Trust Relationship Prediction in Alibaba E- Commerce Platform

N.SRINIVASA RAO [1], KOWRU BHASWANTH [2]

[1] Assistant Professor, DEPT OF MCA, *SKBR PG COLLEGE.* AMALAPURAM, Andhra Pradesh, email:- naagaasrinu@gmail.com

[2] PG Student of MCA, *SKBR PG COLLEGE.* AMALAPURAM, Andhra Pradesh, email:- kowrubobby@gmail.com

---

Abstract - *In our Current trend, the businesses of E-Commerce are booming due to the technological advancements of Mobile-Phones, Laptops etc. But the existing databases can't handle the huge amount of datasets which is supplemented by large number of suppliers and it indicates how to infer trust relationships from Billion-Scale Networked Data to benefit our E-Commerce Business. To prevent these huge dataset related problems, we introduce Big Data technology to capture and analyze several datasets and also undergoing a future proposal of dealing with real-time datasets using Spark technology. Here, the Hadoop tool is used for the analysis of huge amounts of data. Hence, our Analysis provides a comprehensive guide to accurately analyze and handle a huge amount of datasets to overcome the problems of processing time, data consistency and maintenance cost. Now we are going to use the Big Data technology for our sales of mobile phones effectively.*

Key Words*: *Big Data, Hadoop, Map Reduce, Prediction, Visualization

1.INTRODUCTION

E -Commerce has become our vital processes for our modern Commerce or Online Trade involves the purchase and sale of goods, products or services on the Internet. These services provided online over the internet network. These business transactions can be done in four ways: Business to Business (B2B), Business to Customer (B2C), Customer to Customer (C2C), and Customer to Business (C2B). The basic definition of e-commerce is the commercial transaction that taken place over the internet. By 2020, global retail e-commerce can reach up to $27 Trillion. The data which can be beyond the storage capacity and the processing power such a data is called Big Data. Big data means huge volume of data; it is a collection of large datasets that cannot be processed using traditional computing techniques. Big data is not merely a data; rather, it has become a complete topic, involving numerous tools, techniques and frameworks. Normally we work on data of size MB (Wordbook, Excel) or maximum GB (Movies, Codes) but data in Petabytes i.e. 10^15 byte size is called Big Data. It is stated that almost 90% of today's data has been generated in the past 6 years. In this paper we are isolating E-Commerce data by utilizing Hadoop instrument adjacent some Hadoop common systems like hdfs, map reduce, sqoop, hive and pig. By utilizing these devices preparing of information with no confinement is conceivable, no information lost issue, we can get high throughput, upkeep cost comparatively incredibly less and it is an open source programming, it is extraordinary on the majority of the stages since it is Java based. In E-Commerce information is related colossal volume of farthest point of research paper scattering site.

2. LITERATURE REVIEW

With recent advancements in technologies different methodologies have been introduced for analyzing big data. Analyzing the data to provide foresight for business is continuing to play a vital role and various scholars have contributed their ideas to benefit E-commerce business.

Inferring Networks of Substitutable and Complementary Products [1]. In a modern recommender system, it is important to understand how products relate to each other. For example, if a consumer is searching for cell phones, it might make sense to recommend other phones, but if they purchase a phone, we may want to suggest batteries, cases, or chargers instead. These two types of guidelines are referred to as alternatives and complements: alternatives are products which can be purchased in place of each other, while complements are purchased in addition to each other.

Learning Influence Probabilities in Social Networks [2]. The process of power diffusion in social networks has been of immense interest recently. The studies in this area assume a social graph with edges marked with probabilities of control between users has an input to their problems. Never the less until now the question of where the estimates derive or whether they can be derived from real social network has been largely overlooked. So it is important to ask if one can construct models of power from a social graph and a log of behaviour by its users.

This is the main problem attacked in this paper besides proposing models and algorithms to learn the parameters of the model and to check the learned models to make the predictions, we also create techniques to predict the time by which the consumer will be required to perform an action. We validate our ideas and techniques using the Flickr data set consisting of a social graph with 1.3M nodes, 40M edges, and an action log consisting of 35M tuples referring to 300K distinct actions. Beyond showing that a real social network has a genuine impact, we prove that our techniques have excellent predictive output.

Predicting Popularity of Twitter Accounts through the Discovery of Link-Propagating Early Adopters [3]. Due to its dynamicity, new well known records consistently show up and vanish in miniaturized scale blogging administrations. Effective detection of new recordsthat will end up common in the future is a key issue with a variety of applications, such as slant place, viral display and customer suggestions. Estimation of prominence of a record is additionally valuable for approximating the nature of data it posts. Estimation of the nature of data is vital in numerous applications, yet it is for the most part hard to gauge it without human mediation. To tackle this issue, fame based strategies have been broadly utilized. Strategies that gauge data nature of website pages in view of the number of their approaching connections has been effective. Additionally, comparative research was linked effectively to small-scale online journals with linking capacities. These certainties demonstrated that there is high relationship between the notoriety and the nature of data. In this manner, the estimation of forthcoming notoriety of new records, which have not yet settled the prevalence they merit, is additionally helpful for estimation of the quality.

