International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-6, March 2020

Financial Inclusion: An Application of Machine Learning in Collaborative Filtering Recommender Systems

Girija Sankar Das, Bhagirathi Nayak

Abstract: ***In the current scenario in finance, data play a major role for predicting stock market as well as verious financial instruments. For the estimation of financial data, the various algorithms and models have been used. The use of the advising method has been used in this paper. The advising programs are one of the main methodologies used in the present market scenario with machine learning technologies. This paper focuses on the impact of financial inclusion in Odisha using a machine learning approach such as the classification of k-Nearest Neighbors (k-NN). For financial inclusion systems, machine learning has become a commonly used method. The result takes into the ATMs, Banks and BCs ranking in different districts of Odisha. We used the k-Nearest Neighbor's machine learning methodology classification algorithm to characterize the recommendation system based on users of the mentioned populations. Using our approach we equate conventional collective filtering. Our results show that the linear algorithm is more reliable than the current algorithm and is more efficient and stable than current methods.***

Keywords: ***Recommender System, Collaborative filtering, k-Nearest Neighbors.***

I. INTRODUCTION

Financial inclusion has been seen in many developing countries since the last decade as a key priority, given the importance of financial inclusion over social inclusion and inclusive development. The slogan of today's era is financial inclusion. The technology circles have developed a strong interest. Politicians and central bankers from around the world meet in various forums to speak about 'financial inclusion' and create a global structure that is more socially inclusive.

Financial inclusion is described as an efficient and timely accessibility mechanism for financial services and credits, that is necessary to access financial product by vulnerable groups, like the poorer and the low-income groups. Thus, we may be aware that the creation of an integrated financial sector contributes two things to reduce poverty, namely higher economic growth, which decreases deprivation & inequality indirectly and enhances healthcare & living standards through the establishment of sufficient, accessible financial services to poor people.

Manuscript received on March 15, 2020.
Revised Manuscript received on March 24, 2020.
Manuscript published on March 30, 2020.
* Correspondence Author
Mr. Girija Sankar Das*, Research Scholar, Sri Sri University, Cuttack, Odisha Email: girija.s2014ds@srisriuniversity.edu.in
Prof. (Dr.) Bhagirathi Nayak, Professor Sri Sri University, Cuttack, OdishaEmail:bhagirathi.n@srisriuniversity.edu.in

In India, many people, from poor background are not viewed equally by social institutions and financial institutions. In 1904, when the cooperative movement took place in India, the idea of financial inclusion was envisaged. This gained momentum by nationalizing 14 main trade banks in 1969, and the implementation of the lead bank system shortly afterward. In India, "Faster, more inclusive development" was the key slogan of the 11th Five Year Plan (2007-12). Financial sector inclusion, due to poverty, deprivation, and financial exclusion, is a very significant component of inclusive growth. It is a Hercules activity to test financial convergence by creating an index or by generating a range of indices. The existing literature has not been sufficiently rigorous to quantify financial inclusion. This work is not only researching financial inclusion policies and their role for inclusive growth but also suggesting an important component of inclusive growth keeping in view poverty, deprivations, and other financial exclusion socio-economic problems. Creating a new Odisha States Financial Inclusion Ranking with a more comprehensive Financial Inclusion Indexing is objective of our study in this paper. Here we are considering thirty districts of Odisha with urban, semi-urban and rural areas. We also covered population-wise ATMs, Banks, and BCs as of 31.3.2019. The districts are Anugul, Balasore, Bargarh, Bhadrak, Bolangir, Boudh, Cuttack, Deogarh, Dhenkanal, Gajapati, Ganjam, Jagatsinhpur, Jajapur, Jharshuguda, Kalahandi, Kandhamal, Kendrapara, Keonjhor, Khurda, Koraput, Malkangiri, Mayurbhagnj, Nabarangpur, Nayagarh, Nuapada, Puri, Rayagada, Sambalpur, Sonepur, and Sundargarh.

