



Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 116 (2017) 274–283

Procedia
Computer Science

www.elsevier.com/locate/procedia

2nd International Conference on Computer Science and Computational Intelligence 2017, ICCSCI
2017, 13-14 October 2017, Bali, Indonesia

A Review: Big Data Analytics for enhanced Customer Experiences with Crowd Sourcing

Laika Satish*, Norazah Yusof

Faculty of Computing and IT, King Abdul Aziz University, Saudi Arabia

Abstract

The objective of this paper is to analyze and study the ways of considering customer experience big data analytics and incorporate crowd sourcing in the process to increase customer satisfaction and loyalty, revenue and greater employee satisfaction where commerce is immediate and secure. In an era where customer experience play a very important role to improve the necessary target profits and where customers are interacting with many brands across many channels, big data through crowd sourcing gives us a way to focus on the outcomes to increase collaboration coming from expert crowdsource workers, data providers and big data analytics. It is one of the biggest trends that is going to dominate the year 2017-2018 where big data is an exploding area of focus and where business intelligence can be transformed into crowdsourced business intelligence. Since it is nowadays a smart digital world where applications has to support essential workflows, the emerging research areas in big data and crowdsourcing can explore the means to provide big benefits and great potential from a customer service prospective. Findings give us a better understanding how companies are able to understand and monetize their customers better.

© 2017 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of the scientific committee of the 2nd International Conference on Computer Science and Computational Intelligence 2017.

Keywords: Big data, crowd sourcing, Customer Experiences, CX Analysis.

* Corresponding author. Tel.+966-590654342.

E-mail address: lkannan@kau.edu.sa, laika4u@gmail.com

1. Introduction

From past some years, the biggest challenge companies are facing in this area with massive data is translation of business to models. Since it is just an evolving area, it is difficult even to define a problem thoroughly for example non technical staff can't communicate the problem definition to the technical data scientists¹. A large amount of data is produced from structured, semi structured and unstructured sources which makes it very difficult to manage information about customers usage. Mainly it's the need of big Shopping giants like Amazon and Flipkart to process huge amount of data so as to process massive datasets to find hidden patterns. These data have to be extracted

,transformed and loaded to study patterns related to customer behavior and their interactions. Companies can improve the overall performance by using the customer feedbacks. Survey results from big data show them where they stand today in the competitive market. Crowd sourcing is an idea which is different from Big Data, crowdsourcing projects have shown the potential of using a wide group of real people to collect useful and accurate data².

To begin with, there cannot be reliable data management without some human involvement to check for errors and the administrators can easily make their own mistakes and put their own subjective slant on the filing of information². Big data companies need crowdsourcing in their operations to ensure objectivity and diversity, prevent against errors more effectively and let social trends play a part in data examination. This combination of big data and crowdsourcing provides a way to modern data collection for customer experiences to make information more valuable, authentic and manageable. Big benefits can be reaped by pairing up crowdsourcing with big data².

2. Literature Survey with a perspective of big data in customer experiences

In order to get possible required research from the published literature, a systematic literature search capturing big data with crowdsourcing work of the recent three years have been undertaken .i.e. SpringerLink, ScienceDirect, Scopus, Procedia and many international journals as well as the Journal of MIS (JMIS) were searched,

Big data is getting larger and larger day by day and data continues to explode. Assuming valuable customer experiences from large amounts of structured and unstructured data from different sources in different formats require the proper structures and tools. To obtain the maximum business impact, this process requires proper combination of people, process and analytic tools³.To improve lasting relationships, programs that influence big data need to think about more tactical ways about making the customers stay, their loyalty, and relationships. The problems that need to be solved should be focussed not on whether big data can show us what will prompt the customer to do a next purchase, but how can we maintain the loyalty of the customer. It's not about what are the customers total transactions or how much profits the customer is giving but till when the customer will stay, What's his value to the organization is so that the customer won't go to the competitor if the competitor provides a lower best price^{4,5}.

Examples like Amazon and Netflix uses big data in the form of recommendation engines that create value for customers by helping them search what they need. Big data works with crowd sourcing to answer some questions like “what can I learn from other customers?” or “how do I compare with other customers?”.what is their purchase behaviour. What are the experiences of unhappy customers^{6,7}.

