International Journal of Advanced Technology in Engineering and Science
Vol. No.4, Issue No. 08, August 2016
www.ijates.com
ijates
ISSN 2348 - 7550

IMPORTANCE OF BIG DATA AND IT'S FUTURISTIC OPPORTUNITIES

Sarang Sambharia[1], Ridhima Shekhawat[2], Devendra Bhavsar[3]

[1,2]Student - CSE Department, JK Lakshmipat University (India)

[3]Faculty - CSE Department , JK Lakshmipat University (India)

ABSTRACT

This paper basically concentrates on "Importance of Big Data and It's Applications". At first presentation in regards to Big Data and it's significance in different fields has been talked about. Afterwards, we have accumulated case studies which proves the necessity of the concerned topic. Later, in this paper various parameters are marked down which prove to be great assets. The worry of growing and developing data and utilizing it to send to make and catch esteem for people, organizations, groups, and governments inspired this paper which records the parameters and to support the online desire of application behaviour so as to upgrade data access operations. In today's scenario, these parameters have become a necessity for everyone.

Keywords: Big Data, Data Prediction, Data Analysis, Machine Learning, Machine Faults.

I. INTRODUCTION

Big data is a buzzword, catch-phrase, used to describe a massive volume of both structured and unstructured data that is so large that it's difficult to process using traditional database and software techniques. Big data is generated from an increasing plurality of sources, including Internet clicks, mobile transactions, user-generated content, and social media as well as purposefully generated content through sensor networks or business transactions such as sales queries and purchase transactions. In addition, genomics, health care, engineering, operations management, the industrial Internet, and finance all add to big data pervasiveness. These data require the use of powerful computational techniques to unveil trends and patterns within and between these extremely large socioeconomic datasets. On one hand, a wealth of big data can bring us big opportunities. On the other hand, we still do not know how to harness such big amount of data with tremendous complexity, diversity and heterogeneity, yet with high potential values. This makes the data very difficult to process and analyze in a reasonable time.

Data Analysis is a process of examining, cleaning, changing, and displaying data with the goal of achieving useful information, suggesting conclusions, and supporting decision-making. Data analysis is a process of obtaining raw data and converting it into information that is useful for decision-making by various users across the globe. Data is collected and completely analyzed to answer questions, test hypotheses or disprove/prove theories.

Predictive analytics is an area of data mining which deals with extracting useful information from data and using it to predict trends and behavior patterns. Predictive analytics is regularly characterized as foreseeing at a more

itemized level of granularity, i.e., creating prescient scores (probabilities) for every individual hierarchical component. This differentiates it from forecasting. A lot of companies which includes SAS (Predictive Analytics Suite), IBM (IBM SPSS Statistics), and Microsoft (Microsoft Dynamics CRM Analytics Foundation) are providing solutions to enterprises using Data Prediction techniques.

II.DATA PREDICTION AND DATA ANALYSIS IN BIG DATA

Data prediction or data analysis is the practice in which the information is extracted from already existing data with an aim of determining patterns and predicting future outcomes and trends. Predictive analytics does not tell the happenings of the future. It predicts the probable happenings of the future with an acceptable level of reliability, including what-if scenarios and risk assessment.

It is believed that information is the source of power. So extracting information and analyzing it on a large scale is not easy to handle, this large amount of data is commonly known as “big data”. Some of the most wide areas which deal with big data analytics are social networking sites, Businesses, e-commerce,etc. With big data, analysts have more data to work with, as well as the processing power to handle large number of records with a lot many attributes.

Big data can be mainly characterized by three features: Volume, Variety, and Velocity, defined as three “V” dimensions by Meta Group in 2001. Volume stands for the large sized data, which is not easy to process with traditional database systems and single machines. Variety corresponds different types of data such as which is dealt with. Velocity is concerned with the continuous creation of data at faster rates. It is worth noting that “value” is an important quality of big data, but it is not a defining characteristic.

