# A
# Major Project Report

**On**

## “Enhancing Trust With AI: Product Review Analysis And Segregation System”

Submitted in partial fulfillment of the

Requirements for the award of the degree of

**Bachelor of Technology**

**In**

**Computer Science & Engineering -**

**Artificial Intelligence & Machine Learning**

**By**


| | |
|---|---|
| **Y. Rishyendra Kumar** | **20R21A6659** |
| **G. Rohith** | **21R25A6602** |
| **T. Tarakesh** | **21R25A6605** |
| **T. Anurag** | **21R25A6606** |

Under the guidance of

**Dr. Sivakrishna Kondaveeti**

**Associate Professor**

**Department of Computer Science & Engineering – Artificial Intelligence and Machine Learning**

MARRI LAXMAN REDDY GROUP OF INSTITUTIONS

MLR INSTITUTE OF TECHNOLOGY

(UGC AUTONOMOUS)

Affiliated to JNTUH,Approved by AICTE

Laxman Reddy Avenue, Dundigal, Hyderabad-500 043, Telangana, India


**2024**

MARRI LAXMAN REDDY GROUP OF INSTITUTIONS
MLR INSTITUTE OF TECHNOLOGY
(UGC AUTONOMOUS)
Affiliated to JNTUH,Approved by AICTE
Laxman Reddy Avenue, Dundigal, Hyderabad-500 043, Telangana, India

# Department of Computer Science & Engineering-

# Artificial Intelligence & Machine Learning

## CERTIFICATE

This is to certify that the project entitled **"Enhancing Trust with AI: Product Review Analysis and Segregation System"** has been submitted by **Y.Rishyendra kumar (20R21A6659), G.Rohith (21R25A6602), T.Tarakesh (21R25A6605), T.Anurag (21R25A6606)** in partial fulfillment of the requirements for the award of degree of Bachelor of Technology in Computer Science and Engineering from Jawaharlal Nehru Technological University, Hyderabad. The results embodied in this project have not been submitted to any other University or Institution for the award of any degree or diploma.

**Internal Guide** **Head of the Department**

**Project coordinator** **External Examiner**

MARRI LAXMAN REDDY GROUP OF INSTITUTIONS
MLR INSTITUTE OF TECHNOLOGY
(UGC AUTONOMOUS)
Affiliated to JNTUH, Approved by AICTE
Laxman Reddy Avenue, Dundigal, Hyderabad-500 043, Telangana, India

# Department of Computer Science & Engineering-

# Artificial Intelligence & Machine Learning

## DECLARATION

We hereby declare that the project entitled **"Enhancing Trust with AI: Product Review Analysis and Segregation System"** is the work done during the period from **January 2024 to May 2024** and is submitted in partial fulfillment of the requirements for the award of degree of Bachelor of Technology in Computer Science and Engineering from Jawaharlal Nehru Technology University, Hyderabad. The results embodied in this project have not been submitted to any other university or Institution for the award of any degree or diploma.

| | |
|---|---|
| **Y. Rishyendra kumar** | **20R21A6659** |
| **G. Rohith** | **21R25A6602** |
| **T. Tarakesh** | **21R25A6605** |
| **T. Anurag** | **21R25A6606** |

MARRI LAXMAN REDDY GROUP OF INSTITUTIONS

MLR INSTITUTE OF TECHNOLOGY

(UGC AUTONOMOUS)

Affiliated to JNTUH,Approved by AICTE

Laxman Reddy Avenue, Dundigal, Hyderabad-500 043, Telangana, India

# Department of Computer Science & Engineering-

# Artificial Intelligence & Machine Learning


## ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without the mention of people who made it possible, whose constant guidance and encouragement crowned our efforts with success. It is a pleasant aspect that we now have the opportunity to express our guidance for all of them.

First of all, we would like to express our deep gratitude towards our internal guide - **Dr. SIVA KRISHNA KONDAVEETI, Associate Professor, Department of CSE -AIML** for his support in the completion of our dissertation. We wish to express our sincere thanks to **Dr. K. SAI PRASAD,** HOD, Dept. of CSE-AIML and principal **Dr. K. SRINIVAS RAO** for providing the facilities to complete the dissertation.

We would like to thank all our faculty and friends for their help and constructive criticism during the project period. Finally, we are very much indebted to our parents for their moral support and encouragement to achieve goals.

| | |
|---|---|
| **Y. Rishyendra kumar** | **20R21A6659** |
| **G. Rohith** | **21R25A6602** |
| **T. Tarakesh** | **21R25A6605** |
| **T. Anurag** | **21R25A6606** |

MARRI LAXMAN REDDY GROUP OF INSTITUTIONS MLR INSTITUTE OF TECHNOLOGY
(UGC AUTONOMOUS)
Affiliated to JNTUH,Approved by AICTE
Laxman Reddy Avenue, Dundigal, Hyderabad-500 043, Telangana, India

# Department of Computer Science & Engineering-

# Artificial Intelligence & Machine Learning

## ABSTRACT


The "Enhancing Trust with AI: Product Review Analysis and Segregation System" amalgamates cutting-edge technology with user-centric design to equip consumers with tools for navigating the dynamic landscape of online reviews. Through the establishment of a new standard for review evaluation that is strong, reliable, and responsive to the changing needs of modern customers, this innovation has the potential to completely transform the digital market. Our project introduces a pioneering approach at the crossroads of artificial intelligence, natural language processing (NLP), machine learning, and data visualization. In an era where online reviews have a huge impact on customer decisions, this project transforms the review industry by addressing the pressing need for thorough and trustworthy review assessment.

It is now of the utmost importance to guarantee the validity and relevancy of reviews due to the exponential rise of online review sites. The system provides accurate sentiment analysis, classifying reviews into Positive, Negative, and Neutral emotions by using cutting-edge NLP approaches and machine learning algorithms. This study equips users with knowledge, promoting wise choice-making and boosting confidence in review ratings.

## APPENDIX-1

# LIST OF FIGURES

## LIST OF FIGURES

## APPENDIX-2

# LIST OF TABLES

# LIST OF TABLES

# APPENDIX-3

# LIST OF ABBREVIATIONS

# ABBREVIATIONS

| | |
|---|---|
| **SVM** | **Support Vector Machine** |
| **LSTM** | **Long Short-Term Memory** |
| **NLP** | **Natural Language Processing** |
| **TF-IDF** | **Term Frequency Inverse Document Frequency** |
| **NLTK** | **Natural Language Toolkit** |
| **KNN** | **K-Nearest Neighbours** |

# APPENDIX-4

# REFERENCES

## References

[1] Nidhi, A. Patel, “A Survey on Fake Reviews Detection using Machine Learning Techniques”, 2018 4th International Conference on Computing Communication and Automation (ICCCA), doi:10.1109/CCAA.2018.8777594

[2] Rakibul Hassan, Md. Rabiul Islam, “A Supervised Machine Learning Approach to Detect Fake Online Reviews”, 2020 23rd International Conference on Computer and Information Technology (ICCIT), 19-21 December, 2020, doi:10.1109/ICCIT51783.2020.9392727

[3] Andre Sihombing, A.C.M. Fong, “Fake Review Detection on Yelp Dataset Using Classification Techniques in Machine Learning”, 2019 International Conference on contemporary Computing and Informatics (IC3I), doi:10.1109/IC3I46837.2019.9055644

[4] Yin Shuqin, Feng Jing, “Fake Reviews Detection Based on Text Feature and Behaviour Feature”, 2019 IEEE 21st International Conference on High Performance Computing and Communications, doi:10.1109/HPCC/SmartCity/DSS.2019.00277

[5] Aljoharah Almjawel, Sahar Bayoumi, Dalal Alshershi, Soroor Alzahrani, Munirah Alotaibi, “Sentiment Analysis and Visualization of Amazon Book’s Reviews”, 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS), DOI: 10.1109/CAIS.2019.8769589

[6] S. Uma Maheswari, Dr. S.S. Dhenakaran, “Detection of fake and Genuine Reviews with Hybridization of Fuzzy and Neural Networks Techniques”, June 2021

[7] Tanjim Ul Haque, “Sentiment Analysis on Large Scale Amazon Product Reviews”, 2018

[8] Mr. Karthikeyan T, Mr. Karthik Sekaran, Mr. Ranjith D, Mr. Vinoth Kumar V, Mr.Balajee, “personalized content extraction and text classification using effective web scrapping techniques”, 2019

[9] Mr. Navjyotsinh, chirag visani, “A study on different machine learning techniques for spam reviews detection”, 2017

[10] Pansy Nandwani, Rupali Verma, “A review on sentiment analysis and emotion detection from the text”, 2021.

[11] Prof. P.S. Gaikwad, Kaushal Parmar, Rohit Yadav, Datta Supekar, “Implementation of web scraping for e-commerce website”, 2021.

[12] Raheesa Safrin, K.R.Sharmila, T.S.Shri Subangi, E.A.Vimal, “Sentiment analysis on online product review”, 2017.

[13] Tri Astuti, Irnawati Pratika, “Product Review Sentiment Analysis by Artificial Neural Network Algorithm”, 2019.

[14] Beresneva Daria, “Computer-generated Text Detection Using Machine Learning: A Systematic Review”, 2011.

[15] Wenyuan Zhao, “Classification of Customer Reviews on Ecommerce Platforms Based on Naive Bayesian Algorithm and Support Vector Machine”, 2020.

[16] Elshirf Elmurngi, “Fake Reviews Detection using Supervised Machine Learning (IJACSA) International Journal of Advanced Computer Science and Applications”, Vol. 12, No. 1, 2021.

[17] Eka Dyar Wahyuni, Arif Djunaidy, “fake review detection from a product review using modified method of iterative computation framework”.

[18] Mayuri Patil, Snehal Nikumbh, “Fake Product Review Monitoring and Removal for Genuine Product Reviews”, (International Journal of Scientific Research and Engineering Development-– Volume 4 Issue 3, May -June 2021.

[19] N Deshai, B Bhaskara Rao, “Deep Learning Hybrid Approaches to Detect Fake Reviews and Ratings”, (Journal of Scientific & Industrial Research Vol. 82, January 2023).

[20] Ahmed M.Elmogy, Usman Tariq, “Fake Reviews Detection using Supervised Machine Learning”, ((IJACSA) International Journal of Advanced Computer Science and Applications Vol. 12, No. 1, 2021)

# INDEX

# CHAPTER 1

# INTRODUCTION

## 1.1 OVERVIEW

The goal of the "Enhancing Trust with AI: Product Review Analysis and Segregation System" project is to transform e-commerce platforms' customer experiences. This system carefully examines and classifies product reviews using cutting-edge AI algorithms, increasing consumer trust and transparency. By sorting reviews according to sentiment, veracity, and applicability, it enables customers to make wise choices. The project aims to increase user satisfaction and confidence in online shopping environments by using this novel approach.


Figure 1: Detecting fake reviews through analysis

## 1.2 PURPOSE OF THE PROJECT

The critical issue of trust in e-commerce platforms is the focus of the "Enhancing Trust with AI: Product Review Analysis and Segregation System" project. The project aims to improve the transparency and dependability of product reviews by employing cutting-edge AI technologies. In particular, the system aims to offer precise review analysis and classification, helping customers make knowledgeable purchasing decisions. The project's ultimate goal is to build trust between buyers and sellers by using sentiment analysis, authenticity checks, and efficient review segmentation. This will enhance the overall e-commerce experience and allay worries about false or misleading product information.

## 1.3 MOTIVATION

As online shopping grows in popularity, customers are depending more and more on product reviews to guide their decisions. On the other hand, a deluge of reviews combined with problems with bias and authenticity can erode confidence and make decision-making more difficult. By using AI to provide thorough analysis and review categorization, this project aims to allay these worries. The project intends to give customers the confidence to make informed decisions by improving the transparency and dependability of product feedback, eventually building trust and loyalty within the e-commerce community.

# CHAPTER 2

# LITERATURE SURVEY

An extensive literature survey has been conducted by studying existing systems of Certificate verification and generation. A good number of research papers, journals, and publications have also been referred before formulating this survey.

## 2.1EXISTING SYSTEM

Due to the current system's heavy reliance on manual review inspection, errors and inefficiencies occur. Because of its weak ai integration, it is difficult to distinguish between real and fake reviews. The time-consuming nature of this manual method also compromises the system's capacity to deliver trustworthy insights. Moreover, data interpretation is still difficult in the absence of automated visualization tools, which impedes the ability to make wise decisions. All things considered, the existing system is not up to the task of providing reliable and timely analyses of product reviews.

| **1** | | |
|---|---|---|
| **Reference in APA format** | A Survey on Fake Review Detection using Machine Learning Techniques | |
| **URL of the Reference** | **Authors Names and Emails** | **Keywords in this Reference** |
| https://ieeexplore.ieee.org/document/8777594 | Nidhi A. Patel | Fake Review, Sentiment Analysis, Opinion Spam, Fake review detection technique, Machine learning. |
| **The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc)** | **The Goal (Objective) of this Solution & What is the problem that need to be solved** | **What are the components of it?** |
| Machine Learning Techniques<br>• Supervised learning techniques<br>• Semi-supervised learning techniques<br>• Unsupervised learning techniques | The objective of the document is to discuss various techniques and approaches used in detecting fake reviews. It focuses on machine learning-based methods and the different features | The document outlines the steps involved in the machine learning approach for fake review detection, including data collection, data pre-processing, feature extraction and selection, and classifier model construction and testing. |

| | | |
|---|---|---|
| | and classifiers used for fake review detection. | |

**The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process**

Machine learning approach for fake review detection works as follows

| | Process Steps | Advantage | Disadvantage (Limitation) |
|---|---|---|---|
| **1** | Supervised learning techniques | Supervised learning benefits from labelled data (fake or genuine reviews) to train algorithms accurately. Using linguistic features and sentiment scores, it predicts review authenticity effectively. | It needs extensive labelled data, demanding manual labelling of reviews as fake or genuine, a time-consuming and costly process. It struggles with unlabeled data, lacking flexibility for ambiguous reviews. |
| **2** | Semi-supervised learning techniques | Semi-supervised learning classifies fake and genuine reviews without a complete labelled dataset. It uses a small labelled set and a larger unlabeled set, enabling fake review detection with limited labelled data. | Semi-supervised learning assumes labelled positives represent all fake reviews. Inaccurate representation compromises classifier performance. Also, it often needs multiple iterations, making it computationally costly. |
| **3** | Unsupervised learning techniques | Unsupervised learning techniques have the advantage of being able to classify fake and genuine reviews without the need for a labelled dataset. This means that these techniques can be applied to large amounts of unlabeled data, making them more scalable and efficient in detecting fake reviews. | Unsupervised learning's accuracy may be lower than supervised methods due to reliance on broad patterns, struggling with subtle differences. It demands significant computational resources and time for processing large unlabeled datasets. |

**Major Impact Factors in this Work**

| Dependent Variable | Independent Variable | Moderating variable | Mediating (Intervening) variable |
|---|---|---|---|
| The document's dependent variable is the performance of the fake review detection method, determined or forecasted using other variables. | In this document, the independent variables encompass a range of features and techniques utilized for fake review detection, such as linguistic features, behavioural traits, relational aspects, machine learning algorithms, and data mining techniques. | The document does not explicitly mention any moderating variables. | The document does not explicitly mention any mediating variables. However, the type of classifier used for fake review detection, such as naive bayes, support vector machine, decision tree acts as mediating variables |

| Relationship Among the Above 4 Variables in This article |
|---|
| The selected machine learning methodology has a direct impact on the efficiency of fake review detection systems. The type of features used and the classifier selected further affect this influence. As a result, the efficiency of the machine learning technique depends on how well the selected classifier and the provided features. |

<table>
<tr><th>Input and Output</th><th>Feature of This Solution</th><th>Contribution & The Value of This Work</th></tr>
<tr><td><table><tr><th>Input</th><th>Output</th></tr><tr><td>Reviews of the products primarily taken from e-commerce sites.</td><td>The performance of the fake review detection methods.</td></tr></table></td><td>It covers linguistic and textual features, behavioral features, and relational features.</td><td>From this paper we have gained knowledge regarding, classifiers and methods that were used by different machine learning techniques and the challenges that are associated with it.</td></tr>
</table>

| Positive Impact of this Solution in This Project Domain | Negative Impact of this Solution in This Project Domain |
|---|---|
| Review detection enhances the trust by removing the fake reviews, trustworthiness can be significantly improved and it also helps in better decision making, fair business practices. | Review detection can have some negative impact on this domain, which includes false positives which can harm the reputation and trust of the customers. |

| Analyze This Work by Critical Thinking | The Tools That Assessed this Work | What is the Structure of this Paper |
|---|---|---|
| The paper is a survey on fake review detection using machine learning techniques. It categorizes the techniques into supervised, semi-supervised, and unsupervised learning. It also discusses various features and classifiers that can be used to distinguish fake reviews from genuine ones. | None | Abstract<br>I. Introduction<br>II. Related Work<br>III. Machine learning based fake review detection techniques<br>IV. Analysis<br>V. Major challenges<br>VI. Conclusion |

**Diagram/Flowchart**

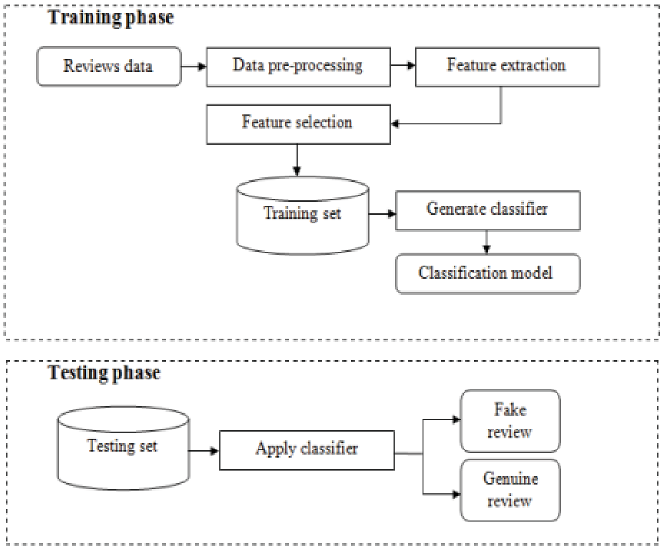



Fig. 1. Machine Learning based Fake Review Detection

| 2 | | |
|---|---|---|
| **Reference in APA format** | A Supervised Machine Learning Approach to Detect Fake Online Reviews | |
| **URL of the Reference** | **Authors Names and Emails** | **Keywords in this Reference** |
| https://ieeexplore.ieee.org/document/9392727 | Rakibul Hassan, Md. Rabiul Islam | supervised learning, support vector machine, naive Bayes, logistic regression, Empath, TF-DF, sentiment polarity. |
| **The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc)** | **The Goal (Objective) of this Solution & What is the problem that need to be solved** | **What are the components of it?** |
| Supervised Machine Learning techniques | The main objective of this document is to introduce a method using supervised machine learning to identify fake online reviews. It explores features like TF-IDF, Empath, and sentiment polarity to create a model that can accurately distinguish between fake and honest reviews. | Content based features.<br>Train-validation set split with a ratio of 75:25 to obtain the train set and validation set.<br>Identification of genre, detecting psycholinguistic behaviour, and categorization of text as features. |
| **The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process** | | |

| | Process Steps | Advantage | Disadvantage (Limitation) |
|---|---|---|---|
| **1** | Feature Selection | Advanced features like TF-IDF, Empath, and sentiment polarity provide a more nuanced understanding of reviews.TF-IDF considers the contextual relevance of words. Words that are frequent in a specific review but rare across all reviews are given higher weight, ensuring the analysis focuses on unique and significant terms. These features enable | The use of advanced features can introduce complexity into the analysis. Understanding and interpreting the output from these features might require a higher level of expertise, calculating TF-IDF values and processing Empath's extensive categories can be computationally intensive. In cases of limited data |

| | | | |
|---|---|---|---|
| | | a deep dive into the emotional aspects of reviews. | availability, these features might not perform optimally. |
| **2** | Supervised Classification | Classifiers such as logistic regression, Naive Bayes, and support vector machine (SVM) were used, when trained on a substantial amount of labelled data, often result in high accuracy. They can handle both numerical and categorical features, making them suitable for diverse applications. Support Vector Machine (SVM) can effectively handle non-linear data by using kernel functions. | Supervised learning relies heavily on labelled data for training. Acquiring and labelling a large dataset can be time-consuming and expensive. If not properly regularized, complex models like SVM can suffer from overfitting. Irrelevant or redundant features can degrade the model's accuracy. Feature selection and engineering are essential but challenging tasks. |
| **3** | Sentiment polarity | Integrating sentiment polarity adds an emotional context to the analysis, allowing the model to grasp the reviewers' emotions. Understanding the emotional tone of reviews provides valuable insights into the user experience. Businesses can utilize sentiment analysis to make informed decisions. | It involves subjectivity, as interpreting the emotional tone of a text can vary among individuals. Sentiment analysis might struggle with contextual ambiguity. Reviews often contain sarcasm or irony, which can be challenging for automated systems to detect. Sentiment analysis often focuses on positive and negative sentiments, neglecting neutral sentiments. |

| **Major Impact Factors in this Work** |
|---|
| |

| Dependent Variable | Independent Variable | Moderating variable | Mediating (Intervening) variable |
|---|---|---|---|
| The dependent variable is the accuracy, which is indicated using F1 score. | The independent variables mentioned in the document include TF-IDF (term frequency-inverse document frequency), Empath categories, and sentiment polarity. These variables are used as features to develop a model for classifying fake and honest reviews. | The document does not explicitly mention any moderating variables.<br>The possible moderating variable is the sentiment polarity of the reviews, which may influence the performance of different classifiers. | The document does not explicitly mention any mediating variables |

| Relationship Among the Above 4 Variables in This article |
|---|
| The dependent variable (review classification) is influenced by the independent variables (TF-IDF, Empath categories, and sentiment polarity), with sentiment polarity potentially moderating the performance of classifiers and TF-IDF feature potentially mediating the accuracy of fake online review detection. This complex relationship shows how various factors work together to distinguish between fake and genuine online reviews. |

| Input and Output | | Feature of This Solution | Contribution in this Work |
|---|---|---|---|
| **Input** | **Output** | | |
| Set of online hotel reviews with labels indicating whether they are fake or truthful. | A classifier that can predict the label of a new review based on its features. | The proposed solution uses content-based and use-behaviour features for classification | The document evaluates the performance of three classifiers: logistic regression, Naive Bayes, and support vector machine, and compares them with previous semi-supervised and supervised techniques. |

| Positive Impact of this Solution in This Project Domain | Negative Impact of this Solution in This Project Domain |
|---|---|
| The use of both content-based and user-behaviour based features can improve the accuracy of fake review detection. | The solution is a supervised learning approach, which relies heavily on the availability of labelled data for training. This could be a limitation in scenarios where labelled data is scarce or expensive to obtain. |

| Analyse This Work By Critical Thinking | The Tools That Assessed this Work | What is the Structure of this Paper |
|---|---|---|
| The high accuracy of the proposed approach suggests that it could be effective in real-world applications. It could help businesses identify fake reviews and make better decisions based on genuine customer feedback. | None | Abstract<br>I. Introduction<br>II. Related Work<br>III. Proposed Work<br>IV. Performance analysis<br>V. Conclusion |
| **Diagram/Flowchart** | | |

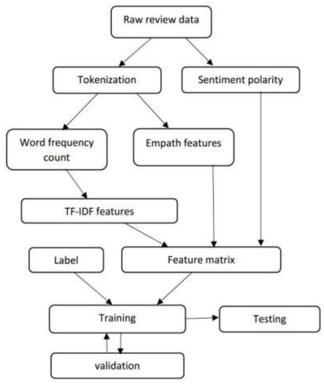


Fig. 1: Proposed classification model

| **3** | | |
|---|---|---|
| **Reference in APA format** | Fake Review Detection on Yelp Dataset Using Classification Techniques in Machine Learning | |
| **URL of the Reference** | **Authors Names and Emails** | **Keywords in this Reference** |
| https://ieeexplore.ieee.org/document/9055644 | Andre Sihombing, A.C.M. Fong | Machine learning, classification, fake reviews detection, online discussion forum |
| **The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc )** | **The Goal (Objective) of this Solution & What is the problem that need to be solved** | **What are the components of it?** |

| | | |
|---|---|---|
| Logistic Regression, Support Vector Machine, Gaussian Naive Bayes, and XGBoost. | The main objective of this document is study of different machine learning classification techniques/models such as Logistic Regression, Support Vector Machine, Gaussian Naive Bayes, and XGBoost for fake review detection. This document provides the insights regarding the effectiveness of classification methods and their potential application in detecting fake reviews. | The workflow in the document has been classified into three parts: data preprocessing, feature engineering and the classification process. In which the components like under-sampling and over-sampling were used for better preprocessing of data. |

**The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process**

| | Process Steps | Advantage | Disadvantage (Limitation) |
|---|---|---|---|
| **1** | Logistic Regression | Logistic Regression is a discriminative classifier, it is mainly used to model the relationship between a dependent variable and one or more independent variables and commonly used for binary classification problems. | It is not a best option in situations where non-linear relationships need to be captured effectively, or for extremely complex datasets with complex patterns. |
| **2** | Gaussian Naive Bayes | Naive Bayes is a simple and computationally efficient algorithm, where it can handle both categorical and continuous input variables. | Naïve Bayes is not suitable for problems with nonlinear boundaries and it is sensitive to the presence of irrelevant features. |
| **3** | Support Vector Machine | SVM can handle both linear and non-linear decision boundaries, through the different kernel functions and it is less prone to overfitting compared to the other algorithms. | SVM can be computationally expensive, especially for large datasets. It may be sensitive to the choice of kernel function. |
| **4** | XGBoost | XGBoost is an ensemble approach where, it improves performance by combining the predictions of several weak learners. It can handle | Similar to the SVM, XGBoost can also be computationally expensive and may require careful tuning of hyperparameters |

| Major Impact Factors in this Work | | | |
|---|---|---|---|
| **Dependent Variable** | **Independent Variable** | **Moderating variable** | **Mediating (Intervening ) variable** |
| The dependent variables in the document are Precision, Recall, F1-Score which are used as evaluation metrics for the different classifiers. | The independent variables mentioned in the document include. Logistic Regression, Support Vector Machine, Gaussian Naive Bayes, and XGBoost | The Ratio of filtered and non-filtered reviews and the feature engineering acts as the moderating variables. | The document does not explicitly mention any mediating variables. However, Length of Reviews, Maximum Review Numbers per Day acts as the mediating variables. |

| Relationship Among The Above 4 Variables in This article |
|---|
| The ratio of filtered to unfiltered reviews is used as a moderating variable, which affects the classification strategies used to address imbalance. The accuracy of classification methods is improved by mediating variables like rating deviation, review length, and maximum reviews per day. |

| Input and Output | Feature of This Solution | Contribution & The Value of This Work |
|---|---|---|
| **Input**: Yelp's dataset<br>**Output**: The performance of the fake review detection methods. | The document has also considered the length of the reviews and found that fake reviews tend to be shorter than genuine ones. | The document aims to compare the performance between four well-known machine learning classification techniques: Logistic Regression, Gaussian Naive Bayes, Support Vector Machine, and XGBoost and determine the most effective approach for the task at hand. |

| Positive Impact of this Solution in This Project Domain | Negative Impact of this Solution in This Project Domain |
|---|---|
| By utilizing the classification techniques mentioned in the document, the research achieved a high F-1 score of 0.9 in prediction, indicating the effectiveness of the approach in identifying fake reviews. | As the document depends upon the labelled data there could be a potential bias in the classification process. If the algorithm contains potential biases, it may filter reviews poorly based on certain characteristics or demographics. |

| | | |
|---|---|---|
| **Analyse This Work By Critical Thinking** | **The Tools That Assessed this Work** | **What is the Structure of this Paper** |
| This paper investigates the use of four well-known machine learning classification techniques to detect fraudulent reviews in online discussion boards. Out of which, XGBoost outperformed the other techniques with the best F-1 score. | None | Abstract<br>I. Introduction<br>II. Literature Review<br>III. Proposed Methodology<br>IV. Evaluation<br>V. Conclusion |
| **Diagram/Flowchart** | | |

