



COMP8230 Mining Unstructured Data

Session 1, Weekday attendance, North Ryde 2020

Department of Computing

Report on Data Mining in Industry

AISHA HASSAN SHAH – 45920842 - Session 1

Submission date -12/April/20

MACQUARIE University Sydney- Australia

Table of Contents

1. Executive Summary -----	3
2. Industry Context—Domain -----	4
3. Motivation Scenario -----	9
3.1 Fraud detection and security-----	9
3.2 Managing Customer data -----	9
3.3 Customer Lifetime Value-----	9
3.4 Real-time and predictive analytics-----	10
3.5 Segmentation of Customer -----	10
3.6 Recommendation Engines-----	10
3.7 Risk Modeling -----	10
4. Problem Statement -----	11
4.1 AS IS Process Model of ABC Emirates bank-----	12
4.2 Gap between the current (problem) state and desired (goal) state of a process-----	13
4.3 TO BE Process Model of ABC Emirates bank	
5. Approach-----	14
5.1 An Overview of Text analysis -----	14
5.2 An Overview of Data Mining techniques -----	15
5.3 An overview of Process flow of ABC Emirates Bank-----	17
6. Results/Prototypes-----	18
7. Conclusion-----	19
8. References -----	20

1. Executive Summary

In this report, the industry domain which has been discussed is “Retail Banking”. Firstly, banking working process and important terminologies used in banking sector are explained to make better understanding of the banking sector. In banking sector, data is generating on daily basis, the most known data which is easily available is transactional data, but in this business study, the main objective is to explore unstructured data comes from various data island and upon retrieval and extraction of unstructured data how we can consumed it properly for betterment of banking sector.

In this technological Era, when online banking is more approachable than person to person banking and customer's activities rate is extensively increasing for accessibility of services and products offered by bank. The main objective is to focus on unstructured data which can be insightful in many overlooked areas in Retail banking such as Fraud detection, risk management and customer relationship and management.

We have case study of ABC Emirates retail bank. This bank is mostly focusing on transactional data and ignoring most important but hidden unstructured data, due to lack of informative data, related to bank and its customers, system has started creating issues. Thus, to resolve ABC Emirates bank issues, the most disruptive technology of unstructured data mining has introduced to the system. This evolved technological approach by implementing(text mining and data mining algorithms), definitely resolve all previous issues of the ABC Emirates bank, gain lucrative insights for revamping three major sector of baking, performance management, CRM, risk management and at the same time provides granular segmentation of customer's data.

2. Industry Context Domain [I]

The Banking sector is the most important participant of the financial services of stabilized and growing economy of any country. Banks are meant to be functional in the channelization of liquid assets for both consumers and businesses. Banking sector has evolved tremendously in 21TH century as it is required by the process of development in banking industry. The main conventional activities of banks were collection of money (savings, salary, payments etc.) this is known as deposit¹, provide money for lending (loans, credit², investments³ etc.) and facilitate services. Banking industry has come a very long way by adding more functional and structural features in banking process without changing its initial practices. In this advanced Technological Era, where the banking is the most adaptable and growing financial institute aims to facilitate maximum categories of monetary processes with operational efficiency and financial stability. In modern Banking, there is diversity of services to cater groups of consumer such as (individuals, sole proprietorship, small businesses, Corporate etc.). Expansion in banking industry brings different types of banking institutes deal with its corresponding customers and provide specialized and unique functional services such as Central bank, Retail bank, Commercial bank, Cooperative bank, Investment bank, specialized bank, National bank etc.)

In this report, the main focus is only stick with **Retail banking**, the exploration of its services, structural operations and different trends in this specific division of banking industry. Retail banking processes are most similar to the traditional banking and its entire focus emphasis to the Retail customers.⁴ Retail banking comprises to offer most flexible financial services and product to the general public (retail customers).

The most prominent and consumed services are mentioned below.

Borrowing loans	Banks are providing specific amount of money by keeping reserves from the borrower with the conditions of interest and time limit .It includes mortgages on personal and professional properties, automobile financing from bank on the basis of net income of account holder and bank will charge back to customer the paid money + interest on any advanced financial aid to the account holder over the certain period of time.
Credit cards ⁵ /Cash credit	This service allows the consumer to spend money in buying goods and services which is paid by bank and consumer can pay back that amount in specific time limit with interest.

¹ A deposit is a financial term that means money held at a bank.

² Borrowing a money from bank with the agreement of paying it back with interest and specific time intervals

³ Investment is the process in which Brower take money from bank under terms and conditions and utilize it as per its convenience or feasibility.

⁴ Retail customers are the individuals, who are saving and depositing their money in the bank and bank is allotting the account to the individuals. Bank is offering multiple services to their account holder to facilitate them such as issuing supplement cards to the family members, credit, debit cards, assigning cheque books and many more services are available for the account holders.

⁵ A credit card issued by bank to the account holder which comes with two option usually either MasterCard or Visa. Account holder is allowed to use it for expenses with certain limit of funds and bank is charging a fee and rate of interest for providing this service to account holder. While customer is using credit card for buying goods or services, the seller sends the bill to the bank, the bank pays the bill after identification of the customer credit card and in the end of the month bank sends the credit card history (bill) to the account holder/customer who is oblige to pay the bill under credit card terms and conditions to the bank.

Cheque book	Bank issues cheque book to the account holder ⁶ . Bank permits to consumer to draw cheque for payment purposes and bank will pay from the holder account after authentic verification procedures.
Foreign Currency Exchange	Banks facilitates to its customer to exchange foreign currency in to local currency
Remittance	Customers can transfer money from one place to another through cheques , draft ,pay orders ,online transfer(bank to bank) etc.
ATM Services	One of the most advanced service provided by bank, without going in to bank, a customer can deposit, withdraw, inquire etc.(by using ATM machine). Availability of 24 hours, fastest and no human interaction.
Debit cards ⁷	Debit cards are issued by bank to account holder, holding funds in the account and can withdraw money from this card or pay their payments directly from card. Verification authenticity is required upon each transaction.
Online banking	This service is the evidence of banking sector transformation and utilization of disruptive technology. With the help of technology, bank offers to account holder to access their bank accounts online via internet. This virtual access of account ,permits account holder to perform all daily basis transaction such as paying bills, credit cards payments, money transfer from one account to another, checking balance etc. Accessibility and availability is highly approachable. There are several devices rather than computer such as mobile, tablets, PDA all can be interfaced between account holder and online banking(internet banking /web banking)
Taking deposits	Deposition of money for savings or carrying for account holder/customer is the basic service of any bank. Banks keep money of depositor in to its allocated account and give full access of this account to account holder. There are generally two kinds of account available for single individuals' current account ⁸ and saving account. ⁹

In retail banking all major and connected physical branches are doing effective operations and serve the customers as its best capacity but the reality is that technology has overcome and impacted the most this particular division of banking. Online banking ¹⁰is one of the key factor which proliferates the platform of retail banking in to another level. As technological trends introduce more flexibility in accessibility of services, Banks launch their mobile application for the ease of customers to log in through mobile application (secure connection) to the bank and access the facilities offered by bank. Online banking and mobile banking both are virtually connected to the bank (front end) where customers can avail core services of retail banking for instance, payment of utility bills, online

6 An individual who is registered a physical account in the bank and authorized by bank to access this account for transaction purposes.

