Unsupervised Learning Based Stock Price Recommendation using Collaborative Filtering

Pratyush Ranjan Mohapatra, Santosh Kumar Swain, Santi Swarup Basa

In this study, 17 stock market data were adopted for long term Prediction of stock price. Now days, Stock market data have got a significant role for invest finance in portfolio management. The various non-linear algorithms and statistical models are used for forecasting of financial data. In this article, we have used application of recommender system for this purpose. We primarily focused on use of machine learning algorithms for developing a stock market data recommender system. Machine learning has become a widely operational tool in financial recommendation systems. Here we considered the daily wise equity trading of Nifty 50 from National Stock Exchange (NSE) of 50 companies in 10 different sectors around 5986 days' transactions as data. We adopted k-Nearest Neighbors classification algorithm to classify users based recommender system. Collaborative filtering method uses for recommend the stock, the performance measure through RMSE, and R2. The result also reveals that k-NN algorithm shown more accuracy as compare to other existing methods.

Keywords : Classification, Collaborative based filtering, k-Nearest Neighbors, Machine learning.

I. INTRODUCTION

The secondary markets are the trade sources of listed securities for the stock market or the equities market. In stock market the Nifty 50 equity trading data is growing exponentially with respect to time. So it is difficult and takes more time to get the right information for which the traders look for. Users can find the information with their interests by identifying a company trading with the help of user interface. Users may not always know their interests beforehand and they might change their interest with passing time. This may require them to change their selection regularly. In present scenarios, machine learning methodologies play an important role to design the recommender systems. Recommender systems provide identifying information by knowing the user's interests from the hints of their interaction with that trading. Our purpose is to develop a recommender system model of finance which will reduce the difficulties of equity traders. Users always expect good recommendation due to the advances in recommender system. The bottleneck of the system is when the threshold of the trading is very low which makes it in-effective or poor quality of recommendation. For example, if a user would like to play a song on a music app and that app is not able to play that particular song then it is dumped by the user. It is the reason why app developing companies are putting lots of effort to make their recommender systems more effective and accurate. What makes the challenge harder is the versatility in the user preferences and availability of choices. In addition to that several influential factors affect the mood of the user and bring more randomness in the behavior of the user. Here we have used collaborating filtering to predict the Nifty 50 equity of daily trade quantity. This article is designed by taking NSE stock market data for collaborative filtering recommender system. The daily wise equity trading of 50 companies in 10 different sectors are considered and each sectors having top five companies.

II. LITERATURE REVIEW

We have gone through different research articles those reflect a relative study on different machine learning techniques used for prediction of stock market. Atsalakis G. S., Valvani K. P. [2] has done a survey on stock market prediction techniques. Chuangguang Huang and JianYin[4] described about the famous procedure of recommender system in collaborative filtering in his article. G. E. Kayakutlu, G. D TU[6] has introduced deep learning in fore casting the stock market and estimated it's effectiveness on the stock price of Google. They have showed that the performance of Recommender Systems (RS) was improved by using that framework. Qian Wang et.all [12] discussed the scalability and accuracy of recommendation systems. To reduce the limitations of recommendation systems, they have proposed a cross breed model based on combined item and demographic information. They have used the genetic algorithm by assigning weights to features of user to know the similarity of users. Then the similarity of users is used to enhance the accuracy and users having similar interests are used to increase the scalability of the system. Chuangguang Huang and Jian Yin [4] suggested a hybrid mechanism using clustering and filtering to resolve cold-start issue in RS. They proposed Association Clusters Filtering (ACF) algorithm which establishes clusters models based on the ratings matrix. For the recommendation of network service, they have also proposed a scheme of collaborative filtering.

