

**Ideation Phase**  
**Literature Survey**

Date	19 September 2022
Team ID	PNT2022TMID21900
Project Name	Global Sales Analytics
Maximum Marks	2 Marks

**DATA-DRIVEN SALES LEADS PREDICTION FOR EVERYTHING AS  
A SERVICE IN THE CLOUD - 2022**

(Chul Sung, Bo Zhang, Chunhui Y. Higgins Yoonsuck Choe)

A cloud platform website, offering a catalog of services, operates under a freemium business model or a free trial business model, aggressively marketing to customers who have previously visited.

However, there are several limitations of existing approaches because of the following challenges heavy customer traffic flows, the noise in user behaviors, a lack of collaboration across stakeholders, class imbalanced customer data (few paying customers vs. high numbers of freemium or trial customers), and unpredictable business environments.

In this paper, we propose a data-driven iterative sales lead prediction framework for cloud everything as a service (XaaS), including a cloud platform or software. In this framework, from the BizDevOps process we collaborate to extract business insights from multiple business stakeholders.

From these business insights, we calculate service usage scores using our RFDL (Recency, Frequency, Duration, and Lifetime) analysis and estimate sales lead prediction based on the usage scores in a supervised manner.

Our framework adapts to a continuously changing environment through iterations of the whole process, maintains its performance of sales lead prediction, and finally shares the prediction results to the sales or marketing team effectively.

A three-month pilot implementation of the framework led to more than 300 paying customers and more than \$200K increase in revenue.

We expect our scalable, iterative sales lead prediction approach to be widely applicable to online or cloud business domains where there is a constant flux of customer traffic. Cloud services enable businesses to track and analyze customer behaviors more quickly and intelligently.

However, to deliver the right data-driven business analytics, instead of just collecting data, we need to understand what data we are going to collect and how customers will use our cloud services. To understand consumer behavior properly, an inter - team collaboration among various stakeholders is critical.

The digital marketing organization within the business conducts user behavior analysis, the data science organization executes a data-driven sales lead analysis, and the sales organization performs the sales lead analysis.

# **DATA ANALYSIS AND VISUALIZATION OF SALES DATASET USING POWER BI – 2022**

(Ms. Sarika Singh , Ms. Lavina Jadhav)

Power BI has completely revolutionized the worlds of business intelligence, data visualization, and analytics. Power BI is an online service that allows users to search for data, transform it, visualize it, and share the reports and dashboards they create with other users in the same or different departments/organizations, as well as the general public.

Power BI is used by over 200,000 organizations in 205 countries as of February 2017. Power BI's Quick Insights feature is an innovative feature that is built on a growing set of advanced analytical algorithms.

This function may be activated with a single click after uploading a data set to Power BI, and it generates a number of reports depending on the data's analysis without any need for human participation.

This also aids in reducing human errors in calculations and statistical techniques, which can lead to research that isn't verifiable. Power BI is simple to use and ready for adoption as a platform for Research Data Analysis, accepting even Excel spreadsheets as input.

The pandemic of the corona virus has hastened the adoption of mobile-friendly analytics and business intelligence platforms. Companies all over the world now want their employees and customers to have access to data and analytics from any location. The mobile BI market will be worth more than \$20 billion by 2024, according to Mordor Intelligence.

Furthermore, mobile devices such as smart phones and tablets are expected to account for more than 72 percent of all internet traffic. As a result, implementing mobile-friendly BI platforms across your organization in 2022 makes even more sense from a business perspective.

The purpose of this paper is to show how Power BI can quickly transform a piece of research data into a set of shareable analytical reports and dashboards. We are all always thinking about the future and what is expected to happen in the coming weeks, months, and even years, and in order to do so, a look into the past is required. Businesses must be able to see their development and the factors influencing their sales.

In this technological era of large-scale data, organizations must reconsider current techniques to better understand clients in order to achieve a competitive advantage in the market.

# **IMPLEMENTATION OF BUSINESS INTELLIGENCE FOR SALES**

## **DATA MANAGEMENT USING INTERACTIVE - 2021**

(Ricky Akbar, Meza Silvana, Mohammad Hafiz Hersyah)

Data Management is one of the crucial processes carried out at XYZ Store to get information about the sale of products. In carrying out its operational activities, XYZ Store uses the Smile Invent application to manage data on products sales transactions.

