

# Data Analytics and Data monitoring Based on Database Recommendation – A Comparison

Pooja Mudgil<sup>\*1</sup>, Paras Jain<sup>2</sup>, Vikas Singh<sup>2</sup>

<sup>\*1</sup>Assistant Professor, Department of Information Technology, BPIT (GGSIPU), Delhi, New Delhi, India

<sup>2</sup>Department of Information Technology, BPIT (GGSIPU), Delhi, India

## ABSTRACT

A comparison of various analysis algorithms based for recommendation systems used in the market and businesses has their usage without considering the fact that if used with correct algorithm can increase its efficiency. Comparison is obtained in this paper after studying and researching various obtained algorithms present currently. This paper presents an introduction to the concept of recommendation systems which are recently working on different domain and then comparison is made, that too employed for scalability, reliability, faster process and efficiency.

**Keywords:** Data Analytics, KDD, java, k-means, Apache Hadoop tool.

## I. INTRODUCTION

Our world has advanced itself with increase in development with technology and constant growth towards science and business field. In these environment, if the interaction with the public is not maintained it can have its major impacts over the corporation and society then, as well. Keeping this in mind, the developers moved to the fact of understanding the choice of people, their liking, their possibility of choosing certain product and have come to many software's that ignited the profit of business handlers through technology via internet, the source of change in society. From the choices of buying products and obtaining products for the warehouse is a concern to business groups. This choice itself is a problem that has been solved with the systems that are used now a days known to be recommendation system that can analyse huge datasets usually available with the business organisation and industries with the optimised options, using opinions of users, public and daily usage or buying over e-commerce helping to gain huge profits by maximum

throughput and efficiency, in turn providing profits to each individual in its own way. In this, companies have been using the algorithms that are usually used in the market without identifying the best method or the best efficient performance gaining factor. Recommendation recommended to correct public with correct product at the particular time can change the factor as well as margins to minimum with much more satisfied customers. There are various methods that could be used by human operations and outcomes through calculations but would not be as perfect and precise to the point till now the developers are able to obtain results with values as through the computer software-based recommendation systems. These could advance the cycle of user's likings, product arrangement and showing the only type of product the customer requires or usually find the need to use of. It gives an easiest method to access the suitable and best option while in hurry. Recommendation system is a newest tool that has helped in navigating people for information related to the ones they prefer or like. Mostly, they are used in the areas of online shopping,

entertainment domains like movie, music yet their applicability is to be need of research. With the advent of big data and analysis, the option of letting things organise in the way the company wants to show to its customers has become much easier. Visitors over internet who e-commercialise led business to gain its explosive growth. There have been many systems developed based being it DevilFinder and Altavista that prioritise and personalise according to the information required to be retrieved as in [1]. This not even increased the demand but excited the developers to think more and scientists worked over the various problems to deal in with it be a medical medicinal based recommendation or home products based, a furnishing company based or graduation books based. The process of all the systems involved occurs through the filtering of the overload information segregating the vital fragmented information that has been collected through various interests, suggestions, surveys and behaviours observed surfed via internet. It provides the ability to let the people feel the comfort ness in interacting with the interface dealing and suggesting their own choices and requirements. This has been worked on using various algorithms usual ones are collaborative filtering, naïve bayes based and KNN, the typical ones used usually in for recommendation systems maybe it for shopping purposes or business products.

## II. LITERATURE SURVEY

A system defined for the handling of problems and decision making, as a means of assisting and augmenting the social process of using suggestions of others to implement and analyse the best or suitable answer to a particular option available and using experiences can handle the problem very easily as in [2]. Approaches have led in binding the exclusive content and services recommendations utilising algorithms like collaborative, content-based and other mixed filtering methods as in [3]. Ringo and

grouplens, type of similar content based filterers are available in the market all based on collaborative filtering methods assisting users to localise the articles as in [4][5]. Even company like amazon uses one of the diversified algos to work and improve their commercial usage with recommendations as in which uses the same collaborative filtering method for overcoming the method of scalability and for similarity basis items recommendation.