Another example is Opower where customers can share their bills with Facebook friends to know the differences among other customers. INRIX, collects traffic data from customers' mobile phones and other sources to provide real-time traffic reports. Zillow combines information to provide consolidated insight about home attributes and values, competitive properties, to buyers, sellers, and agents. Customer behaviour is monitored in real time and analysed to look carefully into processes to assess evidences and yield proper results. We have more examples like Diabetes UK extends their methods for shoppers on the basis of their food shopping history and gives them with advice on how to reduce their diabetes risk^{8,2}.

By targeting customer data many organisations can compare the data on a larger scale and improve their productivity. Viacom, owner of household brands such as Comedy Central, Nickelodeon and MTV, has made a real-time big data analytics platform using Apache Spark and Databricks, through which they can check the quality of videos feeds and reallocates resources in real-time.

Past findings on big data shows that 70% of customer data is never used for making improvements. Only 30 % organizations are able to use CX to know about their organizations pitfalls and help to succeed in today's market^{8,2}.

3. Analysis: A look on the big data basics with crowdsourcing and analysis for customer experiences

Big data is defined by three dimensions Volume, Velocity and variety. Additional dimensions are veracity and value. Using Unstructured data, we have to predict the type of analytics for analysis and find customer behavioural pattern. Then comes Discovery of this golden data^{9, 8, 10, 1, 5}.

For the purpose of processing, they use Hadoop for programming framework with Mapreduce algorithms. The Current Apache Hadoop ecosystem consists of the Hadoop Kernel, Mapreduce, HDFS and numbers of various components like Apache Hive, Base and Zookeeper. Now here crowdsourcing can be combined with big data using Amazon web services technologies like Elastic Map reduce and Mechanical Turk for extracting top K queries of customer data from uncertain data. Crowdsourcing helps in collecting facts from the crowd who may have innovative ideas or subject matter expertise to contribute for customer experiences. Crowdsourcing feedback from your customers lets you know what you can improve [about] your product and how to serve them better," Crowdsourcing will help you learn about their frustrations and what parts of your product don't work as intended.

The process can be divided into Discovering customer experience example using text analytics or through social media, YouTube and take corrective necessary actions. This information has to be transformed and leveraged in various ways using mining, cleansing and modelling. From this sentiment analysis can also be done which can yield us with customer bonds and loyalty. Integration of big data platforms across locations needs to be done.

In Data Interpretation step, visualising data and making data understandable for users is done where data analysis and modelling results are presented to the decision makers to interpret the findings for extracting sense and knowledge⁸.

4. Types of analysis which can be done for customer experiences^{9, 8, 10}

Analytics refers to the methods used to analyze and acquire intelligence from Big data which has to be done with crowdsourcing also. Of the articles in the sample of crowdsourcing with big data, 23% are conceptual articles and deals with foundations of crowdsourcing; 75% follow an experimental ways, where conclusions are drawn from case studies of companies and examples on crowdsourcing projects; the remaining 2% is made up of review papers¹¹

Prescriptive: Through this type of analysis we can know what actions can be taken regarding customer experiences. Through which we can draw rules and make recommendations

Predictive : Using predictive analysis, We can identify history of data patterns of customers and provide information of outcomes in a given situation like forecasting

Diagnostic : Diagnostic analysis differs from predictive analysis in a way which we can do root cause analysis and uncover patterns example assessing the users shares to other users.

Descriptive : Descriptive analysis can be used to mine the customer experiences to know what is happening in real time.

Some variables through which we can analyse customer data are behavioural and lifestyle preferences of customers. Behavioural like frequency of buying or online behaviours and lifestyle like interests of customers and hobbies of customers. Organizations need to monitor their customers across channels of wherever they are storing their data. Modelling can be used to predict the data like response modelling through which we can identify how customers respond to offers. We can employ here crowdsourcing as a type of participative online activity in which an online giant proposes to a group of customers of varying knowledge and preferences to voluntarily undertake some tasks.

