Exploring Influence of Big Data Analytics in Retail Sector

Dr. Gagandeep Jagdev*

Department of Computer Science, Punjabi University Guru Kashi College, Damdama Sahib, Punjab, India.

*Corresponding Author: *Dr. Gagandeep Jagdev, Department of Computer Science, Punjabi University Guru Kashi College, Damdama Sahib, Punjab, India*

Abstract: *The application of Big Data in different sectors have enabled organizations to extract useful results and plan their future proceedings accordingly. The retail sector has also gained a lot with the applications of Big Data. The revenue of major companies have witnessed exponential growth after applying the concept of Big Data into it. In modern times, the data is money. So companies are investing heavily to obtain as much data as possible in order to gain in sight of customer's preferences and priorities. The author focused on different strategies to be followed by the retail sector companies in order to attract customers. It is important for the companies to know that why some customers like their products and equally important to know why some avoid their products. The paper illustrates the use of Apache Hadoop frame work for mining the structured data having relevance with retail sector.*

Keywords: *Apache Hadoop, Big Data, Customer preferences, Retail sector*

1. INTRODUCTION

The data today is no longer that simple as it looked few decades ago. In today's hi-tech world, the data has completely changed it shape and size. Many organizations surviving in different sectors have opted to handle this huge amount of data for their benefits in terms of economy. Such data which is huge in size, is accumulated at very high rate and have different formats, is referred as Big Data. The need of hour is to fetch out the concealed knowledge from this enormous amount of data. The purpose of extracting knowledge is to utilize it in most effective and efficient manner [1]. There is no limit to this data. It multiplies itself exponentially each day. Several websites comes into existence each day. Millions and billions of images, photos, videos, songs, etc., are uploaded on the internet each day. This data needs to be processed and achieve two primary objectives of description and prediction. Description is intended to accomplish analysis and mining of data in order to extract the latest information according to one's preferences. Prediciton focuses on providing answers to the upcoming questions on the basis of past records [2, 3].

The retail sector benefitted a lot from the analysis and mining of Big Data. Different commercial giants have come with huge setups to monitor the preferences and priorities of their customers [4, 5]. It would not be wrong to declare that data is new money. The more data about customers the company have, the more effectively they can have an impact on the living style of their customers. Each organization harness this data with an intention to gain economic growth giving utmost importance to the customer satisfaction [6, 7]. Getting insight of customer's preferences enables companies to target the customers in effective manner. Knowing about customers dislikes is equally important as it is to gain insight of his/her likes.

Every opportunity brings challenges with itself. There are five challenges faced in handling Big Data mentioned below [8, 9] and illustrated in Fig. 1.
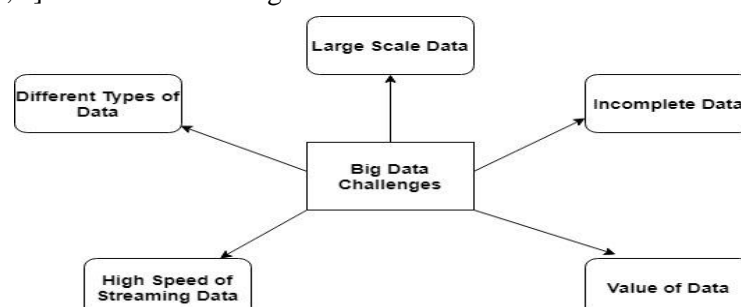

Fig1. *The five challenges faced by Big Dat34a*

- Volume – The terms used to measure data are no longer the same as was a decade ago. Today the data is measured in much higher terms like Coperbytes. The need to handle such huge amount data requires powerful data centers equipped with multiple dominant servers. Coming up with such data centers is not everyone's cup of tea. Only huge giants like Amazon, WalMart, Facebook, etc., can afford establishing such data centers.

- Velocity – The velocity refers to the speed of accumulating and gathering the data. The hardware and software needs to be compatible enough to cope up the speed of accumulation of data from various sources. The web portal "https://www.internetlivestats.com/"illustrates the speed at which data is been generated and transferred over internet.

- Variety – Big Data exists in varied format including text, audio, video, structured, unstructured, and semi-structured data. Because of this vast variety, deciphering of this data further becomes a challenge.

- Veracity – The data available on internet or social media is always doubtful. Until and unless the origin of the data is an authentic source, the big question remains on the preciseness of the data.

- Value – This aspect refers to the concept of how valuable is the data for the companies, organizations, or commercial establishments.

2. INFLUENCE OF BIG DATA IN RETAIL INDUSTRY

Big Data has its deep impact on retail sector. Some prominent aspects influenced are mentioned below.