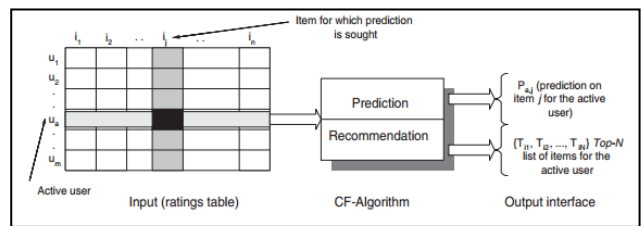


Figure 1 : This shows the collaborative algorithm process as in [6]

Then occurs another algorithm naïve Bayesian with usage of user interface, assisting for tracking as well as prediction of pages for the browsing as interest is based. It is one of learning classifier based model itself that works in real time [7]. As in [8] acts as an intelligent agent working for prediction allowing user to rate and then help in suggesting the suitable option in real time basis. The success has been limited of the popularity of algorithm KNN because of its usage in simple problems. It becomes useless with the data size increase and complexity increase turning the accuracy lowest. There has been development of certain hybrid systems that tried solving problems of KNN and collaborative algorithm. For mitigation some of the problems as said are approached through collaborative filtering but that turns out to be worse, showing cold start with problems in sparsity and scalability. Thus, hybrid collaborative, which combines two or more filtering techniques in different ways in order to increase the accuracy and performance of recommender systems has been proposed as in [9]

Evaluation Criterion	No. of neighbors considered			
	k=6	k=7	k=8	k=9
Hamming Loss	0.197	0.197	0.197	0.197
One-error	0.241	<b>0.239</b>	0.248	0.251
Coverage	6.374	<b>6.302</b>	6.357	6.424
Ranking Loss	0.170	<b>0.168</b>	0.171	0.173
Average Precision	0.758	<b>0.761</b>	0.756	0.755

Figure 2 : The table shows the performance of KNN over the yeast data with different values of k as in [10]

These techniques might harness to strengthen the computation while has been turned by certai hybrid collaborative combinations. If taken an example as given in [11], counts played by music system, recommendation is added with it can refine the definition of allowing people to hear the type of music they like, they feel like listening and try to listen usually, maybe it be of same genre or not.

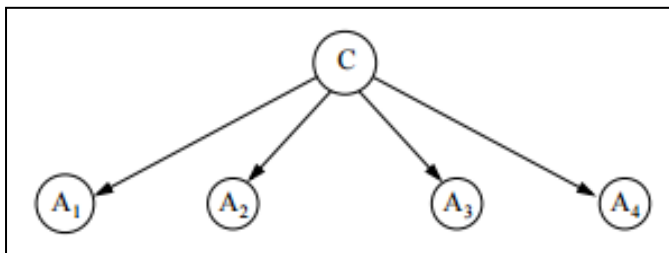


Figure 3 : shows the node connection to its class as direct connection in naïve bayes as [12]

Another classifier Naïve Bayes which has been as defined to prove the best and effective one for most of the applications because of its perfect estimations provided for various univariate probabilities conditional based problems. In paper [13][14] technique has been described that compromises between various number of selected variables and performance classifiers.

### III. RESEARCH WORK

From the obtained literature survey and researches, classification and comparison is required to obtain the best suitable algorithm for the best possible problem. The following table 1 describes the

comparison obtained while working through different papers and survey.

KNN	Collaborative Filtering	Naive Bayes
Discriminative classifier	Classifies on the basis of other's choices	Eager learning classifier
Supervised as well as lazy classifier, proves difficult to use in prediction for real time	Separates on the basis of similar entities	Assumes conditional independence between features and takes probabilistic estimation for each class
Inherent nature of analysing locally	Content based	Keeps on learning overtime
Usually focused on finding similarity	Data sparsity problem occurs	Much faster than KNN
More complex decision boundaries than complex trees	Limited content analysis is possible	Inherent nature for generative classifier
It over fits with more complexity	Shows scalability problems	Can work best in real time
Accuracy decreases with more complexity	Reduce quality of recommend-dations	Its accuracy increases with more learning
Too slow	Slow start	Faster in comparison to KNN
Can be best with PCA, SVD and usually used in small	Similar items based situation are suitable for its usage,	Can be best in situation where new addition of data keeps on

complexity of $O(n^2)$ , $n$ = no, of data points	usually used by companies like amazon	happening for example of email spam with complexity $\theta(1)$
---	---------------------------------------	---

**Table 1 :** The table explains the comparison done to explain the best suitable algorithm for different situation based problems.