5. Discussion

Findings and practical implications :

We can consider Starbucks who can open three branches on the same street. What made them take this step? This coffee giant uses big data with crowd sourcing to determine the success of each new location, taking information on location, traffic, area demographic and customer behaviour into account. so that the crowd will tell more about the information contained within the data points collected.

Such assessment before opening a store means Starbucks can make a fairly accurate estimation of what the success rate will be and choose locations for revenue growth Crowd sourcing thus helps to provide structure (document editing, audio transcription, image annotation) to big data thereby helping analysts to improve predictions by 25% ⁴.

Another example is Refugee crisis in United Nations is being tackled with the intersection of big data and crowdsourcing. United Nations Refugee Agency has partnered with Mindjet’s SpigitEngage platform that uses cloud computing to find out and solve problems through crowdsourcing.

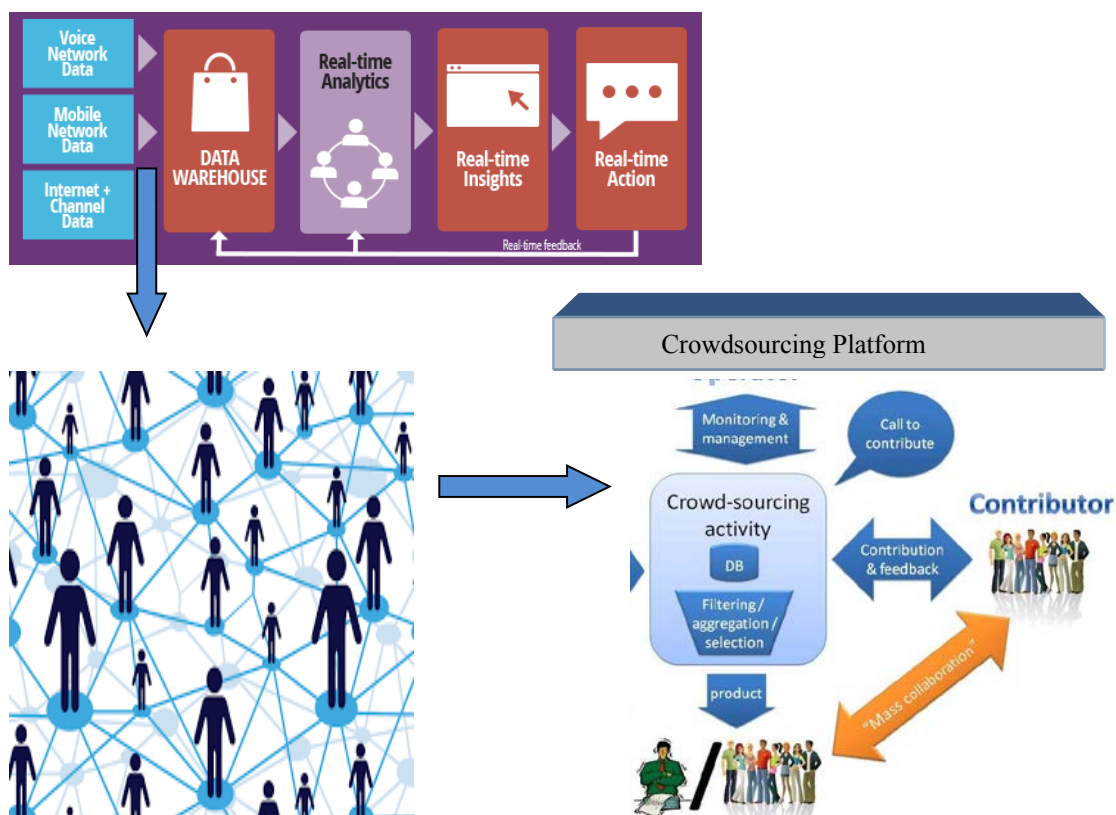
Thus big data with crowdsourcing model can carry out the given tasks fast with accuracy and at a lower cost. Crowdsourcing workers can do big data operations like- data cleansing, data validation, data tagging, normalization and data entry ⁶.

