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 • Supervised learning techniques • Semi-supervised learning techniques • 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 and classifiers used for fake review detection.	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.
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 unlabelled 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 unlabelled 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 unlabelled 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 unlabelled 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.

Input and Output	Feature of This Solution	Contribution & The Value of This Work
Input: Reviews of the products primarily taken from e-commerce sites. **Output:** The performance of the fake review detection methods.	It covers linguistic and textual features, behavioural features, and relational features.	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.

Input	Output
Reviews of the products primarily taken from e-commerce sites.	The performance of the fake review detection methods.

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.

Analyse 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 I. Introduction II. Related Work III. Machine learning based fake review detection techniques IV. Analysis V. Major challenges VI. Conclusion

Diagram/Flowchart

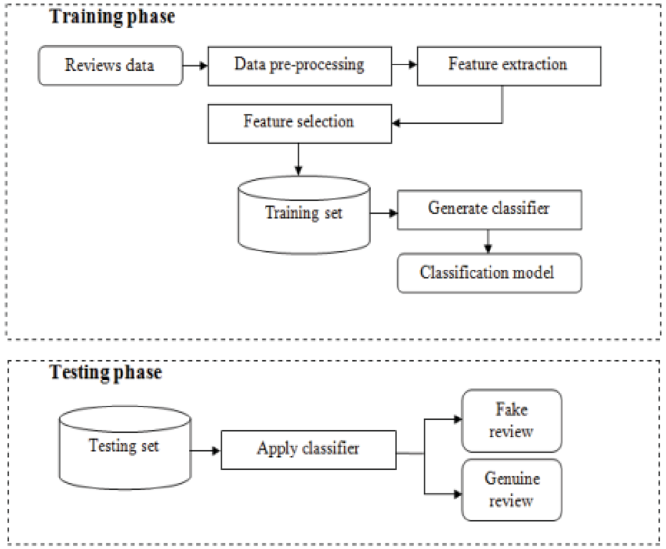


Fig. 1. Machine Learning based Fake Review Detection

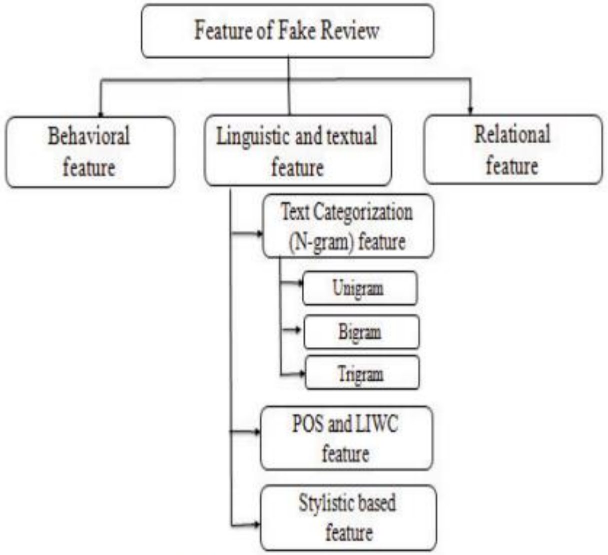


Fig. 2. Types of Fake Review Features

--- End of Paper 1 ---

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. Train-validation set split with a ratio of 75:25 to obtain the train set and validation set. 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	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

		analysis focuses on unique and significant terms. These features enable a deep dive into the emotional aspects of reviews.	can be computationally intensive. In cases of limited data 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. 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**	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.
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.		

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 I. Introduction II. Related Work III. Proposed Work IV. Performance analysis V. Conclusion

Diagram/Flowchart

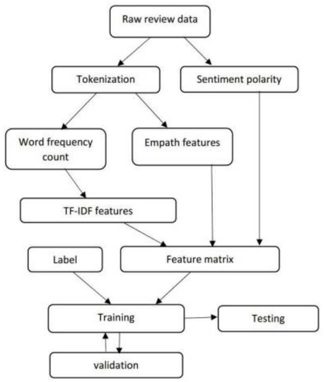


Fig. 1: Proposed classification model

--- End of Paper 2 ---

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	It is not a best option in situations where non-linear relationships need to be

		commonly used for binary classification problems.	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 non linear 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 both classification and regression problems.	Similar to the SVM, XGBoost can also be computationally expensive and may require careful tuning of hyperparameters to achieve optimal performance.

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
<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>	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 I. Introduction II. Literature Review III. Proposed Methodology IV. Evaluation V. Conclusion

Diagram/Flowchart

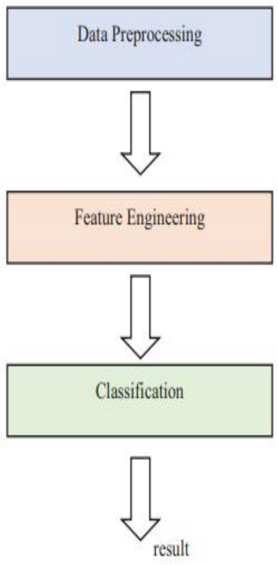


Figure 1. Overall workflow.

--End of Paper 3---

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. Additionally, they can show patterns of similarity and frequency between reviewers and products, which could point to the existence of fake reviews.	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

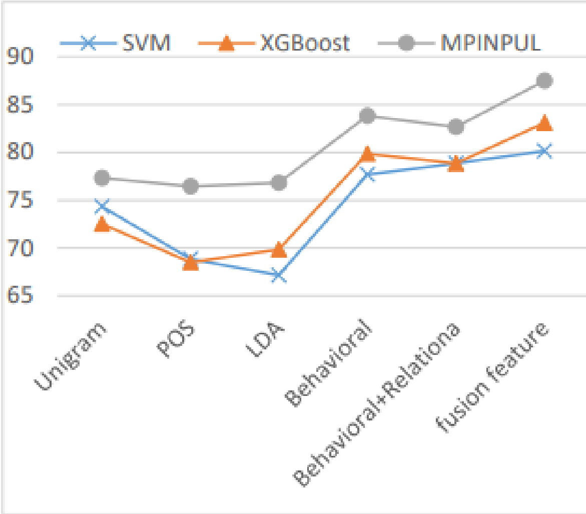
Dependent Variable	Independent Variable	Moderating variable	Mediating (Intervening) variable
The evaluation index which includes accuracy, precision, recall and F1 score acts as dependent variables, which are used to evaluate the performance of classification model.	The major features such as text, behaviour, 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
<table><tr><th>Input</th><th>Output</th></tr><tr><td>Yelp dataset.</td><td>Accuracy of MPINPUL model.</td></tr></table>	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 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.

Analyse This Work By Critical Thinking	The Tools That Assessed this Work	What is the Structure of this Paper
The paper demonstrates the significance of the critic's behaviour in identifying fake reviews and the feasibility and effectiveness of the MPINPUL model.	None	Abstract I. Introduction II. Feature construction III. MPINPUL Classification Model IV. Performance evaluation V. Conclusion

Diagram/Flowchart


--End of Paper 4—

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. It helps in identifying trends and patterns in customer sentiment.	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

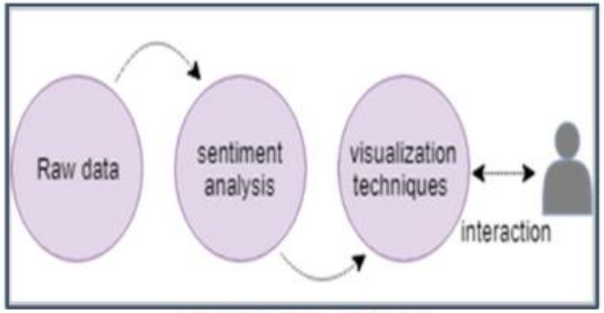
Dependent Variable	Independent Variable	Moderating variable	Mediating (Intervening) variable
The polarity and the rating of each review.	The independent variables in this document are, the book title and different levels of customer satisfaction and feedback	The review's time, which could have an impact 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
<table><tr><th>Input</th><th>Output</th></tr><tr><td>Book's review</td><td>Visual representation of the review</td></tr></table>	The main features of the proposed solution are interactive solution and the sentiment analysis of customer reviews, helping users in making decisions	By summarising viewpoints and emphasising sentiment trends, the proposed solution helps users save time and effort by making it simple to compare books and find the ones with the most positive reviews.

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 I. Introduction II. Feature construction III. MPINPUL Classification Model IV. Performance evaluation V. Conclusion

Diagram/Flowchart
 Figure.1. System Architecture

Version 1.0 _ Week 1		
1		
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 • Word Dictionary. • User Defined Classification. • Fuzzy Logic. • Deep Learning. • Machine Learning.	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 reviews into different categories which include positive, negative, and neutral, and accuracy is been compared with existing methods of ML	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, negative, neutral, Positively fake, negatively fake, and so on.

