Journal of Engineering, Science and Technological Trends (JESTT)

Vol. 3, Issue 2, July 2026
DOI: https://doi.org/10.64060/JESTT3i24

Article

Mitigating the Grey Sheep Problem in Collaborative Filtering via Hybrid Machine Learning Architectures

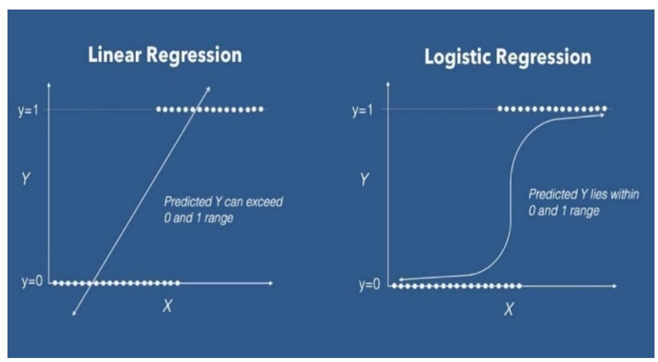
Fiza Noreen[1]*, Qurat Ul Aen Naqvi[1], Awais Rasool[2], Nimra Razzaq[3], Fatima Abbas[1] and Bira Alam[1]

[1]Riphah international university Faisalabad.
[2]University of Lahore, Pakistan
[3]Agriculture University Faisalabad
* Corresponding Email: fiza.noreen@riphahfsd.edu.pk (F. Noreen)

Received: 12 March 2026 / Revised: 23 June 2026 / Accepted: 02 July 2026 / Published online: 10 July 2026

ABSTRACT

In every nation, the major source of entertainment is movies. Worldwide, thousands of people watch movies to feel relaxed and escape from depression. The huge amount of data is an issue that is getting worse in online media. A System is created by using data mining techniques that helps to anticipate future trends, known as a prediction system or filtering system. In this system, the term grey sheep problem refers to users with unique preferences and interests that make it difficult to predict movies. In this research, Machine Learning models are used to identify the grey sheep problem and enhance movie search. A dataset of 15,538 movies was used and different attributes were selected that are publicly available on MovieLens. Then five main data mining techniques were used: Logistic Regression, Support Vector Machine, K-Nearest Neighbour, Naïve Bayes, and Decision Tree. All of the techniques gave their accuracy values, but out of all, Decision Tree gave the highest accuracy of 0.99% for movie prediction for grey sheep users.



Keywords: Machine Learning Algorithms; Movies; Algorithm; Movie Lens; Grey Sheep Users; Hybrid Learning

1. Introduction

The rapid expansion of the internet has brought about a massive amount of data online and it is becoming very hard to access the relevant information of a user in an efficient manner [1]. In order to deal with this problem, prediction systems (recommender systems) have come out as the necessary tools that offer customised recommendations depending on user preference, behaviour, and interactions. These systems are widely implemented in areas of e-commerce, entertainment and social media in order to improve user experience and interaction. There are three broadly used methods of operation of recommenders: content-based filtering (CBF), collaborative filtering (CF) and knowledge-based systems (KBS). Content-based algorithms are based on the characteristics of the item and the history of the users, whereas collaborative filtering is based on the similarities between the users or the items. Knowledge-based systems make use of explicit domain knowledge to produce recommendations. Hybrid methods that have used a combination of these methods in recent years have been proposed to enhance performance and accuracy [2]. The most important difficulty with recommender systems is the ability to deal with various kinds of users, especially grey sheep users. Unlike white sheep users (similar preferences) and black sheep users (minimal or no feedback), grey sheep users have unique and inconsistent preferences, and it is hard to find or create useful recommendations. Such users are usually difficult to manage by traditional recommendation methods because of the poor similarity trends and sparsity of data. Data mining and machine learning have become popular in enhancing the accuracy of recommendations. Clustering, classification, and matrix factorisation are some of the methods that systems use to learn hidden patterns and user preferences based on large datasets. Nevertheless, current methods remain limited to predicting preferences of grey sheep users with reasonable accuracy, which is mostly caused by the scarcity of data and the absence of robust similarity measures. In order to overcome these shortcomings, this study proposes a content-based prediction system that is supplemented with machine learning to better represent the interests of the users and increase the accuracy of recommendations to the grey sheep users. The suggested solution will be built around a process of extracting important features and making use of user behaviour to create more credible and personalised movie recommendations. The primary aim of this research is to enhance the accuracy of predictions for users of the grey sheep and enhance the creation of more effective recommender systems in online media platforms. This section is also supported by some illustrative figures in order to understand the discussed concepts better. Figure 1 shows the category of users as white sheep, black sheep and grey sheep.

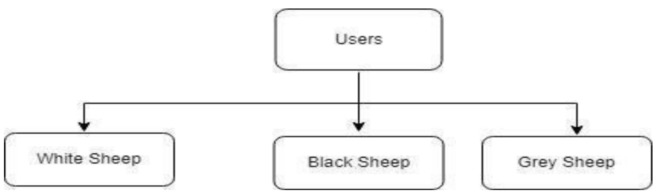


Figure 1: Types of Users.

Figures 2 and 3, as well as 4, show the nature of the recommender systems mechanism of collaborative and content-based filtering, respectively.

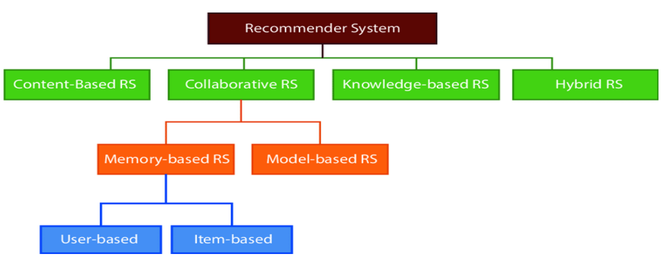


Figure 2: Types of Recommender Systems.

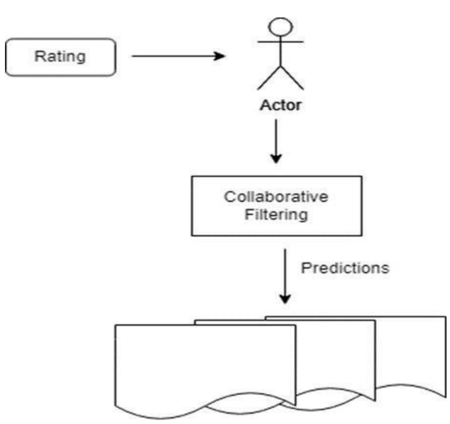


Figure 3: Collaborative Filtering-based Prediction System.