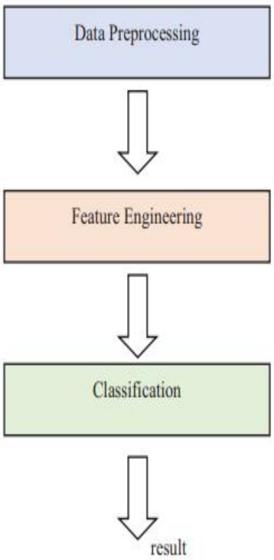



Figure 1. Overall workflow.

| **4** | | |
|---|---|---|
| **Reference in APA format** | Fake Reviews Detection Based on Text Feature and Behaviour Feature | |
| **URL of the Reference** | **Authors Names and Emails** | **Keywords in this Reference** |
| https://ieeexplore.ieee.org/document/8855455 | Yin Shuqin, Feng Jing | Fake reviews, fusion feature, PU-Learning, constrained k-means, classification model |

| The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc ) | The Goal (Objective) of this Solution & What is the problem that need to be solved | What are the components of it? |
|---|---|---|
| This paper suggests the use of multiple features in the MPINPUL (Mixing Population and Individual Nature PU Learning) a model for classifying fake reviews. | The goal of the proposed solution is to develop a PU learning model for the recognition of fake reviews. It utilizes various features such as text, behaviour, and relationship characteristics to accurately identify fake reviews. | The document mainly focused on three major categories for classifying the reviews: text features, behavioural characteristics of reviewers, and relationship characteristics. |

**The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process**

| | Process Steps | Advantage | Disadvantage (Limitation) |
|---|---|---|---|
| 1 | Text features | It offers important details regarding the reviews content, including thematic, lexical, and part-of-speech features. These attributes can be useful in spotting trends and traits that set authentic reviews apart from fraudulent ones. | Since fake reviews frequently mimic genuine comments in vocabulary and language, it can be difficult to distinguish them from real ones just by looking at textual characteristics. |
| 2 | Behavioural Characteristics of Reviewers | Patterns suggestive of fraudulent reviews can be identified by examining reviewers' behaviour, including word choice and the frequency of specific linguistic elements. User behavior characteristics such as maximum text content similarity, common text length, maximum number of comments per day are considered in analysing the reviews. | The fact that these features might not be very robust or discriminative could be a drawback. For instance, if real reviewers regularly comment on comparable goods or services, they may also have high MCS or MDN. |

| | | | |
|---|---|---|---|
| **3** | Relational Characteristics | The complex, multidimensional, and heterogeneous relationships between reviewers, reviews, products, and merchants can be captured by relationship characteristics. | Because online review systems are dynamic, relationship attributes could be noisy, lacking, or inconsistent. To extract and analyse them, they might also need more advanced algorithms and computer power. |

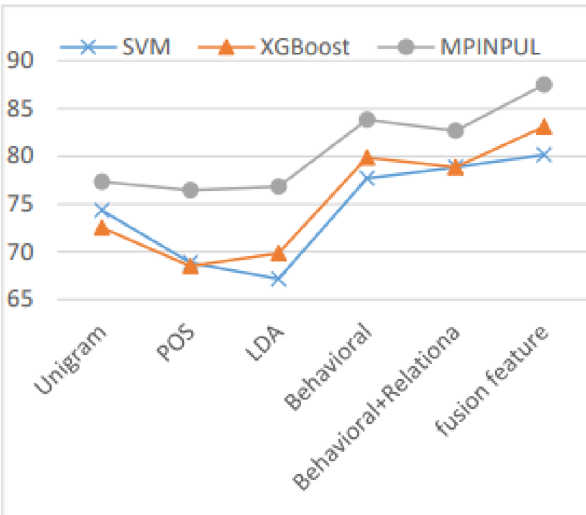
**Major Impact Factors in this Work**

| Dependent Variable | Independent Variable | Moderating variable | Mediating (Intervening ) variable |
|---|---|---|---|
| The evaluation index which includes accuracy, precision, recall and F1 score acts as dependent variables, which are used to evaluate the performance of classification model. | The major features such as text, behaviour, and relationship characteristics, that are used to build MPINPUL classification model acts as independent variables. | The document does not explicitly mention any moderating variables. | The document does not explicitly mention any mediating variables. |

| Relationship Among The Above 4 Variables in This article |
|---|
| The evaluation of the model depends upon the characteristics that were used to build the MPINPUL classification model, which determines the relation between the dependent and the independent variables |

| Input and Output | Feature of This Solution | Contribution & The Value of This Work |
|---|---|---|
| **Input:** Yelp dataset.<br>**Output:** Accuracy of MPINPUL model. | The proposed solution focuses on integrating text, behavioural, and relationship features to build a classification model for fake reviews recognition | The experimental results demonstrate the effectiveness of the MPINPUL model in identifying fake reviews, as it outperforms other single features under fusion feature conditions. |

| Positive Impact of this Solution in This Project Domain | Negative Impact of this Solution in This Project Domain |
|---|---|
| The document has showed that the classification model trained on fusion | The negative impact of proposed solution include misclassification, manipulation and false insights. |

<table>
<tr><td colspan="2">features, which integrate text and behaviour characteristics, is about 10% more accurate than models trained solely on text features.</td><td></td></tr>
<tr><th>Analyse This Work By Critical Thinking</th><th>The Tools That Assessed this Work</th><th>What is the Structure of this Paper</th></tr>
<tr><td>The paper demonstrates the significance of the critic's behaviour in identifying fake reviews and the feasibility and effectiveness of the MPINPUL model.</td><td>None</td><td>Abstract<br>I. Introduction<br>II. Feature construction<br>III. MPINPUL Classification Model<br>IV. Performance evaluation<br>V. Conclusion</td></tr>
<tr><th colspan="3">Diagram/Flowchart</th></tr>
<tr><td colspan="3">
</td></tr>
</table>

| 5 | | |
|---|---|---|
| **Reference in APA format** | Sentiment Analysis and Visualization of Amazon Books' Reviews | |
| **URL of the Reference** | **Authors Names and Emails** | **Keywords in this Reference** |
| https://ieeexplore.ieee.org/document/8769589 | Aljoharah Almjawel, Sahar Bayoumi, Dalal Alshehri, Soroor Alzahrani, Munirah Alotaibi | Text Visualization, Tableau, Rstudio, Amazon Reviews, Opinion Analysis, Sentiment Analysis |
| **The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc)** | **The Goal (Objective) of this Solution & What is the problem that need to be solved** | **What are the components of it?** |
| Interactive Packed bubbles, Linear chart, Stacked bars, and Word-cloud | This document discusses the use of visualization techniques in analysing and summarizing reviews. | Visualization techniques, Sentiment analysis, Tableau and R to provide interactive visualizations. |

**The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process**

| | **Process Steps** | **Advantage** | **Disadvantage (Limitation)** |
|---|---|---|---|
| **1** | Raw data preparation | Meaningful insights and patterns can be more easily extracted from the raw data by cleaning and organizing it. Through the process of eliminating errors, inconsistencies, and missing numbers, the data is made accurate and trustworthy for analysis. | The drawback of raw data preparation is that it can need a lot of time and resources. The procedure includes a number of processes, including data integration, data transformation, and data cleansing, all of which can be quite computationally intensive and complex. |
| **2** | Sentiment analysis process | The advantage of sentiment analysis process is it allows to gain insight into customer opinion by analysing the sentiment expressed in reviews, comments. | Sentiment analysis process is often trained on data from specific languages and cultures, which can introduce bias. |

| | | | |
|---|---|---|---|
| **3** | Visualization techniques | Through the use of visual aids, customers are better able to understand complex data through visualization techniques. It makes easy for consumers to evaluate and comprehend huge amounts of data fast. | Visualizations can be misinterpreted if they are not designed or presented properly. Certain data could be excessively complicated to effectively express visually. |

**Major Impact Factors in this Work**

| Dependent Variable | Independent Variable | Moderating variable | Mediating (Intervening ) variable |
|---|---|---|---|
| The polarity and the rating of each review. | The independent variables in this document are, the book title and different levels of customer satisfaction and feedback | The review's time, which could have an impact on its perspective because customer expectations, tastes, or trends can change over time. | The document does not explicitly mention any mediating variables. However, the summary may influence the sentiment of the review by highlighting the key features or aspects of the book. |

| Relationship Among The Above 4 Variables in This article |
|---|
| The document coneys the relationship between book title, customer satisfaction levels, and feedback on polarity and rating in reviews. The moderating variable, review time, may influence perspectives due to changing trends. |

| Input and Output | | Feature of This Solution | Contribution & The Value of This Work |
|---|---|---|---|
| **Input** | **Output** | The main features of the proposed solution are interactive solution and the sentiment analysis of customer reviews, helping users in making decisions | By summarizing viewpoints and emphasizing sentiment trends, the proposed solution helps users save time books and find the ones with the most positive reviews. |
| Book's review | Visual representation of the review | | |

| Positive Impact of this Solution in This Project Domain | Negative Impact of this Solution in This Project Domain |
|---|---|
| The suggested system is to give book reviews a visual format so that users can examine reviews from customers more efficiently. | As the suggested solution depends on a sentiment analysis system that might have flaws or limitations, it might not accurately convey the genuine opinions of the customers. |

| Analyse This Work By Critical Thinking | The Tools That Assessed this Work | What is the Structure of this Paper |
|---|---|---|
| The paper demonstrates the use Tableau and R together with visualisation techniques for data analysis. It also discusses how the reviews were analysed and presented using a variety of visualisation techniques, including word clouds, packed bubbles, linear charts, and stacked bars. | None | Abstract<br>I. Introduction<br>II. Feature construction<br>III. MPINPUL Classification Model<br>IV. Performance evaluation<br>V. Conclusion |

**Diagram/Flowchart**

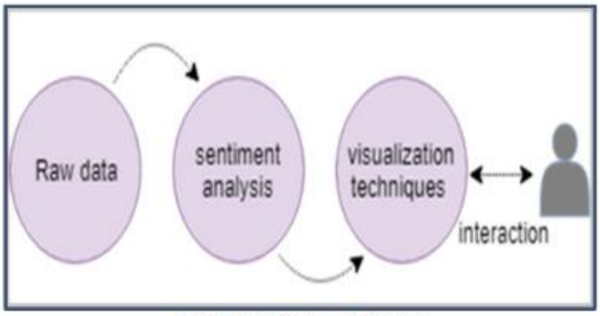



Figure.1. System Architecture

| 6 | | |
|---|---|---|
| **Reference in APA format** | Elshirf elumurni, Abdelouahed gherbi Detecting fake reviews through sentiment analysis using machine learning techniques. | |
| **URL of the Reference** | **Authors Names and Emails** | **Keywords in this Reference** |
| https://www.researchgate.net/publication/325973731_Detecting_Fake_Reviews_through_Sentiment_Analysis_Using_Machine_Learning_Techniques | Elshirf Elmurngi.<br>Abdelouahed gherbi. | Sentiment analysis; fake reviews; naïve bayes; support vector machines; k-nearest neighbour; k-star; Decision tree-ja8 |
| **The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc )** | **The Goal (Objective) of this Solution & What is the problem that need to be solved** | **What are the components of it?** |
| Sentiment Analysis. | The main objective is to classify movie reviews as real reviews or fake reviews using SA algorithms with supervised learning techniques. | This paper consists of Sentiment classification,<br>Feature selection, Detection process. |
| **The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process** | | |

| | **Process Steps** | **Advantage** | **Disadvantage (Limitation)** |
|---|---|---|---|
| **1** | Movie reviews collection<br>The original movie review dataset has been used in order to test methods of classification. | Raw movie reviews are gathered, and the model's predictions directly influence the assessment of the movies, simplifying the process of determining whether a movie aligns with its reviews and is considered good. | Only movie reviews are collected. |
| **2** | Data pre-processing<br>• StringToWordvecto.<br>• Attribute selection.<br>• Feature selection. | It helps in transforming the data before the actual sentiment analysis task. | In the data pre-processing phase, each block relies on its preceding block, and all blocks are interlinked simultaneously. |

| | | | |
|---|---|---|---|
| **3** | Feature selection<br><br>Feature selection is a method employed to pinpoint a subset of features that exhibit strong associations with the target model. | Feature selection is to increase the level of accuracy. | Results differ from one method to the another method. |
| **4** | Sentiment classification algorithms. | It is used in different domains like (commerce, medicine, media).<br><br>It examines data and identify patterns. | It's challenging to find out the exact technique for the model. |
| **5** | Detection process. | it empowers the user to make informed decisions. | The use of a confusion matrix adds complexity to the situation. |

**Major Impact Factors in this Work**

| Dependent Variable | Independent Variable | Moderating variable | Mediating (Intervening ) variable |
|---|---|---|---|
| sentiment classification of movie reviews.<br>Accuracy. | The paper consist of different algorithms such as SVM, NB, KNN-IBK, K-Star, and DT-J48 to determine which algorithm is more accurate in classifying the reviews. | The accuracy of the model relies on a series of interconnected steps in the paper's data pre-processing. Each of these steps builds upon the previous one, collectively contributing to an enhanced model performance. | In this paper, there are no mediating variables; instead, everything is interdependent, with each factor relying on the others. |

| Relationship Among The Above 4 Variables in This article |
|---|
| The relationship between the dependent and moderating variables directly enhances the model's accuracy, with each step being intricately linked. This absence of mediating variables is attributed to the absence of independent variables. |

<table>
<tr><th>Input and Output</th><th>Feature of This Solution</th><th>Contribution & The Value of This Work</th></tr>
<tr><td><table><tr><th>Input</th><th>Output</th></tr><tr><td>Movie review dataset V1.0<br><br>Movie review dataset V2.0</td><td>determines which algorithm is more accurate.</td></tr></table></td><td>The proposed methodology, using the Weka tool and different sentiment classification algorithms, is effective for classifying movie reviews as real or fake.</td><td>This work contributes to the development of techniques for analyzing and classifying sentiment in textual data.<br><br>The effective identification of fake reviews and the model's ability to accurately predict true positive and true negative values on a testing dataset.</td></tr>
</table>

<table>
<tr><th>Positive Impact of this Solution in This Project Domain</th><th>Negative Impact of this Solution in This Project Domain</th></tr>
<tr><td>Sentiment Analysis (SA) has emerged as a subject within text analysis, driven by its potential commercial advantages. Furthermore, the user-generated opinion reviews, categorized as either positive or negative, offer valuable insights for consumers in making product choices.<br><br>Improved Accuracy The experiments conducted in this project have shown that sentiment classification algorithms, particularly SVM.</td><td>In this solution, the accuracy of various supervised algorithms is determined, with each algorithm providing its unique predictions. These predictions vary from one model to another.<br><br>The solution presents various sentiment classification algorithms and methodologies, there are potential limitations and challenges that could impact the accuracy, efficiency, and reliability of sentiment analysis in the project domain of movie reviews</td></tr>
</table>

<table>
<tr><th>Analyse This Work By Critical Thinking</th><th>The Tools That Assessed this Work</th><th>What is the Structure of this Paper</th></tr>
<tr><td>Utilizing supervised algorithms for fake review prediction introduces complexity to the accuracy assessment.<br><br>The importance of sentiment analysis in detecting fake reviews and its potential commercial benefits. They suggest that future research can focus on improving the detection mechanism for fake reviews and evaluating the accuracy of this detection using statistical methods.</td><td>Weka tools.<br><br>String-To-Word Vector filter in Weka was used for transforming the dataset.</td><td>Abstract<br><br>I. Introduction.<br>II. Related Work.<br>III. Methodology.<br>IV. Experiments and results analysis.<br>V. Conclusion and Future work</td></tr>
</table>

| Diagram/Flowchart |
| --- |
|  <br>Fig 1. Steps and techniques used in sentiment analysis. |

| 7 | | |
| --- | --- | --- |
| **Reference in APA format** | Eka Dyar Wahyuni, Arif Djunaidy fake review detection from a product review using modified method of iterative computation framework. | |
| **URL of the Reference** | **Authors Names and Emails** | **Keywords in this Reference** |
| https://www.researchgate.net/publication/303499094_Fake_Review_Detection_From_a_Product_Review_Using_Modified_Method_of_Iterative_Computation_Framework | Eka Dyar Wahyuni .<br>Arif Djunaidy. | Fake reviews, opinion mining, sentiment analysis, test mining, icf. |
| **The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc )** | **The Goal (Objective) of this Solution & What is the problem that need to be solved** | **What are the components of it?** |
| ICF (iterative computation framework) | This research aims to detect fake reviews for a product by using the text and rating property from a review. | The components of the paper include the introduction, the proposed system (ICF++), the methodology used for fake review detection. |

**The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process**

| | Process Steps | Advantage | Disadvantage (Limitation) |
|---|---|---|---|
| **1** | Data pre-processing. | it can help to reduce the time and resources required to train the model | it requires large scale of training data. |
| **2** | ICF++ (iterative computation framework). | This is a process that iteratively determines the honesty value. | As it involves iteration, the time complexity is expected to be relatively high. |
| **3** | Pos tagging (part-of-speech). | It offers valuable linguistic insights and fosters greater precision in understanding language within its context. | POS systems come with several functions – a lot more than a traditional cash register – they're a lot more costly. |
| **4** | Creation of transaction file. | Each row of the file is consist of noun value either (NN,NNP,NNPS,NNS). | It serves as an intermediary component, bridging the connection between the POS tagger and FP-Growth. |
| **5** | FP-growth. | extract information about the features of a product, this study applied the FP-Growth algorithm, which is part of association rule mining techniques. | This process is complex and relies on the utilization of the FP-Growth tree data structure. |
| **6** | Polarity generation. | This procedure is to determine the sentiment expressed in a sentence that includes the attributes identified in the prior step, classifying it as either positive, negative, or neutral. | The terms "funny" and "witty" are individually associated with positive sentiment. However, in the sentence "This movie was actually neither that funny, nor super witty," the combination of these terms results in a negative overall sentiment for the sentence. |

| Major Impact Factors in this Work | | | |
|---|---|---|---|
| **Dependent Variable** | **Independent Variable** | **Moderating variable** | **Mediating (Intervening ) variable** |
| Test Score | Calculation of agreement value | The correlation between test scores and the calculation of an agreement value contributes to enhancing the trustworthiness of reviewers and the product's reliability. | The test score and agreement value are determined based on the calculations of the product's trustworthiness and reliability values. |

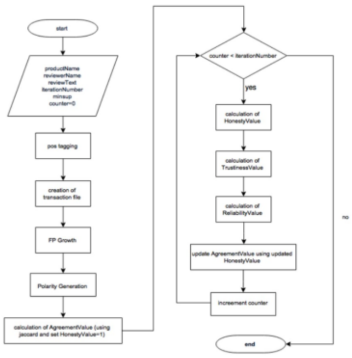
| Relationship Among The Above 4 Variables in This article |
|---|
| The iterative process of calculating trustworthiness, honesty, and product reliability is driven by the connection between test scores and agreement values, ultimately resulting in improved model performance. |

| Input and Output | | Feature of This Solution | Contribution & The Value of This Work |
|---|---|---|---|
| **Input** | **Output** | The paper focuses on feature extraction using the FP-Growth algorithm, polarity generation and sentiment prediction, fake review detection using the ICF algorithm, and the calculation of agreement and honesty values. The paper suggests that incorporating semantic aspects, such as sentiment polarity, can improve the accuracy of fake review detection. | This paper encompasses two iterative approaches, namely ICF and ICF++, which are employed to detect counterfeit reviews, relying on the measures of honesty and the product's reliability. |
| product review from Amazon.com from June 1995 - March 2013, data retrieved from https://snap.stanford.edu/data/web-Amazon.html | It shows accurate method among icf and icf++ based on the calculations. | | |

| Positive Impact of this Solution in This Project Domain | Negative Impact of this Solution in This Project Domain |
|---|---|
| This research is to identify fraudulent reviews for a product by analyzing the textual content and rating associated with each review. This analysis will determine the review's integrity, the reviewer's credibility, and the product's dependability. | The proposed system involves an intricate, iterative process that assesses the honesty value of the product and involves agreement calculations, resulting in high time complexity.<br>The specific equations for calculating the trustworthiness and honesty values are not provided in the given document content. |

| Analyse This Work By Critical Thinking | The Tools That Assessed this Work | What is the Structure of this Paper |
|---|---|---|
| The process involves iterative calculations of four key measures: honesty value, trustworthiness value, and reliability value of the reviews. In the case of ICF++, each calculation and assessment must be achievable, ultimately enhancing the model's accuracy. | • Iterative Computation Framework (ICF)<br>• Part-Of-Speech (POS) Tagger<br>• FPGrowth | Abstract<br>I. Introduction.<br>II. Methodology.<br>III. Results and Discussions.<br>IV. Conclusion |

**Diagram/Flowchart**




| 8 |
|---|

| Reference in APA format | Mayuri patil, snehal nikumbh, Fake product review monitoring and removal for genuine product reviews. | |
|---|---|---|
| **URL of the Reference** | **Authors Names and Emails** | **Keywords in this Reference** |
| IJSRED - Low Publication Fees, Article Publish within 24 hours, Submit Your Research Papers online Publication, | Mayuri patil, snehalnikumbh, aparna parigond, madhavi patil. | Opinion spam, opinion mining, genuine review, fake reviews. |

| The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc ) | The Goal (Objective) of this Solution & What is the problem that need to be solved | What are the components of it? |
|---|---|---|
| fraud risk management system and removal model. | This method identifies fraudulent transactions by assessing user behaviour and network activity, and it then processes these transactions in real-time through Data Mining to make precise predictions regarding suspicious users and transactions. | System architecture input selection , spam detection, spammed content analysis. |

**The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process**

| | Process Steps | Advantage | Disadvantage (Limitation) |
|---|---|---|---|
| **1** | Input Selection:- A specific dataset is employed to discern and distinguish positive from negative reviews by leveraging commonly used keywords found in these reviews. | In this section different types of data sets is used based on this dataset reviews will be categorized as fake or genuine | Datasets which contain symbols, stars, emoji's are not categorized. |
| **2** | Input processing :- input obtained after input selection is processed and readied. | Stop-words will be removed and data will be managed. | Longer words are challenging to categorize and demand additional time. |
| **3** | Spam detection:- nlp is one of the spam detection technique in this paper. | Any dual view data that is any redundant data is removed and also duplicated reviews and unknown reviews are also removed. | It requires IP address of user to categorize it as spam. |

| | | | |
|---|---|---|---|
| **4** | Feature extraction :-where nlp and TF*IDF algorithms are applied. | Predict if reviews are positive or negative and is done using bag pf words. | It can detect only the presence of positive and negative sentiments in reviews and is unable to identify neutral ones. |
| **5** | Spammed content analysis:- Reviews are classified and divided into fake reviews and spammed reviews. | The genuine reviews are visible to the user. | |

**Major Impact Factors in this Work**

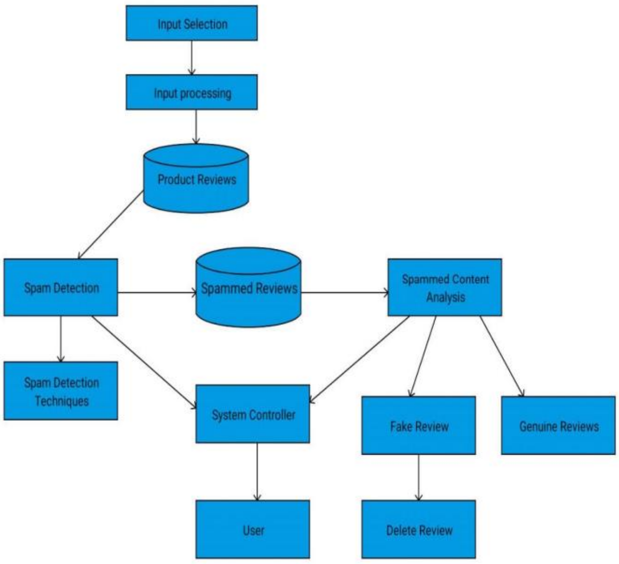
| **Dependent Variable** | **Independent Variable** | **Moderating variable** | **Mediating (Intervening) variable** |
|---|---|---|---|
| Reliability | These include supervised learning, Pu-learning algorithm, TF-IDF(Term frequency-inverse document frequency) | The review's reliability is determined by assessing its content for spam, categorizing it as either a fake review, in which case the administrator will remove it, or routing it to the system controller if it passes the spam check. | The process behind assessing reliability and detecting spammed content involves employing NLP and TF*IDF techniques, which, through sentiment analysis, categorize words or sentences as either positive or negative and determine whether they are spam or fake. |

| **Relationship Among The Above 4 Variables in This article** |
|---|
| Evaluating the trustworthiness of an application or review relies on spammed content analysis, which involves assessing the sentiment conveyed beyond individual words or sentences to categorize them as positive or negative. This analysis ultimately focus on user confidence when making purchases on e-commerce websites or applications. |

<table>
<tr><th>Input and Output</th><th>Feature of This Solution</th><th>Contribution & The Value of This Work</th></tr>
<tr><td><table><tr><th>Input</th><th>Output</th></tr><tr><td>Kaggle review datasets</td><td>Spammed and fake reviews.<br>Genuine review.</td></tr></table></td><td>This practice is commonly referred to as "Opinion Spam," wherein spammers engage in the creation of fake, misleading, or dishonest reviews with the intent of enhancing their product's reputation for financial gain, while also undermining their competitors' products. To address this issue, this paper suggests the development of a fraud risk management system and a removal model.</td><td>The paper facilitates the identification of counterfeit or spam content within product reviews. If a review is determined to be spam or fake, the system outlined in the paper will be able to detect it, subsequently bolstering user confidence in the product.</td></tr>
</table>

| Positive Impact of this Solution in This Project Domain | Negative Impact of this Solution in This Project Domain |
|---|---|
| This paper introduces a sense of trustworthiness within the realm of e-commerce, a domain where many users heavily rely on online reviews as their primary source of information for purchasing products. The proposed system outlined in this paper, with its efficient spam and fake review detection capabilities, significantly enhances the productivity of the company. | The system proposed in the paper employs the system's IP address, potentially giving rise to security concerns for users. |

| Analyse This Work By Critical Thinking | The Tools That Assessed this Work | What is the Structure of this Paper |
|---|---|---|
| In the process of identifying spam and fake reviews, this paper has employed a variety of methods and functions, including NLP techniques for text pre-processing and network-based approaches such as the user's IP address. | Nlp tools<br>Stop-words, Tf*idf, | Abstract<br>1)Introduction.<br>2)Problem definition.<br>3)Literature survey.<br>4)Proposed system.<br>5)Algorithm of proposed system.<br>6)Conclusions. |

| Diagram/Flowchart |
|---|
| <br>Fig. 1System Architecture |

| 9 | | |
|---|---|---|
| **Reference in APA format** | N Deshani, B Bhasakara rao, Deep learning hybrid approaches to detect fake reviews and ratings. | |
| **URL of the Reference** | **Authors Names and Emails** | **Keywords in this Reference** |
| https://nopr.niscpr.res.in/bitstream/123456789/61198/1/JSIR%2082%2801%29%20120-127.pdf | N Deshani.<br>B Bhaskara rao. | CNN-LSTM, Glove,LSTM-RNN, One hot encoding. |
| **The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc )** | **The Goal (Objective) of this Solution & What is the problem that need to be solved** | **What are the components of it?** |
| The paper proposes two novel deep-learning hybrid techniques CNN-LSTM . | Primary goal is to accurately detect fake reviews and what is the main difference between them.<br>Secondary goal is to detect fake ratings and actual ratings-based reviews across the online platform especially Amazon datasets. | This paper consist of data pre-processing, classifiers, model performance. |

| The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process | | | |
|---|---|---|---|
| | **Process Steps** | **Advantage** | **Disadvantage (Limitation)** |
| **1** | Pre-processing:-collection of noise like hyperlinks, HTML tags, unofficial comments and feature extraction. | Valuable insights are gleaned from the content. | Numerous Python libraries have been employed, alongside the need for various natural language processing (NLP) modules. |
| **2** | One hot encoding & glove:- It is a deep learning technique to be applied to sequential classification problems. | Categorical variables as binary vectors to be more expressive and get a better prediction. | Glove model is un-supervised method is trained via least squares using the cost function. |
| **3** | Hybrid classifier :- Deep learning neural network models are used for analysing identifying, and categorizing fraudulent reviews. | CNN-LSTM and LSTM-RNN which increases the accuracy of the recommended hybrid models. | Diverse models are necessary for varying ratings and reviews. |
| **4** | Metrics for model performance:- From the confusion matrices, it is feasible to create a variety of performance measures by basing them on the rates of false-positive and false-negative items. | It demonstrates the precise accuracy, true positive, and true negative values for each model within the system. | More than two function were used like Sensitivity and specificity. |

| Major Impact Factors in this Work | | | |
|---|---|---|---|
| **Dependent** | **Independent Variables** | **Moderating variable** | **Mediating (Intervening ) variable** |
| Performance | LSTM,LSTM-RNN ,one-hot encoder | Based on the LSTM is suggested to predict fake ratings.<br>LSTM-RNN is recommended to detect fake ratings .<br>One hot encoder main strategy is to convert to a numerical vector. | LSTM-RNN is a multilayer perceptron that gains its effectiveness through training on extensive datasets. As the algorithm's performance improves, it enhances the accuracy of results it can predict. |

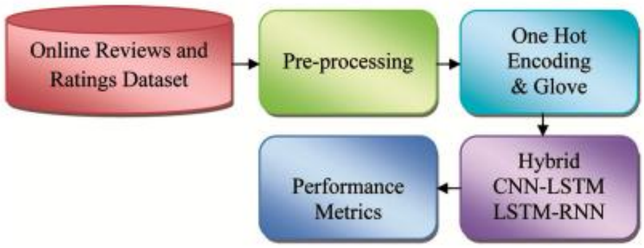
**Relationship Among The Above 4 Variables in This article**

Within this paper, it is asserted that the performance of the model is intricately linked to the interaction of variables at each layer in LSTM and LSTM-RNN. Each node, with its unique weightage, plays a role in enhancing the model's performance.