III. COLLABORATIVE FILTERING

The CF (collaborative filtering hence forth to be referred as CF) is a method of creating recommendation using social intelligence or collective intelligence [7][8]. The recommendation is based upon the clustering of people with common interests. Using applied mathematics the analogous knowledge are discovered by analyzing the patterns among people of common interests and by evaluating the ratings provided by totally heterogeneous users or implicitly through observation of the activity of the different

Revised Manuscript Received on September 03, 2019

***** Correspondence Author

Pratyush Ranjan Mohapatra*, Department of CSE, KIIT Deemed to be University, Bhubaneswar, India. Email: pratyush.mohapatra@gmail.com

Santosh Kumar Swain, Department of CSE, KIIT Deemed to be University, Bhubaneswar, India. Email: sswainfcs@kiit.ac.in

Santi Swarup Basa, Department of CS, North Orissa University, Baripada, Odisha, India. Email: santiswarup.basa@gmail.com

users within the system [9]. This method is drastically different from the opposite and most ordinarily used content based filtering method. In spite of only recommending things as a result of they are just like things a user has liked within the past, things are counselled supported different user's preferences [18]. This information is matched with different users to seek out overlaps of common interests among users. CF is many times referred to as social recommendation, which filters information by using the ratings given by other people considered as social intelligence [13]. For example, people want to watch a movie, might seek for critics from his friends or his social networks. Few friends those have interest in similar type of movies or have already watched the movies may share their suggestion and these critics may be a deciding factor whether to watch the movie or not [11].

A. Collaborative Filtering Method

Analyzing the patterns among people of common interests and by evaluating the ratings provided by totally heterogeneous users or implicitly through observation of the activity of the different users within the system, the varied contents can be explored. Let U- Set of users and I- Set of items recommended to users.

Recommendation task

	I-1	I-2	I-3	I-4
U1	15	11	13	15
U2	?	?	?	12
U3	14	?	13	?

A function using the past data are to be proposed to predict an item i (i∈I) recommended to a user u (u∈U).

User-based k-Nearest Neighbors

	I-1	I-2	I-3	I-4
U1	15	11	13	15
U2	?	?	?	12
U3	14	?	13	?

Then similarity k-users of user a are to be found. Let recommended item is not utilized by user a.

Cosine similarity sim (a,b) = $\dfrac{a.b}{||a||.||b||}$

This algorithm computes cosine or correlation similarity of users or items and recommends items which follows k-NN.

B. K-Nearest Neighbors

k-NN is a very efficient classification algorithm which is extremely used in machine learning. It is a supervised learning algorithm hence it is more reliable also. k-NN is also exhaustively being used in pattern classification, clustering, prediction and intrusion detection systems [14].

No hidden assumptions are to be made about the sample distribution (like the contemporary algorithms such as GMM, that presumes a Gaussian distribution on the sample) which defines it as a non-parametric algorithm and makes it perfectly applicable in solving a wide range of real life problems [10].

IV. EXPERIMENTAL DESIGN

k-NN being a supervised learning algorithm stores the available information and tries to classify the new input by calculating their similarity with the existing cases by using distance functions as tools to measure similarity among patterns. Due to its accuracy of classification and prediction, it is still being used as pattern recognizer and statistical estimator in many applications [19].

Distance functions

Euclidean $$\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \quad (1)$$

Manhattan $$\sum_{i=1}^{k}|x_i - y_i| \quad (2)$$

Minkowski $$\left(\sum_{i=1}^{k}(|x_i - y_i|)^q\right)^{1/q} \quad (3)$$

It can be observed that the above three functions used to measure distance are only applicable with the variables those are of continuous in nature. These three functions do not fit well with variable of categorical type. If the dataset consists of both categorical and numerical variables then the standardization issue for the numerical variable within the range of 0 to 1 may come up.

Hamming Distance

$$D_H = \sum_{i=1}^{k}|x_i - y_i| \quad (4)$$

$$x = y \Rightarrow D = 0$$
$$x \neq y \Rightarrow D = 1 \quad (5)$$

X	Y	Distance
Male	Male	0
Male	Female	1

A. Machine Learning in Financial Engineering

Machine learning in finance proves the superiority of the technique, because finance is the most computationally intensive field where a large number of parameters are there to influence the decision making process [15]. Machine learning techniques have become very popular in prediction and recommendation over the traditional methods. Both supervised and unsupervised models of machine learning are being used for financial prediction [15]. These models are state based models, the econometric models or even the stochastic models are marred by the problems of over fitting, heuristics and poor out of sample results. This is because; the financial domain is highly complex and non-linear with a plethora of factors influencing each other. To solve this, if we look at the research done in Machine Learning in proven fields of investment recommendation, selection of mutual funds investments, prediction of future returns of the stock [16]. In most of the cases, Machine Learning had produced better prediction and classification than its counter statistical estimators due to its self-learning and adaptability capability [17].