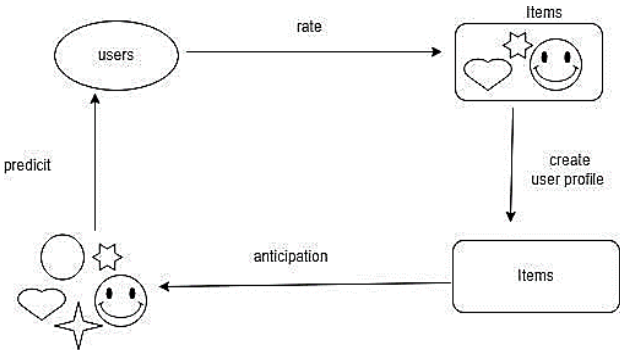


Figure 4: Content-based Filtering-based Prediction System.

Also, Figure 5 gives a descriptive analysis of the prediction systems. To help in understanding the machine learning background, Figure 6 depicts the general machine learning algorithm scheme, Figure 7 depicts the different types of regression and Figure 8 displays clustering methods. All these visuals enhance the clarity of the concepts and give a clear picture of the prediction systems and the machine learning methods explained in this section.

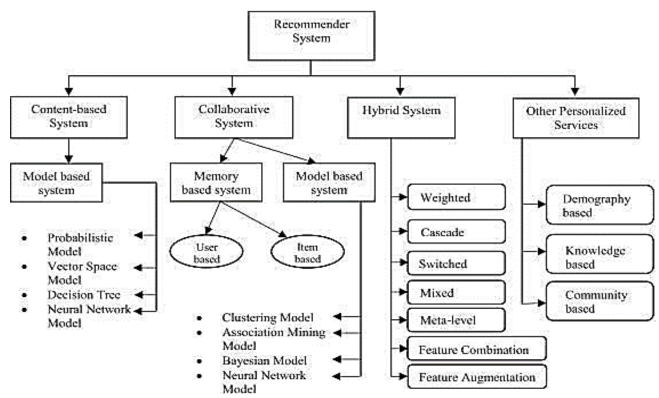


Figure 5: Detailed Classification of the Prediction System.

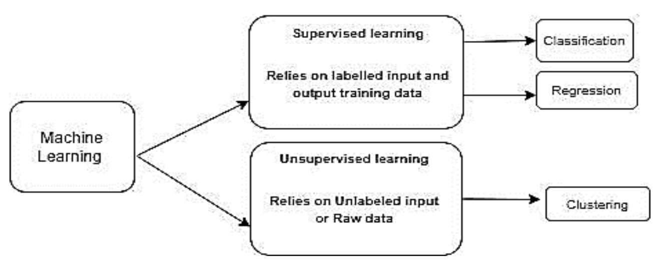


Figure 6: Machine Learning Algorithm Scheme.

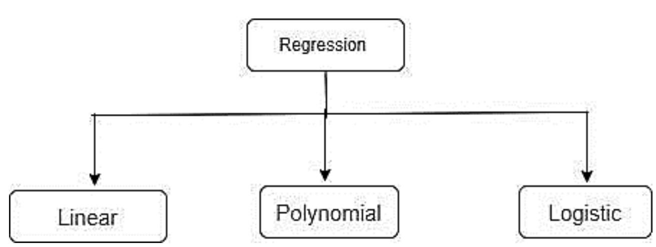


Figure 7: Types of Regression.

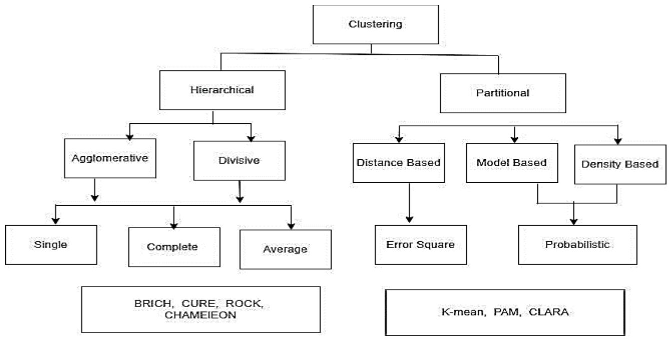


Figure 8: Clustering Types in Machine Learning.

2. Existing Study

The topic of movie recommendation systems has undergone a lot of research, but especially the identification and handling of grey sheep users, whose act of inconsistency or atypical preferences is a thorn in the prediction accuracy. Alabdul Rahman and Viktor [1] introduced the hybrid collaborative filtering model, the Grey-Sheep One-Class Recommendation (GSOR), which is a combination of one-class learning outlier detection and unsupervised learning with better prediction performance, particularly when they use Decision Tree classifiers. Likewise, Ziarani and Ravanmehr [2] have identified the need for novelty and diversity in recommendation systems, where the traditional data mining and machine learning algorithms can at times produce irrelevant predictions, and where contextual serendipity can help reduce the issue. In another strategy, Srivastava, Bala and Kumar [3] utilised psychographic models, the Big Five, Basic Human Values and Needs models to find grey sheep users in the e-commerce platform, and proved to be much more effective than the conventional rating-based methods, based on Amazon datasets. Other scholars were interested in improving clustering-based recommendation system techniques. Rashidi Khamforoosh and Sheikhahmadi [4] suggested the MKMeans++ algorithm, a combination of rating matrices and k-means clustering, demonstrating a better performance of MAE on MovieLens and FilmTrust datasets and the effect of more grey sheep users on prediction errors. On the same note, Jabbar, Mahamud, and Sagban [5] optimised the Ant Colony Optimisation to Clustering (M-ACOC), and it outperformed intra-distance and F-measure in six UCI datasets. There was also a discussion on trust-aware systems, where Gohari, Aliee and Haghighi [6] proposed Significance-Based Trust-Aware Recommendation (SBTAR), which is able to adapt to the changing user preferences using demographic settings and meta-heuristic optimisation, which is more accurate and efficient than current systems. Shokeen and Rana [7] also showed the incorporation of social network functionality in recommender systems, and developed social recommender systems which improve the quality of recommendations and overcome the shortcomings of traditional methods.Evolutionary algorithms and hybrid algorithms have also been promising. Mohammadpour et al. [8] used Genetic Algorithms with Gravitational Emulation Local Search to maximise clustering in collaborative filtering systems, which enhanced predictive performance on MovieLens and FilmTrust data sets. Ahuja, Solanki, and Nayyar [9] applied k-Means and k-NN clustering to movie recommendation, showing lower values of RMSE and fewer clusters on MovieLens. A hybrid collaborative filtering and content-based approach was suggested by Jindal, Agarwal, and Sardana [10] to detect grey sheep users and the authors obtained improved accuracy in their prediction and reduced MAE by 4.62 and 0.7757, respectively. Further studies on collaborative filtering methods were discussed by Ayub et al. [11], who pointed out weaknesses of similarity metrics like Jaccard on person-specific forecasting. To solve the cold-start problem, Gope and Jain [12] classified the solutions to be explicitly or implicitly implemented and proposed combining these methods to get improved recommendations.The problems of the accuracy of prediction and the overload of information have been well-known. Khusro, Ali and Ullah [13] have highlighted that quality prediction systems are required to present the relevant information efficiently and found research opportunities to overcome the current limitations. Models based on trust, e.g. TrustSVD, and Bayesian similarity measures as suggested by Guo, Zhang, and Yorke-Smith [14],[15] enhance performance by combining various sources of information and alleviating sparsity in data. Ghazanfar and Pruegel-Bennett [16],[17] also investigated improvements to K-Means clustering in identifying grey sheep, such as centroid initialization and distance matrix adjustments and Tang et al. [18] applied bio-inspired optimization algorithms, such as C-Ant, Bat, Cuckoo and Wolf Search, to clustering to achieve better convergence and fitness on a variety of datasets Qiu et al. [19] proposed a twofold bootstrapping propagation based opinion mining, which realized notable improvements on precision but Shan and Banerjee [20] experimented with probabilistic matrix factorization models using side information and bias correction to improve prediction accuracy. Nguyen et al. [21] also proposed hybrid recommendation approaches which used fuzzy reasoning and demographic data and Burke [22] performed a comprehensive analysis of hybrid recommender systems which used collaborative filtering as well as knowledge-based methods to overcome the cold-start problem. In general, the literature reveals that collaborative, content-based, clustering, trust-based, evolutionary and hybrid methods of movie recommendation systems have been greatly utilised. Although major strides have been taken to recognise and reduce the impact of the grey sheep users, there are still issues of advancing the accuracy, coverage and efficiency of prediction. The combination of psychographic modelling, trust-awareness, hybrid optimisation and social network capabilities seems to be a good idea to improve the results of recommendations and overcome the weaknesses of current systems.