3. PROPOSED METHOD

Proposed concept deals with providing database by using Hadoop tool, we can analyze with no limitation of data and simply add number of machines to the cluster and we get results with less time, high throughput and maintenance cost is very less and we are using partitions and bucketing techniques in Hadoop.

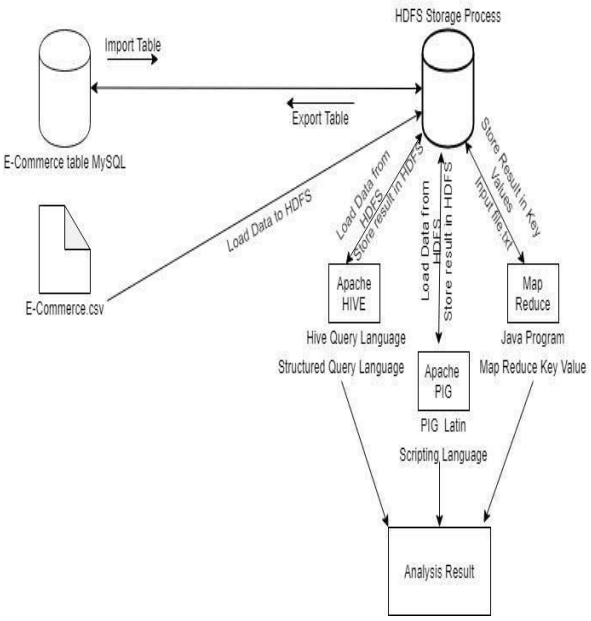

Fig 1. Flow chart of proposed method

A) Existing Application: MySQL

In MySQL is a social database the authority's framework. RDBMS utilizes relations or tables to store E-Commerce information as a cross section of lines by segments with key. With MySQL language, E-Commerce data in tables can be gathered, verified, prepared, recovered, expelled and controlled for the most part for business purpose. Existing thought directs giving backend by utilizing MySQL which contains heap of shortcomings i.e. information hindrance is that dealing with time is high when the information is tremendous and once information is lost we can't recoup so thusly we proposing thought by utilizing Hadoop device.

B) Connector (Sqoop)

Sqoop is a command-line interface application for transferring E-Commerce data between relational databases (MySQL) and Hadoop. Here in MySQL database having E-Commerce data have to import it to HDFS using Sqoop. E-Commerce data can be moved into HDFS/Hive from MySQL and then it will generate the java classes. In previous cases, flow of data was from RDBMS to HDFS. We can import data from HDFS into RDBMS using "Export" function. Sqoop fetches table metadata from MySQL database before exporting. Therefore we need to create a table with the metadata needed first.

C) Analysis Query Language (Hive)

Hive is a Hadoop data warehouse framework running SQL like queries called HQL(Hive query language)

which is converted internally to map reduce jobs. In Hive, E-Commerce data tables and databases are created first and then data is loaded into these tables. Hive as E-Commerce data warehouse designed for managing and querying only structured data that is stored in the tables. Hive organizes E-Commerce data tables into partitions. It is a way to divide a table into related sections, based on partitioned column values. Using partition, it is easy to query a portion of the given dataset. Tables or partitions are sub-divided into buckets, to provide extra structure to the E-commerce data that may be used for more efficient querying. Bucketing works based on some column of a table's hash function value. Large datasets can be processed parallel and applying distributed algorithms on clusters, this programming model is the Map Reduce. Structured and unstructured data is taken by the Map Reduce model in the form of key value pairs. The data is processed in a reliable way.

D) R Visualization

The analyzed data is then visualized using R programming in the form of bar charts and pie charts.