V. METHODS

A. Data Collection

Secondary data has been collected from Nifty 50 daily trade equity from National Stock Exchange (NSE) website. We have collected one year's daily wise data from selected companies of 10 different sector, like i) FMCG ii) Banking iii) IT iv) Pharmaceutical v) Automobile vi) Construction vii) Finance viii) Power Industries ix) Steel and x) Telecom. Each sector has top five companies.

In this study, ratings used for recommendation system, which are on a rating of 1 to 10 (whole-star ratings only) [1][6].We filtered the said data set and took percentage of daily trade quantity of Nifty 50 data to and converted it to rating by given weightage. We used days as user id and companies as item id. After that we did collaborative filtering using rows (users), columns (items) and recommends items as per rating using k-Nearest Neighbors algorithm. There is one file: 'equity_trading.csv' having more than 59 thousand data. Here we are using collaborative filtering for prediction of user ratings, which is shown in figure 1.

Row No.	user_id	item_id	rating
1	DAY-1	ASHOKA LEYLAND	5
2	DAY-1	BAJAJ AUTO	7
3	DAY-1	HERO MOTORS	4
4	DAY-1	MARUTI	5
5	DAY-1	TATA MOTORS	4
6	DAY-1	AXIS BANK	4
7	DAY-1	ICICI BANK	4
8	DAY-1	KOTAK BANK	6
9	DAY-1	YES BANK	4
10	DAY-1	GAMMON INFRASTRUCTURE	6
11	DAY-1	HINDUSTAN	2
12	DAY-1	L & T	5
13	DAY-1	PUNJLLOYD	3
14	DAY-1	EDELWEISS	7
15	DAY-1	MAHINDRA FINANCE	4
16	DAY-1	MFSL	5
17	DAY-1	HINDUSTAN UNILIVER	7

Fig. 1.The initial data set with rating

B. Results & Discussion

For collaborative filtering, we adopt the data and developed k-NN based model. item_id for item information, user_id for user information and rating for label then splitting the total data in to two parts in different ratio. These split data are created a model using k-NN algorithm, which shows in figure 2.

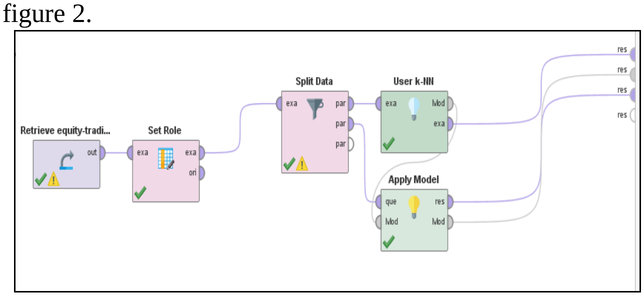


Fig. 2.The Workflow of Collaborative Filtering Model using k-NN

The above item recommendation workflow, the Set Role operator is being used to assign different roles to the attributes appropriate for them. The user_id has been assigned the role of user identification and similarly the item_id has been assigned the role of item identification. Roles for the data attributes must be assigned even though they can ne named arbitrarily. Then the k-NN algorithm has to be trained by setting appropriate roles to the attributes using the training data set available [19]. Here we have used Apply Model operator on the query set for recommendation of new items by applying our trained model. Before applying the model, we have assigned the user identification role to the query set. The operator returns first n-ranked recommended items for example set. After created the said model we checked the performance of the model, which shows in figure 3.

