Major Project Report on

ANALYTICS IN BANKING AND FINANCE INDUSTRY

Submitted By:

Akanksha Gupta

2K18/EMBA/504

Under the Guidance of:

Dr. Shikha N Khera

Assistant Professor

DELHI SCHOOL OF MANAGEMENT

Delhi Technological University

Bawana Road Delhi 110042

CERTIFICATE

This is to certify that Akanksha Gupta, Roll No: 2K18/EMBA/504, student of Masters of Business Administration (Executive 2018-2020) at Delhi Technological University, Delhi has accomplished the project titled "**Analytics in Banking and Finance Industry**" under my guidance and to the best of my knowledge completed the project successfully, for the fulfilment of the course Executive MBA.

Dr. Rajan Yadav	Dr. Shikha N Khera
Head of Department	Asst. Professor
Delhi School of Management	Delhi School of Management
Delhi Technological University	Delhi Technological University

DECLARATION

I, Akanksha Gupta, student of EMBA 2018-2020 of Delhi School of Management, Delhi Technological University, Bawana Road, Delhi-42 declare that Dissertation Report of “Analytics in Banking and Finance Industry” submitted in partial fulfilment of Degree of Masters of Business Administration is an authentic work done by me.

This is to declare that all the work indulged in the completion of this work such as research, data collection, analysis is a profound and honest work of mine. Report is not being submitted to any other University for award of any other Degree, Diploma and Fellowship.

Akanksha Gupta

Place: New Delhi

Date:

ACKNOWLEDGEMENT

This project bears imprint of all those who have directly or indirectly helped and extended their kind support in completing this project. At the time of making this report I express my sincere gratitude to all of them.

I am extremely thankful and obliged to Dr. Shikha N Khera (Project Guide) for practical tips, encouragement to take on challenging assignments and constant guidance since inception, till the completion of the project.

I would also thank Mukul Bapna, Technical Architect, Royal Bank of Scotland, my colleague, for his support and for providing valuable information, which helped me, complete this project successfully.

This project report is a collective effort of all and I sincerely remember and acknowledge all of them for their excellent help and assistance throughout the project.

Akanksha Gupta

EXECUTIVE SUMMARY

One of the most critical IT talks has emerged in the last few years. It is predicted that it produces approx 2.5TB of data worldwide. To determine size of the population, the amount of data exceeding 10 million HD DVDs. The latest data revolution is reaching the developing world in the developed world. In 2016, 7.8 billion international phone calls were registered, and 74% are in developing countries.

The increasing availability of information presents challenges and a lot of opportunities. The biggest challenge is to use this data service and ensuring the privacy. A large part of the most recent data was produced simply because of the interactions with devices phones, internet, shopping, and electronic storage. Individual features can be interpreted from complex algorithms that use this data, which is made possible due to advances in energy analysis.

Analytics is one of the emerging technologies currently in the market that attracts a lot of attention from business, start-ups and media. It has the potential to transform many industries and make processes more democratic, secure, transparent and efficient. With the high volume of data produced every day as a result of record breaking, it is important for every organization to effectively manage security threats and achieve cost efficiency. This is where Analytics, with its promises of providing practical insights from data, is drawing the attention of C-Suite executives. Multiple use cases are also being introduced across the industry as everyone has begun to recognize the disruptive power of this technology. Financial players are the first to take full advantage of this technology even though it is in the nascent stage. Many companies, from a number of non-financial services industries such as telecom Cyber Security, Supply Chain Management, Forecasting, Insurance Industry, Private Transport and Cloud Sharing, Cloud Maintenance, Adult Support, Voting, Management, Energy Management, Real Estate of inventing data analytics that could use cases to properly disrupt their traditional business models or already use their own data analysis cases.

Not surprisingly, the buzzword data analytics of the time attracted the attention of businesses, governments, banks and many people around the world when they saw the beginning of data analysis on the Internet. Also, they are witnessing a shift in the balance of power from the central bodies in the communications and business sectors.

Banks now have an unprecedented opportunity to develop analytical skills to expand their startups. Our hyper-connected world is set to bring a step change in the power of analytics, with nine out of ten of our respondents who believe access to alternative data through open APIs and the rise of the Internet of Things will revolutionize analytics capabilities at banks. Banks can leverage this analytics revolution to transform their customer experience and generate powerful new offers, including customer-related finance and financial management tools designed for someone who adds real value to daily customer lives.

Table of Contents

1. INTRODUCTION

1.1. Introduction to the Topic

Banks record millions of business transactions daily and these entries are real-time in nature. The volume of data generated by banks is not only huge but also real-time in the environment. However, capturing and recording such a large chunk of data is a challenging task for banks. Data analytics help by providing a platform where these transfers can be systematically recorded.

Editing and recording data is useless until there is no plan to use such large data. Therefore, identifying the link between captured data and potential outcomes is a daunting task in today's complex business world. Connectivity can be anything like security & fraud detection, risk management, customer cost analysis & investment pattern, compliance, financial reporting, market segmentation and product customization, etc.

Decades ago, a typical bank customer enters the bank and is greeted by an officer who knows his name, his background and how to use his banking needs. This is a classic model where banks gain and maintain customer trust and use them for a long time. However, circumstances changed. People are often involved in many assignments and travel to different parts of the country. If he someday lives in New Delhi, the next day he may visit Paris in his business assignments. In those cases, it is a challenge for the bank manager to track his interests and where he is to fulfill his needs. Big data provides insight into many areas of individual life including their lifestyle, needs, and preferences for their customers to make it easier for banks to tailor individual services and individual needs.

For a long time, banks have failed miserably to use the information made by their business. Big data has changed the game in changing their business process and made it possible to identify business opportunities and potential threats. In general, banks and financial institutions receive large amounts of data from sources such as log data, referrals, utilities, emails, social media, external feeds, sponsorships, audio, video and other sources.

The introduction of big data in banks has destroyed many business rules and changed the financial services industry. With a huge volume of data connections from countless exchanges, banks are trying to find new business ideas and risk management solutions. Each set of data collected over a period of time tells a unique story and illustrates the fundamental point of the future for a business company to be able to lead this information to reach the competitive edge of the market. Big data analysis can improve the capabilities beyond the risk models used by banks and financial institutions. Big data can also be used in credit management to detect fraudulent signals and the same can be analyzed in real time using artificial intelligence.

The banking and finance industry can't see data analytics on their own. As well as identifying business opportunities, they should identify security threats, fraud detection and possible remedies. In addition, they should try to connect big data across departmental and organizational silas. Many traditional businesses in India have not yet launched their main data operations. The big banks have a limit on making money from those opportunities. Therefore, big data science not only brings new insights to banks, but also enables them to stay a step ahead of the game with advanced technology and analytics tools.

1.2. Industry Profile

Although data analytics can be used in many data processing tasks, payments and fraud reduction are two areas that can be pulled quickly. The banking industry has data that has been used with large gems and unpopular ATMs and credit card details. As banks face increasing pressure to stay profitable, understanding customer needs and choices becomes a key to success. New forms of operational risk management are becoming increasingly accepted by central banks and financial institutions, especially after the Basel II consensus. By utilizing mining data and advanced analytical techniques, banks are better equipped to manage market uncertainty, reduce fraud, and control exposure risk.

According to IBM's Chief Executive Officer Study Guide for IBM 2010, 89 percent of bank managers and financial marketers say the most important thing is to better understand, predict and give customers what they want. Financial metrics and KPIs provide practical steps to summarize your overall bank performance.

But in order to achieve a set of critical success factors that will help banks reach their strategic objectives, they need to go beyond traditional business reporting and sales forecasts. By using data mining and analytical analyzes to extract possible operational insights and inaccessible predictions, banks can obtain information that includes all types of customer behavior, including channel transactions, account opening and closing, automation, fraud and customer movement.

Details about these banking behaviors can be disclosed using multivariate descriptive analytics, as well as predictive analytics, such as credit allocation. Bank analysis, or the use of data mining in banking, can help improve what the banking sector, which understands, acquires and maintains customers. In addition, improvements in risk management, customer insight, risk and fraud enable banks to maintain and grow a more profitable customer base. The importance of these measures is highlighted in the Basel II agreement which clearly emphasizes the need to adopt prudent credit management practices in order to control market uncertainty and reduce the risk of exposure.

While analytics are completely new to the banking world, many banks are preparing for the next analytics push, backed by data load and new tools, advanced technology and risk management, risk management, changing business models, expanding into new markets, re-focusing on customer acquisition are the reasons for many banks look at what modern analytics skills can offer.

Many of the important, recent changes in the banking industry have resulted in a long list of business challenges that the business analysis practice can be set to address. Many financial institutions have quickly recognized and embraced this emerging technology - and are transforming the banking system and providing previously unused banks and financial institutions, weddings and profits. For example, Bank of America Merrill Lynch uses Hadoop technology to manage petabytes of advanced analytics data and new regulatory requirements.

According to Deloitte's research, three business drivers raise the importance of the banking sector analysis:

- **Regulatory reform** – Important laws such as the Dodd-Frank, CARD Act, FATCA (Foreign Account Tax Compliance Act) and Basel III have changed the business environment of banks. Focusing on systematic risk, regulators are pushing banks to better understand the data they have, transform the data into supporting business decisions, and manage risk more effectively. Each request has major changes in data collection, governance and reporting. Over the next several years, regulators will finalize details on the recently enacted law. However, in order to adapt to a radically different regulatory environment, banks must begin to change their business models today.
- **Customer profitability** – Personalized offers are expected to play a big role in attracting and retaining highly profitable customers, but studies show that only a small percentage of banks have strong capabilities in the area. The CARD Act and the Durbin Amendment make it even more important to understand the behavioral economics of each client and find ways to get the wallet in the most profitable categories.
- **Operational efficiency** – Although banks have trimmed a lot of fat over the past few years, still a lot of improvement can be made, including reducing redundant systems, manual tasks and IT costs.

2. LITERATURE REVIEW

In a review, "Fog Entry: Statistics in Learning and Education. EDUCAUSE Review, 46 (5), 30-32" (2011), Nokia, GG, and Long.P describe big data as data points that are larger than standard software tools. capture, store, manage and analyze.

In the report, "TDWI Report for Best Performance: Big Data Analytics (Best Jobs) (pp. 1 - 1). Data Warehouse Institute (TDWI)" (2011), Russom P. writes that for data categorized as big data should have three Vs: Volume, Variety, and Velocity. Many people think big data is just volume, but Russia emphasizes that the other two Vs are important. Big data is not only great, but it varies. and Velocity dictates the speed at which it is produced.One of the reasons that big data stores are that we can produce it very quickly. Russian also explains that volume does not need to refer to terabytes or petabytes. measuring data volume can be the number of files, records, transactions, etc.

In the Gartner.Inc. List, they described big data as high volume prices, high stocks and multimedia data that demanded inexpensive, innovative data processing to gain improved understanding and decision making.

In a conclusion cited in "Big Data Manifestation: A Practical Guide to Transforming Government Business" (2012), the Federal Big Commission of the Tech America Foundation says big data is a term that describes high-speed, complex and diverse data that requires advanced techniques and technologies to enable capture, storage, submit, manage and analyze information.

Ren et al. (2019) recently developed a collection of recent large-scale data definitions for global data sets provided by the International Data Corporation (IDC) (Tien, J.M., 2013). Apart from the large amount of big data, the complex nature of this new data and the difficulty of managing and protecting such data have added to other issues. The idea of big data has been proposed since, so it has become one of the main focus in both the technical and engineering fields (Wang et al., 2016).

LaValle et al. (2011) evaluated the analytical capability of analytics (Big Data AnalyticsC) and described it as the ability to use big data in decision making. A study by Wixom et al. (2013) similarly focused on Big Data AnalyticsC in terms of driving business, recognizing the value of Big Data AnalyticsC in terms of strategy, data management, and human impact by guessing the size of Bigt AnalyticsC data. The study revealed that the introduction of Big Data AnalyticsC leads to increased business value by increasing the speed of decision and allowing the use of big data to spread through business.

Chen et al. (2012) showed that business analytics and related technologies help organizations develop a better understanding of their businesses and markets, while LaValle et al. (2011) have shown that “good decision-making organizations make decisions based on robust analysis at twice the rate of less efficient organizations” (Sharma et al., 2014). Similarly, according to Kiron et al. (2014) Big Data AnalyticsC "is the ability to provide business data using data management, infrastructure (technology) and talent (human resources) to transform a business into a competitor".

A study by Akter et al. (2016) developed a Big Data AnalyticsC strategy based on previous studies that showed the importance of management and technology in a large data center. This study proposed an integrated model for Big Data AnalyticsC and evaluated its impact. Elgendy (2013) has also proposed a framework for Big Plan, Analytics, and Decisions (B-DAD) where major data analysis tools and methods are integrated into the decision-making process.

Elgendy and Elragal (2016) show the decision-making process and how big data analytics can be integrated into it. Using a scientific approach to design, B-DAD can be used to calculate major data and analytics tools at various stages of decision making. Because of this, the additional value gained by integrating big data analytics into the decision-making process can be identified (Elgendy and Elragal, 2014; Elgendy and Elragal, 2016).

Apart from specific challenges, decision-making is supported by advanced technology and tools at each stage of processing and utilizing big data, and the use of big data now plays a major role in making many decisions and predicting domains such as health care, sales, tourism, marketing, finance, and travel (Elgendy and Elragal, 2014).

Use of big data requires decision support, however. The decision-maker should identify the values required and focus on finding the methods, technologies, and tools that allow them to make the best decision; this process then relies on the notion that the decision maker is rational and rational (Wang et al., 2016).

Generally, decision-making occurs at the stage of each major data process, including data storage, data cleansing, data analysis, data visualization, and forecasting. However, it is sometimes difficult to find the right solution for each process, and many technologies and techniques can be used to make decisions in a large data project. Other decision making requires input from many organizations, including data mining, statistics, machine learning, visualization, and social network analysis. Big data tools come in three types of partitioning: batch processing, distributed processing, and hybrid processing tools (Wang, et al., 2016).