7 Debit card is issued by bank to the account holder. This card is like paper money or plastic money, it allows an account holder to pay their bill directly charge money from the bank account.

8 The Current account is the regular account of the account holder in the bank, which permits deposit of money, withdrawals, several payments and transactions on daily basis. On behalf of this account bank releases cheque book and debit card to you to access it for immediate utilization of money.

9 Saving account is also proposed by bank for saving purposes for individuals, it facilitates account holder to deposit or save money for long term and bank will provide interest to the account holder for keeping such big amount in the bank for long term

10 Online banking is the platform provided by bank to the customer to access the services via internet. Authorized web portal (website) is published online by bank for all the customers. Customer can access bank's website via internet(secure connection) and requires to log in with their private login credentials to connect to the virtual bank in order to perform online transactions and monitor the bank account virtually. This is also known as internet or web banking

transactions, monitoring income statement etc. but still there are physical branches exists to offer some services which are not offered by bank online (not on website or either mobile application) and require physical interaction between the customer and the bank (backend processes).

Continuous usage of online and mobile banking can lead to another advanced and robust technology which is known as digital banking. Retail banking is extensively services oriented financial business and customers funding is the most important source of it and in reverse it is bank responsibility to produce dynamic and optimize innovative financial services to its retail customers. Now the present and future of retail banking and all other different type of banking and financial services is to transform in to digitization of all operational functions which will bring the most consistent and convenient availability of services to the customers and it will direct impact on customers experience with the bank, enabling customer purchase decisions, gathering useful data from customer digital activity which will be insightful for analytical purposes, multiple cross platforms can be offered to the customers and many more. This is the banking industry requirement to fulfill the customers need by adapting, focusing and enhancing digital technology which is the best approach to improve efficiency, scalability and agility of the banking system.

Digital banking is wider vision of all (online, mobile, web, internet banking). The main concept is that all the operational and functional services offered by retail banking via online banking, internet banking, physical visits of the customer to the bank will be offered digitally to the customer and there will be no any physical unit of bank exists on back end. Digital banking platforms empowers banks by providing End-to-End business solution with comprehensive digital approach by delivering inside-out of all automated functional operations. It is very advanced level of web based services which requires high level of robust processes automation to facilitate customers with banking products and services and keep updating all the data related to customer activity which is continuously updating and creating data repository of every customer.

Customers activity traffic have been increased whether using online banking and/or digital banking which generates a voluminous amount of data virtually on daily basis. When data is exponentially growing, exists in different formats and resides on multiple online and digital platforms, than these characteristics belong to "**Big data**". This big data has potential to provide most insightful information about customer which can be tool for better understanding of customer needs, buying behaviors, market trends, offered product performance ,planning and production of new products, business analysis for the bank itself and many more predictive and prescriptive analysis for foreseeable future of banking industry. Big data mainly represents three types of data, which are structured data,¹¹unstructured data¹²and semi structured data¹³. In this report we are investigating unstructured data by using big data analytical tool which is data mining.

Data mining techniques explore the data in depth and extract the most appropriate and reliable information from the raw forms of data. Data mining of unstructured data is quite challenging because unstructured data is not having any specific format and data is hidden and scatter in multiple platforms of digital world. In banking industry we need to rectify all possible data repositories where this unstructured data is generating on daily basis and after that data mining techniques can be applied for retrieve of the most filtered, transformed and understandable formats of data to convert it in to formal multifaceted structure of data which will provide us technical

¹¹ Structured data is an organized data and available in tabular format with predefined schema.

¹² Unstructured data is raw form of data(without any format) which is usually text-heavy but sometimes data may appear with dates, numbers and facts as well.

¹³ Semi structured data is not completely table based dataset(non-relational) but it is generalized with markers or classifier .

information of raw data for further operational and analytical processes of decision making and strategy management for banking sector.

Unstructured data can be extracted from many data resources related to the account holders and retail customers of the bank. These data islands are in extremely raw form (irregular and ambiguous) of data but after passing through the proper channel of preprocessing and algorithms of data mining, data will be modified and ready to use for analytical insights.

The most prominent and consumed unstructured data islands for online and digital banking are mentioned below.

Emails	Emails are the text written correspondence between account holder and bank. For data mining purposes, Emails carry a hefty amount of information such as classification of email addresses and text (for email spam detection). In banking sector data of email can provide customer experience and inquiries related to products and services which will enhance the functionality of Marketing through email data and management purposes
Survey responses	Survey responses is also text driven data but after preprocessing and analyzing it properly, can be used for research work in any relevant case study. In banking surveys provide feedback of customer on current services and help to organize new services and identify areas of improvement.
Social media posts (blogging, Facebook, websites, twitter, Instagram, LinkedIn, Pinterest etc.)	Social media data is the representation of human behavior, it is the best opportunity to understand customer sentiments. Availability of this is possible by web scrapping converts unstructured data in to structure data. Banks can use this data for several purposes such as predicting upcoming trends, tracking customer interest, offering promotions and many more.
Customer profile	Customer profile is customer data, which is generating on daily basis (deposits/withdrawals, purchases, online transactions etc.).The collection of this data will enhance the chances of bank to KYC (know your customer), utilization of this type of data which is rich in information is still progressing. Data mining techniques are imposing to retrieve this form of data so it will provide detailed insight of customer day to day activities. To monitor digital activity of customer known as eKYC. Analytics about customer activity is precise approach to control financial crime, money laundering and customer identification.
Transcript of call center (interaction between customer and help desk) Online generated enquires on digital website	Digitally recording of customer calls from call center of bank and online enquiry chat of customer through web portal come under the unstructured forms of data, which can be preprocessed with different algorithms of data mining to convert in to usable

