International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: 25 July 2023 Revised: 12 September 2023 Accepted: 30 September 2023

An Ensemble Model-Based Recommendation Approach for Consumer Decision-Making System

Analp Pathak[1, 2], Dr. B. K. Sharma[3]
[1]Research Scholar, Dr A. P. J. Abdul Kalam Technical University, Lucknow
[2]Department of Information Technology, KIET Group of Institutions, Delhi-NCR, Ghaziabad, India
[3]Principal Scientific Officer and Head Computer Science and Engineering Department, NITRA Technical Campus, Ghaziabad

Abstract: A recommendation system can suggest items aligned with diverse user interests by leveraging multiple sources of information. While many recommendation systems heavily rely on the collaborative filtering (CF) approach—where user preference data is combined with others to predict additional items of potential interest—this study introduces a novel weighted recommendation system to enhance consumer decision-making using CF. The methodology includes the development of equations to calculate the weights for both the product and review, as well as to determine the similarity between consumer reviews. To ensemble the model, Random Forest (RF), Support Vector Machine (SVM), and Logistic Regression (LR) are employed in the methodology. The study considers Ensemble Classifiers (RF+SVM+LR) to implement the results, aiming for improved outcomes compared to prior research. The proposed model is trained and tested using an open-source dataset on Kaggle's website. Numerical analysis of the proposed model reveals superior performance, outperforming conventional methods in terms of accuracy (0.821), precision (0.802), recall (0.821), F-measure (0.833), error rate (0.100), and more.

Keywords*: Recommendation System, CF, Random-Forest, Logistic Regression, Support Vector Machine, and Ensemble Classifiers*

1. Introduction

The information landscape impacting online businesses is rapidly evolving, propelled by the advent of the big data era. As an illustration, Amazon encounters an average of 900 million customers daily. Users engaging with e-commerce platforms are now grappling with the challenge of information overload. Recommendation systems come into play to tackle the issues arising from an overwhelming volume of data. These systems analyse data collected from users' prior interactions to discern their requirements and preferences, subsequently aiding them in making informed decisions about suitable options. Researchers have significantly advanced most research endeavours focused on recommendation systems in recent years. [1][2][3]. One highly effective approach to navigate this overwhelming information is to rely on the recommendations of individuals who, like the average internet user, encounter substantial amounts of data daily. Recommendations can manifest in various forms, such as verbal endorsements, letters of recommendation, media reports, public surveys, travel guides, website reviews, and more. Over the past 15 years, numerous prominent online platforms have integrated recommendation systems to facilitate this inherent social process. The primary objective of these systems is to assist consumers in discovering the most pertinent and valuable information amidst the extensive array of online material. This includes, but is not limited to, news articles, web pages, images, and other content. [4].

1.1 Recommendation System

A recommendation system is a digital tool designed to assist consumers in discovering the most valuable products or services, leveraging insights from the customers' past preferences or tastes. These preferences are often derived from the users' purchase histories. With the increasing prominence of online business, the recommender system emerges as a crucial instrument for executing personalised marketing strategies. A thoughtfully crafted recommender system analyses each consumer's preferences, whether implicitly inferred or explicitly expressed, automatically presenting a curated selection of items or services tailored to individual tastes[5]. Utilising a parallel strategy for generating recommendations offers numerous advantages, with one notable benefit being the ability to deliver results promptly. The parallel execution of algorithms streamlines the process, facilitating output generation with optimal efficiency without compromising performance. These benefits include:

- Efficient Handling of Large Volumes of Data: Parallel processing allows for the swift and efficient handling of substantial data volumes, contributing to increased overall efficiency.
- Diverse Suggestions for a Wide Range of Item Types: The parallel strategy enables the recommendation system to suggest a broad spectrum of item types, enhancing its versatility and applicability.
- Ease of Conversion to Parallel Processing: Converting an existing method to parallel processing is a straightforward process, simplifying the implementation of this strategy [6].

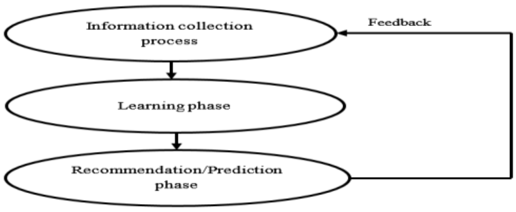


Figure 1. Recommendation System.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

In the realm of making recommendations, two primary perspectives exist:
(a) Content-Based Approach: This approach involves recognising users by attempting to identify their key characteristics, relying on individual information that may be challenging to terminate.
(b) Collaborative Filtering (CF): CF takes advantage of the notion that individuals with common interests in the past are likely to agree on one's tastes in the future [5].

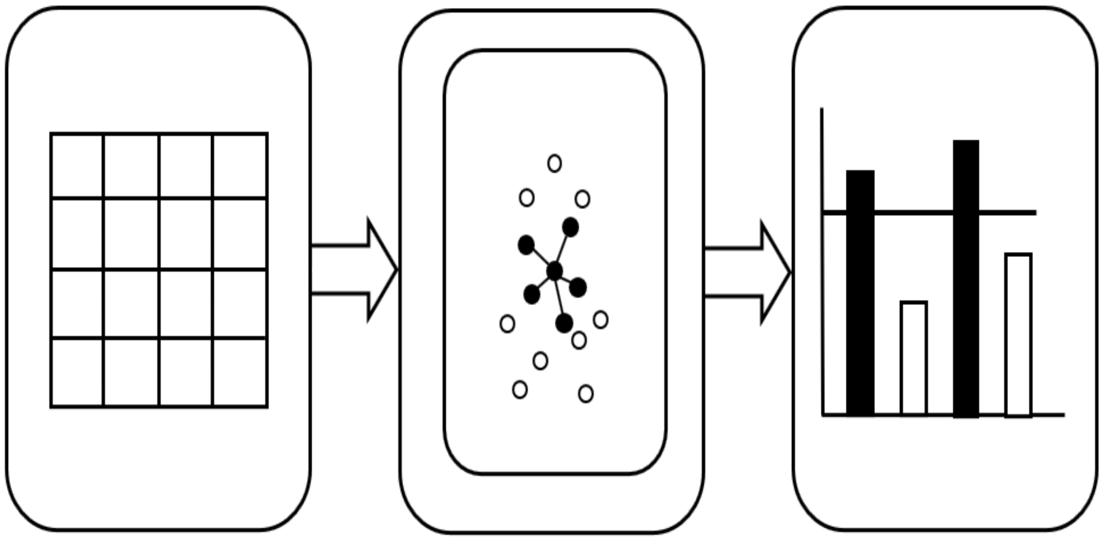
1.2 Collaborative Filtering

Collaborative Filtering (CF) has emerged as a pivotal method in developing personalised recommender systems due to its precision and scalability. The fundamental purpose of CF is to infer users' preferences based on the activity data of both the users themselves and others [7]. It utilises user ratings and comments to decide whether to present specific content to a particular user. CF sifts through a vast user database to identify individuals with preferences similar to the current user's.
There are two primary classes of recommendation strategies: content-based recommendation and CF recommendation approaches, with CF being the more prevalent method. CF can be divided into two main categories—those based on the items themselves and those based on the people who use them.
User-Based CF: The objective here is to identify a group of users with similar interests to the current user.
Item-Based CF: This approach focuses on item similarity, often estimated by analysing user behaviour. However, challenges such as cold starts and data sparsity are associated with both informational "cold starts" and conventional methods' lack of detail. The risk may be mitigated with the accumulation of more diverse data from the internet, including text, images, tag data, specialised demand data, and detailed project information.