Chen et al. (2014) similarly considered the use of big data, opportunities and challenges, and explored many ways to capture big data challenges, such as cloud computing and quantum computing, to evaluate its effectiveness. Wang.et al. (2016) presented an overview of data including four phases: 1 concepts), major data features, and data processing efficiency; (2) state of-the-art techniques for making decisions on big data; (3) the use of big data decisions in the social sciences; and (4) current data challenges and future directions.

3. RESEARCH METHODOLOGY

Research methodology is a very systematic way through which a particular case or problem can be solved. It is a step-by-step , including:

- To explain the problem
- Setting research objectives
- Sources of data
- Data analysis and processing
- Predictive Analytics
- Conclusions and Recommendations
- Research emphasizes scientific thinking and seduction and promotes the development of rational and organizational thinking practices.

<u>Methodology Used:</u>

3.1. Research Objectives

<u>Primary objectives</u>

- To study the different aspects the Data Analytics.
- To study the practical applications of Data Analytics on the Banking & Finance sector.
- To do a practical analysis on one of the Data Analytics Use case in Banking & Finance Industry

<u>Secondary objectives</u>

- To analyse the benefits of Data Analytics in Banking & Finance Industry
- To evaluate the various opportunity where Analytics can be used in the Royal Bank of Scotland.
- To analyse the risks and challenges present in implementing Analytics in any Platform.
- This project aims to identify the Data Analytics development opportunities present in the Organisation which will help to serve the customer more efficiently.
- The use cases can be modified or used in various aspects in the organisation depending on the most viable and feasible option at that time. This research is

carried out and analysed on the basis of the Global market conditions in Finance Industry from January 2020 to May 2020.

3.2. Research Design

The research design will be descriptive followed by partially analytical. The aim of this research is to get an overall view of Data Analytics, its uses and advantages and to identify the potential use cases of analytics in Banking & Finance Industry which can be implemented in the real world. This research used analytical research design where analytical conclusions are required to collect useful insights and where numbers provide a better view and perspective to make important business decisions. It uses statistical tools and techniques to understand the data and therefore offer a better clarification on the subject, and ultimately give a clear picture on the effectiveness and decision to use it based on the impact.

3.3. Data Collection

The research made use of secondary source of data as the data required is confidential for an organisation and cannot be published since it contains Personal Identification Information of several clients. The data has been collected through Kaggle for carrying out this analysis. For this research, data of credit card transactions has been used to identify if a particular transaction qualifies for a probable fraud.

3.4. Scope of Research

This project is based on the analysis done in Banking & Financial Sector. The objective of the research to assess the power of analytics and its emerging practical uses in the Banking & Financial sector. This research aims to identify the potential use cases of Analytics in Banking & Finance Industry which can be implemented in the organisation to serve customer in a more efficient manner. This research includes a creation of proof of concept (POC) for one of the area where analytics can be applied in banking industry.

4. Introduction to Data Analytics

The digital era with its potential and its complexity is the industry and markets facing a large amount of potential data on each purchase. Knowing the importance of the data collected and gaining access to the hidden information creates a new paradigm for this era, which redefines the meaning of organizational power. The power of knowledge leads to organizations aging and reaching goals. Big data analysis compels industries to define, define, predict, provide, and recognize hidden growth opportunities and lead to greater business value. It uses advanced analytical techniques to build knowledge with an increasing amount of data, which will affect the decision-making process in reducing the complexity of the process. It requires sophisticated algorithms and streamlines the process that analyzes and analyzes real-time data and leads to high accuracy analysis. Machine and deep learning share their sophisticated expertise in the process refining the problem approach.

One of the key effects of the digital world is the creation of a large collection of raw data. Managing such valuable assets of different shapes and sizes on an organizational basis' requires the attention of the manager. Big Data has the potential to affect all sectors of society from social to educational and everything in between. As the amount of data increases especially for technology-enabled companies, the issue of managing raw data becomes more important. Dealing with the raw data features such as variety, speed, and large data volume provides advanced tools to overcome the complexity and hidden nature of it. Therefore, big data analysis was proposed for "trial," "simulation," "data analysis," and "monitoring." Machine learning as one of the tools of Big Data Analytics creates a analytical ground for predicting the basis of supervised and unregulated data entry. In fact, a relational relationship has existed between the power of machine learning analytics and data entry; the more precise and accurate the data input, the more effective the analysis. Also, deep readings as a subfield for machine learning are introduced to extract information from hidden data flows.

Big data analysis has the potential to create an effective and efficient value for both the operating system and the strategic plan and plays as a game changer in increasing product productivity.

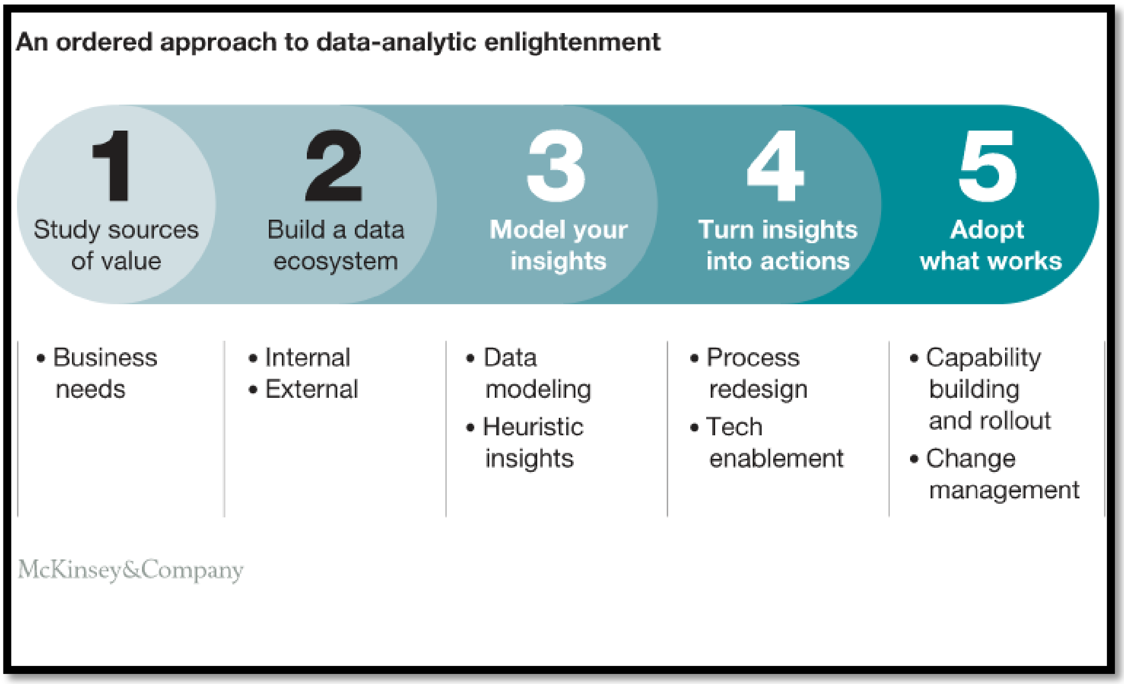


Figure 1: An ordered approach to data- analytic enlightenment

Industry experts believe that big data analytics by the next 'blue sea' brings opportunities for organizations, and is known as the "fourth paradigm".

Machine learning is defined as algorithms that predict data interpretation, followed by algorithm learning in an unnamed program. The three major phases of ML are monitored, monitored, and emphasized in learning, conducted during the "pre-data usage," "learning," and "assessment phase." The earlier configuration is related to the conversion of raw data into a suitable form that can be included in the study section, which includes other levels such as cleaning data, extracting, converting, and merging. In the test phase, data will be used to set up, test performance, test statistics, and measure errors or deviations. This can lead to a change in the selected parameters from the learning process. The first means to analyze the critical aspects to be distinguished from the training data provided. The data included in the training algorithm will then be trained and used in the unregistered data tests. After translating the raw data, a result will be generated, which can be classified as discrete or regression if constant.

ML can be distributed to pattern recognition without a training process, called unsupervised ML. At this stage, when the signal pattern is used for the data entry, an association analysis is performed, and if the hidden data rules have been identified, another ML method will be created, which is a merger. In other words, the primary process for an unsupervised or unified ML is to obtain a group naturally from that information, which is not defined. In this process, the correlation of K by the number of data sets is very similar compared to other clusters that look the same. The three levels of ML that can be controlled are "hierarchical", "fragmented", and "passive" strategies.

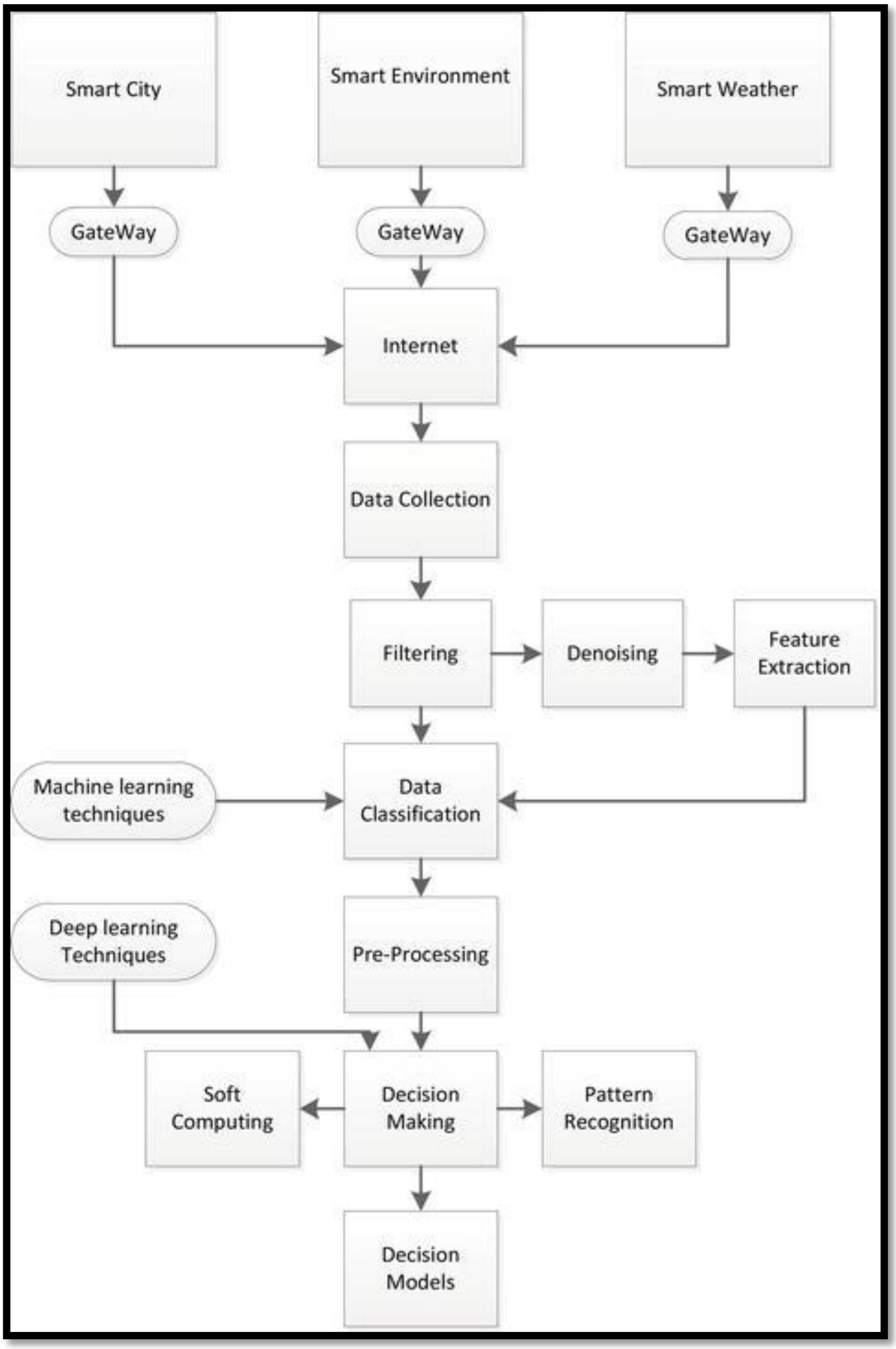


Figure 2: Big data analytics using data received from IOT devices

Data is collected from sensors. Information goes into the filtering process. At this level, an eruption and data purification occur. Also, at this level, the output factor is considered to be the classification stage. After preliminary consideration, decision-making takes place on the basis of a deep learning process.

Deep learning and ML algorithms can be used to analyze data generated by an IoT device, especially in the segmentation and decision-making process.

"Investment in Big Data analysis in the banking sector reached $ 20.8 billion in 2016, according to the 2016 IDC Semi-Annual Big Data and Analytics Spelling Guide. This makes this domain one of the biggest buyers of Big Data services and the ever-hungry market for Big Builders, solutions and bespoke tools."