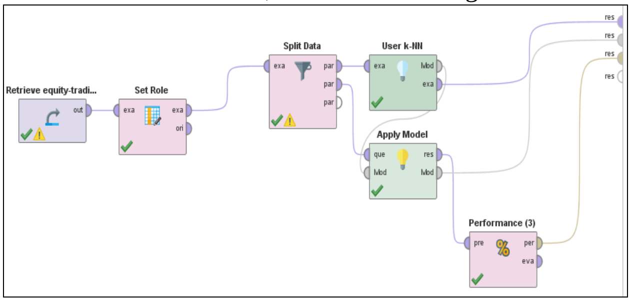


Fig. 3.The Collaborative Filtering Model Performance

C. Performance of the Model

The Performance operator has been used to calculate the error in recommendation. A set containing performance measure is returned as a performance vector by this operator. Both the Performance vector and the output dataset are shown in the figure 4 and 5 respectively. The performance of user-wise prediction is shown in figure 6.

Row No.	user_id	item_id	rating	predicti...
1	DAY-1	HDFC BANK	6	6.954
2	DAY-1	JAYPEE INFRATECH	5	7.393
3	DAY-1	BAJAJ FINANCE	4	4.003
4	DAY-1	PFC	5	3.889
5	Day-1	BRITANIA	6	5.504
6	DAY-1	MARICO	7	6.015
7	DAY-1	HCL	5	5.866
8	DAY-1	INFOSYS	8	6.419
9	DAY-1	SUN PHARMA	4	3.802
10	DAY-1	ZYDUSWELL	8	6.512
11	DAY-1	NTPC	7	6.715
12	DAY-1	TATA POWER	5	4.123
13	DAY-1	GAIL	4	5.149
14	DAY-1	SAIL	2	2.226
15	DAY-1	TATA STEEL	3	2.370
16	DAY-10	BAJAJ AUTO	6	5.126
17	DAY-10	HERO MOTORS	5	6.080

Fig. 4.Result of Collaborative Filtering

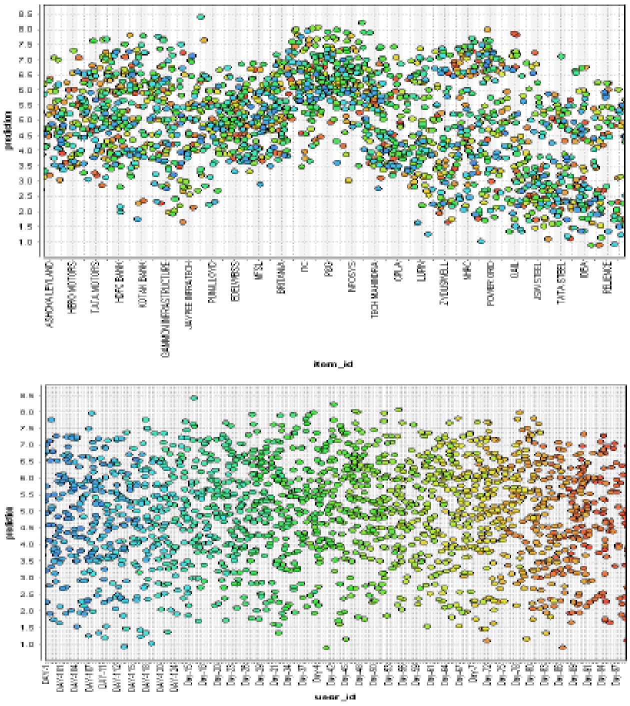


Fig. 5. Scatter plot of Item and User wise prediction

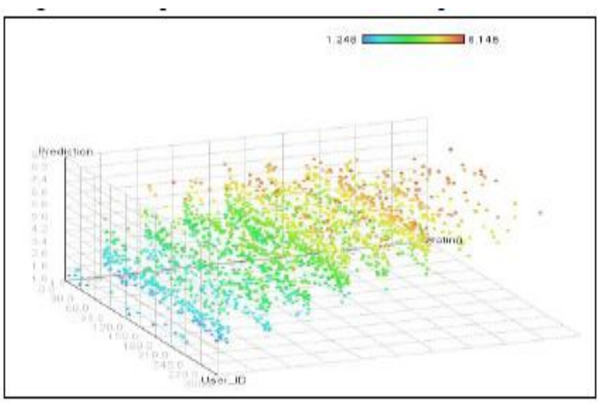
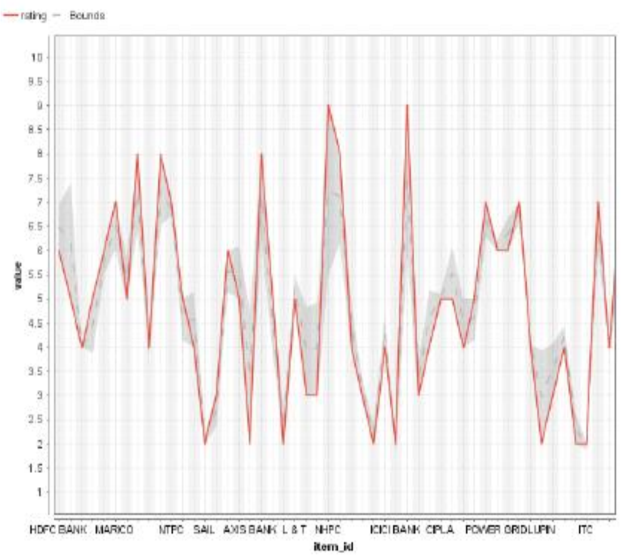


