



# International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 7, Issue 3, March 2019

## Review on Product Recommendation System Naïve Bayes & MapR vide Twitter

Priyanka<sup>1</sup>, Shalini<sup>2</sup>

P.G. Student, Department of Computer Science & Engineering, SRCEM, Palwal, Haryana, India<sup>1</sup>

Assistant Professor, Department of Computer Science & Engineering, SRCEM, Palwal, Haryana, India<sup>2</sup>

**ABSTRACT:** Because of the advent growth and fame of the informal community, people groups and social media like twitter the use of sharing their views, item surveys, appraisals and so on are very important nowadays. This prompts utilize the new factors, for example, relational impact and relational intrigue comparability to settle the virus begin and information sparsely issues in suggestion frameworks and recommendation framework. Variety of elements and aspects are evolved i.e. tweets, re-tweets, user context, machine learning, mapR (Map and Reduce) with be formed under this framework focusing on four social factors, for example, client individual intrigue, client relational intrigue likeness, client social setting and relational impact, combined every one of these components into brought together suggestion demonstrate dependent on probabilistic framework factorization termed to product recommendation system. Consequently, the proposed scheme will provide the effective and potentially best recommendation layout based on the context using Naive Bayes and Map and Reduce (MapR).

**KEYWORDS:** Product Recommendation, Twitter, Machine Learning, Naïve Bayes, Big Data, Hadoop Distributed File System, Map and Reduce.

### I. INTRODUCTION

Big Data portrays strategies and advancements to store, disseminate, oversee and break down extensive measured datasets with high-speed. Enormous information can be organized, unstructured or semi-organized, bringing about lack of ability of ordinary information the executives techniques. Information is created from different distinctive sources and can land in the framework at different rates. So as to process these a lot of information in a reasonable and productive way, parallelism is utilized.. Hadoop is the centre stage for organizing Big Data, and takes care of the issue of making it valuable for investigation purposes. Hadoop is an open source programming venture that empowers the appropriated handling of expansive informational collections with an exceptionally high level of adaptation to internal failure. It alludes to the productive treatment of expansive measure of information that is unimaginable by utilizing customary or regular strategies, for example, social databases or it is a system that is required to deal with the vast measure of information that is created with headways in innovation and increment in the populace. Enormous information stores, recover and adjust these huge informational indexes. For instance, with the coming of brilliant innovation, there is fast increment being used of cell phones because of which substantial measure of information is created each second, so it is difficult to deal with by utilizing customary strategies subsequently to beat this issue enormous information ideas were presented. most investigators and professionals right now allude to informational collections from 30-50 terabytes(10 12 or 1000 gigabytes for every terabyte) to different peta-bytes (1015 or 1000 terabytes for every peta-byte) as large information. Figure No. 1.1 gives a Layered Architecture of Big Data System. The 3Vs of the Big Data System is elaborated for ready reference,

1. **Data Volume:** Information/Data volume alludes to the measure of information. At present, the volume of information put away has developed from megabytes and gigabytes to peta-bytes and should increment to zeta-bytes in the adjacent future.
2. **Data Variety:** Data-Variety alludes to the distinctive sorts of information – content, pictures video, sound, and so on and wellsprings of information. The information being delivered isn't of single classification as it

# International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 7, Issue 3, March 2019

incorporates the conventional information as well as the semi-organized information from different assets like site pages, Web Log Files, internet-based life destinations, email, records.

3. **Data Velocity:** Velocity in Big Data/information is an idea which manages the speed of the information originating from different sources. This mark isn't being restricted to the speed of approaching information yet in addition speed at which the in sequence streams and accumulated and formed.

**HDFS Architecture:** Hadoop incorporates a fault tolerant capacity framework called the Hadoop Distributed File System or HDFS. HDFS can store colossal measures of data, scale up gradually and endure the disappointment of huge pieces of the capacity foundation without losing information. Hadoop makes bunches of machines and facilitates work among them. Bunches can be worked with modest PCs. On the off chance that one falls flat, Hadoop keeps on working the group without losing information or interfering with work, by moving work to the rest of the machines in the bunch. HDFS oversees capacity on the bunch by breaking approaching records into pieces, called "squares," and putting away every one of the squares needlessly over the pool of servers. In the regular case, HDFS stores three complete duplicates of each record by replicating each piece to three unique servers. However below Figure No. 1 depicts the architecture of Hadoop Distributed File System.

