ISSN No:-2456-2165

An Improvised Business Intelligence Recommender System using Data Mining Algorithm

Mustapha Maidawa; A. Y. Dutse; Aminu Ahmad
Department of Management and Information Technology
Abubakar Tafawa Balewa University,
Bauchi

Abdulsalam Ya'u Gital
Department of Mathematical Science
Abubakar Tafawa Balewa University,
Bauchi

Ismail Zahraddeen Yakubu
Department of Computing Technologies
SRM Institute of Science and Technology Chennai,
India

Abstract:- AI allows for a higher quality of recommendation than can be achieved by conventional recommendation methods. This has ushered in a new era for recommender systems, creating advanced observations of the relationship between users and items, presented an expanded understanding of demographic, textural, virtual, and contextual data as well as more intricate data representations. However, the challenge for the recommendation systems is to solve the problems of sparsity, scalability, and cold start. The existing capsule networks take times in training making it a slow algorithm. Also, ignoring the sparsity in the datasets have result to reduction in prediction accuracy. Other works of literature already in existence add column or row meanings to such sparse values. Because the mean disregards the underlying correlation in the data, accuracy is compromised. Hence, this study examined the existing framework and the need to provide a solution to the problem by proposing the inclusion of business intelligence component framework base on recommender system. Therefore, to address these issues, this research proposed a hybrid collaborative base recommendation system using an improved SVD and self-organized map neural network (SOM) to improve cold start, accuracy, speed and sparsity issue of the current recommendations by combining SOM clustering to cluster the dataset, a better SVD to reduce dimensionality and increase sparsity, and a cooperative strategy to address accuracy and sparsity concerns. Experimental result shows that the proposed model has consistently performed better than all the three state-of-the-art methods including the Capsule Neural Network CF algorithm, the KNN CF algorithm and the SVD+SOM clustering base recommender system. This study has proven that data mining can helps companies and business managers to visualize hidden patterns and trends in datasets that were not visible before. Whatever insights are revealed, they make clear decisions that benefit both the company and the customers and the stakeholders they serve.

***Keywords**:- Recommender System, K-Neareast Neighbour, Jaccard Distance, Euclidian Distance and Cosine Distance.*

I. INTRODUCTION

Business Intelligence is a multidisciplinary field that encompasses technology, analytics, data management, and business strategy [1]. It plays a crucial role in enabling data-driven decision-making and assisting businesses in making the most of their data assets to obtain a competitive advantage in the market. The term "business intelligence" (BI) describes the methods, tools, procedures, and practices used to gather, examine, combine, and display business data in order to make decisions. It involves gathering and transforming raw data into actionable insights to support strategic and operational decision-making within an organization. This is achieved using various knowledge extraction techniques [2].

Knowledge extraction, business intelligence (BI), and recommender systems are interconnected in the context of leveraging data to provide valuable insights and personalized recommendations to support decision-making and enhance user experiences [3]. Overall, knowledge extraction provides the foundation for BI by extracting insights from data, and BI, in turn, contributes to recommender systems by providing relevant insights for personalized recommendations. Together, these concepts enable organizations to extract valuable knowledge, make informed decisions, and deliver tailored recommendations to enhance user experiences and drive business growth [4]. It is in the light of these developments that this study proposed an intelligent AI framework base on recommender system for business support system.

However, the challenge for recommendation systems, however, is how to deal with sparsity, scalability, and cold start difficulties[5]. The existing capsule networks take times in training making it a slow algorithm. Also, ignoring the sparsity in the datasets have result to reduction in prediction accuracy. Other works of literature already in existence add column or row meanings to such sparse values. Because the mean disregards the underlying correlation in the data, accuracy is compromised. Hence, this study examined the existing framework and the need to provide a solution to the problem by proposing the inclusion of business intelligence component framework base on recommender system. Therefore, to address these issues, this research proposed a hybrid collaborative base recommendation system using an improved SVD and self-organized map neural network

ISSN No:-2456-2165

(SOM) to improve cold start, accuracy, speed and sparsity issue of the existing recommendation through the integration of SOM clustering for dataset clustering, an enhanced SVD for dimensionality reduction and sparsity, and a cooperative strategy to address the sparsity and accuracy issues.

II. RELATED WORK

Different Recommender System Algorithms, such as Content-Based and Collaborative-Based, have been developed by researchers and data scientists to filter the vast amount of information available on the internet and recommend only the essential and relevant content based on users' personalized interests. These are being used on social networking and e-commerce websites to gather personal data about users and send them recommendations for information and material that is relevant to their interests, activities, and behavior. For example, movies, books, clothes, tweets are being recommended to the people visiting different web sites. Recommender System is an application of Web Mining [6]. For example [7] suggested a productive method based on hierarchical clustering for recommender systems. The findings show that the Chameleon-based recommender system outperforms the K-means-based recommender system in terms of error production. The study tackles the fundamental requirements of modern recommender systems, namely precision and swiftness. However, the running time of chameleon algorithms can further be reduced by using any parallel framework like map reduce.

Moreso, [8] proposed a Movie Recommender System. The system has been developed in PHP and currently uses a simple console-based interface. However, testing the model on a larger data set that will enable more meaningful results.

Similarly, [9] Create a recommendation system that is based on user ratings, then assess it with statistical analysis methods. It was discovered that the clustering accuracy was good and that initial seeding required less rounds to converge. Even when there were more than 100 clusters, Random Forest's processing performance for a large number of labels was. Improvements of.75% have been demonstrated using the suggested approach. The Softmax Regression approach, which is the most efficient among the others, has therefore improved by 0.75%. Nevertheless, when the dataset grows substantially in size, the model will be affected.

Additionally, [10] suggested a method for grouping items based on their metadata. Evaluations are grouped based on the genre of the item. The RMSE is improved between 4.7 and 9.8%, while the MAE is improved between 0.3 and 1.8%. Though for more specialized systems, the weighting approach might be selected based on the system's size and goals, each cluster has its own rating prediction and weighting strategies.

Moreso, [11] suggested a recommender system for the travel sector that makes use of machine learning, prediction, and cluster ensemble approaches. Comparing the cluster ensembles to approaches that depend just on single clustering algorithms, the suggested recommendation method can benefit from higher prediction accuracy. The main restriction, though, is that classic and multi-criteria CF have a scaling problem. Therefore, in order to demonstrate how the incremental learning approach may get beyond the scalability issue, it must be developed and tested on big multi-criteria datasets.