This application has not been able to assist managers in producing the required reports. Therefore, one way to overcome this problem is by implementing the Business Intelligence (BI) application at the XYZ Store by using Interactive Dashboard Visualization.

In implementing the BI application, the BI Roadmap is used as a basis for conducting research starting from the identification of problems to be selected. After that, the planning phase is carried out by evaluating the infrastructure and planning projects.

Then the analysis phase focuses on carrying out a detailed analysis of business problems and opportunities from BI implementation.

Next is the design phase by carrying out the data warehouse design process and ETL using the Pentaho Data Integration (PDI).

Then the implementation phase is carried out, namely the selection and use of BI application tools to perform Data Visualization.

It is hoped that this research can produce reports in the form of Interactive Dashboard Visualization that can be used by store managers to make better decisions.

# **IMPACT OF BIG DATA ANALYTICS ON SALES PERFORMANCE IN PHARMACEUTICAL ORGANIZATIONS: THE ROLE OF CUSTOMER RELATIONSHIP MANAGEMENT CAPABILITIES - 2021**

(Muhammad Shahbaz, Changyuan Gao, Lili Zhai, Fakhar Shahzad, Adeel Luqman, Rimsha Zahid)

In this era of technology development, every business wants to equip its salesforce with a sustainable salesforce automation system to improve sales performance and customer relationship management (CRM) capabilities. This study investigates the impact of big data analytics (BDA) on CRM capabilities and the sales performance of pharmaceutical organizations.

A research model was tested based on 416 valid responses collected from pharmaceutical companies through a structured questionnaire. Structural equation modeling (SEM) was employed using Smart-PLS3 to confirm the contribution of BDA to improving CRM capabilities and sales performance.

The study finds that individual characteristics such as self-efficacy, playfulness, and social norms, along with organizational characteristics such as voluntariness, user involvement, user participation, and management support, are positive predictors of salesforce perception of BDA.

This positive perception of BDA increased the person-technology fit in the salesforce, which ultimately increased the CRM capabilities and sales performance. Introduction Current advances in information technology (IT) and the rising trend of social media have changed the way salespersons perform daily routine activities.

Most often, the salesforce is equipped with a salesforce automation (SFA) system to enhance customer relationship management (CRM) capabilities and sales performance . SFA systems are a set of tools that facilitate organization by providing analyzed information from available data to manage customer relationships and sales-related activities .

An SFA system provides information regarding customer interactions, inventory control, sales forecasting, sales, communication history, and pipeline opportunities to efficiently achieve day-to-day goals .

Organizations annually invest millions of dollars in the implementation of SFA systems to achieve excellent customer relations and sales progress . However, the literature reports that overall, more than 61% of SFA systems fail to meet the current requirements of salesforces.

# **ANALYSIS OF DRUG SALES DATA BASED ON MACHINE LEARNING METHODS - 2020**

(Mohammed A. Al-Gunaid, Maxim V. Shcherbakov, Alla G. Kravets, Vadim I. Loshmanov, Alexandr M. Shumkin, Vladislav V. Trubitsin and Darya V. Vakulenko)

Currently, a data analysis is inalienable part of basic processes of any big company. Building of model of some process allows forecasting its behavior very accurate in specified conditions.

It enables to avoid adverse consequences in case of risks. There are a lot of software packages that implement machine learning methods for forecasting models building exist.

The most common and efficient are linear regression, random forest and artificial neural networks.

The aim of this article is to compare the most known forecasting methods by building data analysis models of medications' sales.

Precise sales forecasting is an important and inexpensive way to increase in profits. In addition, it allows to decrease influence of risks on basic processes inside the company.

According to the results of the built forecast, an opportunity to model the most favorable environment to produce and distribute goods opens. The health of people, who require treatment, depends on timely distribution of medications through sales channels.

Strict limitations on marketing campaigns deprive pharmaceutical distributors of main mechanism of influence on purchasing power of one or the other manufactured and distributed production. In such conditions, it is important to find another ways to increase in profits.

One of that ways is optimization of pharmaceutical preparations distribution. Given approach affects not only an amount of revenue, received by distributors, but also decreases probability of a lack of medications in medical institutions and pharmacies.

The implementation of such like approach is the forecasting of sales taking into account medication distribution influential factors in market conditions.