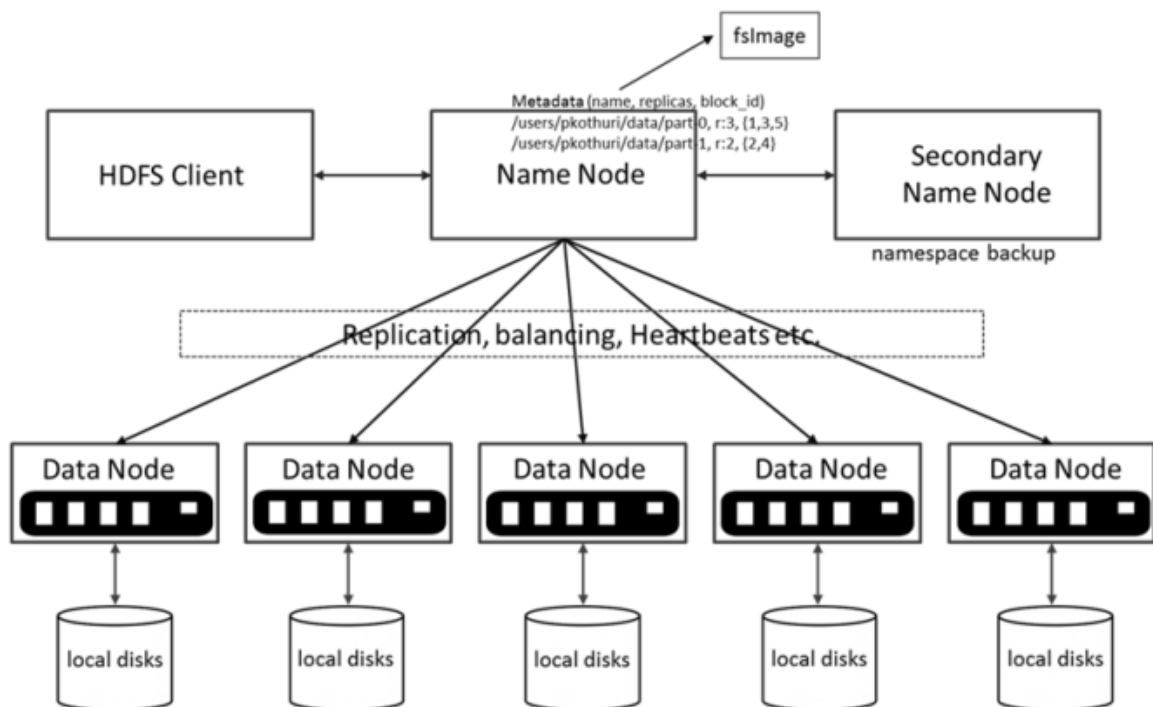


Figure No1. Architecture of Hadoop Distributed File System

**MapR:** Map and Reduce or MapR is major pillar in the Hadoop ecosystem or structure. This system enables the detail of an activity to be connected to a tremendous informational collection, partition the issue and information, and run it in parallel. From an examiner's perspective, this can happen in different measurements. For instance, a huge dataset can be diminished into a littler subset where investigation can be connected. In a customary information warehousing situation, this may involve applying an Extract Transform and Load activity on the information to create something usable by the examiner. In Hadoop, these sorts of tasks are composed as MapReduce employments in Java. There are various more elevated amount dialects like Hive and Pig that make composing these projects simpler. The yields of these occupations can be composed back to either HDFS or put in a customary information distribution centre. There are two capacities in MapReduce as pursue with following Figure No.2