<table>
<tr><th>Input and Output</th><th>Feature of This Solution</th><th>Contribution in This Work</th></tr>
<tr><td><table><tr><th>Input</th><th>Output</th></tr><tr><td>Amazon review datasets.</td><td>Detecting fake online reviews.<br><br>Detecting fake ratings.</td></tr></table></td><td>The Paper proposes two novel deep-learning Hybrid techniques: CNN-LSTM for detecting fake online reviews, and LSTM-RNN for detecting fake ratings in the e-commerce domain.</td><td>Leveraging a multilayer perceptron leads to enhanced accuracy, achieved through extensive training on large datasets, resulting in heightened performance. RNN methods offer efficiency and practicality, potentially making them more suitable for achieving optimal outcomes and maximizing the efficacy of detecting fake online reviews.</td></tr>
</table>

| Positive Impact of this Solution in This Project Domain | Negative Impact of this Solution in This Project Domain |
| --- | --- |
| The paper employs two hybrid models for the identification of counterfeit reviews and ratings, delivering precise outcomes by considering users' historical experiences with the product. | Since it is multilayer perceptron it requires large amount of training data to predict efficiently. |

| Analyse This Work By Critical Thinking | The Tools That Assessed this Work | What is the Structure of this Paper |
| --- | --- | --- |
| The system introduced in the paper is a multilayer perceptron that utilizes techniques such as LSTM-CNN and LSTM-RNN, both of which belong to the family of recurrent neural networks. These algorithms are applied with a variety of activation functions and necessitate extensive training on large datasets to achieve accurate predictions. | Data pre-processing, (NLP) tools lemmatization, tokenization. | • Introduction<br>• Related work.<br>• Proposed hybrid deep learning framework.<br>• Experimental Analysis.<br>• Conclusions. |

| **Diagram/Flowchart** |
| --- |
|   Fig. 2 — Proposed Deep learning hybrid methodology |

| **10** | | |
| --- | --- | --- |
| **Reference in APA format** | Ahmed M.Elmogy, Usman Tariq fake reviews detection using supervised machine learning. | |
| **URL of the Reference** | **Authors Names and Emails** | **Keywords in this Reference** |
| https://thesai.org/Publications/IJACSA | Ahmed M.Elmogy, Usman tariq. | Fake reviews detection, supervised machine learning |
| **The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc )** | **The Goal (Objective) of this Solution & What is the problem that need to be solved** | **What are the components of it?** |
| • Nlp (data pre-processing)<br>• k-nearest neighbour.<br>• logistic regression. | This paper introduces a machine learning method for detecting fraudulent reviews. Alongside the review feature extraction process, the approach incorporates several techniques for feature engineering to capture diverse reviewer behaviours. | This paper contains Data Pre-processing, feature extraction, feature engineering, evolution and testing. |

**The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process**

| | Process Steps | Advantage | Disadvantage (Limitation) |
|---|---|---|---|
| **1** | Data pre-processing | All data are cleaned from the stop words before going forward in the fake reviews detection process. | It's a multi-tiered process in which each level is interconnected and operates simultaneously. |
| **2** | Feature extraction | It is mainly a procedure of removing the unneeded attributes from data that may actually reduce the accuracy of the model | It requires two language models like tri- grams, bigrams. |
| **3** | Comparison of Extracted Features | All these features are taken into consideration to see the effect of the users behaviours on the performance of the classifiers. | It consists of numerous functions aimed at calculating the average for each individual word. |

**Major Impact Factors in this Work**

| Dependent Variable | Independent Variable | Moderating variable | Mediating (Intervening ) variable |
|---|---|---|---|
| identification of fake reviews. | textual features of the reviews.<br>sentiment classification, cosine similarity, and TF-IDF. | The paper doesn't discuss any moderating variables. | The paper doesn't mention mediating variables. Instead, it focuses on extracting textual and behavioural features to identify fake reviews |

**Relationship Among The Above 4 Variables in This article**

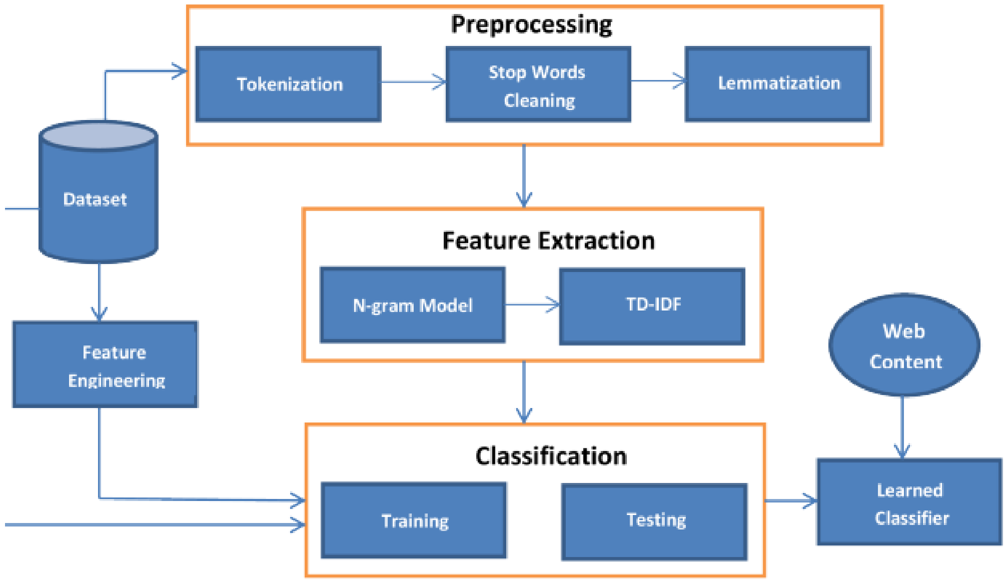
The dependent variable is predicted or measured based on both the independent and behavioural features. This indicates that there is a relationship between the independent variables and the dependent variable and the behavioural features play a role in the performance of the detection process.

<table>
<tr><th>Input and Output</th><th>Feature of This Solution</th><th>Contribution & The Value of This Work</th></tr>
<tr><td><table><tr><th>Input</th><th>Output</th></tr><tr><td>User reviews from e-commerce websites.</td><td>To identity the user’s review whether it’s fake or genuine.</td></tr></table></td><td><ul><li>Different classifiers are implemented in the developed approach.</li><li>The Bi-gram and Trigram language models are used and compared in the developed approach.</li><li>The solution also takes into account aspects related to the reviewers, such as the timing of the reviews and their writing styles, to enhance the identification of fake reviews.</li></ul></td><td><ul><li>This paper illustrates how user behaviour can be discerned based on the reviews they have posted and the historical usage of words within those reviews.</li><li>By examining not just the content of the reviews but also the behaviour of the reviewers, the suggested approach offers a more thorough analysis.</li></ul></td></tr>
</table>

<table>
<tr><th>Positive Impact of this Solution in This Project Domain</th><th>Negative Impact of this Solution in This Project Domain</th></tr>
<tr><td><ul><li>What prompts users to make decisions based on reviews.</li><li>importance of reviews and how they affect almost everything related to web based data.</li><li>it considers not only the key features of the reviews but also the behaviours of the reviewers.</li></ul></td><td>Unable to distinguish whether reviews have been authored by humans or bots.</td></tr>
</table>

<table>
<tr><th>Analyse This Work By Critical Thinking</th><th>The Tools That Assessed this Work</th><th>What is the Structure of this Paper</th></tr>
<tr><td>It comprises distinct sets of blocks, with each block serving a specific function. For example, data pre-processing involves the use of NLP techniques, and feature engineering is applied to uncover user behaviour. The process necessitates the utilization of two language models, namely n-gram and bi-gram models and inclusion of the extracted behavioural features improves the performance of the classifiers.</td><td><ul><li>NLP tools like stop words, lemmatization</li><li>KNN, logistic regression</li></ul></td><td>Abstract<ol><li>Introduction.</li><li>Related work.</li><li>Background.</li><li>proposed approach.</li><li>Experiment results.</li><li>Conclusion</li></ol></td></tr>
</table>

**Diagram/Flowchart**




**11**

| **Reference in APA format** | S. Uma Maheswari,Dr.S.S. Dhenakaran, June 2021, Detection of fake and Genuine Reviews with Hybridization of Fuzzy and Neural Networks Techniques | |
|---|---|---|
| **URL of the Reference** | **Authors Names and Emails** | **Keywords in this Reference** |
| https://www.researchgate.net/publication/352399781 | Uma Maheswari Ph.D. Research Scholar, (E-mail: 17umeshrani@gmail.com). S S Dhenakaran Professor, (E-mail: ssdarvind@yahoo.com), Department of Computer Science, Alagappa University, Karaikudi. | Sentiment Analysis, Classification, Fuzzy Logic, Deep Learning, Neural Networks, Genuine Reviews. |
| **The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ...etc )** | **The Goal (Objective) of this Solution and what is the problem that needs to be solved** | **What are the components of it?** |
| • Text Pre-processing<br>• Word Dictionary. | The main aim of this work is to help customers identify | The author employed four different techniques, User-defined |

| | | |
|---|---|---|
| • User Defined Classification.<br>• Fuzzy Logic.<br>• Deep Learning.<br>• Machine Learning. | fake reviews on social media and websites based on selected features for better decisions on product purchases online and is method of classification categorizing reviews into different categories which include positive, negative, and neutral, and accuracy is been compared with existing methods of ML | Fuzzy Logic, deep learning, and Machine Learning for Sentiment classification and prediction based on the accuracy and f1 score. Furthermore, classification is done according to the score as positive, negative, neutral, Positively fake, negatively fake, and so on. |

**The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process**

Machine learning approach for fake review detection works as follows

| | Process Steps | Advantage | Disadvantage (Limitation) |
|---|---|---|---|
| **1** | User Defined Classification | The author has used user-defined conditional statements for classification which is more accurate with the base conditions and was able to classify as positive and negative reviews. | Based on the conditions of the user and As the size of the data is smaller, the result is accurate, when it comes to the larger data it may take time to execute more number of iteration which increases the time complexity |
| **2** | Fuzzy Logic | Fuzzy logic can be used in uncertain/ambiguous situations. It is a multi-valued mechanism and it can produce higher accuracy in classification. Fuzzy Logic involves the degree of truth and the degree of membership. In other words fuzzy logic is not like binary classification (yes or no) and (0 or 1), it can recognize intermediate multiple values between the range of 0 and 1. | For more accuracy, needs more fuzzy grades which results in increasing exponentially the rule, Lack of real-time response, Restricted number of usage of input variables. |
| **3** | Deep Learning | LSTM which is also known as RNN involves 7 steps, uses the Adam optimizer for performing the multiclass | Deep learning requires a large amount of data. Complex data models require expensive GPUs. |

| | | | |
|---|---|---|---|
| | | classification, and the Cross-Entropy loss function is used to evaluate the network model with this loss value decreased by the optimizer | Overfitting may also occur due to an excess amount of training data. |
| 4 | Machine Learning Algorithms | Five different techniques like NB, SVM, DT, LR, and RF have been implemented and compared to segregate the fake and genuine review classification and prediction with different accuracy scores | As there are five different techniques there is a chance of mixed results and an increase in loss of genuine reviews. |

**Major Impact Factors in this Work**

| Dependent Variable | Independent Variable | Moderating variable | Mediating (Intervening) variable |
|---|---|---|---|
| Overall Accuracy score | Data Entities(IP Address, Location), Machine Learning Algorithms (NB, SVM, DT, LR, RF) | The paper does not explicitly mention a moderating variable. However, one can argue that the "segmentation ratio" might act as a moderating variable. | The paper does not explicitly mention a mediating variable. However, one could consider the "text preprocessing" process as a potential mediating variable. |

**Relationship Among the Above 4 Variables in This article**

The independent variables include the machine learning algorithms used (Naive Bayesian, Random Forest, Decision Tree, Logistic Regression, Support Vector Machines), the data entities like IP Address, location, and the text preprocessing methods.

A potential mediating variable in this context could be "text preprocessing," as it serves to prepare the raw review data, removing punctuation, special symbols, and meaningless vocabulary. This preprocessing step may influence the quality of the data used for training and, subsequently, the classification accuracy (dependent variable).

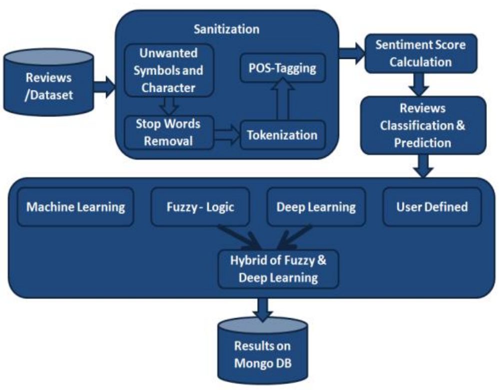
The overall accuracy score of the given data will depend on the proposed algorithms which undergo the Sanitization process and sentimental score calculation..

| Input and Output | Feature of This Solution | Contribution & The Value of This Work |
|---|---|---|

<table>
<tr>
<td>
<table>
<tr><th>Input</th><th>Output</th></tr>
<tr><td>9000 customer reviews primarily from the amazon mobile electronic products</td><td>Categorizing these reviews into groups.</td></tr>
</table>
</td>
<td>It offers a customization feature that enables users to classify the reviews as Fake and Genuine. The solution also relates to the work of classification of customer reviews based on different categories like positive, negative, and neutral through sentiment analysis.</td>
<td>We have seen the different algorithms performing different types of analysis with the same dataset and attributes. Every algorithm follows its own approach for processing the given data but differs in the performance analysis. This paper has proven that performance always differs from the algorithms used with similar data.</td>
</tr>
</table>

<table>
<tr><th>Positive Impact of this Solution in This Project Domain</th><th>Negative Impact of this Solution in This Project Domain</th></tr>
<tr><td>The usage of different methods to achieve higher accuracy like Fuzzy logic, User defined classification shows a positive impact on this project with an increase in accuracy.</td><td>Since this is a performance evaluation of various algorithms, No negative side to this project as all the things used are defined in advance.</td></tr>
</table>

<table>
<tr><th>Analyse This Work by Critical Thinking</th><th>The Tools That Assessed this Work</th><th>What is the Structure of this Paper</th></tr>
<tr><td>This paper has truly enhanced our understanding of how the size of data and methods used, plays a crucial role in influencing the performance of different algorithms.</td><td>Web Browser</td><td>Abstract<br>I. Introduction<br>II. Related Works<br>III. Proposed Work<br>IV. Experimental Analysis<br>V. Conclusion and References</td></tr>
</table>

<table>
<tr><th>Diagram/Flowchart</th></tr>
<tr><td>


</td></tr>
</table>

| 12 | | |
|---|---|---|
| **Reference in APA format** | Pansy Nandwani, Rupali Verma, 2021, A review on sentiment analysis and emotion detection from text. | |
| **URL of the Reference** | **Authors Names and Emails** | **Keywords in this Reference** |
| https://doi.org/10.1007/s13278-021-00776-6 | Pansy Nandwani, Rupali Verma | Affective computing, Natural Language Problems, Opinion mining, Preprocessing. Word embedding |
| **The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ...etc )** | **The Goal (Objective) of this Solution & Whatis the problem that need to be solved** | **What are the components of it?** |
| A review-based approach to find the best method to perform sentiment and emotion analysis on the given text. | The main objective of this paper is to compare the existing techniques for both emotion and sentiment detection and find out the best-performing algorithm with an accuracy score. | This paper includes problem identification, data collection, pre-processing, Feature extraction Model Development, Model assessment, Challenges, and conclusion. |

**The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process**

| | Process Steps | Advantage | Disadvantage (Limitation) |
|---|---|---|---|
| **1** | Preprocessing of the collected data is mainly performed based on three methods which include Tokenization, Stop word Removal, POS Tagging | The quality and significance of the data are achieved by data preprocessing. It helps in increasing the accuracy of the model. | Some of the preprocessing techniques can result in the loss of crucial information for sentiment and emotion analysis. |
| **2** | Feature Extraction | Easy implementation and easy to identify the word count using Bag Of Words (BOW), and n-gram models. | While using Bag Of Words the order of the words in the input sentence may not be the same, which causes semantic errors. |
| **3** | Web Scraping Techniques | SVM can handle both linear and non-linear decision boundaries, through the | SVM can be computationally expensive, especially for large |

| | | | |
|---|---|---|---|
| | | different kernel functions and it is less prone to overfitting compared to the other algorithms. | datasets. It may be sensitive to the choice of kernel function. |
| 4 | Machine Learning-based approach(NB, SVM, DT) | Here, the use of different Machine Learning algorithms helps in improving the accuracy of the model. By using different feature extraction vectors like BOW, and Unigram with Sentiwordnet the accuracy will be higher. | In some cases, machine learning models like SVM, DT, and NB fail to extract some implicit features or aspects of the text. The performance of the Machine learning model depends upon the size of the data and the preprocessing techniques used. |
| 5 | Deep Learning-based approach | With this approach, we can gain insights into the data with the help of the computer, and will be helpful to us in automatic feature extraction. | None |
| 6 | Transfer Learning Approach and Hybrid approach | This method allows the model to reuse the pre-trained models which increases the performance of the model. | None |

**Major Impact Factors in this Work**

| **Dependent Variable** | **Independent Variable** | **Moderating variable** | **Mediating (Intervening ) variable** |
|---|---|---|---|
| Accuracy Score of the sentiment analysis | Data collection, pre-processing, testing performance | The paper does not explicitly mention any moderating variables. However, the process of feature extraction like the usage of Bag Of Words (BOW) and n-gram models acts as the moderating variables. | The paper does not mention any mediating variables. However, the sentiment analysis algorithm itself acts as an intermediary between the independent variables (stages of thesentiment analysis process) and the dependent variable. |

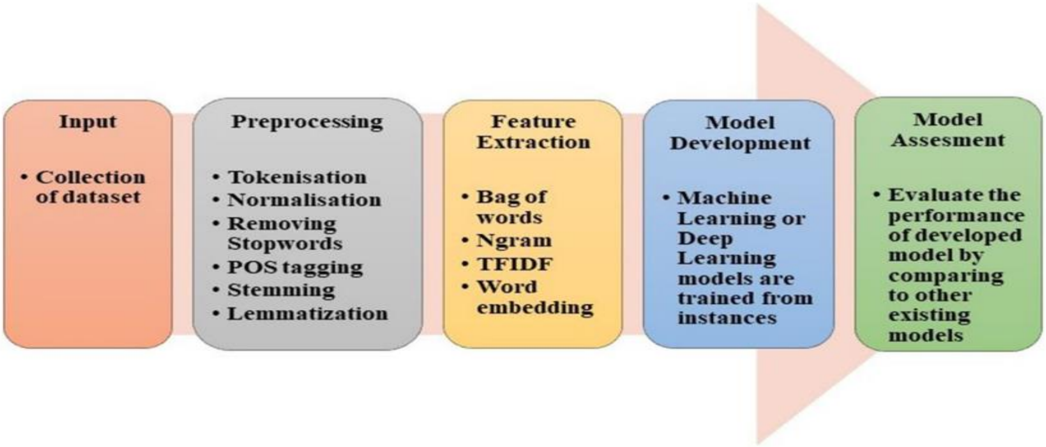
| Relationship Among The Above 4 Variables in This article |
|---|
| The accuracy of the model will analyzed with the help of given data after the preprocessing of the data. First the data is preprocessed and the featured extraction is done with moderating variables.<br>The sentiment of reviews depends on the application of the different lexicon-based and deep learning-based approaches, as the algorithm's output defines the challenges faced during this sentiment analysis. |

| Input and Output | Feature of This Solution | Contribution & The Value of This Work |
|---|---|---|
| **Input**: Preprocessed dataset as the input to perform accuracy test.<br>**Output**: Accuracy score | This solution mainly focuses on sentiment analysis and emotion detection from text which discusses the sentiments and emotions in given input text and addresses the challenges faced in it. | This paper talks about how we can make use of the resources to analyze the text based on the sentiment and emotion pattern. It also provide the accuracy score based on the weights of the terms included |

| Input | Output |
|---|---|
| Preprocessed dataset as the input to perform accuracy test. | Accuracy score |

| Positive Impact of this Solution in This Project Domain | Negative Impact of this Solution in This Project Domain |
|---|---|
| Using the proposed method for sentiment analysis will provide more accuracy in detecting the sentiments and emotions in text, we may find opportunities to further improve the accuracy and effectiveness of our sentiment analysissystem. | The author used different techniques to access the best accurate model which increases the computational costs. |

| Analyse This Work By Critical Thinking | The Tools That Assessed this Work | What is the Structure of this Paper |
|---|---|---|
| This paper provides insights into the levels of sentiment analysis, methodologies, challenges, and applications of sentiment and emotion analysis from text data. However, challenges such as spelling mistakes, new slang, and incorrect grammar usage make sentiment and emotion analysis complex tasks. | None | Abstract<br>I. Introduction<br>II. Background<br>III. Process of sentiment analysis and emotion detection<br>IV. Challenges |

| Diagram/Flowchart |
|---|
|  |



| 13 |
|---|

| | | |
|---|---|---|
| **Reference in APA format** | Tanjim Ul Haque, 2018, Sentiment Analysis on Large Scale Amazon Product Reviews. | |
| **URL of the Reference** | **Authors Names and Emails** | **Keywords in this Reference** |
| https://www.researchgate.net/publication/325756171 | Tanjim Ul Haque, Nudrat Nawal Saber, Faisal Muhammad Shah. | Sentiment Analysis, pool-based active learning, feature extraction, text classification, Machine learning. |
| **The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/)** | **The Goal (Objective) of this Solution & What is the problem that need to be solved** | **What are the components of it?** |
| A machine learning approach to perform sentiment analysis on Amazon product reviews helps to polarize reviews into two different classes. | The main goal of this solution is to calculate the accuracy score of polarizing the reviews into positive and negative using different Amazon product datasets based on some Machine learning algorithms | This paper includes problem identification, data collection, pre-processing, building a Semi-Supervised machine learning model comparing with different ML algorithms, and presenting the results and discussion. |
| **The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process** | | |

| | Process Steps | Advantage | Disadvantage (Limitation) |
|---|---|---|---|
| **1** | **Data Acquisition** is the primary task where the data is collected from different data sources using cloud services like Oracle | Data retrieval is been carried out with the help of online resources, Accurate data is generated with the help of Pool Based Active Learning which provides pre-labeled datasets. | It is hard to gather huge amounts of gold-standard datasets for this purpose as e-commerce sites have thei limitations on giving data publicly. |
| **2** | **Pre-processing** is the next crucial step in this process of polarizing the reviews<br><br>**Feature Extraction**: This involves extracting useful words from the dataset. Two methods used for feature extraction are Bag of Words and TF-IDF. Bag of Words represents a document as a list of its words, while TF-IDF weighs a term's frequency and inverse document frequency to determine its significance. | This results in an optimal solution more quickly compared to traditional gradient descent methods.<br><br>The bag of words approach simplifies text or data by representing it as a collection of its words. This simplification makes it easier to analyze and process the data. | Its effectiveness depends on the specific problem being solved.<br><br>During the feature extraction process, some information may be lost. This is especially true whe using techniques like the bag of words approach, where the order and contex of words are disregarded. This loss of information can affect the accuracy anc completeness of the analysis |
| **3** | **Pull-Based Active Learning**: In the active learning process, a pool of unlabeled data is used. The learning method asks an oracle or user to label a few data points, and classifiers are run to calculate accuracy. If the accuracy is greater than or equal to 90%, the labeled data is combined with pre-labeled data. If not, more data is labeled with the help of the oracle. Once the accuracy is greater than 90%, the data is considered labeled. | Pull-based active learning will be helpful in accelerating the machine learning tasks and it also improves the performance of the model. | The dependency on the external dataset can get the bais or user to label the selected examples. Pull-based active learning implementation is more complex. |

| Major Impact Factors in this Work | | | |
|---|---|---|---|
| **Dependent Variable** | **Independent Variable** | **Moderating variable** | **Mediating (Intervening ) variable** |
| Accuracy Score of the Algorithm used in this paper. | Data(Review ID, Rating, Time and name), pre-processing, Pool based learning | This paper does not contain any Moderating variables excluding the preprocessing and | The paper does not mention any mediating variables. However, the sentiment analysis algorithm itself acts as an intermediary between the independent variable (stages of the sentiment analysis process) and the dependent variable (accuracy performance). |

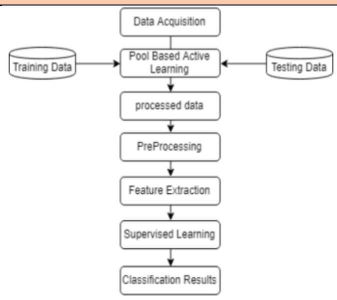
| Relationship Among The Above 4 Variables in This article |
|---|
| These mediating variables could include aspects of the algorithm's architecture, pre-processing steps, or other factors that influence how the algorithm interprets and classifies the sentiment in the reviews.<br>The sentiment of reviews depends on the application of the ANN algorithm, as the algorithm's outpu classifies reviews as positive or negative based on patterns it has learned. |

| Input and Output | | Feature of This Solution | Contribution & The Value of This Work |
|---|---|---|---|
| **Input** | **Output** | **Pool Based Learning** along with Machine Learning algorithms like Linear SVM, Naïve Bayes, Stochastic Gradient Descent, Random Forest | This paper talks about how we can improve the accuracy of the classification while performing a sentiment analysis. The author has used a Pool based active learning strategy on the raw data to make the input data more accurate which parallelly improve the performance of the model with accurate result. |
| Amazon Labeled Dataset after the active learning process | Accuracy of classifier Precision, Recall, F1-Measure for positive and Deceptive values. | | |

| Positive Impact of this Solution in This Project Domain | Negative Impact of this Solution in This Project Domain |
|---|---|
| The high accuracy achieved by the model in classifying the sentiments can help the customers to get a better user experience in finding the right things and also gain insights into customer opinions and preferences. | One of the negative impacts of this solution is that while dealing with active learning and feature extraction techniques, manual labeling is also required to achieve accurate results. |

| Analyse This Work By Critical Thinking | The Tools That Assessed this Work | What is the Structure of this Paper |
|---|---|---|
| The Author has addressed the limitations and challenges of labeling the raw data and preprocessing the data due to the presence of limited standard datasets from e-commerce sites | Active learning algorithms, Supervised learning model. | Abstract<br>I. Introduction<br>II. Related Works and research<br>III. Methodology<br>IV. Results<br>V. Comparative Analysis<br>VI. Conclusion and Future works |