[12] suggested a nine-factor analytic approach, the Genetic Weighted K-Means clustering (GWKMC) clustering methodology, and the current classification algorithm, the Negative Selection Algorithm (NSA). The suggested Recommender System's effectiveness was validated by the carried out trials. Also by using sequential information from users' online page browsing, [12] suggested a revolutionary web-based recommender system. The suggested model's accuracy is over three times more accurate than some of the current systems, as demonstrated by the obvious results. This suggested model has an accuracy of over 33%. However, the research fails to include privacy, trust and social networks with the utilization of hybrid intelligent systems.

[13] Suggested a deep learning neural network framework to produce model-based predictions for the business-user combinations by leveraging reviews in addition to content-based attributes. In comparison to the standalone memory-based collaborative filtering method, the hybrid approach shows great promise as a solution. This technique creates a unified supervised learning model that offers better prediction results than memory-based collaborative filtering recommendation systems by combining content (user and business), collaboration (review and votes), and metadata related to ratings. Though it was not used in this study, geographical data from the companies can be crucial for the creation of recommendation systems that are location aware.

Recently in 2018, By combining the k-means clustering method with the bio-inspired artificial bee colony (ABC) optimization technique, [5] created a hybrid recommender system. When compared to other systems already in use, the method is innovative and produces meaningful fallouts. The results of the experiment on the Movielens dataset shown that, by lowering the cold start issue, the proposed system offers remarkable speed and scalability as well as accurate tailored movie suggestions. On an advanced high-configuration computer, however, the system's performance may be assessed by including other crucial user attributes, such context and privacy with cross-domain data.

Similarly, [14] proposed a clustering strategy to successfully integrate multi-criteria ratings into conventional recommender systems. The results show that the suggested method outperforms the conventional collaborative filtering-based Pearson methodology in terms of accuracy and efficiency. Furthermore, [15] proposes a unique method known as RecDNNing that combines deep neural networks with embedded users and objects. The suggested RecDNNing works better than state-of-the-art methods, according to the experimental findings on MovieLens. Nevertheless, the study does not investigate more sophisticated deep learning techniques to improve the suggestion quality any further.

Recently in 2019, [16] a deep learning-based contextual hybrid method for session-based news recommendation that can make use of a range of information kinds was proposed.

ISSN No:-2456-2165

The advantages of taking into account other forms of information, such as article popularity and recency, are confirmed by the results. The coupling of content and context information with a sequence modeling method based on recurrent neural networks constitutes the work's primary technological contribution. However, outliers in the user profiles were not addressed.

Furthermore, [17] presented MCS optimization strategy to employ the MCFM clustering technique to give the target user an effective suggestion. When applied to grouped data points derived from the proposed MFCM clustering, the suggested MCS optimization technique outperforms alternative optimization algorithms. However, the study is unable to put into practice a user database-equipped web interface with a customized learning model for every user.

Similarly in the same year and same authors, In order to increase RS accuracy, [18] proposes a unique AGNN technique that uses the GA algorithm to adjust the weight of the ANN model and recommend goods to online targeted consumers using a new modified k-means approach. The suggested RS model outperforms the other models in the comparison in terms of recommendation outcomes. However, the research fails to incorporate different machine learning and clustering algorithms and study the comparative results.

Moreover, [18] suggests an Intelligent Recommender System (IRS) built on a Random Neural Network. IRS serves as a conduit between the user and other recommender systems, adjusting itself iteratively based on the perceived relevance of the user. On average, IRS outperforms the Big Data recommender systems after learning iteratively from its customer.

More recently in 2020, [19] use a number of techniques, including the birch, mini-batch, K-Means, mean-shift, affinity propagation, agglomerative, and spectral clustering algorithms, to create groups out of data. The study of clustering performance reveals that the K-Means approach performs well when compared to the birch algorithm (score of 1.24 on the Davies–Bouldin Index) and the Calinski–Harabaz Index (59.41). Nevertheless, a web-based user interface with a user database is not implemented by the research.

III. METHODOLOGY

This research proposes an innovative approach to building a hybrid recommender system that extracts information from relational and non-relational datasets by combining the neural network methodology of Self-Organizing Maps with the collaborative filtering (SVD) method. The new chart (research Model) has the following key component: knowledge base, learning module, clustering, classification, and decision manager.

