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

Vol. 7, Issue 5, May 2019

Product Recommendation System Naïve Bayes & MapR vide Twitter

Priyanka[1], Shalini[2]

P.G. Student, Department of Computer Science & Engineering, SRCEM, Palwal, Haryana, India[1]

Assistant Professor, Department of Computer Science & Engineering, SRCEM, Palwal, Haryana, India [2]

ABSTRACT: As the data or information at Twitter repositories are in high volume therefore to produce the recommendation system on the fly needs the best mechanism to store the tweets, re-tweets and related context in comprehensive manner therefore, big data will be the best exemplary model for such mammoth data based on velocity, volume and variety however, below the same is depicted for consideration to inculcate in future schemes. Huge information alludes to information volumes in the scope of exabytes (1018) and past. Such volumes surpass the limit of current on-line stockpiling frameworks and preparing frameworks. Information, data, and learning are being made and gathered at a rate that is quickly moving toward the exabyte/year extend. Be that as it may, its creation and collection are quickening and will approach the zettabyte/year go inside a couple of years. Volume is just a single part of enormous information; different qualities are assortment, speed, esteem, and multifaceted nature. Capacity and information transport are innovation issues, which appear to be resolvable in the close term, however speak to long haul difficulties that require look into and new ideal models. We dissect the issues and difficulties as we start a collective research program into approaches for huge information examination and structure. 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

In the Internet age, the most serious issue for an individual who needs to purchase something on the web isn't just how to get enough data to make a choice, yet additionally how to take a correct choice with that colossal data. These days, individuals dependably search the Internet to discover the correct items and administrations that they need. Deliberately or unknowingly, they rely upon the recommender framework to beat a data over-burden. Recommender framework has been demonstrated as a significant answer for the data over-burden issues, by giving progressively proactive and customized data administrations to the clients.

Recommender framework gives a recommendation about the items, data or administrations that the client needed to know. It is an astute application to help the client in a basic leadership process where they need to pick one thing among the possibly overpowering arrangement of elective items or administrations. It is likewise a standout amongst the most unmistakable applications substantially affecting the presentation of internet business locales and the segments by and large. Recommender framework has been misused for suggesting books, CDs, films, news, hardware, ventures, money related administrations, and numerous different items and administrations. Recommender framework turns out to be progressively famous even in basic online business sites too.

Recommender framework is utilized in an enormous number of web-based business sites to customize the data for their clients. Despite the fact that the recommender framework proposes things that are determined on the individual's taste, they can likewise be utilized in a more general approach to making every site more client driven. At the point when individuals need to settle on a decision with no close to home learning on the choices, the normal game-plan is to

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

Vol. 7, Issue 5, May 2019

depend on the experience and sentiments of others. Recommender framework accumulates and spares suggestions from individuals who know about the decisions that they confronted and furthermore it esteems their points of view and remembers them as specialists.

Two fundamental elements which show up in any recommender framework are the client (suggestion supplier or proposal searcher) and the thing. A client is an individual who uses the recommender framework giving his assessment about different things and gets proposal about the new things through the framework. In a regular recommender framework, first, the client (proposal supplier) gives some type of contribution to the framework. These information sources can be both unequivocal and certain. At that point, these sources of info are united to shape a portrayal of client's preferences that might be as a framework with thing appraisals or might be as an information structure joining both substance and rating data. The framework will finally process the suggestions utilizing these "client profiles" and "proposal searcher profiles.". Contributions for Recommender System a summed up articulation "All College understudies like Cricket and watching motion pictures" is regularly false. Preferably a recommender framework is to be completely computerized, so it is ready to arrange individuals without the need of direction. The metadata of the things could be drawn upon as a last wellspring of data. The metadata indicates properties of a thing, including class, creator/craftsman, discharge date and so forth. In the extraordinary case, the full substance of a thing (for example the content of a book) can be considered for gathering the things. Such information might be utilized to make speculations dependent on classifications from the thing profile, for instance accepting that on the off chance that somebody loves one thing, at that point the framework would prescribe that he might want all things in that one kind which he enjoyed.

