A
SEMINAR REPORT

ON

”Big Data Analysis for Customer Behaviour”

OF

T.E.(Computer Engineering)
(Academic Year: 2019-2020)

SUBMITTED BY
Deviprasad Pande
Roll No :- T212019

GUIDED BY
Prof. R.P. More

Department of Computer Engineering

Zeal Education Society’s

Zeal College of Engineering & Research Narhe,
Pune-411041

A SEMINAR REPORT

ON

”Big Data Analysis for Customer Behaviour”

Submitted By

Deviprasad Pande

in partial fulfillment for the award of the degree of

Bachelor of Engineering of
Savitribai Phule Pune University

IN

Computer Engineering

Zeal College Of Engineering and Research,Narhe, Pune

2019 - 2020

Zeal Education Society's

Zeal College of Engineering & Research

Department of Computer Engineering

CERTIFICATE

This is to certify that seminar entitled **"Big Data Analysis for Customer Behaviour"**

have successfully completed by **"Deviprasad Pande"** of TE (Comp.Engg) in the academic year 2019-2020 in partial fulfillment as prescribed by the Savitribai Phule Pune University.

Prof.R.P.More	**Dr. S.M.Sangve**
Seminar Guide	HOD

Place: ZCOER, Pune.
Date: / /2020

Acknowledgement

I take this opportunity to thank my seminar guide **Prof.R.P.More** and Head of the Department **Dr.S.M.Sangve** for their valuable guidance and for providing all the necessary facilities, which were indispensable in the completion of this seminar report. I am also thankful to all the staff members of Computer Engineering the Department for their valuable time, support, comments, suggestions and persuasion. I would also like to thank the institute for providing the required facilities, Internet access and important books.

Deviprasad Pande

Abstract

Information technology in this 21st century is reaching the sky with large-scale of data to be processed and studied to make sense of data where the traditional approach is no more effective. Now, retailers need a 360-degree view of their consumers, without which, they can miss competitive edge of the market. Retailers have to create effective promotions and offers to meet its sales and marketing goals, otherwise they will forgo the major opportunities that the current market offers. Many times it is hard for the retailers to comprehend the market condition since their retail stores are at various geographical locations. Big Data application enables these retail organizations to use prior years data to better forecast and predict the coming years sales. It also enables retailers with valuable and analytical insights, especially determining customers with desired products at desired time in a particular store at different geographical locations. In this paper, we analysed the data sets of worlds largest retailers, Walmart Store to determine the business drivers and predict which departments are affected by the different scenarios (such as temperature, fuel price and holidays) and their impact on sales at stores of different locations. This paper has made use of Scala and Python API of the Spark framework to gain new insights into the consumer behaviours and comprehend Walmarts marketing efforts and their data-driven strategies through visual rep- resentation of the analysed data.

Keywords-Big Data Analytics; Hadoop Distributed File Systems; Apache Spark; MapReduce

Contents

List of Figures

Chapter 1

Introduction

Big data is a field that treats ways to analyse, systematically extract information from, or otherwise deal with data sets that are too large or complex to be dealt with by traditional data- processing application software

In this technological era of large scale data, businesses need to rethink on the modern approaches to better understand the customers to gain a competitive edge in the market. Data is worthless if it cannot be analysed, interpreted and applied in context . In this work, we have used the Walmarts sales data to create business value by understanding customer intent (sentiment analysis) and business analytics. A picture speaks a thousand words and business analytics would help paint a picture through visualization of data to give the retailers insights on their business. With these insights the businesses can make relevant changes to their strategy for the future to maximize profits and success.

The big data application enables retailers to use historical dataset to better observe the supply chain, then a clear picture can be obtained about a particular store whether they are making profit or are under loss. When data is properly analysed, we will start to see the patterns, insights and the big picture of the company. Then the required suitable actions can be applied accordingly. This will help optimize operations and maximize sales and profit.

1.1 Related Work

In 2015, Harsoor Patil worked on forecasting Sales of Walmart Store using big data applications: Hadoop, MapReduce and Hive so that resources are managed efficiently. This paper used the same sales data set that we utilized for analysis, however they forecasted the sales for up coming 39 weeks.

