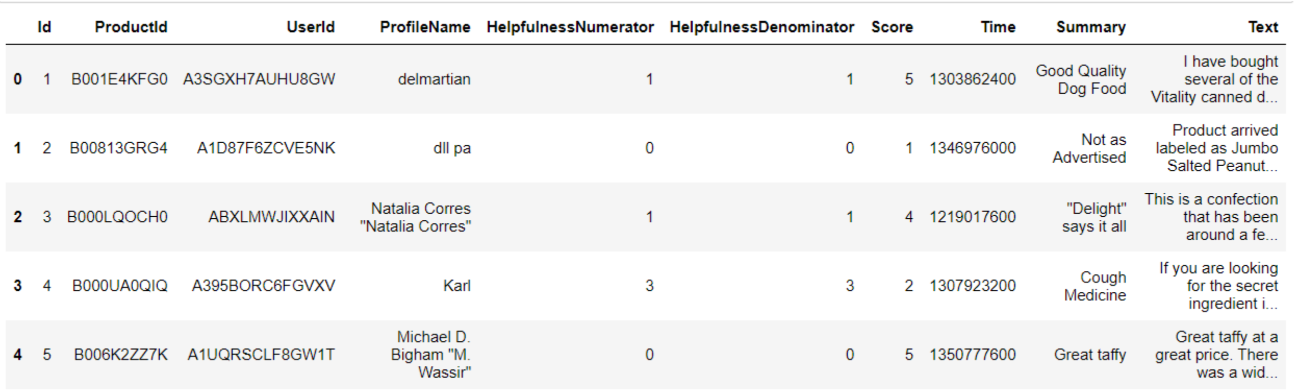


Figure 3.1: Word cloud of the reviews.

Number of Attributes/Columns in data: 10

Attribute Information:

- Id
- ProductId - unique identifier for the product
- UserId - unqiue identifier for the user
- ProfileName
- HelpfulnessNumerator - number of users who found the review helpful
- HelpfulnessDenominator - number of users who indicated whether they found the review helpful or not.
- Score - rating between 1 and 5
- Time - timestamp for the review
- Summary - brief summary of the review
- Text - text of the review



	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time	Summary	Text
0	1	B001E4KFG0	A3SGXH7AUHU8GW	delmartian	1	1	5	1303862400	Good Quality Dog Food	I have bought several of the Vitality canned d...
1	2	B00813GRG4	A1D87F6ZCVE5NK	dll pa	0	0	1	1346976000	Not as Advertised	Product arrived labeled as Jumbo Salted Peanut...
2	3	B000LQOCH0	ABXLMWJIXXAIN	Natalia Corres "Natalia Corres"	1	1	4	1219017600	"Delight" says it all	This is a confection that has been around a fe...
3	4	B000UA0QIQ	A395BORC6FGVXV	Karl	3	3	2	1307923200	Cough Medicine	If you are looking for the secret ingredient i...
4	5	B006K2ZZ7K	A1UQRSCLF8GW1T	Michael D. Bigham "M. Wassir"	0	0	5	1350777600	Great taffy	Great taffy at a great price. There was a wid...

Figure 3.2: First five rows of the Amazon Fine Food Reviews Dataset.

Chapter 4

RESULTS AND ANALYSIS

4.1 Logistic Regression

4.1.1 Accuracy: 86.27%

Classification Report:

Table 4.1: Classification Report for Logistic Regression

	Precision	Recall	F1-Score	Support
Class 0	0.73	0.65	0.69	16181
Class 1	0.51	0.16	0.25	8485
Class 2	0.89	0.97	0.93	89025

Confusion Matrix:

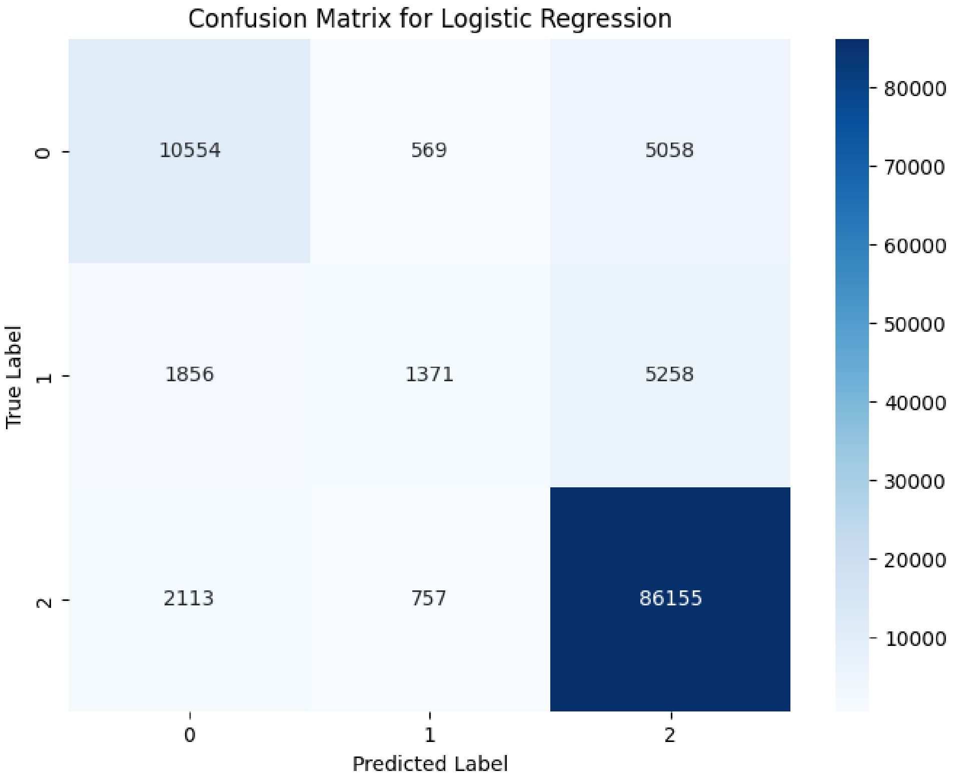


Figure 4.1: Confusion Matrix for Logistic Regression

4.2 Support Vector Machines (SVM)

4.2.1 Accuracy: 86.15%

Classification Report:

Table 4.2: Classification Report for SVM

	Precision	Recall	F1-Score	Support
Class 0	0.72	0.65	0.68	16181
Class 1	0.58	0.09	0.16	8485
Class 2	0.89	0.97	0.93	89025

Confusion Matrix:

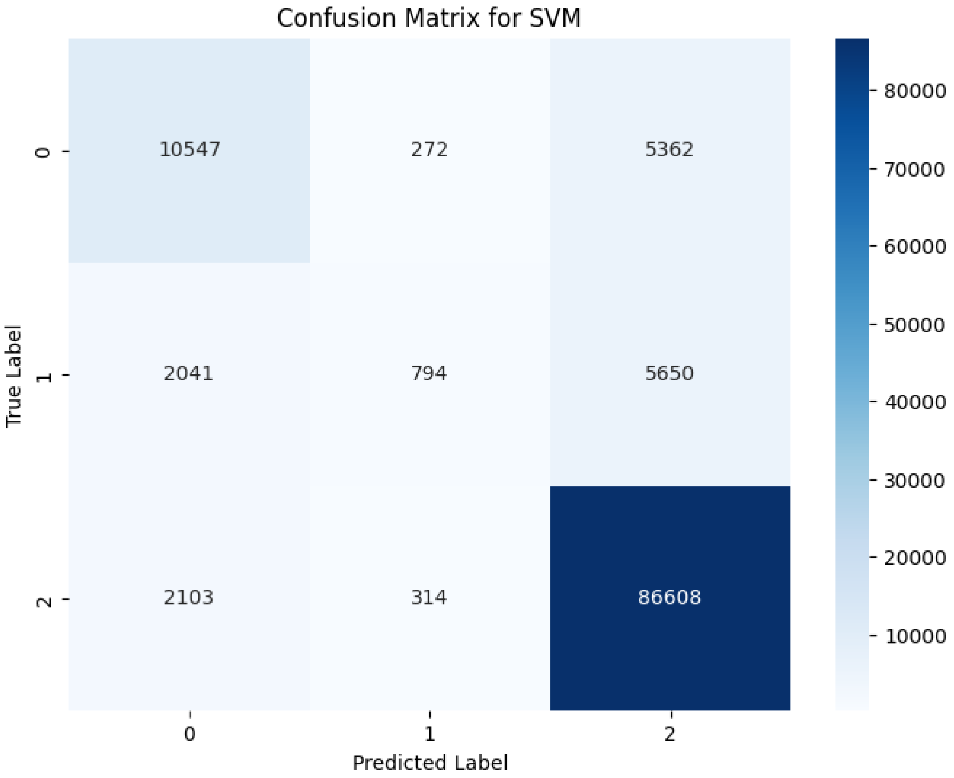


Figure 4.2: Confusion Matrix for SVM

4.3 Gradient Boosting Classifier

4.3.1 Accuracy: 81.91%

Classification Report:

Table 4.3: Classification Report for Gradient Boosting

	Precision	Recall	F1-Score	Support
Class 0	0.81	0.27	0.40	161811
Class 1	0.58	0.04	0.08	8485
Class 2	0.82	0.99	0.90	89025

Confusion Matrix:

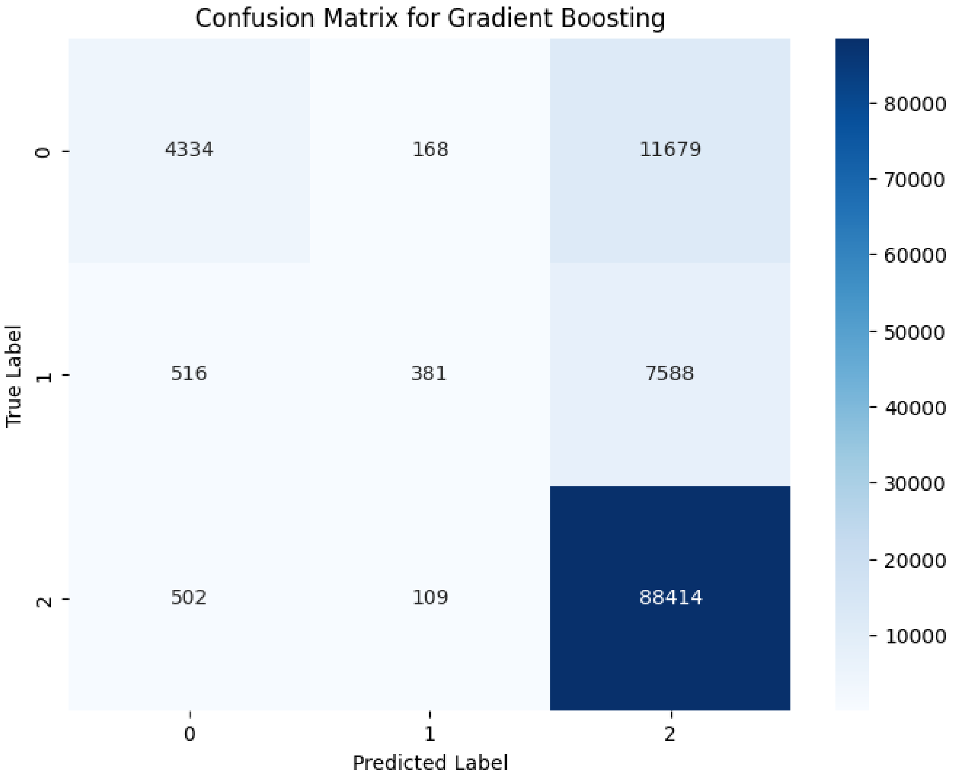


Figure 4.3: Confusion Matrix for Gradient Boosting

4.4 Random Forest Classifier

4.4.1 Accuracy: 89.50%

Classification Report:

Table 4.4: Classification Report for Random Forest

	Precision	Recall	F1-Score	Support
Class 0	0.92	0.61	0.73	16181
Class 1	0.98	0.39	0.55	8485
Class 2	0.89	0.99	0.94	89025

Confusion Matrix:

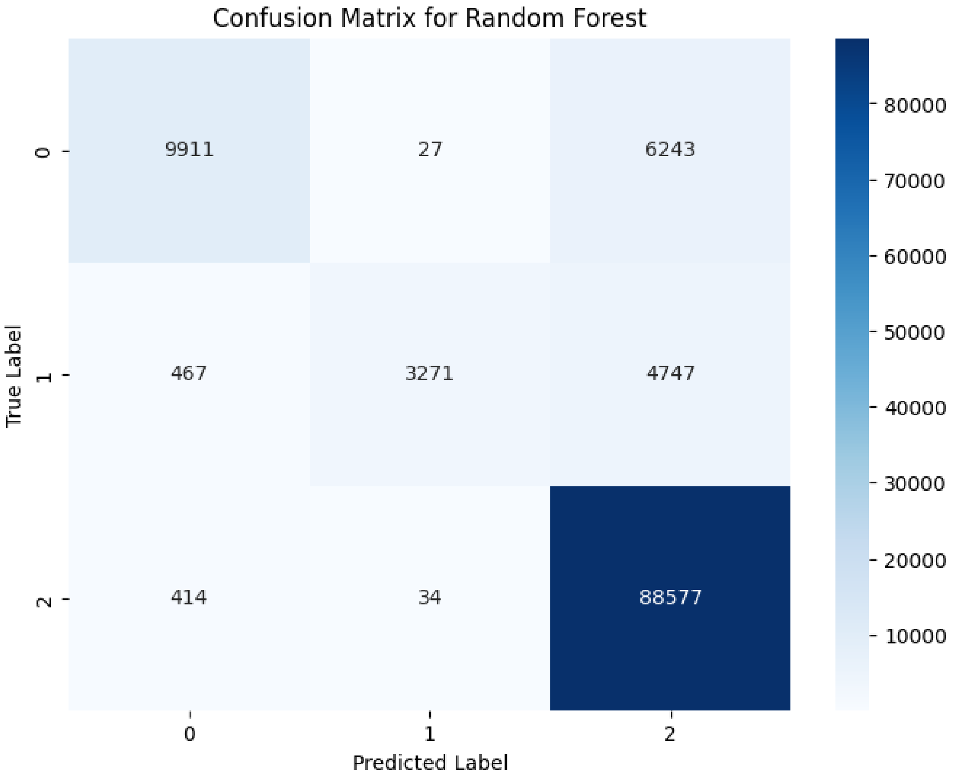


Figure 4.4: Confusion Matrix for Random Forest

4.5 Model Comparison

4.5.1 Model Metrics:

Table 4.5: Model Metrics Comparison

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	86.27%	84.07%	86.27%	84.35%
SVM	86.15%	83.98%	86.15%	83.60%
Gradient Boosting	81.91%	80.14%	81.91%	76.74%
Random Forest	89.50%	90.06%	89.50%	88.15%