**Diagram/Flowchart**




**14**

| Reference in APA format | Mr. Navjyotsinh, Chirag visani, 2017, A STUDY ON DIFFERENT MACHINE LEARNING TECHNIQUES FOR SPAM REVIEWS DETECTION | |
|---|---|---|
| **URL of the Reference** | **Authors Names and Emails** | **Keywords in this Reference** |
| https://www.researchgate.net/publication/318982640 | Mr. Navjyotsinh Jadeja, Chirag visani, navjyotsinh.jadeja@marwadieducation.edu | Text mining, Supervised techniques Support vector Machine, Naïve bayes. |
| **The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/** | **The Goal (Objective) of this Solution & What is the problem that need to be solved** | **What are the components of it?** |

| Model/ Tool/ Framework/ ... etc) | | |
|---|---|---|
| An approach to segregate the reviews in three different categories which are positive, negative, and neutral. This is a study of different machine-learning techniques for detecting spam reviews | To analyze Amazon product reviews and predict the ratings of future reviews It also includes the process of extracting the meaningful narratives | This paper discusses various methods for detecting artificially generated texts on the internet. It explores techniques such as hidden style similarity, frequency counting, linguistic features, and machine learning algorithms. The paper also highlights the challenges and limitations in the field of artificial text detection |
| **The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process** | | |
| **Major Impact Factors in this Work** | | |

| | Process Steps | Advantage | Disadvantage (Limitation) |
|---|---|---|---|
| **1** | 1. Gathering Training Dataset: The first step is to collect a dataset of reviews that are labeled as spam or non-spam. This dataset will be used to train the machine learning model. | A large amount of Amazon data is been collected which gives the data of productrelated reviews | Multiple data frames can be found which are not relevant and decreases the accuracy of the model |
| **2** | Preprocessing the Data: The collected dataset needs to be preprocessed to remove any irrelevant information, such as special characters or stopwords. The text data may also need to be tokenized and normalized. | In this method, the preprocessing improves the quality of the data by cleaning and transforming it into a suitable format for machine learning algorithms. It helps in improving the accuracy and efficiency. | Lack of standardized data and facing difficulties when handling noisy data and Overfitting may occur. |
| **3** | Feature Extraction: Next, similar features need to be extracted from the preprocessed data. These features can include word frequencies, n-grams, or other linguistic features that can help distinguish between spam and non-spam reviews. | In this process, the main advantage is to analyze the pattern and improve the efficiency of the model. | In some cases there will be loss of data and the order of the words may be changed. |
| | Support Vector Machine(SVM) | SVMs are powerful machine learning technique for spam review detection, offering high accuracy and the ability | Support Vectors are computationally expensive and limited to some of the applications. |

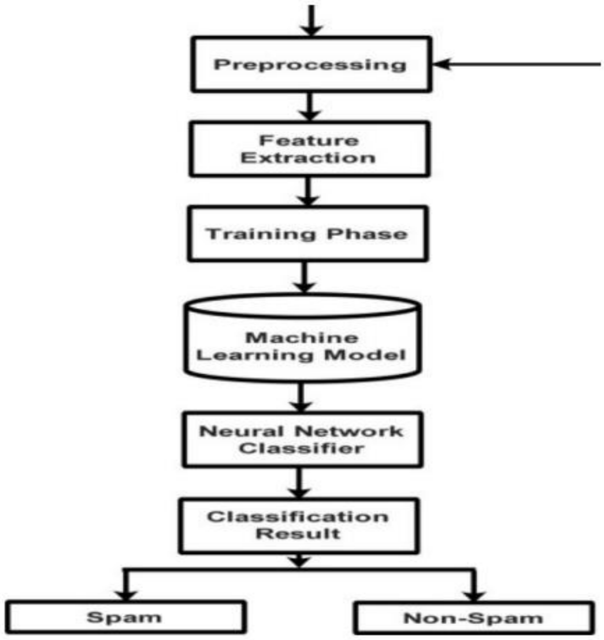
| | | | |
|---|---|---|---|
| | | to handle high-dimensional and non-linear data | |
| | Naive Bayes and Logistic Regression | NB can handle large datasets efficiently and Logistic Regression is capable of handling both numerical and categorical data. | Both NB and LR do not capture the complex data patterns for further analysis |

| Dependent Variable | Independent Variable | Moderating variable | Mediating (Intervening ) variable |
|---|---|---|---|
| accuracy or effectiveness of accuracy | The independent variables in this paper are the set of ML algorithms and the data set. | Length of the review and the support count, time stamp. | In this paper there is n such Mediatin variables concept.inp. |

| Relationship Among The Above 4 Variables in This article |
|---|
| The paper discusses different supervised techniques, with each method having its own set of independent variables. Where the accuracy OS completely dependent on the machine learning algorithms and the algorithmic approach obtained by using parameters of the given data that is length and time stamp of the reviews. With the help of these moderating variables, the independent variables achieve the target. While mediating variables are not explicitly mentioned, they are likely to exist in the research process, influencing how the independent variables impact the effectiveness of spam review detection. |

| Input and Output | Feature of This Solution | Contribution & The Value of This Work |
|---|---|---|
| <table><tr><th>Input</th><th>Output</th></tr><tr><td>Dataset or the reviews</td><td>Classification of the reviews either sail and final lo</td></tr></table> | The paper discusses various methods and techniques for detecting spam reviews using machine learning. Its analyses and detects the review patterns. Using word bigram features and syntactic components for accurate detection. | By experimenting with various machine learning algorithms, this work contributes to the field of detecting spam reviews. The authors are particularly interested in leveraging Twitter as a platform for sentiment analysis and spam review detection. To identify spam reviews, they evaluate supervised and |

<table>
<tr><td></td><td></td><td colspan="2">unsupervised algorithms such as support vector machines (SVM), Nave Bayes classifiers, and logistic regression. The study also emphasises the importance of training data and the necessity for further improvement in spam review detection performance.</td></tr>
<tr><th colspan="2">Positive Impact of this Solution in This Project Domain</th><th colspan="2">Negative Impact of this Solution in This Project Domain</th></tr>
<tr><td colspan="2">The solutions are more helpful to improve the accuracy with the help of Svm or other ml algorithms and effective detection can also be experienced.</td><td colspan="2">There are no such negative impact on the paper, One such thing is that the Limited resources and privacy Concerns.</td></tr>
<tr><th>Analyse This Work By Critical Thinking</th><th colspan="2">The Tools That Assessed this Work</th><th>What is the Structure of this Paper</th></tr>
<tr><td>The paper showsthe importance of considering different strategies and data types when detecting artificial text. The context mainly discusses the use of different machine-learning techniques.</td><td colspan="2">None</td><td>Abstract<br>I. Introduction<br>II. II. Literature Search<br>III. III. Methodology<br>IV. IV. Algorithmic Techniques<br>V. V. Detecting Parameters<br>VI. VI. Results and Discuss<br>VII. VII. Conclusion</td></tr>
<tr><th colspan="4">Diagram/Flowchart</th></tr>
</table>




| 15 | | |
|---|---|---|
| **Reference in APA format** | Prof. P.S.Gaikwad, Kaushal Parmar, Rohit Yadav, Datta Supekar, 2021, IMPLEMENTATION OF WEB SCRAPING FOR E-COMMERCE WEBSITE | |
| **URL of the Reference** | **Authors Names and Emails** | **Keywords in this Reference** |
| https://www.jetir.org/papers/JETIR2106682.pdf | Prof. P.S.Gaikwad, Kaushal Parmar, Rohit Yadav, Datta Supekar | Web scraping, E-commerce, Data extraction, Web crawler |
| **The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc)** | **The Goal (Objective) of this Solution & What is the problem that need to be solved** | **What are the components of it?** |
| • Web Scrapping<br>• MySql<br>• Python<br>• BeautifulSoup<br>• Selenium | This solution aims to improve user convenience by allowing customers to compare products from many e-commerce websites on one page. | • The need for Web Scraping.<br>• Scraping different E-commerce websites.<br>• comparison of product prices. |
| **The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process** | | |

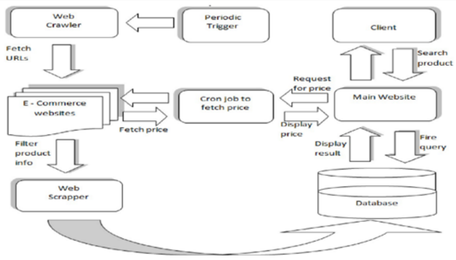
| | Process Steps | Advantage | Disadvantage (Limitation) |
|---|---|---|---|
| **1** | Taking product name as input from the user. | Gives the consumer a customized search experience by letting them identify the product they are interested in. | Depends on the user entering the product name correctly, which could result in inaccurate or inconsistent search results. |
| **2** | Scraping the product details. | Enables quick and efficient extraction of data from various websites, saving time and manual effort. | - |
| **3** | Displaying the information on the user'swindow. | Allows customers to compare and examine product facts from several websites in one place with an easy-to-use interface. | - |

**Major Impact Factors in this Work**

| Dependent Variable | Independent Variable | Moderating variable | Mediating (Intervening) variable |
|---|---|---|---|
| Product Details Displayed | Input Product Name | Web Scraping Tool | MySQL Database |

| Relationship Among the Above 4 Variables in This article |
|---|
| The product details displayed on the user's screen depend on the input product name provided by the user.<br>The choice of web scraping tool (e.g., BeautifulSoup or Selenium) has a effect on the relationship between the input and the displayed product details.<br>The MySQL database serves as an intervening variable, playing a role in storing and retrieving the scraped data before it is displayed to the user. |

| Input and Output | Feature of This Solution | Contribution in This Work |
|---|---|---|

<table>
<tr>
<td><table><tr><th>Input</th><th>Output</th></tr><tr><td>Product name</td><td>to view and compare product details from different websites</td></tr></table></td>
<td>This solution aims to solve the hard and manual task of visiting multiple websites and comparing product data. Users can save time and effort when evaluating products and making informed decisions by having data from multiple e-commerce websites scraped and presented on one page.</td>
<td>By automating the process of comparing product data from several websites on a single platform, this solution attempts to give users a quick and easy approach to making decisions.</td>
</tr>
<tr>
<td>Positive Impact of this Solution in This Project Domain</td>
<td colspan="2">Negative Impact of this Solution in This Project Domain</td>
</tr>
<tr>
<td>It gives users more alternatives by allowing the visibility of various products on a single website.</td>
<td colspan="2">--</td>
</tr>
<tr>
<td>Analyse This Work by Critical Thinking</td>
<td>The Tools That Assessed this Work</td>
<td>What is the Structure of this Paper</td>
</tr>
<tr>
<td>By using a single platform, the solution increases the effectiveness and simplicity of comparing products from many websites.</td>
<td>You tube</td>
<td>Abstract<br>• Introduction<br>• Motivation<br>• System Architecture<br>• Implementation<br>• Results<br>• Conclusion<br>• Future Work<br>VII. Acknowledgment</td>
</tr>
<tr>
<td colspan="3">Diagram/Flowchart</td>
</tr>
<tr>
<td colspan="3"></td>
</tr>
</table>

| 16 | | |
|---|---|---|
| **Reference in APA format** | Mr. Karthikeyan T, Mr. Karthik Sekaran, Mr. Ranjith D, Mr. Vinoth Kumar V, Mr.Balajee, PERSONALIZED CONTENT EXTRACTION AND TEXT CLASSIFICATION USING EFFECTIVE WEB SCRAPPING TECHNIQUES, 2019. | |
| **URL of the Reference** | **Authors Names and Emails** | **Keywords in this Reference** |
| https://www.jetir.org/papers/ JETIR1904I22 .pdf | Mr. Karthikeyan T, Mr. Karthik Sekaran, Mr. Ranjith D, Mr. Vinoth Kumar V, Mr. Balajee | Back-Propagation Neural Networks, Content Retrieval, Machine Learning, Recursive Feature Elimination, Text Classification, Web Harvesting, Web Scraping. |
| **The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc)** | **The Goal (Objective) of this Solution & What is the problem that need to be solved** | **What are the components of it?** |
| Content extraction and Text classification effective web Scraping techniques. | This solution solves major realtime problems like , automated web scraping and classification of data. | • Web scraping<br>• Preprocessing<br>• Feature extraction<br>• Data classification<br>Accuracy score |
| **The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process** | | |

| | Process Steps | Advantage | Disadvantage (Limitation) |
|---|---|---|---|
| **1** | Web Scraping Techniques | With the help of this Web Scraping, the extraction of information is efficient and faster. Multiple pages can be loaded at the same time. | Automated web scraping is not possible when the website is protected with anti-scraping techniques. |
| **2** | Text Preprocessing | The data undergoes NLP operations such as tokenization, Stemming, and bag of words(BOW) which helps to remove the unwanted data and help in increasing the accuracy of the model. | While making use of the bag of words technique the order of the words or grammar may be inappropriate and will be changed |

| | | | |
|---|---|---|---|
| **3** | Feature Extraction | This step is performed on the altered data after the preprocessing step which undergoes the subset generation and learning model. | The extraction of the patterns may be sometimes not accurate due to the missing variables in it. |
| **4** | Logistic Regression – Recursive Feature Elimination (LR-RFE) | RFE has the advantage of considering both features' relevance, redundancy, and interactions. By recursively removing the least important features, RFE can effectively reduce the dimensionality of the dataset while preserving the most informative features. | Can be computationally expensive for large datasets.May not be the best approach for datasets with many correlated features.<br>May not work well with noisy or irrelevant features |

**Major Impact Factors in this Work**

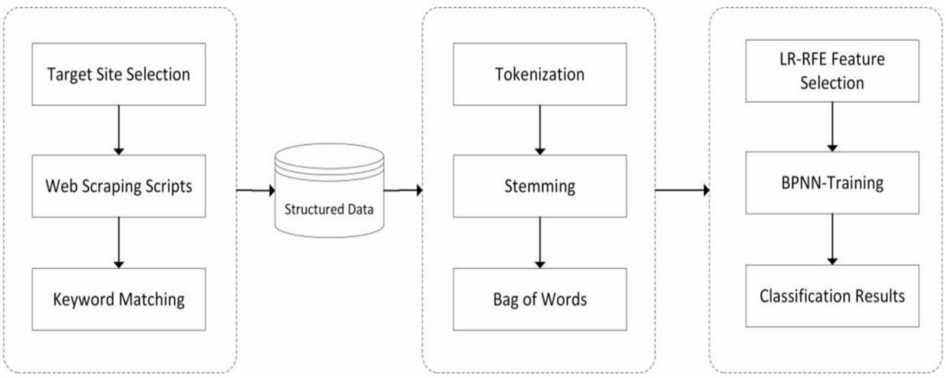
| Dependent Variable | Independent Variable | Moderating variable | Mediating (Intervening) variable |
|---|---|---|---|
| Accuracy Score | Website link, structured data, Machine learning algorithms | Web Scraping Tools like OpenRefine, cURL, Wget | This paper does not contain any mediating variable |

| Relationship Among the Above 4 Variables in This article |
|---|
| The independent variables acts as the key to achieve the accuracy score with the help of the web scraping tool that are acting as the Moderating variables.<br>The web scraping tools are the main source to extract the details from website likes(Independent variable) with meaningful execution and helps to archive the accuracy. |

| Input and Output | Feature of This Solution | Contribution in this Work |
|---|---|---|
| <table><tr><th>Input</th><th>Output</th></tr><tr><td>Product.</td><td>to view and compare product details from different websites</td></tr></table> | This solution addresses the time-consuming and manual effort required to visit websites. | This implementation aims to streamline the process of comparing product details from different websites on a single platform, providing users with a convenient and efficient way to make informed decisions. |

| Positive Impact of this Solution in This Project Domain | Negative Impact of this Solution in This Project Domain |
|---|---|
| This proposed model gives a robust classification model that uses Nlp and machine learning techniques and improves accuracy along with Personalized content extraction with effective text classification | This proposed model does not showcase any negative impacts in the project. The only thing that can be described as a negative impact. |

| Analyse This Work By Critical Thinking | The Tools That Assessed this Work | What is the Structure of this Paper |
|---|---|---|
| the solution improves the efficiency and convenience of comparing products from different websites on a single platform. | None | Abstract<br>I. Introduction<br>II. Background<br>III. Scraping Techniques<br>IV. Materials and Methods<br>V. Results<br>VI. Conclusion |

| Diagram/Flowchart |
|---|
|  |

| 17 | | |
|---|---|---|
| **Reference in APA format** | Tri Astuti, Irnawati Pratika, 2019, Product Review Sentiment Analysis by Artificial Neural Network Algorithm | |
| **URL of the Reference** | **Authors Names and Emails** | **URL of the Reference** |
| https://ijiis.org/index.php/IJIIS/article/vie w/15/14 | Tri Astuti, Irnawati Pratika | https://ijiis.org/index.php/IJIIS/article/vie w/15/14 |
| **The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc )** | **The Goal (Objective) of this Solution & What is the problem that need to be solved** | **The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc )** |
| An Artificial Neural Network (ANN) algorithm to perform sentiment analysis on product reviews. | To understand the difference between Artificial Neural Networks (ANNs) and other machine learning algorithms in the context of sentiment analysis | An Artificial Neural Network (ANN) algorithm to perform sentiment analysis on product reviews. |
| **Major Impact Factors in this Work** | | |

| **Dependent Variable** | **Independent Variable** | **Moderating variable** | **Mediating (Intervening ) variable** |
|---|---|---|---|
| Accuracy Score | Data collection, pre-processing, testing performance | Size of data. | Algorithm used. |

| **Relationship Among the Above 4 Variables in This article** |
|---|
| These mediating variables could include aspects of the algorithm's architecture, pre-processing steps, or other factors that influence how the algorithm interprets and classifies the sentiment in the reviews.<br>The sentiment of reviews depends on the application of the ANN algorithm, as the algorithm's output classifies reviews as positive or negative based on patterns it has learned. |

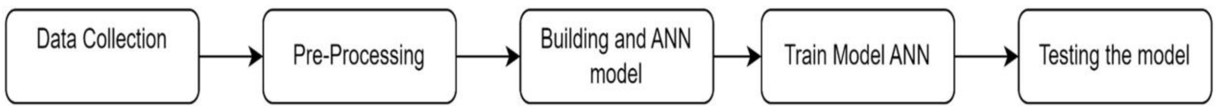
| Input and Output | Feature of This Solution | Contribution in this Work |
|---|---|---|
| **Input:** Numerical vector, which is obtained from the pre-processing of the products.<br>**Output:** The probability of the review being positive, and negative. | The Artificial Neural Network (ANN) algorithm itself is the key component. | This paper talks about how we can use artificial neural networks (ANNs) to perform sentiment analysis in product reviews and do it very accurately. |

| Input | Output |
|---|---|
| Numerical vector, which is obtained from the pre-processing of the products. | The probability of the review being positive, and negative. |

| Positive Impact of this Solution in This Project Domain | Negative Impact of this Solution in This Project Domain |
|---|---|
| It appears that using ANN algorithms is a good strategy, and as artificial intelligence develops more, we might find ways to further boost the precision and efficiency of our sentiment analysis system. | ANN models can be complex, requiring significant computational resources. This complexity can lead to longer processing times and increased hardware and energy costs. |

| Analyse This Work By Critical Thinking | The Tools That Assessed this Work | What is the Structure of this Paper |
|---|---|---|
| ANNs are often considered "black-box" models, meaning it can be challenging to understand how the model arrives at its decisions. This can be a drawback when trying to explain and justify the results to someone. | None | Abstract<br>I. Introduction<br>II. Research Concept<br>III. Results and Discussion<br>VI. Conclusions and Suggestions |

**Diagram/Flowchart**




Fig,3 overall working of the model

| 18 | | |
|---|---|---|
| **Reference in APA format** | Beresneva Daria, 2011, Computer-generated Text Detection Using Machine Learning: A Systematic Review | |
| **URL of the Reference** | **Authors Names and Emails** | **Keywords in this Reference** |
| https://www.researchgate.net/publication/304020905_Computer-Generated_Text_Detection_Using_Machine_Learning_A_Systematic_Review | Beresneva Daria | Artificial content, Generated text, Fake content detection. |
| **The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc )** | **The Goal (Objective) of this Solution & What is the problem that need to be solved** | **What are the components of it?** |
| A detailed examination of methods that can tell if a text is made by a person or created by a machine. | To thoroughly understand various methods and apply benchmark standards to distinguish between text produced by computers and humans. | This paper discusses various methods for detecting artificially generated texts on the internet. It explores techniques such as hidden style similarity, frequency counting, linguistic features, and machine learning algorithms. The paper also highlights the challenges and limitations in the field of artificial text detection. |
| **The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process** | | |

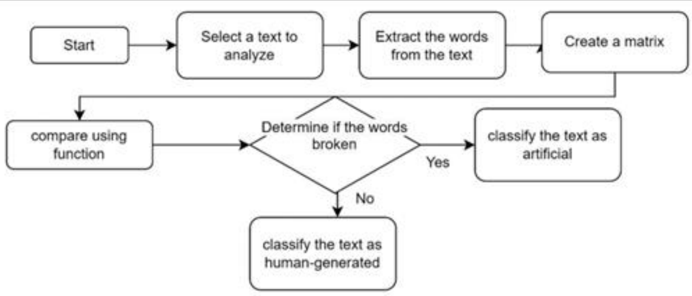
| | Process Steps | Advantage | Disadvantage (Limitation) |
|---|---|---|---|
| **1** | The linguistic features method : is a machine translation detection approach that uses both statistical and linguistic characteristics of the text. It involves the use of classifiers to distinguish between human reference translations and machine translations. | This approach is fully automated and independent of source language, target language, and domain, making it versatile and effective in identifying machine translations. | -- |
| **2** | Artificial content detection using lexicographic features:This method uses | It considers the text's linguistic features, including grammatical words, sentence structure, and word length. | -- |

| | | | |
|---|---|---|---|
| | various linguistic characteristics of the text, such as word and sentence length, grammatical words ratio, dictionary word ratio, vocabulary richness, and more. By training a decision tree using these features, the method can accurately distinguish between human-generated and machine-generated texts. | This can aid in differentiating writing produced by machines and humans. | |

**Major Impact Factors in this Work**

| Dependent Variable | Independent Variable | Moderating variable | Mediating (Intervening ) variable |
|---|---|---|---|
| Accuracy or effectiveness of artificial content detection methods | In the "Frequency counting method", one independent variable is "Cor," which represents the correlation of neighboring words in the text.<br>In the "Linguistic features method" , the independent variables include linguistic and perplexity features extracted from text. | Data Source ,Sample Size | Combination of Features |

| Relationship Among The Above 4 Variables in This article |
|---|
| The paper discusses different methods for artificial content detection, with each method having its own set of independent variables. While moderating and mediating variables are not explicitly mentioned, they are likely to exist in the research process, influencing how the independent variables impact the effectiveness of artificial content detection. The specific moderating and mediating variables would need to be identified and studied in more detail in a research context. |

<table>
<tr><th colspan="2">Input and Output</th><th colspan="2">Feature of This Solution</th><th colspan="2">Contribution & The Value of This Work</th></tr>
<tr><td colspan="2"><table><tr><th>Input</th><th>Output</th></tr><tr><td>Yelp's dataset</td><td>The performance of the fake review detection methods.</td></tr></table></td><td colspan="2">The paper discusses various methods and techniques for detecting artificially generated or fake texts. It covers topics such as scoring penalties for not respecting relationships between words, hidden style similarity measures, clustering algorithms, and linguistic and statistical features for detection. It provides insights into the effectiveness of different methods and their limitations.</td><td colspan="2">The paper provides valuable insights into the detection of artificially generated texts. The frequency counting method and the method of linguistic features offer effective approaches for identifying such content. The numerical results and evaluation of these methods demonstrate their potential</td></tr>
<tr><th colspan="3">Positive Impact of this Solution in This Project Domain</th><th colspan="3">Negative Impact of this Solution in This Project Domain</th></tr>
<tr><td colspan="3">The solutions presented in the document has a positive impact in the project domain by providing accurate, effective, and automated methods for recognizing artificially created text in multiple languages.</td><td colspan="3">--</td></tr>
<tr><th colspan="2">Analyse This Work By Critical Thinking</th><th colspan="2">The Tools That Assessed this Work</th><th colspan="2">What is the Structure of this Paper</th></tr>
<tr><td colspan="2">The paper shows the importance of considering different strategies and data types when detecting artificial text. The evaluation of the frequency counting method shows promising results in detecting automatically generated texts.</td><td colspan="2">--</td><td colspan="2">Abstract<br>I. Introduction<br>II. Literature Search<br>III. The methods of artificial text detection<br>IV. Choosing A Method<br>VI. Conclusion</td></tr>
<tr><th colspan="6">Diagram/Flowchart</th></tr>
<tr><td colspan="6">
</td></tr>
</table>

| 19 | | |
|---|---|---|
| **Reference in APA format** | Raheesa Safrin, K.R.Sharmila, T.S.Shri Subangi, E.A.Vimal, 2017, SENTIMENT ANALYSIS ON ONLINE PRODUCT REVIEW | |
| **URL of the Reference** | **Authors Names and Emails** | **Keywords in this Reference** |
| https://www.irjet.net/archives/V4/i4/IRJET-V4I4598.pdf | Raheesa Safrin, K.R.Sharmila, T.S.Shri Subangi, E.A.Vimal | Sentiment analysis, negation phrase identification, product reviews. |
| **The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ...etc )** | **The Goal (Objective) of this Solution & Whatis the problem that need to be solved** | **What are the components of it?** |
| • Sentiment analysis<br>• Data Collection<br>• Pre-processing and NLP<br>• Feature Labeling.<br>• K-means cluster | To understand K-means clustering along withpart-of-speech tagging to analyze the sentiments in product reviews. | The paper examines today's most advanced techniques for sentiment analysis and presents a new system that involves creating a website, getting user input, and using K-means clustering and part-of-speech tagging to determine the sentiment of the reviews. |
| **The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process** | | |

| | Process Steps | Advantage | **Disadvantage (Limitation)** |
|---|---|---|---|
| **1** | Building a website and getting user input | Simplifies the process of collecting data. | -- |
| **2** | The parts of speech tagging. | By labeling words with their respective partsof speech (nouns, verbs, adjectives, etc.), | limited ability to recognize and interpret sarcasm. |
| **3** | The k-mean clustering. | K-means clustering is relatively simple and computationally efficient, making it suitable for large datasets and providing a quick way togroup similar sentiments in sentiment analysis. | K-means requires specifying the number ofclusters (K) beforehand, which can be challenging, and it is sensitive to the initial placement of centroids. |

| Major Impact Factors in this Work | | | |
|---|---|---|---|
| **Dependent Variable** | **Independent Variable** | **Moderating variable** | **Mediating (Intervening ) variable** |
| Recall AccuracyPrecision | Text data size | Number of clusters in K-meansclustering | Pre-processing steps |