Their strategy included the collection of huge Sales data and transferred on HDFS and performed Map Reduce which later due to enormous data size, proved difficult to draw conclusion.

In 2013, Katal, Wazid, Goudar performed thorough studies about handling a Big Data; their issues, challenges, various tools and good practices. Technical challenges like scalability, fault tolerance, data quality and heterogeneous data processing was also mentioned.

Chapter 2

Literature Survey

2.1 Survey

There are other companies who are constantly rising as well and would give a tough competition in the future if one does not stay to the top of their game. In order to do so, they will need to understand their business trends, the customer needs and manage the resources wisely. In this era when the technologies are reaching out to new levels, Big Data is taking over the traditional method of managing and analyzing data. These technologies are constantly used to understand complex datasets in a matter of time with beautiful visual representations. Through observing the history of the companys datasets, clearer ideas on the sales for the previous years was realized which will be very helpful to the company on its own. Additionally, seasonality trend and randomness and future forecasts will help to analyse sale drops which the companies can avoid by using a more focused and efficient tactics to minimize the sale drop and maximize the profit and remain in competition.

Chapter 3

Concept and Specification

3.1 Hardware and Software

The tools and techniques used for this work includes the collection of Huge Walmart sales datasets stored in CSV format. This paper Apache Spark with a build version of Hadoop leveraging HDFS as a data storage option. Apache Spark is a framework capable of handling both batch and stream processing on the same application at the same time. Our development tools include InteliJ Idea Community Edition and iPython Notebook. InteliJ Idea was integrated with Spark instead of using the traditional Spark shell.

After we configured our environment, our first task was to load the files as spark dataframes. Dataframe is a distributed collection of data organized into named columns which is equivalent to tables in RDMS . The spark dataframe API was designed to make big data processing simple for a wider audience and also it supports distributed data processing in general purpose programing languages like Scala, Python and Java.

3.2 Algorithm

Their strategy included the collection of huge Sales data and transferred on HDFS and performed Map Reduce which later due to enormous data size, proved difficult to draw conclusion. Thus Hive processing was done to calculate average sales feature for all 45 stores and 99 departments. Machine learning algorithm, R programming was used for statistic computing. Henceforth, Holt Winters was used for training dataset provided by Walmart and then sales prediction was done. Subsequently the predicted sales were given graphical representation using Tableau interactive

data visualization.

InteliJ Idea was integrated with Spark instead of using the traditional Spark shell. After we configured our environment, our first task was to load the files as spark dataframes.

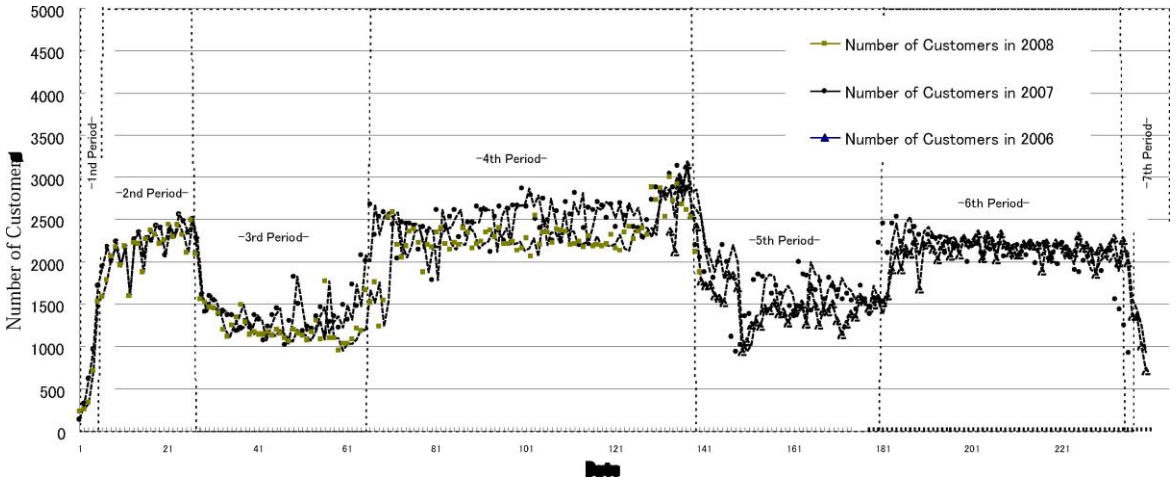

Figure 3.1: The number of customers in 15-minute intervals

Chapter 4

Methodology

Retailers first priority is usually to understand their customers to be able to satisfy their needs so that these customers will return to the store for future needs, thus increasing the product demands and adding to the business value. These businesses want this information to plan where and when to invest profitably.