The Process (Mechanism) of this Work; Means How the Problem has been Solved & Advantages & Disadvantages of Each Step in This Process

	Process Steps	Advantage	Disadvantage (Limitation)
1	The primary emphasis of the analysis is placed on reviews related to Amazon electronic products.	90000 reviews from different customers had been gather which helps in dealing with different word ambiguation and helps in achieving high accuracy.	Limited and restricted customers information is been shared, Privacy of the customers is the main reason behind the limitations in shared data. Example: Transaction details

2	Sanitization.	This process helps in the removal of unwanted symbols and characters and also helps in tokenization along with POS–tagging.	While dealing with similar words many dictionaries are been created that may occupy more space and increase space complexity.
3	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
4	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.

5	Deep Learning	LSTM which is also known as RNN involves 7 steps, uses the Adam optimizer for performing the multiclass classification, and the Cross-Entropy loss function is used to evaluate the network model with this loss value decreased by the optimizer	Deep learning requires a large amount of data. Complex data models require expensive GPUs. Overfitting may also occur due to an excess amount of training data.
6	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 textpreprocessing 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, theaccuracy of the classification (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.

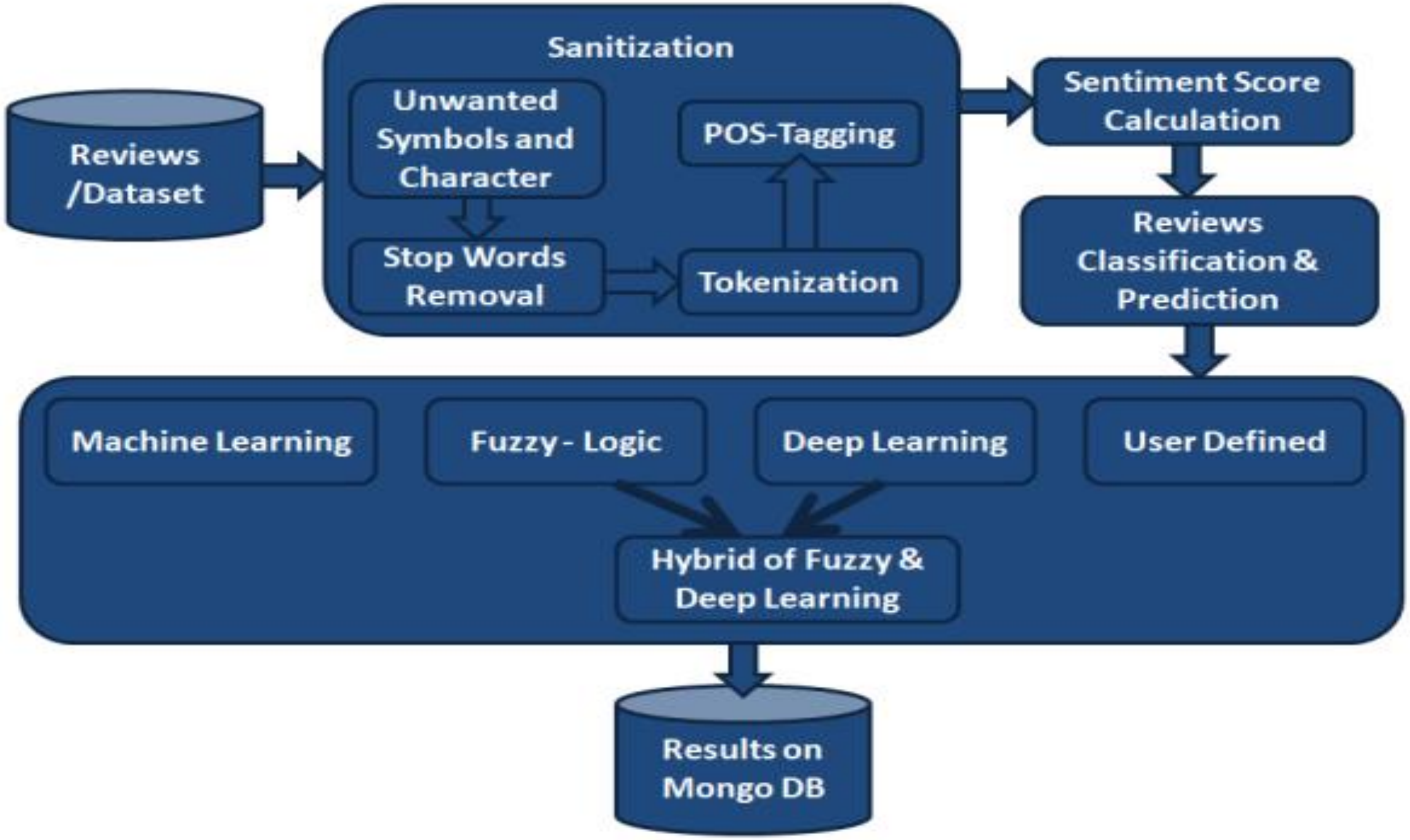
The moderating variable might be the "segmentation ratio" of the training and test data sets. It influences the performance of classification by altering the proportion of test data.

Input and Output	Feature of This Solution	Contribution & The Value of This Work
<table><tr><th>Input</th><th>Output</th></tr><tr><td>9000 customer reviews primarily from the Amazon mobile electronics products</td><td>Categorizing these reviews into groups: Genuine reviews, and fake reviews such as positive, negative, and Neutral.</td></tr></table>	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.	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.

Positive Impact of this Solution in This Project Domain	Negative Impact of this Solution in This Project Domain
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.	Since this is a performance evaluation of various algorithms, No negative side to this project as all the things used are defined in advance.

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 understanding of how the size of data and methods used, plays a crucial role in influencing the performance of different algorithms.	Web Browser	Abstract I. Introduction **II. Related Works** **III. Proposed Work** **IV. Experimental Analysis** **V. Conclusion** and **References**

Diagram/Flowchart



---End of Paper 1-

2		
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/ ... etc)	**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 been 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 their limitations on giving data publicly.
2	**Pre-processing** is the next crucial step in this process of polarizing the reviews	This results in an optimal solution more quickly compared to traditional gradient descent methods.	Its effectiveness depends on the specific problem being solved.

	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.	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.	During the feature extraction process, some information may be lost. This is especially true when using techniques like the bag of words approach, where the order and context of words are disregarded. This loss of information can affect the accuracy and completeness of the analysis
	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 variables (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.

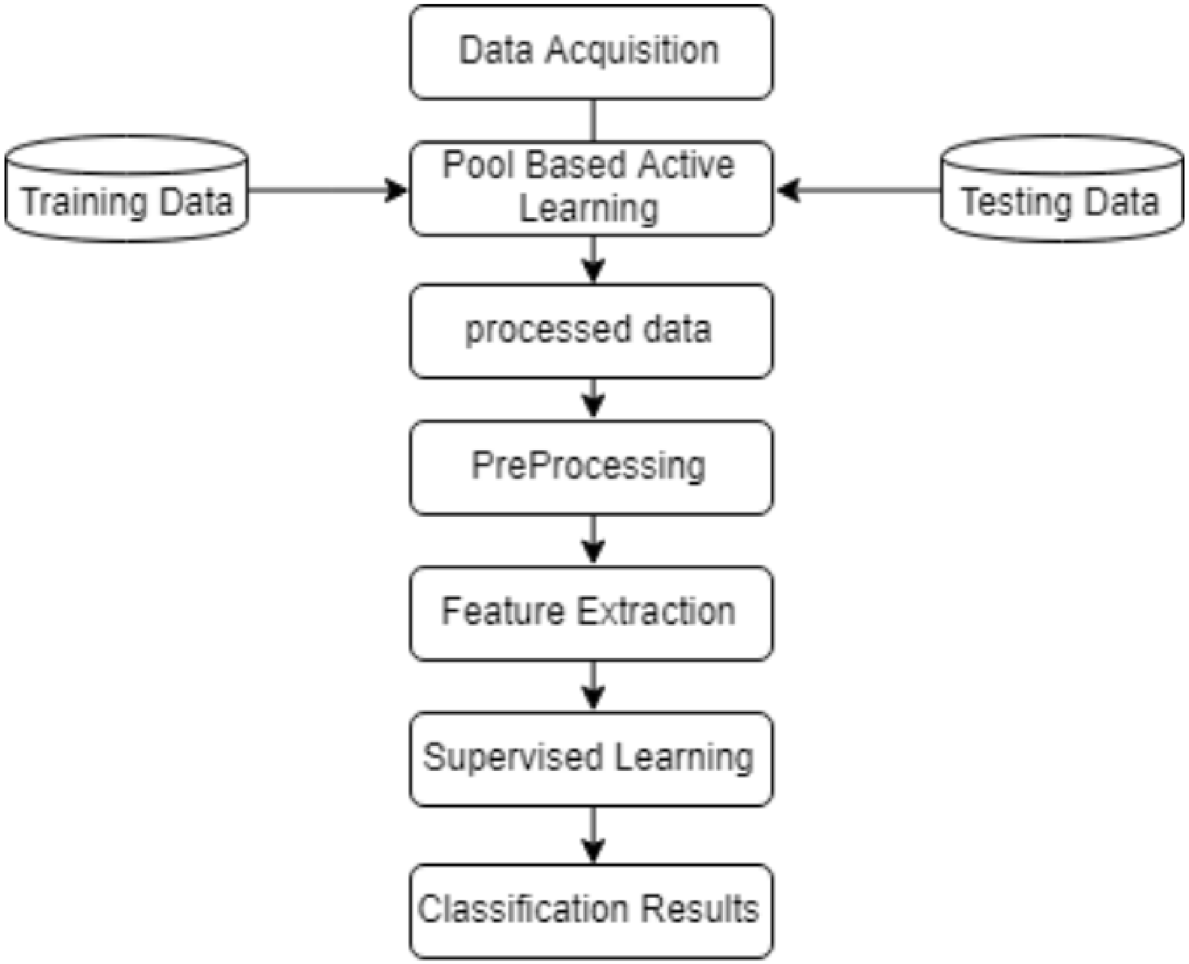
The sentiment of reviews depends on the application of the ANN algorithm, as the algorithm's output classifies reviews as positive or negative based on patterns it has learned.