- The companies needs to maintain a balance between supply and demand of manufactured products. The supply of the products in the market should be strictly as per the demand of that products [10]. Excess supply or restraint supply, both have adverse effect on effective working of company and disturbs the supply chain.

- The application of different algorithms concerned with mining Big Data enables the organizations to predict the future customer inclinations [11].

- It is important for retailers to know that which product helps them to gain maximum profits. The continuous monitoring of the market enables to have deep insight and decide when to higher or lower the price of different products [12]. The factors deciding the variation in the price can be geographical, seasonal, or competitive. Few big giants in the retail sector have even opted to shift to private cloud in order to enhance their overall growth.

- The power of technology have enabled companies to develop applications compatible for mobile devices [13]. The customers don't need to visit market and can check out the product to be bought online at the convenience of their home, workplace etc.

3. RESULTS AND IMPLEMENTATION

This section deals with handling unstructured data which is gathered in the form of tweets. The task have been accomplished by creating a twitter API to collect tweets related to popular mobile manufacturing companies like Oppo, Vivo, and Samsung. The program have been written in Python to collect the tweets and later on these collected tweets are analyzed via another program created using Java and run on Eclipse Oxygen. Some of the collected tweets are mentioned below as sample.

@oppo Dear Oppo Team,

*I have Oppo F11 pro which I purchased in June https://t.co/ifAnwKRWtR January 2020 my front c…
https://t.co/oHgy5muxHw*

Sentiment(polarity=0.0, subjectivity=0.0)

@oppo battery is draining fast after updating color7. So kindly solve this issue soon in f11 pro

Sentiment(polarity=0.4, subjectivity=0.75)

@oppo OPPO F11 PRO . THE CAMERA BEAST. BEST CAMERA PHONE IN LOWEST PRICE WITH POP UP CAMERA. ULTA FULL SCREEN DISPL… https://t.co/Qenmgk98Tx

Sentiment(polarity=0.675, subjectivity=0.42500000000000004)

@oppo @oppomobileindia Unable to reach to customer care / service center for fault in my in-warranty oppo f11 pro phone. Kindly guide.

Sentiment(polarity=0.0, subjectivity=0.5)

@oppo Hello,, am using Oppo F11 Pro on Android 10..

And today out of nowhere,, whenever I touch my screen,, it sho… https://t.co/gYYD9HAQzi

Sentiment(polarity=0.0, subjectivity=0.0)

@oppomobileindia @oppo I'm not getting cell reception in my f11 pro. When are we releasing VoWiFi update?? It's rea… https://t.co/FOBTbqsVuZ

Sentiment(polarity=0.0, subjectivity=0.0)

@oppooppo f11 pro version 10 update

Sentiment(polarity=0.0, subjectivity=0.0)

two beautiful things in single frame..??? ?????????????????

#outofthephone

shot on oppo f11 pro @oppo @OPPOCareIN… https://t.co/3wZ1M1dnCW

Sentiment(polarity=0.3892857142857143, subjectivity=0.6071428571428571)

RT @iamaditya650: @oppomobileindia @oppo

My front camera of Oppo f11 pro stopped working. Pop up camera is working fine but still rear ca…

Sentiment(polarity=0.4166666666666667, subjectivity=0.5)

RT @techno_drift: Vivo X50 Pro coming with Gimbal like camera #Vivo #vivox50Pro @Vivo_India @vivo

Visit our website at https://t.co/6l7qxgf…

Sentiment(polarity=0.0, subjectivity=0.0)

Vivo X50 Pro coming with Gimbal like camera #Vivo #vivox50Pro @Vivo_India @vivo

Visit our website at… https://t.co/OsbqY55wOf

Sentiment(polarity=0.0, subjectivity=0.0)

issport local to vocal. boycott all chinese made product @vivo @oppo @Xiaomi other all product... china… https://t.co/lSzfeuJjtM

Sentiment(polarity=-0.041666666666666664, subjectivity=0.125)

@1_VigNaN_7 @Lenovo @oppo @realmemobiles @vivo @Xiaomi Completely respect your point.

Kindly have a look -… https://t.co/4klKmUfM6R

Sentiment(polarity=0.35, subjectivity=0.65)

@1_VigNaN_7 @GooglePlay @GoogleIndia @oneplus @Xiaomi @vivo @oppo @Lenovo @realmemobiles Get well soon??