# International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 7, Issue 3, March 2019

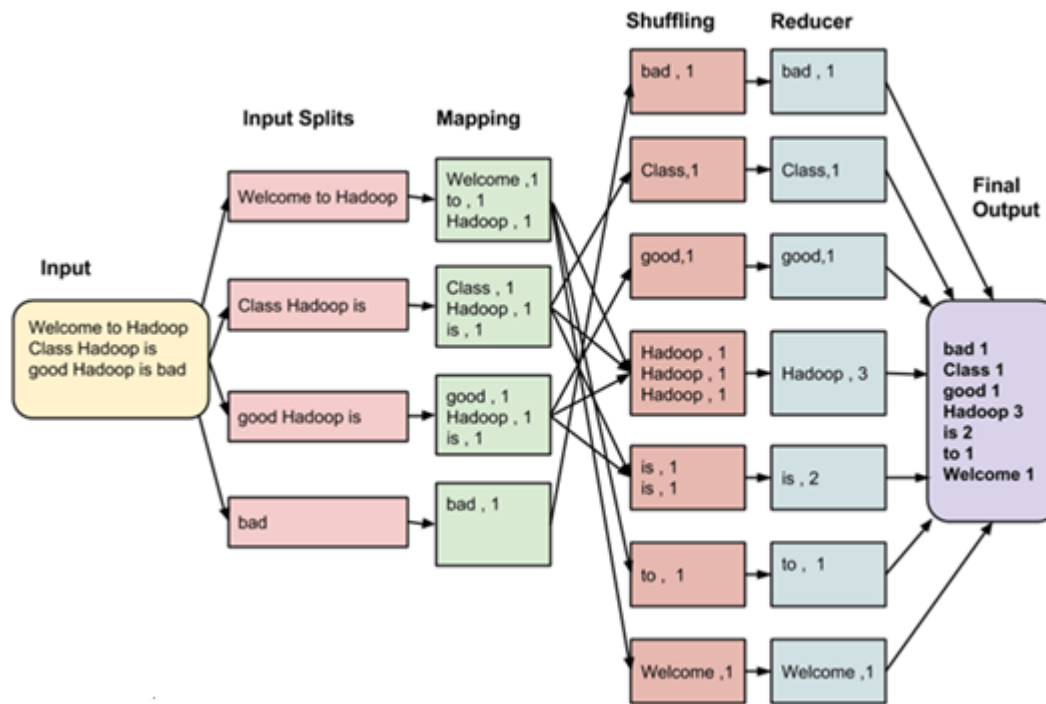


Figure No. 2 Architecture of MapR (Map and Reduce)

**Naïve Bayes:** Naïve Bayes classifiers have a spot with a gathering of fundamental probabilities based classifiers which apply Bayes' theory with strong presumptions between the segments. They are exceedingly versatile, requiring the amount of parameters lineary with respect to the amount of components (highlights/indicators) in a learning issue. Greatest probability preparing was performed on this classifier by evaluating a shut-shape articulation which takes direct time, instead of using exorbitant iterative hypothesis as used by various diverse sorts of classifiers. We then compute the division of the tweets involving and supplementary classified as follows:

- polarization = 0: Neutral Review/Opinion/Recommendation
- polarization < 0: Negative Review/Opinion/Recommendation
- polarization > 0: Positive Review/Opinion/Recommendation

## II. RELATED WORK

**Kaveri Roy, Aditi Choudhary and J. Jayapradha [1]** depicts that, the data/information mining is a cross-disciplinary field that focuses on finding properties of informational collections. There are distinctive ways to deal with finding properties of informational collections and Machine Learning is one of them. AI is a sub-field of information science that centers around planning calculations that can gain from and make expectations on the information. With the expansion in the interest for the online business sites, heaps of data emerges because of which the clients face trouble in finding the applicable data coordinating their inclinations. Hence, we speak to a framework which will prescribe comparative nourishment items to the client dependent on his buy. The Food Product will be prescribed dependent on the everyday wellbeing illnesses of the client. The client profile is shaped in which wellbeing intricacy of the client is there. The dataset for Recommendation System contains 2075 sustenance things. We will apply K-Means algorithm/calculation to understand the Recommendation System. We will likewise actualize Machine Learning calculations, for example, Support Vector Machine (SVM) and Random Forest. What's more, the examination among SVM and Random Forest is performed and SVM beats Random Forest calculation as it demonstrates an expansion in the execution.



# International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 7, Issue 3, March 2019

**Tanvir Habib Sardar, Zahid Ansari [2]** depicts that, there have been developing premiums in the region of recommender frameworks utilizing AI procedures. As there are an incredible number of express and understood highlights that can be utilized for assessing client inclination, it requires adaptable and precise calculations alongside a framework with high accessibility and adaptability. Substituting least square lattice (ALS) calculation is an upgraded form of dormant factor models utilizing framework factorization with great versatility and prescient precision. Apache Spark is an open-source conveyed stage for preparing huge information, accomplishing great speed and adaptability reasonable for iterative AI calculations. Amazon offers distributed computing administrations with different usefulness including information stockpiling and handling motors and is very accessible and adaptable. In this examination, we connected the ALS calculation utilizing Apache Spark running on an Amazon Web Service (AWS) Elastic Map Reduce (EMR) bunch for prescribing an item with a decent exactness and improved versatility.