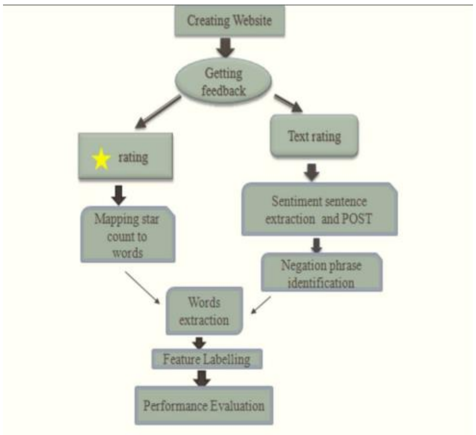
| Relationship Among The Above 4 Variables in This article |
|---|
| The link between the independent variable (text data) and the dependent variable (review classification) is influenced by various mediating variables that are associated with the preprocessing of the data. Furthermore, the number of clusters in the K-means clustering may also moderate the reviews' categorization and, thus, affect the final classification of the reviews. |

| Input and Output | | Feature of This Solution | Contribution in This Work |
|---|---|---|---|
| **Input** | **Output** | | |
| Text data collected through the website. | Classification of reviews based on sentiment analysis | A dedicated website created to collect user reviews on a certain product. | Creating a website for data collection is a valuable idea, and incorporating various techniques to enhance overall accuracy is also commendable. |

| Positive Impact of this Solution in This Project Domain | Negative Impact of this Solution in This Project Domain |
|---|---|
| Using more than one technique for classification may provide more accurate results. | None. |

| Analyse This Work By Critical Thinking | The Tools That Assessed this Work | What is the Structure of this Paper |
|---|---|---|
| The combination of various methods, like sentiment analysis and point-of-sale tagging, improves the accuracy of the client sentiment classification process. This improved accuracy provides more accurate insights into the opinions of customers regarding the products, which is important in the e-commerce industry. | None | Abstract<br>• Introduction<br>• Related Work<br>• Proposed Method<br>• Implementation<br>• Conclusion<br>• Performance Evaluation<br>• References |

**Diagrams/Flowcharts**




**20**

| **Reference in APA format** | Wenyuan Zhao, 2020,Classification of Customer Reviews on Ecommerce Platforms Based on Naive Bayesian Algorithm and Support Vector Machine | |
|---|---|---|
| **URL of the Reference** | **Authors Names and Emails** | **Keywords in this Reference** |
| https://iopscience.iop.org/article/10.1088/1742-6596/1678/1/012081/pdf | Wenyuan Zhao | Machine learning, evaluation metrics, Classifiers. |
| **The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ...etc )** | **The Goal (Objective) of this Solution & What is the problem that need to be solved** | **What are the components of it?** |
| • Text Pre-processing<br>• Word Segmentation.<br>• Scikit-learn Library.<br>• Naive Bayes Classifier (NBC)<br>• Support Vector Machine (SVM)<br>• Python. | The objective is to assess the effectiveness ofvarious classification models in categorizing reviews into multiple groups, including positive, negative, and neutral, using both training and testing data. | The author employed the Naïve Bayes Classifier and Support Vector Machine, which are supervised learning techniques commonly utilized for classification tasks. These two algorithms were compared in terms of various metrics to evaluate their performance. |

| The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process | | | |
|---|---|---|---|
| | **Process Steps** | **Advantage** | **Disadvantage (Limitation)** |
| **1** | The primary emphasis of the analysis is placedon reviews related to Chinese e-commerce. | is often determined by a combination of phrases, rather than individual words, which isdistinct from English, where each word has its own meaning. | segmentation is not precise, it can lead to difficulties in accurately extracting and interpreting the context, potentially impacting the performance of text analysis and classification tasks. This is in contrast to English, where individual words typically have more distinct meanings, making segmentation less critical. |
| **2** | The Support Vector Machine. | SVM demonstrates superior performance in terms of recall rate and accuracy, making it a strong choice for tasks where precision and completeness in classification are critical.<br>SVM is particularly convenient when the classification task involves separating data into two classes, making it a favorable choicefor such scenarios. | SVM can be computationally intensive, especially when dealing with a large number of features or categories. This can lead to longer training times and resource requirements.<br>SVM is sensitive to the scale of input features. It often requires feature scaling,and improper scaling can impact its performance.<br>SVM may struggle with noisy or overlappingdata. In such cases, it may lead to suboptimal results. |
| **3** | The Naïve Bayes Classifier. | NBC excels in terms of classification speed, which is advantageous when handling a largenumber of reviews or when strict accuracy requirements are not a priority.<br>NBC is more practical for multi-class classification tasks, making it suitable for | NBC assumes that attributes are independent of each other. This independence assumption can lead tosuboptimal results when dealing with correlated features.<br>NBC may not capture complex relationships in |

| | | | |
|---|---|---|---|
| | | scenarios where the data needs to be categorized into multiple categories. | data as effectively as other. |

**Major Impact Factors in this Work**

| Dependent Variable | Independent Variable | Moderating variable | Mediating (Intervening ) variable |
|---|---|---|---|
| classification accuracy | Data Size, Text Preprocessing, Segmentation Ratio. | segmentation ratio. | text preprocessing. |

**Relationship Among The Above 4 Variables in This article**

The independent variables include the machine learning algorithms used (Naive Bayesian and Support Vector Machines), the data size, and the textpreprocessing methods.

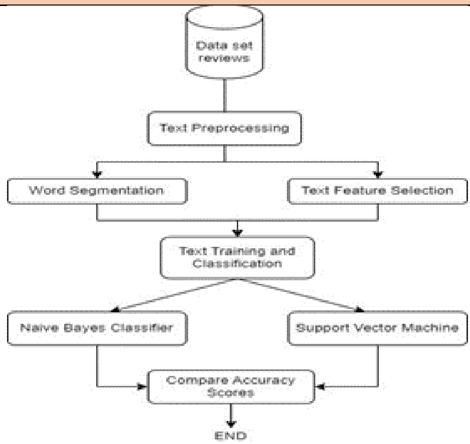
A potential mediating variable in this context could be the "text preprocessing," as it serves to prepare the raw review data, removing punctuation, special symbols, and meaningless vocabulary. This preprocessing step may influence the quality of the data used for training and, subsequently, theaccuracy of the classification (dependent variable).

| Input and Output | | Feature of This Solution | Contribution & The Value of This Work |
|---|---|---|---|
| **Input** | **Output** | Offering a customization feature that enables users to select reviews associated with specificsentiments. | From the paper, we've gained knowledge about SVM and NBC, as well as their pros and cons. Whatshould be the optimal size for the training and testing datasets to achieve the highest efficiency. |
| Reviews of products primarily from the Chinese e-commerce platform, particularly Alibaba. | Categorizing these reviews into two groups: positive and negative. | | |

| Positive Impact of this Solution in This Project Domain | Negative Impact of this Solution in This Project Domain |
|---|---|
| Machine learning algorithms are big channelings in the current researchand eyeing this area makes sense win right direction. | Since this is an assessment of the performance of different algorithms, there isn't anything to be concerned about because everything is predefined. |

| Analyse This Work By Critical Thinking | The Tools That Assessed this Work | What is the Structure of this Paper |
|---|---|---|
| This paper has truly enhanced our understandingof how the size of data plays a crucial role in influencing the performance of different algorithms. | Stack overflow. | Abstract<br>I. Introduction<br>II. Method<br>III. Experiment<br>IV. Results and Discussion<br>V. Conclusion and References |

| Diagram/Flowchart |
|---|



## 2.2 COMPARISION TABLE

| Author | Year | Approach | Description |
|---|---|---|---|
| Beresne va Daria | 2011 | Linguistic features method | A detailed examination of methods that can tell if a text is made by a person or created by a machine. The frequency counting method and the method of linguistic features offer effective approaches for identifying such content. |
| Eka Dyar Wahyuni,Arif Djunaidy | 2016 | Iterative Computation Framework (ICF) | The criteria used by human evaluators and the application of the ICF and ICF++ models for identifying fraudulent reviews are all covered in this paper. |
| Raheesa Safrin, K.R. Sharmila T.S.Shri Subangi, E.A. Vimal | 2017 | Feature Labelling, K-means clusters | The author examines advanced sentiment analysis methods, evaluating review sentiments with part-of-speech tagging. |
| Mr. Navjyots inh Jadeja, Chirag visani, | 2017 | Naïve Bayes, SVM | In this paper a comparative analysis of both the algorithms is performed based on the classification of the reviews. The author finds that Naïve Bayes is more efficient than SVM. |
| Tanjim Ul Haque, Nudrat Nawal Saber | 2018 | Pool Based Active learning | The author has used a Pool-based active learning strategy on the raw data to make the input data more accurate to improve the performance of the model. |
| Mr. Karthike yan T, Mr. Karthik Sekaran, | 2018 | Web Scraping Techniques with LR-RFE (Recursive Feature Eliminati-on ) | Website data scraped, stored, preprocessed, feature selection with LR-RFE, categorized using Back Propagation into positive, negative, neutral reviews. |

| | | | |
|---|---|---|---|
| Nidhi A. Patel | 2018 | Machine Learning Techniques | The objective of this paper is to examine several methods and strategies for identifying fraudulent evaluations. It emphasizes the various features and classifiers used for fake review detection, as well as machine learning-based techniques. |
| Tri Astuti, Irnawati Pratika | 2019 | ANN, The conjugate Scale Gradient Method | This author developed an ANN model and Conjugate method to perform sentiment analysis on product reviews with high accuracy. |
| Andre Sihombi -ng A.C.M Fong | 2019 | Gaussain Naïve Bayes and XGBoost | Study explores Logistic Regression, SVM, Naive Bayes, and XGBoost for fake review detection, assessing effectiveness and applications. |
| Aljoha ah Almjawe , Sahar Bayoumi, Dalal Alshehr Soroor | 2019 | Interactive Packed bubbles, Linear chart, Stacked bars, and Word-cloud | This document discusses the use of visualization techniques in analyzing and summarizing reviews. |
| Yin Shuqin, Feng Jing | 2019 | MPINPUL(Mixing Population and Individual Nature PU Learning) | Proposed PU learning model identifies fake reviews using text, behavior, and relationship features for accurate recognition. |
| Wenyua n Zhao | 2020 | Naive Bayes Classifier, SVM | This author evaluates how well different classification models classify reviews into three groups: Positive, Negative, and Neutral. |
| Rakibul Hassan, Md.Rabiul Islam | 2020 | Supervised Machine Learning techniques | This Document introduces supervised machine learning to identify fake online reviews using TF-IDF, Empath, and sentiment features |
| Prof. P.S.Gaik wad, Kaushal Parmar, Rohit Yadav, Datta Supekar | 2021 | Web Scraping | This approach takes the product name as input, scraps the details from different sources, and displays the information on the user's window. |

| | | | |
|---|---|---|---|
| Mayuri patil, snehal Nikumb h | 2021 | Spam Detection, NLP | This paper shows Calculated TF-IDF values, multiplied frequency and IDF matrices, implemented system controller for genuine reviews post spam detection. |
| Uma Mahesh wari, Dr.S.S. Dhenaka ran | 2021 | Fuzzy logic, Deep Learning, and user-defined classification | Classified reviews with Fuzzy logic, user-defined classification, compared accuracy with regular DL and ML algorithms |
| Elshirf elmurngi | 2021 | Sentiment Analysis | Paper stresses textual reviews' importance in reputation models, advocating their consideration for a comprehensive understanding, surpassing mere ratings. |
| Pansy Nandwa ni, Rupali Verma | 2021 | Review Based Approach | Author used lexicon-based, ML, DL, and Transfer Learning for sentiment analysis and emotion detection in text, employing multiple stages. |
| Ahmed M.Elmo gy, Usman Tariq | 2021 | K-Nearest Neighbour | This paper says that the content of the reviews is processed using natural language processing to extract key features like Sentiment analysis and linguistic patterns. |
| N Deshai, B Bhaskar a Rao | 2023 | CNN- LSTM | Author used lexicon-based, ML, DL, and Transfer Learning for sentiment analysis and emotion detection in text, employing multiple stages. |

## 2.3 WORK EVALUATION TABLE

| Author Name and Year | Work Goal | System's Components | System's Mechanism | Features /Characteristics | Advantages | Limitations /Disadvantages | Results |
|---|---|---|---|---|---|---|---|
| Nidhi A. Patel | The objective of the document is to discuss various techniques and approaches used in detecting fake reviews. It focuses on machine learning-based methods and the different features and | The document outlines the steps involved in the machine learning approach for fake review detection, including data collection, data pre-processing, feature extraction. | Machine Learning Techniques<br>Supervised learning techniques<br>Semi-supervised learning techniques<br>Unsupervised learning techniques | It covers linguistic and textual features, behavioural features, and relational features. | Review detection enhances the trust by removing the fake reviews, trustworthiness can be significantly improved and it also helps in better decision making, fair business practices. | Review detection can have some negative impact on this domain, which includes false positives which can harm the reputation and trust of the customers. | The performance of the fake review detection methods. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | classifiers used for fake review detection. | | | | | | |
| Rakibul Hassan, Md. Rabiul Islam | The main objective of this document is to introduce a method using supervised machine learning to identify fake online reviews. It explores features like TF-IDF, Empath, and sentiment | Content based features.<br>Train-validation set split with a ratio of 75:25 to obtain the train set and validation set.<br>Identification of genre, detecting psycholinguistic behaviour, and categorization of text as features | Supervised Machine Learning techniques | Content based features.<br>Train-validation set split with a ratio of 75:25 to obtain the train set and validation set.<br>Identification of genre, detecting psycholinguistic behaviour, and categorization of text as features. | The use of both content-based and user-behaviour based features can improve the accuracy of fake review detection. | The solution is a supervised learning approach, which relies heavily on the availability of labelled data for training. This could be a limitation in scenarios where labelled data is scarce or expensive to obtain. | A classifier that can predict the label of a new review based on its features. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | polarity to create a model that can accurately distinguish between fake and honest reviews. | | | | | | |
| Andre Sihombing, A.C.M. Fong | The main objective of this document is study of different machine learning classification techniques/ models such as | The workflow in the document has been classified into three parts: data preprocessing, feature engineering and the classification process. In which the components like under- | Logistic Regression, Support Vector Machine, Gaussian Naive Bayes, and XGBoost. | The document has also considered the length of the reviews and found that fake reviews tend to be shorter than genuine ones. | By utilizing the classification techniques mentioned in the document, the research achieved a high F-1 score of 0.9 in prediction, indicating the effectiv | As the document depends upon the labelled data there could be a potential bias in the classification process. If the algorithm contains potential biases, it may filter reviews poorly based on certain characteristics or demographics. | The performance of the fake review detection methods. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Logistic Regression, Support Vector Machine, Gaussian Naive Bayes, and XGBoost for fake review detection. This document provides the insights regarding the effectiveness of classification method. | sampling and over-sampling were used for better preprocessing of data. | | | eness of the approach in identifying fake reviews. | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Yin Shuqin, Feng Jing | The goal of the proposed solution is to develop a PU learning model for the recognition of fake reviews. It utilizes various features such as text, behaviour, and relationship characteristics to accurately identify fake | The document mainly focused on three major categories for classifying the reviews: text features, behavioural characteristics of reviewers, and relationship characteristics. | This paper suggests the use of multiple features in the MPINPUL (Mixing Population and Individual Nature PU Learning) a model for classifying fake reviews. | The proposed solution focuses on integrating text, behavioural, and relationship features to build a classification model for fake reviews recognition | The document has showed that the classification model trained on fusion features, which integrate text and behaviour characteristics, is about 10% more accurate than models trained solely on text features. | The negative impact of proposed solution include misclassification, manipulation and false insights. | Accuracy of MPINPUL model. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | reviews. | | | | | | |
| Aljoharah Almjawel, Sahar Bayoumi, Dalal Alshehri, Soroor Alzahrani, Munirah Alotaibi | This document discusses the use of visualization techniques in analysing and summarizing reviews. | Mobile user, Mobile network operator Blockchain Instant messanger Encryption/decryption Backup and restoration. | Visualization techniques, Sentiment analysis, Tableau and R to provide interactive visualizations. | Interactive Packed bubbles, Linear chart, Stacked bars, and Word-cloud | The suggested system is to give book reviews a visual format so that users can examine reviews from customers more efficiently. | As the suggested solution depends on a sentiment analysis system that might have flaws or limitations, it might not accurately convey the genuine opinions of the customers. | Visual representation of the review |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| S. Uma Maheswari, Dr.S.S. Dhenakarn, June 2021 | The main aim of this work is to help customers identify fake reviews on social media and websites based on selected features for better decisions on product purchases online and is method of classification categorizing | The author employed four different techniques, User-defined Fuzzy Logic, deep learning, and Machine Learning for Sentiment classification and prediction based on the accuracy and f1 score. Furthermore, classification is done according to the score as positive, | We have seen the different algorithms performing different types of analysis with the same dataset and attributes. Every algorithm follows its own approach for processing the given data but differs in the performance analysis. This paper has | It offers a customization on feature that enables users to classify the reviews as Fake and Genuine. The solution also relates to the work of classification of customer reviews based on different categories like positive, negative, and neutral | The usage of different methods to achieve higher accuracy like Fuzzy logic, User defined classification shows a positive impact on this project with an increase in accuracy. | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | reviews into different categories which include positive, negative, and neutral, and accuracy is been compared with existing methods of ML | negative, neutral, Positively fake, negatively fake, and so on. | proven that performance always differs from the algorithms used with similar data. | through sentiment analysis. | | | |
| Mr. Navjyotsinh, Chirag visani, 2017 | To analyze Amazon product reviews and predict the ratings of | This paper discusses various methods for detecting artificially generated texts on the internet. It | By experimenting with various machine learning algorithms, this work contributes to the field of detecting | The paper discusses various methods and techniques for detecting spam reviews using machine learning. | The solutions are more helpful to improve the accuracy with the help of Svm or other ml | There are no such negative impacts on the paper, One such thing is the Limited resources and privacy Concerns. | The paper showsthe importance of considering different strategies and data types when detecting artificial text. The context mainly |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | future reviews It also includes the process of extracting the meaningful narratives | explores techniques such as hidden style similarity, frequency counting, linguistic features, and machine learning algorithms. The paper also highlights the challenges and limitations in the field of artificial text detection | spam reviews. The authors are particularly interested in leveraging Twitter as a platform for sentiment analysis . | Its analyses and detects the review patterns. Using word bigram features and syntactic components for accurate detection | algorithms and effective detection can also be experience ed | | discusses the use of different machine-learning techniques. |
| Vinoth Kumar V, Mr.Balajee | This solution solves major realtime problems like , automated web scrapi | Web scraping • Preprocessing •Feature extraction • Data classification Accuracy score | This implementatio n aims to streamline the process of comparing product details from different websites | This solution addresses the time-consuming and manual effort required to visit websites . | This proposed model gives a robust classification model that uses Nlp and machine learnin | This proposed model does not showcase any negative impacts on the project. The only thing that can be described as a negative impact is the consuming and manual effort required to visit websites. | This solution improves the efficiency and convenience of comparing products from different websites on |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | ng and classification of data. | | on a single platform, providing users with a convenient and efficient way to make informed decisions. | | g techniques and improves accuracy along with Personalized content extraction with effective text classification | | a single platform. |
| Tanjim Ul Haque, 2018 | The main goal of this solution is to calculate the accuracy score of polarizing the reviews | This paper includes problem identification, data collection, pre-processing, building a Semi-Supervised | This paper talks about how we can use artificial neural networks (ANNs) to perform sentiment analysis in product reviews and doit | Pool Based Learning along with Machine Learning algorithms like Linear SVM, Naïve Bayes, Stochastic Gradient Descent, | The high accuracy achieved by the model in classifying the sentiments can help the customers to get a better user | One of the negative impacts of this solution is that while dealing with active learning and feature extraction techniques, manual labeling is also required to achieve accurate results. | The Author has addressed the limitations and challenges of labeling the raw data and preprocessing the data due to the presence of limited standard datasets from e- |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | into positive and negative using different Amazon product datasets based on some Machine learning algorithms | machine learning model comparing with different ML algorithms, and presenting the results and discussion | very accurately | Random Forest | experience in finding the right things and also gain insights into customer opinions and preferences. | | | commerce sites |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Pansy N andwani, Rupali Verma, 2021 | The main objective of this paper is to compare the existing techniques for both emotion and sentiment detection and find out the best-performing algorithm with an accuracy score | This paper includes problem identification, data collection, pre-processing, Feature extraction Model Development, Model assessment, Challenges, and conclusion based on the weights of the terms included | This paper talks about how we can make use of the resources to analyze the text based on the sentiment and emotion pattern. It also provide the accuracy score | This solution mainly focuses on sentiment analysis and emotion detection from text which discusse sthe sentiments and emotions in given input text and addresses the challenges faced in it | Using the proposed method for sentiment analysis will provide more accuracy in detecting the sentiments and emotions in text, we may find opportunities esto further improve the accuracy and effectiveness of our sentiment | The author used different techniques to access the best accurate model which increases the computational costs. | This paper provides insights into the levels of sentiment analysis, methodologies, challenges, and applications of sentiment and emotion analysis fromtext data. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | analysis system. | | |
| Elshirf elmurngi. Abdelouahed gherbi. | To classify movie reviews as real reviews or fake reviews using sa algorithms with supervised learning techniques. | Sentiment classification, Feature selection, detection process. | Sentiment analysis. | Different sentiment classification algorithms, is effective for classifying movie reviews as real or fake. | Improved accuracy the experiments conducted in this project have shown that sentiment classification algorithms, particularly svm. | System proposed in the paper employs the system's ip address, potentially giving rise to security concerns for users. | The solution presents various sentiment classification algorithms and methodologies, there are potential limitations and challenges that could impact the accuracy |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Eka dyar wahyuni . Arif djunaidy . | To detect fake reviews for a product by using the text and rating property .from a review. | Proposed system (icf++), the methodology used for fake review detection, the evaluation strategy, and the results. | Iterative computation framework (icf). | Paper suggests that incorporating semantic aspects, such as sentiment polarity, can improve the accuracy of fake review detection. | This analysis will determine the review's integrity, the reviewer's credibility, and the product's dependability. | Specific equations for calculating the trustworthiness and honesty values are not provided in the given document content. | The specific equations for calculating the trustworthiness and honesty values are not provided in the given document content. |
| Mayuri patil , snehal nikumbh . | This method identifies fraudulent transactions by assessing user behaviour and netwo | Architecture input selection , spam detection, spammed content analysis. | Fraud risk management system and removal model | Opinion Spam wherein spammers engage in the creation of fake, misleading, or dishonest reviews with the intent of enhancing their product's | With its efficient spam and fake review detection capabilities, significantly enhan | Paper employs the system's IP address, potentially giving rise to security concerns for users. | This analysis ultimately focus on user confidence when making purchases on e-commerce websites or applications. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | rk activity, | | | reputation for financial gain, while also undermining their competitors' products. | ces the productivity of the company. | | |
| N deshani. B bhaskaraao. | Detect fake reviews and what is the main difference between them. Secondary goal is to detect fake rating | Data pre-processing, classifiers, model performance. | Hybrid deep learning methods (cnn-lstm) (rnn-lstm) | Cnn-lstm for detecting fake online reviews, and lstm-rnn for detecting fake ratings in the e-commerce domain. | The identification of counterfeit reviews and ratings, delivering precise outcomes. | It requires large amount of training data to predict efficiently. | RNN methods offer efficiency and practicality, potentially making them more suitable for achieving optimal outcomes and maximizing the efficacy of detecting fake online reviews. |

# CHAPTER 3

# PROPOSED SYSTEM

## 3.1 PROPOSED SYSTEM

The proposed system revolutionizes the consumer experience by providing users with robust insights essential for making informed purchasing decisions.  It streamlines the decision-making process by effectively evaluating large amounts of reviews, making it easier to choose the products that best fit unique needs and preferences. By classifying reviews into three sentiment categories Positive, Negative, and Neutral the system increases customer trust in the reliability and validity of online reviews. This classification scheme effectively reduces doubt and doubt, allowing consumers to have faith in the feedback of real customers. In addition, the system uses visualization techniques and thorough analyses of review data to prioritize user convenience and comprehension. Users can quickly extract pertinent information from the large amount of data that is available by using clear and natural insights, which improves their overall experience and facilitates decision-making.

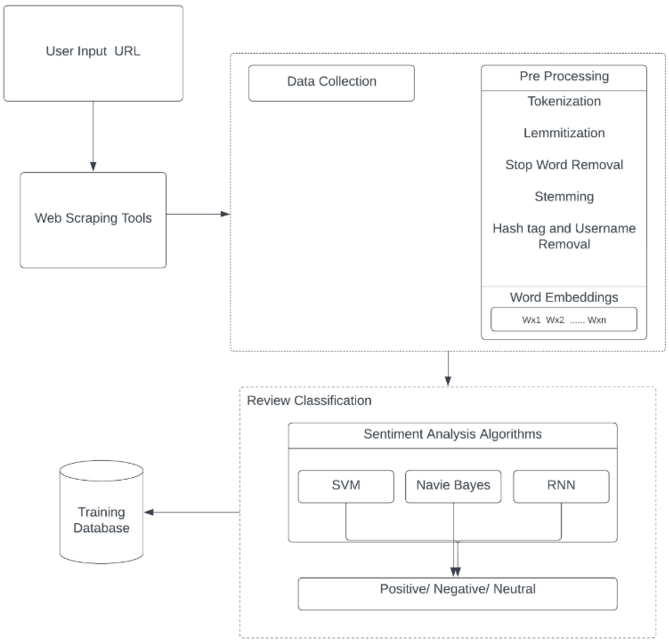



Figure 2: Flowchart outlining proposed solution overview

## 3.2 ADVANTAGES OF PROPOSED SYSTEM

The proposed system has the following advantages:

- Reduces consumer review analysis time by 50%, allowing for quicker, more informed purchase decisions.
- Simplifies review assessment, empowering informed purchasing decisions based on sentiment analysis.
- Provides reliable sentiment analysis that goes beyond surface-level keywords by leveraging powerful Natural Language Processing (NLP) methodologies and machine learning algorithms.
- Enhances user experience through comprehensive analysis and visualization techniques.

## 3.3 SYSTEM REQUIREMENTS

The system requirements for the development and deployment of the project as an application are specified in this section. These requirements are not be confused with the end-user system requirements. There are no specific, end-user requirements as the intended application is cross-platform and is supposed to work on devices of all form-factors and configurations.

### 3.3.1 SOFTWARE REQUIREMENTS

Below are the software requirements for application development:

1. Operating System: Windows 10 or macOS Catalina (or later versions).
2. Python: Version 3.6 or higher.
3. Anaconda: Python distribution for package management and environment setup.
4. Integrated Development Environment (IDE): Visual Studio Code for coding HTML, CSS, and JavaScript.
5. Flask: Python web framework for developing the backend of the application.
6. Web Browser: Google Chrome, Mozilla Firefox, Microsoft Edge, or Brave Browser with extension support for testing and running the web application.
7. Graphics Library: Matplotlib or Plotly for generating bar graphs for visualization.

8. Text Analysis Libraries: NLTK (Natural Language Toolkit), SpaCy, or TextBlob for sentiment analysis and fake review detection.
9. Machine Learning Framework: TensorFlow or scikit-learn for building and training AI models.
10. Database: SQLite or MySQL for storing review data and classification results.
11. Version Control: Git for managing source code and collaboration with team members.

### 3.3.2 HARDWARE REQUIREMENTS

Hardware requirements for application development are as follows:

1. CPU- intel i5 or higher
2. RAM – 8 GB or higher
3. Storage - adequate storage to store model checkpoints, etc. A solid SSD is recommended.

### 3.3.3 IMPLEMENTATION TECHNOLOGIES

**Web Scraping Module (Beautiful Soup):**

Beautiful Soup, a Python library, facilitates web scraping by parsing HTML and XML documents. It aids in extracting structured data from web pages, including product reviews, by navigating the HTML tree and retrieving relevant information such as review text, ratings, and timestamps. With its powerful features, Beautiful Soup streamlines the process of collecting data from online sources, enabling efficient analysis and segmentation of product reviews.

**NLP (Natural Language Processing):**

Natural Language Processing (NLP) techniques enable the processing and analysis of textual data. It encompasses various tasks like tokenization, part-of-speech tagging, and named entity recognition. In the context of product reviews, NLP assists in extracting meaningful insights, identifying sentiment, and detecting key phrases or topics. By leveraging NLP, the system can interpret and categorize reviews accurately, facilitating informed decision-making for businesses and consumers alike.