5.1 Practical Implementation approached :

Some questions that comes into our mind are

- 1.Where do we store our customer data for analysis ?
- 2.How do we pull all data together for understanding our customers ?
- 3.How can we use crowdsourcing and analysis example like predictive analysis, customer churn analysis , sentiment analysis ,Behaviour patterns, text analytics ,Demographic data analysis ,Transactional data analysis ? This step also leads us to make leveraging life changing Use cases
- 4.Visualization of data all on a single dashboard .

Fig.1 shows incorporating crowd sourcing in the big data process for customer Analytics.



Crowd Sourcing (Workers)

Main requirement is it should be able to handle massive data and provide operations on it per second. Unstructured data can be stored in Hadoop servers like Apache Hadoop which can store data in cluster ,NOSQL like

MongoDB, MPP databases, HIVE for distributed data ,SQOOP to transfer data from relational database ,SAAS Solutions.^{11, 12, 13}

For this we have to understand and navigate federated data sources ,Here we can crowd source data also for big data .Traditional sources like content repository for documents and shared operational information. Also information from governing systems operating inside and outside the organizations ,streaming data ,Unstructured data from text analytics engine. With crowdsourcing ,the enormous massive data can be split as work among crowd workers c_w which are network of experts in BI and the crowdsourcing System will make associations so that the job can be done with total assurance .The n number of projects p which when divided into tasks given to workers W_t where the workers are customers with n ways to become involved with every task which may offer rewards R.A pre-screening procedure can make it sure that right person is assigned the right task.^{15, 16, 17}

Elastic Map reduce algorithms are used here to scale data from multiple processing platforms for crowdsourcing. The input data is passed to the mapper function. The mapper processes the data and creates several small chunks of data. We can pass the data through a crowdsourcing platform for top Queries over uncertain data. After cleansing step ideas become more organized so there is no need to analyse unrelated comments /Suggestions . The Reducer processes the data that comes from the mapper to form appropriate results about the customer. The data can come from CRM data servers, ERP data servers, App server logs ,Click stream data ,web access logs, Social media logs, call centres data ,legacy systems. The data can be divided into 2 ways static data and dynamic data .Static data like customer's residency permit number ,name and other details which are same across multiple data sources .and dynamic data like where the customer shops or what he purchases more or what are his lifestyle preferences, The data can also come from his shoppers card like sensor data .Here we can also pull data from crowd sourcing engines and use intelligent systems . We can achieve real time recommendations from crowd sourced data^{11, 12, 13, 18, 19} .All this will influence our decisions in the next step.

For analysis we can run analysis algorithms on all the customer data from the above step using Hadoop distribution to initially filter and clean important events. This step will help us to drive insights from unstructured customer data. Text analytics can be used here. Example data can come from feedback of customers so that organizations can understand problems before they become big. Sentiment analysis to identify forms of communication to make improvements in customer satisfaction. Crowdsourcing a campaign analysis to see customer segments in more detail. Customer usage analysis with segmentation to get a better understanding of customers loyalty and create relationships among crowds for likes and dislikes, behavioural analysis for example a customer started shopping on a website but later leaves the page and goes after sometime if he forgets where he found that item on which website, behavioural analysis can help him track his history of data or another example like Customer conversations on social media about a purchase or comments on a YouTube channel about a product or know shopping behaviour to find gold or platinum customers or positive/negative experiences on twitter. Like this We can link all the information about a customer to extend it to new use cases. We can provide a consistent view of the customer by integrating dynamic recommendations and collaboration tools to reveal more about the customer and to enable positive customer outcomes and to provide indexing and searching on data. This step will help us to increase market penetration and to reduce customer churn by churn modelling for example to know about purchase frequency which will also help us to make retention strategies in real time marketing^{11, 12 20, 21} .

We should be able to view the data real time per hour or even per second or detailed report analysis also taking into account different locations and different attributes which is also known as 360⁰ view of the customer example monthly churn graphs from multiple channels. This step should show customer records in different formats. Fig 2 and Fig 3. depicts this. It should show customer's interaction history with outcomes. This will help us to know customer satisfactions and dissatisfactions about the products. From crowd sourcing perspective it should show positive and negative outcomes of product promotions ,customer loyalty and retentions . These visualizations should be able to give recommendations how to promote products and new customer acquisitions so to enhance customer support and co learning of different aspects. Through detailed visualization we can trouble shoot errors and do anomaly detections ,can conduct Omni channel pricing and promotions and payment flow analysis^{14 , 22} .

