Pragmatic Approach of big data analytics

Yogita Rawat[1], Debodutta Sinha[2], Neha Katwale[3]
[1, 2, 3] Dept. of IT & Analytics
[1, 2, 3] ITM Business School

Abstract- *This paper presents the concept, characteristics, the need of big data analytics in every organization to support organizational decision making and problem solving, improve operational efficiency, drive new revenue and gain competitive advantage over business rivals during regular period and specially to be ahead in this era of fierce competitive world so that the organization can survive during turbulent times. There are different types of analytics application to consider descriptive statistics which focuses on describing something that has already happened, Predictive analytics which is used to make predictions about unknown future events and prescriptive analytics which subsumes the results of predictive analytics to suggest actions that will best take advantage of predicted scenarios. These are used at different places and time depending upon the application.*

Keywords- Tableau, JasperSoft, BI, Sentiment Analysis

I. INTRODUCTION

The idea for big data analytics have been around for years but many organizations have now started to understand that if they use the data that comes into their businesses, they can apply analytics and can get significant value from it. But even years before anyone spoke the term “big data”, business-es were still using some sort of analytics with the help of software like spreadsheets to uncover in-sights and know about the trends. But due to the emergence and advancement now in analyz-ing big data, there are new benefits which will ben-efit the organization such as speed and time.

Whereas a few years ago, a business would have gathered information, run analytics and unearthed information that could be used for future decisions, today that business can identify insights for imme-diate decisions. The ability to work faster – and stay agile – gives organizations a competitive edge they didn’t have before.

II. NEED OF BIG DATA ANALYTICS IN ORGANIZATIONS

Figure 1. Big data in different sectors [3]

I. Cost Reduction: Big data technologies such as Hadoop and cloud based analytics bring significant cost advantages when it comes to storing large amounts of data and they can identify producing more efficient and effective ways of doing busi-ness.

II. Faster, Better decision making: With the help of three types of analytics i.e. predictive, descrip-tive and prescriptive analytics and the tools, busi-nesses can analyse information immediately and can make decisions quickly.

III. New products and services: With the ability to gauge customers’ needs and satisfactions through analytics comes the power to give custom-ers what they want.

III. COMPANIES USING BIG DATA ANALYTICS

1. Bank of America

Today the bank is focusing on big data, but with a focus on a (having different things working togeth-er as one unit) approach to customers and a (having different things working together as one unit) or-ganizational structure. It thinks of big data in three different "buckets"--big transactional data, data about customers, and (without rules, schedules, etc.) data. The focus is on the first two categories. With a very large amount of customer data across many channels and relationships, the bank (in the past) was unable to carefully study all the custom-ers at once, and depended on

well-thought-out samples. With big data technology, it can more and more process and analyse data from its full customer set. Other than some experiments with analysis of (without rules, schedules, etc.) data, the primary focus of the bank's big data efforts is on under-standing the customer across all channels and inter-actions, and presenting consistent, appealing offers to customer parts.

2. Macys.com

Within the Macys organization, the "Customer insights " group deals with these issues, but it also has a "Business insights " group (focused primarily on supporting and measuring activity around the marketing calendar) and a "Data Science" organiza-tion. The latter addresses more leading-edge (hav-ing to do with measuring things with numbers) ways of doing things involving data mining, marketing, and experimental design. Macys uses a vari-ety of leading-edge technologies for big data, most of which are not used in other places within the company. They include open-source tools like Ha-doop, R, and Impala, as well as (bought) software such as SAS, IBM DB2, Vertica etc.

3. Amazon

Amazon.com Inc. collects, store, process and analyze personal information of its customers to know how customers are spending their money. The com-pany uses predictive analytics for targeted market-ing to increase customer satisfaction and build company loyalty.

Personalised recommendation in shopping- It uses a collaborative filtering engine. It analyses your pre-vious purchase history and what items you intend to buy by checking your wish list, your most searched items. It also checks your what products you have reviewed or given rating, this type of in-formation is used by the amazon for recommending additional products that other customers might buy while purchasing same items.

Book recommendation -In a similar manner amazon also uses big data for recommending books you might be intrested in reading by analysing what their customers has bought in the past and which items they have in their virtual shopping cart. They also check the items you have either rated, liked or both.