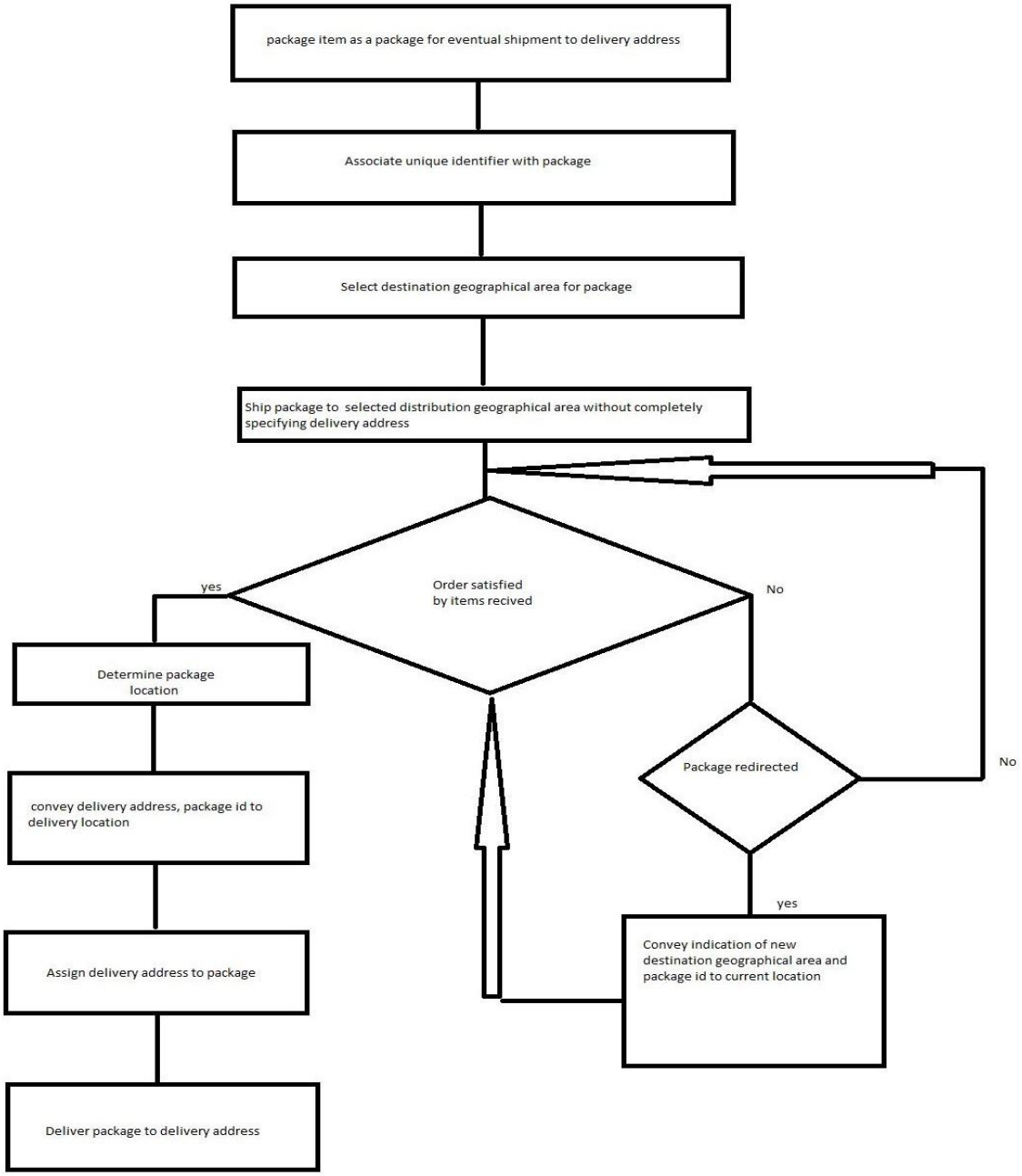
We will discuss some of the big business which deals with big data and shows us how important Big Data, data prediction and data analysis is, in today’s world and the future.