# **PERFORMANCE ANALYSIS OF SALES BIG DATA PROCESSING USING HADOOP AND HIVE IN CLOUD ENVIRONMENT - 2020**

(Hanindia Prami Swari, Kadek Susila Satwika,Putu Susila Handika)

Nowadays, big data gains much attention from academics and IT industries. This is due to the extraordinary current growth of data that must be accompanied by a variety of qualified data storage and processing techniques to overcome the 5 V's challenge of big data. This research is aimed to conduct a performance analysis of big data processing.

The sales data will be processed in a parallel scheme on the cloud server and then managed using Hadoop and hive. The research shows that the more VMs used, the lower processing time needed, but this is inversely proportional to the CPU time required.

Whereas, from the side of block size testing the research result shows that the decrease in the time of query execution is very visible by the change in the use of block size from 2MB to 4MB and 8MB, but the change in the blocksize size from 4MB to 8MB does not significantly affect the speed of query execution.

The growth of data is increasing rapidly day by day due to the rapidly increasing population, sensors usage, use of social media, the gains on IoT project etc. Many datasets have specific degree of heterogeneity in types, structures, semantics, organizations, granularity, and also accessibility. Data representation is an important part in big data, - and it aims to make data more useful for the analysis with computer and the understanding of users .

Through the presence of IoT, cloud computing and Artificial Intelligence, corporation generates more data than ever before.

The existence of internet and social media contributes to causing situations to disconcert significantly. Every second, approximately 6.000 tweets on average are tweeted on Twitter.

This means that over 350.000 tweets sent every minute, there are 500 million tweets sent in a day and approximately 200 billion tweets sent every year. In almost all organizations, the usage of applications and web servers are generating an overwhelming amount of data logs.

# **PERFORMANCE ANALYSIS OF SALES BIG DATA PROCESSING USING HADOOP AND HIVE IN CLOUD - 2020**

(Hanindia Prami Swari, Kadek Susila Satwika,Putu Susila Handika)

Nowadays, big data gains much attention from academics and IT industries. This is due to the extraordinary current growth of data that must be accompanied by a variety of qualified data storage and processing techniques to overcome the 5 V's challenge of big data. This research is aimed to conduct a performance analysis of big data processing.

The sales data will be processed in a parallel scheme on the cloud server and then managed using Hadoop and hive. The research shows that the more VMs used, the lower processing time needed, but this is inversely proportional to the CPU time required.

Whereas, from the side of block size testing the research result shows that the decrease in the time of query execution is very visible by the change in the use of block size from 2MB to 4MB and 8MB, but the change in the blocksize size from 4MB to 8MB does not significantly affect the speed of query execution. The growth of data is increasing rapidly day by day due to the rapidly increasing population, sensors usage, use of social media, the gains on IoT project etc. Many datasets have specific degree of heterogeneity in types, structures, semantics, organizations, granularity, and also accessibility.

Data representation is an important part in big data, - and it aims to make data more useful for the analysis with computer and the understanding of users.

Since several decades ago, the growth of data generated and stored to meet business needs in the corporate has developed rapidly.

Through the presence of IoT, cloud computing and Artificial Intelligence, corporation generates more data than ever before.

The existence of internet and social media contributes to causing situations to disconcert significantly. Every second, approximately 6.000 tweets on average are tweeted on Twitter.

This means that over 350.000 tweets sent every minute, there are 500 million tweets sent in a day and approximately 200 billion tweets sent every year.



In almost all organizations, the usage of applications and web servers are generating an overwhelming amount of data logs.

Besides, a large number of other systems also contribute to enterprise data growth. Facebook, WhatsApp, and other social media have the same velocity. With such a huge amount of information, corporations need a system that is capable of processing process a huge volume of data and producing valuable insights.

The “big data” construct is not only about the novelty of data and technology, but is also concerning a new frame of mind Big data projects have become an inseparable part of a business. Nevertheless, it does not mean that big data is easy to build, let alone to manage.

The rapid increase in the size of data and the diversity of the data form is a challenge on data acquisition, storage, management and analysis. Traditional database schemes relying on data storage methods with a relational based approach (RDMS) will certainly not be able to overcome this problem.

It can happen because RDBMS is only able to store and manage structured data, whereas if we talk about big data, this will be related to a very large amount of data and heterogeneity of big data.

In addition, RDBMSs are increasingly utilizing more and more expensive hardware Researchers have conducted various studies to find solutions to this problem from various perspectives.