**TF-IDF (Term Frequency-Inverse Document Frequency):**

TF-IDF (Term Frequency-Inverse Document Frequency) is a statistical measure used for feature extraction in text analysis. It calculates the importance of a word in a document relative to a corpus of documents. By considering both the frequency of a term in a document and its rarity across the corpus, TF-IDF identifies significant terms that characterize the content of each review. This technique aids in capturing the essence of reviews and distinguishing relevant features for analysis.

$$\mathrm{TF}(t,d) = \frac{\text{Number of times term } t \text{ appears in document } d}{\text{Number of terms in document } d}$$

$$\mathrm{IDF}(t,D) = \log\left(\frac{\text{Number of documents containing term } t}{\text{Total number of documents in the corpus } D}\right)$$

$$\text{TF-IDF}(t,d,D) = \mathrm{TF}(t,d) \times \mathrm{IDF}(t,D)$$

Figure 3: TF-IDF formula

**SVM (Support Vector Machine):**

Support Vector Machine (SVM) is a supervised machine learning algorithm employed for classification tasks. It learns to classify data points by finding the hyperplane that best separates them into different classes. In the context of product review analysis, SVM can be trained to classify reviews into categories such as positive, negative, or neutral sentiments, as well as distinguishing between genuine and fake reviews

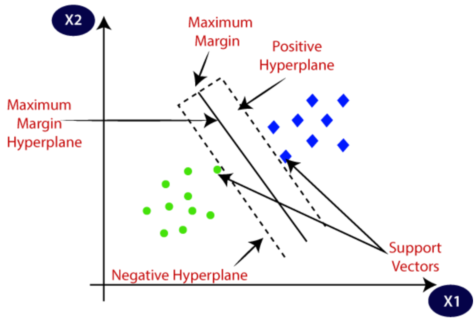



Figure 4: Support vector machine hyperplane

**Sentiment Analysis:**

Sentiment Analysis is a subfield of Natural Language Processing focused on determining the sentiment expressed in text. It employs various techniques to classify text into positive, negative, or neutral sentiments based on the conveyed emotions or opinions. In the context of product reviews, sentiment analysis helps businesses understand customer satisfaction levels, identify areas for improvement, and make data-driven decisions. By gauging sentiment, businesses can effectively respond to customer feedback and enhance overall user experience.

**Visualization:**

Visualization techniques play a crucial role in presenting analysis results in a visually interpretable format. Bar graphs, pie charts, and word clouds are commonly used visualization methods in product review analysis. They provide intuitive representations of sentiment distributions, review frequencies, and significant terms extracted from the data. Visualization enhances comprehension and insight generation, enabling stakeholders to grasp key trends, patterns, and insights derived from the analysis effectively.

# CHAPTER 4

# SYSTEM DESIGN

## 4.1 PROPOSED SYSTEM ARCHITECTURE

The proposed system, "Enhancing Trust with AI: Product Review Analysis and Segregation System," aims to develop a comprehensive web application named Classer. Classer will be a versatile platform accessible across various devices and operating systems. Leveraging advanced AI algorithms, Classer will analyze product reviews, categorize them as positive, negative, or neutral, and identify fake reviews generated by bots or computers. The system will enhance trust by providing users with reliable insights into product feedback, promoting informed decision-making.

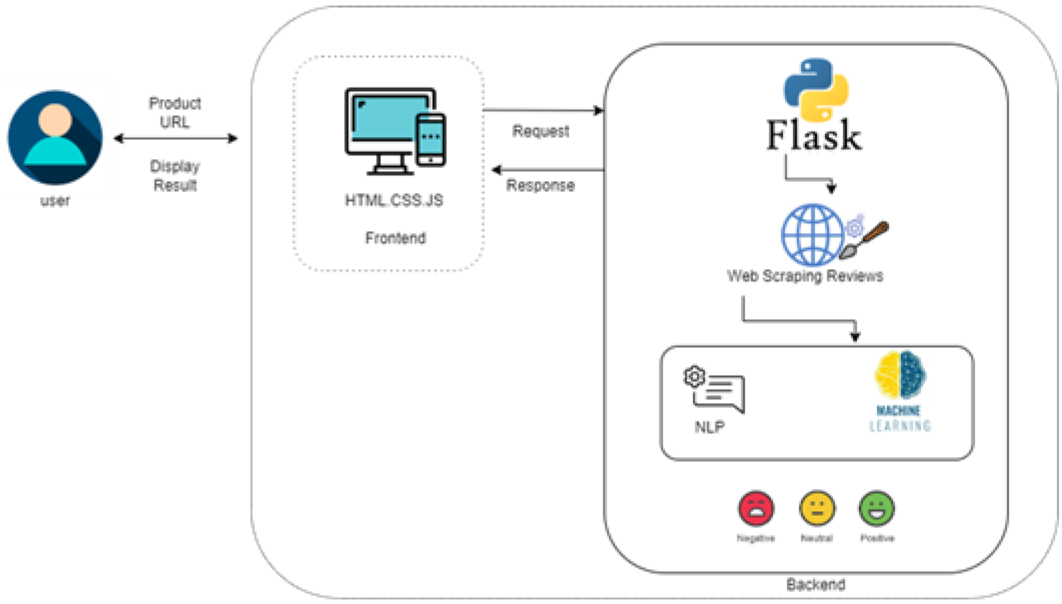



Figure 5: Proposed Architecture

## 4.2 APPLICATION MODULES

The application on an overall involves four main modules, which cater to the four main functions of this application, i.e., to classify and identify fake reviews.

### 4.2.1 Web Scraping Module:

**Purpose**: The Web Scraping Module is responsible for extracting reviews and relevant information from online sources based on the provided.

**Functionality:** Sends HTTP requests to the product URL to fetch the HTML content of the product page. Parses the HTML content to identify and extract review data, including review text, ratings, and timestamps. Utilizes URL manipulation techniques to navigate to specific review sections or pages within the website.

**Implementation Details:** Utilizes web scraping libraries like BeautifulSoup to parse HTML content. Implements robust error handling and retry mechanisms to handle potential network issues or website changes

### 4.2.2 Sentiment Analysis Module:

**Purpose:** The Sentiment Analysis Module analyzes the sentiment of the collected reviews to determine whether they are positive, negative, or neutral.

**Functionality:** Applies sentiment analysis techniques to classify review text into sentiment categories based on the emotions expressed. Utilizes machine learning algorithms such as Support Vector Machines (SVM) or lexicon-based approaches like VADER (Valence Aware Dictionary and sentiment Reasoner). Calculates sentiment scores or probabilities for each review, indicating the degree of positivity or negativity.


Figure 6: Labeling reviews according to their sentiments

**Implementation Details:** Trains machine learning models on labeled review datasets to classify sentiment accurately. Incorporates pre-trained sentiment analysis models like NLTK (Natural Language Toolkit) for efficient sentiment analysis. Considers context and language nuances to improve the accuracy of sentiment classification, especially for informal or domain-specific text.

### 4.2.3 Fake Review Detection Module:

**Purpose:** The Fake Review Detection Module identifies and flags potential fake or fraudulent reviews among the collected data.

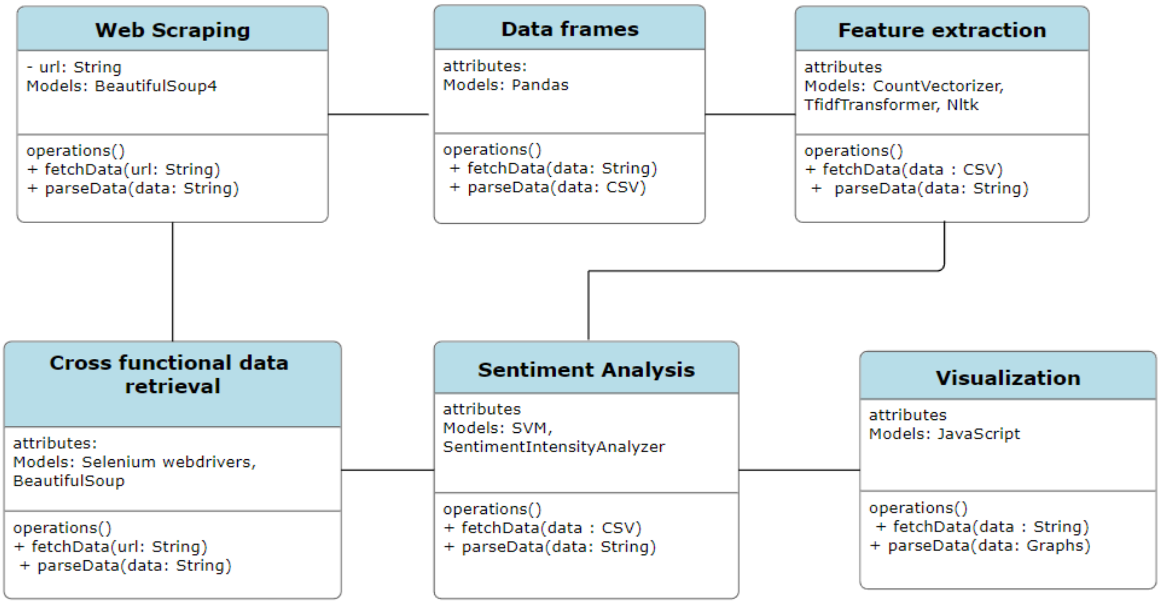



Figure 7: Proposed System UML Chart

**Functionality**: Analyzes review characteristics such as review length, language patterns, and abnormal review behavior to detect anomalies indicative of fake reviews. Utilizes machine learning or statistical techniques to identify patterns associated with fake reviews, such as repetitive language or unusual review timing. Integrates with external databases or services to cross-reference review data with known sources of fake reviews.

**Implementation Details:** Implements robust outlier detection algorithms to identify suspicious review patterns or outliers in the dataset. Utilizes domain-specific heuristics or rules to flag reviews that exhibit characteristics commonly associated with fake or deceptive behavior. Regularly updates detection algorithms and rulesets to adapt to evolving tactics used by fake review perpetrators.

### 4.2.4 Data Visualization Module:

**Purpose:** The Data Visualization Module presents the processed review data and sentiment analysis results in a visually appealing and informative manner.

**Functionality:** Generates interactive charts, graphs, or visualizations to summarize review sentiments, trends, and distributions. Provides filtering and drill-down capabilities to explore review data based on various criteria such as product categories or time periods. Integrates with frontend UI components to display dynamic and responsive visualizations that enhance user understanding and engagement.

**Implementation Details:** Utilizes data visualization libraries like Matplotlib, Plotly to create rich and interactive visualizations. Customizes visualization styles and layouts to align with the application's branding and user experience guidelines. Implements data caching and optimization techniques to ensure smooth rendering and responsiveness, especially for large datasets or complex visualizations.


Figure 8: Visualizing data patterns through pie charts

## 4.3 UML Diagrams

UML stands for Unified Modelling Language. UML is a standardized fashionable-cause modelling language in the subject of object-oriented software engineering. In its modern shape, UML comprises of two essential components: a Meta-model and a notation. The Unified Modelling Language is a standard language for specifying, Visualization, Constructing and documenting the artifacts of software program machine, in addition to for commercial enterprise modelling and other non-software systems. The UML uses more often than not graphical notations to express the design of software program projects.

### 4.3.1 Use Case Diagram

In the Unified Modeling Language (UML), a use case diagram is a behavioral diagram that stems from use-case analysis. Its number one objective is to provide a visual summary of a gadget's capability, showcasing actors, their objectives (portrayed as use cases), and any relationships amongst those use cases. The fundamental aim of a use case diagram is to demonstrate which device capabilities are accomplished for each actor worried, while additionally illustrating the jobs played via these actors within the gadget.

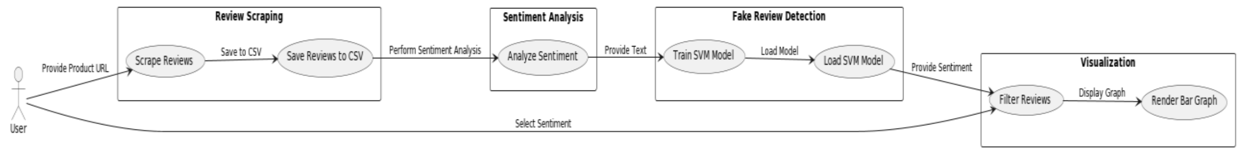



Figure 9: Use case Diagram

### 4.3.2 Class Diagram

In software engineering, a class diagram within the Unified Modeling Language (UML) is a static shape diagram that delineates the architecture of a machine. It achieves this by using illustrating the training within the gadget, inclusive of their attributes, operations (or techniques), and the connections between those classes. This diagram elucidates the distribution of statistics among lessons and clarifies which elegance is responsible for housing unique records.

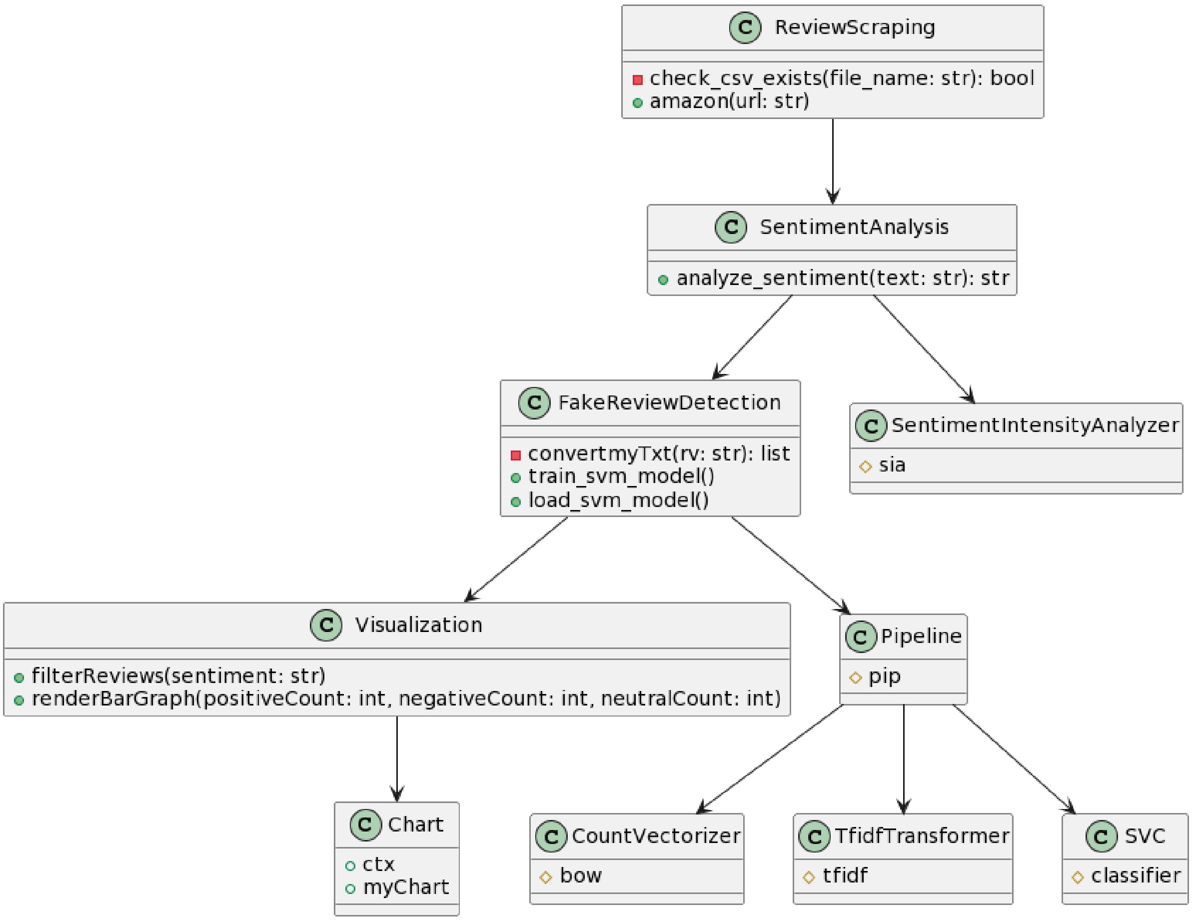



Figure 10: Class Diagram

### 4.3.3 Sequence Diagram

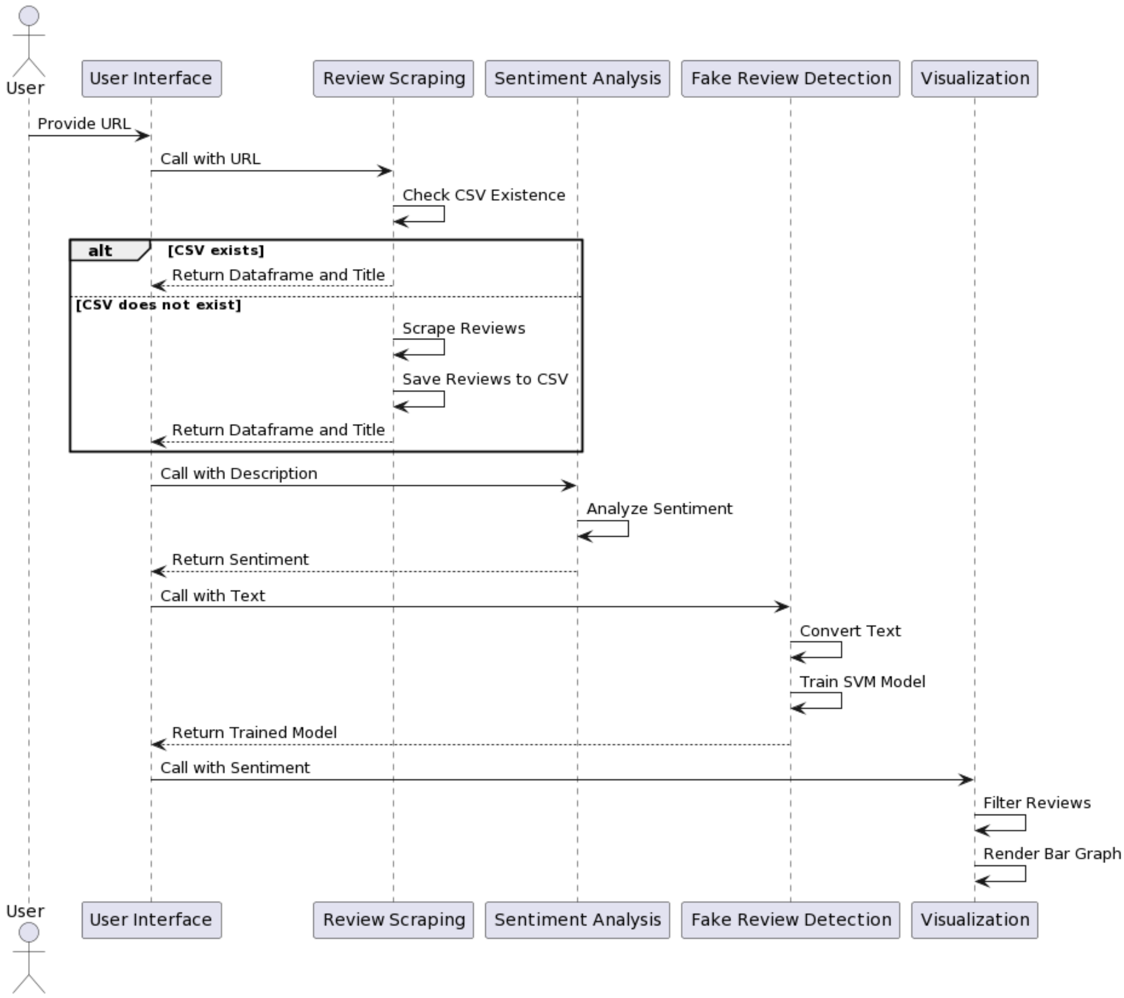



Figure 11: Sequence Diagram

### 4.3.4 Activity Diagram

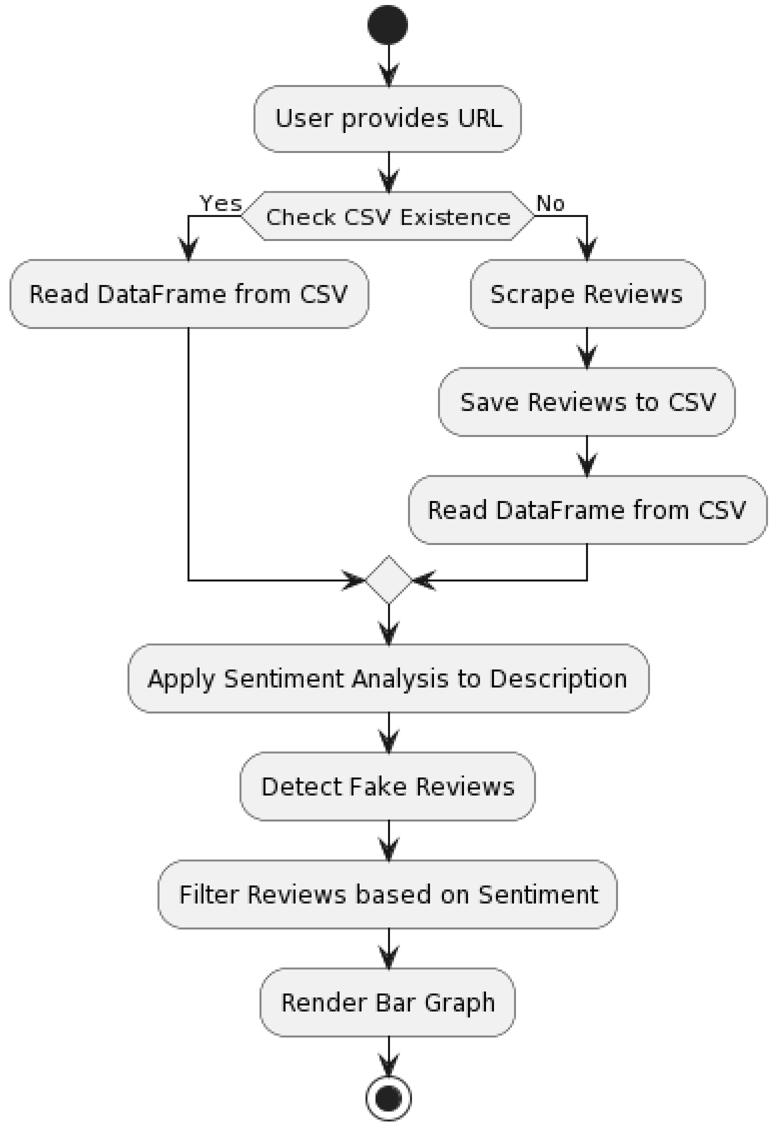



Figure 12: Activity Diagram

# CHAPTER 5

# IMPLEMENTATION

## 5.1 BRIEF EXPLANATION OF IMPLEMENTATION

In implementing the project, a multifaceted approach is adopted, combining various technologies and methodologies to achieve the desired functionalities. The project's backbone lies in Python, a versatile and widely-used programming language well-suited for tasks ranging from web scraping to machine learning. Python's rich ecosystem of libraries and frameworks, including BeautifulSoup, Scrapy, NLTK, and scikit-learn, facilitates efficient data extraction, sentiment analysis, and fake review detection.

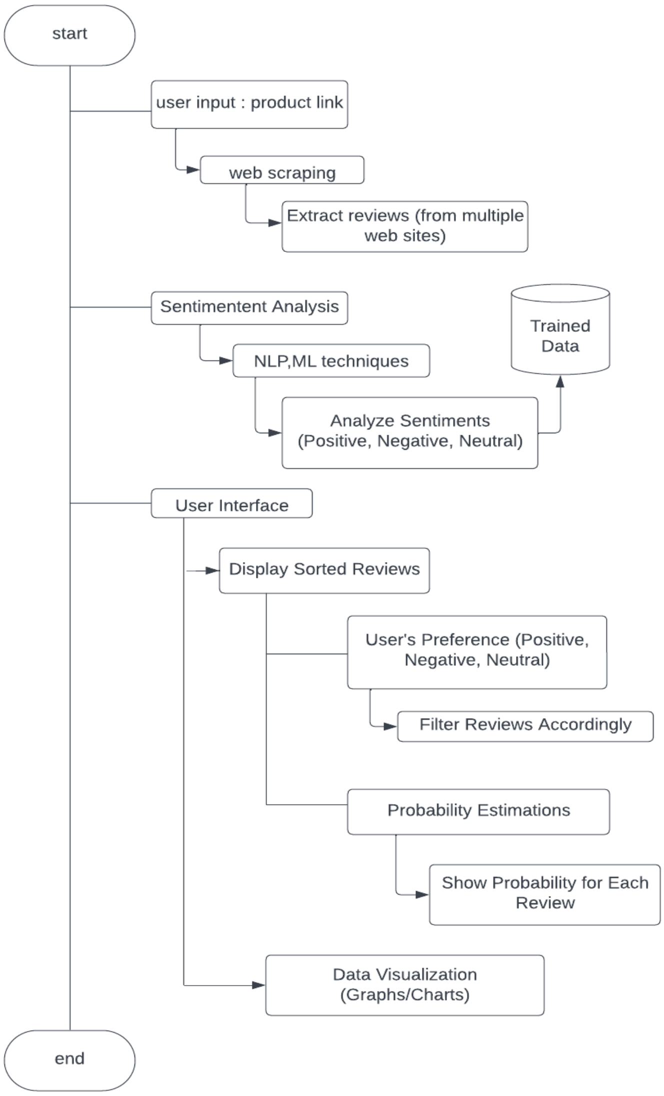



Figure 13: Detailed Implementation flowchart

The project's implementation involves meticulous attention to detail in each module's development. For instance, in the Web Scraping Module, careful consideration is given to handling HTTP requests, parsing HTML content, and navigating product pages to extract review data accurately and efficiently. Similarly, in the Sentiment Analysis Module, rigorous testing and evaluation of machine learning models ensure reliable sentiment classification, leveraging techniques like SVM trained on labelled datasets to achieve high accuracy.

Furthermore, the project's success hinges on effective integration and coordination between modules. Seamless communication between the Web Scraping, Sentiment Analysis, and Fake Review Detection Modules is essential for the smooth flow of data and accurate analysis results. This integration is achieved through well-defined interfaces and standardized data formats, enabling seamless data exchange and collaboration between different components of the system. Overall, the project's implementation reflects a meticulous and systematic approach, leveraging Python's capabilities and fostering synergy between diverse technologies to deliver robust and reliable functionality for analysing product reviews.

Finding real reviews among the millions of evaluations is difficult. Due to a dearth of reliable data that has been identified as authentic or fraudulent for training purposes, a significant number of duplicate and almost duplicate reviews were found in the instance of 4.8 million reviews and 1.14 million reviewers on Amazon.com. Because each user has unique strategies and habits it might be difficult to discern between bot-generated and human-generated content when examining posted evaluations. Most false review detections are done manually. The majority of false review detections are made by hand.