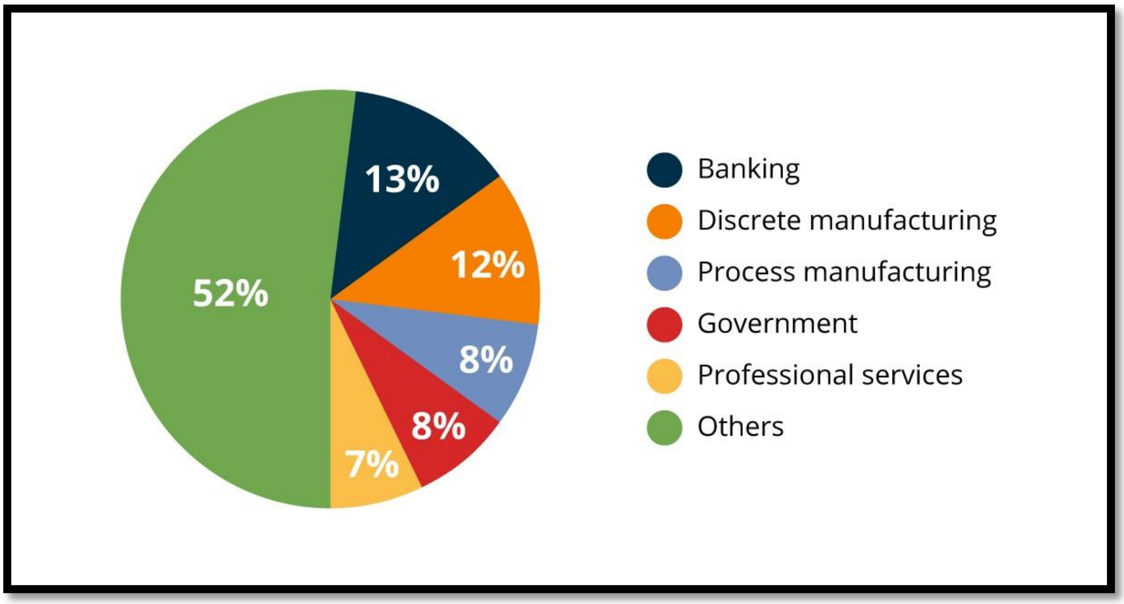


Figure 3: Investment in Big Data Analytics across different sectors
Source: PwC FinTech Report, 2016

"Within this wealth of investments, the allocation of funds is mostly targeted to customer support, risk assessment, decision-making support and researching for new profit opportunities along with investing in new markets, as the PwC Global FinTech Report, published March 2016, showed."

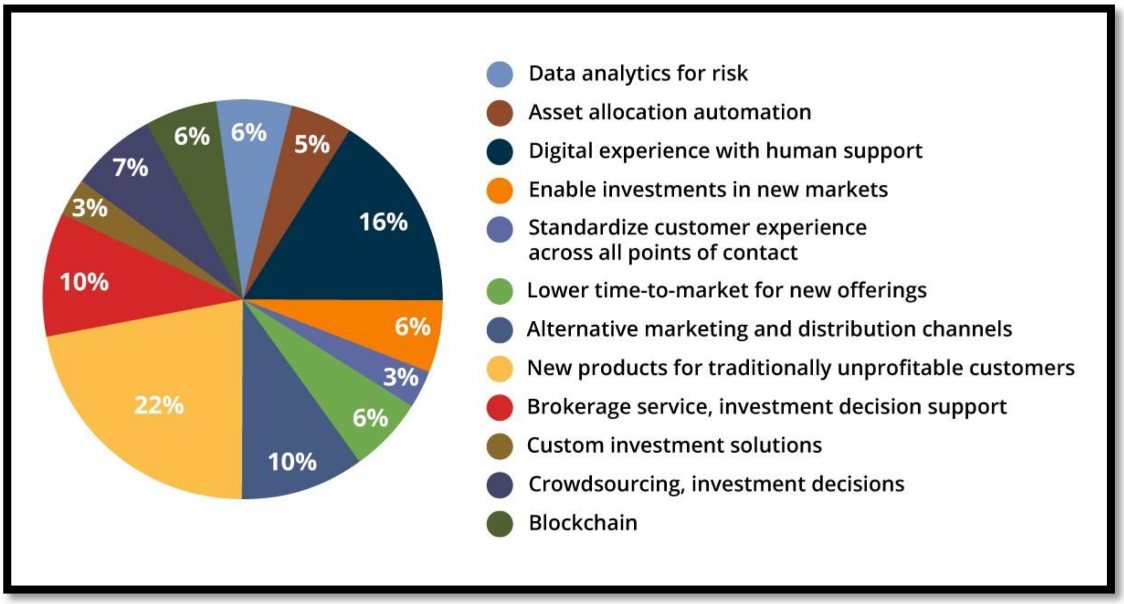


Figure 4: Different areas where data analytics is used in Industry,
Source: PwC FinTech Report, 2016

"The trend is growing and it is even bigger in 2017. The amount of data generated each second is expected to grow by 700% in 2020, according to GDC prognosis. The financial and banking data will stay as one of the major contributors of this Big Data flood, and being able to process it means being competitive among the banking and finance industry."

4.1. Types of Analytics

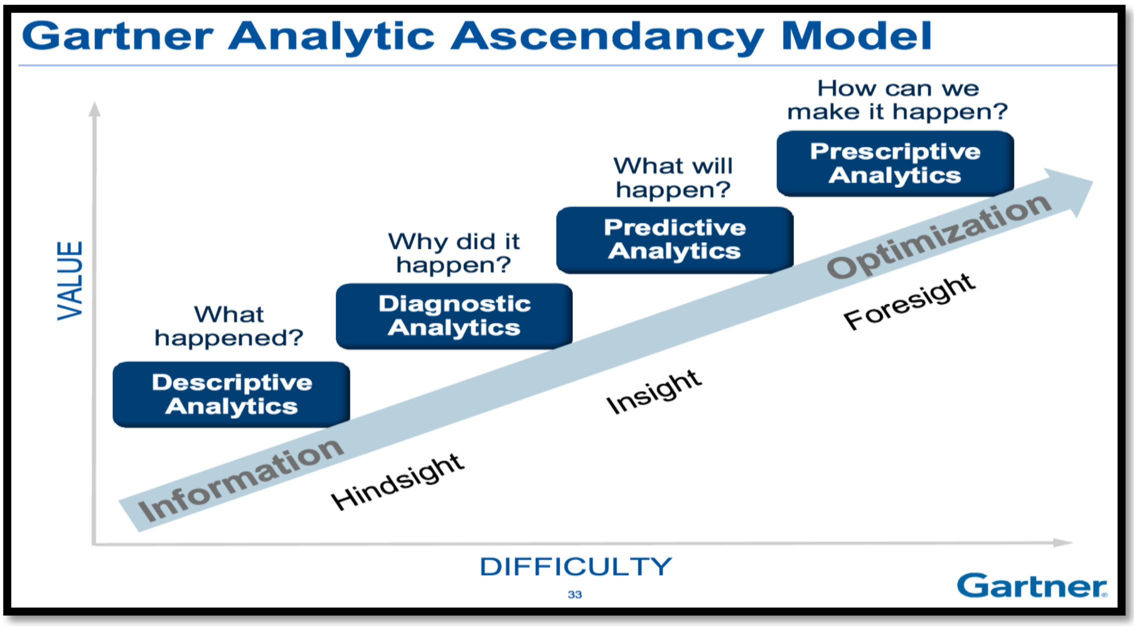


Figure 5: Different Big Data Analytics Stages,
Source: Gartner

- **Descriptive Analytics**

  Descriptive analytics answers the question of *what happened.* It is a preparatory step in data processing that gives a summary of data from past periods to provide useful insights and prepare the data for further analysis.

- **Diagnostic Analytics**

  At this stage, historical data can be measured against other data to answer the question of *why something happened.* Diagnostic analytics gives in-depth insights into a particular problem. At the same time, a company should have detailed information at their disposal; otherwise, data collection may turn out to be individual for every issue and time-consuming. Diagnostic analysis allows companies to identify anomalies, for examples sudden spikes in sales on a given day or torrential changes in website traffic.

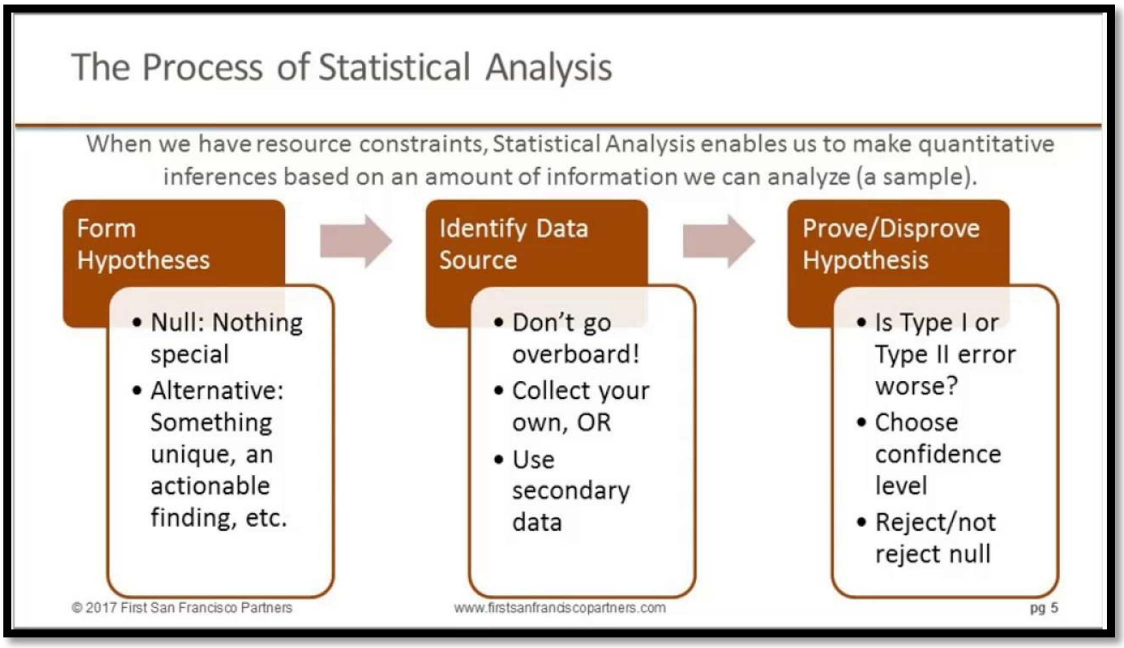
  Here, data analysts need to single out the right data sets to help them explain the anomaly. Searching for the answer often involves drawing information from external sources. When the needed data is on the table, the analysts establish causal relationships and use different types of data analytics (probability theory, regression analysis, filtering, and other) to find the answer.

- **Predictive Analytics**

  Predictive analytics tells *what is likely to happen.* It uses the findings of descriptive and diagnostic analytics to detect clusters and exceptions, and to predict future trends, which makes it a valuable tool for forecasting. Predictive analysis involves advanced tools and technologies and should be based on a big amount of solid data (internal and external) to yield a reliable result. Importantly, achieving an effective forecast depends on a wide array of factors, including the level of volatility of the situation. Predictive analytics also require continuous involvement from the data science team. Predictive Analytics Techniques includes:

  - Statistical Analysis of data
  - Machine learning models

- Data Mining



- **Prescriptive analytics**

The purpose of prescriptive analytics is to literally prescribe *what action to take* to eliminate a future problem or take full advantage of a promising trend. It is a data analysis type that uses advanced technology heavily to find the best solution based on data provided from predictive analytics. Thus, prescriptive analytics would determine what a company could do with a problem or trend foreseen by predictive analytics. Like predictive analytics, prescriptive analysis needs its own business logic and algorithms. As for prescriptive analytics techniques, machine learning is one of the most common.

On one side, prescriptive analytics techniques can be used to gain highly rich insights in customer behaviour across industries. On the other, machine learning algorithms can be trained to analyse stocks markets and automate human decision making by presenting decisions based on large amounts internal and external data. In any case, prescriptive analytics are a costly investment: the investors need to be confident that the analysis yields substantial benefits.

4.2. Benefits of Analytics in Banking and Finance Industry

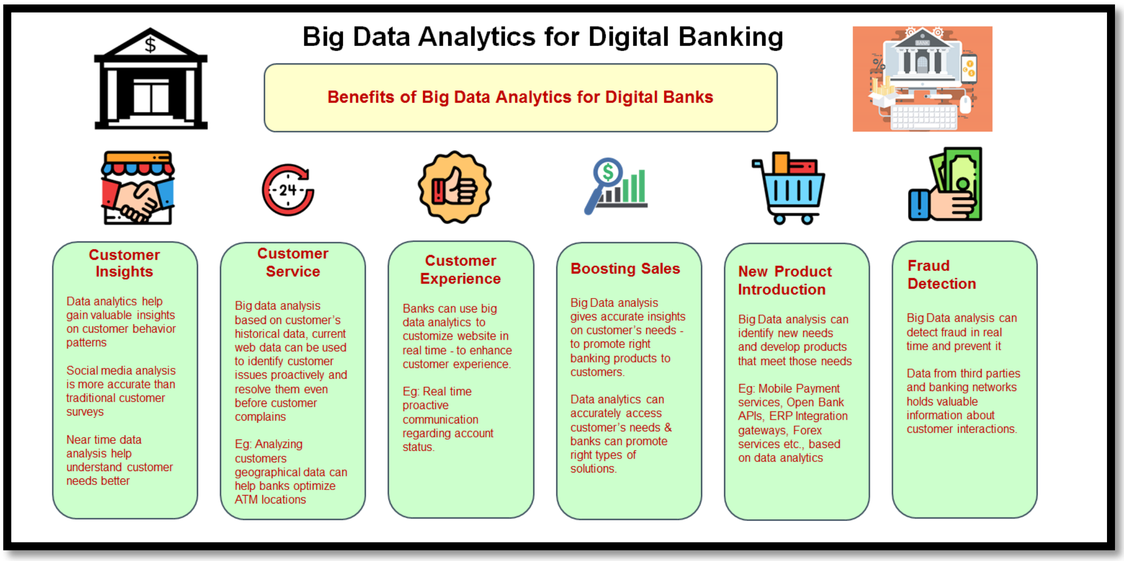


Figure 6: Big Data Analytics in Digital Banking

The benefits that big data analytics can have on financial and banking services companies are:

1. **Customer Profiling**

   Customer segmentation enables the banks and financial institutions to separate their customers into separate categories by demography, but it lacks the granularity these institutions require to understand their customers' needs. Instead, hence institutions should use big data in banking to take segmentation to the next level by building detailed customer profiles. These profiles should include for a variety of factors, including:

   - The demography
   - No of accts they possess
   - What all products they have in present
   - What offers they have declined previously
   - What products they like to purchase
   - Their major life goals and events
   - Relationship with other customer
   - Attitude toward the banking and the finance industry as a whole
   - Behavioural aspects etc.