CF methods use a person's past choices and the opinions of individuals with similar tastes to suggest new items or assess the usefulness of items like hotels to the user. In a typical CF setup, there are m user lists (U = {$u_1, u_2, ..., u_m$}) and n item lists (I = {$i_1, i_2, ..., i_n$}). The set of items on which each user u_i has commented is designated by I_{ui}. Ratings, usually on a numerical scale, express users' thoughts, and these can also be derived from user behaviour such as purchasing patterns, time spent on the site, and connection behaviour. It's important to note that if $I_{ui} \subseteq I$, then I_{ui} might be a collection of zero elements. For a key user, $U_a \subseteq I$, also referred to as the active agent, the goal of a CF process is to determine the likelihood of an item, and this likelihood may take one of two forms. [8].

Recommendation: It provides a list of the top N hotels according to the user's overall satisfaction with each establishment $I_r \subset I$. $I_r \subseteq I_{u_a} = \phi$ must include on the suggested list both things and hotels that the present consumer has not previously bought. In certain circles, it is also referred to as a Top-N recommendation method interface [9].

Prediction: $P_{a,j}$ is a numeric number that reflects the anticipated possibility of the item $i_j \notin I_{u_a}$ for the active user u_a, depending on the user's behavior. This value is expressed as a percentage, and it may range from 0 to 1. The information that was supplied by u_a indicates that this forecasted number is somewhere within the same range of 1 to 5 [8].
The CF procedure is shown in Figure 2. It begins with the inputting of data, which is followed by the setting of parameters according to the format of the neighbours in the third stage. In the third stage, it makes suggestions for the production of brand-new things.



Rate input data　　Neighbours format　　Generation Recommended

Figure 2. The CF process [9].

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: 25 July 2023 Revised: 12 September 2023 Accepted: 30 September 2023

2. Related Work

In this section some related work based on a recommendation system using a collaborative filter is discussed below:

Zhao [10] This paper introduces a hybrid recommendation system by combining the traditional Collaborative Filtering (CF) approach with a subsystem based on cluster analysis using a genetic algorithm. The conceptualisation and implementation of this system are detailed in the paper. Notably, the system employs a unique strategy where only users belonging to the clusters within the user base are considered when determining the nearest neighbours. As a result, the system transforms into a Genetic Clustering-based CF Recommender System.

Key highlights and findings from various experiments include:

Performance Improvement: The hybrid system is designed to enhance the traditional CF recommendation system by incorporating genetic clustering. The genetic clustering-based CF recommender system demonstrates improved response times compared to the classic CF recommendation system.

Cluster-Based Nearest Neighbours: The system focuses on identifying nearest neighbours only within the clusters formed by the genetic algorithm. This approach adds a layer of sophistication to the recommendation process, potentially improving the accuracy and relevance of recommendations.

Scalability Advantage: Through experiments, it is observed that the response time of the classic CF recommendation system scales linearly with the number of consumers.

In contrast, the response time of the genetic clustering-based CF recommendation system remains constant, irrespective of the size of the user base.

This innovative combination of genetic clustering and collaborative filtering not only improves recommendation accuracy but also demonstrates efficiency gains in terms of response times, particularly as the user base grows.

Sharma et. al., [11] The proposed hybrid recommendation system employs a prediction algorithm to generate book suggestions. The system's three stages involve a combination of two types of filtering: content-based filtering and Collaborative Filtering (CF). Here is a breakdown of the system's stages:

User Profile Comparison (Collaborative Filtering): In the first stage, the system compares the current user's profile to a database of users to identify individuals with similar characteristics. Collaborative filtering is utilised to find users with comparable preferences and behaviours.

Item Selection for Comparable Users (Content-Based Filtering + Collaborative Filtering): In the second stage, the system utilises both the user's profile and the contents of the items to select candidate items for each user identified in the first stage.

This stage involves a combination of content-based filtering and collaborative filtering, leveraging user profiles and item characteristics.

Prediction Value Calculation (Collaborative Filtering): Once the candidate items are identified, the Resnick prediction equation is employed to calculate the prediction value for each item. The prediction values serve as the basis for recommending items to the end user.

Evaluation of Hybrid Filtering Strategy: The proposed system's performance is assessed by comparing it with state-of-the-art recommendation techniques, including traditional Collaborative Filtering and content-based filtering.

Experimental results demonstrate the superiority of the suggested hybrid filtering strategy over both traditional Collaborative Filtering and content-based filtering.

In summary, the hybrid recommendation system effectively combines Collaborative Filtering and content-based filtering to provide book suggestions. The system's multi-stage approach leverages user profiles, item contents, and prediction equations to offer personalised and accurate recommendations, outperforming traditional recommendation techniques in experimental evaluations.

Tahira et. al., (2022) [12] The developed recommender system operates within the context of the Internet of Things (IoT) and utilises online customer evaluations to match product characteristics significant to the buyer. The algorithm employed follows a specific process:

Product Analysis: The algorithm analyses the product to identify its key characteristics that hold significance for the consumer.

Aspect-Based Sentiment Classification: After identifying important product aspects, the system conducts aspect-based sentiment classification. It locates these aspects in customer evaluations and assigns a sentiment score to each of them, indicating the sentiment expressed in reviews.

Weighted Evaluation Based on Credibility: The credibility of each customer review is considered in the algorithm, determining the weight assigned to each review. Weighted evaluations contribute to a more nuanced understanding of customer sentiments.

Experimental Research: Experimental research is conducted to investigate how the algorithm's impact varies with different types of items. The focus is on understanding how the recommender system performs with hedonistic and utilitarian items, recognising that the product's nature influences recommender systems' effectiveness.

The overarching goal is to tailor recommendations based on product characteristics and consumer sentiments expressed in reviews. The consideration of credibility and the experimental exploration of different product types contribute to a more context-aware and effective recommender system within the Internet of Things environment.

Han et al. [13] introduce the Multilayer Fuzzy Perception Similarity (MFPS) algorithm, aiming to perceive and interpret user similarities to enhance the subjective quality of recommendations. Notably, this study marks the first application of triangular fuzzy numbers in Recommender Systems (RS). The key features of the MFPS algorithm are:

Triangular Fuzzy Numbers: The study pioneers using triangular fuzzy numbers in Recommender Systems, providing a novel approach to handling uncertainty and subjectivity in user preferences.