**Sahil Pathan, Nitin Yadav, Karan Panjwani, Shreyas Lokhande, Bhushan Thakare [3]** depicts that, the suggestion is the most vital element in any sort of online client driven application. Such Systems are utilized to expand the development of online organizations. On account of a business undertaking client's information is accessible in a huge volume of information. With the goal that they can perform different information mining calculation to remove the vital information. Consequently this concentrates information can be valuable to discover clients explicit propensities, buy designs, clients most loved class and utilizing this sort of data proposal frameworks computes recommendation for the client. On the off chance that these suggestions are sufficient, at that point it in the end expands client's advantage. This framework when actualized on a nearly new Hyper-Local Based Services showcase, it would help develop this market.

**Prof. Alka Leekha Shreyas Upadhye, Pratik Ahire, Pranav Pawar [4]** depicts that, The suggestion is the most vital element in any sort of online client driven application. Such Systems are utilized to expand the development of online organizations. On account of a business undertaking client's information is accessible in a huge volume of information. With the goal that they can perform different information mining calculation to remove the vital information. Consequently this concentrates information can be valuable to discover clients explicit propensities, buy designs, clients most loved class and utilizing this sort of data proposal frameworks computes recommendation for the client. On the off chance that these suggestions are sufficient, at that point it in the end expands client's advantage. This framework when actualized on a nearly new Hyper-Local Based Services showcase, it would help develop this market.

### III. PROPOSED FRAMEWORK

The proposed framework for the product recommendation system using Map and Reduce and Naïve Bayes using Twitter will form the following steps to attain the potentially efficient and accurate recommendation based on contextual reference made by the user or consumer :-

1. Contextual Query Reference.
2. Establishing the connectors with Twitter Cloud and transformation services like flume to Fetch Tweets and Re-Tweets based on Contextual Query.
3. Storing Corpus on Hadoop Distributed File System (HDFS),
4. Removing noise or Pre-Processing of Data.
5. Forming Schema using Apache Hive and Migrating Data to Hive Repository using MapR.
6. TF-IDF formation for specific keywords
7. Polarization Generation in respect to Contextual Query for Product Recommendation.
8. Using Naïve Bayes to Form Relation with Contextual Query Generated DataSets by Twitter and providing best recommended product therein.

The below diagram depicts the workflow model of the proposed scheme the above mentioned steps are been elaborated using pictographic scenarios for better understanding and ready reference.

# International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: [www.ijirccce.com](http://www.ijirccce.com)

Vol. 7, Issue 3, March 2019

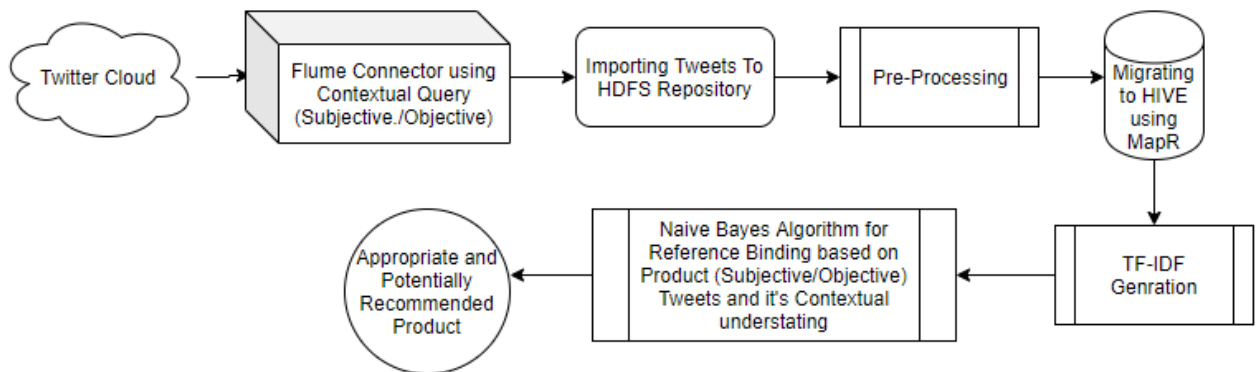


Figure No.3 Work Flow of Proposed Scheme for Product Recommendation System.