Anticipatory Shipping Model:

Amazon’s patented anticipatory shipping model.It uses big data for predicting what products you are likely to purchase and when you may buy them and where you might need the products. The items are sent in advance to a local distribution centre or warehouse or even in trucks until an order arrives so they will be ready for shipping immediately. This way amazon uses predictive analytics to increase its product sales and decreasing overall expense.

Supply Chain Optimization:

Amazon is linked with their manufacturers and tracks their inventory. Amazon also uses big data for selecting warehouse closest to the vendor and the customer to reduce shipping costs. Other than this it also uses graph theory which helps in decid-ing the best delivery schedule, route and product groupings to further reduce shipping expenses.

4. General electric

The company is putting sensors on gas turbines, jet engines, and other machines. There aim is to identify ways to improve machine efficiency thereby reducing cost, time and resources. To han-dle this huge amount of data, GE developed a cloud-based software platform named Predix that could provide machine operators and maintenance engineers with real-time information to schedule maintenance checks, improve machine efficiency, and reduce downtime. This information can be used for GE's own product development activities and to lower costs in its service agreements which the companies provide to their customers with the machine. Using this data, they could find interest-ing and unique patterns in the data and to predict when the time for maintenance.

IV. VARIOUS BIG DATA ANALYTICS TOOLS

1. HADOOP

Hadoop is an open source software framework for distributed storage of very large datasets on computer clusters. It also helps in scaling up our data up and down without having to worry about hardware failures. Hadoop provides massive amounts of storage for any kind of data, enormous processing power and the ability to handle virtually limitless concurrent tasks or jobs.

List of some of the companies using Hadoop tool:

Facebook: Hadoop is used as a source for report-ing and machine learning.

Twitter: Hadoop is used since 2010 to store and process tweets.

LinkedIn: LinkedIn's data flow through Hadoop clusters.

EBay: EBay has one of the largest Hadoop clusters in the industry that run prominently on MapReduce jobs. Hadoop is used at EBay for search optimiza-tion and research.

Fox Audience Network: Hadoop is used for log analysis and machine learning.

2. JASPER SOFT BI SUITE:

Jaspersoft is one of the open source tool for pro-ducing reports from database columns and visualize data stored in big data engine. It is compatible to take data from many of the major storage platforms such as Hadoop, MongoDB, Cassandra, Redis, Ri-ak, CouchDB, and Neo4j.It's beneficial when we can generate reports from the big data wherever it is stored and this can be carried out by using Jasper reports server. Jaspersoft can generate reports, rep-resentation in dashboards, analytics, and data inte-gration. It has interactive reporting features where in individual can filter inline, sort, hide/move col-umn, conditional formatting, change chart types, export to xls etc. User can also drill down, change parameters or charts, HTML 5 visualisation can also be done in the interactive dashboard. Browser based drag and drop designer, visualize in table or chart format, save and share insights with others can also be done in the adhoc analysis.

3. PENTAHO:

Pentaho is another software (raised, flat supporting surface) that began as a report creat-ing engine; it is like jaspersoft, branching into big data by making it easier to soak up (like a towel) information from the new sources. We can hook up Pentaho's tool to many of the most popular nosql (computer files full of information) such as mon-godb and Cassandra. Once the (computer files full of information) are connected we can drag and drop the columns into views and reports as if the Information came from SQL (computer files full of information).

Pentaho also provides software for drawing HDFS file data and hbase data from Hadoop groups. One of the more interesting tools is the graphical pro-gramming (connecting point/way of interacting with something) Known as either Kettle or Pentaho Data (Combination of different things together that work as one unit). It has a bunch of built-in mod-ules that we can drag and drop onto a picture, then connect them.

4. TABLEAU:

Tableau Desktop is a visualization tool that makes it easy to look at our data in new ways, then slice it up and look at it in a different way. We can even mix the data with other data and examine it in yet another light. The tool is optimized to give us all the columns for the data and let us mix them before stuffing it into one of the dozens of graphical tem-plates provided. [1]

5. SKYTREE:

Skytree is more focused on the guts than the shiny GUI. Skytree Server is optimized to run several classic machine-learning algorithms on your data using an implementation the company claims can be 10,000 times faster than other packages. It can search through your data looking for clusters of mathematically similar items, then invert this to identify outliers that may be problems, opportuni-ties, or both. The algorithms can be more precise than humans, and they can search through vast quantities of data looking for the entries that are a bit out of the ordinary. This may be fraud -- or a particularly good customer who will spend and spend [5]. The free version of the software offers the same algorithms as the proprietary version, but it's limited to data sets of 100,000 rows. This should be sufficient to establish whether the soft-ware is a good match.