<table>
<tr><th colspan="2">Input and Output</th><th>Feature of This Solution</th><th>Contribution & The Value of This Work</th></tr>
<tr><td>Input</td><td>Output</td><td rowspan="2">Pool Based Learning along with Machine Learning algorithms like Linear SVM, Naïve Bayes, Stochastic Gradient Descent, Random Forest</td><td rowspan="2">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.</td></tr>
<tr><td>Amazon Labeled Dataset after the active learning process</td><td>Accuracy of classifier Precision, Recall, F1-Measure for positive and Deceptive values.</td></tr>
</table>

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 I. Introduction II. **Related Works and research** III. **Methodology** IV. **Results** V. **Comparative Analysis** VI. **Conclusion and Future works**

Diagram/Flowchart



Fig,3 overall working of the model

--End of Paper 3--

3		
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 • Preprocessing • Feature extraction • Data classification • 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. May not work well with noisy or irrelevant features
5	Classification involves some methods like Back Propagation Neural Networks (BPNN) and ML algorithms for comparative analysis.	BPNN highly adaptable and efficient, and don't require prior knowledge about the network.	Here Unclean data can affect the backpropagation algorithm when training a neural network used for data mining.

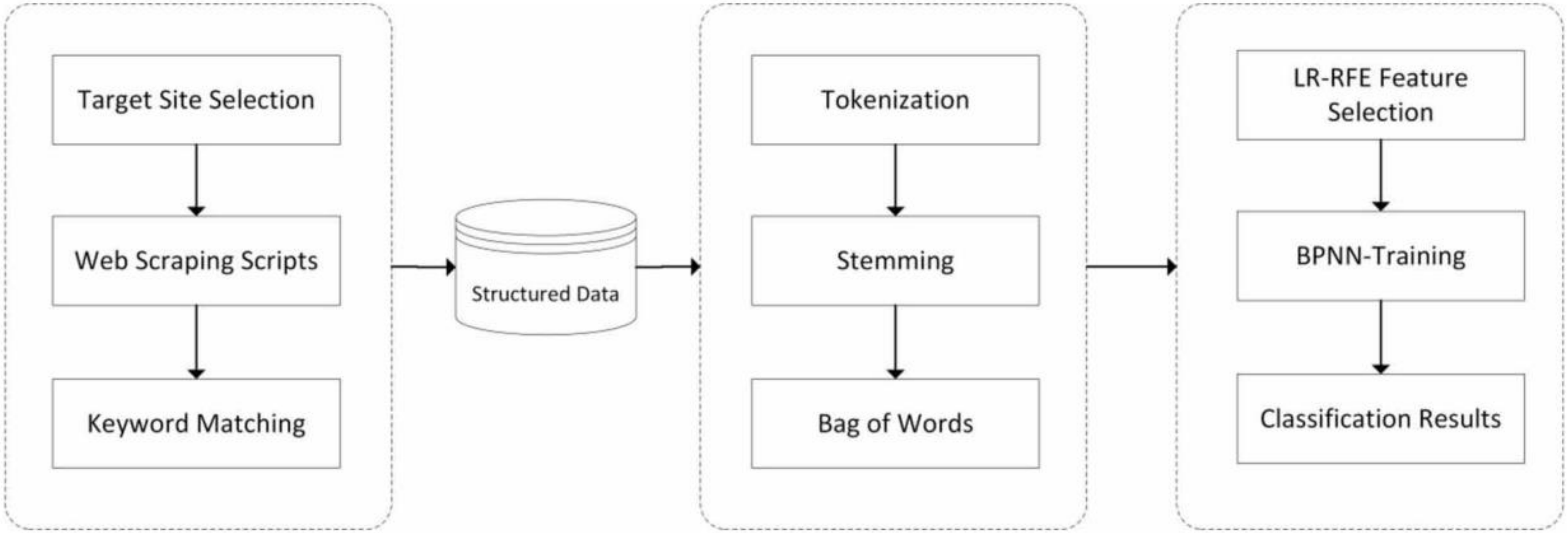
Major Impact Factors in this Work

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.
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.

<table>
<tr><th>Input and Output</th><th colspan="2">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>Product</td><td>to view and compare product details from different websites</td></tr></table></td><td colspan="2">This solution addresses the time-consuming and manual effort required to visit websites.</td><td>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.</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">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</td><td colspan="2">This proposed model does not showcase any negative impacts in the project. The only thing that can be described as a negative impact is the</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 solution improves the efficiency and convenience of comparing products from different websites on a single platform.</td><td colspan="2">None</td><td>Abstract
I. Introduction
II. Background
III. Scraping Techniques
IV. Materials and Methods
V. Results
VI. Conclusion</td></tr>
<tr><th colspan="4">Diagram/Flowchart</th></tr>
<tr><td colspan="4"></td></tr>
</table>



--End of Paper 2—

4		
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.eAdu	Text mining, Supervised techniques Support vector Machine, Naïve bayes.
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?**
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		

	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 product-related reviews that	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 to handle high-dimensional and non-linear data	Support Vectors are computationally expensive and limited to some of the applications.
	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.

Major Impact Factors in this Work

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 no such Mediating variables concept.inp

Relationship Among The Above 4 Variables in This article

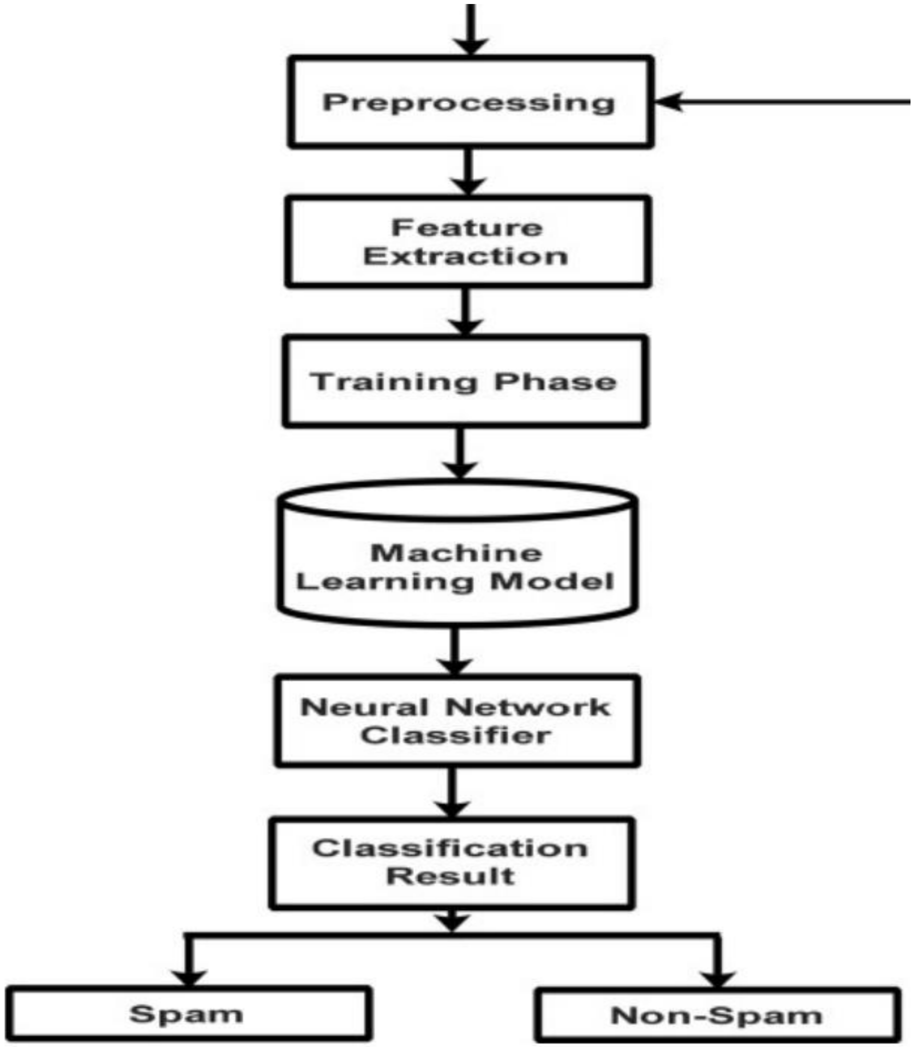
the paper discusses different supervised techniques, with each method having its own set of independent variables. Where the accuracy os completely dependent on the Machine learning algorithms and the algorithmic approach obtained by using parameters of the given data that is length and time stamp of 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.