## 5.2 SOURCE CODE

**index.html**

```
<!doctype html>
<html lang="en">
 <head>
  <meta charset="utf-8">
  <meta name="viewport" content="width=device-width, initial-scale=1">
  <meta name="description" content="">
  <meta name="author" content="">
  <title>Classer</title>
  <!-- CSS FILES -->
  <link rel="preconnect" href="https://fonts.googleapis.com">
  <link rel="preconnect" href="https://fonts.gstatic.com" crossorigin>
  <link
href="https://fonts.googleapis.com/css2?family=Montserrat:wght@500;600
;700&family=Open+Sans&display=swap" rel="stylesheet">
  <link href="{{ url_for('static', filename='css/bootstrap.min.css')
}}" rel="stylesheet">
  <link href="{{ url_for('static', filename='css/bootstrap-icons.css')
}}" rel="stylesheet">
  <link href="{{ url_for('static', filename='css/templatemo-topic-
listing.css') }}" rel="stylesheet">
 </head>
 <body id="top">
  <main>
   <nav class="navbar navbar-expand-lg">
    <div class="container">
     <a class="navbar-brand" href="index.html">
      <img src="static\images\classer_crop.png" width=220px
height=60px>
     </a>
     <div class="d-lg-none ms-auto me-4">
      <a href="#top" class="navbar-icon bi-person smoothscroll"></a>
     </div>
     <button class="navbar-toggler" type="button" data-bs-
toggle="collapse" data-bs-target="#navbarNav" aria-
```

```
controls="navbarNav" aria-expanded="false" aria-label="Toggle
navigation">
      <span class="navbar-toggler-icon"></span>
     </button>
     <div class="collapse navbar-collapse" id="navbarNav">
      <ul class="navbar-nav ms-lg-5 me-lg-auto">
       <li class="nav-item">
        <a class="nav-link click-scroll" href="#section_1">Home</a>
       </li>
       <li class="nav-item">
        <a class="nav-link click-scroll" href="#section_2">About</a>
       </li>
       <li class="nav-item">
        <a class="nav-link click-scroll" href="#section_3">How it
works</a>
       </li>
       <li class="nav-item">
        <a class="nav-link click-scroll" href="#section_4">FAQs</a>
       </li>
      </ul>
      <div class="d-none d-lg-block">
       <a href="#top" class="navbar-icon bi-person smoothscroll"></a>
      </div>
     </div>
    </div>
   </nav>
   <section class="hero-section d-flex justify-content-center align-
items-center" id="section_1">
    <div class="container">
     <div class="row">
      <div class="col-lg-8 col-12 mx-auto">
       <h1 class="text-white text-center">Review. Analyze. Decide</h1>
       <h6 class="text-center">platform for your convenient
shopping</h6>
       <form method="post" action="new_page" class="custom-form mt-4
pt-2 mb-lg-0 mb-5" role="search" id="analyzeForm">
        <!-- Your existing form elements -->
        <div class="input-group input-group-lg">
         <span class="input-group-text bi-search"
id="basicaddon1"></span>
```

```
          <input name="keyword" type="search" class="form-control"
id="keyword" placeholder="Paste your URL here ..." aria-
label="Analyze">
          <button type="submit" class="form control">Analyze</button>
         </div>
        </form>
       </div>
      </div>
     </div>
    </section>
    <section class="featured-section">
     <div class="container">
      <div class="row justify-content-center">
       <div class="col-lg-4 col-12 mb-4 mb-lg-0">
        <div class="custom-block bg-white shadow-lg">
         <div class="d-flex">
          <div>
           <h5 class="mb-2">Classer!</h5>
           <p class="mb-0">Welcome to our innovative cross-functional
website, where we bring together the best of Amazon and Flipkart
reviews under one roof. </p>
          </div>
         </div>
         <img src="static\images\image1.png" height=355px width=355px>
         </a>
        </div>
       </div>
       <div class="col-lg-6 col-12">
        <div class="custom-block custom-block-overlay">
         <div class="d-flex flex-column h-100">
          <img src="static\images\bg_img.png" class="custom-block-image
img-fluid" alt="">
          <div class="custom-block-overlay-text d-flex">
           <div>
            <h5 class="text-white mb-2">Your Trusted Guide</h5>
            <p class="text-white">With our dedication to transparency
and accuracy, we aspire to be more than just a website. We aim to be
your steadfast companion, guiding you towards products that align with
your needs and preferences. </p>
            <a href="#section_3" class="btn custom-btn mt-2 mt-lg-
3">Learn More</a>
           </div>
```

```
          </div>
          <div class="social-share d-flex">
           <p class="text-white me-4">Share:</p>
           <ul class="social-icon">
            <li class="social-icon-item">
             <a href="#" class="social-icon-link bi-twitter"></a>
            </li>
            <li class="social-icon-item">
             <a href="#" class="social-icon-link bi-facebook"></a>
            </li>
            <li class="social-icon-item">
             <a href="#" class="social-icon-link bi-pinterest"></a>
            </li>
           </ul>
           <a href="#" class="custom-icon bi-bookmark ms-auto"></a>
          </div>
          <div class="section-overlay"></div>
         </div>
        </div>
       </div>
      </div>
     </div>
    </section>
    <section class="explore-section section-padding" id="section_2">
     <div class="container">
      <div class="row">
       <div class="col-12 text-center">
        <h2 class="mb-4">About </h1>
       </div>
      </div>
     </div>
     <div class="container">
      <div class="row">
       <div class="col-12">
        <div class="tab-content" id="myTabContent">
         <div class="tab-pane fade show active" id="design-tab-pane"
role="tabpanel" aria-labelledby="design-tab" tabindex="0">
          <div class="row">
           <div class="col-lg-4 col-md-6 col-12 mb-4 mb-lg-0">
            <div class="custom-block bg-white shadow-lg">
             <a href="/topics_detail">
              <div class="d-flex">
```

```
              <div>
               <h5 class="mb-2">Review</h5>
               <p class="mb-0">Filter products by review
legitimacy</p>
              </div>
             </div>
             <img
src="static\images\topics\undraw_Remote_design_team_re_urdx.png"
class="custom-block-image img-fluid" alt="">
            </a>
           </div>
          </div>
          <div class="col-lg-4 col-md-6 col-12 mb-4 mb-lg-0">
           <div class="custom-block bg-white shadow-lg">
            <a href="/topics_detail">
             <div class="d-flex">
              <div>
               <h5 class="mb-2">Analyze</h5>
               <p class="mb-0">Find out in seconds if product reviews
are reliable, with our detailed review report</p>
              </div>
             </div>
             <img
src="static\images\topics\undraw_Redesign_feedback_re_jvm0.png"
class="custom-block-image img-fluid" alt="">
            </a>
           </div>
          </div>
          <div class="col-lg-4 col-md-6 col-12">
           <div class="custom-block bg-white shadow-lg">
            <a href="/topics_detail">
             <div class="d-flex">
              <div>
               <h5 class="mb-2">Decide</h5>
               <p class="mb-0">Experience seamless shopping with
us</p>
              </div>
             </div>
             <img src="static\images\topics\colleagues-working-cozy-
office-medium-shot.png" class="custom-block-image img-fluid" alt="">
            </a>
           </div>
```

```
          </div>
         </div>
        </div>
   </section>
   <section class="timeline-section section-padding" id="section_3">
    <div class="section-overlay"></div>
    <div class="container">
     <div class="row">
      <div class="col-12 text-center">
       <h2 class="text-white mb-4">How does it work? </h1>
      </div>
      <div class="col-lg-10 col-12 mx-auto">
       <div class="timeline-container">
        <ul class="vertical-scrollable-timeline" id="vertical-
scrollable-timeline">
         <div class="list-progress">
          <div class="inner"></div>
         </div>
         <li>
          <h4 class="text-white mb-3">Paste link to analyze</h4>
          <p class="text-white">Begin by pasting the link of the
product you wish to evaluate. Our website will swiftly process the
provided URL to extract and scrutinize the associated reviews.</p>
          <div class="icon-holder">
           <i class="bi-search"></i>
          </div>
         </li>
         <li>
          <h4 class="text-white mb-3">Get a review report</h4>
          <p class="text-white">Our system meticulously analyzes the
reviews associated with the product. Within moments, you'll receive a
comprehensive report highlighting key insights and trends extracted
from the reviews.</p>
          <div class="icon-holder">
           <i class="bi-filter-circle-fill"></i>
          </div>
         </li>
         <li>
          <h4 class="text-white mb-3">Visualize & Understand</h4>
          <p class="text-white">Dive into the report with ease as it's
presented with intuitive visualizations. Our clear and concise
graphics make it simple to grasp the analysis findings, allowing you
```

```
to discern genuine feedback from potentially misleading or fake
reviews.</p>
          <div class="icon-holder">
           <i class="bi-pie-chart-fill"></i>
          </div>
         </li>
        </ul>
       </div>
      </div>
      <div class="col-12 text-center mt-5">
       <p class="text-white"> Want to learn more? <a href="#"
class="btn custom-btn custom-border-btn ms-3">Check out Youtube</a>
       </p>
      </div>
     </div>
    </div>
   </section>
   <section class="faq-section section-padding" id="section_4">
    <div class="container">
     <div class="row">
      <div class="col-lg-6 col-12">
       <h2 class="mb-4">Frequently Asked Questions</h2>
      </div>
      <div class="clearfix"></div>
      <div class="col-lg-5 col-12">
       <img src="static\images\faq_graphic.jpg" class="img-fluid"
alt="FAQs">
      </div>
      <div class="col-lg-6 col-12 m-auto">
       <div class="accordion" id="accordionExample">
        <div class="accordion-item">
         <h2 class="accordion-header" id="headingOne">
          <button class="accordion-button" type="button" data-bs-
toggle="collapse" data-bs-target="#collapseOne" aria-expanded="true"
aria-controls="collapseOne"> What is Classer? </button>
         </h2>
         <div id="collapseOne" class="accordion-collapse collapse
show" aria-labelledby="headingOne" data-bs-parent="#accordionExample">
          <div class="accordion-body"> Welcome to Classer, your go-to
platform for discerning genuine product feedback from deceptive or
fake reviews. At Classer, we understand the challenges consumers face
when navigating the vast landscape of online reviews. That's why we've
```

```
developed an innovative solution that leverages advanced algorithms to classify and filter reviews with unparalleled accuracy. </div>
         </div>
        </div>
        <div class="accordion-item">
         <h2 class="accordion-header" id="headingTwo">
          <button class="accordion-button collapsed" type="button" data-bs-toggle="collapse" data-bs-target="#collapseTwo" aria-expanded="false" aria-controls="collapseTwo"> What types of products does Classer analyze? </button>
         </h2>
         <div id="collapseTwo" class="accordion-collapse collapse" aria-labelledby="headingTwo" data-bs-parent="#accordionExample">
          <div class="accordion-body"> Classer can analyze reviews for a wide range of products across different categories, including electronics, apparel, household goods, beauty products, and more. Our platform is flexible and can adapt to various product types. </div>
         </div>
        </div>
        <div class="accordion-item">
         <h2 class="accordion-header" id="headingThree">
          <button class="accordion-button collapsed" type="button" data-bs-toggle="collapse" data-bs-target="#collapseThree" aria-expanded="false" aria-controls="collapseThree"> Is my personal information secure when using Classer? </button>
         </h2>
         <div id="collapseThree" class="accordion-collapse collapse" aria-labelledby="headingThree" data-bs-parent="#accordionExample">
          <div class="accordion-body"> Yes, Classer takes user privacy and data security seriously. We adhere to strict privacy policies and employ robust security measures to safeguard user information. Your personal data is encrypted and protected according to industry best practices. </div>
         </div>
        </div>
       </div>
      </div>
     </div>
    </div>
   </section>
  </main>
  <footer class="site-footer section-padding">
```

```
   <div class="container">
    <div class="row">
     <div class="col-lg-3 col-12 mb-4 pb-2">
      <a class="navbar-brand mb-2" href="index.html">
       <img src="static\images\classer_blue.png" width=220px
height=60px>
      </a>
     </div>
     <div class="col-lg-3 col-md-4 col-6">
      <h6 class="site-footer-title mb-3">Resources</h6>
      <ul class="site-footer-links">
       <li class="site-footer-link-item">
        <a href="#" class="site-footer-link">Home</a>
       </li>
       <li class="site-footer-link-item">
        <a href="#" class="site-footer-link">How it works</a>
       </li>
       <li class="site-footer-link-item">
        <a href="#" class="site-footer-link">FAQs</a>
       </li>
       <li class="site-footer-link-item">
        <a href="#" class="site-footer-link">Contact</a>
       </li>
      </ul>
     </div>
     <div class="col-lg-3 col-md-4 col-6 mb-4 mb-lg-0">
      <h6 class="site-footer-title mb-3">Information</h6>
      <p class="text-white d-flex mb-1">
      <p>Major Project</p>
      </p>
      <p class="text-white d-flex">
      <p>CSM | Team-04</p>
      </p>
     </div>
     <div class="col-lg-3 col-md-4 col-12 mt-4 mt-lg-0 ms-auto">
      <p class="copyright-text mt-lg-5 mt-4">Copyright © CSM || Team-
04 || MLRIT.
     </div>
    </div>
   </div>
  </footer>
  <!-- JAVASCRIPT FILES -->
```

```
  <script src="{{ url_for('static', filename='js/jquery.min.js')
}}"></script>
  <script src="{{ url_for('static',
filename='js/bootstrap.bundle.min.js') }}"></script>
  <link rel="stylesheet" href="{{ url_for('static',
filename='css/bootstrap.min.css') }}?v=1.0">
  <script src="{{ url_for('static', filename='js/jquery.sticky.js')
}}"></script>
  <script src="{{ url_for('static', filename='js/click-scroll.js')
}}"></script>
  <script src="{{ url_for('static', filename='js/custom.js')
}}"></script>
 </body>

</html>
```

**newpage.html**

```
<!DOCTYPE html>
<html lang="en">
 <head>
  <meta charset="UTF-8">
  <meta name="viewport" content="width=device-width, initial-
scale=1.0">
  <title>Submit Page</title>
  <!-- CSS FILES -->
  <link rel="preconnect" href="https://fonts.googleapis.com">
  <link rel="preconnect" href="https://fonts.gstatic.com" crossorigin>
  <link
href="https://fonts.googleapis.com/css2?family=Montserrat:wght@500;600
;700&family=Open+Sans&display=swap" rel="stylesheet">
  <link href="{{ url_for('static', filename='css/submit-page.css') }}"
rel="stylesheet">
  <!-- ... other CSS files ... -->
  <style>
   table {
    border-collapse: collapse;
    width: 100%;
   }

   th,
   td {
    text-align: left;
    padding: 8px;
```

```
    }
   </style>
  </head>
  <body class="body">
   <nav class="navbar navbar-expand-lg">
    <div class="container">
     <a class="navbar-brand" href="/">
      <img src="static\images\classer_crop.png" width=220px
height=60px>
     </a>
    </div>
   </nav>
   <br>
   <div class="container_for_title">
    <div class="title">
     <h1>{{ Name }}</h1>
    </div>
   </div>
   <br>
   <div class="container_for_options">
    <div class="selection">
     <button class="button all"
onclick="filterReviews('All')">All</button>
     <button class="button positive"
onclick="filterReviews('Positive')">Positive</button>
     <button class="button negative"
onclick="filterReviews('Negative')">Negative</button>
     <button class="button neutral"
onclick="filterReviews('Neutral')">Neutral</button>
    </div>
   </div>
   </div>
   <div class="table">
    <table>
     <thead>
      <tr>
       <th>Stars</th>
       <th>Title</th>
       <th>Description</th>
       <th>label</th>
       <th>Sentiment</th>
      </tr>
```

```
    </thead>
    <tbody> {% for index, row in csv_data.iterrows() %} <tr
class="review-row" data-sentiment="{{ row['Sentiment'] }}">
      <td>{{ row['Stars'] }}</td>
      <td>{{ row['Title'] }}</td>
      <td>{{ row['Description'] }}</td>
      <td>{{ row['label'] }}</td>
      <td>{{ row['Sentiment'] }}</td>
     </tr> {% endfor %} </tbody>
   </table>
   <ul class="list">
    <li>OR - Original</li>
    <li>CG - Computer Generated</li>
   </ul>
  </div>
  <!-- Graphs Container -->
  <div class="container_for_graph">
   <div class="graphs">
    <h2>Graphs</h2>
    <canvas id="sentimentChart" width="500" height="300"></canvas>
   </div>
  </div>
  <!-- JAVASCRIPT FILES -->
  <script src="https://cdn.jsdelivr.net/npm/chart.js"></script>
  <script src="{{ url_for('static', filename='js/jquery.min.js')
}}"></script>
  <script src="{{ url_for('static',
filename='js/bootstrap.bundle.min.js') }}"></script>
  <link rel="stylesheet" href="{{ url_for('static',
filename='css/bootstrap.min.css') }}?v=1.0">
  <script src="{{ url_for('static', filename='js/jquery.sticky.js')
}}"></script>
  <script src="{{ url_for('static', filename='js/click-scroll.js')
}}"></script>
  <script src="{{ url_for('static', filename='js/custom.js')
}}"></script>
  <script src="{{ url_for('static', filename='js/script.js')
}}"></script>
 </body>
</html>
```

**topic-details.html**

```
<!doctype html>
```

```
<html lang="en">
 <head>
  <meta charset="utf-8">
  <meta name="viewport" content="width=device-width, initial-scale=1">
  <meta name="description" content="">
  <meta name="author" content="">
  <title>Classer</title>
  <!-- CSS FILES -->
  <link rel="preconnect" href="https://fonts.googleapis.com">
  <link rel="preconnect" href="https://fonts.gstatic.com" crossorigin>
  <link
href="https://fonts.googleapis.com/css2?family=Montserrat:wght@500;600
;700&family=Open+Sans&display=swap" rel="stylesheet">
  <link href="{{ url_for('static', filename='css/bootstrap.min.css')
}}" rel="stylesheet">
  <link href="{{ url_for('static', filename='css/bootstrap-icons.css')
}}" rel="stylesheet">
  <link href="{{ url_for('static', filename='css/templatemo-topic-
listing.css') }}" rel="stylesheet">
  <!--

TemplateMo 590 topic listing

https://classer.com

-->
 </head>
 <body id="top">
  <main>
   <nav class="navbar navbar-expand-lg">
    <div class="container">
     <a class="navbar-brand" href="index.html">
      <img src="static\images\classer_crop.png" width=220px
height=60px>
     </a>
     <div class="d-lg-none ms-auto me-4">
      <a href="#top" class="navbar-icon bi-person smoothscroll"></a>
     </div>
     <button class="navbar-toggler" type="button" data-bs-
toggle="collapse" data-bs-target="#navbarNav" aria-
controls="navbarNav" aria-expanded="false" aria-label="Toggle
navigation">
```

```
      <span class="navbar-toggler-icon"></span>
     </button>
     <div class="collapse navbar-collapse" id="navbarNav">
      <ul class="navbar-nav ms-lg-5 me-lg-auto">
       <li class="nav-item">
        <a class="nav-link click-scroll" href="#section_1">Home</a>
       </li>
       <li class="nav-item">
        <a class="nav-link click-scroll" href="#section_2">About</a>
       </li>
       <li class="nav-item">
        <a class="nav-link click-scroll" href="#section_3">How it
works</a>
       </li>
       <li class="nav-item">
        <a class="nav-link click-scroll" href="#section_4">FAQs</a>
       </li>
      </ul>
      <div class="d-none d-lg-block">
       <a href="#top" class="navbar-icon bi-person smoothscroll"></a>
      </div>
     </div>
    </div>
   </nav>
   <header class="site-header d-flex flex-column justify-content-
center align-items-center">
    <div class="container">
     <div class="row justify-content-center align-items-center">
      <div class="col-lg-5 col-12 mb-5">
       <h2 class="text-white">Get your review report here!!</h2>
       <div class="d-flex align-items-center mt-5">
        <a href="#topics-detail" class="btn custom-btn custom-border-
btn smoothscroll me-4">Read More</a>
        <a href="#top" class="custom-icon bi-bookmark
smoothscroll"></a>
       </div>
      </div>
      <div class="col-lg-5 col-12">
       <div class="topics-detail-block bg-white shadow-lg">
        <img
src="static/images/topics/undraw_Remote_design_team_re_urdx.png"
class="topics-detail-block-image img-fluid">
```

```
        </div>
       </div>
      </div>
     </div>
    </header>
    <section class="topics-detail-section section-padding" id="topics-
detail">
     <div class="container">
      <div class="row">
       <div class="col-lg-8 col-12 m-auto">
        <h3 class="mb-4">Introduction to Classer</h3>
        <p>Welcome to Classer, your go-to platform for discerning
genuine product feedback from deceptive or fake reviews. At Classer,
we understand the challenges consumers face when navigating the vast
landscape of online reviews. That's why we've developed an innovative
solution that leverages advanced algorithms to classify and filter
reviews with unparalleled accuracy.</p>
        <blockquote> "Reviews analyzed. Decisions made confidently."
</blockquote>
        <div class="row my-4">
         <div class="col-lg-6 col-md-6 col-12">
          <img src="static/images/businesswoman-using-tablet-
analysis.jpg" class="topics-detail-block-image img-fluid">
         </div>
         <div class="col-lg-6 col-md-6 col-12 mt-4 mt-lg-0 mt-md-0">
          <img src="static/images/colleagues-working-cozy-office-
medium-shot.jpg" class="topics-detail-block-image img-fluid">
         </div>
        </div>
        <p>Our website employs cutting-edge technology to analyze
review patterns, linguistic cues, and sentiment indicators, enabling
us to identify and categorize fake reviews with precision. </p>
       </div>
      </div>
     </div>
    </section>
    <section class="section-padding section-bg">
     <div class="container">
      <div class="row justify-content-center">
       <div class="col-lg-5 col-12">
        <img src="static/images/rear-view-young-college-student.jpg"
class="newsletter-image img-fluid" alt="">
```

```
      </div>
      <div class="col-lg-5 col-12 subscribe-form-wrap d-flex justify-
content-center align-items-center">
       <form class="custom-form subscribe-form" action="#"
method="post" role="form">
        <h4 class="mb-4 pb-2">Get Newsletter</h4>
        <input type="email" name="subscribe-email" id="subscribe-
email" pattern="[^ @]*@[^ @]*" class="form-control" placeholder="Email
Address" required="">
        <div class="col-lg-12 col-12">
         <button type="submit" class="form-control">Subscribe</button>
        </div>
       </form>
      </div>
     </div>
    </div>
   </section>
  </main>
  <footer class="site-footer section-padding">
   <div class="container">
    <div class="row">
     <div class="col-lg-3 col-12 mb-4 pb-2">
      <a class="navbar-brand mb-2" href="index.html">
       <img src="static/images/classer_blue.png" width=220px
height=60px>
      </a>
     </div>
     <div class="col-lg-3 col-md-4 col-6">
      <h6 class="site-footer-title mb-3">Resources</h6>
      <ul class="site-footer-links">
       <li class="site-footer-link-item">
        <a href="#" class="site-footer-link">Home</a>
       </li>
       <li class="site-footer-link-item">
        <a href="#" class="site-footer-link">How it works</a>
       </li>
       <li class="site-footer-link-item">
        <a href="#" class="site-footer-link">FAQs</a>
       </li>
       <li class="site-footer-link-item">
        <a href="#" class="site-footer-link">Contact</a>
       </li>
```

```
        </ul>
      </div>
      <div class="col-lg-3 col-md-4 col-6 mb-4 mb-lg-0">
        <h6 class="site-footer-title mb-3">Information</h6>
        <p class="text-white d-flex mb-1">
        <p>Major Project</p>
        </p>
        <p class="text-white d-flex">
        <p>CSM | Team-04</p>
        </p>
      </div>
      <div class="col-lg-3 col-md-4 col-12 mt-4 mt-lg-0 ms-auto">
        <p class="copyright-text mt-lg-5 mt-4">Copyright © CSM || Team-
04 || MLRIT.
      </div>
    </div>
   </div>
  </footer>
  <!-- JAVASCRIPT FILES -->
  <script src="{{ url_for('static', filename='js/jquery.min.js')
}}"></script>
  <script src="{{ url_for('static',
filename='js/bootstrap.bundle.min.js') }}"></script>
  <link rel="stylesheet" href="{{ url_for('static',
filename='css/bootstrap.min.css') }}?v=1.0">
  <script src="{{ url_for('static', filename='js/jquery.sticky.js')
}}"></script>
  <script src="{{ url_for('static', filename='js/click-scroll.js')
}}"></script>
  <script src="{{ url_for('static', filename='js/custom.js')
}}"></script>
 </body>
</html>
```

**submit-page.css**

```
/* submit-page.css */
/* Style for navigation bar background color */
html {
    -ms-overflow-style: none;  /* Internet Explorer 10+ */
    scrollbar-width: none;
}
html::-webkit-scrollbar{
```

```
    display: none;  /* Safari and Chrome */
}
.body{
    background-image: url('static/images/Premium Photo _ Online
shopping icon on smart phone for global concept.jpg');
    background-color: #f5f5f5
}
.navbar {
    background-color: #27277e; /* Replace with your preferred shade of
blue */
    display: flex;
    justify-content: center;
    align-items: center;
}
.container_for_title {
    display: flex;
    justify-content: center; /* Horizontally center the child elements
*/
    align-items: center; /* Vertically center the child elements */
    height: 50px;
}
.container_for_options {
    display: flex;
    justify-content: space-around; /* Horizontally center the child
elements */
    align-items: center; /* Vertically center the child elements */
    height: 100px;
}
.selection button{
    margin: 0 30px;
}
.button {
    display: inline-block;
    padding: 15px 25px;
    font-size: 15px;
    cursor: pointer;
    text-align: center;
    text-decoration: none;
    outline: none;
    border: none;
    border-radius: 15px;
    box-shadow: 0 3px #c0c0c0;
```

```css
.button:hover {
    background-color: #3e8e41;
}
.button:active {
    background-color: #3e8e41;
    box-shadow: 0 3px #b7b7b7;
    transform: translateY(4px);
}
.button.all {
    background-color: #f8b946; /* Orange for 'All' */
}
.button.positive {
    background-color: #3eecad; /* Green for 'Positive' */
}
.button.negative {
    background-color: #f94b4b; /* Red for 'Negative' */
}
.button.neutral {
    background-color: #5190c4; /* Blue for 'Neutral' */
}
.list {
    list-style-type: none; /* Removes default bullet points */
    padding: 0; }
.list li {
    display: inline-block; /* Display list items side by side */
    margin-right: 10px; /* Add some space between list items */
}
.container_for_graph {
    display: flex;
    justify-content: center; /* Horizontally center the child elements
*/
    align-items: center; /* Vertically center the child elements */
    height: 100vh; /* Set the container height to the full viewport
height */
}
.graphs {
    width: 700px;
    height: 400px;
}
/* Additional styles for the page if needed */
```