II. LITERATURE REVIEW

The literature suggests different approaches, including the use in econometric estimation of various dimensions of financial inclusion. “Beck et al. (2006)” made one of the first attempts to measure the financial sector's growth across continents. The authors developed new banking sector outreach metrics for three forms of deposits, loans, and payments (availability, accessibility, and eligibility). It can be difficult to integrate these components to determine the relative advances made by nations. A complex, open financial system lowers knowledge and transaction costs, has an impact on saving rates, investment decisions, technological development and so on.

Conceptual models used to demonstrate that a lack of exposure to financing would contribute to deprivation traps and inequality “Aghion & Bolton (1997), Banerjee & Newman (1993), Galor & Zeira (1993)”.

In the structured financial intermediary use of banking information, "Honohan (2008)" calculated a fraction of the adult population that accounts for more than nations, followed by association with inequality "Gini coefficient" and poverty. "Pande & Burgess (2003)", shows that the policy of rural Indians has decreased rural poverty considerably and improved non-agricultural employment. The plan of the Indian regional division extension has been dramatically reduced. Since India is a highly diversified economy and society, sufficient attention must be paid to financial inclusion assessment. Several researchers have sought to quantify other dimensions of financial inclusion.

The financial inclusion index was developed by "Sarma (2008)", using composite banking variables such as account numbers, number of bank branches and overall credit, and depositing GDP proportions into 55 countries. To calculate financial inclusion and then seek to classify the ties between financial and economic inclusion "Mehrotra et al. (2009)" developed the "Financial Inclusion Index (FII)".

In several papers and literature, we analyze the different dimensions of proposed structures and address some similar works in this section. Some collective filtering methods by the name of the author and the title of the article are represented by "Qian Wang, Xianhu Yuan, Min Sun" "Collaborative Filtering Recommendation Algorithm based on Hybrid User Model". Collaborative filtering confronts difficulties of adaptability and suggestion precision so the paper proposes a crossbreed user model to expel some of its downsides. The recommender framework in light of this model not just holds the benefit of suggestion accuracy in memory-based strategy, additionally has the scalability in the same class as a model-based technique. By "Chuangguang Huang and Jian Yin, Effective Association Clusters Filtering to Cold–Start Recommendations". This article focuses on the best way to combat the issue of cold starting in the modern Recommendations System (RS) analysis. For real electronic RS, CF does essentially not cope with the cold start issue in sparsity tests. The best-known technique for RS is collaboration filtering (CF) CF. The paper proposed a new filtering group classification.

1. Collaborative filtering

Collaborative filtering (CF) is a tool for producing suggestions using social intelligence. CF is a shared filtering mechanism. The recommendation is focused on the separation of individuals with shared interests. The study of the models of common interests among individuals and by analyzing ratings given by totally heterogeneous users or indirectly by examining the behaviors of the different users within the system shows analogous information using applied mathematics. This approach is quite different from the other form of filtering content and is most widely used. Despite recommending only things as a result, like items a user enjoyed in the past, different user habits are suggested. In collaboration with multiple applications, this knowledge searches for overlaps in users' mutual interest. CF is often called a social evaluation, which collects information by scores provided as social intelligence by other citizens. For eg, people would like to watch a film and could look for feedback from their peers or social networks. Few friends those have an interest in a similar type of movies or has already watched the movies may share their suggestion and these critics may be a deciding factor whether to watch the movie or not.

2. Collaborative filtering Methods

Methods of collective filtering Recommended products that related users like, allow for the discovery of different content recommendations.

Recommendation task

	I-1	I-2	I-3	I-4
U1	5	3	1	5
U2	?	?	?	2
U3	4	?	3	?

The U app package and several items that I must prescribe to the customers. Learn a feature based on the expected utility of any object $i(\epsilon I)$ in the past for every consumer $u(\epsilon U)$.

User-based k-Nearest Neighbors

	I-1	I-2	I-3	I-4
U1	5	3	1	5
U2	?	?	?	2
U3	4	?	3	?

Calculate user similarity, consider k users close to the user a. Not read by a reader a recommended book. The similarity of cosine.

$$sim(a,b) = \frac{a \,.\, b}{\|a\| \,.\, \|b\|}$$

This algorithm calculates the cosine or correlation similarity of rows (users) and columns (items) are k -Nearest Neighbors.