Simulation is a general methodology designed to be used to evaluate and estimate effects on system performance. In the logistics and service industries, simulation is used as a powerful analyzing and design tool in performance management.

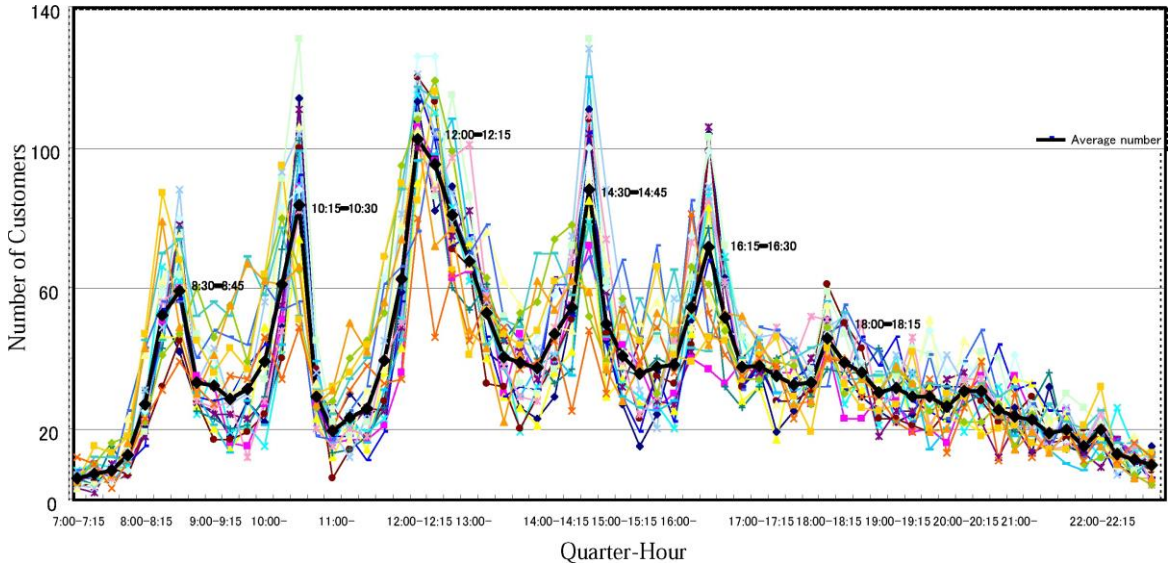

Figure 4.1: The waves of customers in 15-minute intervals

Chapter 5

Conclusion

Wal-Mart is the number one retailer in the USA and it also operates in many other countries all around the world and is moving into new countries as years pass by. There, are other companies who are constantly rising as well and would give Walmart a tough competition in the future if Walmart does not stay to the top of their game.

In order to do so, they will need to understand their business trends, the customer needs and manage the resources wisely. In this era when the technologies are reaching out to new levels, Big Data is taking over the traditional method of managing and analyzing data. These technologies are constantly used to understand complex datasets in a matter of time with beautiful visual representations.

Chapter 6

References

1 M. Franco-Santos and M. Bourne, The impact of performance targets on behaviour: a close look at sales force contexts, Research executive summaries series, vol. 5, 2009.

2 D. Silverman, Interpreting Qualitative Data: Methods for Analyzing Talk, Text and Inter- action 3rd Ed. Text and Interaction, Sage Publications Ltd: Methods for Analyzing Talk, 2006.

3 UBM. (2003) Big Data analytics: Descriptive vs. predictive vs. prescriptive. [Accessed 17 September 2017]. [Online]. Available: www.informationweek.com/about-us/d/d-id/705542

4 A. S. Harsoor and A. Patil, Forecast of sales of walmart store using Big Data application, International Journal of Research in Engineering and Technology, vol. 4, p. 6, June 2015.