Customer 360° View

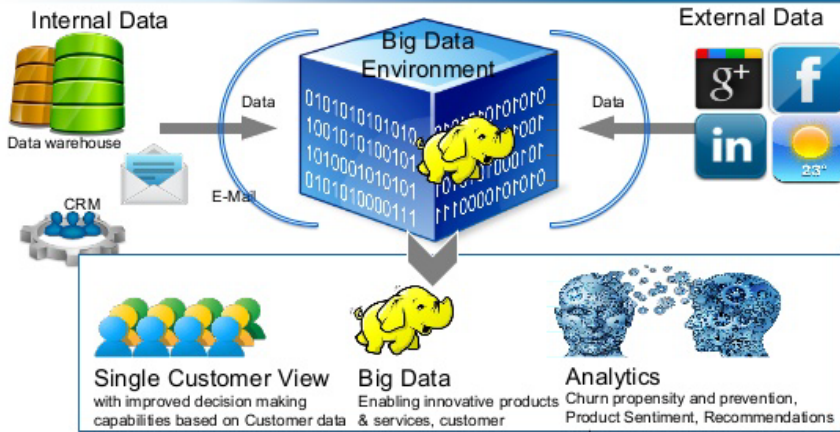


Fig 2. 360 degree view of the customer

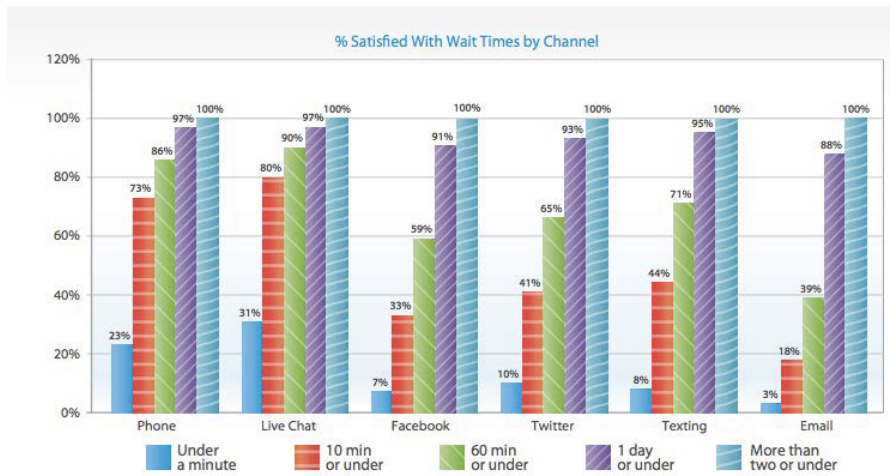


Fig 3. Example of Live visualization of Big Data from different sources

Some Frameworks which are under development /developed are developed by companies Terradata, ZoomData, Cloudera, HortonWorks, Datameer ^{15, 16, 17}.

5.2 Case studies :

Lufthansa group ,this leading airlines uses Terradata software to manage its vast sector of airlines and services and to integrate data to achieve operational excellence to maximize revenue. They created a common data language from multiple data sources since they believe that revenue can't be generated alone from data but it needs big data with crowd sourcing analytics to understand their data, So that they are able to look at each customers travel experiences ¹⁵.Carrefour group of markets, Toysrus, Ebay use Terradata Enterprise management model ,Demand chain analytics in big data. Their demand forecast was not consistently accurate, resulting in mistakes when predicting quantities to meet seasonal and consumer demand requirements. Opportunities existed to reduce excessive handling costs at store level and the amount of deep markdowns of unsold or shop-worn merchandise. The pairing gave them a broad range

of analytics designed around all aspects of managing Demand Forecasting, Inventory Replenishment, and Supply Chain Management ¹⁵.