3. K-Nearest Neighbors

K-NN is an extremely effective grading algorithm that is used widely in machine learning. The learning algorithm is supervised and is, therefore, more accurate. In the pattern classification, cluster, estimation and intrusion detection programs K-NN is also used comprehensively. There are no hidden assumptions about sample distribution (like current algorithms like GMM, which presumes that there is a Gaussian distribution in this sample) which describe the sample as a non-parametric algorithm and makes it perfect for solving a wide variety of problems in real life.

K-NN, a supervised learning algorithm, stores the available information and attempts, by using distance functions as tools for measuring similarities among patterns, to identify the newer information. It is still used as a pattern recognizer and statistical estimator in various applications because of its classification and prediction precision.

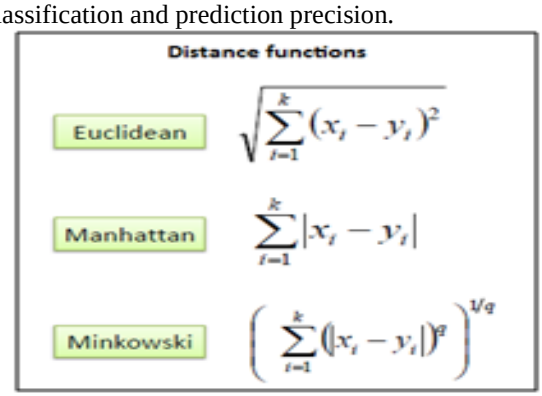


Figure 1: Calculate the distance functions

The above three functions for measuring distance can be found only with the variables that are constant in existence. The three functions don't suit the categorical variable well. If both categories and numerical variables comprise the data set, then the standardizing problem that occurs for the numerical variable between 0 and 1. The following formula shows the hamming distance.

Hamming Distance

$$D_H = \sum_{i=1}^{k} |x_i - y_i|$$

$$x = y \Rightarrow D = 0$$

$$x \neq y \Rightarrow D = 1$$

X	Y	Distance
Male	Male	0
Male	Female	1

Figure 2: formula of Hamming Distance

4. Machine Learning in Financial Inclusions

The application of machine learning in financial inclusion shows the technique's dominance as finance is the most complex field computationally where a vast range of parameters affect decision-making. In the prediction and suggestion over conventional approaches, machine learning techniques have become very popular. Machine learning algorithms are used for financial prediction, both supervised and unsupervised. Such models, which are state-based, economic models or even stochastic, are plagued by problems of over-fitting, heuristics and poor sample efficiency. This is why, with many influencing factors, the financial sector is extremely complex and non-linear. To resolve this condition, we analyze research conducted in Machine Learning invalidated areas of investment evaluation, collection of stock mutual funds, and delivery of potential inventory returns. Machine learning had in most cases been better measured and graded by self-learning and adaptability than its counter-statistic estimators.

5. Data Collection & Analysis

Here we use historical information from the Odisha website of Banking Data. Data have been collected from thirty districts of Odisha. in three different sectors such as Urban, Semi-Urban, and Rural areas. We have collected all ATMs, Banks, and BCs from all district wise with thousand Sq. Kilometer and 100000 population-wise. We require scores of 1 to 10 (total star ratings only), for Recommender Systems. We filtered the collected data and translated the data into a ranking of the specified weight. Districts have been used as a user ID and the Number of ATMs, Banks, and BCs are as the item ID. We then filtered collaboratively using rows (users), Columns (items) and suggests items according to ranking with the algorithm k-Nearest Neighbors. There is one file that has over 30 districts. We use collective filtering here to forecast user ratings which show in the following figures.