Enhancement of Similarity Techniques: A selection of cutting-edge similarity techniques is improved nonlinearly to imbue them with the capability to understand and incorporate human emotions.
This enhancement contributes to the algorithm's ability to capture subjective aspects of user preferences.
Layered Structure for Subjective Perception: The algorithm employs a layered structure designed to enhance the RS's capacity to perceive similarities across user qualities subjectively. This multi-layered approach aims to provide a nuanced understanding of user preferences and similarities.
Effectiveness and Consistency: The results of the studies demonstrate that MFPS outperforms other competitor baselines in terms of effectiveness, explicability, and consistency of the suggestions it provides. The algorithm's capability to consider subjective factors in user similarities contributes to its success in delivering high-quality recommendations.
The MFPS algorithm introduces a novel approach to user similarity perception in Recommender Systems, leveraging triangular fuzzy numbers and a multilayered structure. The demonstrated effectiveness, explicability, and consistency in providing recommendations highlight the algorithm's potential to enhance the subjective quality of suggestions.

Mohd Sabri and Nurul [14] designed and evaluated a book recommendation system using the item-based Collaborative Filtering (CF) method. The recommendation system successfully forecasted book suggestions with an acceptable F-measure of 80.38%. Collaborative filtering, particularly the item-based approach, effectively generated accurate book recommendations. Key Points are as follows:
Effectiveness of CF: Collaborative Filtering (CF) is highlighted as a widely modified and extensively utilised strategy in recommendation systems. The study focuses on the item-based CF approach, showcasing its effectiveness in accurately forecasting book suggestions.
Dataset and Cross-Validation: The dataset for this investigation was obtained from the Kaggle website, where it had undergone the rigorous 10-fold cross-validation method. A random selection of one thousand data points was made, with nine hundred used for training and one hundred reserved for testing.
Evaluation Metrics: Precision, Recall, and F-measure were employed as metrics to assess the performance of the book recommendation prototype. These metrics comprehensively evaluate the system's ability to make accurate and relevant book suggestions.
Performance Metrics: The book recommender system demonstrated satisfactory performance, as indicated by an F-measure value of 80.38%. This suggests that the system effectively balances precision and recall in recommending books.
In conclusion, the study presents a successful implementation of a book recommendation system using item-based Collaborative Filtering. The choice of evaluation metrics and the achieved F-measure value of 80.38% indicate the system's ability to provide accurate and meaningful book suggestions.

Bi et al. [15] devised a recommendation system based on deep neural networks, leveraging various features such as item average rating, user primary data (gender, age, profession, user ID), and item basic data (name, category, ID). The algorithm uses deep neural networks to construct a regression model predicting users' evaluations. The fundamental concept involves using four distinct types of neural network layers to build user and item feature matrices using user and item data, respectively. Three tests were conducted with data directly sourced from the Movie Lens website to validate the effectiveness of the proposed method. Key Components and Findings are as follows:
Feature Utilisation: The recommendation system incorporates various features, including item average rating, user basic data (gender, age, profession, user ID), and item basic data (name, category, ID).
These features contribute to constructing user and item feature matrices for predicting user evaluations.
Deep Neural Networks: The algorithm utilises deep neural networks to create a regression model capable of anticipating user evaluations. The use of multiple layers in the neural network facilitates the extraction of complex patterns from user and item data.
Evaluation and Testing: The proposed method undergoes three tests using data directly obtained from the Movie Lens website. The tests are designed to validate the system's effectiveness compared to state-of-the-art Collaborative Filtering (CF) recommendation algorithms.
Addressing Challenges: Experiments demonstrate that the proposed method not only outperforms existing CF recommendation algorithms but also addresses challenges such as data sparsity and the cold-start problem. These findings suggest that the deep neural network-based approach is effective in handling common issues encountered in recommendation systems.
In summary, the recommendation system based on deep neural networks showcases the effective integration of various features and the application of advanced neural network architectures. The method's success is demonstrated through testing, surpassing existing CF algorithms and addressing challenges associated with data sparsity and the cold-start problem.

Chen et al. [16] devised a recommendation system based on deep neural networks, leveraging various features such as item average rating, user basic data (gender, age, profession, user ID), and item basic data (name, category, ID). The algorithm utilises deep neural networks to construct a regression model that predicts users' evaluations. The fundamental concept involves the use of four distinct types of neural network layers to build user feature matrices and item feature matrices using user and item data, respectively. Three tests were conducted with data directly sourced from the Movie Lens website to validate the effectiveness of the proposed method. Key Components and Findings are as follows:
Feature Utilisation: The recommendation system incorporates various features, including item average rating, user basic data

(gender, age, profession, user ID), and item basic data (name, category, ID). These features contribute to constructing user and item feature matrices for predicting user evaluations.

Deep Neural Networks: The algorithm utilises deep neural networks to create a regression model capable of anticipating user evaluations. The use of multiple layers in the neural network facilitates the extraction of complex patterns from user and item data.

Evaluation and Testing: The proposed method undergoes three tests using data directly obtained from the Movie Lens website. The tests are designed to validate the system's effectiveness in comparison to state-of-the-art Collaborative Filtering (CF) recommendation algorithms.

Addressing Challenges: Experiments demonstrate that the proposed method not only outperforms existing CF recommendation algorithms but also addresses challenges such as data sparsity and the cold-start problem.

These findings suggest that the deep neural network-based approach is effective in handling common issues encountered in recommendation systems.

In summary, the recommendation system based on deep neural networks showcases the effective integration of various features and the application of advanced neural network architectures. The method's success is demonstrated through testing, surpassing existing CF algorithms and addressing challenges associated with data sparsity and the cold-start problem.

Lai et. al., [17] presents an innovative approach to rating prediction by employing a deep learning model with semantic components based on attention-based gated recurrent units (GRUs). The proposed two-step process focuses on extracting feature aspects from user preferences, integrating word attention techniques with review semantics. Key Steps and Findings are as follows:

Bidirectional GRU for Key Phrases Extraction: The first step involves using a bidirectional GRU neural network to extract key phrases from user evaluations.

The network utilises the word attention technique to identify significant phrases in user reviews.

Aspect-Based Attention Semantic Vectors: In the second step, user reviews are analysed into individual words, employing Latent Dirichlet Allocation (LDA) and attention weights of selected words.

This process generates aspect-based attention semantic vectors from the reviews, capturing the semantic nuances of user preferences.

Prediction Using XGBoost: The aspect-based attention semantic vectors are combined with the Extreme Gradient Boosting (XGBoost) technique for predicting user preference ratings. XGBoost is a gradient-boosting algorithm known for its high performance and efficiency.

Experimental Results and Effectiveness: Experimental results demonstrate the effectiveness of the proposed strategy in achieving higher prediction accuracy compared to conventional methods. The semantic components and attention mechanisms contribute to a more nuanced understanding of user preferences, leading to improved predictions.

The study showcases a sophisticated approach to rating prediction, leveraging deep learning models with attention-based GRUs and semantic components. The combination of bidirectional GRUs, attention mechanisms, and XGBoost demonstrates improved accuracy in predicting user preferences compared to traditional methods.

3. Background Study

The widespread use of social networking sites notably fuels the growing interest in review-based recommender systems. This surge in interest stems from the desire to leverage the valuable insights embedded in users' written reviews. In response, the current study introduces a Collaborative Filtering (CF) recommendation system enhanced by sentiment analysis.

To achieve this, sentiment analysis is applied to a dataset comprising 7,210 reviews of 221 novels from the Amazon website. User feedback is gathered through an ensemble of models, employing a weighted vote classifier approach for ensemble modelling. Java Web Crawlers extract necessary information from Amazon.com, focusing on customer comments about individual book titles like "Business Intelligence."