6. SPLUNK:

Splunk is a bit different from the oth-er options. It's not exactly a report-generating tool or a collection of AI routines, although it accom-plishes much of that along the way. It creates an index of your data as if your data were a book or a block of text. Yes, databases also build indices, but Splunk's approach is much closer to a text search process. Splunk takes text strings and search around in the index. We might type in the URLs of important articles or the IP address. Splunk finds them and packages them into a timeline built around the time stamps it discovers in the data. All other fields are correlated, and you can click around to drill deeper and deeper into the data set. While this is a simple process, it's quite powerful if you're looking for the right kind of needle in your data feed. If you know the right text string, Splunk will help you track it. Log files are a great applica-tion for it.

V. FINDINGS

According to the analysis done from "http://insidebigdata.com/2016/12/21/big-data-industry-predictions-2017/" by using "Antconc" tool we found that 73% people from different or-ganizations are talking about big data in today's world. The keyword analysis done in the antconc is shown as below:

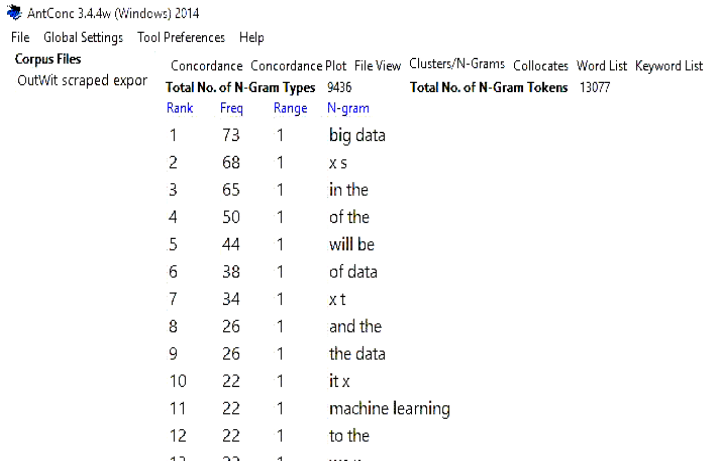

Figure 2. Keywords Frequency List

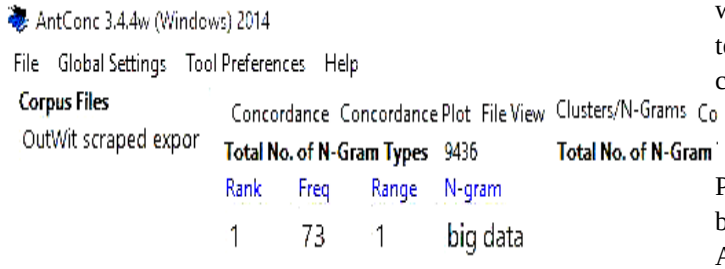

Figure 3. 'Big Data' Keyword Frequency

Some of the comments made by diffrerent people from different organizatons are mentioned below:

"IT becomes the data hero. By empowering the or-ganization to make data-driven decisions at the speed of business, IT will emerge as the data hero who helps shape the future of the business.

Francois Ajenstat, Chief Product Officer at Tab-leau"

"In 2017, we are going to see analytics do more than ever to drive customer satisfaction. As the world of big data exploded, business leaders had a false comfort in having these mammoth data lakes which brought no value on their own when they were sitting unanal
yzed.

Ketan Karkhanis, SVP and GM of the Salesforce Analytics Cloud" said "In the past, big data analy-sis has lagged in implementation compared to other parts of the business being transformed by ad-vanced technology such as supply chains. I believe companies will begin to place different data storage systems into the hands of end users in a fast and efficient manner that has user self-direction and flexibility, democratizing data analysis.

Chuck Pieper, CEO, Cambridge Semantics"

"The battleground for data-enriched CRM will on-ly continue to heat up in 2017. Data is a great way to extend the value proposition of CRM to busi-nesses of all sizes, especially those in the small-to mid-size range. The key, of course, is seeing which providers provide the most seamless and most sen-sible use cases out of the box for their customers.