Open in Turbo Prep Auto Model Filter

Row No.	D_ID	DISTRICTS	TOTAL_BAN...	BANK_ID	RATINGS
1	1	ANGUL	247	1	4
2	2	BALASORE	298	2	5
3	3	BARGARH	157	3	3
4	4	BHADRAK	219	4	4
5	5	BOLANGIR	186	5	3
6	6	BOUDH	47	6	0
7	7	CUTTACK	594	7	7
8	8	DEOGARH	40	8	0
9	9	DHENKANAL	132	9	2
10	11	GANJAM	532	11	7
11	13	JAJPUR	290	13	5
12	14	JHARSUGUDA	132	14	2
13	15	KALAHANDI	139	15	2
14	17	KENDRAPARA	174	17	3
15	18	KEONJHAR	260	18	5

Figure 3: The initial ATMs data set with a rating

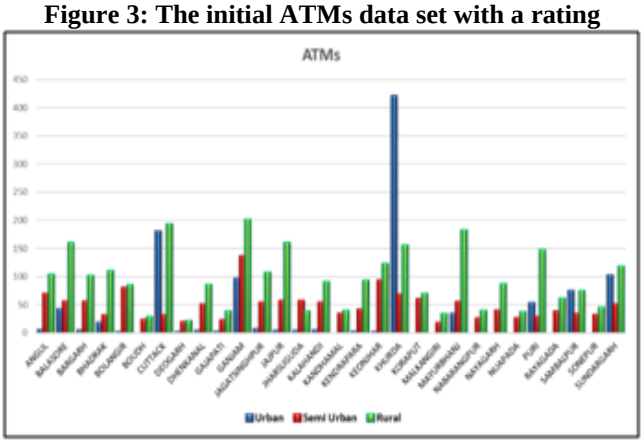


Figure 4: The initial ATMs data set from different sector

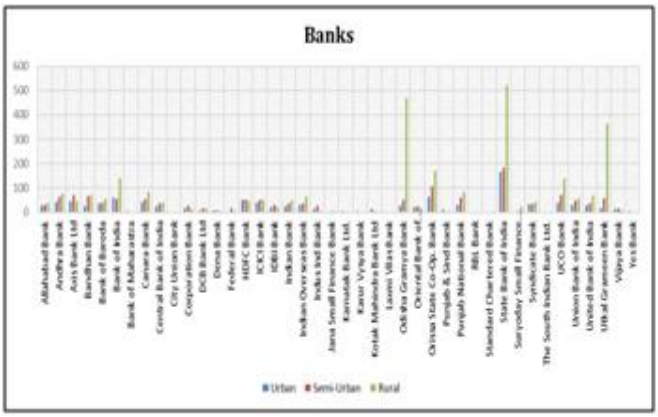


Figure 5: The initial Banks data set from different sector

The above "figure 4" shows that the total number of ATMs situated in different areas like Urban, Semi-urban, and Rural areas. Here we can observe the highest ATMs are situated in the Urban area of Khurda district, then Cuttack and Sundargarh District. Whereas other Rural Areas, the highest ATMs in Ganjam, Cuttack, and Mayurbhanj. Whereas we can see the number of banks is more in rural areas than the Urban and Semi-Urban areas, that shows in "figure 5". The highest number of State Banks of India, Odisha Gramya Bank then Utkal Grameen Bank. The total number of ATMs, Banks, and BCs graphical represents are given bellow.

Here we used the data set and construct the layout using the k-NN algorithm for collaborative filtering. We have selected the said data set and, given the weightage of label ranking. A model using the k-NN algorithm is generated using these divided results. The model shows in figure 9.