Various approaches, including text normalisation and ensemble techniques, are employed for sentiment analysis. The study reveals that incorporating sentiment analysis into recommender systems enhances their performance, particularly in suggesting popular items. This sentiment analysis integration influences users' likelihood to recommend items positively.

By demonstrating significant gains in the effectiveness of recommender systems through the incorporation of sentiment analysis, this research underscores the potential impact and relevance of leveraging user sentiments for improved recommendations. [18].

4. Problem Formulation

Customers, equipped with an Internet connection, now can shop for necessities at any time and from any location of their choice. Many individuals opt for online retail platforms such as Amazon for their product purchases. One distinctive feature of these platforms is the ability for users to provide written feedback on products, potentially influencing the purchasing decisions of other consumers. However, the sheer volume of data generated by these online interactions poses a challenge. Extracting meaningful information from such a massive dataset becomes challenging when the content of documents is still being determined. Users require assistance formulating relevant queries, categorising relevant documents, and identifying patterns within the data. The rapid proliferation and diversification of information on the internet, coupled with the continuous introduction of new e-commerce products (including purchasing items, product comparisons, various auctions, etc.), often overwhelms customers. This information overload can lead to suboptimal decisions and judgments. Significantly, this phenomenon negatively impacts profits, as customers may need help navigating the vast array of options and making informed choices.

5. Research Objectives

The objectives of the study can be summarised as follows:

- Review and Evaluate Previous Studies:
- Investigate and analyse existing research on recommendation systems for consumer decisions.
- Assess the methodologies, strengths, and limitations of previous studies in this domain.
- Develop a Novel Recommendation Technique:
- Formulate and implement a new recommendation technique by introducing an innovative formula.
- Utilise Collaborative Filtering (CF) to enhance consumer decision-making.
- Demonstrate Model Robustness:
- Establish the robustness of the proposed recommendation model through rigorous evaluation.
- Conduct a comparative analysis with a conventional model to assess accuracy and other key performance parameters.
- Provide empirical evidence and insights into the effectiveness of the novel recommendation technique.

These objectives collectively aim to contribute to the field of recommendation systems, offering a novel approach to enhance consumer decisions and providing a thorough evaluation framework for the proposed model.

6. Research Methodology

The examination of the designed architecture is a focal point in the research methodology context. In the realm of research methodology, the term "research methodology" encompasses the systematic process by which authors articulate the specifics of how they intend to conduct their studies. This process is integral to ensuring a reasonable and systematic approach to addressing a research problem. Authors commonly provide a concise elucidation of their chosen methodology to ensure precision, reliability, and the accomplishment of stated goals and objectives in their research. This methodological approach extends beyond the mere consideration of data itself; it encompasses the origins of the data, potential applications, and the methods employed for data acquisition. Following this comprehensive methodological framework, ensemble models are employed to amalgamate various components of the methodology. Subsequently, a recommendation system is modelled, leveraging the insights derived from the ensembled methodology, with the ultimate aim of enhancing consumer decision-making.

6.1 Technique Used

In this section, a brief description of all the techniques which are taken into consideration is given below:

Ensemble Classifiers: Ensemble Methods play a pivotal role in constructing the sentiment analysis model in this study. The classification of comments is achieved through an ensemble modelling technique that incorporates various classification methods. Ensembles, at a meta-algorithmic level, amalgamate insights from multiple intelligent models into a unified prediction algorithm. The primary objective of ensemble algorithms is to enhance overall performance by leveraging the collective capabilities of several weak learners. Different ensemble approaches, such as bagging, boosting, and stacking, bring unique focuses to the task at hand. Bagging aims to reduce variance, boosting seeks to alleviate bias, and stacking strives to elevate prediction accuracy. The core principle of ensemble techniques involves the amalgamation of multiple classifiers to achieve superior results compared to any individual classifier operating alone. In this study, various models, including Random Forest (RF), Support Vector Machine (SVM), and Logistic Regression, are employed as part of the ensemble technique. These algorithms contribute to the prediction phase of supervised learning, collectively enhancing the sentiment analysis model's robustness and effectiveness.

Random Forest is a powerful ensemble learning method introduced by Leo Breiman. It comprises a series of unpruned regression or classification trees, each constructed using a random subset of samples from the training data. The key idea behind Random Forest is to introduce randomness at various stages of the model-building process to enhance robustness and reduce overfitting. In the induction process of each tree within the Random Forest:

- Random subsets of features are chosen during the tree-building process.
- These subsets, also known as feature subsets, introduce diversity among the trees.
- Each tree is constructed independently.

After the individual trees are built, the predictions from the ensemble are aggregated to make the final prediction. For classification tasks, a majority vote is often used, while for regression tasks, the predictions are typically averaged. This aggregation process further contributes to the model's generalisation and predictive performance. Random Forest's ability to handle high-dimensional data, reduce variance, and mitigate overfitting makes it a popular and effective choice in various machine learning applications. Its robustness and capacity to deal with diverse datasets contribute to its widespread use in both academic research and practical implementations. Each tree is nurtured in accordance with the construction process of a Random Forest involves the following steps:

- Bootstrapped Sampling: Choose N examples at random from the original data with replacement, creating a bootstrapped sample. This bootstrapped sample serves as the training set for building a tree.

- Feature Randomisation: For each node in the tree, specify a parameter, m, such that m << M, where M is the total number of input variables. Choose m variables at random from the M available variables.
- Best Split Selection: Utilise the m randomly selected variables to find the best split for the node. This split is used to divide the node into child nodes.
- Tree Growth Process: Maintain the value of m throughout the growth process. Maximise the potential size of each tree during growth. No pruning is performed during the entire process.
- Performance and Generalisation: Random Forest typically demonstrates a significant performance improvement compared to single-tree classifiers like C4.5. The generalisation error rate is favorable when compared to Adaboost, and Random Forest is often more robust to noise.

Figure 3 depicts a block diagram of the Random Forest methodology, showcasing the various stages involved in its construction and the overall ensemble learning process[19]. This approach, characterised by bootstrapped sampling and feature randomisation, contributes to the robustness and effectiveness of Random Forest in handling diverse datasets.

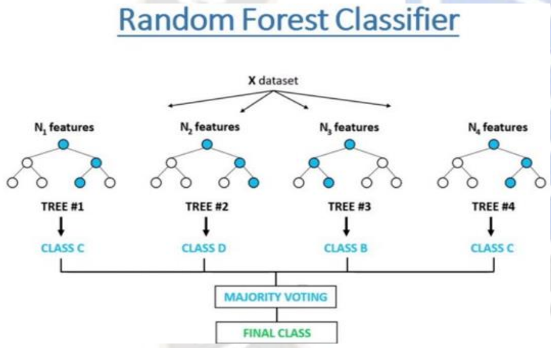


Figure 3. Random Forest [20].

Support Vector Machine: It combines supervised regression and classification learning methods. As a result, a higher dividing hyperplane can be constructed by projecting the input vector to a higher-dimensional space [21]. Figure 4, an SVM trained on instances from 2 has maximum edge hyperplane classes.