	data which helps to identify CRM performance of the banking sector.
Financial news	Online Financial News feed is external unstructured data source which provides analytical view on strategy management and plan arrival of new products and services to the customer, which enlightens the customer for better choices.
Open data	Open data is the external unstructured data source, available in text format and can access, alter and share freely from the sources (wiki data, governmental data etc.).The format of the data is machine readable and can easily processed .It is easily available without any copyright issues (public data), so for the purpose of predictive analysis and new upcoming concepts and trends can retrieve easily. Open data is useful in banking industry for research purposes and continuous mining insightful data for future business and customer benefits related trends.
IoT (internet of things)	IoT is the terminology used when all the physical devices (mobile, tablet, PDA, Sensors, gaming apps, Smart TVs electronic gadgets etc.) attached to the internet and all the activity and readings are keep updating saving in the forms of log files. This phenomena provides digital intelligence to the devices and the collection of data can be utilize in any research work and study. In banking sector customers are using online platforms continuously to correspond their respective working companies with their electronic devices .These interactive activities are used keep saving data related to customer profiles which will be helpful for banks to understand their customers.
Videos, images, live streaming etc.	Videos, images, pictures, live streaming are the source of unstructured data and equally important as text data. Picture recognition or photo identity is most necessary requirement in banking sector, it helps in security management. There are many more advantages, system can gain from video and image data.

3. Motivation Scenario

Banking sector is evolving with the progression and advancement of technology. The phases have been discussed how Retail banking modified in to online banking and online banking is transforming in to digital banking. This structural and operational development of banking sector have come up with diversified forms of data .The huge volume of unstructured text data retrieving from multiple platforms related to banking customers, businesses, employee performance, functional processes which will provide more depth of understanding of data and create more chances to enhance productivity, security, effectiveness in decision making in banking sector. The main motivation of studying unstructured data of banking sector is to work on some neglected areas in retail banking which can be improved efficiently by utilizing big data. (Unstructured data). [\[2\]](#) [\[3\]](#)

3.1 Fraud detection and security

Banking customers are vigorously using online banking(web banking ,internet banking) and access banking application(bank website and App),visits multiple websites with different technological devices for buying purposes(online transactions) and using their debit and credit cards details which lead to jeopardize customer's bank account security(including chances of losing or stolen credit/debit card information by phishing¹⁴). This is bank responsibility to identify proactively any unwanted purchase on customer's credit or debit card or any unusual online transactions occur. As soonest as bank finds out the activity of fraud (irregular transactional patterns), it can be easily preventive without much loss.

The retrieval of all unstructured data (related to customer's bank account and online activities) applying data mining techniques (clustering, classification, SVM and many more), to transform the scattered or noisy data in a format where data is continuously updating for the precise insights on customer's activity. There are procedures (Fraud detection algorithms) can investigate data information (retrieved unstructured data after preprocessing) to capture unwanted purchases, or unusual increase in number of accounts registered, in small time span with similar information.

3.2 Managing Customer data

Banks are continuously storing, managing and monitoring customer data but these large data marts are working regularly as required by the system there is no new next step, which should be taken for Customer wellbeing.

With the help of big data analytical approach, the unknown data (unstructured data) laying on multiple data sources can be obtained, preprocesses, transformed and feed in to analytical processes to know more about their targeted valuable customer .At the same time, system is well aware about the market trends so it will be in benefits of banks to offer latest products to the customers and generate revenue opportunities.

Multiple approaches can be done on clustered and classified data of customers while offering general products and services to the customers, bank can purpose personalized offers(by predicting the choice) of individuals(as monitoring ,managing and segregating the activity of each customer).This is personalized banking and this approach helps to build up, bank and customers relationship.

3.3 Customer Lifetime Value

After this progressed adaptation of big data retrieval and collection of data. Banks are opened to perform for seeking new analytical procedures which will enhance funds generation for bank and at the same time will cater the bank customers at their best capacity because customers are most

¹⁴ Phishing is a cybercrime which involves producing fake website where users provide their personal details (asked by website)online by mistakes.

important assets for the bank . Those individuals who have saved and keep investing big chunk of money in the bank and automatically become a part of maximum funds contribution for the bank can easily eligible for Customer Lifetime value. Data mining of unstructured data provides 360 view on customer data and system can easily locate the Customers which can be qualified for discounts from bankside in future financial schemes/revenues). This is possible with predictive analysis to calculate the lifetime value of customers, assigning the funds and engaging them in long term commitment with the bank.

3.4 Real-time and predictive analytics

As availability of complex, hidden and unknown data is possible with mining of unstructured data in banking sector, so this is the time to perform analytics and drive effective decision making strategies which improve productivity and performance of banking sector overall. In this scenario Real time analytics provide granular data insight to rectify the issues that are effecting the business. While predictive analytics behaves as a problem solver to resolve the issues by recommending the correct technique. Two analytical techniques contribute substantially in work flow of banking processes with experiencing hurdles in current system (real-time) and potentially predict better process (predictive) for smooth operations in future.

3.5 Segmentation of Customer

We are in most digitalized and advanced phase of banking and all types of data (structured, unstructured and semi structured) from big data is organized and easy to manage, retrieve and modify for insightful analysis (hypothetically) .This is the best practice to divide the customers in groups according to their similar commonalities and in order to approach them appropriately. There can be various segmentation of customers according to their behavior patterns or any general feature similarities. These segmentations are highly productive for CRM and all structural and functional operations which are introduced after successful digital banking.

3.6 Recommendation Engines

With the help of big data and data mining techniques, banking system has command on collection of data and vigilant departments are scrutinizing data on daily basis. So this is system's prime responsibility to offer best and concise products and services to the customer. According to the observations, customer transactional activities, personal and professional information are suggesting which kind of services and promotions should be offered to the customer. Banks are the partners with many cooperate companies and for the generation of revenue and growth banks behave as mediator to offer companies products to the relevant customers (who can be interested in).Banks are also offering those products to the customer which are generally in trend in the market and most likely to be purchased.

3.7 Risk Modeling

Banking is the business which provides several financial services and in return funding itself by the payments paid by customers to the bank in the form of interest rate, services charges, credit/debit cards bills and many more. In real time scenario, when any customer applied for credit card, while processing an application for the eligibility of assigning a credit card or not, as there is always a risk factor exists whether the customer will repay the money to the bank. This is a complicated task for the bank.

With the help of Risk modeling of big data, we have an opportunity to explore bank customers in detailed and find out the customer's credit archives and other transactional payments, this will provide an idea to the bank to apply a threshold limit while issuing a credit card to different groups of customers and develop more precise strategies to reinforce lending criteria. Risk

modelling is also measuring overall functioning of all departments in the bank and keep analyzing and monitoring the track of overall performance of the bank.