2. **Tailoring the Customer Experience for each Individual**

   Nearly one-third of customer expects the companies with which they do business to know personal information about them; in fact, m customers abandoned business relationship because of a lack of personalization in the service they received. For all its talk of relationship banking, the financial services industry isn't exactly known for its high level of personalized service. For those banks and credit unions that hope to not just survive, but thrive, a banking analytics-oriented shift in perspective and tailor-made customer experience are absolute necessities.

3. **Understand the buying pattern of Customers**

   Almost all of big data in banking is generated by customers, either through interactions with sales teams and service representatives, or through transactions. Although both forms of customer data have immense value, data generated through transactions offer banks a clear view into their customers' spending habits and, over time, larger behavioural patterns.

4. **Identifying Opportunities: Upselling and Cross-selling of Products**

   Businesses are 60%–70% more likely to sell to existing customers than they are to prospects, which means cross-selling and upselling present easy opportunities for banks to increase their profit share and hence opportunities are made even easier by big data analytics in banking.

5. **Reducing the risk of Fraudulent Behaviour**

   Identity fraud is one of the fastest-growing forms of fraud, with 16.7 million victims in 2017 alone. Monitoring customer spending patterns and identifying unusual behaviour is one way in which banks can leverage big data to prevent fraud and make customers feel more secure.

5. Future of Data Analytics

“In the mid-90s, Bill Gates said that 'banking is necessary, banks are not.’ This idea has deepened among the people over the last decade, with opinion turning against banks after the financial crisis of 2008 and technology opening up a wide range of new options for financial industry.” This has enabled several start-ups to enter the technology sector at an unprecedented rate, causing a huge disruption. In a recent PricewaterhouseCoopers survey of more than 1,300 financial industry executives, “88% said they feared their business was at risk to standalone financial technology companies in areas such as payments, money transfers, and personal finance, and 51% said they believe they could lose as much as 40% of their revenue to standalone FinTech firms”.

However, despite this, banks are still in place, and they are still same as they were twenty years ago. In order to stay relevant, they have worked hard to harness the digital revolution and completely re-imagined their role and the customer experience, often working alongside start-ups to do so.

One of the main benefits is that traditional banks have a large amount of financial information they hold about their millions of customers. They also have a building and a capital to exploit you. Speaking at a recent Google Cloud Next conference, Darryl West, HSBC's Chief Information Officer, explained that, In addition to our $ 2.4 trillion asset value on our valuation page, we have a large asset field [type] of our data. And what has been happening over the past three years is huge growth in the size of our data assets. Our customers are embracing digital channels more aggressively and we collect more information about how our customers interact with us. As a bank, we need to work with partners to allow us to understand what's happening and to present information so that we can do better business and create better customer experiences. '

The power of data analysis is evident in the financial sector. According to the latest World Semi-Annual Big Data and Analytics Spelling Guide from IDC, global revenue for big data and business analytics (Big Data Analytics) will increase from $ 130.1 billion in 2016 to more than $ 203 billion by 2020. And it's a bank. to lead the case, the IDC estimates that the industry spent about $ 17 billion on big data and business solutions in 2016.

Data requests and banking analytics are endless. They can use big personalization data, enabling them to offer products and services made to specific consumers in real time. For example, when buying an international airline or car, the bank sends insurance offers to cover these products. In the future, such applications can be expanded even further. One way this can be done is when you get a big bill, the bank can send a text message as you receive a loan to cover your expenses. The algorithm would calculate which interest rate would be most appropriate based on your historical borrowing patterns and your perception of you as a credit risk, before transferring payment overseas.

The data will also mean that banks can accurately measure the risk of lending to a customer. Predictive analytics models such as the FICO scoring system can analyze consumer credit history, a loan or credit application, and other information to check that a consumer will pay on time. They can also join in structured customer feedback through social media comments and other informal data to create a perfect customer profile, thus reducing the risk around payment.

One of the most important ways banks will be able to use their data in the future is to train machine learning algorithms that can make many processes work. Artificial intelligence (AI) solutions have the potential to change the way banks deal with regulatory compliance issues. According to Rahul Singh, vice president of financial services for IT services provider HCL Technologies, 'We are already facing the use of AI and analytical advance in a clearinghouse where technology can bring down false benefits, allowing for risk-based approaches and risk avoidance. 'A 2015 report from McKinsey & Company revealed that many European banks have already moved from traditional statistical analysis to machine learning, with many showing 10% new product sales and average spending decreased 20% as a result.

Organizations need to reverse this myopic mindset by encouraging the formation of cross-banking analytics team. This, in turn, enables the Analytics Center of Excellence to provide a pan-organizational understanding, while at the same time assessing the capabilities of existing technologies such as machine learning and computer literacy. Similarly, solution providers need to critically evaluate and provide pan-Organization solutions that address the challenges of the organization as a whole.

5.1. Analytics in Financial Sector

Using analytics-driven strategies and tools, banks are able to unlock the potential of big data, and to great effect: "Businesses that are able to quantify their gains from analysing big data reported an average 8% increase in revenue and a 10% reduction in overall costs, according to a 2015 survey from BARC (Business Application Research Center, Germany). To better illustrate just how financial institutions can take advantage of big data and big data analytics in banking, we'll follow the journey of a fictional customer, Dana, who recently opened a primary checking account with America One, a fictional bank."

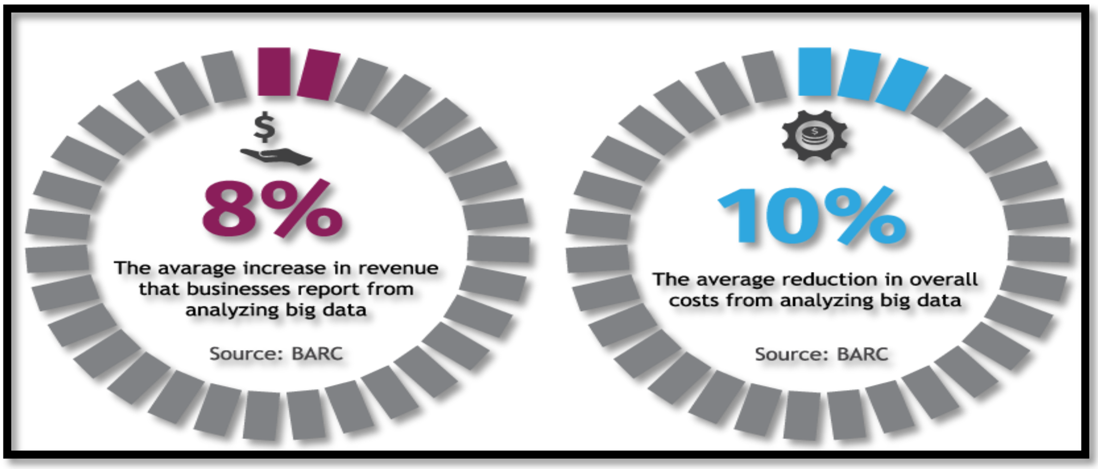


Figure 7: Benefits from Big Data Analytics Implementation
Source: BARC, Germany

The BFSI industry holds the highest market share among all other end-use verticals of Big Data Analytics. Inadequate availability of skilled staffs, lack of business support, and continuous technical problems with database software are barriers to the adoption of big data analytics.

BFSI is a data-driven industry with a massive volume of data which are generated through ATM transactions, cash transactions, account opening, internet banking, online shopping, and others. The need to deliver customized and customer-centric services and offers are driving the demand for big data analytics in the BFSI industry.

Adoption of Big Data Analytics in BFSI sector has helped the professionals to create enhanced user experience, focus on customer acquisition & retention, and also create omni-channel platforms. This industry is highly regulated with several compliance and is more prone to cases of fraud.

Technology is driving the growth of every industry. Similarly, the deployment of Big Data Analytics technology has helped to reduce the instances of possible fraud and related scenarios drastically. It is also playing a significant role in income tax departments to unearth the cases of fraudulent IT claims. Other segments which are benefiting from Big Data Analytics are insurance companies and the banking industry. Insurance companies are investing highly in Big Data Analytics to deal with false claims to assess and underwrite risks. Now, banks can achieve their business goals efficiently, limit customer attrition, and deliver customer-centric services.

With Big Data Analytics, now banks understand their customer's needs better and find the best way to fulfil the same. According to McKinsey, "30% of the bank's works can be automated through this technology, and the key lies in big data analytics."

5.2. Applications of Analytics in Banking & Finance Industry

5.2.1. Customer Data Management

Banks are required to collect, analyse, and store huge amount of data. But viewing this as just compliance exercise, employing machine learning and data science tools can transform this into a possibility to learn more about their clients to drive new revenue opportunities.

In recent times, digital banking has become more popular and is widely used. This generates several terabytes of data; therefore the first step of data scientist's team is to isolate the useful data. Post that, being equipped with information about customer behaviours, interactions, and preferences.

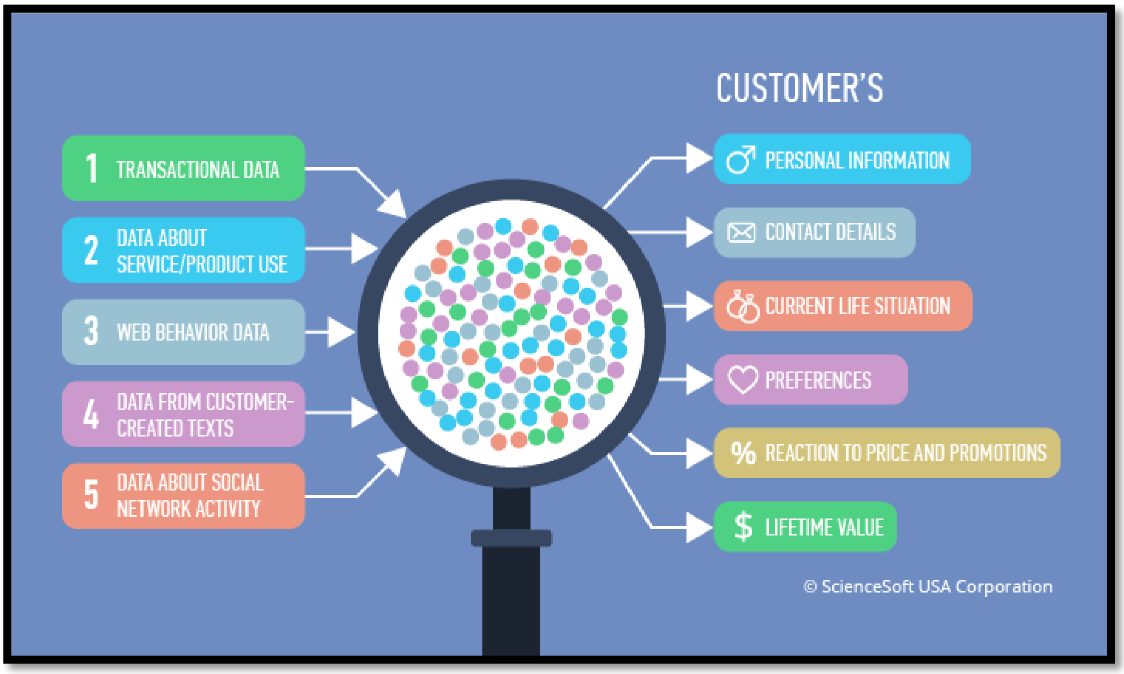


Figure 8: Big Data Analytics for Managing Customer Data
Source: ScienceSoft USA Corporation

Analytics can be employed in the following use cases which will help a bank understand its customer in an efficient manner:

- Deliver customized customer service services, using customer satisfaction data, preferences, purchase history, demographic history, and behavior to better understand their needs. This insight can help you align your products and services and provide highly targeted, personalized service that enhances customer satisfaction and retention.
- Provide the next best product to buy using in-depth data to directly engage customers and prospects into categories based on their profiles and potential needs. Use this information to grow more sales and marketing opportunities, which can be created at the right time through the right channel.
- Provide robo-assist advisory services to assist clients with investment decisions by providing peer-to-peer comparisons or customer-related

portfolio advice. A robo-advisor can manage portfolios without human influence, supporting investment decisions on algorithms developed for risk profiles for clients.

- Create personalized financial management, which gives consumers a complete overview of their finances and provides forward-looking advice. Identify investment opportunities based on customer risk profiles and existing finances, suggest repaying a mortgage, or use previous use data to understand trends and promote better ways to save customers.
- Provide discussion sessions that address customer needs and inquiries, customer journey through process steps, provide predictive messages and behavioral insights, and perform tasks such as cash transfers or balance inquiries. Over time, chat conversations gather behavioral data from users and read relevant responses to user requests.

5.2.2. Risk Modelling

Risk modelling is a crucial area in banking industry. It helps formulating new strategies for assessing the performance. Risk Modelling is one of the important aspects. It allows banks to analyse how the loan will be repaid.

- With Risk Modeling, banks are able to analyze the default rate and come up with strategies to strengthen their lending programs. With the help of Big Data and Data Science, the banking industry is able to analyze and classify lenders before installing loans in a high-risk environment. Ranger Modeling is also applicable to the complete operation of a bank where analytical tools are used to measure the performance of banks and keep track of their performance.
- Some of the scenarios used when analytics are not employed make risk models for the banking industry are:
- Provide early warning forecasts using obligation analysis to identify potential disclosures before default. You can also work with clients to manage their debts and reduce bank exposure.
- Predicting credit risk and recommend effective remedial strategies by identifying fraudulent lenders and identifying self-financing customers.