2.1. Considering “Amazon”

Data prediction and analysis makes it possible for amazon to sell you stuff before you buy it.[3] Amazon is considering another stage in it's journey for worldwide trade mastery. They're working an arrangement that would transport items to you before you even buy them since Amazon comprehends what you need superior to anything you do.

That is general thought behind another patent for "anticipatory shipping" found by The Verge that frameworks an odd pre-perception like framework that Amazon arrangements to actualize to trim conveyance times. Amazon will transport items to various regions of the nation in light of the shopping propensities for the general population who live there. The new framework will utilize "past hunts and buys, lists of things to get, and to what extent the client's cursor floats over a thing online," to figure out where the items will go, The Verge clarifies.Amazon petitioned for the patent, formally known as "method and system for anticipatory package shipping,” in 2012. The patent outline depicts a technique for shipping a package of one or more products "to the destination geographical area without completely specifying the delivery address at time of shipment," with the final destination defined end route.

[4]That's where the algorithm comes in. According to the patent, this forecasting model uses data from your previous Amazon activity, including the time spent on site, length of views, connections clicked and floated over, shopping basket movement and lists of things to get. Whenever possible, the calculation additionally sprinkles in certifiable data gathered from client phone request and reactions to showcasing materials, among different components. Together, this can give "choice backing for theoretical transportation of things," per the patent.

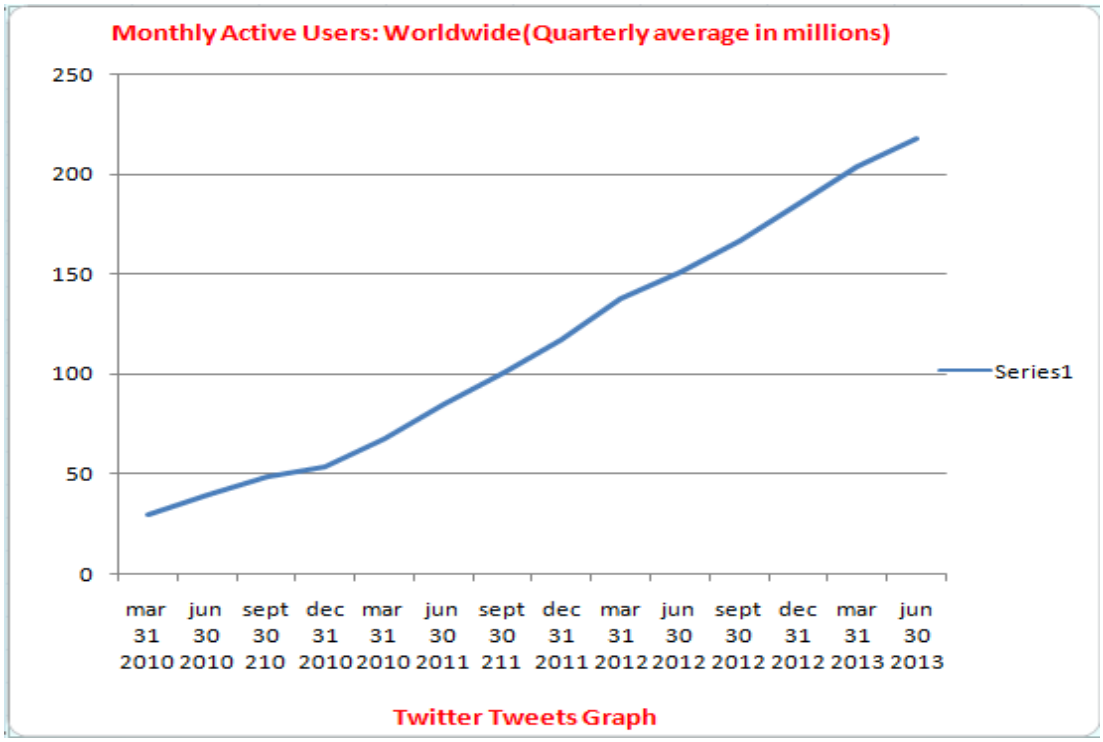


Amazon's framework includes two PCs, one that recognizes the general shipping location and a second that sits tight for the conveyance location to be settled. Theoretically, if inhabitants of Des Moines, Iowa, purchase a great deal of scarves in January, a neighborhood satisfaction focus may top off with a gathering of scarves yet