Fig. 6.3D plot User wise prediction



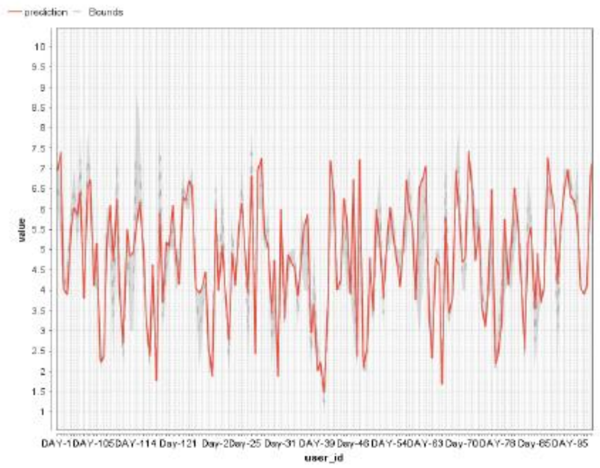


Fig. 7. Line graph of Item and User wise prediction

D. Root Mean Square Error (RMSE)

The RMSE of a prediction model is calculated by using Equ. 6.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_{obs,i} - X_{model,i})^2}{n}} \quad (6)$$

Where X_{model} estimated variable at time/place i .
X_{obs} is observed values

PerformanceVector

```
PerformanceVector:
RMSE: 1.148
MAE: 0.896
NMAE: 0.224
```

VI. CONCLUSION

Recommender systems occupied a key role in making predictions in the information and decision-overwhelmed world. It has not only changed the perspective of decision making by introducing group intelligence but also helped the implementers to maximize the financial gain by applying such social recommendation into e-commerce and finance. We also presented here how recommender system can use in finance [20][21]. This research may further used by companies equity recommendation for financial recommendation model, which may be utilized for recommending the different company's trade equity to the end user for understand the financial stock market. The result shows that the model can be used for large dataset by extending number of users. The result also reveals that k-NN algorithm revealed more accuracy as compared to other existing methods.

REFERENCES

1. H. A. Abdou, W. M. Abdallah, J. Mulkeen, C. G. Ntim, and Y. Wang (2017), "Prediction of financial strength ratings using machine learning and conventional techniques," *Investment Management and Financial Innovation*, 14(4), pp. 194-211.
2. G. S. Atsalakis and K. P. Valavanis (2009)," Surveying stock market forecasting techniques," *Soft computing method, Expert Systems with Applications: An International Journal*, 36(3), pp. 5932-5941.