Sentiment(polarity=0.0, subjectivity=0.0)

RT @Anjali00851154: #OperationHaftaVasooli

#SaveThem

#AtmaNirbharBharatAbhiyan

#AtmaNirbharBharat

#Lockdown4

For performing the sentiment analysis on the above mentioned tweets, two dictionaries have been created, one comprising of positive words and another comprising of negative tweets. Each word is accompanied with the numerical figure which have been allotted to the word based on the effective intensity of the word. The list of some positive words is shown below.

abound 2abounds 2abundance 2abundant 3accessible 1accessible 1acclaim 2acclaimed 2acclamation 3accolade 4accolades 4accommodative 3accommodative 3accomplish 2accomplished 2accomplishment 2accomplishments 2accurate 3accurately 3achievable 3achievement 3achievements 3achievable 2acumen 4adaptable 4adaptive 3adequate 2adjustable 1admirable 2admirably 4 admiration 2admire 1admirer 3admiring 2admiringly 4

The list of few negative words is mentioned below.

abnormal -2abolish -3abominable -2abominably -3abominate -3abomination -3abort -1aborted -1aborts -2abrade -3abrasive -2abrupt -1abruptly -1abscond -2absence -2absent-minded -2absentee -2absurd -2absurdity -3absurdly -3absurdness -3abuse -2abused -1abuses -1abusive -abysmal -2abysmally -3abyss -3accidental -

4accost -2accursed -3accusation -4accusations -1accuse -2accuses -2accusing -3accusingly -1acerbate - 2acerbic -2acerbically -3ache -1ached -1aches -1achey -1aching -1acrid -2acridly -2acridness -2acrimonious - 2acrimoniously -3acrimony -3adamant -2adamantly -3addict -2addicted -3addicting -1addicts -1admonish - 2admonisher -2admonishingly -2admonishment -2admonition -3adulterate -1adulterated -2adulteration -3 adulterer -2

After constructing the two dictionaries the jave program is executed. Each tweet is analyzed when read. Each tweets individual score is calculated on reading that particular tweet and on the basis of the existence of positive and negative words in it. These individual tweet score is summed up after reading each tweet and the final score is obtained after running the the entire file in which tweets are placed. The overall score decides the overall sentiment. If the score ends up as a positive value, it can be concluded that people's sentiments are on positive side and if final score comes out to be less than 0, then it means that the public sentiment is on negative side.

The executed tweets along with current score and overall score are shown below.

{@oppo Dear Oppo Team,=0.0}

Current Counter score=0.0

{I have Oppo F11 pro which I purchased in June https://t.co/ifAnwKRWtR January 2020 my front c… https://t.co/oHgy5muxHw=0.0}

Current Counter score=0.0

{Sentiment(polarity=0.0, subjectivity=0.0)=0.0}

Current Counter score=0.0

{=0.0}

Current Counter score=0.0

{=0.0}

Current Counter score=0.0

{@oppo battery is draining fast after updating color7. So kindly solve this issue soon in f11 pro=1.0}

Current Counter score=1.0

{Sentiment(polarity=0.4, subjectivity=0.75)=0.0}

Current Counter score=1.0

{=0.0}

Current Counter score=1.0

{=0.0}

Current Counter score=1.0

{@oppo OPPO F11 PRO . THE CAMERA BEAST. BEST CAMERA PHONE IN LOWEST PRICE WITH POP UP CAMERA. ULTA FULL SCREEN DISPL… https://t.co/Qenmgk98Tx=0.0}

Current Counter score=1.0

{Sentiment(polarity=0.675, subjectivity=0.42500000000000004)=0.0}

Current Counter score=1.0

{=0.0}

Current Counter score=1.0

{=0.0}

Current Counter score=1.0

{@oppo @oppomobileindia Unable to reach to customer care / service center for fault in my in-warranty oppo f11 pro phone.
Kindly guide.=3.0}

Current Counter score=4.0

{Sentiment(polarity=0.0, subjectivity=0.5)=0.0}

Current Counter score=4.0

{=0.0}

Current Counter score=4.0

{=0.0}

Current Counter score=4.0

{@oppo Hello,, am using Oppo F11 Pro on Android 10..=0.0}

Current Counter score=4.0

{=0.0}

Current Counter score=4.0

{And today out of nowhere,, whenever I touch my screen,, it sho… https://t.co/gYYD9HAQzi=0.0}

Current Counter score=4.0

{Sentiment(polarity=0.0, subjectivity=0.0)=0.0}

Current Counter score=4.0

.

.

.

.

.