none would be transported until a real request is gotten at Amazon's home-based dispatching framework. At the point when a request is put, the thing would as of now be most of the way to its destination, slicing conveyance time to as meager as one day. It could be considerably snappier, contingent upon the client's vicinity to an Amazon appropriation focus.

2.2. Another Application Would Be Social Networking Websites Like Twitter

Considering "Twitter"[1][2] Every second, averagely, around 6,000 tweets are tweeted on Twitter, which corresponds to over 350,000 tweets per minute, **500 million tweets per day** and around 200 billion tweets per year. Given below is the chart which shows the number of tweets per day throughout Twitter's history



The first tweet was sent on March 21, 2006 by Jack Dorsey, the maker of Twitter. It assumed control three years, until the end of May 2009, to achieve the billionth tweet. Today, it takes under two days for one billion tweets to be sent.

In Twitter's short history, we went from 5,000 tweets for each day in 2007 to 500,000,000 tweets for each day in 2013, which speaks to a six requests of extent increment. The middle of the road steps were 300,000 tweets for each day in 2008, 2.5 million tweets for every day in 2009, 35 million tweets for every day in 2010, 200 million tweets for every day in 2011, and 340 million tweets for every day when Twitter commended its 6th year on March 21, 2012.

From Twitter's dispatch in 2006 and until 2009, the volume of tweets developed at progressively high rates, drawing closer a 1,400% increase in every day volume year to year and around 1,000% addition in yearly volume.

III. OPPORTUNITIES IN BIG DATA

There are many great apportunities ahead of us regarding "Big Data" in various fields. But the most interesting and futuristic approach that drawn us towards itself was "IOT(Internet of Things) with Machine learning and further using it to detect faults in big machines either mechanical, electrical or computerized machines. This idea will reduce the man power and human dependency and further, use the machine learning to make a smart and intelligent move in detecting machine faults and eventually fixing them on their own.
First we have to be clear with some common terms regarding this idea, which are as follows:

3.1. Fault detection, isolation, and recovery (FDIR)

It is a [5]subfield of control engineering which concerns itself with monitoring a system, identifying when a fault has occurred, and pinpointing the type of fault and its location. Two methodologies can be recognized: A direct pattern recognition of sensor readings that demonstrate a fault and an analysis of the inconsistency between the sensor readings and expected values, recieved from some model. In the last case, it is ordinary that a fault is said to be detected if the disparity orresidual goes over a specific limit. It is then the task of fault isolation to sort the kind of fault and its location in the machinery.
"Fault detection is necessary to avoid hazards and accidents to a certain limit and considering risk factor in any machine dependant organization."