By understanding this, banks can optimize collection strategies and improve payment rates over time.

- Improve collection and recovery rates. To reduce fraud, credit card issuers can use account pattern recognition technology and develop guidelines for communicating fraudulent account strategies.
- Predict the risks of guessing at individual customers and recommend effective retention strategies to improve customer loyalty. Identify at-risk customers and take immediate action to save them.
- Identify financial crimes such as fraud, money laundering, or terrorism financing activities by identifying transaction inconsistencies or suspicious activities using transactions, customer, blacklist, and country data.

Figure 9: Big Data Analytics for Risk Modelling
Source: Royal Bank of Scotland

5.2.3. Real Time and Predictive Analytics

The growing importance of financial banking analysis cannot be underestimated. Machine learning algorithms and data science techniques can greatly improve the bank's analytics strategy because all bank usage cases are highly correlated with analytics. As the availability and variety of information increases rapidly, analytics are becoming more sophisticated and intuitive.

The potential amount of available data is astounding: the amount of logical data showing actual signals, not just sound, has increased exponentially over the past few years, while the cost and size of data systems have decreased. Separating relevant background information contributes to successful problem-solving and smart strategic decisions. Real-time analytics help to understand the problem that is holding a business, while mathematical analytics help in choosing the right way to solve it. The best results can be achieved by integrating analytics into the flow of banking to avoid potential problems in advance.

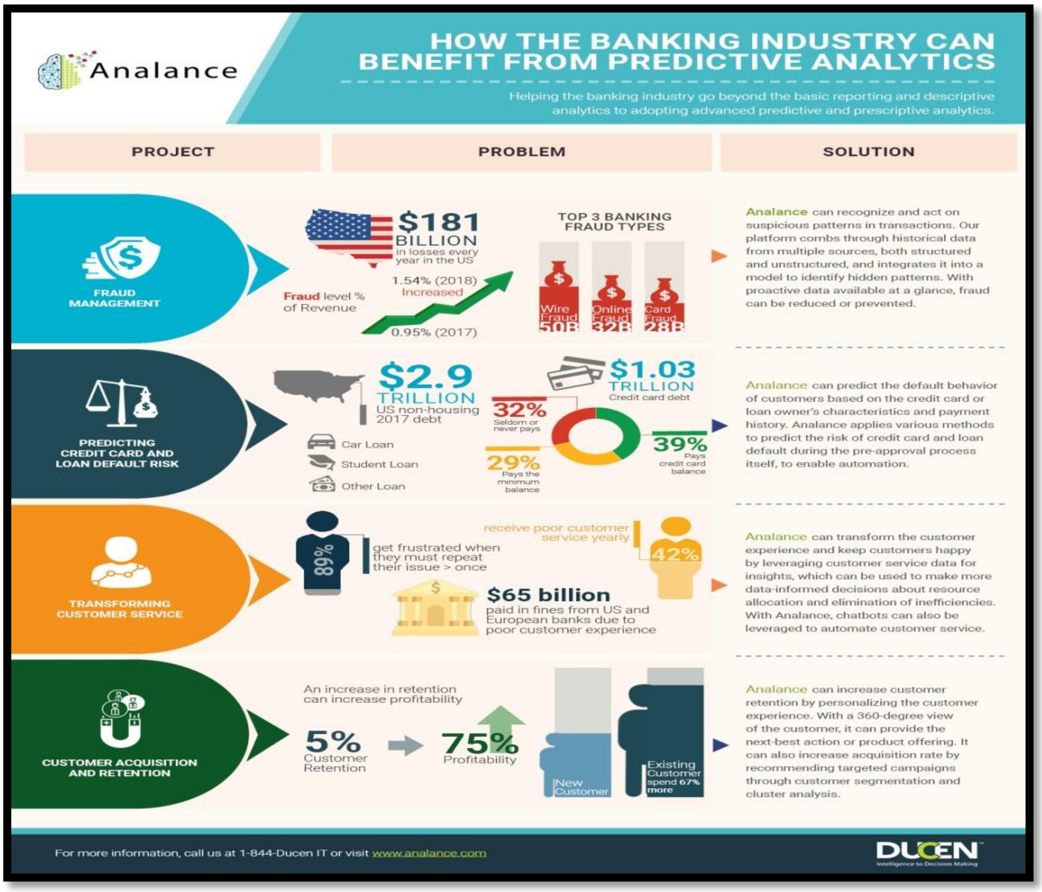


Figure 10: Predictive Analytics in Banking Industry
Source: DUCEN

Some of the use cases used in this area are:

- Algorithmic trading based on deep learning, computer efficiency, and geographical positioning can bring a second-time advantage to automated trading.
- Online customer risk assessment uses the app and customer data in automated credit decisions based on information such as age, income, address, guarantee, loan size, work experience, rate, and transaction history.

- Customer complaints management uses information from various communication channels to understand why customers complain, identify dissatisfied customers, find out the causes of problems, and respond quickly to affected customers.
- Query Feedback uses data from customer engagement channels to automatically submit and answer questions while spending a few resources on crafts.

5.2.4. Fraud Detection

Knowing the normal ways to spend money helps raise a red flag in the event of an upset. If a careful investor who chooses to pay with his card is trying to withdraw all cash from his / her ATM account, this could mean that the card has been stolen and used by fraudsters. A phone call from a bank asking for permission to perform such work helps a simple understanding if it is a legitimate claim or fraudulent cardholder act. Analyzing other types of transactions helps reduce the risk of fraudulent activities.

The fast-growing digital world offers us many benefits but on the other hand, it breeds different types of fraud. Our personal data is more at risk of cyber attacks than ever before and is the biggest challenge that a banking organization faces. Using Big Data Analytics and other Machine Learning Algorithms, organizations are now able to detect deception before it can be implemented. This is done by identifying unusual user spending patterns, predicting unusual user activities, etc.

With advances in machine learning, it has become easier for companies to detect deception and inconsistency in exchange patterns. Fraud detection involves monitoring and analyzing user activity to detect any common or malicious behavior. With the increasing reliance on online and offline transactions, the amount of fraud has increased dramatically.

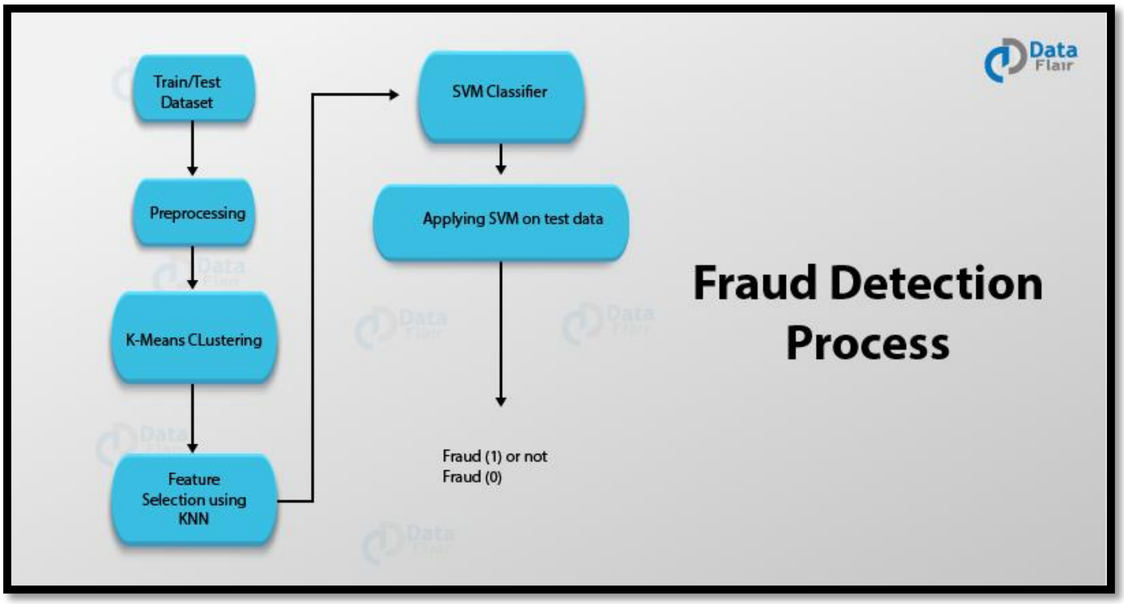


Figure 11: Fraud Detection Using Analytics
Source: DataFlair

Using data science, industries can leverage machine learning and predictive analytics to create interactive tools that will help identify various trends and patterns in the fraud detection system. There are various algorithms such as K integration, SVM methods that are useful for building a platform for detecting patterns of non-exchange activities. The fraud detection process includes -

- Obtaining data samples for model training.
- Training our model in the information provided. The training process involves the use of several machine algorithms for feature selection and additional classification.
- Testing and submitting our model

For example, two algorithms such as K-means integration and SVM can be used for data processing and prediction. K-methods can be used for feature selection and then SVMs are used in the classification data for the class of trick or otherwise..

5.3. Data Analysis: Credit Card Fraud Detection using R

Approach to carry out the Predictive analytics for identify if a credit card transaction is fraudulent or not is specified below:

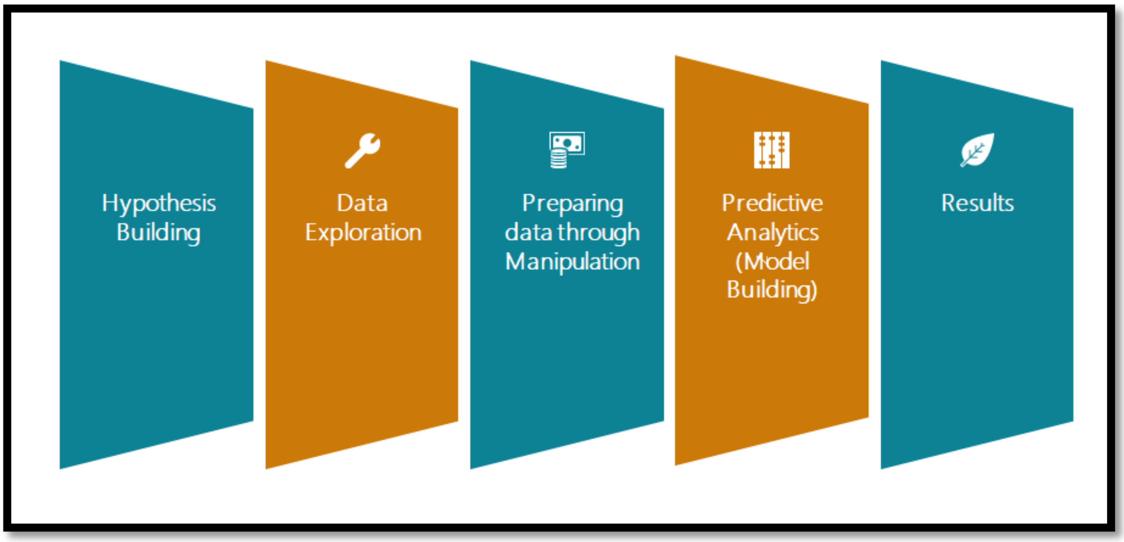


Figure 12: Stages of Data Analytics

1. <u>**Hypothesis Statement**</u>

   The Hypothesis statement to carry out the analysis is:

   Ho: Credit card transaction is fraudulent

   H1: Credit card transaction is not fraudulent

2. <u>**Data Exploration**</u>

   Data exploration involves importing the data and to perform basic operation to understand the important features and the type of data which we have for analysis. It involves gaining some basic insights with the help of RStudio. I will be using RStudio for the analysis related to identifying if a credit card transaction is fraudulent. In this I will explore the data that is contained in the dataframe. Then data will be displayed using the head() function as well as the tail() function.

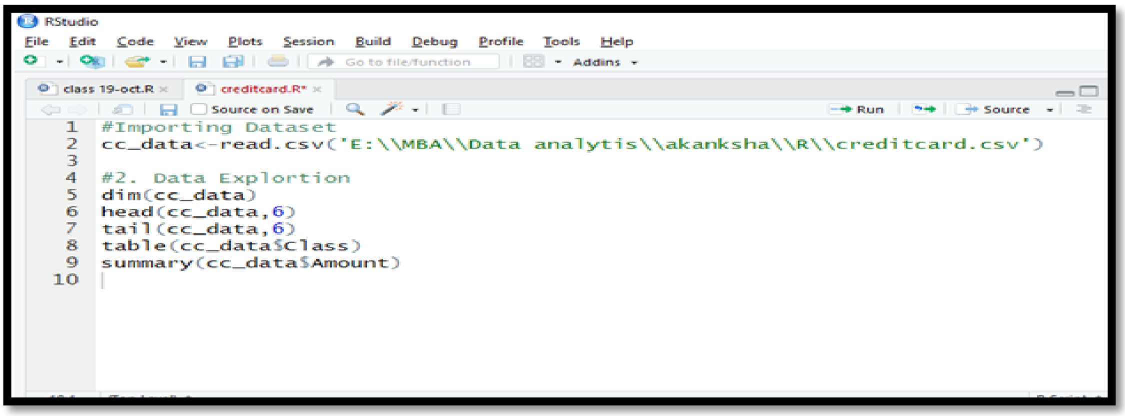


Figure 13: Importing the dataset

Importing the dataset and checking the dimensions:

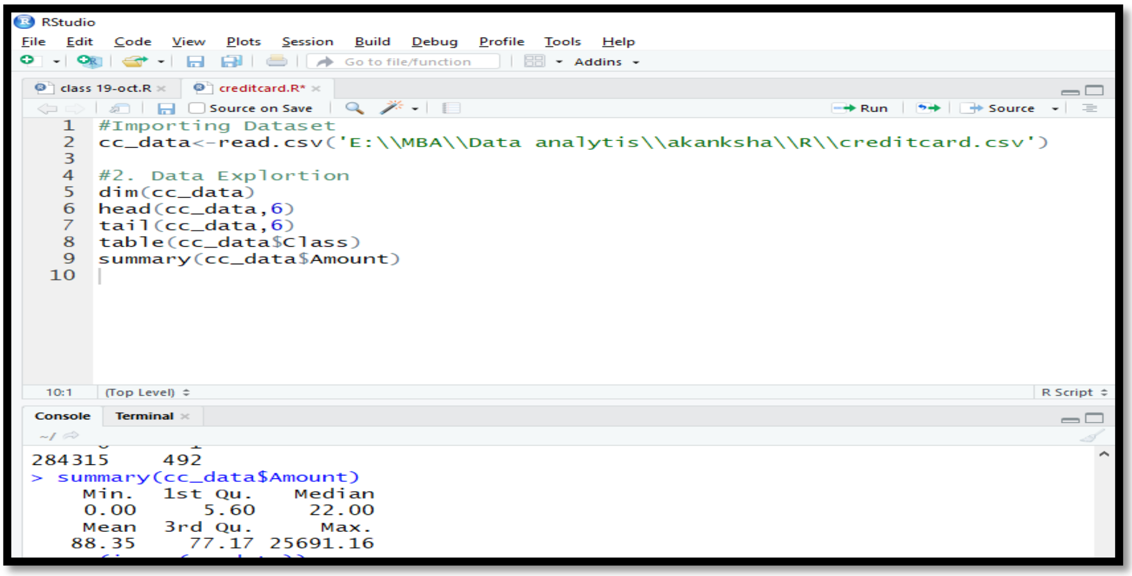
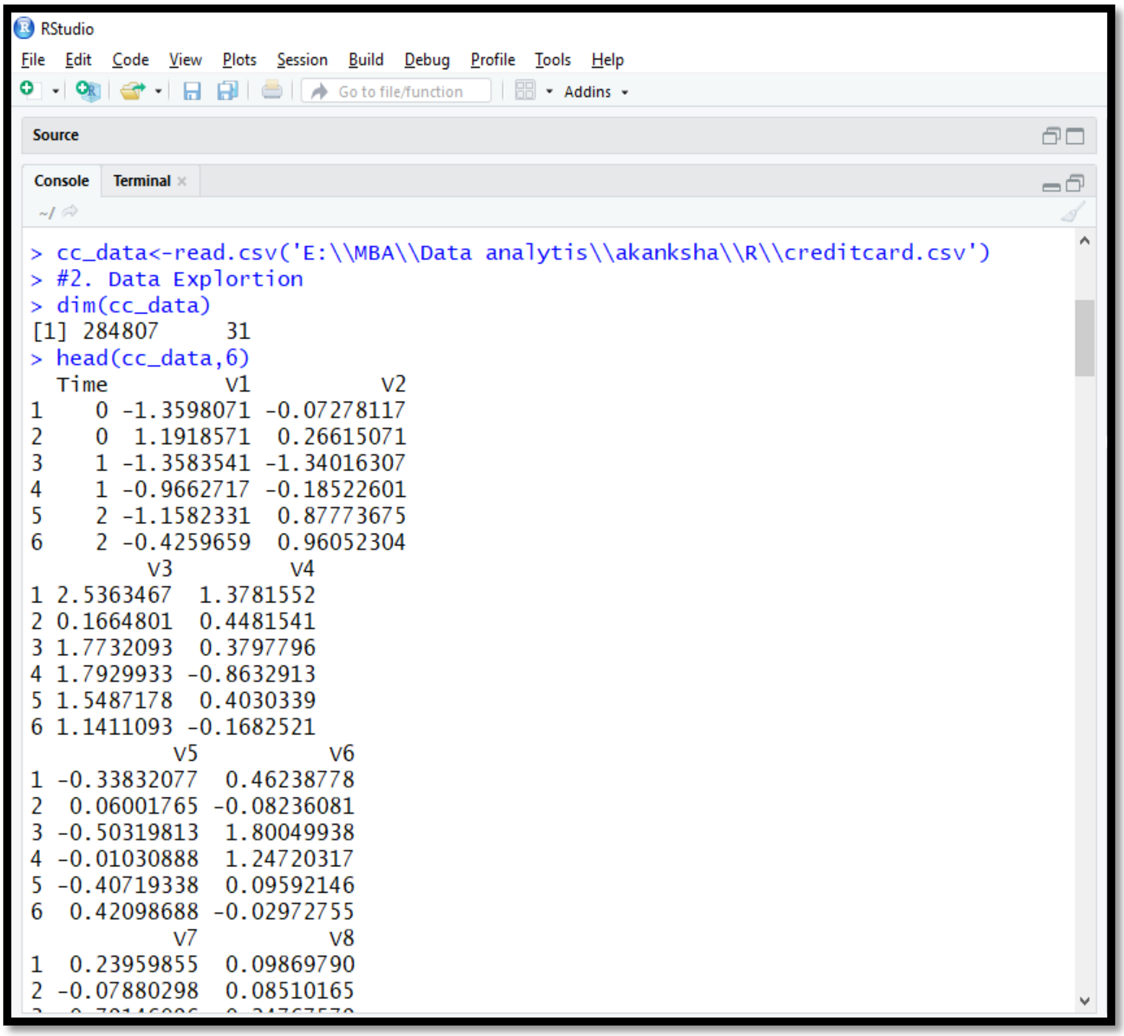


Figure 14: Getting insights from the dataset

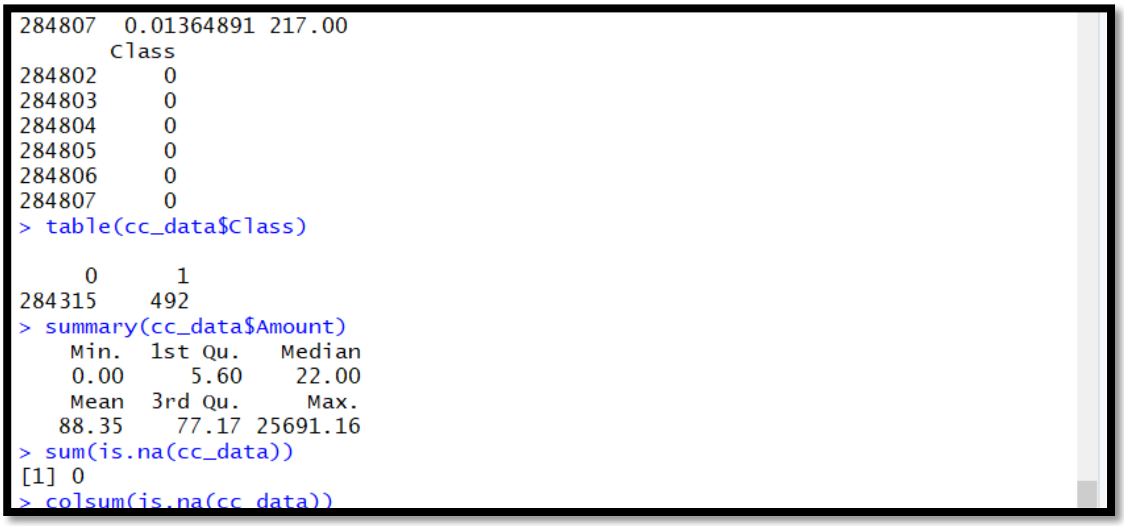
Get basic insights from the data; check the values and type of data in the dataset. Analyzing the different features of the dataset and identify the import features to be used in data modelling.



```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function   Addins
Source
Console   Terminal
~/
> cc_data<-read.csv('E:\\MBA\\Data analytis\\akanksha\\R\\creditcard.csv')
> #2. Data Explortion
> dim(cc_data)
[1] 284807      31
> head(cc_data,6)
  Time         V1          V2
1    0 -1.3598071 -0.07278117
2    0  1.1918571  0.26615071
3    1 -1.3583541 -1.34016307
4    1 -0.9662717 -0.18522601
5    2 -1.1582331  0.87773675
6    2 -0.4259659  0.96052304
         V3         V4
1 2.5363467  1.3781552
2 0.1664801  0.4481541
3 1.7732093  0.3797796
4 1.7929933 -0.8632913
5 1.5487178  0.4030339
6 1.1411093 -0.1682521
           V5          V6
1 -0.33832077  0.46238778
2  0.06001765 -0.08236081
3 -0.50319813  1.80049938
4 -0.01030888  1.24720317
5 -0.40719338  0.09592146
6  0.42098688 -0.02972755
           V7          V8
1  0.23959855  0.09869790
2 -0.07880298  0.08510165
```

Figure 15: Getting insights from the Credit Card dataset

Identifying if the data has any Missing values:



```
284807  0.01364891 217.00
       Class
284802     0
284803     0
284804     0
284805     0
284806     0
284807     0
> table(cc_data$Class)

     0      1
284315    492
> summary(cc_data$Amount)
    Min.  1st Qu.   Median
    0.00     5.60    22.00
    Mean  3rd Qu.     Max.
   88.35    77.17 25691.16
> sum(is.na(cc_data))
[1] 0
> colsum(is.na(cc_data))
```

Figure 16: Field Analysis of Credit Card dataset

3. **Data Manipulation**

This includes scaling various fields in the dataset to make it comparable with other features. Scaling is also known as feature standardization. With the help of scaling, the data is structured according to a specified range. Therefore, there are no extreme values in our dataset that might interfere with the functioning of our model.

```
> #3. Data Manipulation
> cc_data$Amount=scale(cc_data$Amount)
> NewData=cc_data[,-c(1)]
> head(NewData)
          V1          V2        V3         V4          V5          V6
1 -1.3598071 -0.07278117 2.5363467  1.3781552 -0.33832077  0.46238778
2  1.1918571  0.26615071 0.1664801  0.4481541  0.06001765 -0.08236081
3 -1.3583541 -1.34016307 1.7732093  0.3797796 -0.50319813  1.80049938
4 -0.9662717 -0.18522601 1.7929933 -0.8632913 -0.01030888  1.24720317
5 -1.1582331  0.87773675 1.5487178  0.4030339 -0.40719338  0.09592146
6 -0.4259659  0.96052304 1.1411093 -0.1682521  0.42098688 -0.02972755
           V7          V8         V9         V10        V11         V12
1  0.23959855  0.09869790  0.3637870  0.09079417 -0.5515995 -0.61780086
2 -0.07880298  0.08510165 -0.2554251 -0.16697441  1.6127267  1.06523531
3  0.79146096  0.24767579 -1.5146543  0.20764287  0.6245015  0.06608369
4  0.23760894  0.37743587 -1.3870241 -0.05495192 -0.2264873  0.17822823
5  0.59294075 -0.27053268  0.8177393  0.75307443 -0.8228429  0.53819555
6  0.47620095  0.26031433 -0.5686714 -0.37140720  1.3412620  0.35989384
         V13        V14        V15        V16         V17         V18
1 -0.9913898 -0.3111694  1.4681770 -0.4704005  0.20797124  0.02579058
2  0.4890950 -0.1437723  0.6355581  0.4639170 -0.11480466 -0.18336127
3  0.7172927 -0.1659459  2.3458649 -2.8900832  1.10996938 -0.12135931
4  0.5077569 -0.2879237 -0.6314181 -1.0596472 -0.68409279  1.96577500
5  1.3458516 -1.1196698  0.1751211 -0.4514492 -0.23703324 -0.03819479
6 -0.3580907 -0.1371337  0.5176168  0.4017259 -0.05813282  0.06865315
          V19         V20          V21          V22         V23         V24
1  0.40399296  0.25141210 -0.018306778  0.277837576 -0.11047391  0.06692807
```

Figure 17: Scaling the Amount field of credit card dataset

4. **Data Modelling**

After standardizing the dataset, I have split the dataset into training set as well as test set with a split ratio of 0.80. This means that 80% of our data will be attributed to the train_data whereas 20% will be attributed to the test data. Post this dimensions will be found using the dim() function.

```
> library(caTools)
> #4. Predictive Analytics( Building a Model)
> data_sample = sample.split(NewData$Class,SplitRatio=0.80)
> train_data = subset(NewData,data_sample==TRUE)
> test_data = subset(NewData,data_sample==FALSE)
> dim(train_data)
[1] 227846     30
> dim(test_data)
[1] 56961    30
```

Figure 18: Dividing the dataset into Training and test to apply model

Once the dataset is divided into train and test, I have applied the logistic regression model on the train dataset. A logistic regression is used for modelling the outcome probability of a class such as pass/fail, positive/negative and in this case – fraud/not fraud.

```
Source
Console   Terminal ×
~/
Error in eval(family$initialize) : y values must be 0 <= y <= 1
> Logreg<-glm(Class~.,test_data,family = binomial())
Warning message:
glm.fit: fitted probabilities numerically 0 or 1 occurred
> summary(Logreg)

Call:
glm(formula = Class ~ ., family = binomial(), data = test_data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.6608  -0.0266  -0.0164  -0.0100   4.3378

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -8.941249   0.355625 -25.142  < 2e-16 ***
V1           0.125299   0.091146   1.375 0.169224
V2           0.004881   0.138043   0.035 0.971796
V3          -0.103212   0.094658  -1.090 0.275547
V4           0.763008   0.155609   4.903 9.42e-07 ***
V5           0.234373   0.131867   1.777 0.075512 .
V6          -0.001626   0.135301  -0.012 0.990412
V7          -0.086595   0.140937  -0.614 0.538932
V8          -0.159724   0.104492  -1.529 0.126369
V9          -0.284515   0.235831  -1.206 0.227651
V10         -0.787979   0.209503  -3.761 0.000169 ***
```