{@SparkNZ @Samsung @SamsungMobile can you enable The FM Radio on the Samsung Galaxy S9+ I need it to listen to a drive in cinema on Thursday.=0.0}

Current Counter score=40.0

{Sentiment(polarity=0.0, subjectivity=0.0)=0.0}

Current Counter score=40.0

{=0.0}

Current Counter score=40.0

{=0.0}

Current Counter score=40.0

{My first @VRChat experience in actual VR!! Thanks to several plug ins, my @Samsung S10, and the easy compatibility… https://t.co/uLaOZwETKz=1.0}

Current Counter score=41.0

{Sentiment(polarity=0.17666666666666667, subjectivity=0.29333333333333333)=0.0}

Current Counter score=41.0

{=0.0}

Current Counter score=41.0

{=0.0}

Current Counter score=41.0

{Hi @SamsungIndia please direct me on how to remove/detach the 'dedicated sink' in your actiwash+ wobble technology… https://t.co/cZqEHPROc2=1.0}

Current Counter score=42.0

{Sentiment(polarity=0.1, subjectivity=0.4)=0.0}

Current Counter score=42.0

{=0.0}

Current Counter score=42.0

{=0.0}

Current Counter score=42.0

{@Samsung @SamsungIndia =0.0}

Current Counter score=42.0

{Air conditioner not working.=0.0}

Current Counter score=42.0

{Technician Samsung is sending can't seem to work out the probl… https://t.co/bf3LACFZKB=0.0}

Current Counter score=42.0

{Sentiment(polarity=0.0, subjectivity=0.0)=0.0}

Current Counter score=42.0

{=0.0}

Current Counter score=42.0

{=0.0}

Current Counter score=42.0

{RT @Samsung: The Electrocardiogram feature on #GalaxyWatchActive2 uses advanced sensor technology on the watch and will enable users to mea…=1.0}

Current Counter score=43.0

{Sentiment(polarity=0.4, subjectivity=0.6)=0.0}

Current Counter score=43.0

{=0.0}

Current Counter score=43.0

{=0.0}

Current Counter score=43.0

{@SamsungMobile @Samsung I'm taking a swing. I have a charity fof fight cancer called @MusiCuresCancer. Would yall b… https://t.co/PzLB3VytZg=-2.0}

Current Counter score=41.0

{Sentiment(polarity=0.0, subjectivity=0.0)=0.0}

Current Counter score=41.0

{=0.0}

Current Counter score=41.0

{=0.0}

Current Counter score=41.0

{RT @UNIT9: Winner of 2020 @TheWebbyAwards in Best Art Direction is Samsung: KaDeWe, Windows Into The Future! Thank you all who appreciated…=0.0}

Current Counter score=41.0

{Sentiment(polarity=0.5, subjectivity=0.2125)=0.0}

Current Counter score=41.0

{=0.0}

Current Counter score=41.0

{=0.0}

Current Counter score=41.0

{RT @Cryptogoldn: Tron is listed as an "ECO partner" for Samsung Developers SDK=0.0}

Current Counter score=41.0

{https://t.co/Dmco4m5UJU=0.0}

Current Counter score=41.0

{=0.0}

Current Counter score=41.0

{Samsung Blockchain SDKs bring devel…=0.0}

Current Counter score=41.0

{Sentiment(polarity=0.0, subjectivity=0.0)=0.0}

Current Counter score=41.0

{=0.0}

Current Counter score=41.0

{=0.0}

Current Counter score=41.0

{RT @Samsung: The #GalaxyZFlip's sweeper technology helps to keep the dust and dirt out of the device's Hideaway Hinge and remains flexible…=0.0}

Current Counter score=41.0

{Sentiment(polarity=0.0, subjectivity=0.0)=0.0}

Current Counter score=41.0

{=0.0}

Current Counter score=41.0

{=0.0}

Current Counter score=41.0

{RT @Samsung: From password logins to blockchain keys, many of us regularly store valuable information on our smartphones. See how #GalaxyS2…=0.0}

Current Counter score=41.0

{Sentiment(polarity=0.25, subjectivity=0.28846153846153844)=0.0}

Current Counter score=41.0

{=0.0}

Current Counter score=41.0

{=0.0}

Current Counter score=41.0

{RT @Samsung: The #GalaxyA21s packs Samsung's most essential mobile innovations into a stunning, holographic design. See here for details on…=0.0}

Current Counter score=41.0

{Sentiment(polarity=0.3333333333333333, subjectivity=0.6)=0.0}

Current Counter score=41.0

{=0.0}

Current Counter score=41.0

{=0.0}

Current Counter score=41.0

{RT @Samsung: Samsung Smart City in Gumi is the first workplace in South Korea that manufactures finished products to earn validation at the…=0.0}