Deep learning, matrix factorisation, ensemble learning and hybrid recommendation models have been increasingly applied in recent years to recommender systems research. While these methods have shown good results, the current study concentrates on the effectiveness of classical supervised machine learning methods for grey sheep user identification, thus offering a good baseline for further comparison.

3. Methodology

Grey sheep users can be defined as persons with unique tastes whose interests are not easily predictable in a system. Their preferences are unusual; thus, it is challenging to determine connections with other users even though they are proactive in offering feedback and expressing opinions. The chapter describes the approach planned to improve the search for movies among the grey sheep users by the use of machine learning. The suggested research will integrate various machine learning algorithms and a tailored movie dataset to enhance the quality of prediction and recommendations. The methodology is a complete procedure of the research which involves the proposed computational model, dataset structure, tools and technologies, evaluation criteria and the machine learning methods that are applied to classify data. The suggested computational model of computation, as shown in Figure 9, Research Proposed Method, exploits a dataset that was obtained from MovieLens (2019) and includes various attributes. Data preprocessing is used to remove the features that are directly associated with the preferences of the user's movies, which allows identifying grey sheep users correctly. The process of data collection is organised to collect and assess variables in the publicly available dataset, which is converted into a CSV to be compatible with machine learning algorithms. Data preprocessing is used to clean, integrate, transform and sample raw datasets, which usually have inconsistencies, noise, or missing values. Cleaning will eliminate duplications, outliers and noisy entries, whereas integration will consolidate various data sources, and transformation will normalise and discretise data to an appropriate format to be analysed. Sampling minimises the computational cost by choosing representative subsets and overfitting is reduced to ensure that the models are not learning noise instead of generalizable patterns, as shown in Figure 10: Data Overfitting.

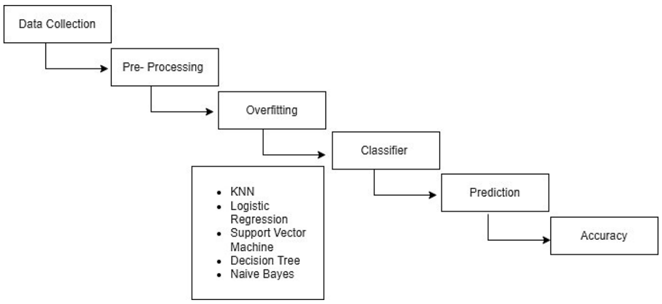


Figure 9: Research Proposed Method.

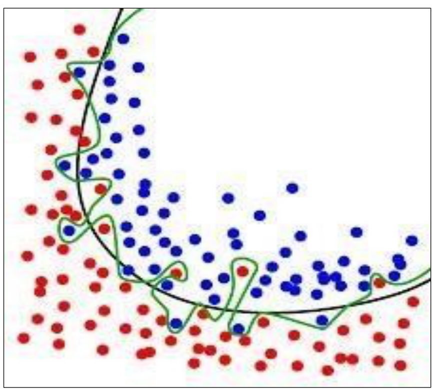


Figure 10: Data Overfitting.

The process of data mining incorporates several supervised learning methods, such as Logistic Regression (LR), K-Nearest Neighbour (KNN), Naive Bayes (NB), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF) and Gradient Boosting (GB). All the algorithms are trained and tested with various parameters to establish their predictive qualities, especially for grey sheep users. The last dataset format presented in Table 1: Final Dataset Structure incorporates the features of Movie ID, Rating, Views, Title and user Labels (grey or white), which are essential to supervised learning and model testing.

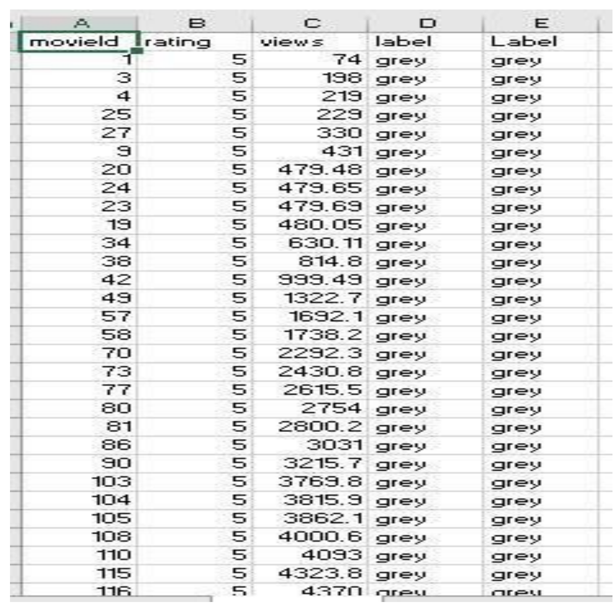
A rule-based labelling strategy was used to identify grey sheep users. A movie was considered a grey sheep user if it had fewer than 20,000 views and had a rating of 3 or higher. These films suggest users who are different from the mainstream since they give positive ratings to relatively obscure material. The users that do not meet this criterion were classified as white sheep users.