4.5.2 Model Comparison Plot:

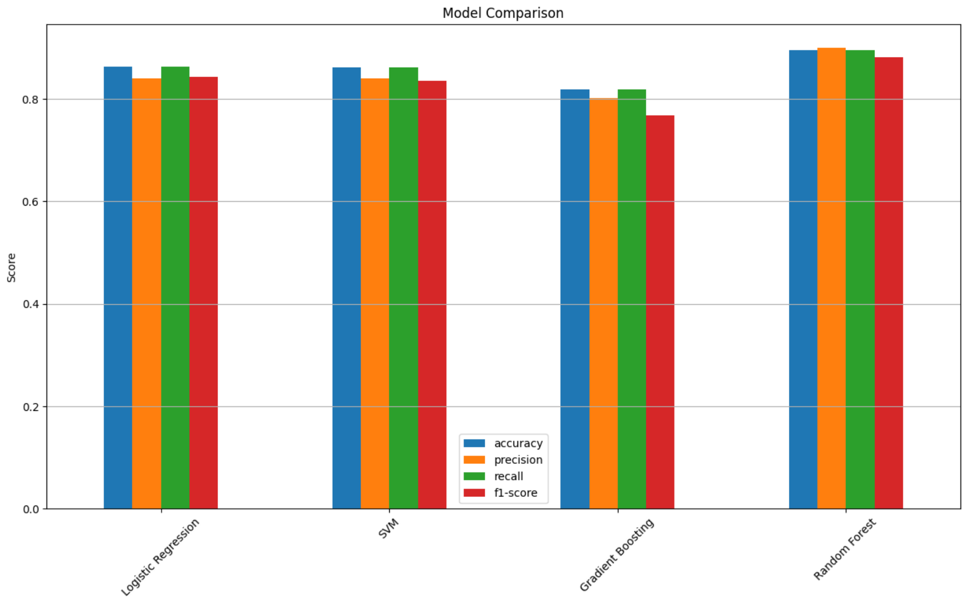


Figure 4.5: Model Comparison Plot

4.6 Feature Importance (Random Forest)

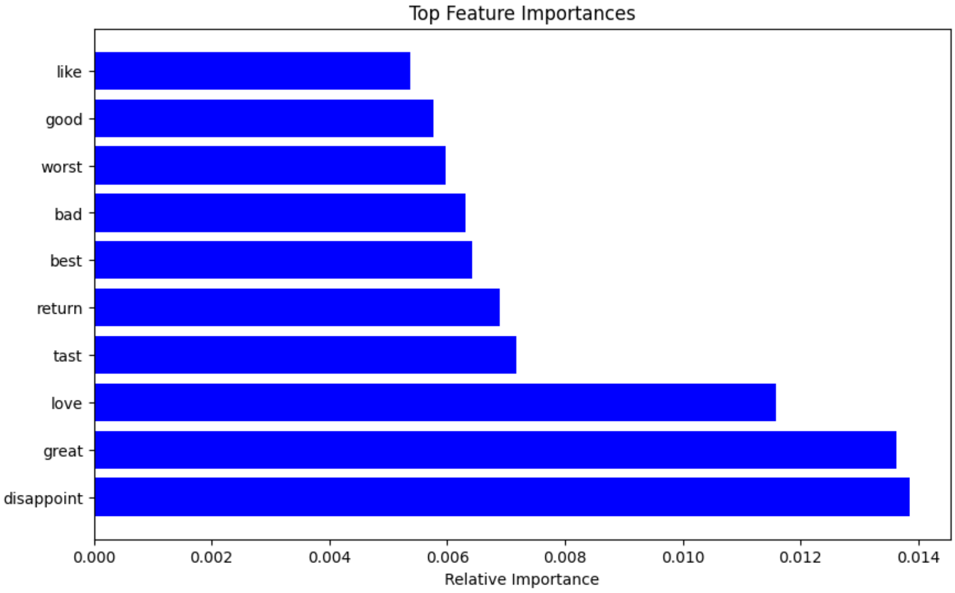


Figure 4.6: Feature Importance for Random Forest

4.7 Conclusion

The presented ensemble learning algorithms—Gradient Boosting, Support Vector Machines (SVM), and Random Forest—stand as formidable tools in the realm of machine learning. Each algorithm possesses unique characteristics that cater to diverse problem domains. Gradient Boosting excels in sequential error correction, SVM provides geometrically motivated hyperplane solutions, and Random Forest showcases robustness against overfitting.

For our specific application, the choice of Random Forest proves particularly compelling. Random Forest's resilience to overfitting, capacity to handle noisy data, and ability to provide insightful feature importance rankings align seamlessly with the challenges of our dataset. In scenarios where interpretability and robust performance are paramount, Random Forest emerges as a pragmatic choice over the other algorithms. As machine learning evolves, Random Forest continues to demonstrate its versatility and reliability, making it a valuable asset in our pursuit of accurate and interpretable predictive modeling.

REFERENCES

[1] S. T. K. Shivaprasad and J. Shetty, ”Sentiment Analysis of Product Reviews: A Review,” NMAM Institute of Technology Nitte, 2017. [Online]. Available: https://doi.org/10.1109/ICICCT.2017.7975207

[2] M. Mashaabi, A. Alotaibi, H. Qudaih, R. Alnashwan, and H. Al-Khalifa, ”Natural Language Processing in Customer Service: A Systematic Review,” King Saud University, Riyadh, 2022. [Online]. Available: https://doi.org/10.48550/arXiv.2212.09523

[3] G. Kaur and A. Sharma, ”Comparison of Different Machine Learning Algorithms for Sentiment Analysis,” Symbiosis International University, Pune, 2022. [Online]. Available: https://doi.org/10.1109/ICSCDS53736.2022.9760846

[4] D. Khurana, A. Koli, K. Khatter, and S. Singh, ”Natural language processing: state of the art, current trends, and challenges,” 2022. [Online]. Available: https://doi.org/10.1007/s11042-022-13428-4

[5] G. Papadia, M. Pacella, M. Perrone, and V. Giliberti, ”A Comparison of Different Topic Modeling Methods through a Real Case Study of Italian Customer Care,” Algorithms, vol. 16, no. 94, 2023. [Online]. Available: https://doi.org/10.3390/a16020094

[6] P. Dhanalakshmi, G. A. Kumar, B. S. Satwik, K. Sreeranga, A. T. Sai and G. Jashwanth, ”Sentiment Analysis Using VADER and Logistic Regression Techniques,” 2023 International Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS), Coimbatore, India, 2023, pp. 139-144, doi: 10.1109/ICISCoIS56541.2023.10100565.

[7] Wankhade, M., Rao, A.C.S., Kulkarni, C. A survey on sentiment analysis methods, applications, and challenges. Artif Intell Rev 55, 5731–5780 (2022). https://doi.org/10.1007/s10462-022-10144-1