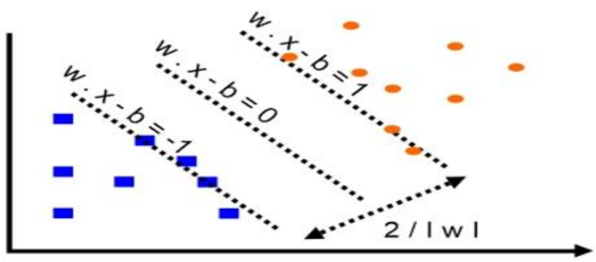


Figure 4. SVM [21].

Logistic Regression: The LR method is widely used for linear classification. It enables the formation of a multivariate regression by allowing a connection to arise between a variable that is independent and dependent variables. LR is the framework of multivariate analysis that may be used to forecast the existence or absence of a function or consequences based on the values of several different predictor variables in a series. This can be helpful in several different contexts [22].

$$\text{Log} = \left[\frac{p}{1-p}\right] = \beta_0 + \beta(\text{Age}) \quad (1)$$

Where p represents the probability, and β_0 indicates the intercept value. With the assistance of a line of regression, the LR splits the data into two distinct categories,
as seen in Figure 5 [22].

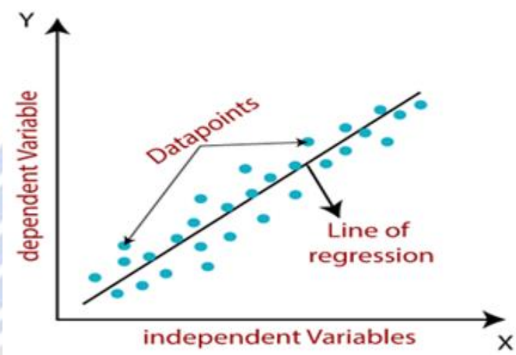


Figure 5. Logistics Regression [22].

6.2 Proposed Methodology

The architecture of the proposed research work is shown in Figure 6, and the work process of this methodology is given in the steps below.

Step 1: Data Collection

Data collection is the first step of this work. In this step, data is collected from the websites regarding users (U_1, U_2, U_3, … U_n), products (P_1, P_2, P_3, … P_n), and reviews of the users. The reviews considered are given for the same products by all the users or customers.

Step 2: Pre-processing of the Data

After collecting data, pre-processing of this data is done. Pre-processing of the data is the most crucial step. In the methodology, pre-processing is done using case folding, tokenisation, and replacement of noisy data to enhance the methodology's performance.

Step 3: Concatenation of the Reviews and Products

In this step 3, after pre-processing the data, the concatenation of the reviews and products is done. Products and the reviews of the products given by several users are linked.

Step 4: Calculate the Weight for both Product and Review

After the concatenation of the product and reviews in the previous step, the weight is calculated for products and reviews of the user by using a newly created formula as given below:

$$W^{P}_{(u_p)} = R_{P_{U_i}} / \sum_{U_i \in U}^{n} R_P, \quad (2)$$

Where,
$W^{P}_{(u_p)}$= Denotes the total weight of the product (P) and review by users
R_P = Review (R) of Product given by user (U_i) where i = 1, 2, 3, … n

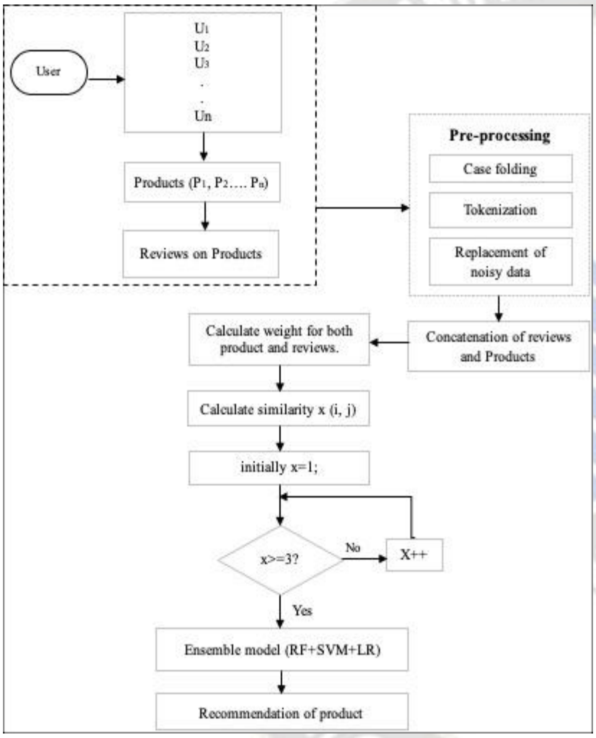


Figure 6. Block Diagram of Proposed Methodology

Step 5: Calculate the Similarity of users' reviews.

In this step 5, the similarity of users' reviews is calculated by using a formula as given below:

$$Sim\,U(i,j) = (1-d) + d \sum_{U_j \in U} Sim\,(U_j) \times R_{P_{U_i}} \quad (3)$$

Where,
$Sim\,U(i,j)$= Denotes the similarity of two consumers' reviews on the same products
d = Dampening factor

Step 6: Ensemble Model

Ensemble models are applied to the data which is obtained in step 5 by calculating similarities of users' reviews. RF, SVM, and LR are employed for modelling with the help of a few classifier algorithms, which collaborate to assign labels to the feedback.

Step 7: Recommendation of products

This is the last step of the whole process. After completing all the processes products are recommended to the consumers as per their needs.

6.3 Proposed Algorithm

ALGORITHM: CONSUMER DECISION USING CF		
Start		
INPUT: →	**PRODUCT:** $P_{i\epsilon 1,2,\ldots n}$ **CONSUMER:** $U_{j\epsilon 1,2,\ldots n}$	User per review
CO-RELATION: →	⟦PRODUCT ↔ CONSUMER⟧	$P_{1,2,3,\ldots U_i \neq j}$
OUTPUT: →	PRODUCT ADVOCACY based on TOP-RATED REVIEWS.	

Phase – I: Data Collection

Step 1: Data collection should be accurate based on consumer reviews [U_j] based on each product [P_i].
Step 2: Data contains each product's [P_i] review, available on every e-commerce website.

Phase – II: Data Pre-processing

Step 3: The review for each product and its conciseness will be in Text format. So, it is a Text classification problem.
Step 4: Each review will be cleaned with punctuation, escape sequence, stop words, emoji, unwanted spaces, and digits, then apply WordNet Lemmatization.
Step 5: Each review's conciseness will be converted into a numerical value with either OneHotEncoding.
Step 6: The rating will be visualised with a histogram plot.

Phase – III: Concatenated Review with weight Calculation

Step 7: Each product's threshold will be a minimum with three ratings.
Step 8: Each product's review's conciseness will be considered the top two most frequent ratings.
Step 9: The product will be concatenated with its topmost reviews.
Step 10: To determine each product's review-dominancy, TERM FREQUENCY AND INVERSE DOCUMENT FREQUENCY (TF-IDF) will be calculated to understand the context of the corpus, which can be calculated as:

$$W^{P}_{(u_R)} = \frac{R_{P_{U_i}}}{\sum_{U_i \in U}^{n} R_P}$$

where,
$W^{P}_{(u_R)}$ → denotes the total Weight of the product (P_i) per user's review.
R_P → Review (R) of the Product given by the user (U_i)

Phase – IV: Interlinked between more than two webpages

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: 25 July 2023 Revised: 12 September 2023 Accepted: 30 September 2023

Step 11: The maximum occurrence of n-gram tokens for each product's review will be considered a dominant specification.