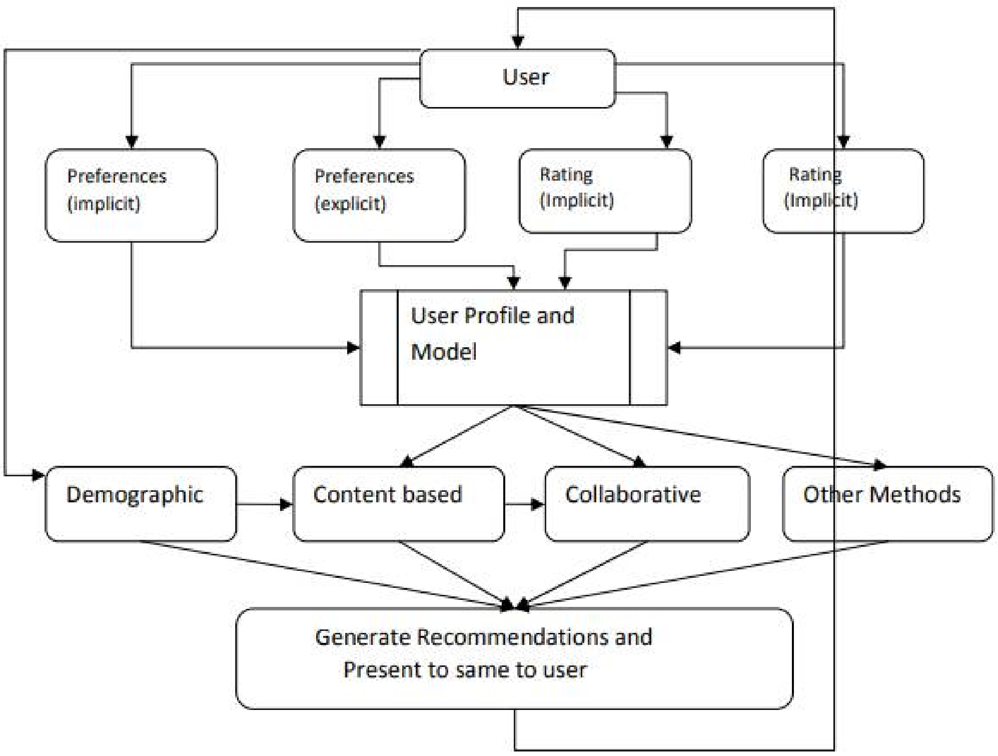

Figure No.1 General Frame work of Recommender System

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

Vol. 7, Issue 5, May 2019

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 emerge 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 depends 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 the 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.

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 the client's advantage. This framework when actualized on a nearly new Hyper-Local Based Services showcase, it would help develop this market.

III. PROPOSED ALGORITHM

MapReduce: For the presence of massive tweets, how to carry on the rapid establishment of inverted index is an important research point, because if the inverted index update speed, is based on the keyword search results returned cannot contain additional or modified page, may also return to the old or already nonexistent page. Let us consider the use of MapReduce parallel to generate all the tweets of the inverted index for

First of all, we need to pre-processing i.e. extracted every word in the document, and the calculation of the word in the document appeared a number of times. In Map phase, we each encounter a word, just output < word document, < ID >, 1. Because the Hadoop platform to ensure all the same key corresponding to the value list will be the same received by the reducer node, so the REDUCE stage, can be collected in the same word all relevant information, and they add up to get the sum of.

Psuedo code for Mapper and Reducer :

Mapper :

map(DocumentID, DocumentText)

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

Vol. 7, Issue 5, May 2019

```
foreach word W in DocumentText
emit(W, DocumentID);
done
Reducer :
reduce(word, values):
foreach DocumentID in values:
AddToOutputList(DocumentID);
done
emit_Final(FormattedDocumentIDListForWord);
```

MAPPER algorithm description and analysis are as follows:

```
1: class MAPPER
2: method MAP(ID, Document)
3: H _ new ASSOCIATIVEARRAY
4: for all term t _ doc d do
5: H{t} _ H{t} + 1
6: for all term t _ H do
7: EMIT(tuple <t, n>, tf H{t})
```

In the MAP phase, each computing node input a document and document content d id n First of all, the map function will create an associative container H, Then the document D appear in each of a word will be in H T, total T occurrence. The H of every element output. The output of the key is a two tuple < T ID n > value, document, document frequency list.

REDUCER algorithm description and analysis are as follows:

```
1: class REDUCER
2: method INITIALIZE
3: tprev _ null
4: P new POSTINGLIST
5: method REDUCE(tuple <t,n>, tf [f])
6: if t != tprev AND tprev != null
7: then EMIT(term t, postings P)
8: P.RESET()
9: P.ADD(<n,f>)
10: tprev _ t
11: method CLOSE
12: EMIT(term t, postings P)
```

In the REDUCE phase, each computing node receives a key is a two tuple < T ID n > words, document, value is the corresponding document frequency list. The INITIALIZE function is first created a list PostingList, subsequently received on each input processing. If the input is a new word, it will have the output list. If you still keep a distance is the same word, the document ID and document frequency is added to the list PostingList. Finally all the words and the corresponding PostingList full output.

Naiev Bayes: Naïve Bayes Classification algorithm is used for the purpose of classification of given trained model. It is the probabilistic approach to the text classification. Here the class labels are known and the goal is to create probabilistic models, which can be used to classify new texts. It is specifically formulated for text and makes use of text specific characteristics. The Naïve Bayes classifier treats each tweet as a "bag of words" free of stop words. Illustration is baaed opinion as a matter of fact outcome probabilities in the drifting aspects, and the MARGINAL probabilities of Naïve Bayes with Classification (from integrating out the other variable from the joint) on the side and bottom the below steps depicts the solution under the below mentioned scheme for opinion results.