<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>Dataset or the reviews</td><td>Classification of the reviews either sail and final lo</td></tr></table></td><td>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.</td><td>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 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. Overall, this work sheds light on the use of machine learning approaches for detecting spam reviews and recommends future research directions.</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 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>--There are no such negative impact on the paper, One such thing is that the Limited resources and privacy Concerns.</td></tr>
</table>

Analyse This Work By Critical Thinking	The Tools That Assessed this Work	What is the Structure of this Paper

The paper shows the importance of considering different strategies and data types when detecting artificial text. The context mainly discusses the use of different machine-learning techniques.	NONE	Abstract I. Introduction II. Literature Search III. Methodology IV. Algorithmic Techniques V. Detecting Parameters VI. Results and Discuss VII. Conclusion
Diagram/Flowchart		



-End of Paper4—

5		
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 & What is 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	**The lexicon-based approach** maintains a word dictionary in two different types which are 1)Dictionary-based approach and 2)Corpus-based approach	This method is more efficient in sentiment calculations by using the sum and mean of sentiment values. It adopts faster .	The dictionary-based approach does not consider the context around the sentiment word thus it leads to less efficiency.

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.	
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.	
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 that allow data refinement.	The paper does not mention any mediating variables. However, the sentiment analysis algorithm itself acts as an intermediary between the independent variables (stages of the sentiment analysis process) and the dependent variable (accuracy performance).

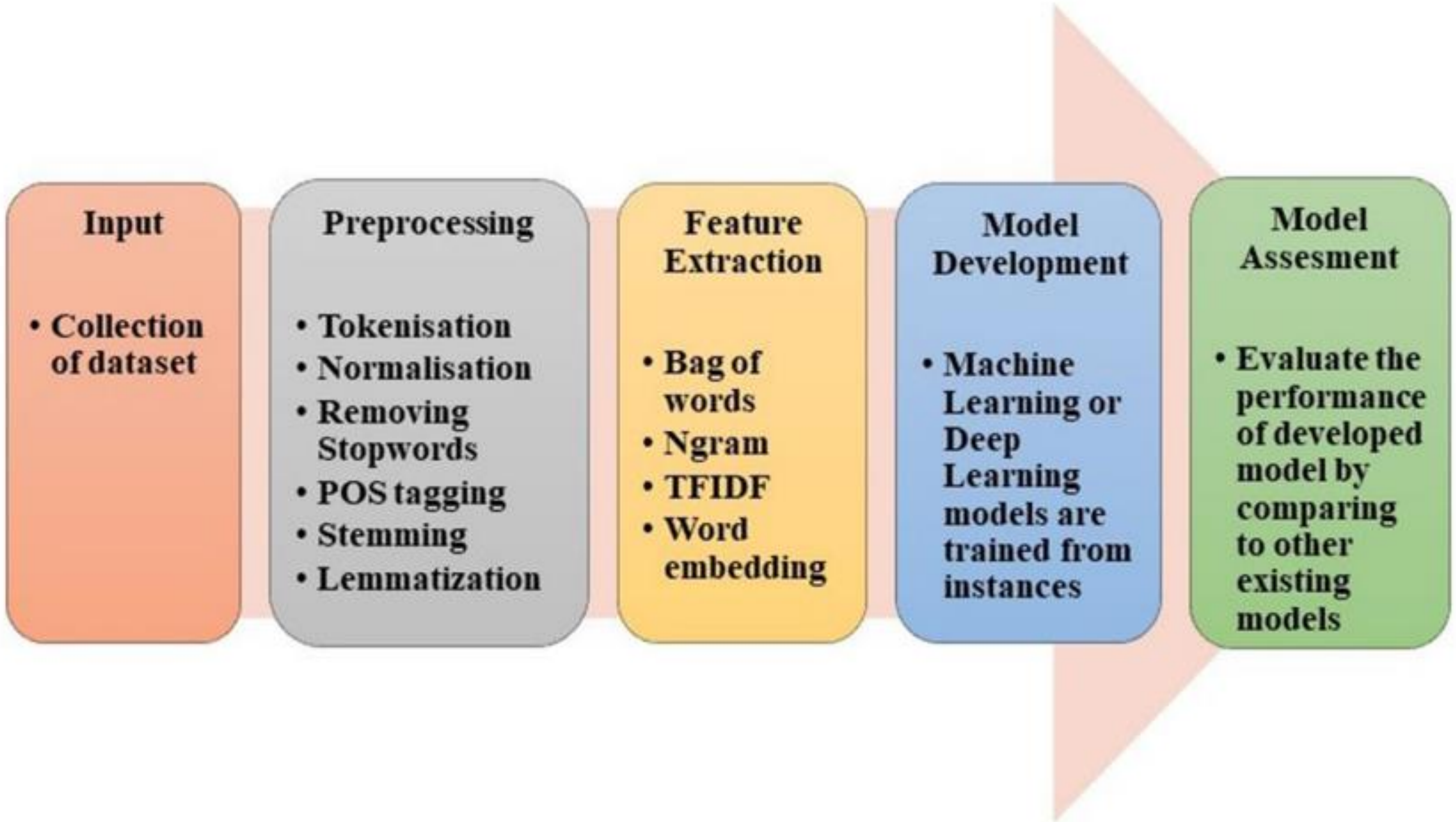
Relationship Among The Above 4 Variables in This article
The accuracy of the model will analyzed with the help of given data after the preprocessing of the data. First the data is preprocessed and the featured extraction is done with moderating variables. The sentiment of reviews depends on the application 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
<table><tr><th>Input</th><th>Output</th></tr><tr><td>Preprocessed dataset as the input to perform accuracy test.</td><td>Accuracy score</td></tr></table>	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

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 I. Introduction **II. Background** **III. Process of sentiment analysis and emotion detection** **IV. Challenges** **V. Conclusions and Suggestions**

Diagram/Flowchart



Fig,3 overall working of the model

--End of Paper 3--

1		
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 • MySql • Python • BeautifulSoup • 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. • Scraping different E-commerce websites. • 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.	-
4	Displaying the information on the user's window.	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

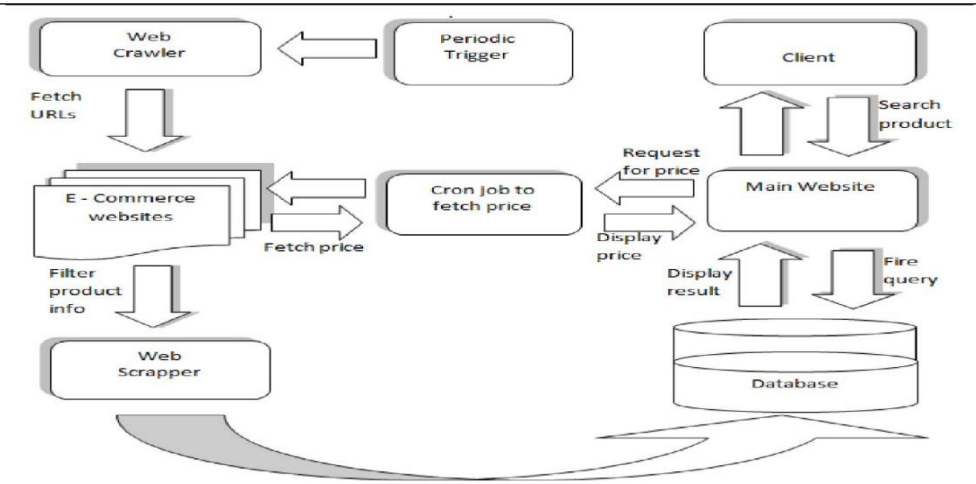
The product details displayed on the user's screen depend on the input product name provided by the user.

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.

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.