4. Problem Statement

ABC Emirates bank is well known and one of the reputable bank in the country. Bank operational functions are based on Retail banking and trying to facilitate their customer at its best capacity. Bank has well developed (full setup) which is required for retail banking, legitimate bank website and mobile application facilities too. Customers are accessing bank services online and highly active with online transactions and Ecommerce businesses. And if there is any extraordinary situation, Customer can visit the local branch for face to face consultations. Most of the data is transactional data (OLTP¹⁵), allocated in the forms of tables and internal departments operations are executing on routine basis, performing operational analysis (reporting) on relational database (bank's internal data) is satisfying bank's higher management and it seems everything goes perfect.

But as Customers are increasing and utilization of online banking increases, this is the starting point where bank starts realizing missing out data due to unwanted complains from customer side and when bank is trying to adapt high level of data analytical and visualization tool, to see the performance of banking system and monitoring of Customer activities for better understanding of the situation in the system, it is difficult to acquire, due to the transactional and operational data of the customers are not enough available(sample size, features, instances etc.) to perform in-depth insight on the data which can draw any decision making conclusions, mainly for the strategy management, Security measures and Customer relation management, which are the prerequisite steps towards the betterment of the bank and customers.

The most prominent issues are persisting in the ABC Emirates banking system are mentioned below.

- Bank is receiving complaints about the unknowns transactions are being carried out on credit cards of some customers.
- Some customers are not able to pay their dues to the bank and in this case bank capitalization is suffering and after internal audits in the finance department, it shows that AR(account receivable) is less than AP(account payable)
- CRM department is facing number of unsatisfactory complains (on phone, Email and personal visits of the customers), Customers are not happy with the offers and packages purposed by bank. (Services and products).
- Strategy and Marketing departments are performing very feeble and revenue generation is highly effecting due to lack of focusing on productive measures which will go in favor of bank rehabilitation.

¹⁵ OLTP (online transactional processing) is the process, works for transactional data. In this process data can be add, delete and update. It supports transitional oriented data applications on the internet

4.1 AS IS Process Model of ABC Emirates bank

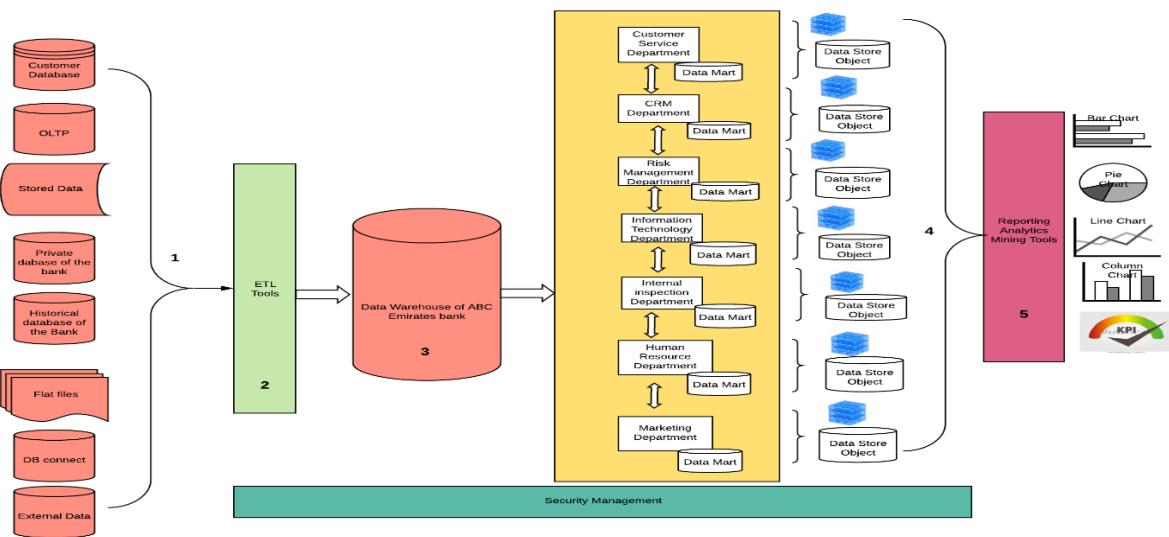


Fig A: AS IS Process Model of ABC Emirates bank

4.2 Gap between the current (problem) state and desired (goal) state of a process

The major gap between the current system and desired system is data. In Current system data feeding is not adequate (amount of data, retrieved properly from maximum data resources) to the corresponding departments and highly impacting on performance management and improve productivity.

In purposed system to the bank, the main and imperative steps are to focus on big data analytics and trying to extract all possible 80%untapped (unstructured data)from various data platforms which can be most in contact with the customers.

Upon successful extraction of data (batch, micro batch, real time) it needs to be laying down under staging phase, for further cleansing of data. After preprocessing, data is ready for next step Extraction, transformation and loading in to most impartment settlement of data in to different type of data formats (Transactional and operational data which we have imported from AS IS system of ABC Emirates bank, SQL databases for structured data and NOSQL databases for unstructured and semi structured data).Now the most innovative approach for unstructured data is going to initiate and that is Text analysis¹⁶. Text analysis has different approaches for contextualization of text¹⁷, these segmentations of raw unstructured data (laying on numerous data islands) add value in to the data. Now system is becoming smart, it has knowledge graph¹⁸of the unstructured text data which is most filtered, precise, non-redundant, and flexible form of text data, ready for further procedures. Thus, data mining algorithms¹⁹prepared to apply on ‘signal data’²⁰ for modelling and Analytical tasks.

This evolved technological approach, definitely resolve all previous issues of the ABC Emirates bank, gain lucrative insights for revamping three major sector of baking, performance management, CRM, risk management and at the same time provides granular segmentation of customer’s data.

16 Text analysis is resolving text in order to obtain machine readable facts. In fact, the main reason is to convert text in to structured format.

17 Contextualization of text means, conversion of raw text in to meaningful and cognitive formats for analyzing purposes.

18 Knowledge graph illustrates a map of interlinked explanation of entities (edges labelled with relations.)

19 Data mining algorithms are mathematical procedures, every procedure has certain limitation and after applying it, data can transform in to various forms

20 Signal data means useful data, free from noise(irregularities and ambiguities)

The most prominent resolved issues are mentioned below.