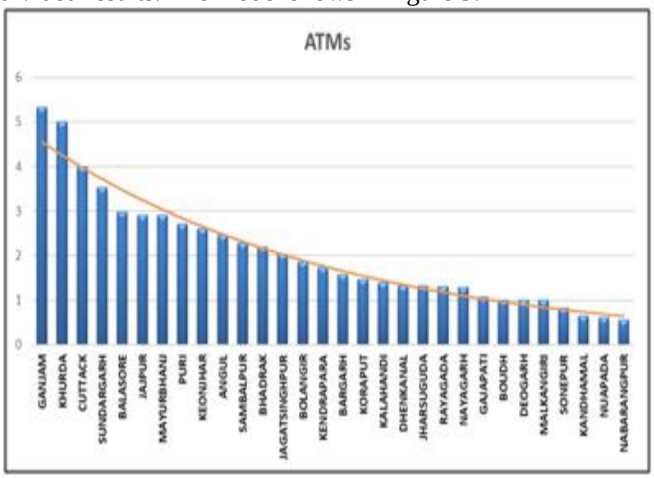


Figure 6: The ratings of the ATMs

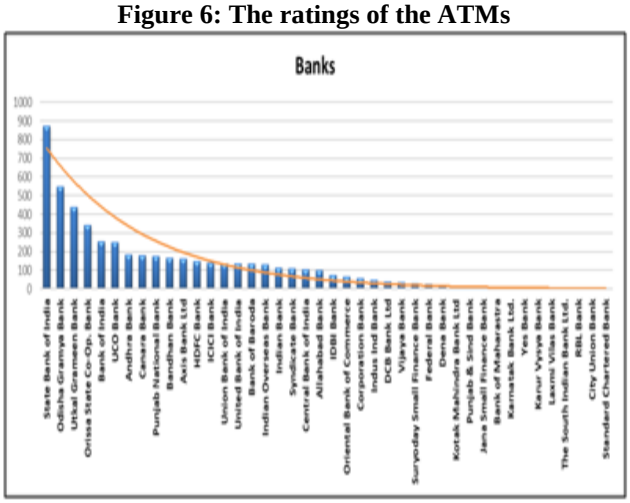


Figure 7: The total Banks

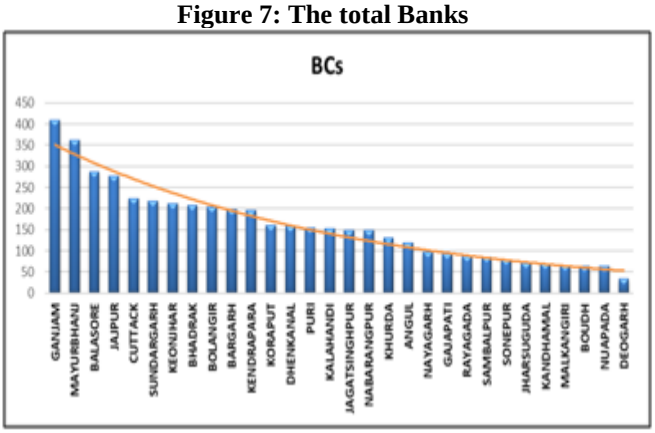


Figure 8: The total BCs

The above figures have shown the total number of ATMs, Banks, and BCs of thirty districts of Odisha. According to these data, we have developed a Collaborative filter model using the k-NN algorithm that shows in the following figure.

7.1 Collaborative filtration process workflow with k-NN

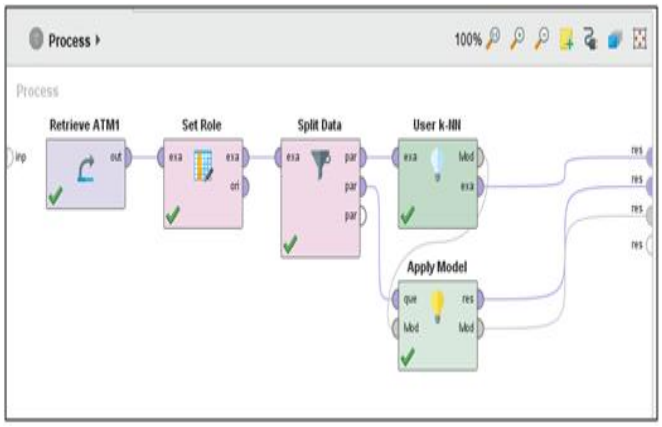
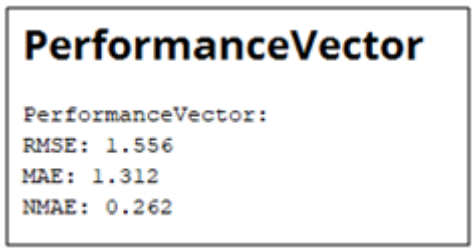


Figure 9: Collaborative filtration process workflow with k-NN

The workflow of the above model is used by the Defined Position operator to delegate different functions to the correct attributes. The function of user recognition was allocated to the user ID and the item ID identification role was assigned. Data attribute functions have to be assigned although they can randomly be named. The k-NN algorithm needs then to be conditioned by defining appropriate attributes roles using the available training data set. For the suggestion of new items, we used our qualified guide for our application process operator. Application Guide operator allocated the user ID function to the database collection before using the pattern. For each consumer in the database set, the application model operator returns a sample collection that includes the first n-grade recommendations. We analyzed the performance of the experiment after the development of this model. Before using the template, we have assigned the user ID function to the database array. The application model operator returns a list of samples for each user in a database set that includes the initial n-grade guidance. After creating this model, we evaluated the success of the study in the following figures.