<table>
<tr><th colspan="2">Input and Output</th><th colspan="2">Feature of This Solution</th><th>Contribution in This Work</th></tr>
<tr><td colspan="2"><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 colspan="2">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><th colspan="3">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="3">It gives users more alternatives by allowing the visibility of various products on a single website.</td><td colspan="2">--</td></tr>
<tr><th colspan="2">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 colspan="2">By using a single platform, the solution increases the effectiveness and simplicity of comparing products from many websites.</td><td colspan="2">You tube</td><td>Abstract
Introduction
Motivation
System Architecture
Implementation
Results
Conclusion
Future Work
Acknowledgment</td></tr>
</table>

Diagram/Flowchart



SYSTEM ARCHITECTURE

--End of Pape1--

2		
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, productreviews.
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 • Data Collection • Pre-processing and NLP • Feature Labeling. • K-means cluster	To understand K-means clustering along with part-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 parts of speech (nouns, verbs, adjectives, etc.),	limited ability to recognize and interpret sarcasm.

		sentiment analysis algorithms can assign moreaccurate sentiment scores to words.	--
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 Accuracy Precision	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
<table><tr><th>Input</th><th>Output</th></tr><tr><td>Text data collected through the website.</td><td>Classification of reviews based on sentiment analysis</td></tr></table>	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 • Introduction • Related Work • Proposed Method • Implementation • Conclusion • Performance Evaluation • References

Diagram/Flowchart

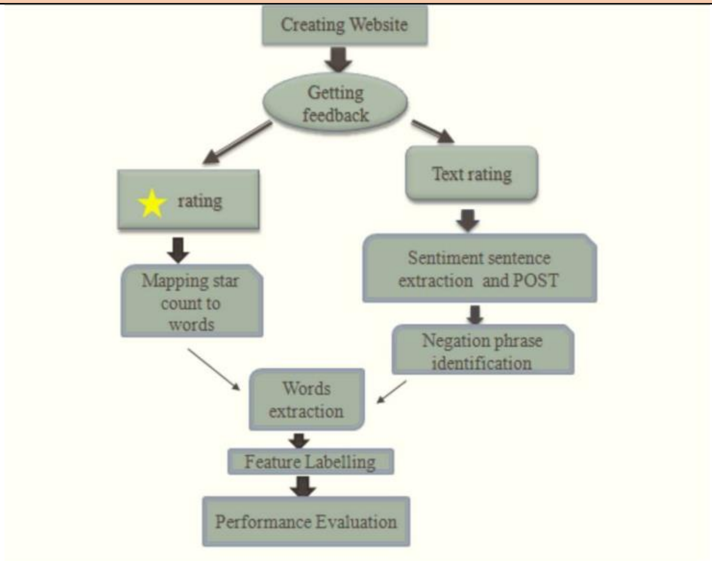



Fig-2 flowchart that depicts proposed process.

--End of Paper 2—

3		
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**	**Keywords in this Reference**
https://ijiis.org/index.php/IJIIS/article/view/15/14	Tri Astuti, Irnawati Pratika	Sentiment Analysis, ANN, Product Review
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?**
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	This paper includes problem identification, data collection, pre-processing, building an ANN model 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	Building a Artificial Neural Networks (ANNs).	Building an Artificial Neural Network (ANN) for sentiment analysis allows for high accuracy in classifying consumer responses.	It may still be challenging to analyze a large volume of data effectively.
2	The Conjugate Scale Gradient Method	This results in an optimal solution more quickly compared to traditional gradient descent methods.	--

Major Impact Factors in this Work

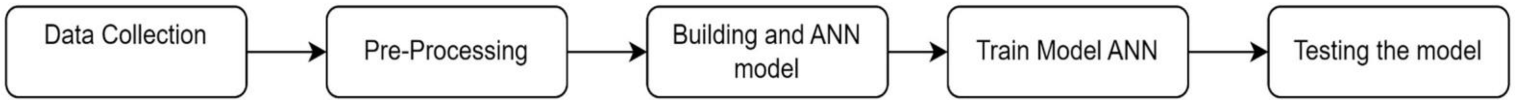
Dependent Variable	Independent Variable	Moderating variable	Mediating (Intervening) variable
Accuracy Score	Data collection, pre-processing, testing performance	Size of data.	Algorithm used.

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. 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 & The Value of This Work
Input	**Output**	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.
Numerical vector, which is obtained from the pre-processing of the product review data.	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 I. Introduction II. Research Concept III. Results and Discussion IV. Conclusions and Suggestions

Diagram/Flowchart


Fig,3 overall working of the model

--End of Paper 3--

4		
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 frequency counting method : is used to detect whether a text is automatically generated by a machine translation system or written/translated by a human. This method	The advantage of this method is that it can identify patterns in word pair distribution to detect artificial text.	The disadvantage is that it requires a large amount of training data and may not be

	analyzes the correlations of neighboring words in the text and measures the degree of "compatibility" of words based on their frequency of occurrence.		effective against more sophisticated generation strategies.
2	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.	--
3	Artificial content detection using lexicographic features:This method uses 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.	It considers the text's linguistic features, including grammatical words, sentence structure, and word length. This can aid in differentiating writing produced by machines and humans.	--

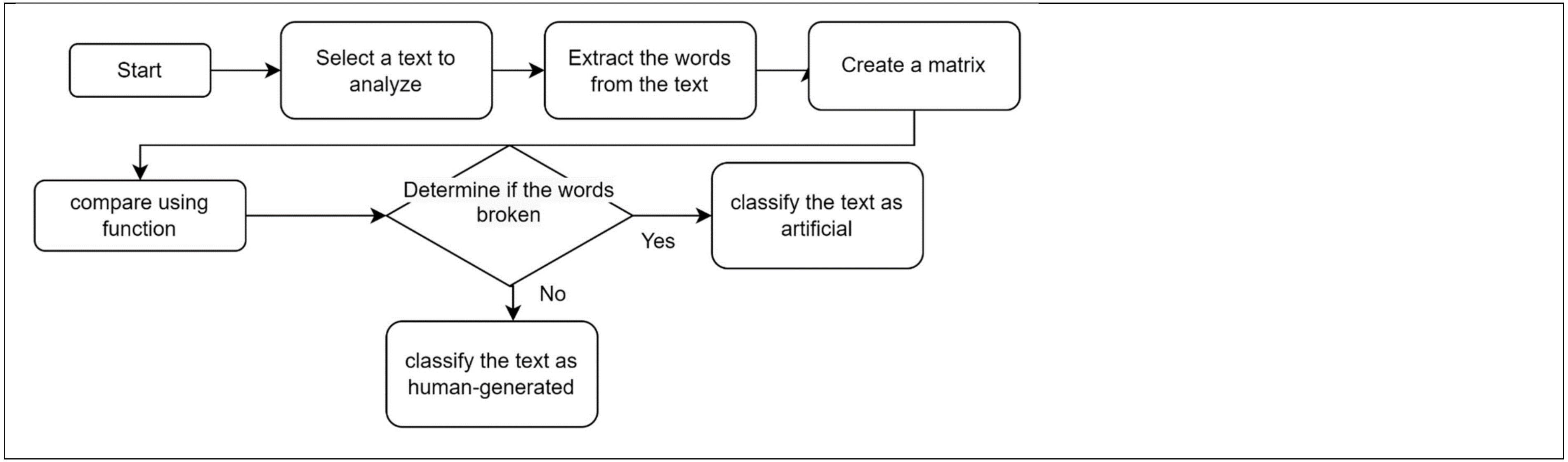
Major Impact Factors in this Work

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. In the "Linguistic features method" , the independent variables include linguistic and perplexity features extracted from text. In the "Artificial content detection using lexicographic features" , independent variables consist of various lexicographic features.	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>Input and Output</th><th colspan="2">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>Sample text</td><td>Categorize the text as either machine-generated or human-generated.</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>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="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 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="2">--</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 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>Abstract
I. Introduction
II. Literature Search
III. The methods of artificial text detection
IV. Choosing A Method
V. Conclusion</td></tr>
<tr><th colspan="4">Diagram/Flowchart</th></tr>
<tr><td colspan="4"></td></tr>
</table>



--End of Paper 4—

5		
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 • Word Segmentation. • Scikit-learn Library. • Naive Bayes Classifier (NBC) • Support Vector Machine (SVM) • Python.	The objective is to assess the effectiveness of various 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 placed on reviews related to Chinese e-commerce.	The accuracy of word segmentation plays a critical role in the effectiveness of Chinese text feature selection and training. This is because in Chinese text, the significance of a paragraph	The drawback of the above point is that in Chinese text, the reliance on accurate word segmentation is essential for understanding the meaning of a paragraph. If the word

		is often determined by a combination of phrases, rather than individual words, which is distinct 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. SVM is particularly convenient when the classification task involves separating data into two classes, making it a favorable choice for 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. SVM is sensitive to the scale of input features. It often requires feature scaling, and improper scaling can impact its performance. SVM may struggle with noisy or overlapping data. 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 large number of reviews or when strict accuracy requirements are not a priority. NBC is more practical for multi-class classification tasks, making it suitable for scenarios where the data needs to be categorized into multiple categories.	NBC assumes that attributes are independent of each other. This independence assumption can lead to suboptimal results when dealing with correlated features. NBC may not capture complex relationships in data as effectively as other

		Despite potential limitations related to attribute independence, NBC is still valuable in the classification of customer reviews on e-commerce platforms due to its speed and suitability for multi-category tasks.	models, particularly when dealing with non-independent features. NBC can perform poorly when dealing with imbalanced datasets where one class has significantly fewer instances than the others.