3.2. Machine learning and data mining

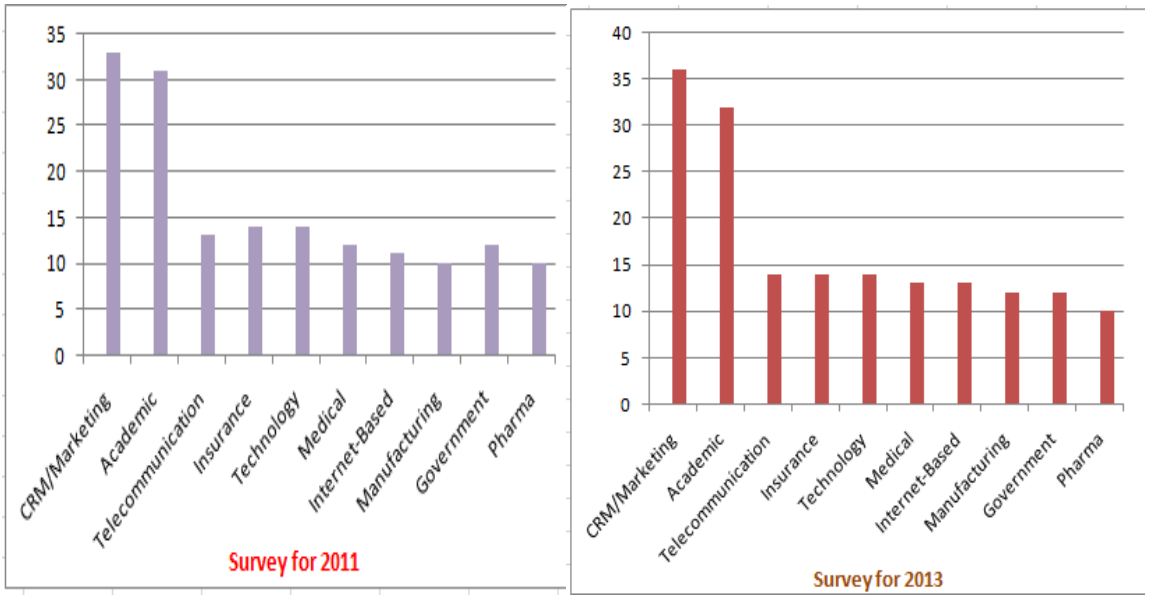
Early data analysis techniques were oriented toward extracting quantitative and statistical data characteristics. These techniques facilitate useful data interpretations and can help to get better insights into the processes behind the data. Although the traditional data analysis techniques can indirectly lead us to knowledge, it is still created by human analysts.
To go beyond,[6] a data analysis system has to be equipped with a substantial amount of background knowledge, and be able to perform reasoning tasks involving that knowledge and the data provided. In effort to meet this objective, scientists have swung to thoughts from the machine learning field. This is a characteristic wellspring of thoughts, since the machine learning task can be portrayed as turning background knowledge and cases (input) into knowledge (output).
If data mining results in finding significant patterns, data transforms into information. Information or patterns that are novel, legitimate and possibly valuable are not only information, but rather knowledge. One talks about finding knowledge, before covered up in the tremendous measure of data, however now uncovered.

3.3. Machine learning in Fault Detection in Machines – "Predictive Maintainance"

As awareness increases of the Internet of Things, more possibilities open up to connect physical objects with one another, making possible access to object-related data, such as sensor data and reports. This represents a huge amount of data that can be used in many ways. One possibility is predictive maintenance, where the data is used to detect upcoming failures before they occur, and schedule maintenance actions. The task of predictive maintenance is to determine when a machine must be maintained. Assessing the failure that occurs is called 'failure type detection' in this thesis. Failure type detection and predictive maintenance are related to each other. But there is also a difference between them. For example, sometimes it is not possible to detect the type of failure but an anomaly can be discerned. Thus, the task of predictive maintenance can be fulfilled, as well as there is no failure type detection possible. On the other hand, failures that occur could be detected and distinguished from one another. In this case, the failure type detection works, but by the time the failure is detected, it is too late to schedule any maintenance action. It is very obvious that for machine learning, the huge amount of data i.e. Big Data is majorly required. As machines will only learn much with much data provided to it. More the data, more intelligent is the machine. We have come up with an idea of using machine learning in fault detection in other machines, making it easy and convenient for us, reducing man power and making wide use of technology. A good technique in this area can lower the costs of damage, improve security and reduce the number of unnecessary maintenance actions.

Machine learning will accelerate, and will be increasingly applied within the faulty, fraud and risk sectors. Data-scientist demand and supply continues to work towards equilibrium. Advanced techniques will start to be applied within faults, fraud and risk that improve models and allow acceleration towards more real-time analysis and alerting. This acceleration will come from education and real-world applications of market leaders.

This idea of ours is hoped to be applied in the future for the concerned topic, making us step into the future.