All of the proposed framework is based on a hybrid machine learning structure with several processing stages in the same recommendation framework. The architecture comprises data preprocessing for datasets, feature extraction, rule-based grey sheep labelling, supervised machine learning-based classification and recommendation-based analysis. In this work, the hybrid architecture is not an ensemble classifier, but rather the combination of the concepts of the recommendation system with different machine learning processing steps. In particular, the framework integrates two stages of data pre-processing, feature extraction, rule-based grey sheep labelling and supervised machine learning grey sheep classification in one workflow. The proposed hybrid architecture is not a combination of more than one classifier as an ensemble model, but the combination of recommendation system methodology and machine learning techniques.

Table 1:
Final Dataset Structure

Attributes	Meaning	Example
Movie Id	Movie unique ID	3176
Rating	Rating is given by the user	5
Views	How many users view	7049,957
Title	Movie title	Wild Horses (Caballossalvajes) (1995)
Labels	Label users' type	Grey or white

The training models require labelled datasets, which allow algorithms to identify patterns and make correct inferences, and the result can be observed in Figure 11: Output of Label Data. The notebook is based on Python, which is the main program to be used to run and analyse.



movieId	rating	views	label	Label
1	5	74	grey	grey
3	5	198	grey	grey
4	5	219	grey	grey
25	5	229	grey	grey
27	5	330	grey	grey
9	5	431	grey	grey
20	5	479.48	grey	grey
24	5	479.65	grey	grey
23	5	479.69	grey	grey
19	5	480.05	grey	grey
34	5	630.11	grey	grey
38	5	814.8	grey	grey
42	5	999.49	grey	grey
49	5	1322.7	grey	grey
57	5	1692.1	grey	grey
58	5	1738.2	grey	grey
70	5	2292.3	grey	grey
73	5	2430.8	grey	grey
77	5	2615.5	grey	grey
80	5	2754	grey	grey
81	5	2800.2	grey	grey
86	5	3031	grey	grey
90	5	3215.7	grey	grey
103	5	3769.8	grey	grey
104	5	3815.9	grey	grey
105	5	3862.1	grey	grey
108	5	4000.6	grey	grey
110	5	4093	grey	grey
115	5	4323.8	grey	grey
116	5	4370	grey	grey

Figure 11: Output of Label Data.

Machine learning algorithms because it is cross-platform and has extensive data analysis libraries. Parameters are set to assess model performance, including ways to identify overfitting or underfitting. The data is divided into training and testing sets, usually in an 80:20 proportion in order to make sure that models can be used effectively even on unseen data. To compute the results of classification, the confusion matrix is used and in the context of classification results, the accuracy, precision and recall are mathematically defined by Equation (1), Equation (2) and Equation (3), respectively, which give quantitative measures of the model performance.

$$\text{Accuracy} = \frac{TN + TP}{(TN + FP + TP + FN)} \quad (1)$$

$$\text{recision} = \frac{TP}{(TP + FP)} \quad (2)$$

$$\text{Recall} = \frac{TP}{(TP + FN)} \quad (3)$$

The research applies seven machine learning methods to the prediction of grey sheep users. Logistic Regression (LR) models are relations between the dependent and independent variables and their mathematical model is presented in Equation (3) and the comparison of the linear and logistic regression is shown in Figure 12: Linear Regression Vs Logistic Regression Graph.

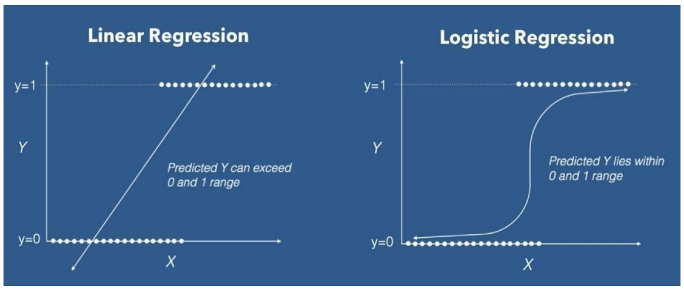


Figure 12: Linear Regression Vs Logistic Regression Graph.

Support Vector Machine (SVM) algorithms form hyperplanes to classify classes both in linear and non-linear space, as shown in Figure 13: SVM Classification of Data and Figure 14: SVM Classification of Non-Linear Data. Decision Trees (DT):

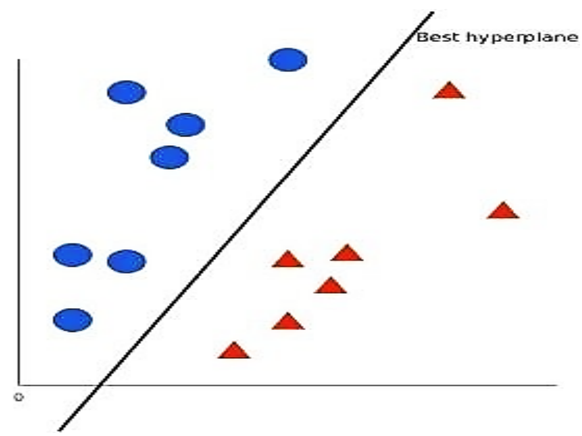


Figure 13: SVM Classification of Data.

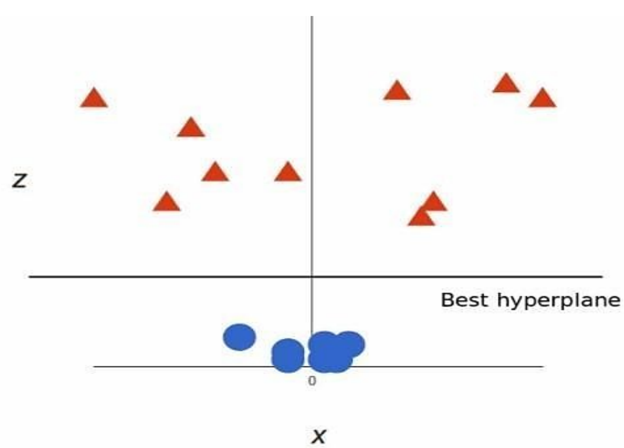


Figure 14: SVM Classification of Non-Linear Data.