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 text preprocessing 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, the accuracy of the classification (dependent variable).

The moderating variable, might be the "segmentation ratio" of the training and test data sets. It influences the performance of classification by altering the proportion of test data.

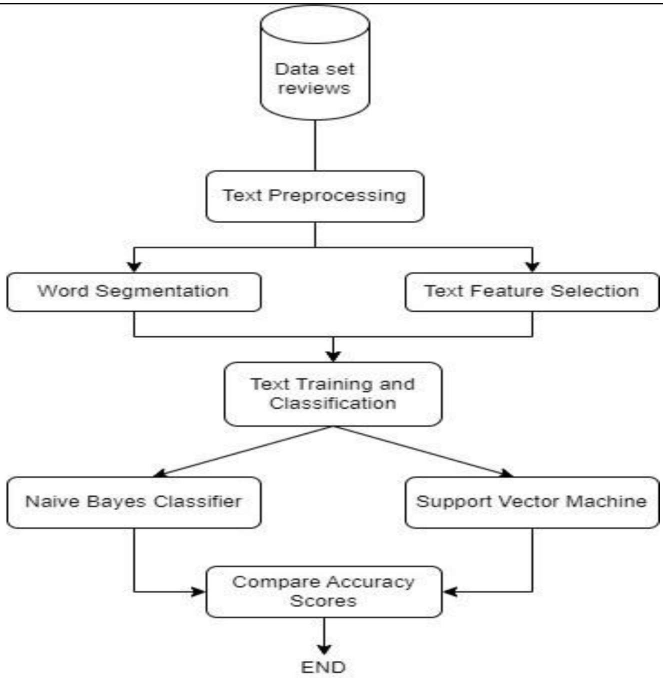
Input and Output	Feature of This Solution	Contribution & The Value of This Work
(see table below)	Offering a customization feature that enables users to select reviews associated with specific sentiments.	From the paper, we've gained knowledge about SVM and NBC, as well as their pros and cons. What should be the optimal size for the training and testing datasets to achieve the highest efficiency.

Input	Output
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 research and 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 understanding of how the size of data plays a crucial role in influencing the performance of different algorithms.	Stack overflow.	Abstract I. Introduction **II. Method** **III. Experiment** **IV. Results and Discussion** **V. Conclusion** and **References**

Diagram/Flowchart



Overall architecture

End of Paper5

1		
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. 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, 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 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	Only movie reviews are collected.

		aligns with its reviews and is considered good.	
2	Data pre-processing • StringToWordvecto. • Attribute selection. • 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 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). 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. Accuracy.	The paper consist of different algorithms such as SVM, NB, KNN-IBK, K-Star, and DT-J48 to determine	The accuracy of the model relies on a series of interconnected steps in the paper's data pre-processing.	In this paper, there are no mediating variables; instead,

	which algorithm is more accurate in classifying the reviews.	Each of these steps builds upon the previous one, collectively contributing to an enhanced model performance.	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.

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

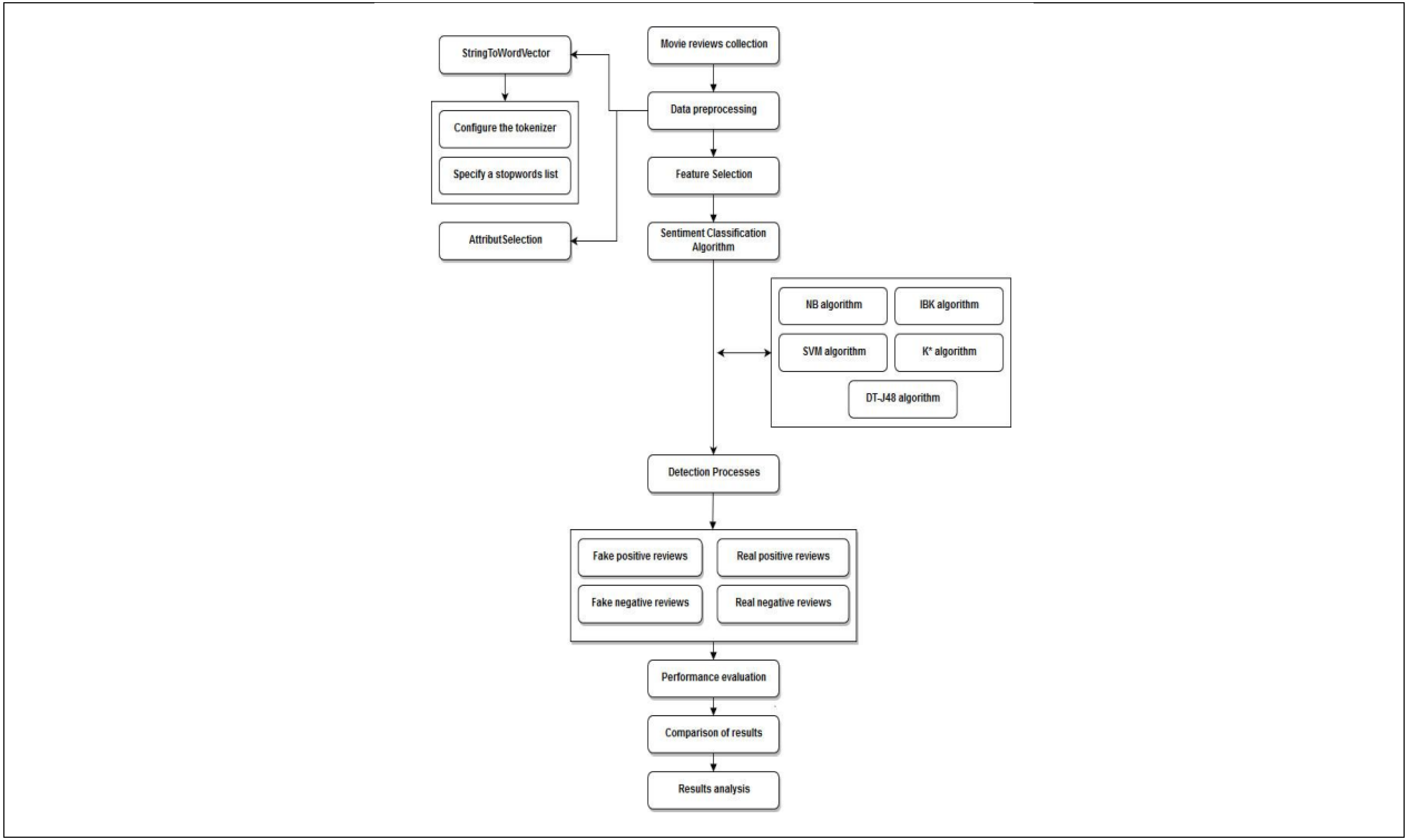
Positive Impact of this Solution in This Project Domain	**Negative Impact of this Solution in This Project Domain**
Sentiment Analysis (SA) has emerged as a subject within text analysis, driven by its potential commercial advantages. Furthermore, the user-generated opinion reviews, categorized as either positive or negative, offer valuable insights for consumers in making product choices.	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.

<table>
<tr><td colspan="2">Improved Accuracy The experiments conducted in this project have shown that sentiment classification algorithms, particularly SVM.</td><td colspan="2">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>
<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>Utilizing supervised algorithms for fake review prediction introduces complexity to the accuracy assessment.

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 colspan="2">Weka tools.

String-To-Word Vector filter in Weka was used for transforming the dataset.</td><td>Abstract

I. Introduction.
II. Related Work.
III. Methodology.
IV. Experiments and results analysis.
V. Conclusion and Future work</td></tr>
<tr><th colspan="4">Diagram/Flowchart</th></tr>
<tr><td colspan="4">Fig 1. Steps and techniques used in sentiment analysis.</td></tr>
</table>


Movie reviews collection
StringToWordVector
Data preprocessing
Configure the tokenizer
Specify a stopwords list
Feature Selection
AttributSelection
Sentiment Classification Algorithm
NB algorithm
IBK algorithm
SVM algorithm
K* algorithm
DT-J48 algorithm
Detection Processes
Fake positive reviews
Real positive reviews
Fake negative reviews
Real negative reviews
Performance evaluation
Comparison of results
Results analysis

--End of Paper 1-

2		
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 . 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 evaluation strategy, and the results obtained from the experiment.
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	The test score and agreement value are determined based on the

		value contributes to enhancing the trustworthiness of reviewers and the product's reliability.	calculations of the product's trustworthiness and reliability values.