American Airlines, Shoepassion.com, Priceline.com, National Instruments uses Datameer's big data customer analytics. It provides the easy access point to unlock the power of Hadoop. They used Hadoop for showing business value in a large Enterprise area. They used advanced analytics to gain insights from a multitude of data sources which helped them to calculate important metrics . Big data analytics helped retailers to make sense of their data by optimizing merchandising tactics, personalize the in-store experience with loyalty apps and drive timely offers to attract consumers to complete purchases. Shoepassion.com used marketing metrics model to understand customer buying behaviour and use this information to create cross sell recommendations. They used Datameer's FlipSide framework model which is an automatically generated visual profile of all the data, showing trends, anomalies, and outliers of the vast data and is available at every step of their analytic cycle ^{15, 24, 22, 25 28, 29}

Table1:Comparative analysis of software's ^{24, 25, 23}

Terradata	Zoomdata	Datameer
Linearly scalable, Extensive parallel processing with high fault tolerance and data protection with extensive cloud support but not suitable for OLTP databases .It's very expensive and a difficult software and complex analytics and concurrency operations. There are certain disadvantages on indexes, Name standardisation is not easy Sometimes tangling in queries can cause delayed processing and building complex models is difficult	Uses Google cloud support with analytics .Simple to use and less costly than Terradata simplified data exploration ,analysis with new real streaming models but doesn't support profit analysis ,predictive analysis and trend indicators	Specialities here are this software targets promotions provides an open architecture with operational analysis and fraud detection. It supplies a large set of algorithms to create predictive models but less matured reporting tools, Needs to adopt quickly to market trends .

Table2:Analysis of features that can be provided when using big data with crowdsourcing for customer retention

Data cleansing and Analysis	Segmentation	Process Automation and crowd sourcing	Data Exploration
Knowledge discovery, Parametric Queries, Group information ,Normalization on basis of ranges . Amazon Mechanical Turk good for these kinds of tasks. Crowd Source workers will discover errors and present accurate information, along with the	Visual, complex, Parametric segmentation using crowdsourcing the data from the first step will then be segmented into significant groups with weights with similar attributes. This segmentation will	Crowd sourcing recency ,frequency, Social Media analysis ,Churn Analysis, Forecasting, Task Scheduling, Monetary values. Workflow models have to be created with automated	Visualizations ,Representations, Data trees and value charts ,Statistical graphs .

<p>source, the worker model will be presented with some items and asked to verify some attributes or all information made known. The worker model can be given recommended list of sources as well as a list of invalid data sources. and this information has to be verified. If a worker conclude that the information is incorrect, the worker will provide the proper information and the source it came from, letting us update the customer massive database with accurate data which will then follow the types of analysis mentioned above.</p>	<p>describe very valuable information about how profit process or marketing channels, should be planned in order to achieve greater customer insights.</p>	<p>algorithms and analytical approaches and the crowdsourced data needs to be trained well with machine learning algorithms.</p>	
---	--	--	--

5.3 Limitations :

Still models are less to process data totally error free which are coming from different sources because the data is numerous in semi structured and unstructured ways. Data Scientists to process and work with the data is very less. After all customer insights with crowd sourcing gives us **what is happening** in real time. It doesn't answer the question **why it is happening**. There are very less models under research to show why this is happening .This is totally up to the managers to figure out using their own statistics analysis and find solutions. Many customers may operate across devices, and it is very difficult to tell for any given touch point sequence how distributed the way actually resides. Data can be collected only for online customers with crowd sourcing not for offline customers who do direct cash transactions. Sometimes visualizations are difficult to understand by domain experts. The companies are not able to do big learning's from customer data. Sometimes the tools we use to gather big data sets for customers are imprecise and more research needs to be done. Customer data is threatened by the need for data privacy. Big data methods are still prone to data infringements and breaches.

6. Conclusions: The idea in this paper can act as a guide for incorporating crowdsourcing with Big data in customer analytics which can have the potential to bring together a large group of crowd workers who are customers themselves on the same platform when there is an issue that affects them all. Crowdsourcing issues for customer retentions are mostly focussed around problems with the worth, correctness, and aggregation of data especially when the data is massive. However, these issues can be addressed in future research through proper planning and with an understanding of the final goal of crowdsourcing in big data for customer retention .