The relationship between big data technology and computing technology is very close, in that the cloud is a solution to the limitations in the provision of physical servers for companies that are not focused on IT.

The main purpose of cloud computing is to allow the use of large computing resources. Data stored and integrated into the cloud computing system will be on one centralized server.

# **RESEARCH ON REFINED SALES MANAGEMENT, DATA ANALYSIS AND FORECASTING UNDER BIG DATA - 2020**

(Wenhui Shan)

This article analyzes the key points of refined sales management under big data. The main points of sales management include how to establish a sales management organization, how to improve the sales management information system, how to improve the evaluation management system, and how to strengthen internal sales control.

Combining the key points of data analysis under big data, the author studies the establishment of data warehouse, data cleaning and mining, the establishment of data prediction models, and the arrangement of model analysis results.

The purpose of this article is to help people give full play to the advantages of big data technology applications and promote the healthy development of the enterprise economy. In the context of the rapid improvement of social and economic levels, the number of products produced by enterprises is increasing, and the frequency of product updates is also rapidly increasing.

However, there are big differences in product quality at the application stage. In the era of diversification, it is difficult to achieve sales growth by relying solely on technological innovation.

After the transformation of the enterprise, the application advantages of big data technology are used to carry out refined management of the entire sales link.

At the same time, doing a good job of data analysis and forecasting can not only reduce the cost of product sales, but also increase product sales and accelerate the economic development of enterprises.

# **DATA ANALYSIS AND VISUALIZATION OF SALES DATA - 2019**

(Kiran Singh, Rakhi Wajgi)

Data is being generated very rapidly due to increase in information in everyday life. Huge amount of data get accumulated from various organizations that is difficult to analyze and exploit.

Data created by an expanding number of sensors in the environment such as traffic cameras and satellites, internet activities on social networking sites, healthcare database, government database, sales data etc., are example of huge data.

Processing, analyzing and communicating this data are a challenge. Online shopping websites get flooded with voluminous amount of sales data every day. Analyzing and visualizing this data for information retrieval is a difficult task.

Therefore a system is required which will effectively analyze and visualize data. This paper focuses on a system which will visualize sales data which will help users in applying intelligence in business, revenue generation, and decision making, managing business operation and tracking progress of tasks.

Data visualization is a process which aims to communicate data effectively and clearly to the user through graphical representation. Effective and efficient data visualization is the key part of the discovery process.

It is the intermediate between the human intuition and quantitative context of the data, thus an essential component of the scientific path from data into knowledge and understanding. It is a powerful new technology having a great potential to help researchers as well as companies for building revenue decision.

# **WALMART'S SALES DATA ANALYSIS- A BIG DATA ANALYTICS PERSPECTIVE - 2019**

(Manpreet Singh, Bhawick Ghutla, Reuben Lilo Jnr, Aesaan F S Mohammed,  
Mahmood A Rashid)

Information technology in this 21st century is reaching the skies 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 year's data to better forecast and predict the coming year's 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 world's 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.

We have made use of Scala and Python API of the Spark framework to gain new insights into the consumer behaviours and comprehend Walmart's marketing efforts and their data-driven strategies through visual representation of the analyzed data.000

We all are constantly thinking about the future and what is expected to happen in the coming weeks, months and even years, and to be able to do so, a look at the past is mandatory. Business needs to be able to see their progress and the factors affecting their sales.

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 Walmart's 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. Most of the raw data, particularly large scale datasets do not offer value in its unprocessed state. By applying the right set of tools, we can pull powerful insights from this stockpile of bits. The main focus here is to read and analyse the Walmart's available datasets to produce insights and the company's overall overview. The retail stores sell products and gain profit from it. There are a lot of subsidiaries of the stores network which are scattered on various geographical locations.

As the network of stores is huge and located at different geographical locations, the company would not fully understand the customer needs and market potentials at these various locations. In this work, we used the gathered store sales datasets of Walmart to understand the factors affecting the sales for example, the un-employment rate, fuel prices, temperature and holidays in the different stores located at different geographical locations so that the resources can be managed wisely to maximize on the returns.

These insights can help retailers comprehend market conditions of the various factors affecting sales for example Easter holiday would induce a spike in sales and retailers can better allocate resources (supply of goods and human resources). Thus, customer demands are observed accordingly based on the above factors.

Moreover, 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.