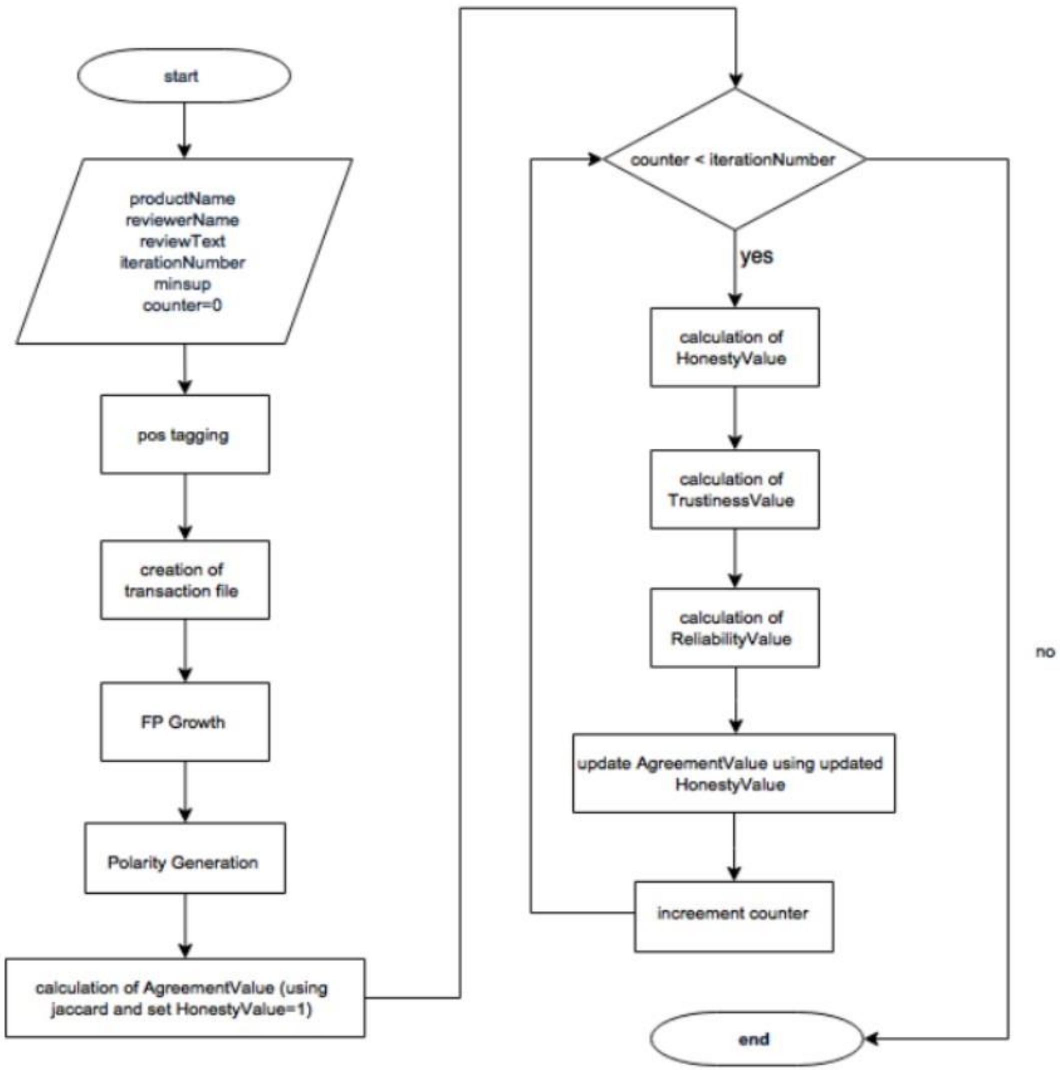
Relationship Among The Above 4 Variables in This article
The iterative process of calculating trustworthiness, honesty, and product reliability is driven by the connection between test scores and agreement values, ultimately resulting in improved model performance.

Input and Output	Feature of This Solution	Contribution & The Value of This Work
Input: product review from Amazon.com from June 1995 - March 2013, data retrieved from https://snap.stanford.edu/data/web-Amazon.html **Output**: It shows accurate method among icf and icf++ based on the calculations.	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.

Input	Output
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

<table>
<tr><td colspan="2">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.</td><td>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.
The specific equations for calculating the trustworthiness and honesty values are not provided in the given document content.</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 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.</td><td>• Iterative Computation Framework (ICF)
• Part-Of-Speech (POS) Tagger
• FPGrowth</td><td>Abstract
I. Introduction.
II. Methodology.
III. Results and Discussions.
IV. Conclusion</td></tr>
<tr><th colspan="3">Diagram/Flowchart</th></tr>
</table>


start
productName
reviewerName
reviewText
iterationNumber
minsup
counter=0
pos tagging
creation of transaction file
FP Growth
Polarity Generation
calculation of AgreementValue (using jaccard and set HonestyValue=1)
counter < iterationNumber
yes
calculation of HonestyValue
calculation of TrustinessValue
calculation of ReliabilityValue
update AgreementValue using updated HonestyValue
increement counter
no
end

Version 2.0 Week 2		
3		
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, Low Cost Publication fees 700, Publish paper in 4 Hr]()	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	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

		routing it to the system controller if it passes the spam check.	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.

Input and Output	**Feature of This Solution**	**Contribution & The Value of This Work**
<table><tr><th>Input</th><th>Output</th></tr><tr><td>Kaggle review datasets</td><td>Spammed and fake reviews. Genuine review.</td></tr></table>	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.	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.

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	The system proposed in the paper employs the system's IP address, potentially giving rise to security concerns for users.

proposed system outlined in this paper, with its efficient spam and fake review detection capabilities, significantly enhances the productivity of the company.	

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 Stop-words, Tf*idf,	Abstract 1)Introduction. 2)Problem definition. 3)Literature survey. 4)Proposed system. 5)Algorithm of proposed system. 6)Conclusions.

Diagram/Flowchart

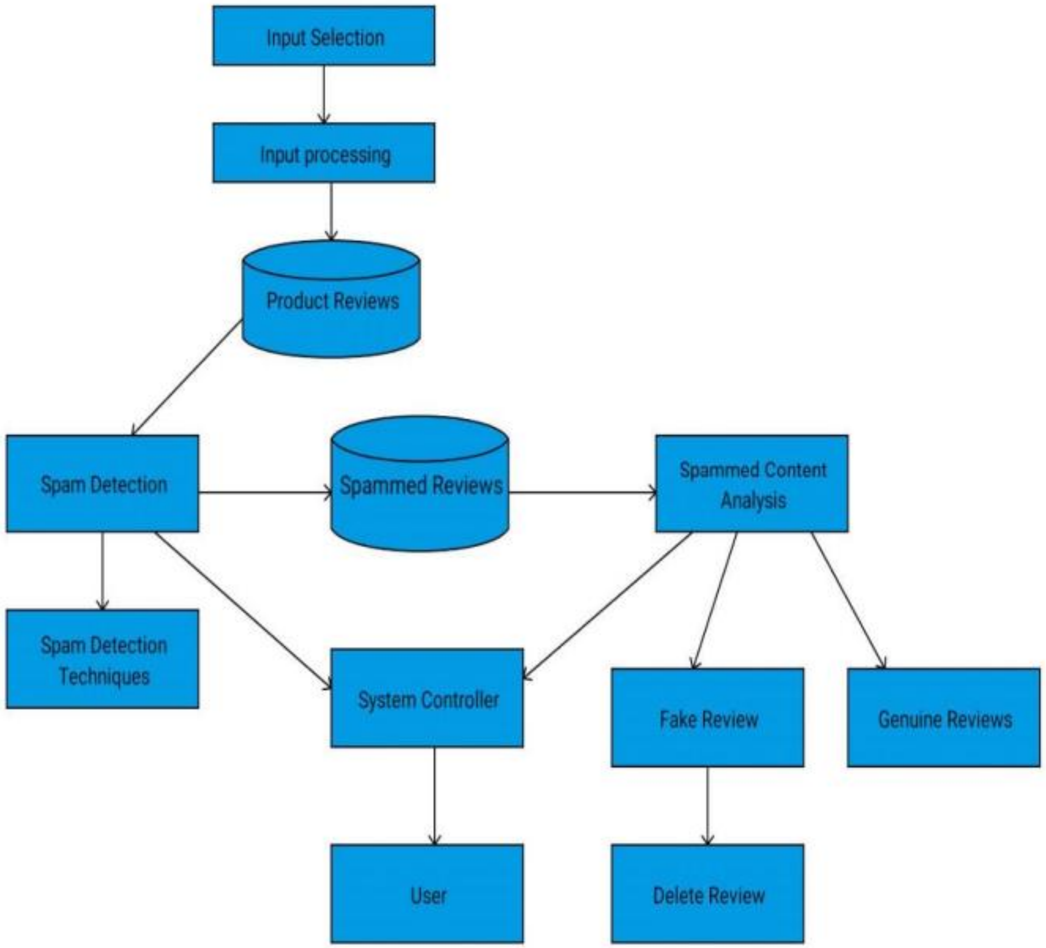


Fig. 1System Architecture

4		
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. 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. 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. LSTM-RNN is recommended to detect fake ratings . 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.		
Input and Output	**Feature of This Solution**	**Contribution in This Work**
<table><tr><th>Input</th><th>Output</th></tr><tr><td>Amazon review datasets.</td><td>Detecting fake online reviews. Detecting fake ratings.</td></tr></table>	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.	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.