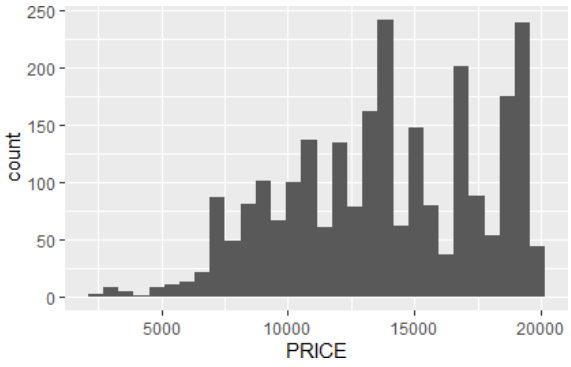

Fig 2. Bar chart depicting count sold in the price ranges

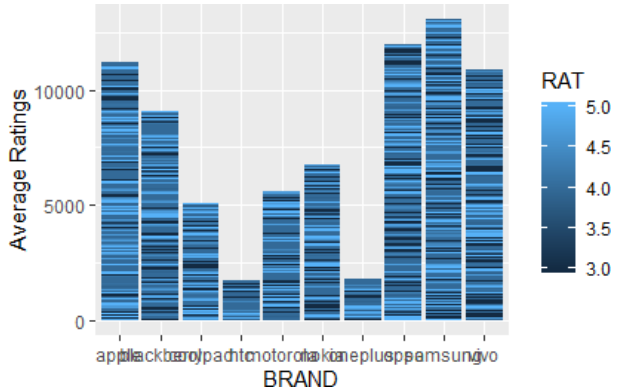

Fig 3. Bar Chart depicting ratings for brands

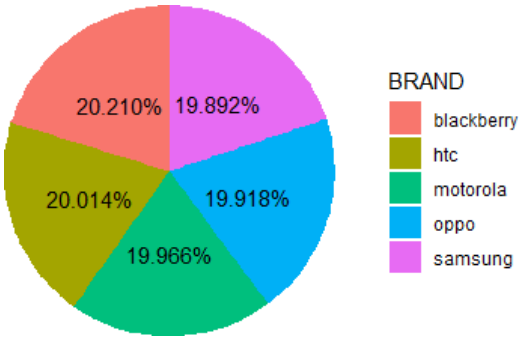

Fig 4. Pie Chart depicting top 5 brands

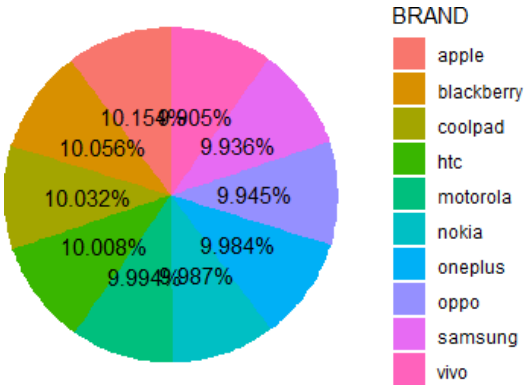

Fig 5. Pie Chart depicting ratings of top 10 brands

4. CONCLUSION

In this paper, we showed an examination on E-business data and gauge concerning explore paper about mobile thing. To examination the E-Commerce information in Hadoop natural system to improve the business subject to number of things sold. Hadoop condition is having hive, pig, Map Reduce instruments for preparing whether yield will set aside less effort to process and result will be very fast. Hence in this undertaking beginning at now E-Commerce information which is typically going to store in RDBMS going to less execution starting now and into the foreseeable future by utilizing Hadoop device quicker and effectively managing the information.

ACKNOWLEDGEMENT

I would like to express my sincerest regards to my project guide **N.SRINIVASA RAO** for his valuable inputs and able guidance throughout the project and the study. My sincere thanks to the technical and support

staff in SKBR PG COLLEGE for thehelp they provided.

REFERENCES

[1] Julian McAuley, Rahul Pandey, Jure Leskovec, "Inferring Networks of Substitutable and Complementary Products" in 21st ACM, 2015

[2] Amit Goyal, Francesco Bonchi, Laks V. S. Lakshmanan,"Learning Influence Probabilities In Social Networks.",WSDM 2010.

[3] D. Imamori and K. Tajima, "Predicting popularity of twitter accounts through the discovery of link-propagating early adopters" in CoRR, 2015, p. 1512.

[4] R. Peres, E. Muller, and V. Mahajan, "Innovation diffusion and new product growth models: A critical review and research directions," International Journal of Research in Marketing, vol. 27, no. 2, pp. 91 – 106, 2010.

[5] L. A. Fourt and J. W. Woodlock, "Early prediction of market success for new grocery products." Journal of Marketing, vol. 25, no. 2, pp. 31 – 38, 1960.

[6] B. W. O, "Reference group influence on product and brand purchase decisions," Journal of Consumer Research, vol. 9, pp. 183–194, 1982.

[7] J. J. McAuley, C. Targett, Q. Shi, and A. van den Hengel, "Imagebased recommendations on styles and substitutes," in SIGIR, 2015, pp. 43–52.

[8] E. M.Rogers, Diffusion of Innovations. New York: The Rise of High- Technology Culture, 1983.

[9] K. Sarkar and H. Sundaram, "How do we find early adopters who will guide a resource constrained network towards a desired distribution of behaviors?" in CoRR, 2013, p. 1303.

[10] N. V. Nielsen, "E-commerce: Evolution or revolution in the fastmoving consumer goods world," nngroup. com, 2014.