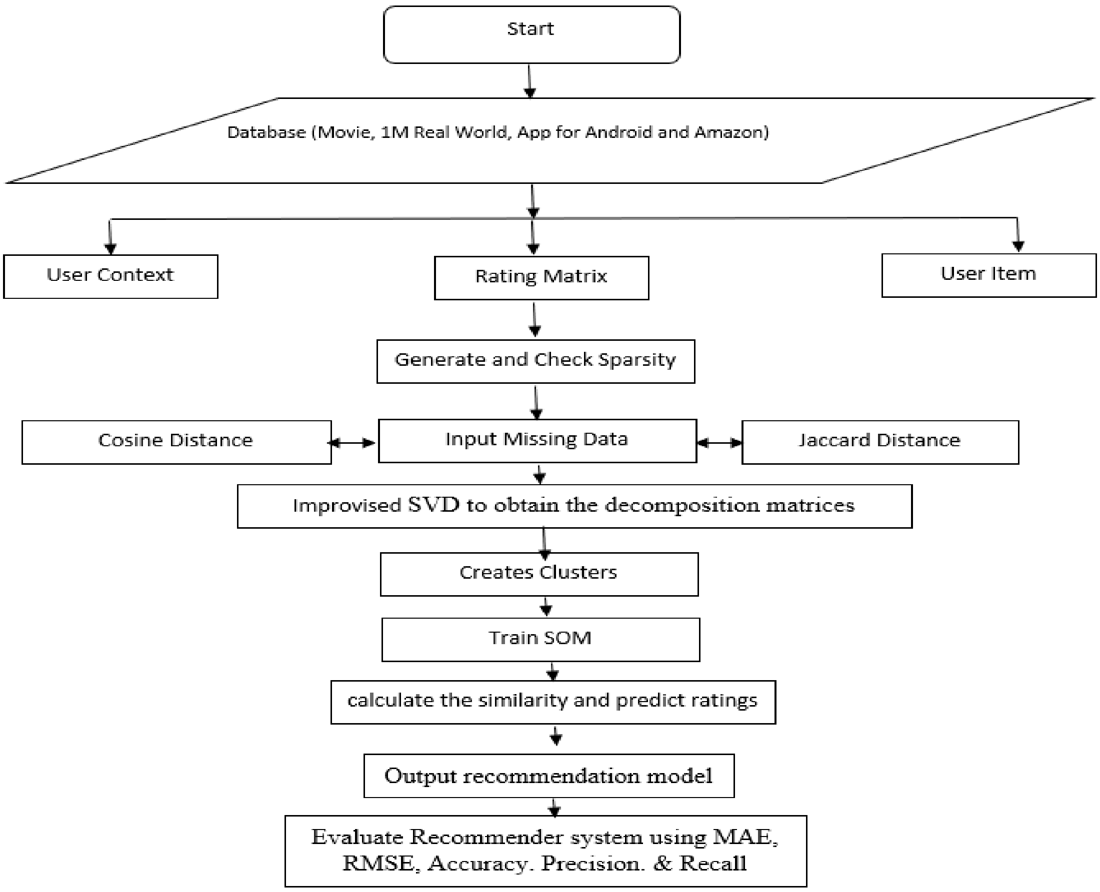

Fig. 1: System architecture

There is a lot of information in this business world. In the corporate world, information must be kept up to date for decision-making. Online analytical processing (OLAP) and online transaction processing (OLTP) are the two forms of data used in decision making. Whereas the latter just includes current business activities, the former includes historical data about the company going all the way back to its founding. Based on these two types of data, Utilizing a cutting-edge approach, we build a hybrid recommender system that blends the power of self-organizing maps and collaborative filtering expertise and user knowledge to improve business intelligence based on frequent item set mining and clustering. We can run the decision-making process. This model consists of three main phases: data preprocessing base on three main distance metric approach to create an improved SVD for filling data sparsity, followed by clustering, classification, and finally the recommendation phase is constituted.

On and above this, one serious drawback of the existing recommendation approaches is the fact that they do not give insights into the rating patterns. For example, after looking into several ratings from a user one cannot get an understanding on what might have motivated the user to give such ratings. We We We believe that the user rating matrix's singular value decomposition (SVD) will provide us with insights into these underlying ideas or patterns. The standard SVD does, however, have a flaw in that it performs poorly in cases when the rating matrix has missing values.

Furthermore, real-world recommender systems must deal with a significant degree of sparsity or unobserved value. Since zero values in a rating matrix have a specific meaning, estimating the missing values to be zero results in inaccurate predictions. Therefore, a different strategy would be to imputationally fill in the missing data using a prediction method. The most popular method is to insert column or row means into those values. Although recentering the matrix first would be a more efficient approach, the entire matrix so acquired may then be utilized for SVD computations. But mean loses accuracy since it doesn't take into account the underlying correlation of the data.

To overcome the aforementioned issues, we pre-processed the data by first determining the K-nearest neighbors of a specific user using three key distance measures, and then we filled in the missing rating values of the user in question based on their ratings.

The main aim of this phase is to improve the performance of SVD algorithm. For this purpose, we calculated the three different distance metrics namely, cosine, Jaccard and Euclidean.

Following the clustering phase, the chosen cluster is subjected to a similarity computation. The similarity between users and objects is determined by this similarity metric. Using the KNN base method, the outcome will be utilized to forecast the ratings of a missing value. This is intended to increase the accuracy of the outcome and lessen the sparsity of the ratings, thereby the improved SVD and self-organizing map neural network clustering technique is applied on the data sets to recognize and group the instances in the datasets. The SOM, SVD and KNN technique distributes input vectors into separated clusters by means of similarity and distance measurement. All input vectors are assembled into distinct centers by means of minimizing objective function. To create a ranked list of items, dimension reduction techniques like SVD will be used to the output of the similarity measurements. To determine how similar the target user or thing is to other users or products, it is based on previous ratings.

In the classification phase, collaborative filtering algorithm base on neural network (SOM) architecture is applied to instances of the database to extract the knowledge learned. To build a model, each classifier is used to classify the training data sets independently. The developed models are then used to make predictions for the testing data sets. Creating recommendations and predictions for the intended user is the final step. Recommendations are generated by the system based on the expected item frequency. The propose algorithm is depicted in below.

Algorithm 1: The Proposed hybrid SVD-KNN base SOM Recommender System

Input: User-Item rating matrix A, User-Item Context.

Output: Recommendation

Step 1: Input user item matrix including users' ratings data;

Step 2: Check Sparsity and input missing data based on new Jaccard & Euclidian base SVD Algo.

Step 3: Create user clusters using clustering SOM

Step 4: Calculate similarity of each cluster & compute the rating prediction.

Step 5: Use SVD to obtain the decomposition matrices of each cluster;

Step 6: For each matrix obtained from the decomposition step, apply context (user or item context) and calculate the similarity;

Step 7: Output Recommendations

Step 8: Evaluate Recommender system using MAE, RMSE, ACC. PREC. & REC.

A. Evaluation and Test

The recommender system has an answer thanks to the hybrid algorithm that has been suggested. This approach may be verified in a variety of experimental contexts. This assignment uses a several datasets from ecommerce industry set to test a proposed algorithm to encourage consumers to buy a product. This algorithm helps active users find the items they want to buy from their business. E-commerce businesses that use the recommendation system are Amazon.com, CDNOW.com, Drugstore.com, eBay, MovieFinder.com and Reel.com.

The offline evaluation method for in -domain recommender systems is an easier way to evaluate recommender systems. In this method, the data set containing user information, items and evaluations is divided into a training set and a validation set, a model training set and a model test validation set. System performance is further evaluated with a set of validations using different evaluation techniques. The offline evaluation method is the easiest to use because it provides an opportunity to weigh the recommended algorithms differently from each other. Thus, the performance of the proposed algorithm can be compared with the existing methodology using various metrics such as accuracy, recall and precision. This can be done by looking for qualities in different numbers of neighbors, iterations and clusters to ensure that the performance of the proposed method is better than the existing method.

The accuracy of the rating prediction was measured using Five (5) commonly used evaluation indicator namely Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Recall, F-metric, and Precision. The MAE and RMSE were obtained by calculating the difference between the true rating and the predicted rating. A smaller value corresponds to higher accuracy of the rating prediction. MAE can be calculated by the following formula.

$$MAE\,(Pred\,, act) = \sum_{i=1}^{N} |\frac{Pred_{u,i} - act_{u,i}}{N}| \quad \ldots(4)$$

The precision calculates the percentage of the item that matters in the outcome that was obtained. While the F-score or F-measure is a measure of a test's accuracy. Both metrics should be used in common. Precision and F-score can be calculated with the following formulas.

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

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

$$\text{F1} = 2* \frac{Precision * Recall}{Precision + Recall} \quad \ldots(7)$$

IV. RESULT

The proposed recommender system was developed using Anaconda and Jupyter Notebook and run-on Intel® Pentium Duo Core 1.6GHz system having a memory of 8GB, 500GB HDD and Windows 10 64 bit. The proposed recommender system is evaluated using MovieLens 100k. A comparison is made between the proposed recommender system and three of the most advanced traditional model-based CF methods. They include, the Capsule Neural Network CF algorithm, the KNN CF algorithm and the EM clustering base recommender system. As stated earlier, the accuracy, F-score, precision and recall are measured and evaluated in relation to other state-of-the-art methods.