This is based on hierarchical structures to group data according to information gain and entropy, and the classification output is provided in Figure 15:

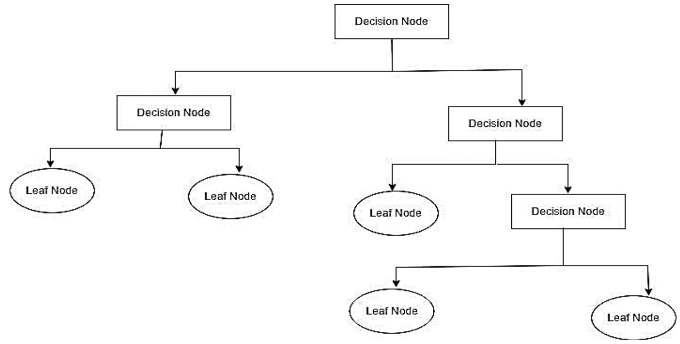


Figure 15: Classification of Decision Tree.

Classification of Decision Tree. K-Nearest Neighbour (KNN) is a classifier that allocates a label based on the proximity of the neighbouring points, which is computed with the help of Euclidean distance (Figure 16: Calculation of Euclidean Distance) and illustrated in Figure 17: K-Nearest Neighbour Showing Nearest Neighbour.

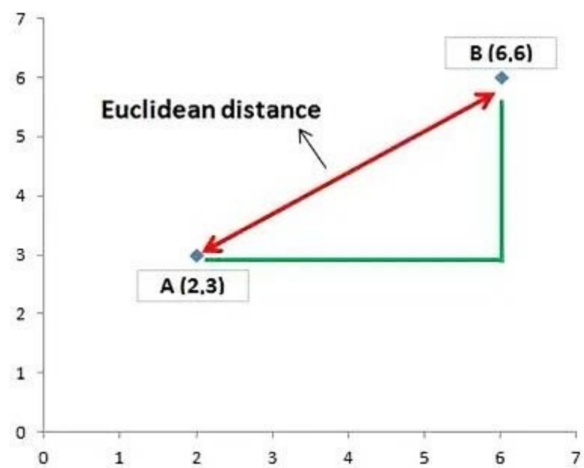


Figure 16: Calculation of Euclidean Distance.

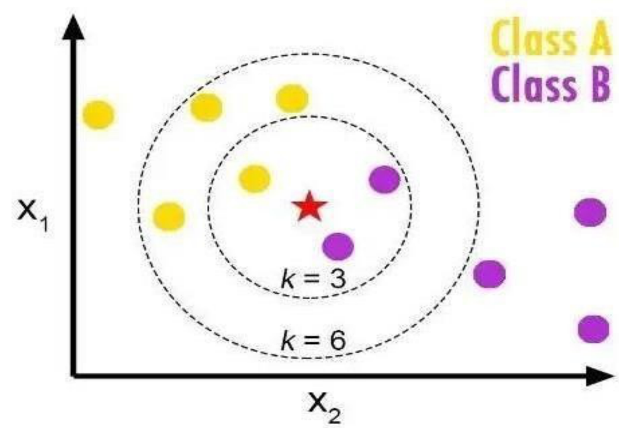


Figure 17: K-Nearest Neighbour Showing Nearest Neighbour.

Naive Bayes (NB) classifiers are classifiers that assume feature independence to scale in terms of efficiency and scalability (Figure 18: Naïve Bayes Classifier).

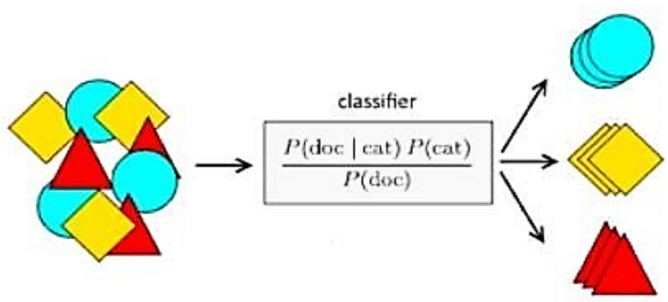


Figure 18: Naïve Bayes Classifier.

Random Forests (RF) combine the decisions of several Decision Trees, taking advantage of bootstrap sampling and randomness of features and the model operation is shown in Figure 19: Random Forest Classifier.

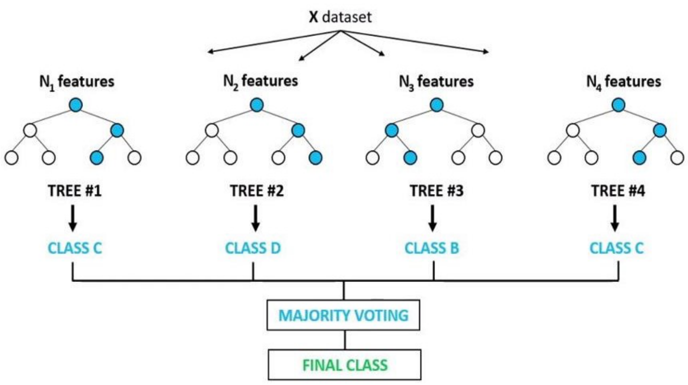


Figure 19: Random Forest Classifier.

Gradient Boosting (GB) models are optimised sequentially to minimise bias and variance, and the workflow is illustrated in Figure 20: Working Flow of Gradient Boosting Classifier.

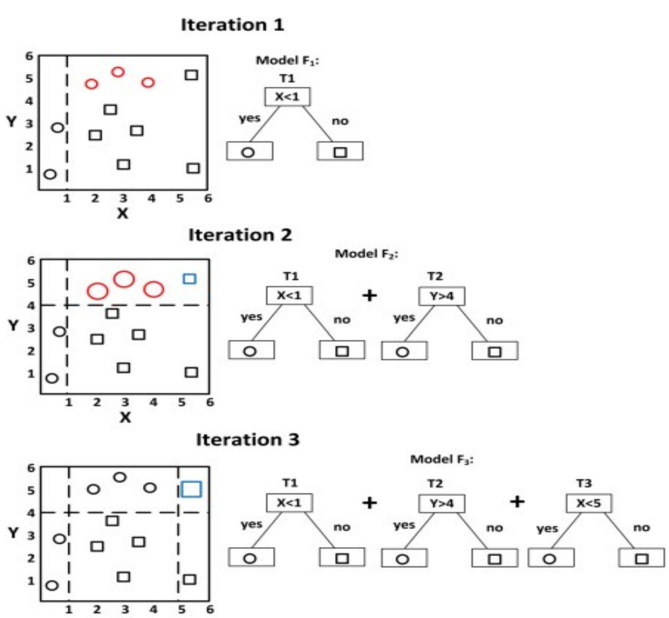


Figure 20: Working Flow of Gradient Boosting Classifier.