Step 12: For each product [P_i] minimum of three reviews ($R_{P_{U_j}}$) will be extracted to find the occurrence of the most frequent tokens. The damping factor, d, is the likelihood that a user will click on a link, and (1-d) is for non-direct connections to any webpage.

$$Sim\ (i,j) = (1-d) + d \sum_{U_j \in U} Sim\ (U_j) \times W^{P}_{(u_p)}$$

where,
$Sim\ U(i,j) \rightarrow$ similarity of two consumers' product review
d = Dampening factor

Phase – V: Ensemble algorithms to find the best algorithm

Step 13: Ensemble Methods build sentiment analysis models. Ensembles combine classifiers to improve outcomes. RANDOM FOREST (**RF**), SUPPORT VECTOR MACHINE (**SVM**), and LOGISTIC REGRESSION(**LR**) are used.

Step 14: Provide classification report.

Step 15: Calculate the best **A**ccuracy, **P**recision, **F**1 – **S**core, and **R**ecall for the best algorithm.

End

7. Results and Dataset Description

This section presents the outcomes derived from the proposed methodology, accompanied by a concise overview of the dataset employed for training and testing purposes. A comparative analysis is conducted, pitting the proposed model against another conventional model to assess its efficiency.

- Dataset:

The dataset utilised in the proposed methodology is the "Amazon Products Recommendation 2016-2017," an open-source dataset readily available on Kaggle's website. This extensive dataset, initially released by Amazon in 2017, serves multiple computer vision applications, including instance segmentation, object identification, critical point recognition, and semantic segmentation.

Comprising 34,661 reviews, each corresponding to a unique consumer, this dataset offers a rich source of information. The reviews focus on various electronic devices available on Amazon, spanning two brands: Amazon and Amazon digital services.

- Proposed Model Comparison:

To gauge the effectiveness of the proposed model, a comparative analysis is conducted against another conventional model. The evaluation is based on key performance metrics, shedding light on the efficiency and prowess of the proposed model about established methodologies.

This comprehensive examination aims to provide insights into the robustness and applicability of the proposed model within the context of the Amazon Products Recommendation dataset. [23].

Indeed, evaluating the performance of a recommender system is a crucial step in ensuring its successful deployment. Various well-established metrics are commonly employed to analyse recommender systems' precision and overall effectiveness. The following key measurements offer insights into different aspects of a recommender system's performance:

Accuracy: This metric gauge the overall correctness of the system's predictions. It measures the proportion of correctly predicted recommendations to the total number of recommendations. Accuracy of the model a formula is used which is given below;

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

Precision: Precision focuses on the accuracy of the positive predictions made by the recommender system. It is calculated as the ratio of true positive predictions to the sum of true positives and false positives. Precision is calculated by;

$$Precision = \frac{TP}{(TP+FP)} \quad (5)$$

Where,
TP = True positive value,
FP = False Positive value.

Recall: Recall, also known as sensitivity or true positive rate, assesses the system's ability to capture all relevant instances. It is calculated as the ratio of true positives to the sum of true positives and false negatives. Recall obtained by using.

$$Recall = \frac{TP}{(TP+FN)} \quad (6)$$

Where,
TP = True positive value,
FN = False Negative

Error Rate: The error rate provides an overall measure of how often the recommender system's predictions deviate from the actual outcomes. It is calculated as the ratio of the total number of incorrect predictions to the total number of predictions.

F-Measure: The F-measure is a balanced metric that combines precision and recall, offering a comprehensive view of the system's performance. It is particularly useful when there is an inherent trade-off between precision and recall. The F1-score measured by using the formula is;

$$F-Measure = \frac{(2 * Precision * Recall)}{(Precision + Recall)} \quad (7)$$

By considering these diverse metrics, evaluators can gain a nuanced understanding of how well a recommender system is performing and where improvements may be needed. These measurements collectively contribute to a robust assessment of the recommender system's efficacy in providing accurate and relevant recommendations.

The results which are calculated for the proposed methods are given in Result-1, Result-2, Result-3, Result-4, and Result-5 as shown below.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

Result-1: Random Forest

This result presents the results of calculating the precision, recall, F1-measure, and accuracy of the proposed RF for the products and reviews that are taken from the dataset. The values of the specified parameters that are computed for the RF in this work are shown in table 1 below. A graph representation of this result is shown in figure 7 which is given below table 1.

Table 1. Calculated Values of parameters for RF

Technique	Precision	Recall	F1-Measure	Accuracy
RF	0.791	0.78	0.783	0.799

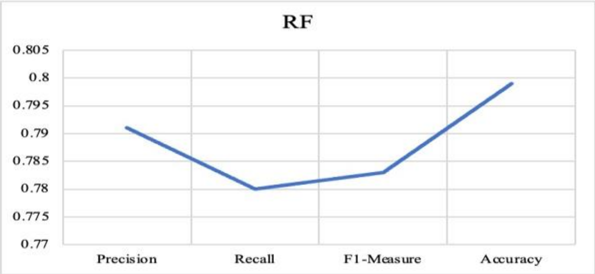


Figure 7. Graph of Random Forest Results

Result-2: Support Vector Machine

The value of the precision, recall, f1-measure, and accuracy are calculated for the proposed SVM in this result in the context of reviews and products. The table shows the values of the parameters which are calculated for the SVM in the work are shown in table 2 and the graph representation of this result is shown in figure 8 as given below.

Table 2. Calculated Values of parameters for SVM

Technique	Precision	Recall	F1-Measure	Accuracy
SVM	0.775	0.761	0.762	0.785

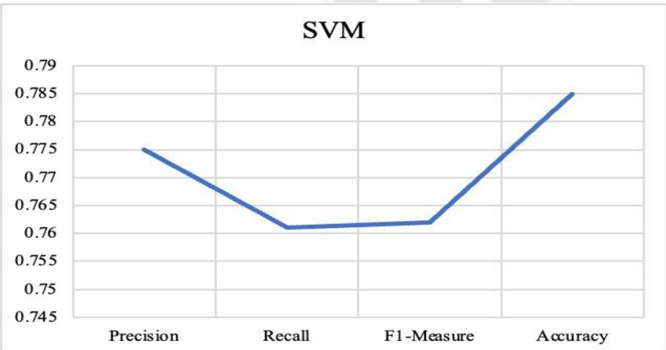


Figure 8. Graph of Support Vector Machine results

Result-3: Logistic Regression

In the context of reviews and products, the value of the precision, recall, f 1-measure, and accuracy are determined for the suggested LR in this result. The values of the parameters that are computed for the LR in this work are displayed in table 3, and a graph representation of this result is shown in figure 9 which is presented below.