1. Say, opinion type = w and drifting outcome = p.

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

Vol. 7, Issue 5, May 2019

2. P(w,p) is the joint probabilities and P(p) and P(w) are the marginal's.
P(w,p) = P(w|p) * P(p) = P(p|w) * P(w).
3. From the center cells we have P(w,p) and from the side/bottom we get P(p) and P(w).
4. Depending on what you need to calculate, it follows that:
(1): P(w|p) = P(w,p) / P(p)
(2:)P(p|w) = P(w,p) / P(w), which is what you did with P(opinion, yes) = 3/14 and P(w) = 5/14, yielding (3/14) (14/5), with the 14's cancelling out.

IV. SIMULATION AND RESULTS

Product Recommendation System predicts the assessment of the client whether the client is in floating mode, positive or negative based on the tweet_id of the client on live social twitter information. Likewise to anticipate the general assessment of clients in various areas in a specific time stamp in a specific setting and delineated in diagram structure.

User Name	User Tweets	Behavior Analyses	
skj8728	RT @republic: #KarnatakaFloorTest | Journalists in sit-in protest outside Karnataka Vidhana Soudha on account of not being permitted to ent?	Pos :1 Neg :5 Neu :0	User Negative
mohitsmartlove	RT @republic: #KarnatakaFloorTest | Journalists in sit-in protest outside Karnataka Vidhana Soudha on account of not being permitted to ent?	Pos :1 Neg :5 Neu :0	User Negative
KrishnaBPrasad	Retweeted Ankit Lal (@AnkitLal): Over 2,400 VVPAT machines malfunctioned in Karnataka. 2400x1200 = 28,80,000 ap? https://t.co/DPG3yjVi1w	Pos :1 Neg :2 Neu :0	User Negative
Iamthe_dude	RT @ArvindKejriwal: VVPAT machines is not rocket science. Our country has capability to launch satellites. Can?t we manufacture functioning?	Pos :3 Neg :2 Neu :1	User Positive
online_deepakk	RT @ArvindKejriwal: VVPAT machines is not rocket science. Our country has capability to launch satellites. Can?t we manufacture functioning?	Pos :3 Neg :2 Neu :1	User Positive
saisharan1990	RT @UnSubtleDesi: The governor?s decision was legitimate in Goa and Karnataka. Here are the precedents. Congress has reduced itself to a bu?	Pos :2 Neg :0 Neu :0	User Positive
AyamAtmaBrahm	RT @PRSLegislative: Explained: How 220 MLAs will vote for or against BS Yeddyurappa today https://t.co/tbDgdMLikl	Pos :5 Neg :3 Neu :0	User Positive
RangerPaatil	RT @ArvindKejriwal: VVPAT machines is not rocket science. Our country has capability to launch satellites. Can?t we manufacture functioning?	Pos :3 Neg :2 Neu :1	User Positive

Figure 2: Recommendation Analysis based on MapR and Naïve Bayes

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

Vol. 7, Issue 5, May 2019

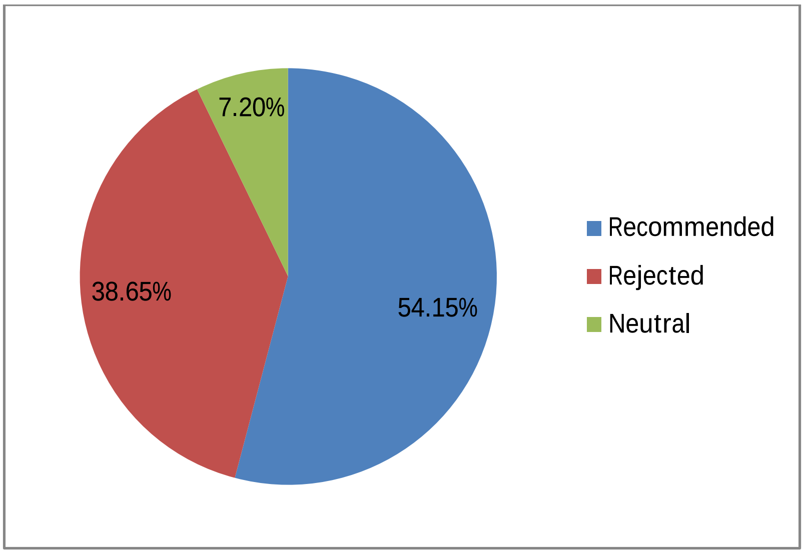

Figure 3: Recommendation Analysis based on MapR and Naïve Bayes whereas based on Opinion Counts the product iPhoneXR is recommended to purchase

V. Conclusion and Future scope

The above-proposed framework alludes 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 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 using the polarity will be extracted. The recommendation depicts the term like product recommended, product not recommended and the product holds the neutral advice.

For future scope the recommendation system can be inculcated as the part of the twitter eco-system which will save lot to potential time and resource for the comparison based on rational analysis model which may be delivered to the user based on sentiments, opinions and recommendation to the masses on the instant note.

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 2018Available 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

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

Vol. 7, Issue 5, May 2019

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-toitem 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)
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
16. https://www.qubole.com/resources/solution/ecomme rce-big-data/