Figure 19: Applying Logistic regression on the credit card Test dataset

```
V18         -0.265554   0.281569  -0.943 0.345619
V19          0.022382   0.205383   0.109 0.913221
V20         -0.527764   0.175266  -3.011 0.002602 **
V21          0.425960   0.162726   2.618 0.008854 **
V22          0.678402   0.287003   2.364 0.018091 *
V23         -0.065774   0.132998  -0.495 0.620921
V24          0.426365   0.298192   1.430 0.152764
V25          0.171485   0.316230   0.542 0.587625
V26          0.257849   0.420062   0.614 0.539323
V27         -1.014157   0.235065  -4.314 1.60e-05 ***
V28         -0.429252   0.220194  -1.949 0.051244 .
Amount       0.293836   0.230031   1.277 0.201470
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1443.40  on 56960  degrees of freedom
Residual deviance:  437.75  on 56931  degrees of freedom
AIC: 497.75

Number of Fisher Scoring iterations: 12

>
```

Figure 20: Summarising the results of Logistic regression

Once the model is summarised, plot will be visualised:

```r

library(scales)
library(caTools)
#4. Predictive Analytics( Building a Model)
data_sample = sample.split(NewData$Class,SplitRatio=0.80)
train_data = subset(NewData,data_sample==TRUE)
test_data = subset(NewData,data_sample==FALSE)
dim(train_data)
dim(test_data)

#Logistic reg Building Model

Logreg<-glm(Class~.,test_data,family = binomial())
summary(Logreg)
plot(Logreg)

return

library(pROC)
lr.predict <- predict(Logreg,train_data, probability = TRUE)
```

Figure 21: Summarising the results of Logistic regression

Output Screenshots:

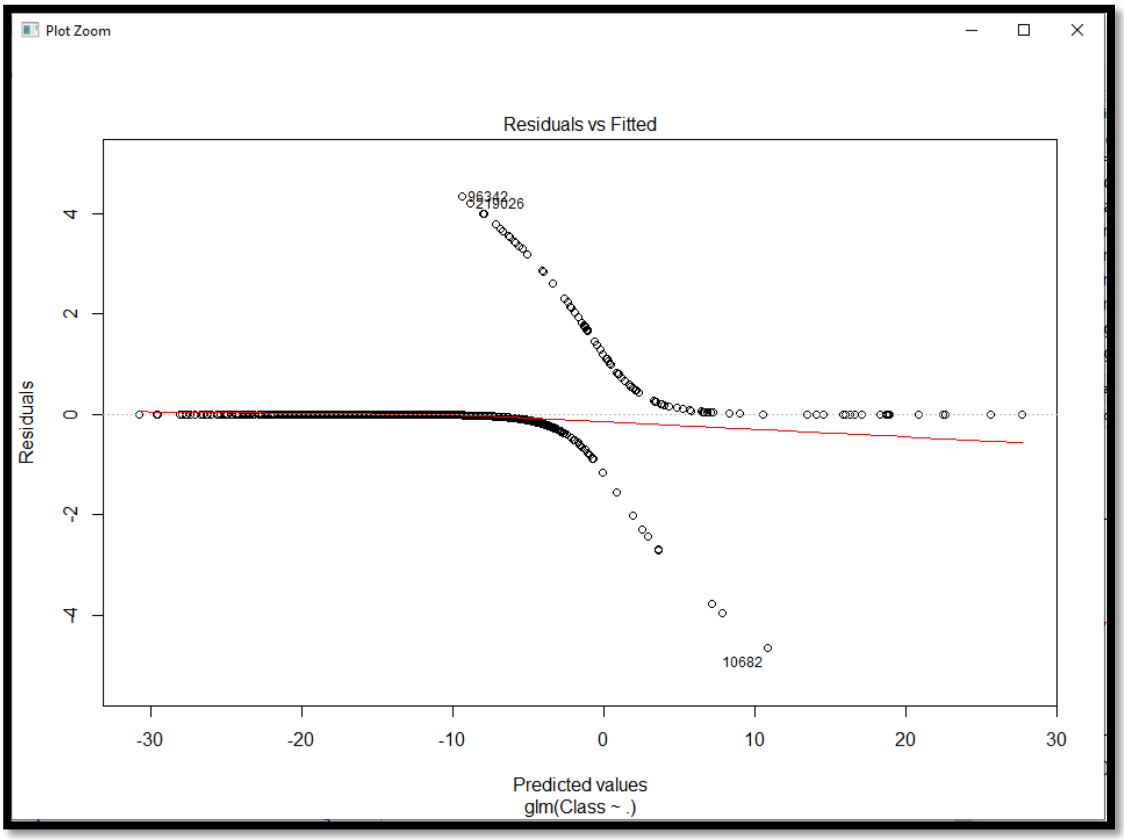


Figure 22: Residual Vs Fitted Plot

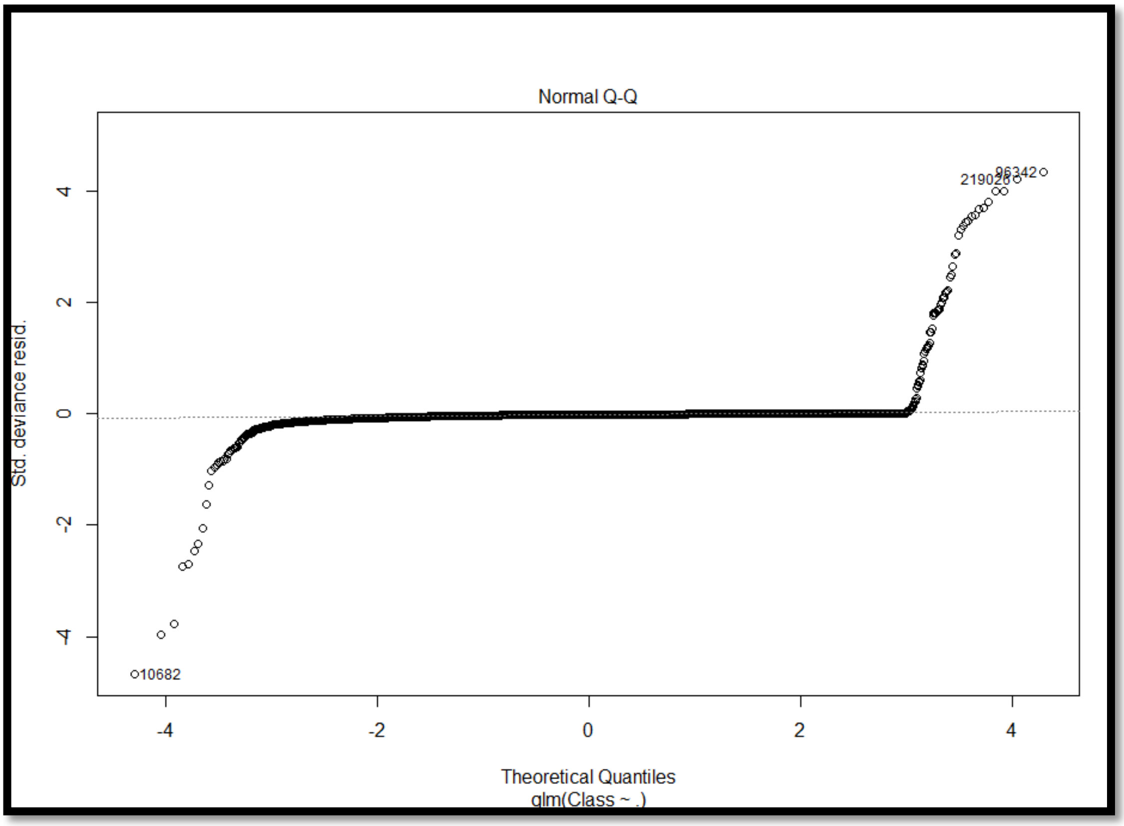


Figure 23: Normal Q-Q plot

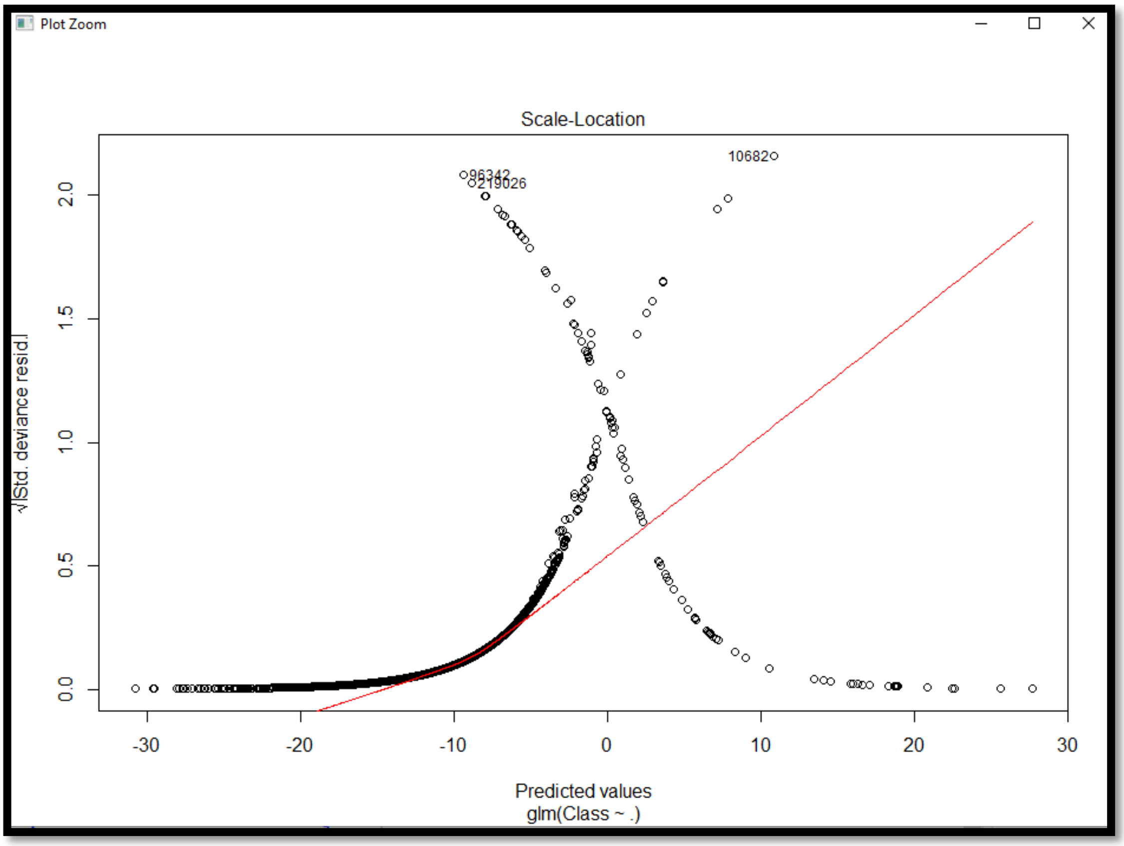


Figure 24: Scale Location Plot

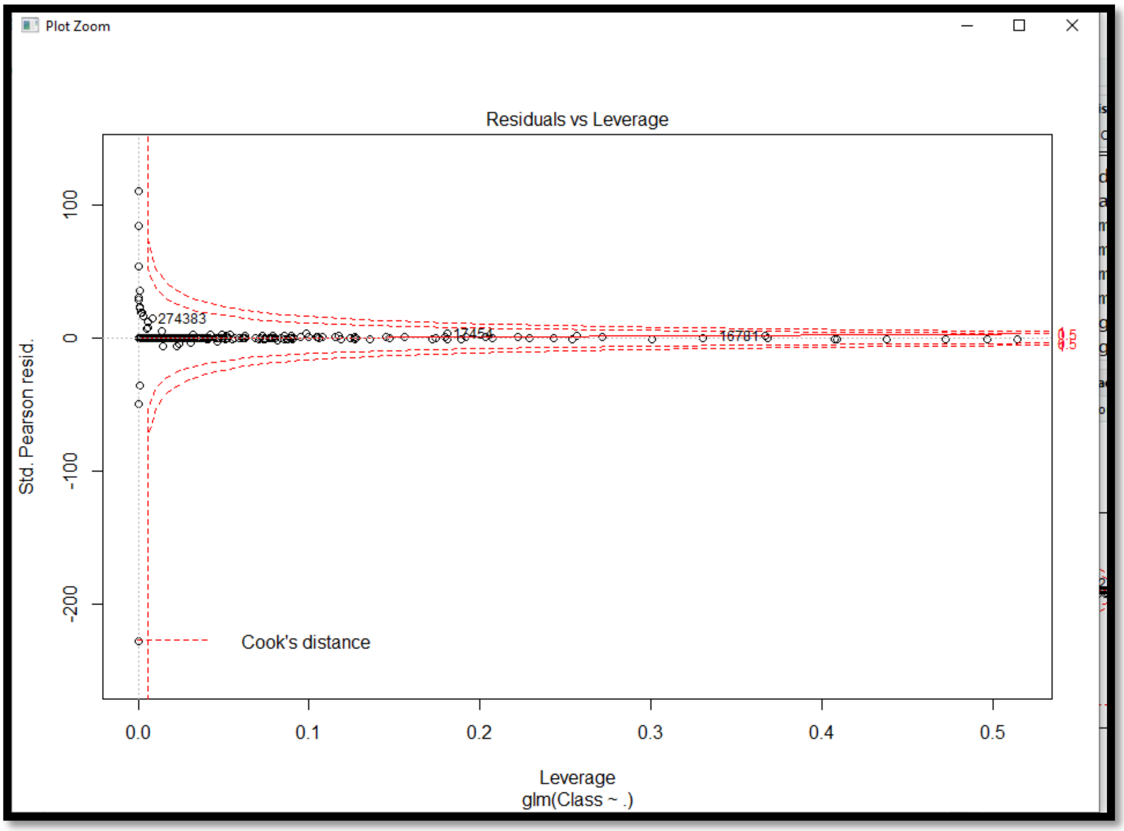


Figure 25: Residual Vs Leverage Plot

In order to assess the performance of model, delineating the ROC curve is required. ROC is also known as Receiver Optimistic Characteristics. For this, plotting ROC curve will help to analyse the performance of model.