Table 3. Calculated Values of parameters for LR

Technique	Precision	Recall	F1-Measure	Accuracy
LR	0.798	0.791	0.781	0.811

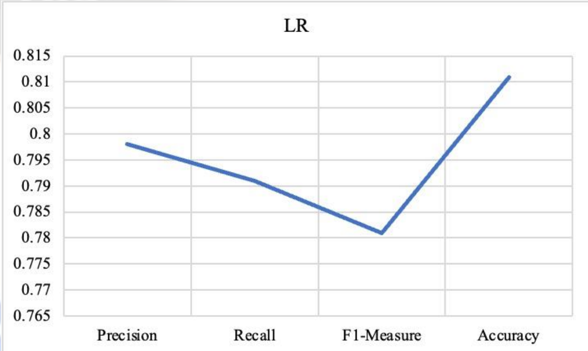


Figure 9. Graph of Logistic Regression results

Result-4: Ensemble Classifiers

The presented results pertain to the ensemble classifiers, where Random Forest (RF), Support Vector Machine (SVM), and Logistic Regression (LR) are amalgamated in an ensemble model to achieve superior outcomes compared to prior research efforts. Specifically, the evaluation focuses on precision, recall, F1-measure, and accuracy in the context of reviews and products.
Table 4 furnishes a comprehensive overview of the ensemble model's computed values for these performance metrics. It encapsulates precision, recall, F1-measure, and accuracy, providing a detailed snapshot of the ensemble classifier's effectiveness.

Figure 10 complements the tabulated information by presenting the results' graphical representation. This visual representation aids in conveying the trends and comparative performance of the ensemble classifiers across the specified metrics.

Together, these results and visualisations serve as a comprehensive assessment of the proposed ensemble classifiers, demonstrating their efficacy in enhancing precision, recall, F1 measure, and accuracy in reviewing and product recommendations.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: 25 July 2023 Revised: 12 September 2023 Accepted: 30 September 2023

Table 4. Calculated Values of parameters for Ensemble Classifiers

Technique	Parameters			
	Precision	**Recall**	**F1-Measure**	**Accuracy**
Ensemble Classifiers	0.811	0.798	0.809	0.821

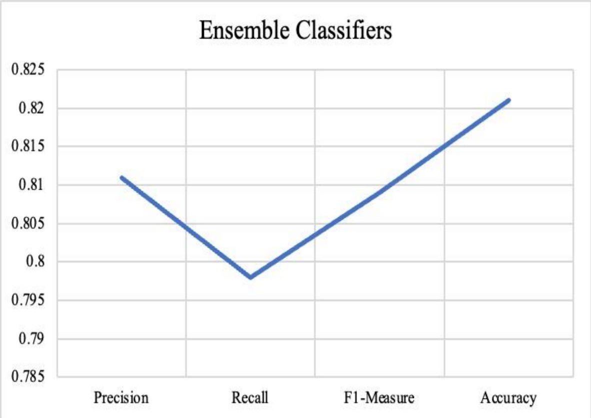


Figure 10. Graph of Ensemble Classifier results

Result-5: Error rate of all techniques

In this result, firstly, the error rate of all techniques is calculated separately. Then, the error rate of the ensemble model (RF+SVM+LR) is calculated to get a lower error than the previous work. The error rate of all these techniques is depicted in Table 5, and a graph representation of this result is shown in Figure 11 below.

Table 5. Calculated Values of Error Rate of all the techniques

Technique	Error Rate
RF	0.108
SVM	0.153
LR	0.1097
Ensemble Classifiers (RF+SVM+LR)	0.100

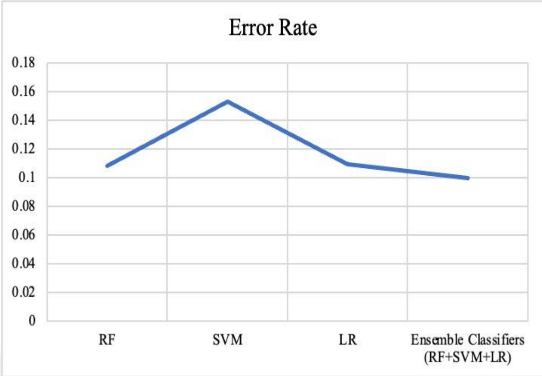


Figure 11. Graph of Error rate results for all techniques

These results were obtained by repeating the model for each classifier 3 times. It was revealed that the ensemble model with an accuracy rate of 95.2% had higher accuracy than the other algorithms for modelling.

8. Comparative Analysis

In this section, a comprehensive comparison is conducted between the proposed model and other conventional methods, including Random Forest (RF), Support Vector Machine (SVM), and Logistic Regression (LR). The comparison is based on critical positive metric parameters, encompassing accuracy, precision, recall, F1-measure, and error rate. Figure 12 visually compares the proposed model with conventional techniques, showcasing precision, recall, F1-measure, and accuracy. The results indicate that the proposed model consistently outperforms all other methods across these metrics. Figure 13 illustrates the comparison based on error rate, highlighting the proposed model's lower error rate compared to all other methods. Tables 6 and 7 provide a comprehensive overview of the comparison results. Table 6 outlines the precision, recall, F1-measure, and accuracy, emphasising the superior performance of the proposed model. Meanwhile, Table 7 details the error rate, further corroborating that the proposed model exhibits a lower error rate than other conventional techniques. These findings collectively underscore the effectiveness and superiority of the proposed model in enhancing key performance metrics compared to established conventional methods.

Table 6. Comparison of Results

Parameters	Models						
	MNB [18]	**MLP [18]**	**LR [18]**	**RF**	**SVM**	**LR**	**Proposed Ensemble Classifiers (RF+SVM+LR)**
Precision	0.783	0.753	0.783	0.798	0.775	0.798	0.802
Recall	0.774	0.748	0.774	0.791	0.761	0.791	0.821
F1-measure	0.777	0.747	0.777	0.821	0.762	0.781	0.833
Accuracy	0.7907	0.7599	0.7911	0.925	0.785	0.811	0.821

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

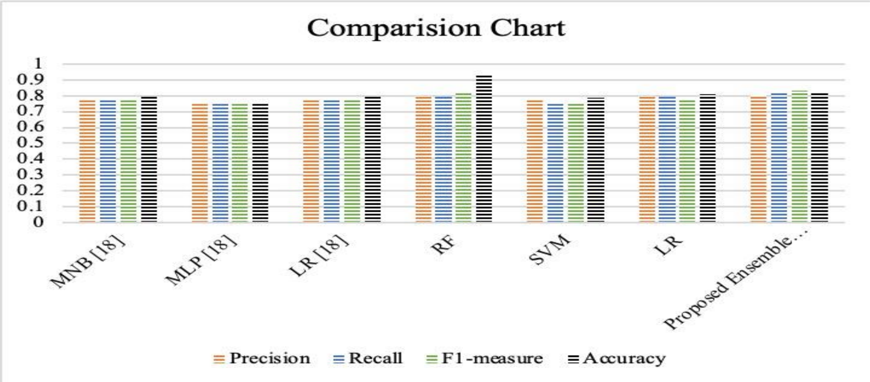


Figure 12. Comparison of the proposed work's performance to that of similar current schemes in terms of Precision, Recall, F1-measure, and Accuracy

Table 7. Error Rate

Parameter	Models						
	MNB [18]	MLP [18]	LR [18]	RF	SVM	LR	Proposed Ensemble Classifiers (RF+SVM+LR)
Error Rate	0.207	0.239	0.208	0.108	0.153	0.1097	0.100