## IV. CONCLUSION

The above-proposed framework will form and produce the efficient and potentially effective product recommendation system using a machine learning technique namely Naïve Bayes where the flume-like connectors will establish the secured connection with Twitter Cloud Repository to gather the data information based on a contextual query raised by user or client. Thereinafter the data or tweets will be migrated to HDFS repository, herein using the Hadoop Ecosystem the noise or stop words will be omitted and feature generation will be done subsequently, using HIVE the contextual data will be formed under the schema and MapR will provide the distinct word sets with counts herein using Term Frequency Invert Document Frequency (TF-IDF) the frequency of phrase will be evaluated and at last using Naïve Bayes the contextual binding or relational binding will be amalgamated with polarity to get the effective and accurate information for recommendation system. Therefore, we will create a progressively viable and precise-proposed framework for future uses in product recommendation scenarios.

## REFERENCES

1. Kaveri Roy, Aditi Choudhary and J. Jayapradha , Product Recommendations Using Data Mining And Machine Learning Algorithms, ARPN Journal of Engineering and Applied Sciences ©2006-2017 Asian Research Publishing Network (ARPN). All rights reserved.
2. Tanvir Habib Sardar, Zahid Ansari. An Analysis Of Mapreduce Efficiency In Document Clustering Using Parallelk-Means Algorithm, Computer Science And Engineering, P.A. College Of Engineering, Mangalore, India Received 28 October 2017; Revised 18 February 2018; Accepted 20 March 2018 Available Online 17 May 2018.
3. Sahil Pathan, Nitin Yadav, Karan Panjwani, Shreyas Lokhande, Bhushan Thakare, Recommendation System Application using Naïve Bayes' Algorithm, June 2016, IJIRT, Volume 3 Issue 1.
4. Prof. Alka Leekha Shreyas Upadhye, Pratik Ahire, Pranav Pawar, Recommendation System for Ecommerce using Big Data. International Journal of Scientific Research Engineering & Technology (IJSRET), ISSN 2278 – 0882, Volume 6, Issue 4, April 2017
5. Nada Elgendy and Ahmed Elragal, Big Data Analytics: A Literature Review Paper, P. Perner (Ed.): ICDM 2014, LNAI 8557, pp. 214–227, 2014. Springer International Publishing Switzerland 2014
6. Meghna Khatri, A Survey of Naïve Bayesian Algorithms for Similarity in Recommendation Systems, Volume 2, Issue 5 International Journal of Advanced Research in Computer Science and Software Engineering, May 2012
7. Stanford large network dataset collection. <http://snap.stanford.edu/data/index.html>.
8. Lars Backstrom and Jure Leskovec. Supervised random walks: Predicting and recommending links in social networks. Proceeding of WSDM 2011, pages 635–644, 2011.
9. Jure Leskovec, Lada A. Adamic, and Bernardo A. Huberman. The dynamics of viral marketing. ACM Transactions on the Web (ACMTWEB), 1(1), 2007. Greg Linden, Brent Smith, and Jeremy York. Amazon.com recommendations: Item-to-item collaborative filtering. IEEE Internet Computing, 7(1):76–80, 2003.
10. Tao Zhou, Jie Ren, Matus Medo, and Yi-Cheng Zhang. Bipartite network projection and personal recommendation. Physical Review E, page 76, 2007
11. Chen, J., Nairn, R., Nelson, L., Bernstein, M., Chi, E.: Short and Tweet: Experiments on Recommending Content from Information Streams. In: The 28th International Conference on Human Factors in Computing Systems (2010)
12. Choudhary, A., Hendrix, W., Lee, K., Palsetia, D., Liao, W.K.: Social media evolution of the Egyptian revolution. Communications of ACM 55(5), 74–80 (May 2012)



ISSN(Online): 2320-9801  
ISSN (Print) : 2320-9798

# International Journal of Innovative Research in Computer and Communication Engineering

*(A High Impact Factor, Monthly, Peer Reviewed Journal)*

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 7, Issue 3, March 2019

13. De Francisci Morales, G., Gionis, A., Lucchese, C.: From Chatter to Headlines: Harnessing the Real-Time Web for Personalized News Recommendation. In: The 5th ACM International Conference on Web Search and Data Mining (2012)
14. Garcia, R., Amatriain, X.: Weighted Content Based Methods for Recommending Connections in Online Social Networks. In: The 2nd ACM Workshop on Recommendation Systems and the Social Web. Barcelona, Spain (June 2010).
15. [http://www.smartdatacollective.com/seanmallonbizd aquk/410001/5-benefits-big-data](http://www.smartdatacollective.com/seanmallonbizd%20aquk/410001/5-benefits-big-data)
16. <https://www.qubole.com/resources/solution/ecommerce-big-data/>