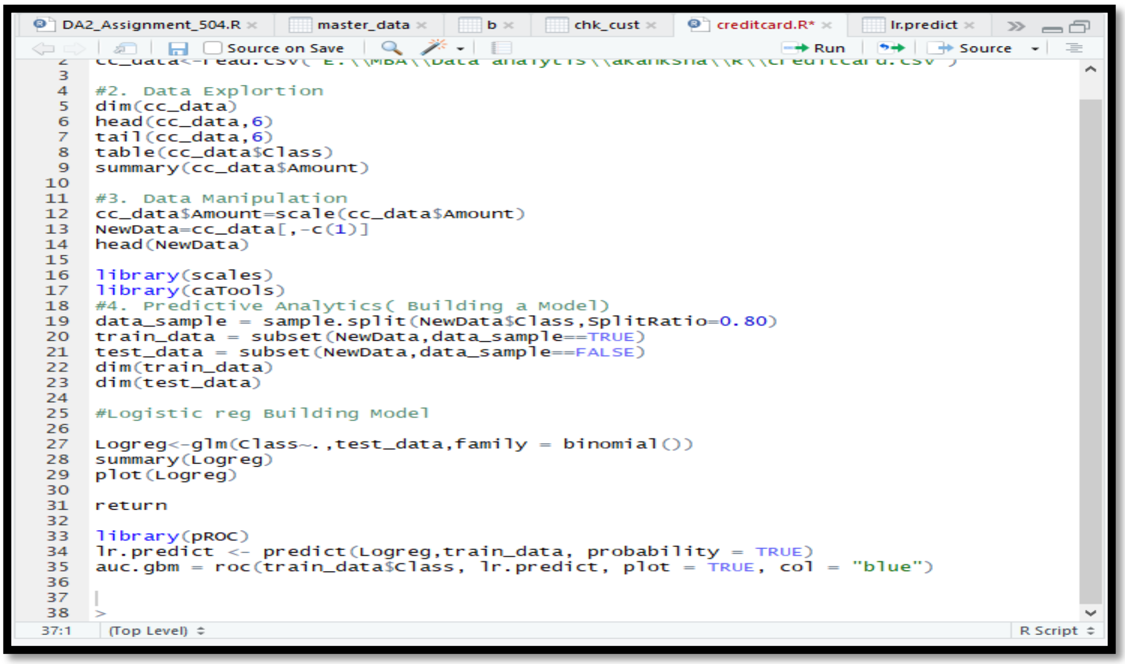


Figure 26: Plotting ROC curve

Screenshot for ROC Curve of the Logistic regression Model:

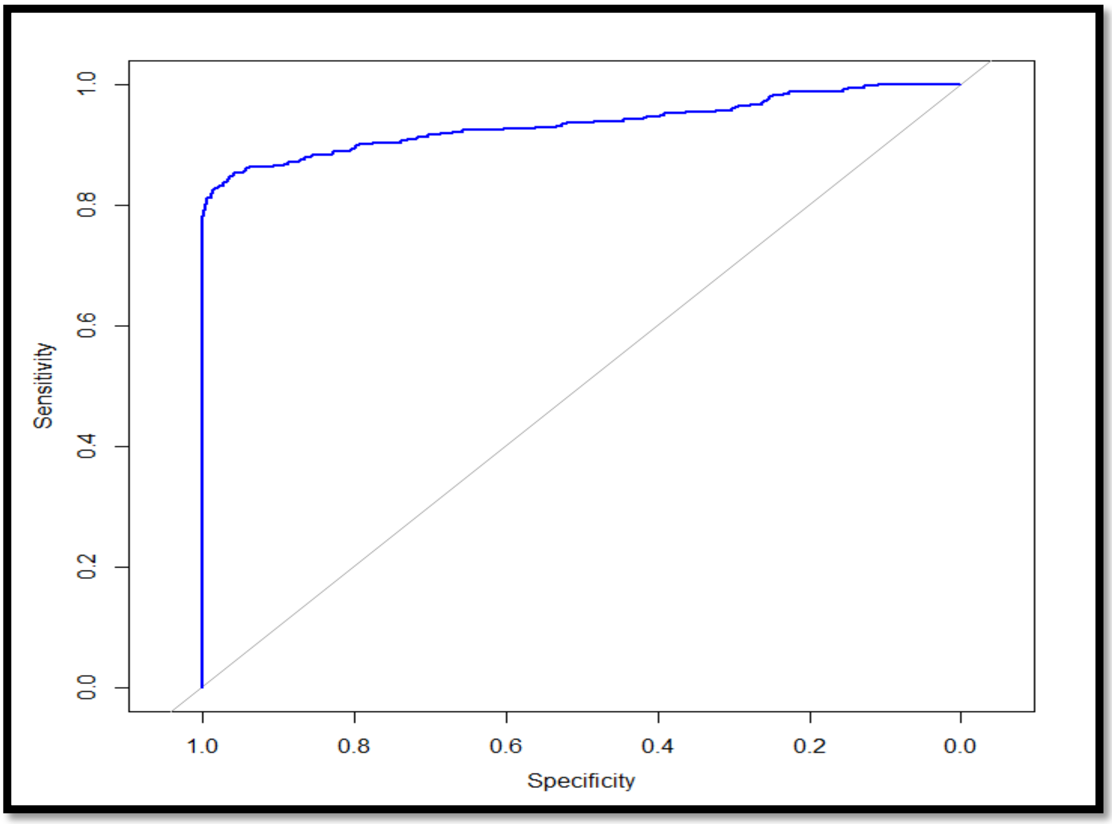


Figure 27: ROC Curve

For extensive study, Complete R programming code has been attached as Annexure-A.

5.4. Challenges in Implementing Big Data Analytics

In this world, a large amount of data is generated every second. This data makes it challenging to store, manage, use and analyze it. Even big business businesses are struggling to find ways to make this huge amount of data work. Today, the amount of data produced by big business enterprises is growing, at an average of 40 to 60% a year. Simply setting this large amount of data will not be all that effective and that is why organizations are looking at options such as data pools and big data analysis tools that can help them manage big data on a large scale. Some of the challenges you faced in using Big Data analytics are:

1. The Need to Sync to All Separated Data Sources

As data sets become larger and more varied, there is a greater challenge to integrate them into the analytics platform. If this is ignored, it will create gaps and lead to inaccurate messages and insights.

2. Major Shortage of Understanding Specialists

Data analysis is important to generate the resulting amount of data produced every minute. With the rise of specific data, a great need for big data scientists and Big Data analysts has been created in the market. It is important for business organizations to hire a data scientist with various skills as the job of a data scientist is multifaceted. Another major challenge facing businesses is the lack of professionals who understand Big Data analysis. There is a sharp shortage of data scientists compared to the vast amount of data being produced.

3. Obtain meaningful information about the use of Big Data Analytics

It is important for business organizations to get important information from Big Data analytics, and it is also important that only the right department receives this information. The biggest challenge facing companies in Big Data analytics is addressing this gap effectively.

4. Finding Voluminous Data in The Big Data Platform

No wonder data is growing every day. This simply indicates that business organizations need to manage large amounts of data every day. The amount and variety of data available these days can surpass any data engineer which is why it is considered important to make data access easier and easier for product owners and managers.

5. Uncertainty of Data Control Data

With the rise of Big Data, new technologies and companies are being built every day. However, the biggest challenge facing companies in Big Data analytics is to find out which technology will be best for them unless new problems and risks are introduced.

6. Data storage and quality

Business organizations are growing at a rapid pace. With the rapid growth of large companies and organizations, it is increasing the amount of data being produced. Saving this huge amount of data has become a real challenge for everyone. Popular data storage options such as data pools / repositories used for collecting and storing large amounts of informal and organized data in its traditional way. The real problem arises when pools / data in the repository tries to aggregate random and unrelated data from various sources, encountering errors. Missing data, inconsistent data, logical arguments, and repetitive data all result in data quality challenges.

7. Security and confidentiality of information

Once business enterprises have figured out how to use Big Data, it brings them many opportunities and opportunities. However, it also includes the potential risks associated with big data when it comes to confidentiality and data security. Big Data tools are used to analyze and store data sources. This ultimately poses a high risk of disclosure, making it vulnerable. Thus, an increase in the number of statistical data increases the concern for privacy and security.

5.5. Limitations of Research

The data used for carrying out the analysis is not the real data, instead a dummy dataset received from external website has been used to carry out the analysis in order to avoid any statutory compliance issue and to be complaint against the General Data Protection Regulation (GDPR).

Due to unavailability of data and time constraint for the research, predictive analytics has been carried out only for on the Use Case of Banking Industry. The Use Case which has been analysed as part of this research is Credit Card Fraud Detection.

6. CONCLUSION

There are many possible ways that analytics can make government more accountable, transparent, efficient and fraud-proof, which include contract management, electronic voting and health care. There are already several pilot projects in different countries regarding the use of analytics in e-health, e-resident systems, elections and especially land and property registration. A prominent country which has already several applications of blockchain technology in use is Estonia. Other countries include for example Sweden, Hong Kong, Ghana, Kenya, Nigeria or Georgia. However, despite these pilot projects, analytics is still in its infancy, so that there are still unknown factors and vulnerabilities.

The societal demand for a trustworthy public sector resonates until today. This need also includes issues such as better quality public services – fairness and customer service standards in public service provision. Informants mentioned establishing trust in governance, accessing timely and accurate information, unlinking public sector and politics as some of the key needs under this header.

7. BIBLIOGRAPHY

- Sorournejad, S., Zojaji, Z., Atani, R. E., & Monadjemi, A. H. (2016). A survey of credit card fraud detection techniques: Data and technique oriented perspective. *arXiv preprint arXiv:1611.06439*.
- Thomas, T., Vijayaraghavan, A. P., & Emmanuel, S. (2020). Support Vector Machines and Malware Detection. In Machine Learning Approaches in Cyber Security Analytics (pp. 49-71). Springer, Singapore.
- Wedge, R., Kanter, J. M., Rubio, S. M., Perez, S. I., & Veeramachaneni, K. (2017). Solving the" false positives" problem in fraud prediction. arXiv preprint arXiv:1710.07709.
- PayPal Inc. Quarterly results https://www.paypal.com/stories/us/paypalreports-third-quarter-2018-results
- Rajani, S., & Padmavathamma, M. (2012). A model for rule based fraud detection in telecommunications. International Journal of Engineering Research & Technology (IJERT), 1(5).
- Kramer, O. (2016). Scikit-learn. In Machine learning for evolution strategies (pp. 45-53). Springer, Cham.
- Paysim - Synthetic Financial Datasets For Fraud Detection https://www.kaggle.com/ntnu-testimon/paysim1
- Business Software themes for 2014. Wells Fargo Securities. Equity Research; 2014
- Turner, D., Schroeck, M., & Shockley, R. (2013). Analytics: The real-world use of big data in financial services. *IBM Global Business Services*, *27*.
- Srivastava, U., & Gopalkrishnan, S. (2015). Impact of big data analytics on banking sector: Learning for Indian banks. Procedia Computer Science, 50, 643-652.
- How can financial services industry unlock the value of big data. PricewaterhouseCoopers; 2013.
- Big Data: The next big thing. Nasscom; 2012.
- Floyer, D. (2013). Financial comparison of big data MPP solution and data warehouse appliance. Wikibon Article.
- Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. MIS quarterly, 1165-1188.

- Oracle Financial Services. Initial Steps on the Journey through Big Data for Financial Services Institutions; 2012
- Pivotal Case Study – China CITIC Bank. Driving Revenue and Reducing Risk; 2013.
- PWC. FS Viewpoint - How the financial services industry can unlock the Big Data; 2013.
- Emcien. Big Data for Banks - 5 Profitable Use Cases
- Laney, D. (2013). Meta Group: Application Delivery Strategies.
- Sobolevsky, S., Sitko, I., Des Combes, R. T., Hawelka, B., Arias, J. M., & Ratti, C. (2014, June). Money on the move: Big data of bank card transactions as the new proxy for human mobility patterns and regional delineation. the case of residents and foreign visitors in spain. In 2014 IEEE international congress on big data (pp. 136-143). IEEE.
- Agarwal, R., & Dhar, V. (2014). Big data, data science, and analytics: The opportunity and challenge for IS research.
- Hoffmann, A. O., & Birnbrich, C. (2012). The impact of fraud prevention on bank-customer relationships. International journal of bank marketing.
- Barker, K. J., D'amato, J., & Sheridon, P. (2008). Credit card fraud: awareness and prevention. Journal of financial crime.
- Cloudera. Identifying Fraud, Managing Risk and Improving Compliance in Financial Services; 2013.
- Capgemini. Big Data Alchemy: How can Banks Maximize the Value of their Customer Data; 2014

Annexure-A (R Code)

```
#Importing Dataset
cc_data<-read.csv('E:\\MBA\\Data analytis\\akanksha\\R\\creditcard.csv')

#2. Data Explortion
dim(cc_data)
head(cc_data,6)
tail(cc_data,6)
table(cc_data$Class)
summary(cc_data$Amount)

#3. Data Manipulation
cc_data$Amount=scale(cc_data$Amount)
NewData=cc_data[,-c(1)]
head(NewData)

library(scales)
library(caTools)
#4. Predictive Analytics( Building a Model)
data_sample = sample.split(NewData$Class,SplitRatio=0.80)
train_data = subset(NewData,data_sample==TRUE)
test_data = subset(NewData,data_sample==FALSE)
dim(train_data)
dim(test_data)

#Logistic reg Building Model

Logreg<-glm(Class~.,test_data,family = binomial())
summary(Logreg)
plot(Logreg)

return

library(pROC)
lr.predict <- predict(Logreg,train_data, probability = TRUE)
auc.gbm = roc(train_data$Class, lr.predict, plot = TRUE, col = "blue")
```

The R code file is also attached for reference.

creditcard.R