**topic-listing.css**

```
// Function to filter reviews based on sentiment
.section-padding {
  padding-top: 100px;
  padding-bottom: 100px;
}

.section-bg {
  background-color: var(--section-bg-color);
}

.section-overlay {
  background-image: linear-gradient(15deg, #92d0f5 0%, #27277e 100%);
  position: absolute;
  top: 0;
  left: 0;
  pointer-events: none;
  width: 100%;
  height: 100%;
  opacity: 0.85;
}

.section-overlay + .container {
  position: relative;
}
```

```
.tab-content {
  background-color: var(--white-color);
  border-radius: var(--border-radius-medium);
}

.nav-tabs {
  border-bottom: 1px solid #ecf3f2;
  margin-bottom: 40px;
  justify-content: center;
}

.nav-tabs .nav-link {
  border-radius: 0;
  border-top: 0;
  border-right: 0;
  border-left: 0;
  color: var(--p-color);
  font-family: var(--title-font-family);
  font-size: var(--btn-font-size);
  font-weight: var(--font-weight-medium);
  padding: 15px 25px;
  transition: all 0.3s;
}
```

```
.nav-tabs .nav-link:first-child {
  margin-right: 20px;
}

.nav-tabs .nav-item.show .nav-link,
.nav-tabs .nav-link.active,
.nav-tabs .nav-link:focus,
.nav-tabs .nav-link:hover {
  border-bottom-color: var(--primary-color);
  color: var(--primary-color);
}


/*---------------------------------------
  CUSTOM ICON COLOR
-----------------------------------------*/
.custom-icon {
  color: var(--secondary-color);
}


/*---------------------------------------
  CUSTOM BUTTON
-----------------------------------------*/
.custom-btn {
```

```
  background: var(--custom-btn-bg-color);
  border: 2px solid transparent;
  border-radius: var(--border-radius-large);
  color: var(--white-color);
  font-size: var(--btn-font-size);
  font-weight: var(--font-weight-semibold);
  line-height: normal;
  transition: all 0.3s;
  padding: 10px 20px;
}

.custom-btn:hover {
  background: var(--custom-btn-bg-hover-color);
  color: var(--white-color);
}

.custom-border-btn {
  background: transparent;
  border: 2px solid var(--custom-btn-bg-color);
  color: var(--custom-btn-bg-color);
}

.custom-border-btn:hover {
  background: var(--custom-btn-bg-color);
  border-color: transparent;
```

```
  color: var(--white-color);
}


.custom-btn-bg-white {
  border-color: var(--white-color);
  color: var(--white-color);
}



/*---------------------------------------
  SITE HEADER
-----------------------------------------*/
.site-header {
  background-image: linear-gradient(15deg, #92d0f5 0%, #27277e 100%);
  padding-top: 150px;
  padding-bottom: 80px;
}


.site-header .container {
  height: 100%;
}


.breadcrumb-item+.breadcrumb-item::before,
.breadcrumb-item a:hover,
.breadcrumb-item.active {
```

```
  color: var(--white-color);
}

.site-header .custom-icon {
  color: var(--white-color);
  font-size: var(--h4-font-size);
}

.site-header .custom-icon:hover {
  color: var(--secondary-color);
}
/*---------------------------------------
  NAVIGATION
-----------------------------------------*/
.sticky-wrapper {
  position: absolute;
  top: 0;
  right: 0;
  left: 0;
}

.sticky-wrapper.is-sticky .navbar {
  background-color: var(--secondary-color);
}
```

```css
.navbar {
  background: transparent;
  border-bottom: 1px solid rgba(128, 208, 199, 0.35);
  z-index: 9;
}

.navbar-brand,
.navbar-brand:hover {
  font-size: var(--h3-font-size);
  font-weight: var(--font-weight-bold);
  display: block;
}

.navbar-brand span {
  font-family: var(--title-font-family);
}

.navbar-expand-lg .navbar-nav .nav-link {
  border-radius: var(--border-radius-large);
  margin: 10px;
  padding: 10px;
  text-align: center;
}

.navbar-nav .nav-link {
```

```
  display: inline-block;
  color: var(--white-color);
  font-family: var(--title-font-family);
  font-size: var(--menu-font-size);
  font-weight: var(--font-weight-medium);
  text-transform: uppercase;
  letter-spacing: 0.5px;
  position: relative;
  padding-top: 15px;
  padding-bottom: 15px;
}

.navbar-nav .nav-link.active,
.navbar-nav .nav-link:hover {
  color: var(--primary-color);
}

.navbar .dropdown-menu {
  background: var(--white-color);
  box-shadow: 0 1rem 3rem rgba(0,0,0,.175);
  border: 0;
  display: inherit;
  opacity: 0;
  min-width: 9rem;
  margin-top: 20px;
```

```
  padding: 13px 0 10px 0;
  transition: all 0.3s;
  pointer-events: none;
}

.navbar .dropdown-menu::before {
  content: "";
  width: 0;
  height: 0;
  border-left: 20px solid transparent;
  border-right: 20px solid transparent;
  border-bottom: 15px solid var(--white-color);
  position: absolute;
  top: -10px;
  left: 10px;
}

.navbar .dropdown-item {
  display: inline-block;
  color: var(--p-bg-color);
  font-family: var(--title-font-family);
  font-size: var(--menu-font-size);
  font-weight: var(--font-weight-medium);
  text-transform: uppercase;
  letter-spacing: 0.5px;
```

```
  position: relative;
}

.navbar .dropdown-item.active,
.navbar .dropdown-item:active,
.navbar .dropdown-item:focus,
.navbar .dropdown-item:hover {
  background: transparent;
  color: var(--primary-color);
}

.navbar .dropdown-toggle::after {
  content: "\f282";
  display: inline-block;
  font-family: bootstrap-icons !important;
  font-size: var(--copyright-font-size);
  font-style: normal;
  font-weight: normal !important;
  font-variant: normal;
  text-transform: none;
  line-height: 1;
  vertical-align: -.125em;
  -webkit-font-smoothing: antialiased;
  -moz-osx-font-smoothing: grayscale;
  position: relative;
```

```
  left: 2px;

  border: 0;

}
```

**script.js**

```
// Function to filter reviews based on sentiment
function filterReviews(sentiment) {
    var rows = document.querySelectorAll('.review-row');

    // Initialize counts for positive, negative, and neutral reviews
    var positiveCount = 0;
    var negativeCount = 0;
    var neutralCount = 0;

    rows.forEach(function(row) {
        var rowSentiment = row.getAttribute('data-sentiment');

        if (sentiment === 'All' || rowSentiment === sentiment) {
            row.style.display = 'table-row';

            // Increment respective count based on sentiment
            if (rowSentiment === 'Positive') {
                positiveCount++;
            } else if (rowSentiment === 'Negative') {
                negativeCount++;
            } else if (rowSentiment === 'Neutral') {
                neutralCount++;
            }
        } else {
            row.style.display = 'none';
        }
    });

    // Render bar graph using counts
    renderBarGraph(positiveCount, negativeCount, neutralCount);
}

// Function to render the bar graph
function renderBarGraph(positiveCount, negativeCount, neutralCount) {
```

```
    const ctx =
document.getElementById('sentimentChart').getContext('2d');

    const myChart = new Chart(ctx, {
        type: 'bar',
        data: {
            labels: ['Positive', 'Negative', 'Neutral'],
            datasets: [{
                label: 'Number of Reviews',
                data: [positiveCount, negativeCount, neutralCount],
                backgroundColor: [
                    'rgba(50, 205, 50, 0.5)',  // Green for positive
                    'rgba(255, 99, 71, 0.5)',   // Red for negative
                    'rgba(135, 206, 250, 0.5)'
                ],
                borderColor: [
                    'rgba(75, 192, 192, 1)',
                    'rgba(255, 99, 132, 1)',
                    'rgba(255, 206, 86, 1)'
                ],
                borderWidth: 1
            }]
        },
        options: {
            scales: {
                yAxes: [{
                    ticks: {
                        beginAtZero: true
                    }
                }]
            }
        }
    });
}
```

**click-scroll.js**

```
//jquery-click-scroll
//by syamsul'isul' Arifin

var sectionArray = [1, 2, 3, 4, 5];

$.each(sectionArray, function(index, value){
```

```
     $(document).scroll(function(){
         var offsetSection = $('#' + 'section_' + value).offset().top
- 75;
         var docScroll = $(document).scrollTop();
         var docScroll1 = docScroll + 1;


         if ( docScroll1 >= offsetSection ){
             $('.navbar-nav .nav-item .nav-
link').removeClass('active');
             $('.navbar-nav .nav-item .nav-
link:link').addClass('inactive');
             $('.navbar-nav .nav-item .nav-
link').eq(index).addClass('active');
             $('.navbar-nav .nav-item .nav-
link').eq(index).removeClass('inactive');
         }

     });

    $('.click-scroll').eq(index).click(function(e){
        var offsetClick = $('#' + 'section_' + value).offset().top -
75;
        e.preventDefault();
        $('html, body').animate({
            'scrollTop':offsetClick
        }, 300)
    });

});

$(document).ready(function(){
    $('.navbar-nav .nav-item .nav-link:link').addClass('inactive');
    $('.navbar-nav .nav-item .nav-link').eq(0).addClass('active');
    $('.navbar-nav .nav-item .nav-
link:link').eq(0).removeClass('inactive');
});
```

**app.py**

```
from flask import Flask, redirect, render_template, request, url_for
from main import mainfun as mf
from fake import convertmyTxt
```

```
import pandas as pd

app = Flask(__name__)

@app.route('/')
def index():
    return render_template('index.html')

@app.route('/new_page', methods=['POST', 'GET'])
def new_page():
    input_text = request.form['keyword']
    csv_path, Title= mf(input_text)
    df = pd.read_csv(csv_path)
    return render_template('new_page.html', submitted_text=input_text,
csv_data=df, Name=Title)

@app.route('/topics_detail')
def topics_detail():
    return render_template('topics_detail.html')



if __name__ == '__main__':
    app.run(debug=True)
```

**fake.py**

```
import pandas as pd
import string
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import CountVectorizer,
TfidfTransformer
from sklearn.pipeline import Pipeline
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score
import nltk
from nltk.corpus import stopwords
from joblib import dump, load

def convertmyTxt(rv):
    np = [c for c in rv if c not in string.punctuation]
    np = ''.join(np)
    return [w for w in np.split() if w.lower() not in
stopwords.words('english')]
```

```
def train_svm_model():
    dataframe = pd.read_csv('fake reviews dataset.csv')  # Replace
with your actual dataset filename
    dataframe.dropna(inplace=True)
    dataframe['length'] = dataframe['text_'].apply(len)

    x_train, x_test, y_train, y_test =
train_test_split(dataframe['text_'], dataframe['label'],
test_size=0.25)

    pip = Pipeline([
        ('bow', CountVectorizer(analyzer=convertmyTxt)),
        ('tfidf', TfidfTransformer()),
        ('classifier', SVC())
    ])

    nltk.download('stopwords')

    pip.fit(x_train, y_train)
    supportVectorClassifier = pip.predict(x_test)
    accuracy = accuracy_score(y_test, supportVectorClassifier)
    print(f"Accuracy: {accuracy}")

    # Save the trained pipeline to a file
    dump(pip, 'svm_model.joblib')

    return pip

def load_svm_model():
    # Load the trained pipeline from the file
    return load('svm_model.joblib')

if __name__ == "__main__":
    trained_pipeline = train_svm_model()
```

**main.py**

```
import amz
import flk
import sentiment
from fake import convertmyTxt
from fake import load_svm_model
```

```
import os
from bs4 import BeautifulSoup
import urllib
import requests


def mainfun(url):
    #url=input("Enter url: ")


    df ,Title= amz.amazon(url)
    def check_csv_exists(file_name):
        return os.path.isfile(file_name)
    csv_file_name = Title+'.csv'
    if check_csv_exists(csv_file_name):
        return csv_file_name, Title
    else:
        # Load the trained pipeline from the file
        trained_pipeline = load_svm_model()

        # Test with a custom sentence lable prediction
        print("processing")
        df['label'] = trained_pipeline.predict(df['Description'])

        #prediction = trained_pipeline.predict([custom_sentence])

        #print(f"Predicted Label: {prediction[0]}")


        #sentiment analysis
        df['Sentiment'] =
df['Description'].apply(sentiment.analyze_sentiment)

        print(df)
        csv_file_path=Title+'.csv'
        with open(csv_file_path, 'w') as csvfile:
            pass

        # Use the to_csv method to convert and save the DataFrame to a
CSV file
        df.to_csv(csv_file_path, index=False)
        return csv_file_name, Title
```

# CHAPTER 6

# RESULTS

The obtained results after the development of the Classer application are as follows.

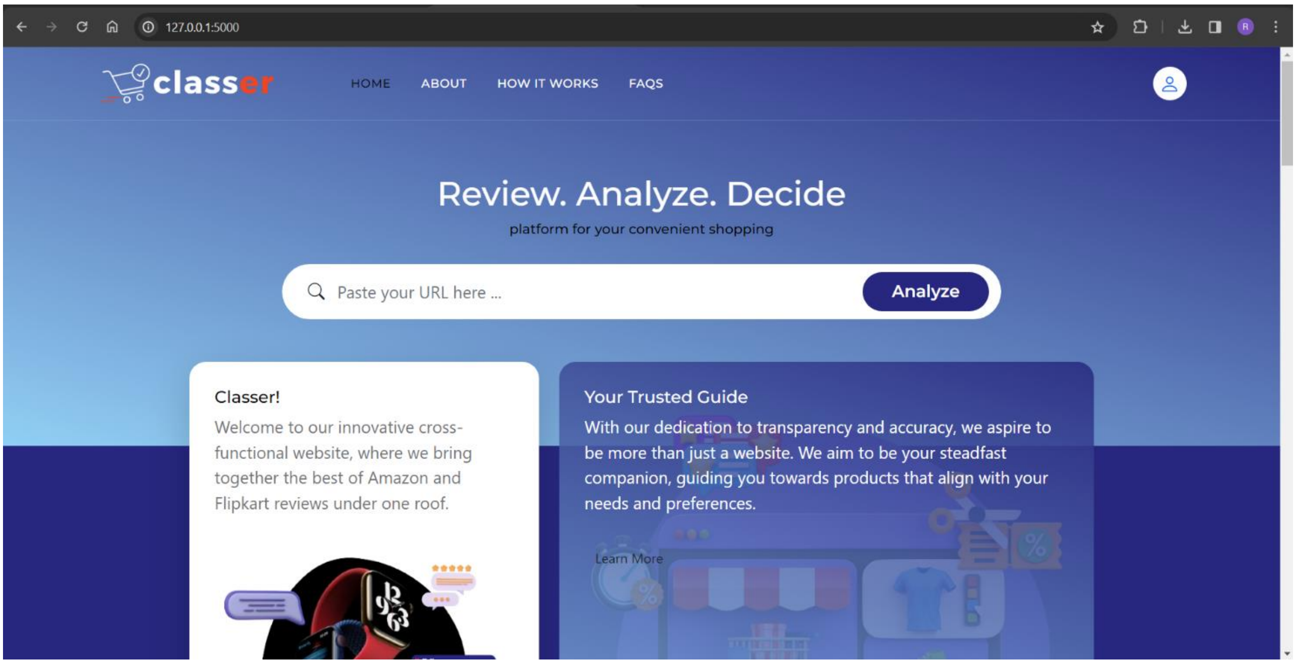



Figure 14: Classer Home page

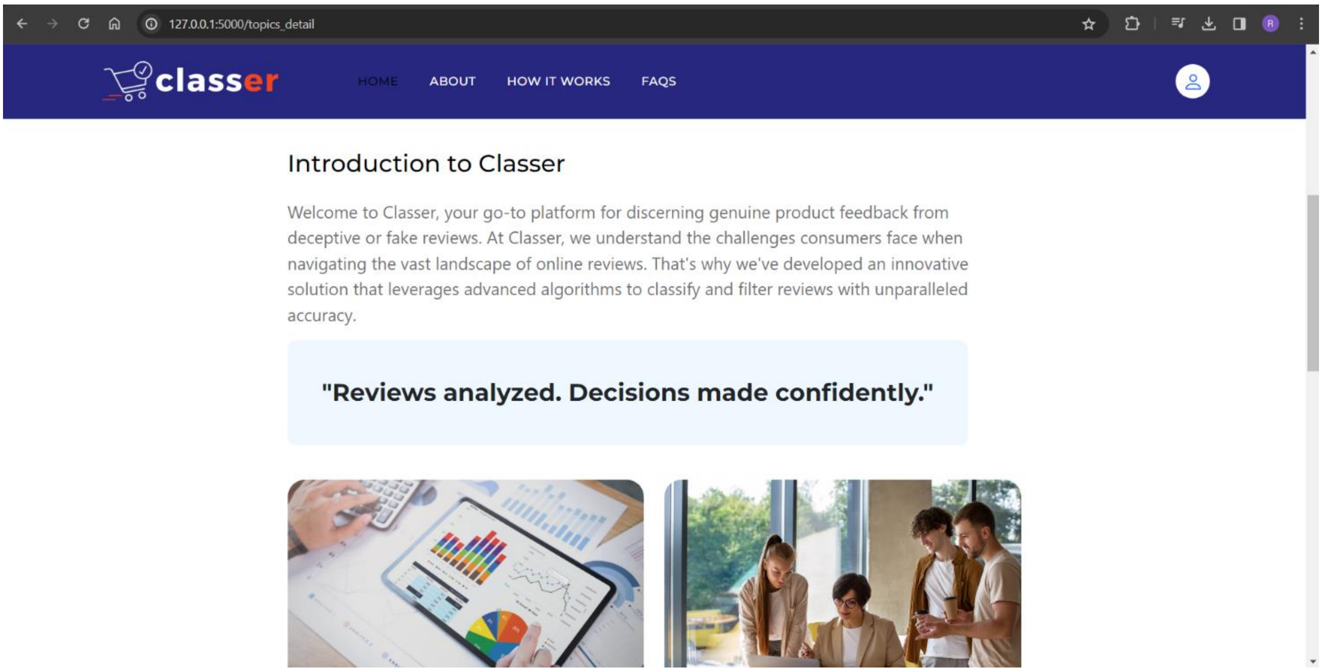



Figure 15: Classer Web application description

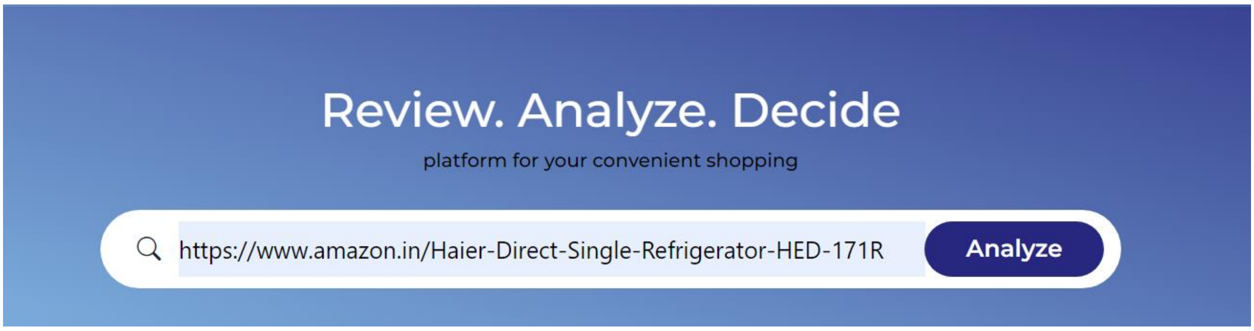
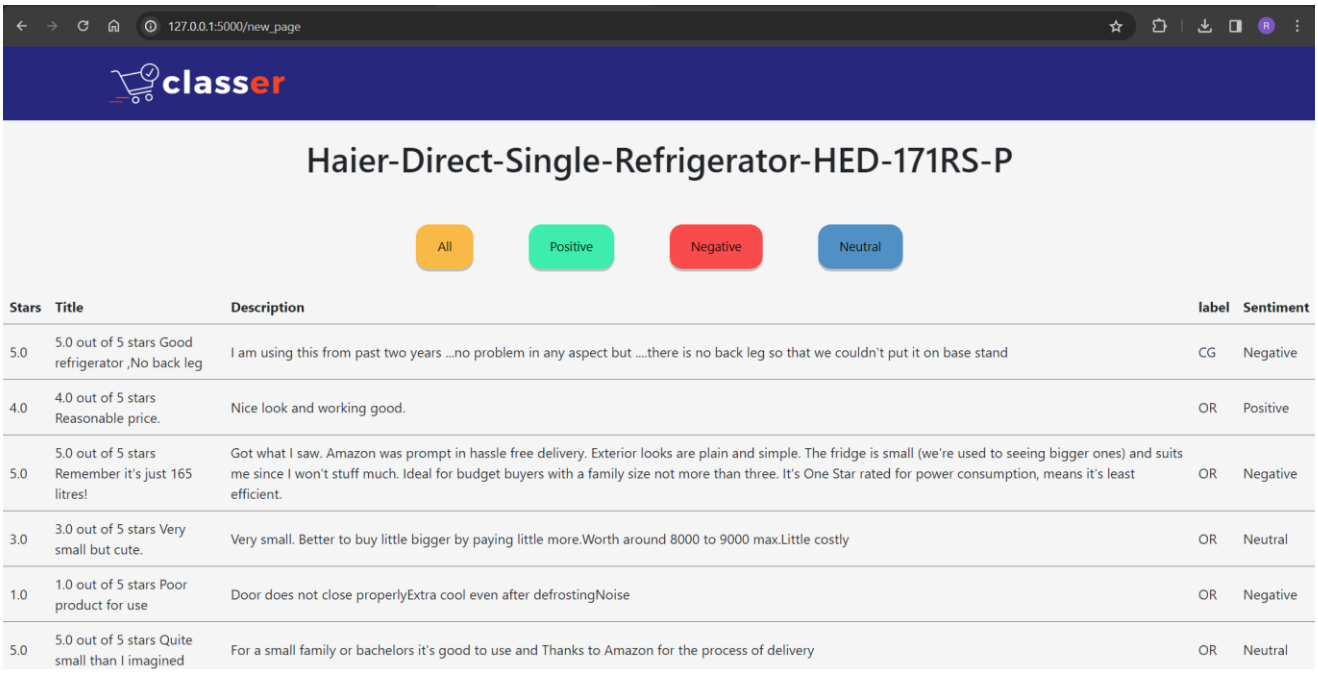



Figure 16: Search bar to analyze product reviews




| Stars | Title | Description | label | Sentiment |
|---|---|---|---|---|
| 5.0 | 5.0 out of 5 stars Good refrigerator ,No back leg | I am using this from past two years ...no problem in any aspect but ....there is no back leg so that we couldn't put it on base stand | CG | Negative |
| 4.0 | 4.0 out of 5 stars Reasonable price. | Nice look and working good. | OR | Positive |
| 5.0 | 5.0 out of 5 stars Remember it's just 165 litres! | Got what I saw. Amazon was prompt in hassle free delivery. Exterior looks are plain and simple. The fridge is small (we're used to seeing bigger ones) and suits me since I won't stuff much. Ideal for budget buyers with a family size not more than three. It's One Star rated for power consumption, means it's least efficient. | OR | Negative |
| 3.0 | 3.0 out of 5 stars Very small but cute. | Very small. Better to buy little bigger by paying little more.Worth around 8000 to 9000 max.Little costly | OR | Neutral |
| 1.0 | 1.0 out of 5 stars Poor product for use | Door does not close properlyExtra cool even after defrostingNoise | OR | Negative |
| 5.0 | 5.0 out of 5 stars Quite small than I imagined | For a small family or bachelors it's good to use and Thanks to Amazon for the process of delivery | OR | Neutral |

Figure 17: Classified resultant reviews of product

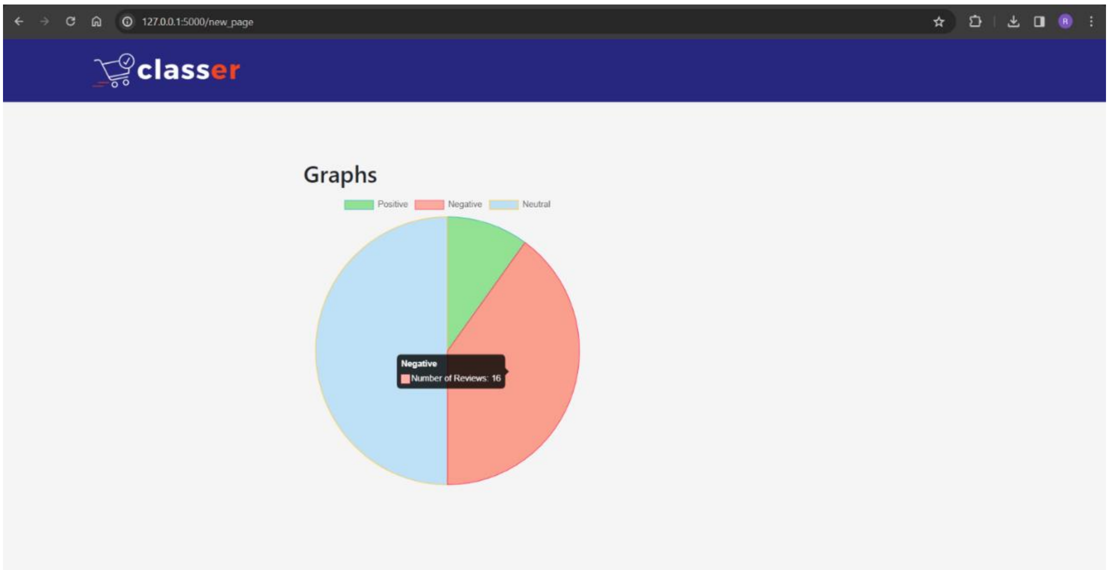



Figure 18: Visualizing results through pie charts

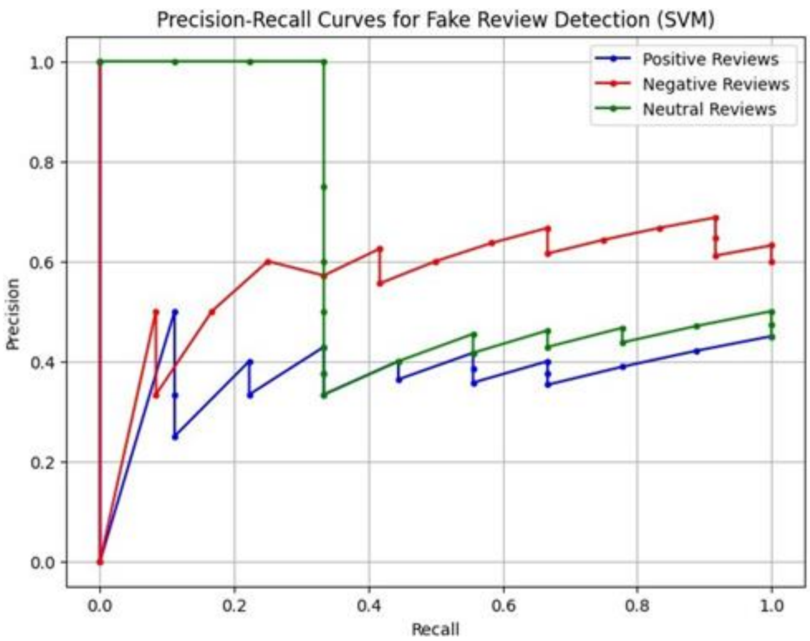
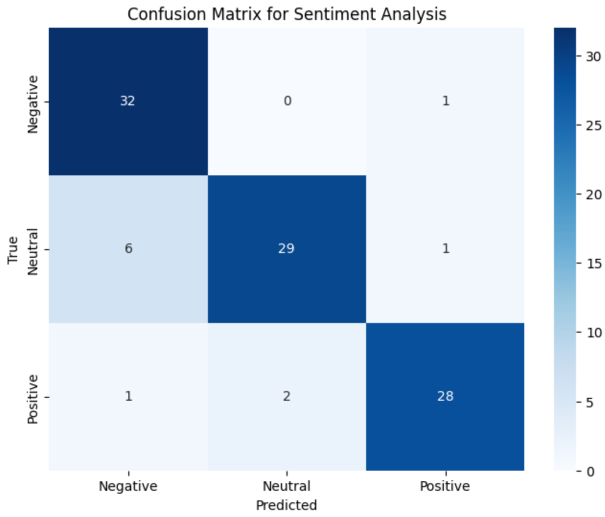



Figure 19: Image of Precision and Recall Curve



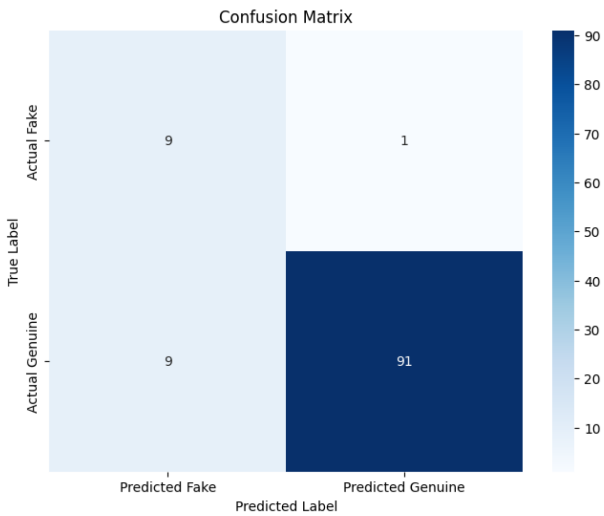




Figure 20: Confusion matrix for Sentiment Analysis and SVM

# CHAPTER 7

# CONCLUSION

The document provides valuable insights into the future enhancements in technology for e-commerce review analysis, emphasizing the potential for more advanced machine learning techniques and improvements in user interface design to further enhance user experiences.

Advanced machine learning algorithms hold the promise of revolutionizing sentiment analysis by enabling more nuanced and accurate interpretation of reviews. Future enhancements may involve the development of deep learning models capable of understanding complex linguistic structures and contextual cues, thereby improving the classification of reviews into nuanced categories beyond just positive, negative, or neutral sentiments. These advancements would provide users with more nuanced insights into product reviews, allowing for a deeper understanding of customer sentiment and preferences.

## FUTURE ENHANCEMENTS AND DISCUSSIONS

Our research has demonstrated how technology can revolutionize customer decision-making in the e-commerce space. Our model accelerated review aggregation, sentiment analysis, and authenticity assessment by combining web crawling, machine learning, and natural language processing. This approach provides visitors the ability to confidently and quickly browse through a variety of reviews. The advantages are obvious: better understanding, more trust, and well-informed decisions. Future developments in machine learning for more complex sentiment analysis and UI improvements should improve user experiences. This study highlights how technology is changing the way that e-commerce reviews are analyzed, providing insight into a future in which customers have even more power and confidence when making purchases online.