- Due to proper screening of all transactions carried out by customers are being monitored and records are keep updating, if there is any abnormal activity is registered (related to credit card utilization)immediately transactions are stopped and alerted send to customer for identification.
- Risk modelling is successfully implemented so customer credit history is easily acquire with other investigation criteria to check the customer's credibility before allotting any grant to the customer.
- CRM department is thriving the most due to extensive grip on all unstructured data of customer, most suitable products and services have been offered on personalized level and by looking at market trends and suggestion, famous and reputable products and services are also being offered to the customers by marketing banners or in recommendation engines.
- All these changes bring tremendous growth in bank's revenue generation.IT team is more focused and providing all possible and accurate analytics to the strategy and marketing management teams to generate new strategical plans and proposals for bank's higher management and stakeholders.

4.3 TO BE Process Model of ABC Emirates bank

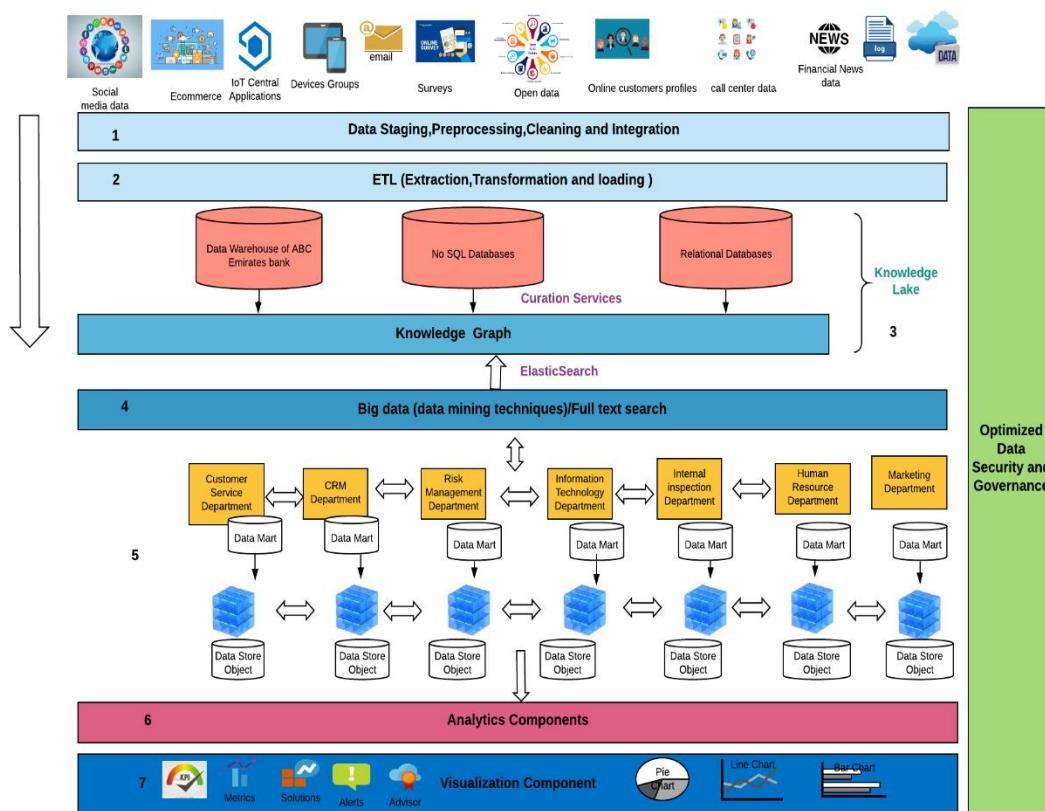


Fig B: TO BE Process Model of ABC Emirates bank

5. Approach

The rate of technological consumptions increasing with time, this is directly impacting almost all business sectors especially banking sector. Customers are more prone towards internet banking for performing all banking tasks rather than go and visits banking departments frequently. The trend of customer's behavior is changing and this is the time for banking sector to focus and implement modern IT advanced and disruptive paradigm which ensures improvement in internal and external processes of bank, Customers related services, generate new revenue growth plans. These all milestones can be done with only one significant approach and that is acquiring relevant data (80% unstructured data) from all possible data sources which are suspected to be linked with Customers, employees, businesses, vendors, partners and banking processes.

The first approach is to capture all unstructured data, which will be exposed in Text mining²¹ process to explore and derive high quality data (useful information) from the text. Text mining is integrated and incorporated with the tools of data mining techniques, machine learning and Analysis methodology (e.g. NLP). When text mining and NLP both engage together for mining unstructured data always bring most filtered and informative data landscape for further analysis. This combination (text mining and NLP) is mainly used for extraction of data from social media, where most of the data related to customer's activity and comments which represents customer's sentiments, moods, emotions and individual's opinion on different products and services. This kind of analysis is really beneficial for CRM and marketing department of banking sector.