7.2 Root Mean Square Error (RMSE)

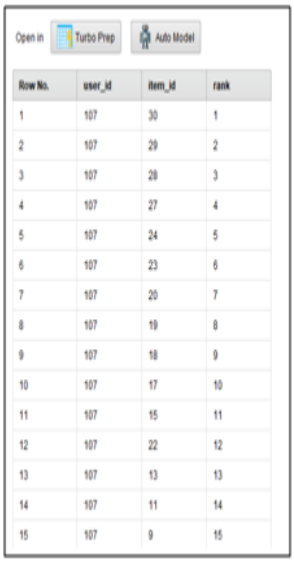
The RMSE of a prediction model concerning the estimated variable X*model* is defined as the square root of the mean squared error, where X*obsis* observed values and X*model* is modeled values at time/place *i*. It can be observed that Lesser values of RMSE and MAE give better accuracy. MAE is usually less than or equal to RMSE. The model set of expected results. It shows results for RMSE, MAE, and Normal Mean Absolute Error (NMAE). Fewer values from RMSE and MAE are shown to increase precision. Usually, MAE is equivalent to or less than



RMSE. If the two metrics are identical, all errors have the same value.

Figure 10: The statistics

In the above figure 10, we can observe the total number of Banks, minimum 40 numbers and maximum 1140 numbers and average 250 numbers. Total rating minimum 0 and maximum 10. The total districts 30 numbers. The result of the analysis is showing in figure 11, as per the Ranking of Indexing of Financial Inclusion (IFI).



Row No.	user_id	Item_id	rank
1	107	30	1
2	107	29	2
3	107	28	3
4	107	27	4
5	107	24	5
6	107	23	6
7	107	20	7
8	107	19	8
9	107	18	9
10	107	17	10
11	107	15	11
12	107	22	12
13	107	13	13
14	107	11	14
15	107	9	15

Districts	Ranking
KHURDA	1
CUTTACK	2
GANJAM	3
SUNDARGARH	4
BALASORE	5
JAJPUR	6
MAYURBHANJ	7
PURI	8
KEONJHAR	9
ANGUL	10
SAMBALPUR	11
BHADRAK	12
JAGATSINGHPUR	13
BOLANGIR	14
KENDRAPARA	15
BARGARH	16
KORAPUT	17
KALAHANDI	18
DHENKANAL	19
JHARSUGUDA	20
RAYAGADA	21
NAYAGARH	22
GAJAPATI	23
SONEPUR	24
KANDHAMAL	25
NUAPADA	26
NABARANGPUR	27
BOUDH	28
MALKANGIRI	29
DEOGARH	30

Figure 11: The Ranking of Indexing of Financial Inclusion (IFI)

III. CONCLUSION

The paper suggested a recommendation system, which has given the Indexing of Financial Inclusion (IFI) in thirty districts of Odisha. Khurda District has situated in rank one, Cuttack has in the second rank, whereas Deogarh is in bottom rank. The user preferences by combining interactive filters and computer teaching to supplement the traditional system of recommenders. This model can be used for testing various suggested performance evaluation algorithms. This paper only used a single model (e.g. k-NN), but for more consistent suggestions it is possible to apply more heuristic algorithms belonging to computer education on the same platform. This paper's expertise can be used to address other suggestion problems and improve the forecast quality and make good decisions in the future.