As stated earlier, we pre-processed the data with Jaccard distance and Euclidian distance in terms of rating matrix. The aim here is to improve the performance of SVD algorithm. Therefore, in this subsection, we evaluate the performance vised SVD data imputation scheme against the normal SVD in terms of in terms of correct prediction using user-based similarity. In this article, we suggest three distinct techniques based on the improvised SVD+SOM. We quantify standard prediction accuracy using the MAE and RMSE between the anticipated and real ratings, and we give the experimental findings for each method's performance. By calculating the overlap, the decision support accuracy (Precision, Recall, and F-1) was further utilized to compare the suggested items with the pertinent ones. Performance Evaluation for All Methods at K=10.

In this subsection, we furthermore evaluate the performance of all the methods at k=10. Table 1 depict the emphatical results achieved in terms of MAE, RMSE, Precision, recall and F-score for the correct prediction using user-based similarity at K=10.

Table 1: Performance achieved for K=10 for Movie datasets

Method	**MAE**	**RMSE**	**Precision**	**Recall**	**F-Score**
Euclidian improvised SVD+SOM	0.7748	0.9964	0.7106	0.6793	0.6946
Jaccard improvised SVD+SOM	0.7306	0.9433	0.7452	0.7197	0.7322
Cosine Improvised SVD+SOM	0.7834	1.0032	0.7167	0.6540	0.6839

From Table 1 it is noticed that at k=10 Jaccard's improvised model attained the best and most stable predictive accuracy (MAE and RMSE) and decision support accuracy (Precision, Recall and F-Score). This result can better be analyzed using the following figures presented below.

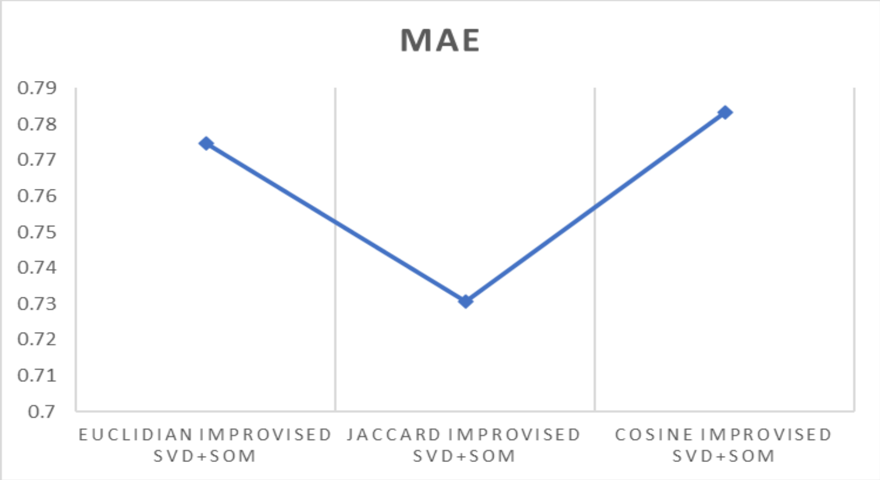

Fig. 2: MAE for all three methods at k=10

As sated earlier, This gap between the real and anticipated ratings was used to calculate the MAE. Greater accuracy in rating prediction is indicated by a lower number. Thus, for the MAE, lower values indicate less errors and better prediction accuracy. From Figure 2 it is seen that the Jaccard's distance model attained the best and most stable MAE by achieving the lowest MAE of 0.7306 followed by the Euclidean distance model which has 0.7748. the cosine distance model was the least performing model by attaining highest errors of 0.7834.

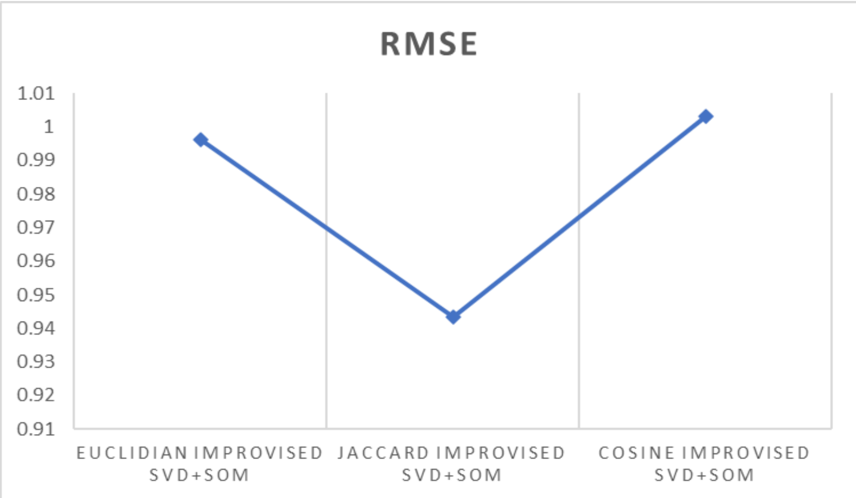

Fig. 3: RMSE for all three methods at k=10

Similarly, the RMSE was obtained by calculating the difference between the true rating and the predicted rating. A smaller value corresponds to higher accuracy of the rating prediction. Thus, for the RMSE, lower values indicate less errors and better performance by the model. From Figure 3. it is seen that the Jaccard's distance model perform better by attaining the least error RMSE of 0.9433 followed by the Euclidean distance model which has 0.9964. The cosine distance model was the least performing model by attaining highest errors of 1.0032.