7. References

1. Nada Elgendy and Ahmed Elragal (. Big_Data_Analytics_A_Literature_Review_Paper. ResearchGate. 2014 August.

2. Jeffrey Spiess YT,D,PS,LP. Using big data to improve customer experience and business performance. IEEE Explore digital library. 2014.
3. Unleashing the potential of big data. In World Summit on big data and Organization Decision; 2013.
4. Thau B. forbes.com. [Online].
5. Thomas DM. A Review paper on BIG Data. Springer. 2014.
6. Arora S. martechnadviser.com,Recommendation Engines: How Amazon and Netflix Are Winning the Personalization Battle. [Online].; 2016.
7. Ebner K,BT,&UN. Think big with Big Data: Identifying suitable Big Data strategies in corporate environments. Proceedings of the 47th International Conference on System Science (pp. 3748–3757). Washington: DC: IEEE Computing Society. 2014.
8. Uthayasankar Sivarajah MMK,ZI. Science Direct Journal. 2017.
9. Kaur M. Big Data and Methodology. International Journal of Advanced Research in Computer Science and Software Engineering. 2014.
10. Harshawardhan S. Bhosale1 PDPG. A Review Paper on Big Data and Hadoop. International Journal of Scientific and Research Publications. 2014; Volume 4(10).
11. Antonio Ghezzi DG. Crowdsourcing: A Review and Suggestions for Future Research. Wiley Online Library. 2017 January.
12. Borthakur D. HDFS Architecture Guide. In.; 2008.
13. 360-degree customer view. Hortonworks.
14. Customer 360-Degree View Reduce Customer Churn and Identify New Revenue Opportunities, blueprint for big data success. Hitachi; 2015.
15. Gao RDEaJX. A Review of Crowdsourcing Literature Related to the Manufacturing Industry. Journal of Advanced Science Vol. 4, No. 3, May 2016. 2016 may; 4(3).
16. Amrollahi A. A Process Model for Crowdsourcing: Insights from the Literature on Implementation. In ResearchGate; 2016; griffith university,Australia.
17. Alvin Tarrell NT. Crowdsourcing: A Snapshot of Published Research. In Proceedings of the Nineteenth Americas Conference on Information Systems,; 2013; Chicago, Illinois.
18. Buettner R. A Systematic Literature Review of Crowdsourcing Research. In Proceedings of the 48th Hawaii International Conference on System Sciences -; 2015; Germany.
19. Maryam M Najafabadi FVTMKNRSRWaEM. Deep learning applications and challenges in big data analytics. Journal of Big Data. 2015.
20. Big_data:_changing_the_way_businesses_operate/EY-Insights-on-GRC-Big-data. 2014.
21. <http://www.computerweekly.com/podcast/Big-data-storage-Defining-big-data- and-the-type-of-storage-it-needs>. [Online].
22. https://www.pentaho.com/sites/default/files/uploads/resources/pentaho_customer_360-degree_view.pdf. [Online].
23. Dholakia RRDaN. Scholarly Research in marketing Trends and challenges in the era of marketing. 2012.
24. Sap.com. [Online].
25. Datameer.com. [Online].
26. Zoomdata.com. [Online].
27. Terrdata.com. [Online].
28. The Big Data Revolution Opportunities, Challenges and Vision. FUJITSU Technology Solutions; 2016.
29. Sunil Erevelles NF. Big Data Consumer Analytics and the Transformation of Marketing. Elsevier journal. 2015.
30. Ghalayini DNaM. Develop insight driven customer experiences using big data and advanced analytics. 2017.
31. Vossen G. Big data as the new enabler in business and other intelligence. Springer. 2014.

32. Bhatia DS. Big Data and New Product Innovation. 2015.
33. Big Data vs. Crowdsourcing Ventures - Revolutionizing Business Processes..
34. <https://www.icas.com/ca-today-news/10-companies-using-big-data>. [Online].
35. Konstantin Shvachko HKSRRRC. The Hadoop Distributed File System storageconference.us. In ; 2010; Yahoo! Sunnyvale, California USA.
36. Dawar N. Use Big Data to Create Value for Customers, Not Just Target Them. Harvard Business Journal. 2016.
37. Smith A. Big Data – The Case for Customer Experience. ; 2016.