REFERENCES

1. Q. Wang, M. Sun, X. Yuan, (2010) "Collaborative Filtering Recommendation Algorithm based on Hybrid User Model", FSKD.
2. D J. Bobadilla, F. Ortega, A. Hernando, A. Gutierrez, (2013), "Recommender Systems Survey", Knowledge-Based Systems, 46, 109-132.
3. P. Resnick, H. R. Varian, (1997), "Recommender Systems", Communications of the ACM, 40, 56-58.
4. G. Zhuo, Jingyu Sun and Xueli Yu, (2007), "A Framework for Multi-Type Recommendations", Eighth International Conference on Fuzzy Systems and Knowledge Discovery,
5. Deshpande M. and Karypis, G., (2004), "Item-based top-n recommendation algorithms", ACM Trans. Inf. Syst. 22, 1, 143-177.
6. B. J., Heckerman D., and Kadie C., (1998), "Empirical analysis of predictive algorithms for collaborative filtering", In Proceedings of the Fourteenth Annual Conference on Uncertainty in Artificial Intelligence, page 4352.
7. B. N. Miller, J. A. Konstan, J. Riedl, PocketLens, (2004), "Toward a personal recommender system", ACM Transactions on Information Systems (TOIS), 22, 437-476.
8. Rajes Kumar Ojha, Dr. B. Nayak, (2018), "Application of Machine Learning in Collaborative Filtering Recommender Systems", 7 (4.38) 213-215.
9. Y. Jiang, J. Liu, M. Tang and X. (Frank) Liu, (2011), "An Effective Web Service Recommendation Method based on Personalized Collaborative Filtering", IEEE International Conference on Web Services.
10. M. O. Mohany, N. Hurley, N. Kushmerick and G. Silverstre, (2004), "Collaborative Recommendation: A Robustness Analysis", ACM Transactions on Internet Technology Vol. 4, 344-377.
11. X. N. Lam, T. Vu, T. Duc Le, A. Duc Duong, (2008), "Addressing Cold-Start in Recommendation Systems", Proceedings of the 2nd international conference on Ubiquitous information management and communication, 208-211.
12. Hao Ma, (2013), "An Experimental Study on Implicit Social Recommendation", ACM 978-1-4503-2034.
13. Ricci F, Rokach L, Shapira B, Kantor PB (eds) (2011), "Recommender Systems Handbook. Springer

AUTHORS PROFILE

Prof. (Dr.) Bhagirathi Nayak, Professor, Sri Sri University An immensely diligent and experienced Educationist, **Prof. Bhagirathi Nayak** carries a rich 25 years of competence in teaching at the Indian Institute of Technology, Kharagpur, India. A highly result-oriented professional, Prof. Nayak owns incredible expertise in the areas of Science and Engineering, Bioinformatics, Management, Database Designer and Developer, Teaching, Academic Administration and Pedagogical activities. The impeccable educator, Prof. Nayak has also been associated with Sri Sri University, Cuttack from August 2014. Prof. Nayak has been awarded several laurels for his contribution to the education sector, few being: Awarded by Dr. A.P.J. Abdul Kalam Research Centre, for his immense contribution to Data Science, on 22nd July 2019, at Calcutta; Awarded with 'Best Professor of Data Science in Odisha' on 18th August 2018 at Bengaluru by 'India Education Awards 2019', awarded with 'Best Researcher Award' on 15th September 2019 on Engineering Day by Odisha Technical Conclave 2019, and also awarded "Research & Innovation Award of the Year" on the occasion of National Summit on Women & Education Empowerment 2020 on 17th January 2020 at New Delhi in the august guest presence of Shri. Nitin Gadkari. He also bags the credit of achieving five International Patents in the area of Computers Science and Engineering. The extensive researching competencies of Prof. Nayak is a proud author of two books, published in 'MacMillan Publishers' and 'Allied Publishers', and has also authored two books of 'Linux Operating Systems' and 'Information Security' for Odisha State Open University.

Mr. Girija Sankar Das Mr. Das Ph.D. Student of Sri Sri University, Cuttack, Odisha. He is working as a General Manager (Finance) at Rourkela Steel Plant of Steel Authority of India Limited (SAIL), a MAHA RATNA PUBLIC SECTOR ENTERPRISE and having over 28 years of experience at the Senior Executive level in the Finance department with a successful background in generating increased profits, Budget & Costing, Financial Analysis Investment Analysis, Working Capital Management and establishing financial control. Accomplished Manager with excellent interpersonal and communication skills.