Thus, from the analysis, it is easier to say that, at K=10, the Jaccard-SVD model performed better than all the three methods used in the study in terms of the predictive accuracy (MAE and RMSE). However, to further ensure the generalization of the prediction performance by each model, we further evaluate the proposed models using decision support accuracy such as precision, recall and F-score as presented in the following figures. For the decision support accuracy measures (Precision, Recall and F-score), it plays an important role for the multi-criteria recommender evaluations. Many metrics for this purpose are well known from the information retrieval area. The precision here measures the portion of items that are relevant within the received result. F-measure is a measure of a test's accuracy. For this research, the precision recall and F1 scores were reported within the range of 0-1. Higher values close to 1 means better decision support accuracy and lower values close to 0 means poor decision support accuracy. For better understanding and more intuitive discussion for decision support accuracy, the results are plotted in Fig. 4

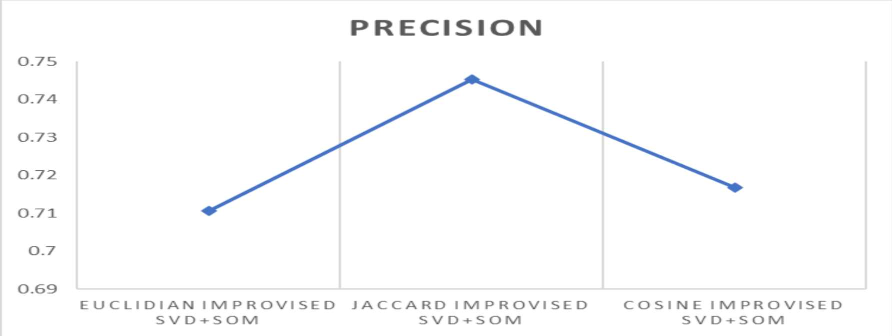

Fig. 4: Precision for all three methods at k=10

As previously said, the accuracy measures the part of the item that is significant to the result that was received. Precisions tell you how exact and precise your model is, as well as how many of the projected positives really turn out to be positives. A useful metric to identify large false positive costs is precision. The accuracy recall and F1 scores for this study were provided in the 0–1 range. Greater precision accuracy is indicated by values nearer 1, while poorer precision accuracy is indicated by values nearer 0. From Figure 4. the Jaccard base SVD model attain the highest score of 0.7452. however, the cosine base SVD came second by attaining precision score of 0.7167 while the least performing was the conventional Euclidean distance base SVD model which attains precision score of 0.7106. Similarly, the performance in terms of recall is depicted in Figure 5.

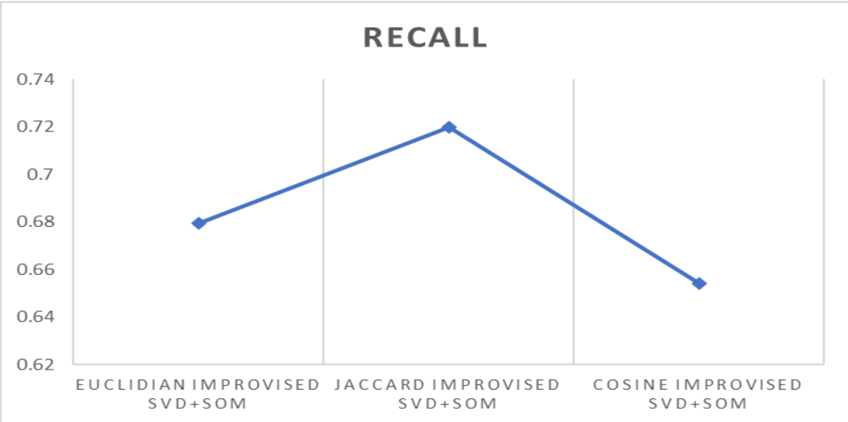

Fig. 5: Recall for all three methods at k=10

Recall, as previously said, is to determine the percentage of real positives that were accurately detected. The recall ratings for this study were provided in the 0–1 range. Better recall accuracy is indicated by higher values near 1 while worse remember accuracy is shown by lower values near 0. From Figure 5 the Jaccard base SVD model attain the highest score of 0.6793. however, the cosine base SVD came second by attaining precision score of 0.6540 while the least performing was the conventional Euclidean distance base SVD model which attains precision score of 0.7197. Finally, the performance in terms of F-score is depicted in Figure 6.

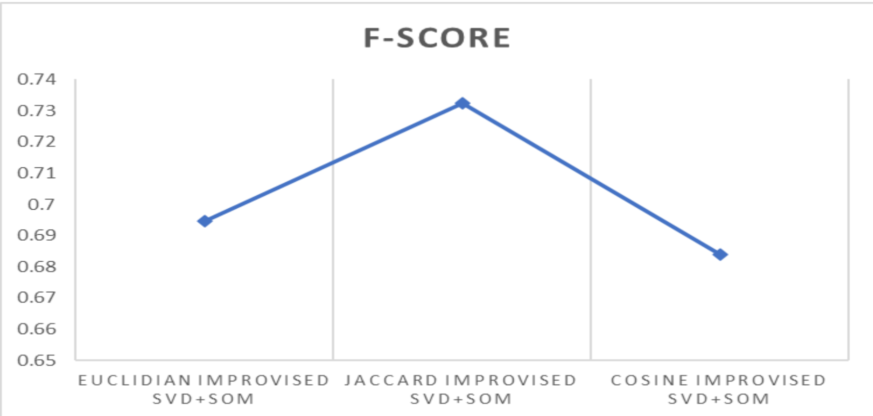

Fig. 6: F-score for all three methods at k=10

The F-Measure, as previously stated, is a function of both accuracy and recall; hence, it may be more appropriate to utilize in situations when we want to strike a balance between the two and there is an equitable distribution of classes (many real negatives). Thus, in this research, A test's accuracy is gauged by its F-measure, which was reported in the range of 0 to 1. A better F-score is indicated by higher values near 1, whereas a poorer F-score is indicated by lower values near 0. From Figure 6. the Jaccard base SVD model attain the highest F-score of 0.7322. however, the conventional Euclidean base SVD came second by attaining F-score of 0.6946 while the least performing was the cosine distance base SVD model which attains F-score of 0.6839. The summary for the performance of the proposed model across all the five metrics used in this study at k=10 is further depicted in 7 for easy comparison.