#### IV. CONCLUSION

According to the obtained collection of data variants and comparison based on surveys and researches with the researched work, concludes us with the fact that each and every algorithm is concise with the different sets of problems occurring in the software environment varying with the requirements.

KNN is best suitable in situations where data size does not increase and simple problem is present while collaborative can be best suitable in situations where there occurs similar entities and user choices are concerned but the scalability is less only.

Consequently, Naïve Bayes came out to be best usual algorithm for the datasets where there is increase in new information, there is recommendation requirement while the tasks are still going on and keeping the tasks and situations updated.

Thus, every algorithm which was considered in this paper has its own usage with the varying situations and problems.

#### V. REFERENCES

- [1]. F.O. Isinkaye, Y. O. Folajimi, B.A. Ojokoh, Recommendation systems: principles, methods, and evaluation, published in Science direct journal of Egyptian Informatics
- [2]. P.Resnick, H.R. Varian, Recommender systems published in Commun ACM, 40(3)(1997), pp. 56-58, 10.1145/245108.24512
- [3]. A.M. Acilar, A. Arslan, A collaborative filtering method based on Artificial Immune Network published in nExp Syst appl, 36(4)(2009), pp. 8324-8332
- [4]. L.S. Chen, F.H. Hsu, M.C. Chen, Y.C. Hsu, Developing recommender systems with the consideration of product profitability for sellers published in Int. J Inform Sci, 178(4) (2008), pp. 1032-1048
- [5]. G. Adomavicius, A. Tuzhilin, Toward the next generation of recommender system. A survey of the state-of-the-art and possible extensions, published in IEEE Trans Knowl Data Eng, 17 (6) (2005), pp. 734-749
- [6]. Mukund Deshpande and George Karypis, Item-Based Top-N Recommendation Algorithms, University of Minnesota, Department of Computer Science Minneapolis, MN 55455
- [7]. P. Langley, W. Iba, and K. Thompson. An analysis of Bayesian classifiers. In 10th national conference on Artificial Intelligence, pages 223–228. AAAI Press, 1992.
- [8]. M.J. Pazzani, A framework for collaborative, content-based and demographic filtering, Artific Intell Rev, 13 (1999), pp. 393-408 No. 5(6)
- [9]. IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, VOL. SMC-15, NO. 4, JULY/AUGUST 1985 A Fuzzy  $\Lambda$ -Nearest Neighbor Algorithm JAMES M. KELLER, MICHAEL R. GRAY

- [10]. G. Murat, G.O. SuleCombination of web page recommender systems, *Exp Syst Applicat*, 37 (4) (2010), pp. 2911-2922
- [11]. Isabelle Guyon and Amir Saffari, Compression-Based Averaging of Selective Naive Bayes Classifiers, *Journal of Machine Learning Research* 8 (2007) 1659-1685
- [12]. Konstan I, Stathopoulos V, Jose JM. On social networks and collaborative recommendation. In: *The proceedings of the 32nd international ACM conference (SIGIR'09)*, ACM. New York, NY, USA; 2009. p.195–202.
- [13]. The Optimality of Naive Bayes Harry Zhang Faculty of Computer Science University of New Brunswick Fredericton, New Brunswick, Canada
- [14]. J.A. Hoeting, D. Madigan, A.E. Raftery, and C.T. Volinsky. Bayesian model averaging: A tutorial. *Statistical Science*, 14(4):382–417, 1999.
- [15]. P. Langley, W. Iba, and K. Thompson. An analysis of Bayesian classifiers. In *10th national conference on Artificial Intelligence*, pages 223–228. AAAI Press, 1992.

**Cite this article as :**

Pooja Mudgil, Paras Jain, Vikas Singh, "Data Analytics and Data monitoring Based on Database Recommendation - A Comparison", *International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT)*, ISSN : 2456-3307, Volume 5 Issue 2, pp. 1166-1170, March-April 2019. Available at doi : <https://doi.org/10.32628/CSEIT1952312>  
Journal URL : <http://ijsrcseit.com/CSEIT1952312>