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	Data pre-processing, (NLP) tools lemmatization, tokenization.	• Introduction • Related work. • Proposed hybrid deep learning framework. • Experimental Analysis.

variety of activation functions and necessitate extensive training on large datasets to achieve accurate predictions.		• Conclusions.

Diagram/Flowchart

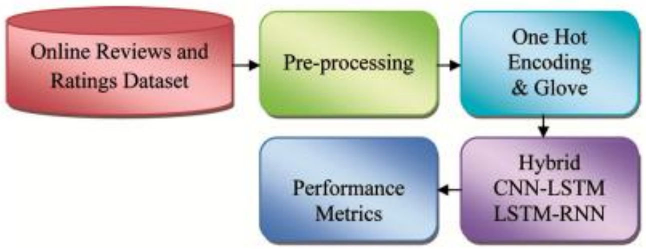


Fig. 2 — Proposed Deep learning hybrid methodology

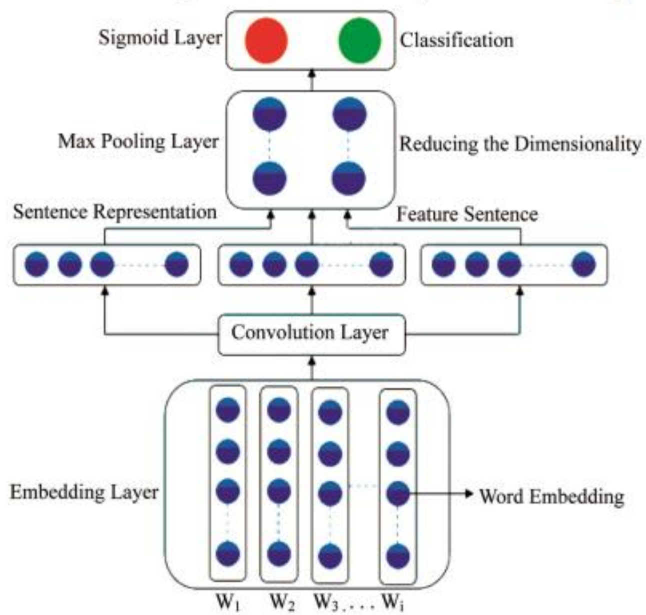


Fig. 3 — An instance of unreliable online reviews described by LSTM-RNN models

5		
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) • k-nearest neighbour. • 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. 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.

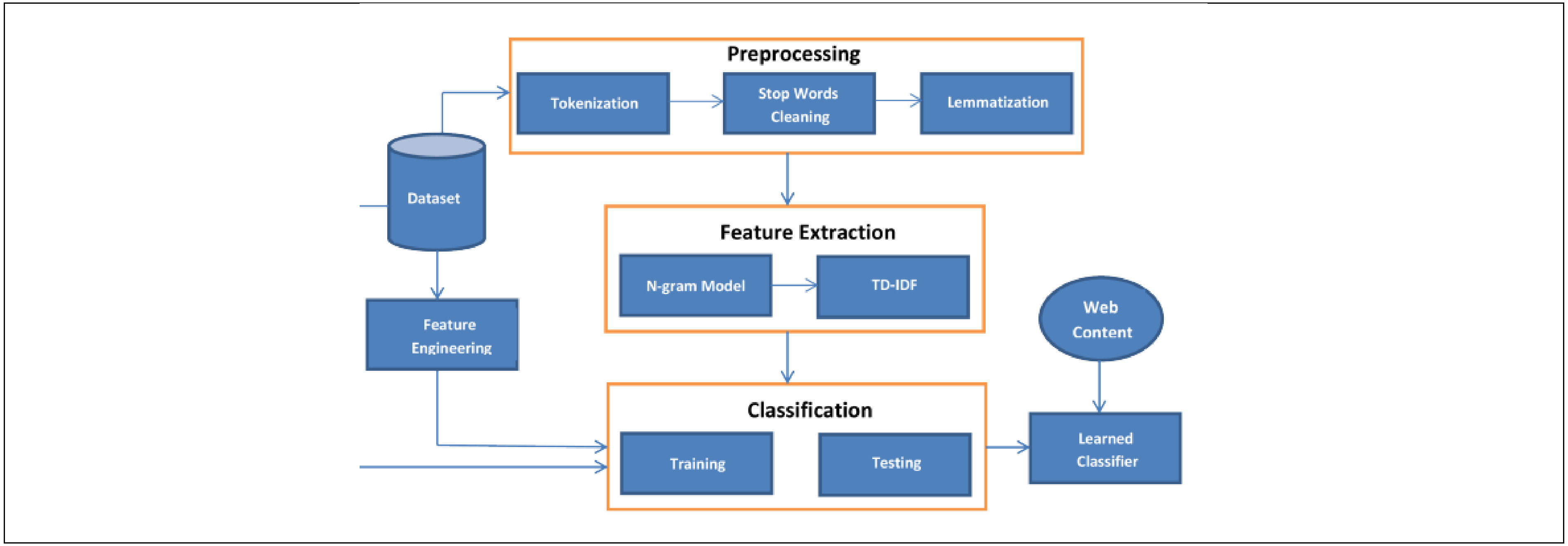
Input and Output	Feature of This Solution	Contribution & The Value of This Work
<table><tr><th>Input</th><th>Output</th></tr><tr><td>User reviews from e-commerce websites.</td><td>To identity the user's review whether it's fake or genuine.</td></tr></table>	• Different classifiers are implemented in the developed approach. • The Bi-gram and Trigram language models are used and compared in the developed approach. • The solution also takes into account aspects related to the reviewers, such as	• 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. • By examining not just the content of the reviews but also the behaviour of the

	the timing of the reviews and their writing styles, to enhance the identification of fake reviews.	reviewers, the suggested approach offers a more thorough analysis.

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

Analyse This Work By Critical Thinking	**The Tools That Assessed this Work**	**What is the Structure of this Paper**
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, increasing the F1-score by 3.80%.	• Nlp tools like stop words, lemmatization • Knn, logistic regression	Abstract 1) Introduction. 2) Related work. 3) Background. 4) proposed approach. 5) Experiment results. 6) Conclusion

Diagram/Flowchart



Preprocessing
Tokenization
Stop Words Cleaning
Lemmatization
Dataset
Feature Extraction
N-gram Model
TD-IDF
Web Content
Feature Engineering
Classification
Training
Testing
Learned Classifier

Work Evaluation Table

<Use the same factors you have used in "Work Evaluation Table" to build your own "Proposed and Previous comparison table ">

Author Name and Year	Work Goal	System's Components	System's Mechanism	Features /Characteristics	Cost	Speed	Security	Performance	Advantages	Limitations /Disadvantages	Platform	Results
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.				Based on the classification algorithm Ex:-SVM	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.		
Eka Dyar Wahyuni . Arif Djunaidy.	To detect fake reviews for a product by using the text and rating property	proposed system (ICF++), the methodology used for fake review detection	Iterative computation framework (ICF).	paper suggests that incorporating semantic aspects, such as sentiment polarity, can improve the accuracy of fake				Performance is Based on the honesty and trustworthy values be	This analysis will determine the review's integrity, the reviewer's credibility, and the product's	specific equations for calculating the trustworthiness and honesty values are not provided in the given document content.		

	.from a review.	, the evaluation strategy, and the results.		review detection.					dependability.			
Mayuri patil, snehalnikumbh.	This method identifies fraudulent transactions by assessing user behaviour and network activity,	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 reputation for financial gain, while also undermining their competitors' products.				Spammed content analysis show the performance	with its efficient spam and fake review detection capabilities, significantly enhances the productivity of the company.	paper employs the system's IP address, potentially giving rise to security concerns for users.		
N Deshani. B Bhaskara ao.	detect fake reviews and what is the main difference	data pre-processing , classifiers, model	Hybrid deep learning methods **(**CNN-LSTM)	CNN-LSTM for detecting fake online reviews, and LSTM-RNN for detecting				**(**CNN-LSTM) (RNN-LSTM)	The identification of counterfeit reviews and	it requires large amount of training		

	between them. Secondary goal is to detect fake ratings.	performa nce.	**(**RNN-LSTM)	fake ratings in the e-commerce domain.					ratings, delivering precise outcomes by considering users' historical experiences with the product	data to predict efficiently.		
Ahmed M.Elmogy, Usman tariq.	a machine learning method for detecting fraudulent reviews.	a machine learning method for detecting fraudulent reviews. Alongside the review feature extraction process.	Supervised algorithms (classifiers, svm, knn)	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.				Svm provide accurate performance	importance of reviews and how they affect almost everything related to web based data. it considers not only the key features of the reviews but also the behaviours of the reviewers.	Unable to distinguish whether reviews have been authored by humans or bots.		