Current Counter score=41.0

{Sentiment(polarity=0.23214285714285715, subjectivity=0.48809523809523814)=0.0}

Current Counter score=41.0

{=0.0}

Current Counter score=41.0

{=0.0}

Current Counter score=41.0

Final Counter score=41.0

Sentiment Analysis ended on a positive note

4. CONCLUSION

The research work conducted in the paper demonstrated how to collect and analyze the tweets in order to perform sentiment analysis on them. The final answer obtained after executing the entire process helps to reach a definite conclusion related to the sentiments of the people. The program is capable of handling large number of tweets. In future, many more such analysis can be conducted on different categories of products related to retail sector like refrigerators, air conditioner's, LEDs, LCDs, etc. The retail companies and even a layman, both can make use of the programming model to find out the sentiment status of particular product in the market.

REFERENCES

[1] Gagandeep Jagdevetal., ― A Comparative study of Conventional Data Mining Algorithms against Map-Reduce Algorithm‖, in International Journal of Advance Research in Science and Engineering (IJARSE), ISSN(O) – 2319-8354, ISSN(P) –2319-8346,Volume–06, Issue–05, May2017.

[2] N. Marzand J. Warren. Big Data: Principles and best practices of scalable real time data systems. Manning Publications, 2013.

[3] Gagandeep Jagdevetal.,― Analyzing Maneuver of Hadoop Framework and Map R Algorithm Proficient in supervising Big Data‖, in International Journal of Advanced Technology in Engineering and Science (IJATES), ISSN–2348-7550,Volume–05, Issue–05, May2017.

[4] R. Smolanand J. Erwitt. The Human Face of Big Data. Sterling Publishing Company Incorporated, 2012.

[5] Jagdev,G.(2018). Augmenting Revenue Growth in Retail Segment via Data Mining. International Journal of Research Studies in Computer Science and Engineering (IJRSCSE), 5(3), pp.1-8. http://dx.doi.org/10.204 31/ 2349-4859.0503001.

[6] Katal,A., Wazid,M., & Goudar, R.H.(2013). Big Data: Issues, Challenges, Tools and Good Practices. IEEE 404-409.

[7] Dr.Gagandeep Jagdev et.al (2017). ― Big Data in Retail Sector–An Evolution that Turned to a Revolution‖, International Journal of Research Studies in Computer Science and Engineering (IJRSCSE), ISSN:2349-4840(P),ISSN:2349-4859(O), Volume4, Issue4, pp.43-52, DOI: http://dx.doi.org/10.20431/2349-4859.040 4006.

[8] Kaisler, S., Armour, F., Espinosa J.A., & Money, W.(2013). Big Data: Issues and Challenges Moving Forward, International Conference on System Sciences (pp.995-1004). Hawaii: IEEE Computer Soceity.

[9] Dr.Gagandeep Jagdevetal.,―A Study of Clustering and Classification Techniques involved in Data Mining‖ , in International Journal of Advanced Technology in Engineering and Science (IJATES), ISSN–2348-75 50,Volume–05,Issue–05,May2017.

[10] Undefined by Data: A Survey of Big Data Definitions, by Jonathan Stuart Ward and Adam Barker, School of Computer Science, University of St. Andrews, UK, 2013. Available at: http://arxiv.org/pdf/ 1309. 5821v1.pdf.

[11] Gagandeep Jagdevetal., ― Association of Big Data with Map-Reduce Technology Augments for Economic Growth in Retail‖, in International Journal of Engineering Technology Science and Research (IJETSR), ISSN:2394-3386, Volume4, Issue2, February2017.

[12] http://mike2.openmethodology.org/; LastAccessed:20th November2018.

[13] https://neilpatel.com/blog/retailers-are-using-big-data/;LastAccessed:20th November2018

AUTHOR'S BIOGRAPHY

Dr. Gagandeep Jagdev, is working in the capacity of a faculty member in Dept. of Computer Science, Punjabi University Guru Kashi College, Damdama Sahib, Punjab, India. His total teaching experience is more than 12 years and has 132 international and national publications in reputed journals and conferences to his credit. He is also a member of editorial board of several international peer reviewed journals and has been active TPC member of several international and national conferences conducted by renowned universities and academic institutions. He has been awarded Best Research Paper award at NCEEITET – 2015 held at Govt. College of Engineering &Technology, Jammu. He has also been awarded with Best Research Paper Award by National Institute of Technical Teachers Training and Research (NITTTR), Chandigarh at International Conference ICCCN – 2017 conducted by Dept. of CSE. His field of expertise is Big Data, Data Mining, Image processing, ANN, Biometrics, RFID, Cloud Computing, Cloud security, Cryptography, MANETs, and VANETs.