All these algorithms are used to identify grey sheep users using various attributes so that the predictive accuracy can be maximised and overcome the problem of overfitting or data sparsity. Overall, the chapter outlines the systematic research process in prediction and classification of grey sheep users on the basis of machine learning. It involves extensive data preprocessing, proper labelling, and using various supervised learning algorithms. Evaluation parameters such as accuracy, precision, recall, and confusion matrices give quantitative information about the model performance, which makes the proposed methods robust and effective to identify unique user preferences.

4. Results and Discussion

In this chapter, the experimental assessment and a thorough discussion of the suggested method of detecting the users of grey sheep in movie recommendation systems are provided. The research uses the MovieLens dataset comprising 1,002,009 ratings given by 6,040 users on almost 3,900 films and longer data of 15,512 films with ratings and the number of views. The data were preprocessed and classified into grey sheep users and white sheep users based on an IF condition whereby users with high ratings and fewer views were classified as grey sheep users and 7,708 white sheep users, respectively, as shown in Figure 21: Distribution of Users in the CSV Dataset. The dataset was separated into 80% training and 20% testing, with X being the columns (except the label) and Y being the label column. Data was processed through the Pandas library with the labelled CSV dataset and model training and testing were done after splitting the data.

$$\text{Grey Sheep} = \begin{cases} 1, \text{ if Rating} \geq 3 \text{ and Views} < 20{,}000 \\ 0, \text{ Otherwise} \end{cases}$$

To achieve consistency across all classifiers, the present study used the commonly used train-test evaluation protocol of 80:20. Another approach to providing robustness analysis is provided by k-fold cross-validation, but this was not investigated in the present study and has been recognised as a key area for future research.

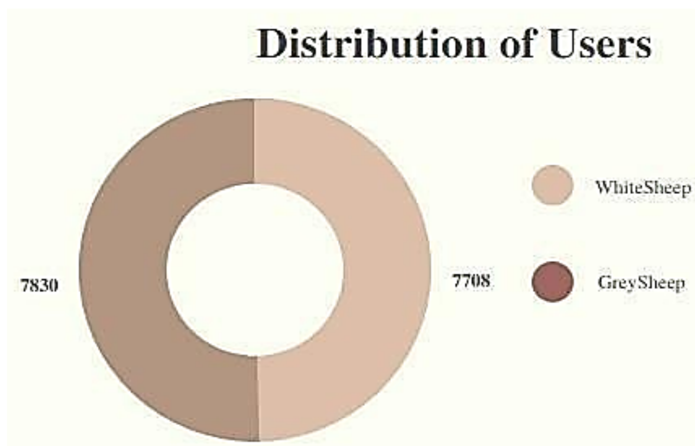


Figure 21: Distribution of Users in the CSV Dataset.

To assess the performance of the proposed method, several machine learning algorithms were used, such as Naive Bayes, K-Nearest Neighbour (K-NN), Logistic Regression (LR), Support Vector Machine (SVM) and Decision Tree (DT). The accuracy, F1-score, precision and recall were used to measure the performance of these algorithms. The results of the precision evaluation are given in the Figure 22: Evaluation of Precision Along K-NN, Figure 23: Evaluation of Precision Along with SVM, Figure 24: Evaluation of Precision Along LR Similarly, F1-score evaluations are shown in Figure 25: Evaluation of F1 Along with Naïve Bayes, Figure 26: Evaluation of F1 Along with SVM, Figure 27: Evaluation of F1 Along K-NN, demonstrating the comparative performance of each algorithm across different evaluation metrics.

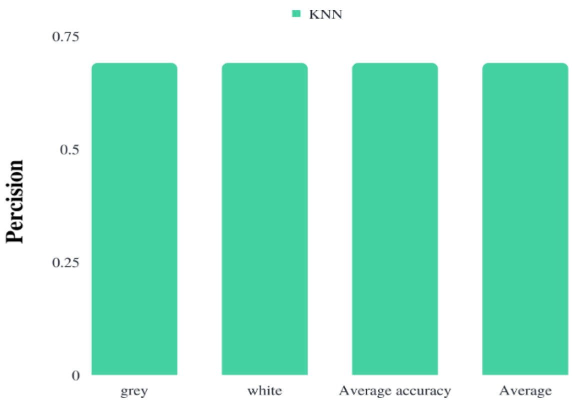


Figure 22: Evaluation of Precision Along LR.

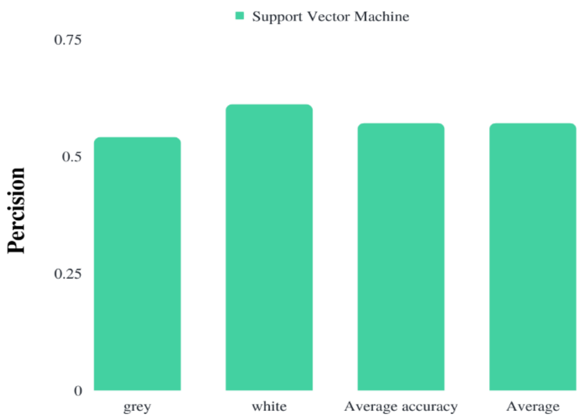


Figure 23: Evaluation of Precision Along K-NN.

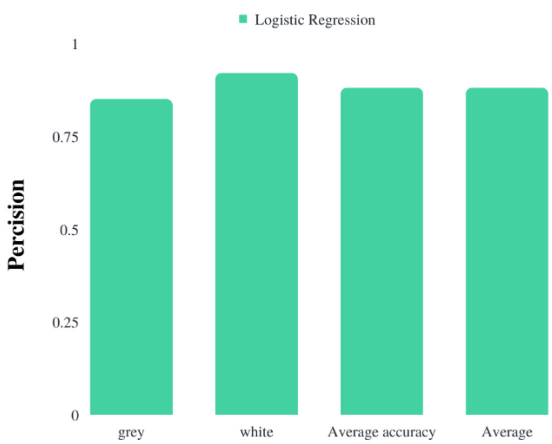


Figure 24: Evaluation of Precision Along LR.