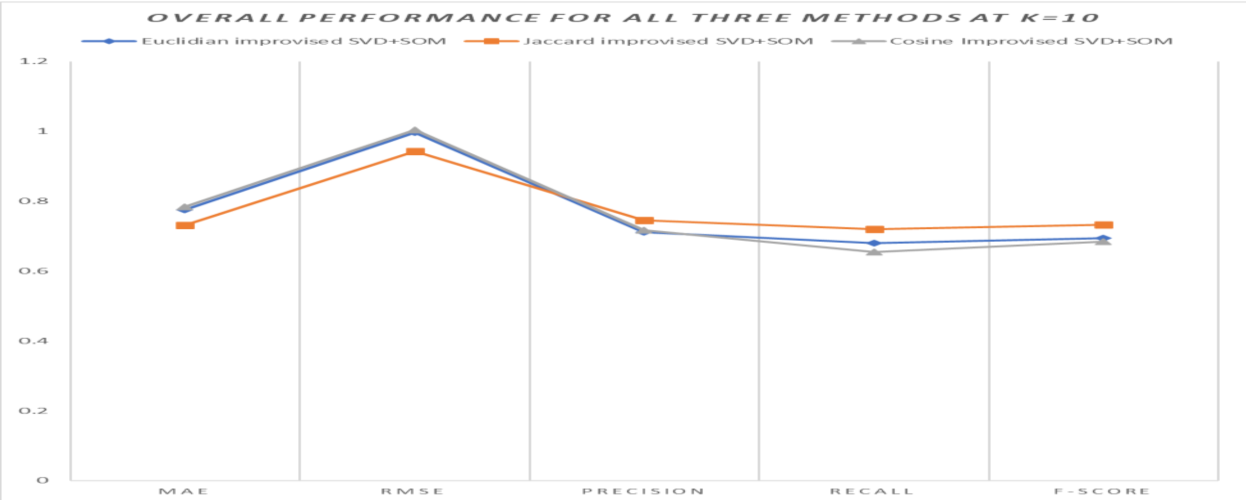

Fig. 7: Overall Performance for all three methods at k=10

From Fig. 7, the proposed Jaccard base SVD model attains the better performance in terms of all the five metrics (MAE, RMSE precision, recall and F-score) at K= 10. This was followed by the proposed cosine Improvised SVD+SOM which attains second best in terms of precision and recall, while for the MAE and F-score, the conventional Euclidean base SVD+SOM attain the second-best MAE, RMSE and F-scores respectively.

Performance Evaluation for All Methods at K=20

We furthermore evaluate the performance of all the methods at k=20. Table 2 depict the emphatical results achieved in terms of MAE, RMSE, Precision, recall and F-score for the correct prediction using user-based similarity at K=20.

Table 2: Performance achieved by all methods for K=20

Method	**MAE**	**RMSE**	**Precision**	**Recall**	**F-Score**
Euclidian improvised SVD+SOM	0.7741	0.9961	0.7107	0.6794	0.6947
Jaccard improvised SVD+SOM	0.7302	0.9429	0.7452	0.7198	0.7323
Cosine Improvised SVD+SOM	0.7830	1.0028	0.7168	0.6541	0.6840

From Table 2. it is noticed that at k=20 Jaccard's improvised model attained the best and most stable predictive accuracy (MAE and RMSE) and decision support accuracy (Precision, Recall and F-Score). It was also noticed that there is general improvement across all the three methods proposed in the study at k values increase from 10 to 20. This result can better be analyzed using the following figures presented below.

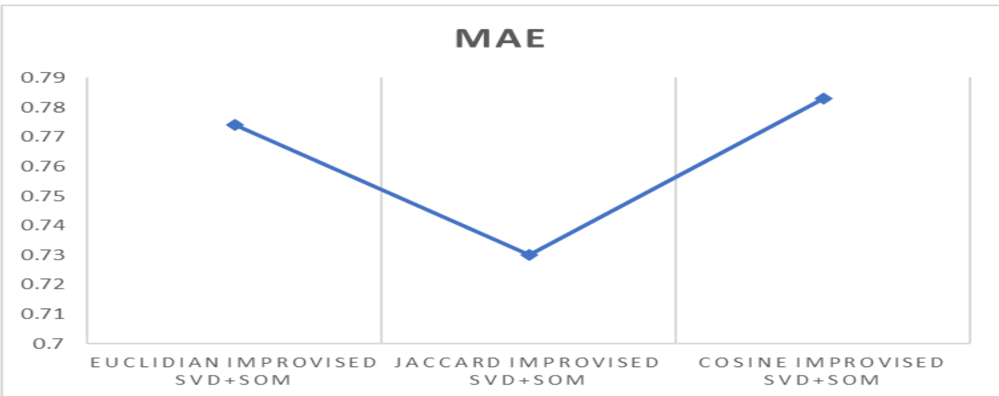

Fig. 8: MAE for all three methods at k=20

ISSN No:-2456-2165

As sated earlier, the MAE was obtained by calculating the difference between the true rating and the predicted rating. A smaller value corresponds to higher accuracy of the rating prediction. Thus, for the MAE, lower values indicate less errors and better prediction accuracy. From Figure 8. it is seen that the Jaccard's distance model attained the best and most stable MAE by achieving the lowest MAE of 0.7302 followed by the Euclidean distance model which has 0.7741. The cosine distance model was the least performing model by attaining highest errors of 0.7830.

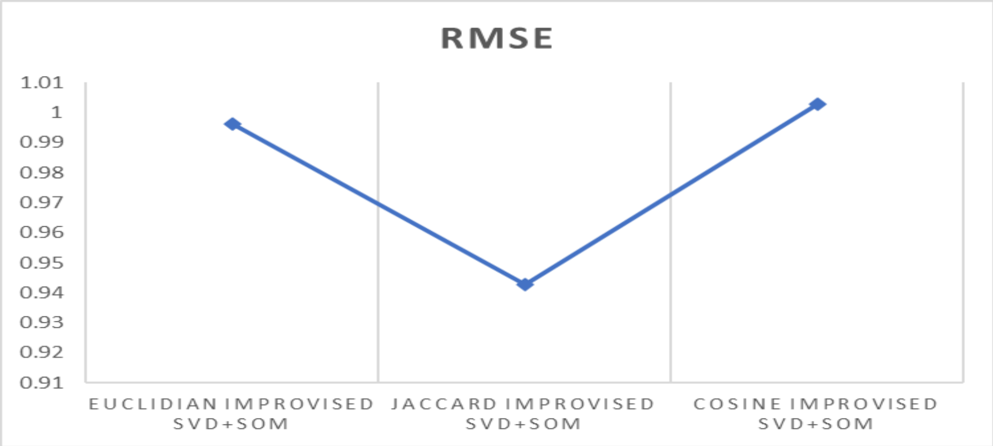

Fig. 9: RMSE for all three methods at k=20

In the same way, the difference between the real and anticipated ratings was calculated to get the RMSE. Greater accuracy in rating prediction is indicated by a lower number. Thus, for the RMSE, lower values indicate less errors and better performance by the model. From Figure 9 it is seen that the Jaccard's distance model perform better by attaining the least error RMSE of 0.9429 followed by the Euclidean distance model which has 0.9961. The cosine distance model was the least performing model by attaining highest errors of 1.0028.

Thus, from the analysis, it is easier to say that, at K=20, the Jaccard-SVD model performed better than all the three methods used in the study in terms of the predictive accuracy (MAE and RMSE). However, to further ensure the generalization of the prediction performance by each model, we further evaluate the proposed models using decision support accuracy such as precision, recall and F-score as presented in the following figures. It has a significant impact on the multi-criteria recommender assessments for the decision support accuracy metrics (Precision, Recall, and F-score). Numerous measures for this aim are widely recognized from the field of information retrieval.