Martin Schneider, Vice President of Corporate Communications, SugarCRM"

"In 2017 (and 2018), streaming analytics will be-come a default enterprise capability, and we are going to see widespread enterprise adoption and implementation of this technology as the next big step to help companies gain a competitive ad-vantage from their data.

According to Mr. Anand Venugopal, Head of Product, StreamAnalytix, Impetus Technologies, Big Data will be subsumed into the topic of AI, as big data is an enabler of AI not an end in itself. The lack of focus on big data will let the field mature with only the serious players and result in much better business results.

Anil Kaul, Co-Founder and CEO of Absolutdata"

After analyzing all the comments made by the peo-ple of different organizations it is being observed that:

- People are very much interested in big data analytics in today's world. Business and IT leaders are using big data analytics to iden-tify trends, detect patterns and find other valuable information from the ocean of in-formation available.
- Only big data technologies are not suffi-cient to make the organization successful, skilfull resources are also required to carry out an effective and efficient work.
- Many companies want to include big data in their companies but as due to the lack of skillfull people in the field of big data ana-lytics so they are in dilemma whether to in-clude it or not in the organization.
- Many fortune 500 brands are using Big Da-ta Analytical tools in their companies which brought huge success in the organization.
- Many companies have ideas and initiatives around big data but no proper execution plan.

1. Sentiment Analysis:

After analyzing the data from "http://insidebigdata.com/2016/12/21/big-data-industry-predictions-2017/"

We found out that the organizations and the people of various organizations have positive perception towards inclusion of big data analytics in the or-gnizations for data quality improvement, faster ex-ecution, customer satisfaction, making more profits and lot more.

The positive information percentage is 48.4%.

The percentage of people who think that the inclu-sion of big data analytics will be a problem in the organization or its not the right time to execute big data analytics in the organization is 15.9% i.e. the negative information percentage.

The neutral information percentage of the same is 35.7% .

From which we can analyze that people are in di-lemma to include the big data analytics in the or-ganization or not.

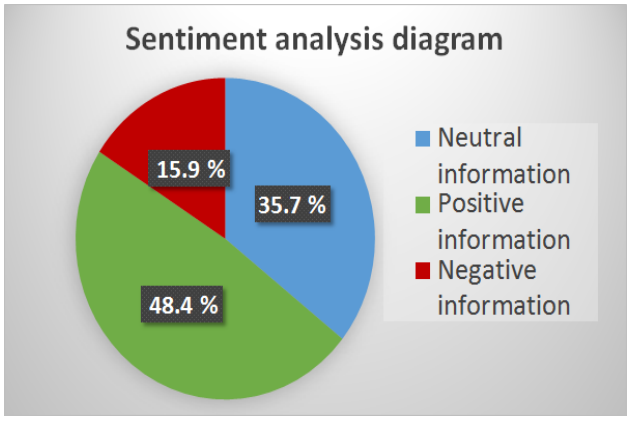

Figure 4. Sentiment Analysis

2. Positive Sentiment Analysis:

The factors for positive sentiment analysis are as follows:

- Improvement in technology make things easier and automatic
- Providing personalization, flexibility and agility
- Democratization of Data Analysis
- Technology empowering business users
- Enhanced customer satisfaction and experi-ence
- Data quality improvements

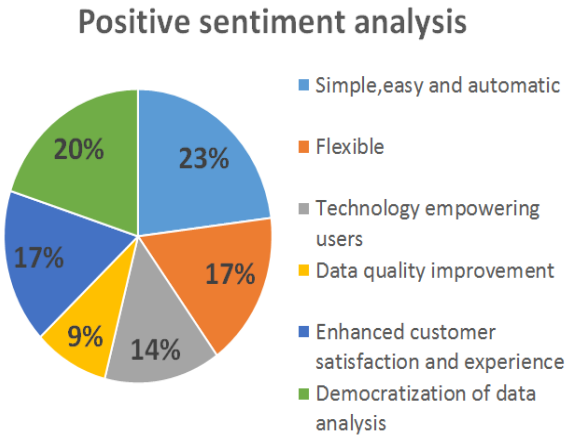

Figure 5. Positive Sentiment Analysis

3. Negative Sentiment Analysis:

The factors for negative sentiment analysis are as follows:

- Implementation lagging
- Lack of skillfull resources
- Managing the systems and data i.e security
- Complexity
- High cost of entry and ongoing expenses

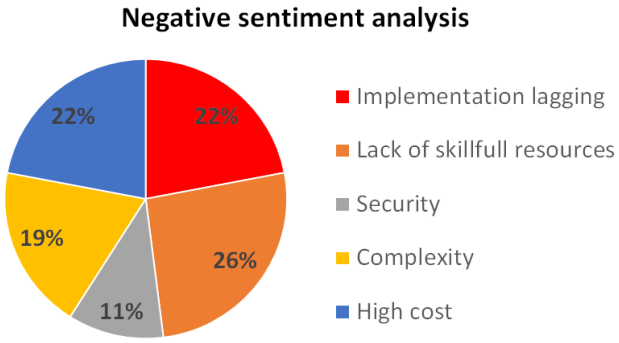

Figure 6. Negative Sentiment Analysis

Apart from that, after doing the keyword analysis we found out that there are some keywords in the text document which are being repeated most which is shown in the following table:

Table 1.

KEYWORDS	FREQUENCY
Data	197
AI [artificial Intelligence]	89
Business	43
Analytics	42
Companies	38

The following picture is the representation of the above data visually:

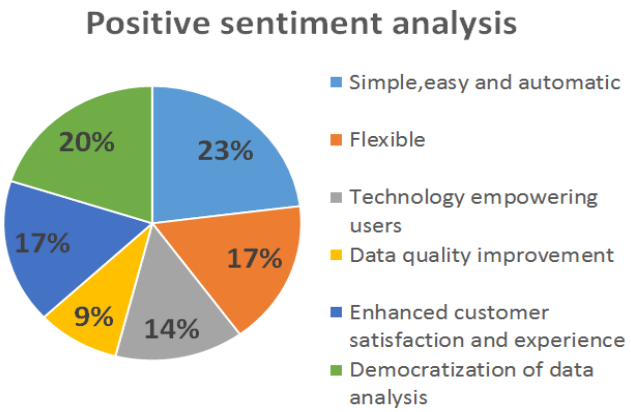

Figure 7. Keyword Analysis [11]

VI. CONCLUSION

This paper focuses on the contribution of big data analytics in various organizations and how it will help in today's world and can predict about the fu-ture so that it can help the organization in turbulent times.

There are various tools of big data analytics which can be used in every organization and can improve the business.

The leaders of every organizations are interested in inclusion of big data analytics to their industries but due to the lack of skillfull resources and lack of proper implementation plan, many organizations are still in the thinking process and not actually includ-ing big data in the organization.

REFERENCES

[1] Richard Wesley, Mathew Eldridge, Pawel T. Teelecki, "An analytic data engine for visualization in tableau" Tableau Software Seattle, Proceeding SIGMOD'11 Proceesings of the 2011 ACM SIG-MOD International Conference on Management of data Pages 1185-1194

[2] Available at: http://blogs.systweak.com/2016/10/big-data-a-future-nightmare/

[3] Available at: http://www.tcs.com/SiteCollectionDocuments/White-Papers/Predictive-Analytics-HR-0115-1.pdf

[4] Available at: https://blog.galaxyweblinks.com/big-data-with-bigger-benefits/

[5] Details available at: http://www.skytree.net/

[6] Agrawal D., Bernstein P., Bertino E., Davidson S., Dayal U., Franklin M, Widom J. (2012)," Challenges and Opportunities with Big Data": A white paper prepared for the Computing Community Consortium committee of the Computing Research Association.

[7] Details available at: http://insidebigdata.com/2016/12/21/big-data-industry-predictions-2017

[8] Ray M. Chang, Robert J. Kauffman, YoungOk Kwon "Understanding the paradigm shift to computational social science in the presence of big data".

[9] Eleanor O'Neill, "companies that are using big data" Available on site: https://www.icas.com/ca-today-news/10-companies-using-big-data

[10] 1st Nada Elgendy, 2nd Ahmed Elragal, "Big Data Analytics in Support of the DecisionMaking Process" A white paper prepared for Conference: CENTERIS - Conference on Enterprise Information Systems, At Porto, Portugal

[11] Keyword analysis tool: https://voyanttools.org/?corpus=80a9de0aa6f0311afaed02dbe9899d78