3. A. Carzaniga, M. J. Rutherford and A. Wolf (2004), "A Routing Scheme for Content-Based Networking," *IEEE INFOCOM*, pp. 918-928.
4. Huang Chuangguang and Yin Jian (2010), "Effective association clusters filtering to cold-start recommendations," *Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2010)*, pp. 2461-2464.
5. Zhuo Guangping, Sun Jingyu and Yu Xueli (2011), "A framework for multi-type recommendations," *Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2011)*, pp. 1884-1887.
6. E. Guresen, G. Kayakutlu and T. U. Daim (2011), "Using artificial neural network models in stock market index prediction," *Expert Systems with Applications: An International Journal*, 38(8), pp. 10389–10397.
7. K. Maglaughlin and D. Sonnenwald (2002), "User Perspectives on Relevance Criteria," *American Society for Information Science and Technology*, 53(5), pp. 327-342.
8. Liew Jim Kyung-Soo; and Mayster Boris (2017), "Forecasting ETFs with Machine Learning Algorithm," *The Journal of Alternative Investments Winter 2018*, 20(3), pp. 58-78.
9. D. Olson and C. Mossman (2003), "Neural network forecasts of Canadian stock returns using accounting ratios," *International Journal of Forecasting*, 19(3), pp. 453–465.
10. SachiNandan Mohanty, j. Rejnia Pravin, K. Vinoth Kumar, K. C. Ramya, S. Sheeba Rani, S. K. Lakshmanaprabu (2019), "Optimal rough fuzzy clustering for user profile ontology based web page recommendation analysis," *Journal of Intelligent Fuzzy Systems*, DOI-10.3233/JIFS-179078.
11. R. D. McLean and J. Pontiff (2016), "Does academic research destroy stock return predictability?," *The Journal of Finance*, 71(1),pp. 5–32.
12. Wang Qian, Yuan Xianhu and Sun Min(2010), "Collaborative filtering recommendation algorithm based on hybrid user model," *Seventh International Conference on Fuzzy Systems and Knowledge Discovery*, 6, pp. 1985-1990.
13. Rose et al (2007), "Cobra: Content-based Filtering and Aggregation of Blogs and RSS Feeds," *USENIX NSDI*, pp. 29-42.
14. J. Bobadilla, F. Ortega, A. Hernando and A. Gutierrez (2013), "Recommender systems survey," *Knowledge-Based Systems*, 46, pp. 109-132.
15. F. Ricci, L. Rokach, B. Shapira, P. B. Kantor (eds) (2011), "Recommender Systems Handbook," *Springer US*.
16. P. Resnick and H. R. Varian (1997), "Recommender systems," *Communications of the ACM*, 40(3), pp. 56-58.
17. M. Deshpande and G. Karypis (2004), "Item-based top-N recommendation algorithms, *ACM Transactions on Information Systems (TOIS)*, 22(1), pp. 143-177.
18. J. S. Breese, D. Heckerman and C. Kadie (1998), "Empirical analysis of predictive algorithms for collaborative filtering," *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence (UAI' 98)*, pp. 43-52.
19. B. N. Miller, J. A. Konstan and J. Riedl (2004), "PocketLens: Toward a personal recommender system," *ACM Transactions on Information Systems (TOIS)*, 22(3), pp. 437-476.
20. B. Nayak and CA. VijayaBatth (2015), "Association Rules Mining using Apriori algorithm for work-related beliefs of Generation X and Generation Y," *International Journal of Scientific Research And Management*, 3(7), pp. 3324-3330.
21. Jiang Yechun, Liu Jianxun, Tang Mingdong and Liu Xiaoqing (2011), "An Effective Web Service Recommendation Method based on Personalized Collaborative Filtering,"*2011 IEEE International Conference on Web Services*, pp. 211-218.

AUTHORS PROFILE

Pratyush Ranjan Mohapatra currently pursuing PhD under KIIT Deemed to be University, India. He is member of Institute Of Engineers and ISTE. He obtained BE (Computer Sc & Engineering) from BPUT Odisha & MTech from Berhampur University, India. His research mainly focused in Machine Learning, Business Data Analysis & Computational Intelligence. He has published a few research papers in different International and National Journals.

Santosh Kumar Swain currently working as Professor in School of Computer Engineering, KIIT Deemed to be University, India. He is having 12 years of experience in the field research. He has published more then 30 research papers in various international and national journals. He has guided 3 PhD scholars and authored 2 books. His research is focused in Software Engineering, Web Engineering, Data Mining & Machine Learning.

Santi Swarup Basa is currently working as Assistant Professor in Department of Computer Science, North Orissa University. He completed his M.Tech (Comp.Sc.) from F.M.University, Balasore and completed his Ph.D from North Orissa University, Baripada. His Research areas are Software Engineering and Soft Computing. He has published a few research papers in different International and National Journals.