5.1 An Overview of Text analysis

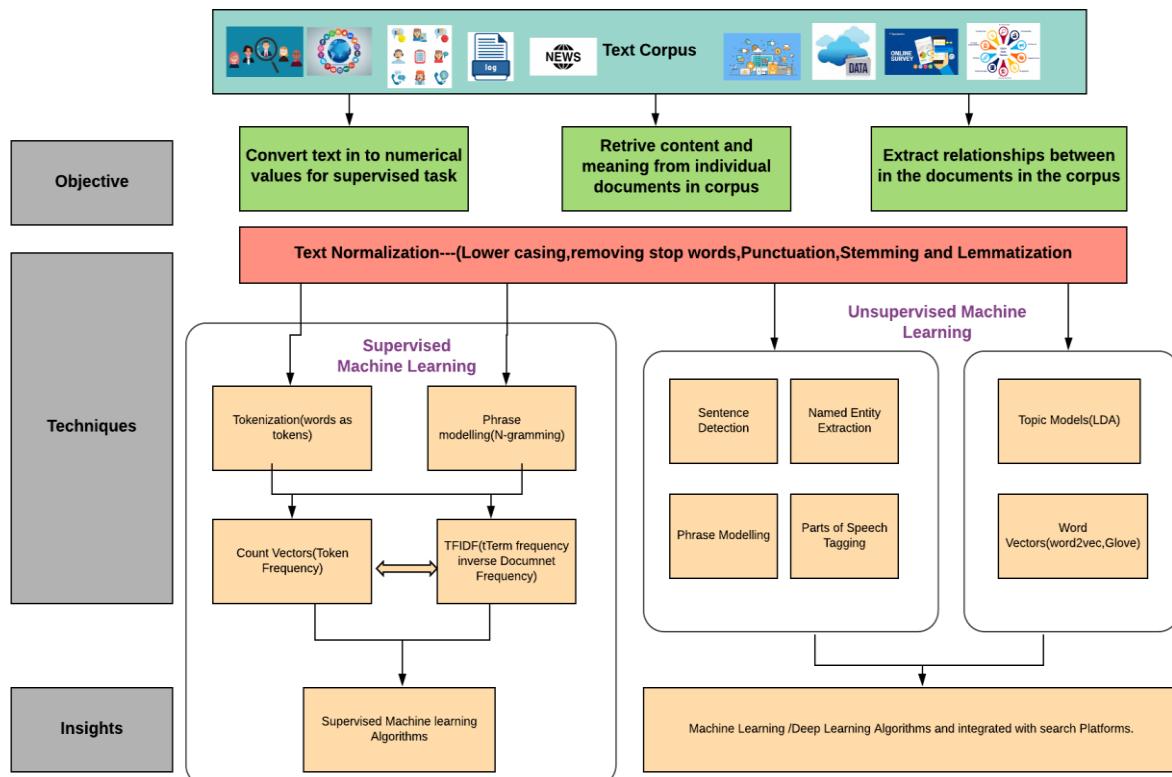


Fig C: An overview of Text Analysis

²¹ Text mining has come up with fundamental steps. These are, collection of unstructured data from multiple source of repository, find and remove anomalies by cleaning and preprocessing, (it can be done with text mining tools and application), transform all relevant processed data from unstructured to structured formats, in the end storage repositories are available to store valuable information(structured form) for performing analysis purposes.

In Fig C, the Text corpus represents, staging phase of unstructured data resources after extraction, they all are residing as large body of texts in a structured set format for next step processing. Now NLP²²(Natural language processing) is being applied on extracted data. NLP main objective is to break down text data in to small, understandable and relatable elements and trying to find out the patterns, connections, relations and meaning of the data. There are many NLP procedures but basic and most important are tokenization²³, text parsing²⁴, part-of-speech tagging²⁵, Named Entity recognition²⁶, lemmatization /stemming²⁷, topic modelling²⁸, TFIDF²⁹and recognition of semantic relationships. [5]

The most prominent text mining techniques used in banking sector are mentioned below.

Sentiment Analysis	CRM (Customer relationship management) is the most integral and sensitive department in banking sector. Customer's feedbacks, satisfaction reports and retention can be gauged by sentiment Analysis. This kind of analysis helps management to decide most likeable products and services by customers and generates alert, if there are negative opinions related to any specific product or service to take further actions. (decision making)
Free Text Analysis	Free text exists in the form of extra information, comments, additional suggestions provided by customers in online survey, e-performance evaluation forms etc. This extra information always brings some extra advantage while making business decisions related to any product consumption and needs improvisation on specific feature of that product for the betterment.
Context Based Analysis	During text analysis, it is very important to understand the meaning of Customers words wisely. Content based analysis is cognitive interpretation of text to comprehend the text data in accurate meaning and best suits to the real scenario.
Statistical Text Analysis	Statistical text analysis is used when data is too noisy, grammatically wrong, incorrect spellings ,icons ,symbols, characters and any other text mining techniques are not providing meaningful data, so this analysis can be drawn some useful information in the form of patterns and trends

5.2 An Overview of Data Mining techniques

With the presence of big data, it looks possible to explore and exploit the data (unstructured data) properly for the utilization of it for solving many challenging areas in banking sector. Data mining techniques are going to apply on extracted, modified and transformed data for retrieval of

22 NLP is an integral of text mining that deals with linguistic analysis that helps machines to read text data meaningfully

23 Tokenization is the process to break down a string in to hierarchical backwards pattern (e.g. paragraph can be break in to sentences and sentences can break in to words).

24 Text parsing is used to find out syntactic structure of the text by observing its words with grammar

25 This is tagging on parts of speech in a tuple which recognizes noun ,adjective ,verb and so on

26 It is popular technique applied on Named entities to categorize them under various predefined classes(proper noun etc.)

27 Stemming and Lemmatization deal with number of words, due to inflections they are different. Stemming detaches suffixes/prefixes/affixes from the word and lemmatization is to stick with based form of actual word.

28 This is mainly dealing with unsupervised learning technique, focuses on scanning of documents, looking for most relatable words and phrases and start making clustering on basis of similarities with in them.

29 TF-IDF (Term frequency-inverse Dense Frequency),this technique manages balance between number of words and how frequently they are occurring, if frequency rate is high of any word, it will be least important and not included for further procedures.

comprehensive analysis on most important targeted divisions of banking and implement stringent rules against unwanted activities and apply preventive measures to acquire better outcomes for performance management, CRM and risk management.

The most prominent data mining techniques are used in banking sector for following criteria are mentioned below.

- **Fraud detection and security management**

Extensive flow of big data is more vulnerable with growing and complex high tech cybercrimes.

Data mining techniques are effective in monitoring unwanted fraud behaviors. In supervised learning³⁰, data mining techniques for security and fraud management is “Classification³¹” and under classification umbrella, implementation of Decision trees³², Support vector machine³³, Neural Networks³⁴, Naïve Bayes³⁵ and in Unsupervised learning³⁶, K-means clustering³⁷ and association rule mining³⁸ are most relevant and effective algorithms should apply on big data for effective and accurate outcomes.

Application of these algorithms according to their learning categories (supervised/unsupervised) or pair up with each other is possible. For example K means clustering and SVM are appropriate for data preprocessing and classification, after that K-means can be applied for feature selection and after that SVMs are used for final classification between fraudulent class and genuine class.

- **Risk management**

Risk management on big data is highly important for funding (finance and controlling department) And also assessing overall risk factors (bank efficiency evaluation, AR department update and warnings of failing banking) in the banking sector. The most appropriate algorithms for finding defaulter list and segregate authentic customers with the implementation of Decision trees, Neural Networks ,SVM, Naïve Bayes and linear regression(for predicting only one target value) and K – means Clustering. These algorithms are also providing key vigilant alerts for all risking factors in the system.

- **Customer Relationship Management**

Data mining techniques are widely consumed in banking sector for marketing and Customer relationship management (customer's management, segmentation, profiling).These techniques have capacity to gauge customer's behavioral patterns, targeting capable customers, precise understanding of customer, retention customer satisfaction and many more and equip marketing and strategic management departments to design multifaceted plan for customers and bank.

The most prominent data mining techniques used for CRM are mentioned below.

For customer profiling, the most relevant algorithms are Classification (Decision Tress, Neural Networks) and K means clustering. Customer segmentation is possible with K-means clustering

³⁰ Supervised learning is the data mining task in which you train a model by labelling the data and training set will predict the outcome of untrained data

³¹ Classification is supervised learning, in this approach, the program will learn from the input data and by its experience it classify the new observations

³² Decision Tress are supervised learning, in training dataset input and output is classified already, and data is keep splitting according to the specific parameter.(two features of decision tree are decision nodes and leaves).