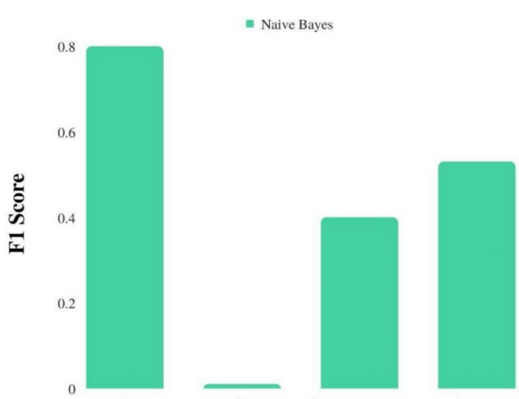


Figure 25: Evaluation of F1 Along with Naïve Bayes.

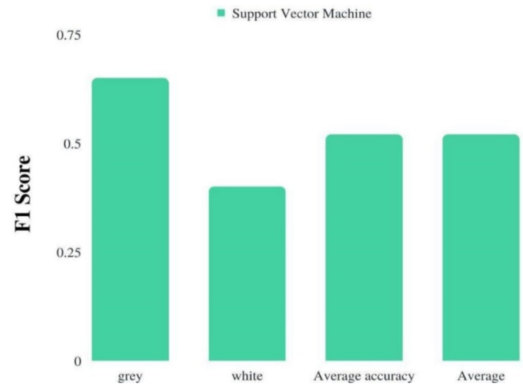


Figure 26: Evaluation of F1 Along with SVM.

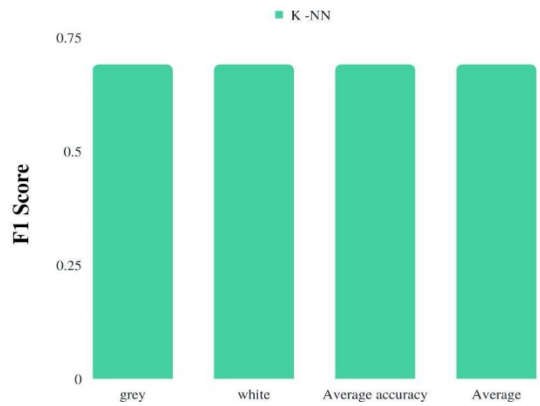


Figure 27: Evaluation of F1 Along with K-NN.

The standard formula used to calculate accuracy is Accuracy = (TN + TP) / (TN + FP + TP + FN) and the results of the comparisons are summarised in Table 2: Accuracy Comparison and graphically presented in Figure 28: Gained Accuracy After Comparing Different Algorithms. These findings have shown that the Decision Tree algorithm was most accurate at 0.99 with perfect recall, precision and F1-score and next was the Logistic Regression at 0.88 and then SVM. Moreover, the efficiency of the given method was also confirmed by the comparison with the existing methods, which are presented in Table 3: Comparison Table and Figure 29 Comparison Graph with Previously Used Algorithms, where the given technique was more efficient in terms of accuracy than MKMeans++, SFLA, PCC and Collaborative Filtering methods.

Table 2:
Accuracy Comparison

Algorithm	Accuracy	Recall	F1 Score	Precision
NB	0.66	0.67	1.00	0.80
K-NN	0.69	0.69	0.70	0.69
LR	0.88	0.85	0.93	0.89
SVM	0.55	0.54	0.82	0.65
DT	0.99	1.0	1.0	1.0

The performance reported is measured on experiments carried out under a uniform experimental setup with the MovieLens dataset. The Decision Tree classifier outperformed the other classifiers because the features used for this classification were selected (Movie ID, Rating, Views, Title and Labels) and the labelling strategy used in this study was well-defined. The results reported are valid only for the data set used and are not representative of all data sets for recommender systems.

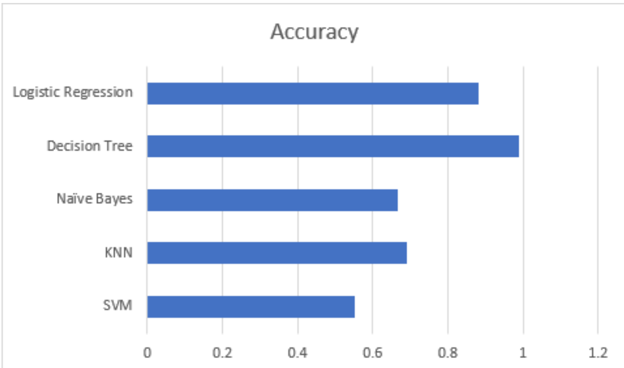


Figure 28: Gained Accuracy After Comparing Different Algorithms.

Table 3:
Comparison Table

Sr. No.	Author	Method	Accuracy
1	(Rashidi, Khamforoosh, & Sheikhahmadi, 2022)	MKMeans++, FirefyMkMeans++	68%
2	(Gohari, Aliee, & Haghighi, 2020)	Shuffled Frog Leaping Algorithm (SFLA), Significance-Based Trust-Aware Recommendation (SBTA)	60%
3	(Reddy, Kanmani, & Surendiran, 2020)	Pearson correlation coefficient (PCC)	48%
4	(Gras, Brun, & Boyer, 2016)	Collaborative Filtering	90%
5	(Proposed Method)	Decision Tree, Logistic Regression, K-Nearest Neighbour, Naïve Bayes, Support Vector Machine	99% 88% 69% 66% 55%

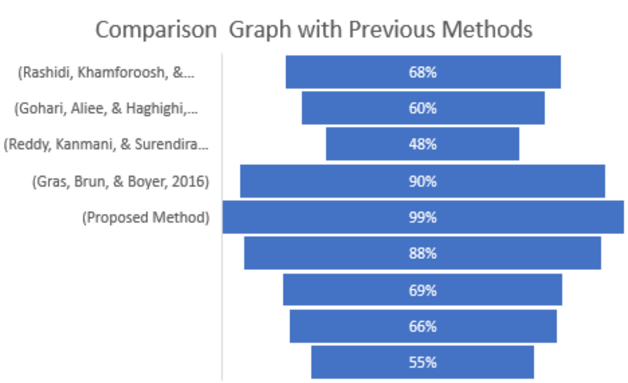


Figure 29: Comparison Graph with Previously Used Algorithms.

Another issue that the study deals with is imbalanced datasets, where grey sheep users are the minority group, by using both machine learning, anomaly detection, and one-class classification methods to enhance the predictive accuracy. Also, sample profiles of grey sheep users can be found in Table 4: Profile Sample of Grey Sheep Users with a focus on such attributes as movie ID, title, year, awards and IMDB ratings. Generally, the results show that the proposed intelligent recommender system can be used as an effective tool for improving the accuracy of the predictions and also in identifying grey sheep users, thus making the proposed system a great tool compared to the current system.