Here, the accuracy quantifies the percentage of relevant elements in the obtained result. An indicator of a test's accuracy is the F-measure. The accuracy recall and F1 scores for this study were provided in the 0–1 range. Better decision support accuracy is indicated by higher values near 1, while poorer decision support accuracy is indicated by lower values near 0. The findings are shown in Fig. 10 to provide a clearer understanding and more intuitive discussion of decision support accuracy.

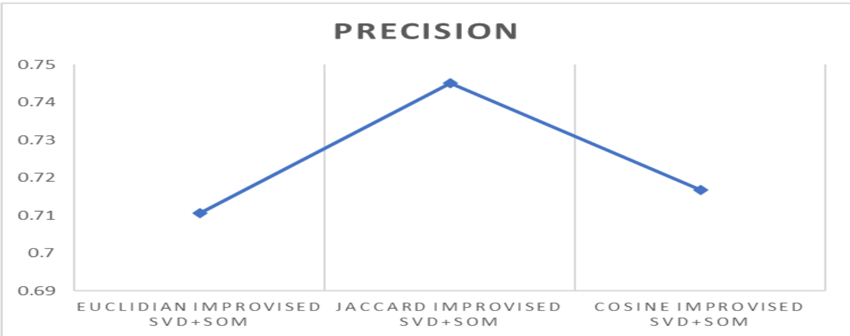

Fig. 10: Precision for all three methods at k=20

As previously said, the accuracy measures the part of the item that is significant to the result that was received. Precisions tell you how exact and precise your model is, as well as how many of the projected positives really turn out to be positives. A useful metric to identify large false positive costs is precision. The accuracy recall and F1 scores for this study were provided in the 0–1 range. Better precision accuracy is indicated by higher values near 1, while worse precision accuracy is shown by lower values near 0. From Figure 10. the Jaccard base SVD model attain the highest score of 0.7452. however, the cosine base SVD came second by attaining precision score of 0.7168 while the least performing was the conventional Euclidean distance base SVD model which attains precision score of 0.7107. Similarly, the performance in terms of recall is depicted in Figure 11.

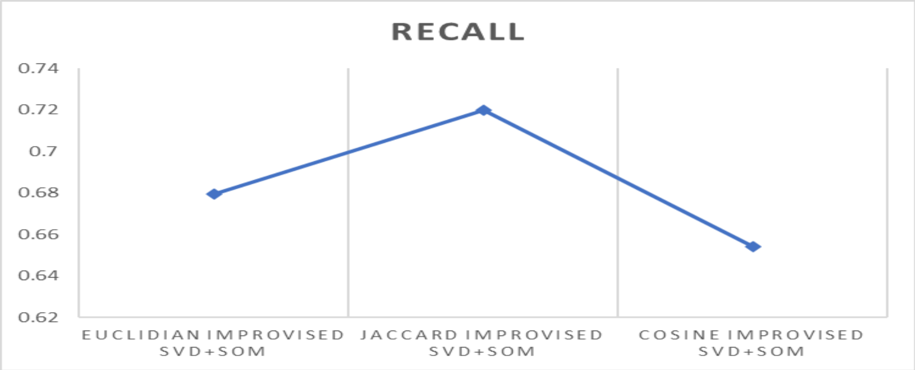

Fig. 11: Recall for all three methods at k=20

Recall, as previously said, is to determine the percentage of real positives that were accurately detected. The recall ratings for this study were provided in the 0–1 range. Better recall accuracy is indicated by higher values near 1 while worse remember accuracy is shown by lower values near 0. From Figure 11. the Jaccard base SVD model attain the highest score of 0.7198. however, the Euclidean base SVD came second by attaining precision score of 0.6794 while the least performing was the cosine distance base SVD model which attains precision score of 0.6541. Finally, the performance in terms of F-score is depicted in Figure 12.

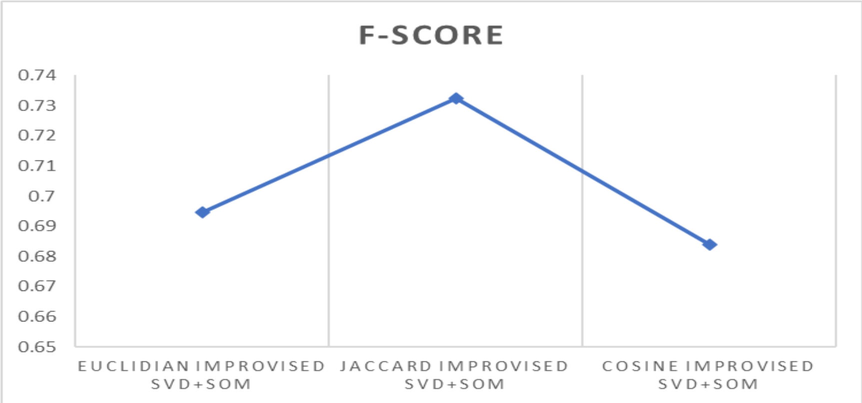

Fig. 12: F-score for all three methods at k=20

The F-Measure, as previously stated, is a function of both accuracy and recall; hence, it may be more appropriate to utilize in situations when we want to strike a balance between the two and there is an equitable distribution of classes (many real negatives). Therefore, in this study, the F-measure, which falls between 0 and 1, represents the accuracy of a test. An F-score that is closer to 1 indicates a better one, while one that is closer to 0 indicates a bad one. From Figure 12. the Jaccard base SVD model attain the highest score of 0.7323. however, the conventional Euclidean base SVD came second by attaining F-score of 0.6947 while the least performing was the cosine distance base SVD model which attains F-score of 0.6840. The summary for the performance of the proposed model across all the five metrics used in this study at k=20 is further depicted in 13 for easy comparison.