³³ SVM is a supervised machine learning and can be applicable for both classification and regression tasks, the basic idea is to maximize the margin around the boundary of the data points and generates the classes of the data point (classes of data points which have similarities between them)

³⁴ Neural networks are group of algorithms, the internal structure of neural networks is like human brain neurons network and made for recognizing patterns.

³⁵ Naive Bayes algorithm works as Bayes theorem to classify objects.

³⁶Unsupervised learning is data mining task, in this task, training is not required, model will do its own learning by observing and discover useful information (data is unlabeled).

³⁷ K-means algorithm is repetitive function tries to find out the partitions of the dataset without overlapping into K-clusters.

³⁸ This algorithm finds frequent patterns, correlations, structures and linking between datasets of the databases

algorithm. Customer satisfaction can be explore with K-means clustering and classification (Neural Networks) .Customer life time value can be achieved by utilizing classification (Decision Tree, Neural Networks, Naive Bayes, Linear Regressionism) and K –means Clustering. These are the combinations of algorithms individually applied on each subdivided category of CRM.

In [Fig D](#), the process flow of ABC Emirates bank is visually display, in which the main focus is to emphasize the acquisition of unstructured data, after retrieval data should pass through the text mining and NLP procedures and extraction has been take place then loading the table into appropriate data repositories then new platform, Knowledge Lake has been introduced ,for cognitive assistance of all data at this stage must be most modified and transformed format of raw data(which we retrieved from many data islands), for going in to further procedures of data mining techniques. The data mining technique is directly connected to banks all departments and both are working together continuously by sharing new predictions on specific datasets and as data is continuously coming ,mining techniques are keep training and recognizing patterns of data and pushing the information in to respective departments for further analysis ,visualization and strategy management.

5.3 An overview of Process flow of ABC Emirates Bank [\[4\]](#)

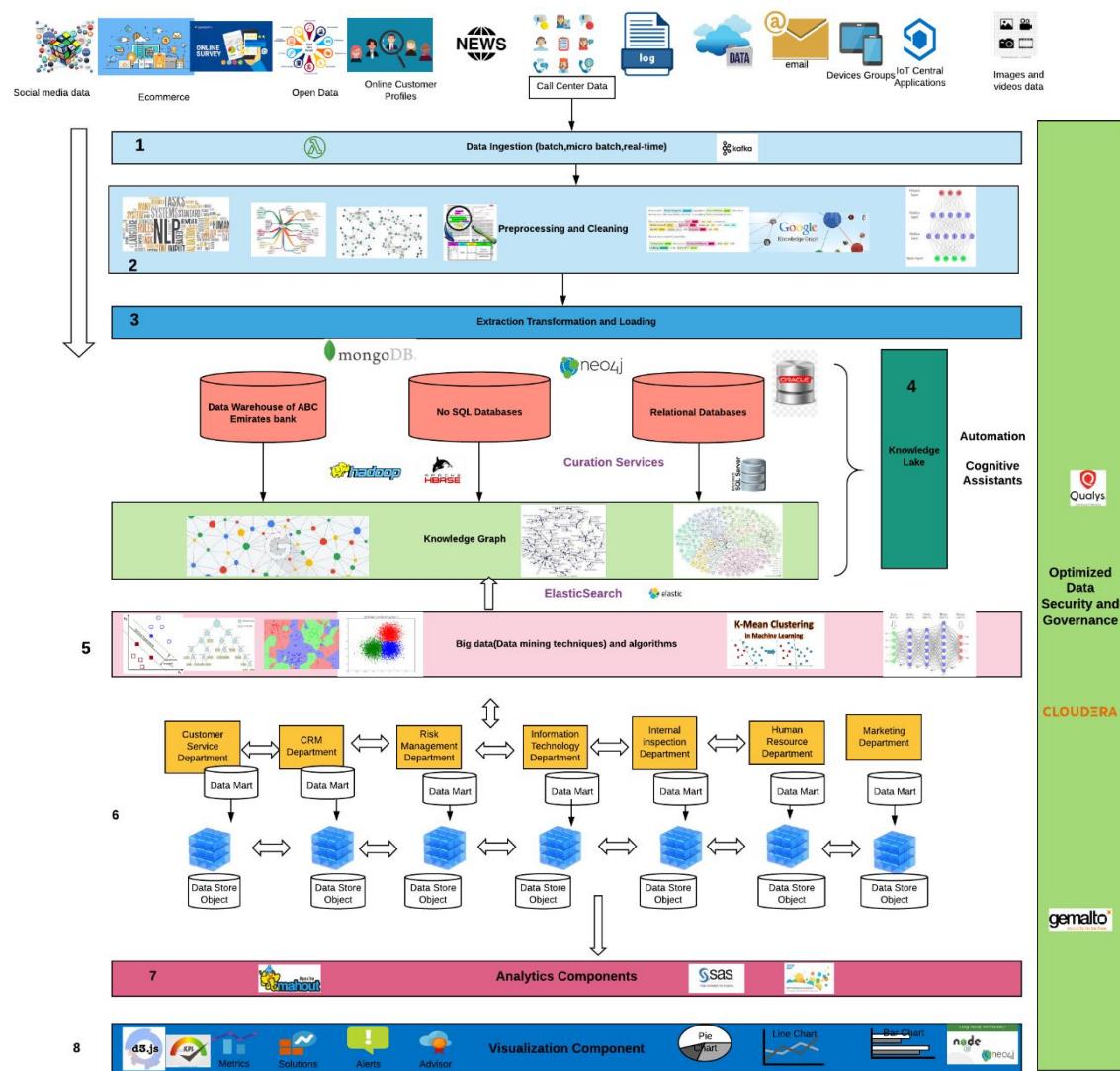
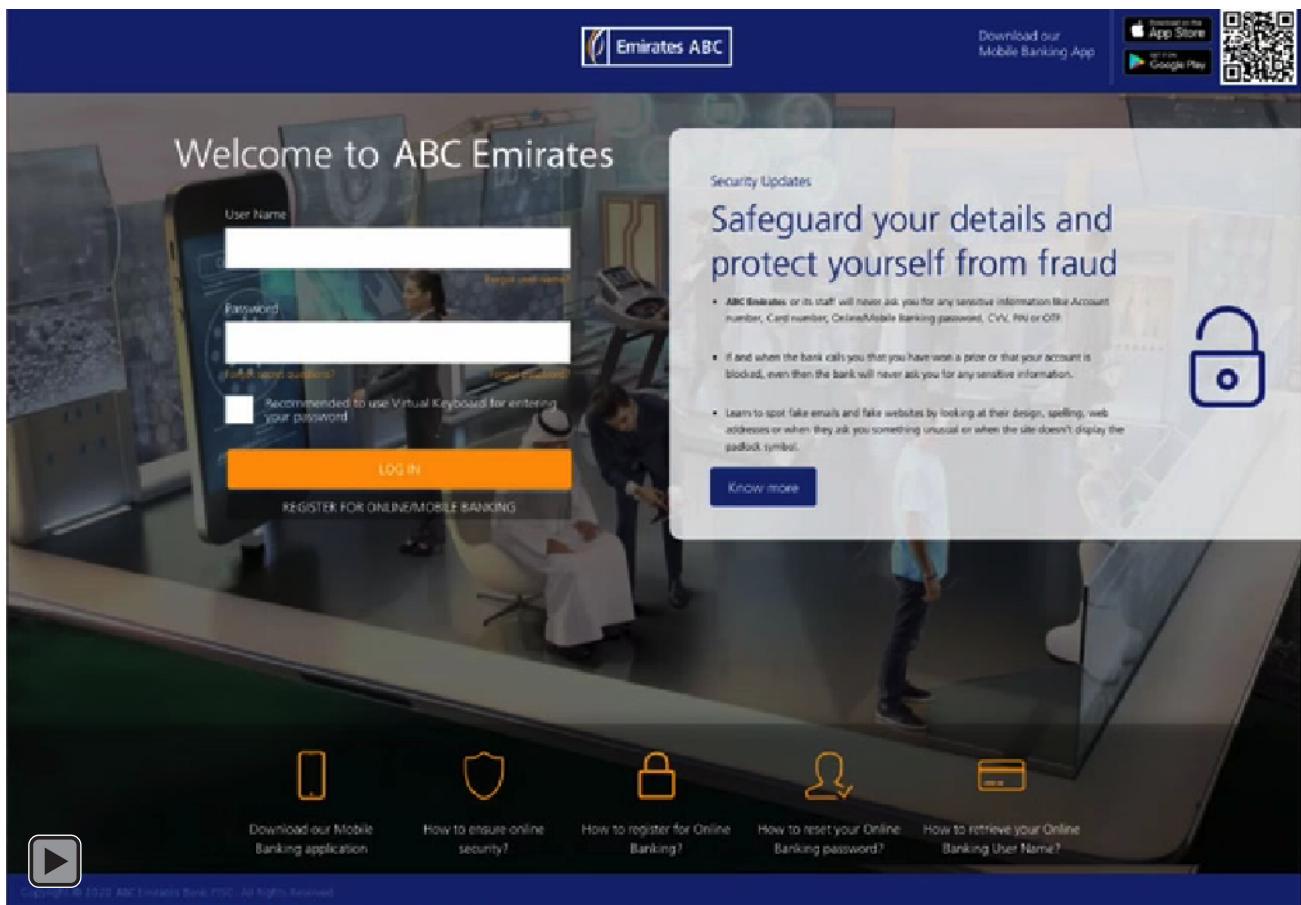