This study aimed to evaluate the performance of machine learning classifiers against popular classifiers and use widely accepted evaluation metrics such as accuracy, precision, recall and F1-score. Other measures like ROC-AUC and statistical significance testing are recognised as useful additions and have been identified as future work.

Table 4:
Profile Sample of Grey Sheep User

Movie Id	Title	Year	Awards	IMDB.com
73,023	Crazy Heart	1996	12 Gener: Drama-Romance	82,278
74,545	Ghost Writer	2010	23 Gener: Drama - Mystery Thriller	153,234
78,574	Ghost Writer	2010	29 Gener: Drama-Mystery Thriller	132,434
81,158	Restrepo	2010	10 Genre: Documentry-war	21,606
81,788	The Next Three Days	2010	0 Gener: Drama-crime	176,901
82,667	I Saw The Devil	2010	15 Gener: Crime- thril	103,699

5. Conclusion and Future Work

The study on the users of grey sheep is summarised, and the unique preferences used by these users are described to ensure that proper prediction is difficult to achieve in the recommendation systems. The study emphasises the role of predictive systems in aiding decision-making in a wide range of areas and introduces a unique approach aimed at enhancing the movie recommendations to such users. It is possible to apply and evaluate multiple machine learning algorithms and the best accuracy of 0.99 with perfect recall and F1-score was achieved on Decision Tree, whereas Logistic Regression (0.88), K-Nearest Neighbour (0.69) and Support Vector (0.55) had lower accuracy. The dataset used in the study is the publicly available MovieLens dataset, where data preprocessing, feature selection and implementation were conducted in Python within a Jupyter Notebook environment. The results indicate that the suggested method has a significant positive effect on prediction accuracy and Decision Tree is the most successful method to identify grey sheep users. Additionally, the study proposes future studies, such as incorporating information on trust and distrust, adding contextual variables (of time, place and user emotions) and testing the methodology with other data sets to further validate and compare the results. Further testing of the proposed framework will also be carried out in future using multiple benchmark datasets to assess the robustness of the framework through k-fold cross-validation.

Declaration

AI Disclosure: N/A

Availability of Data and Materials: All data is provided in the paper.

Competing Interests: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Consent to Publish: The authors agree to publish the version of the manuscript in this journal.

Ethical Issues: There are no ethical issues.

Funding Statement: N/A.

References

1. Aamodt, A., & Plaza, E. (1994). Case-based reasoning: Basic issues, approaches and methods. AI Communications, 7(1), 39-59.
2. Ahirwadkar, B., & Deshmukh, S. (2019). Recommender systems using deep neural networks. International Journal of Innovative Technology and Exploring Engineering, 8(12), 4838-4842.
3. Ahuja, R., Solanki, A., & Nayyar, A. (2019). Movie recommender system based on KMeansclustering and KNearest Neighbor. In 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 10th - 11th January, India, (pp. 263-268). IEEE.
4. Ajit, P. (2016). Machine learning algorithms for prediction of employee turnover rate in organizations. Algorithms, 4(5), 5-12.
5. Alabdulrahman, R., & Viktor, H. (2021). Meeting different needs: One-class machine learning for grey sheep users recommender systems. Expert Systems with Applications, 166(7), 114-061.
6. Anonymous (2014) MovieLens, Avilable online at, courses, Group Lens (https://grouplens.org/datasets/movielens/), visited on 04-JAN-2023.
7. Ayub, M., Ghazanfar, M. A., Maqsood, M., & Saleem, A. (2018). A jaccard base similarity measure to improve performance of CF based recommender systems. I n International Conference on the Information Networking (ICOIN), 10th - 12th January, Thailand, (pp. 1-6). IEEE.
8. Aziz, N., Akhir, E. A. P., Aziz, I. A., Jaafar, J., Hasan, M. H., & Abas, A. N. C. (2020). Gradient boosting algorithms in development of artificial intelligence systems for monitoring and prediction. International Conference on Computational Intelligence (ICCI), 08th - 09th October, Malaysia, (pp. 11- 16). IEEE.
9. Balabanović, M., & Shoham, Y. (1997). Fab: content-based, collaborative recommendation. Communications of the ACM, 40(3), 66-72.
10. Billsus, D., & Pazzani, M. J. (2000). User modeling for adaptive news access. User Modeling and User-Adapted Interaction, 10(2), 147-180.
11. Bishop, C. M., & Nasrabadi, N. M. (2006). Pattern recognition and machine learning Springer, 4(5), 738.
12. Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). Recommender systems survey. Knowledge-Based Systems, 46(5), 109-132.
13. Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). Training an optimal margin classifier: A support vector machine. In Fifth Annual Workshop on Computational Learning Theory, 66(1), 144-152.
14. Boutilier, C., Brafman, R. I., Hoos, H. H., & Poole, D. (1999). Reasoning With Conditional Ceteris Paribus Preference Statements. UAI, 99(2), 71-80.
15. Breese, J. S., Heckerman, D., & Kadie, C. (2013). Collaborative filtering for imputation. ArXiv Preprint ArXiv:10(8),1301-7363.
16. Breiman, L. (2001). Random forests. Machine learning, 45, 5-32.
17. Burke, R. (2002). Hybrid recommender systems: Survey and experiments. Burke, R. (2007). Hybrid web recommender systems. The adaptive web, 77(1), 377
18. Elsalamony, H. A. (2014). Data mining techniques for bank direct marketing. International Journal of Computer Applications, 85(7), 12-22.
19. Felfernig, A. and Burke, R. (2008). Technology and research challenges of constraint-based recommender systems. In 10th International Conference on Electronic Commerce,67(8), 19th August, Austria, (pp. 1-10).
20. Adomavicius, G., & Tuzhilin, A. (2005). The next generation of recommender systems: A survey of the state-of-the-art and future directions. IEEE Transactions on knowledge and Data Engineering, 17(6), 734-749.

Author(s) Bio

Fiza Noreen from Riphah international university Faisalabad.
Email: fiza.noreen@riphahfsd.edu.pk

Qurat Ul Aen Naqvi from Riphah international university Faisalabad.
Email: quratulaen.naqvi@riphahfsd.edu.pk

Awais Rasool from University of Lahore, Pakistan.
Email: awais.rasool@se.uol.edu.pk

Nimra Razzaq from Agriculture University Faisalabad.
Email: csnimra@gmail.com

Fatima Abbas from Riphah International University Faisalabad.
Email: Fatima.abbas@riphahfsd.edu.pk

Bira Alam from Riphah International University Faisalabad.
Email: Bira.alam@riphahfsd.edu.pk