The above graph shows growth in the big data in various fields in the year 2011 and 2013 respectively

IV. ADVANTAGES OF BIG DATA AND ANALYTICS

Some of the basic, but much needed and vital advantages of Big Data collection, it's prediction and analysis which might really help us stepping into the future are as follows:

4.1. Cost Reduction

Big Data technologies like Hadoop and cloud-based analytics can give considerable cost advantages. While correlations between big data technology and customary architectures (data stockrooms and bazaars specifically) are troublesome on account of contrasts in usefulness, a price comparison alone can recommend order-of-magnitude improvements. Settled firms like Citi, Wells Fargo and USAA all have significant. Hadoop ventures in progress that exist alongside existing stockpiling and processing capabilities for analytics. While the long haul part of these technologies in a venture architecture is misty, it's reasonable that they will play a perpetual and critical part in helping organizations oversee big data.

4.2. Faster, better decision making

Analytics has constantly included endeavors to enhance decision making, and big data doesn't change that. Substantial associations are looking for both speedier and better decisions with big data, and they're discovering them. Driven by the velocity of Hadoop and in-memory analytics, a few organizations we inquired about were centered around accelerating existing decisions.

For instance, Caesars, a leading gaming organization that has long embraces analytics, is presently grasping big data analytics for quicker decisions. The organization has data about its clients from its Total Rewards loyalty program, web clickstreams, and real-time play in slot machines. It has customarily utilized all those data sources to comprehend clients, yet it has been hard to incorporate and follow up on them in real time, while the client is as yet playing at a slot machine or in the resort.

4.3. New Products And Services

Perhaps the most interesting use of big data analytics is to create new products and services for customers. Online companies have done this for a decade or so, but now predominantly offline firms are doing it too. GE, for example, has made a major investment in new service models for its industrial products using big data analytics.

Verizon Wireless is also pursuing new offerings based on its extensive mobile device data. In a business unit called Precision Market Insights, Verizon is selling information about how often mobile phone users are in certain locations, their activities and backgrounds. Customers thus far have included malls, stadium owners and billboard firms.

V. CONCLUSION AND FUTURE CHALLENGES

These examples make clear that big data analytics projects are delivering value.The only disadvantage of big data is that it is really very big and complex to handle. Considering it's value in the market, currently, there is no approach for the feature extraction of physical properties with regard to failure type detection. Such a technique must convert the data given into the physical properties of a component. There is also no known feature

selection technique designed for failure type detection and predictive maintenance, in which the features, with the most physical relevance. With respect to upcoming failures, are selected. Investigations into the requirements of a feature selection technique, and how to implement them, could be another area of future work.

REFERENCES

1. www.internetlivestats.com
2. www.blog.twitter.com
3. www.mashable.com
4. www.forbes.com/sites/onmarketing
5. www.wikipedia.com
6. www.multiply.at
7. Chikhale Rupali, "A Data Mining: Overview to Distributed Systems", International Journal of Scientific and Research Publications, Volume 4, Issue 6, June 2014.
8. Raghavendra Kune1, Pramod Kumar Konugurthi, Arun Agarwal, Raghavendra Rao Chillarige and Rajkumar Buyya, " The anatomy of big data computing", In Proceedings of Wiley Online Library (wileyonlinelibrary.com)., 9 October 2015.
9. Drineas P, Krishnamoorthy M, Sofka M and Yener B, "Studying e-mail graphs for intelligence monitoring and analysis in the absence of semantic information", In Proceedings of the Symposium on Intelligence and Security Informatics, Vol.3073, June 2004
10. Saroiu S, Gummadi K, Dunn R, Gribble S and Levy H, "An analysis of internet content delivery systems",In Proceedings of OSDI, Vol.36, Issue SI, December 2002.
11. www.predictiveanalyticsworld.com
12. Koutroumpis, P., & Leiponen, A. 2013. Understanding the value of (big) data. In Proceedings of 2013 IEEE international conference on big data. 38–42. SiliconValley, CA, October 6–9, 2013. Los Alamitos, CA: IEEE Computer Society Press.