Fig D: An overview of Process flow of ABC Emirates Bank

6. Results and Prototypes



The above display is an adaptation of original Results of the system which is purposed in problem statement and has been discussed in Approach section. It is a Video file (**PRESS THE PLAY BUTTON**), First part shows the Application of new ABC Emirates bank and then graphical representation of each department showing improvement due to the implementation of new system.

In CRM department, comparative bar chart represents “Services” of bank offered to customers. It shows with time, the consumption of services increases and there is decrement in number of complaints registered by customers. In “Product” line graph shows as system is reading and keep segregating customers’ data and performing classification and clustering(back end), the variety of different products are offering to customers, and this time, it is more personalized and favorably selected by customers. In Risk modelling the line graphs illustrates, due to strict measurements taken by risk management, credit card usage activities monitored and department is imposing limits on credit availability on credit cards. In the pie chart of fraud detection and security, it can be seen clearly that number of frauds are declining and security is getting robust against those malicious online activities.

With this new approach, system is improving and working efficiently for key functional operational departments of ABC Emirates bank.

7. Conclusion

By utilizing most worthy data (unstructured data) which is hidden and growing exponentially on daily basis and containing very valuable information of customers' data. By acquiring this data with the help of many processes and evaluates this information to understand current and predict future of ABC Emirates bank business scenario. But use of internet, evolving technology are constantly progressing and it is difficult to say that any implemented system is perfect and working efficiently. However, there are some ideas which need to exploit more such as banking system will completely transform in to [Digitalized banking](#) (which is still not yet done 100%). With digital banking is will be easy to introduces "Banking as a service" BaaS. Banking sector is more service oriented business and serves the customer is always its first priority, So Digital banking will open more doors of online accessibility and definitely new doors will open for big data and Data mining technologies (Still there are requirements to comprehend, more data islands containing unstructured data such as cloud computing and IoT). The complete accessibility of Customers 'data will enhance the opportunities of implementing Artificial intelligence using deep learning for predictive analysis and in future this advanced level of data and system will provide us platform for prescriptive analysis.

There is also one area where we need to put more efforts and that is Analytical tools and trends. Mining of unstructured data is progressing day by day and to see the End results there should be more robust Platforms with more functionality to represent such a huge data. In this report, in CRM section, still there are corners which can be turned over to work on it for more Powerful and robust system such as Customer churn management (preventing customers from switching banks). However obtaining meaningful insights from unstructured data is challenging and requires human interventions to explore the power of these technologies, which can do calculations but cannot interpret the outcomes.

8. References

1. Heffernan, S. (2005). Modern Banking. In S. Heffernan, Modern Banking (pp. 1-682). England: John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ,
<https://www.academia.edu/35734316/Heffernan - Modern Banking>
2. Wizards, A. (2016, Jan). *Top 9 Data Science Use Cases in Banking*. Retrieved from ActiveWizards:
<https://activewizards.com/blog/top-9-data-science-use-cases-in-banking/>
3. Flair, D. (2018). *Intriguing Applications of Data Science in Banking*. Retrieved from Data Flair: <https://data-flair.training/blogs/data-science-in-banking/>
4. Hossein Hassani 1, X. H. (2018). *Digitalisation and Big Data Mining in Banking*. UK: MDPI.
5. Chandershekher Joshi, L. D. (2013). *Leveraging Unstructured Text Data for Banks The imperative mining unstructured data*. India: ResearchGate.