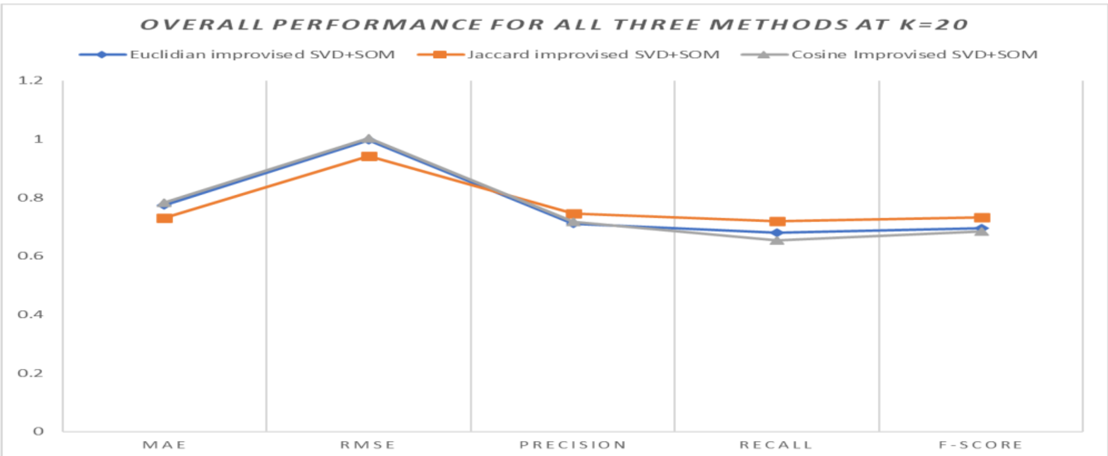

Fig. 13: Overall Performance for all three methods k=20

From Fig. 13, the proposed Jaccard base SVD model attains the better performance in terms of all the five metrics (MAE, RMSE precision, recall and F-score) at K= 20. This was followed by the Euclidean Improvised SVD+SOM which attains second best in terms of MAE, RMSE, recall and F-score with the exception of only the precision, where the proposed cosine base SVD+SOM attain the second-best score respectively.

V. CONCLUSION

In order to address the accuracy and sparsity issues with the current recommendation, this study proposed a hybrid collaborative base recommendation system that combines self-organized map neural networks (SOM) and an improved SVD. SOM clustering is used to group the dataset, an improved SVD is used to reduce dimensionality and improve sparsity, and a collaborative approach is used to address the accuracy and sparsity issues. The distance matrices were used to find K-nearest neighbors using the KNN algorithm from the fast KNN library.

The goal is to provide a way to enhance the SVD matrix's performance in the context of recommender systems. We use the suggested model in a variety of recommendation situations and datasets, including group, joint, and social recommendations. We evaluate the strategy over a range of K values for which the improvised method performs noticeably better than the standard SVD. Experimental result shows that, as the number of K values increase from 10 to 20, the predictive accuracy (MAE and RMSE) increases significantly across the three models respectively. On evaluating the performance of the proposed improvised SVD data imputation using two different distance metrics namely, cosine, Jaccard and Euclidean. the experimental result shows that, by using both cosine and Jaccard distance to improvised the SVD, the error has been further reduced compare to the conventional Euclidean SVD approach. Jaccard and Cosine distance base data preprocessing on the SVD has significantly improve on the general performance of the normal SVD algorithm in terms both predictive accuracy and decision support accuracies respectively. This result can be attributed to the fact that the improvised SVD preprocessing approach has significantly addressed the existing data sparsity issue of recommender system to some extend as against the conventional SVD preprocessing approach. This makes it easy for the SOM to cluster and generates the recommendations to the users with high precision and accuracy.

However, in our further study, we will evaluate the performance of the proposed model using K =30 and other different recommender datasets.

ACKNOWLEDGMENT

I wish to appreciate the expert support of my supervisors Prof. A. Y. Dutse, Prof. Abdulsalam Ya'u Gital & Prof. Aminu Ahmad.

REFERENCES

[1]. Liang, T.-P. and Y.-H. Liu, Research landscape of business intelligence and big data analytics: A bibliometrics study. Expert Systems with Applications, 2018. 111: p. 2-10.
[2]. Zheng, W., Y.-C.J. Wu, and L. Chen, Business intelligence for patient-centeredness: a systematic review. Telematics and Informatics, 2018. 35(4): p. 665-676.
[3]. Chen, H., R.H. Chiang, and V.C. Storey, Business intelligence and analytics: From big data to big impact. MIS quarterly, 2012: p. 1165-1188.
[4]. Niu, Y., et al., Organizational business intelligence and decision making using big data analytics. Information Processing & Management, 2021. 58(6): p. 102725.
[5]. Danlami, M., et al. Ensemble-Based Context-Aware Recommender System Using Clustering And Singular Value Decomposition. in 2023 3rd Asian Conference on Innovation in Technology (ASIANCON). 2023. IEEE.
[6]. Katarya, R., Movie recommender system with metaheuristic artificial bee. Neural Computing and Applications, 2018. 30(6): p. 1983-1990.
[7]. Gupta, U. and N. Patil. Recommender system based on hierarchical clustering algorithm chameleon. in 2015 IEEE international advance computing conference (IACC). 2015. IEEE.
[8]. Kumar, M., et al., A movie recommender system: Movrec. International Journal of Computer Applications, 2015. 124(3).
[9]. Ajesh, A., J. Nair, and P. Jijin. A random forest approach for rating-based recommender system. in 2016 International conference on advances in computing, communications and informatics (ICACCI). 2016. IEEE.
[10]. Frémal, S. and F. Lecron, Weighting strategies for a recommender system using item clustering based on genres. Expert Systems with Applications, 2017. 77: p. 105-113.
[11]. Nilashi, M., et al., A recommender system for tourism industry using cluster ensemble and prediction machine learning techniques. Computers & industrial engineering, 2017. 109: p. 357-368.
[12]. Soundarya, V., U. Kanimozhi, and D. Manjula, Recommendation System for Criminal Behavioral Analysis on Social Network using Genetic Weighted K-Means Clustering. J. Comput., 2017. 12(3): p. 212-220.
[13]. Paradarami, T.K., N.D. Bastian, and J.L. Wightman, A hybrid recommender system using artificial neural networks. Expert Systems with Applications, 2017. 83: p. 300-313.
[14]. Wasid, M. and R. Ali, An improved recommender system based on multi-criteria clustering approach. Procedia Computer Science, 2018. 131: p. 93-101.
[15]. Zarzour, H., Z.A. Al-Sharif, and Y. Jararweh. RecDNNing: a recommender system using deep neural network with user and item embeddings. in 2019 10th International Conference on Information and Communication Systems (ICICS). 2019. IEEE.

ISSN No:-2456-2165

[16]. Gabriel de Souza, P.M., D. Jannach, and A.M. Da Cunha, Contextual hybrid session-based news recommendation with recurrent neural networks. IEEE Access, 2019. 7: p. 169185-169203.

[17]. Selvi, C. and E. Sivasankar, A novel optimization algorithm for recommender system using modified fuzzy c-means clustering approach. Soft Computing, 2019. 23(6): p. 1901-1916.

[18]. Serrano, W., Intelligent recommender system for big data applications based on the random neural network. Big Data and Cognitive Computing, 2019. 3(1): p. 15.

[19]. Cintia Ganesha Putri, D., J.-S. Leu, and P. Seda, Design of an unsupervised machine learning-based movie recommender system. Symmetry, 2020. 12(2): p. 185.