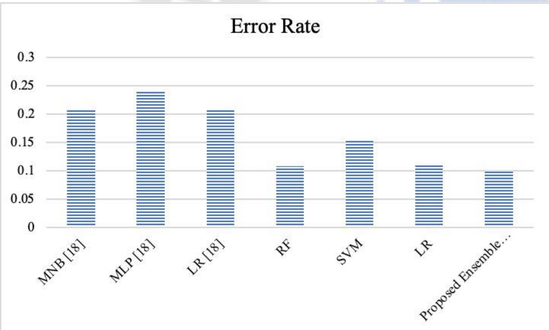


Figure 13. Comparison graph of Error rate among all techniques

9. Conclusion and Future Work

This study focuses on introducing innovative group trust models to enhance consumer decision-making. Recommendation systems have been widely employed across various applications, aiding in item recommendations (e.g., movies or music) and alleviating the information overload challenge by suggesting items likely to interest consumers. This work uses Collaborative Filtering (CF) to devise a novel weighted recommendation system to improve consumer decision-making. Traditionally, recommendation systems heavily rely on CF, where customer preference data is combined with that of other users to predict additional items of potential interest to the consumer. The unique contribution of this study lies in developing a weighted recommendation system using CF, incorporating a formula to calculate the weight for both the product and the review. A calculation is also introduced to assess the similarity between different consumer reviews. The ensemble model is composed of a Random Forest (RF), Support Vector Machine (SVM), and Logistic Regression (LR). Ensemble Classifiers, comprising RF, SVM, and LR, are considered in the practical implementation to achieve superior outcomes compared to previous research. The proposed model undergoes training and evaluation using a publicly accessible open-source dataset from the Kaggle website. The study includes a comparative analysis of the results, demonstrating that the suggested model outperforms traditional approaches across various metrics, including accuracy (0.821), precision (0.802), recall (0.821), F-measure (0.833), and error rate (0.100). Future work may involve evaluating alternative algorithms for performance comparison, aiming to identify the most effective recommendation system, particularly for online shopping websites.

References

[1]. Cui, Laizhong, Linyong Dong, Xianghua Fu, Zhenkun Wen, Nan Lu, and Guanjing Zhang. "A video recommendation algorithm based on the combination of video content and social network." Concurrency and Computation: Practice and Experience 29, no. 14 (2017): e3900.

[2]. Liu, Mengsi, Weike Pan, Miao Liu, Yaofeng Chen, Xiaogang Peng, and Zhong Ming. "Mixed similarity learning for recommendation with implicit feedback." Knowledge-Based Systems 119 (2017): 178-185.

[3]. Ding, Linlin, Baishuo Han, Shu Wang, Xiaoguang Li, and Baoyan Song. "User-centered recommendation using us-elm based on dynamic graph model in e-commerce." International Journal of Machine Learning and Cybernetics 10 (2019): 693-703.

[4]. Bouras, Christos, and Vassilis Tsogkas. "Improving news articles recommendations via user

clustering." International Journal of Machine Learning and Cybernetics 8 (2017): 223-237.

[5]. Kim, Byung-Do, and Sun-Ok Kim. "A new recommender system to combine content-based and collaborative filtering systems." Journal of Database Marketing & Customer Strategy Management 8 (2001): 244-252.

[6]. Sivaramakrishnan, N., and V. Subramaniyaswamy. "GPU-based Collaborative Filtering Recommendation System using Task parallelism approach." 2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics, and Cloud)(I-SMAC) I-SMAC (IoT in Social, Mobile, Analytics, and Cloud)(I-SMAC), 2018 2nd International Conference on. IEEE, 2018

[7]. Deng, Xiaoyi, Fuzhen Zhuang, and Zhiguo Zhu. "Neural variational collaborative filtering with side information for a top-K recommendation." International Journal of Machine Learning and Cybernetics 10 (2019): 3273-3284.

[8]. He, Xiangnan, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. "Neural collaborative filtering." In Proceedings of the 26th International Conference on World Wide Web, pp. 173-182. 2017.

[9]. Wang, Wei, et al. "Trust-enhanced collaborative filtering for personalised point of interests recommendation." IEEE Transactions on Industrial Informatics 16.9 (2019): 6124-6132.

[10]. Zhao, Yan. "Design of Garment Style Recommendation System Based on Interactive Genetic Algorithm." Computational Intelligence and Neuroscience 2022 (2022).

[11]. Sharma, Sunny, Vijay Rana, and Manisha Malhotra. "Automatic recommendation system based on hybrid filtering algorithm." Education and Information Technologies (2022): 1-16.

[12]. Tahira, Anum, Walayat Hussain, and Arif Ali. "Based Recommender System for Hedonic and Utilitarian Products in IoT Framework." In IoT as a Service: 7th EAI International Conference, IoTaaS 2021, Sydney, Australia, December 13–14, 2021, Proceedings, pp. 221-232. Cham: Springer International Publishing, 2022.

[13]. Han, Di, Yijun Chen, and Shuya Zhang. "Implicit social recommendation algorithm based on multilayer fuzzy perception similarity." International Journal of Machine Learning and Cybernetics 13, no. 2 (2022): 357-369.

[14]. Mohd Sabri, Norlina, and Nurul Azeymasnita Jaffar. "Book recommendation based on collaborative filtering technique." ESTEEM Academic Journal 18 (2022): 92-103.

[15]. Bi, Jian-Wu, Yang Liu, and Zhi-Ping Fan. "A deep neural networks-based recommendation algorithm using user and item basic data." International Journal of Machine Learning and Cybernetics 11 (2020): 763-777.

[16]. Chen, Jianrui, Bo Wang, Zhiping Ouyang, and Zhihui Wang. "Dynamic clustering collaborative filtering recommendation algorithm based on double-layer network." International Journal of Machine Learning and Cybernetics 12 (2021): 1097-1113.

[17]. Lai, Chin-Hui, Duen-Ren Liu, and Kun-Sin Lien. "A hybrid of XGBoost and aspect-based review mining with attention neural network for user preference prediction." International Journal of Machine Learning and Cybernetics 12 (2021): 1203-1217.

[18]. Abbasi, Fatemeh, and Ameneh Khadivar. "Collaborative Filtering Recommendation System through Sentiment Analysis." *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* 12, no. 14 (2021): 1843-1853.

[19]. Ali, Jehad, Rehanullah Khan, Nasir Ahmad, and Imran Maqsood. "Random forests and decision trees." *International Journal of Computer Science Issues (IJCSI)* 9, no. 5 (2012): 272.

[20]. https://medium.com/analytics-vidhya/random-forest-classifier-and-its-hyperparameters-8467bec755f6

[21]. Katzfuss, Matthias, Jonathan R. Stroud, and Christopher K. Wikle. "Understanding the ensemble Kalman filter." The American Statistician 70, no. 4 (2016): 350-357.

[22]. Ribokaitė, Lina. "Outlier detection in multidimensional streaming data." Ph.D. diss., Vilniaus universitetas, 2021.

[23]. https://www.kaggle.com/code/haojie98/amazon-product-recommendations/data+9