Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 50 (2015) 643 – 652

Procedia Computer Science

2nd International Symposium on Big Data and Cloud Computing (ISBCC'15)

Impact of Big Data Analytics on Banking Sector: Learning for Indian Banks

Utkarsh Srivastava[a], Santosh Gopalkrishnan[b]*

[a]*MBA-I (Finance), Symbiosis Institute of Business Management, Symbiosis International University, Lavale, Pune 412115, India*
[b]*Research Scholar, Dr. D.Y. Patil University Pune & Assistant Professor, SIBM, Symbiosis International University, Lavale, Pune 412115, India*

Abstract

The big data revolution happening in and around 21st century has found a resonance with banking firms, considering the valuable data they've been storing since many decades. This data has now unlocked secrets of money movements, helped prevent major disasters and thefts and understand consumer behaviour. Banks reap the most benefits from big data as they now can extract good information quickly and easily from their data and convert it into meaningful benefits for themselves and their customers.

Banks internationally are beginning to harness the power of data in order to derive utility across various spheres of their functioning, ranging from sentiment analysis, product cross selling, regulatory compliances management, reputational risk management, financial crime management and much more. Indian banks are catching up with their international counterparts; however a lot of scope remains.

This paper aims to capture how big data analytics is being successfully used in banking sector, with respect to following aspects:

1. Spending pattern of customers
2. Channel usages
3. Customer Segmentation and Profiling
4. Product Cross Selling based on the profiling to increase hit rate
5. Sentiment and feedback analysis
6. Security and fraud management

* Corresponding author. Tel.: +91-976-654-5439; fax: +91-20-391-160-60.
E-mail address: santoshgopalkrishnan@sibmpune.edu.in

Peer-review under responsibility of scientific committee of 2nd International Symposium on Big Data and Cloud Computing (ISBCC'15)
doi:10.1016/j.procs.2015.04.098

The data used is secondary data from a bank while the analysis is of primary nature. This study reveals some of the best practices being adopted by banks globally, and can be replicated by Indian banks to enhance their financial service offerings to customers.

Peer-review under responsibility of scientific committee of 2nd International Symposium on Big Data and Cloud Computing (ISBCC'15)

Keywords: Big data Analytics; Banking; Fraud Management; Feedback Analysis; Customer Satisfaction Index Measurement; Product Cross-Selling; Customer Behaviour Analysis; Spending Pattern Analysis; Transactional Analysis

1. Introduction

The big data revolution happening in and around 21st century has found a resonance with financial service firms, considering the valuable data they've been storing since many decades. And even though the collection of this data was unplanned, since accounting system has always been historical in nature, the potential unlocked by big data analytics exceeds any expectation previously expected from this historical record set. This data has now unlocked secrets of money movements, helped prevent major disasters and thefts and understand consumer behavior. Banks reap the most benefits from big data as they now can extract good information quickly and easily from their data and convert it into meaningful benefits for themselves and their customers.

Financial firms are looking forward to application of big data in spheres like front office risk management to back office trade operations.[4] Before we delve into the most affected areas in BFSI, let us also have a look at what really is big data all about.

1.1. Impact of Big Data on Banking Institutions and major areas of work

Finance industry experts define big data as the tool which allows an organization to create, manipulate, and manage very large data sets in a given timeframe and the storage required to support the volume of data, characterized by variety, volume and velocity.[12]

Below we look at the major areas where big data is being utilized by financial institutions which are ramping up their enterprise risk management frameworks to help improve enterprise transparency, auditability, and executive oversight of risk.[5]

1.1.1. Customer Centric

Client experience closed feedback loop	Customer life event analysis
Next best offer	Real time allocation based offerings
Sentiment analysis-enabled strategy management	Sentiment analysis-enabled lead/referral management
Quality of lead analytics	Micro-segmentation
Customer Gamification	Sentiment analysis-enable sales forecasting

1.1.2. Risk Management

Following are the ways in which data analysis is being used to find out and evaluate financial crime management (FCM) solution rules, by early detection of the correlation between financial crime and attributes of the transaction, or series of transactions.

MIS/ Regulatory reporting	Disclosure reporting
Real time keyboard conversation tracking	Anti-money laundering

1.1.3. Transactions

Transactions and trading, when followed over a period of time, tend to reveal a lot of information about the nature of trade, log analytics, trading sentiments and other aspects. Banks and other financial institutions leverage Big Data under this header in following ways –

IVR analysis	B2B merchant insights
Real time capital calculations	Log analytics

2. Use case from Banking Sector - Problem statement and available data

The bank under consideration is a bank in the middle-east. Identity has been concealed to prevent confidential information from leaking out, and henceforth we will refer to this bank as XYZ Bank. It's been in operation since the past 20 years, and has had trouble reviving its profit margins post the 2008 financial crisis. From 2011 onwards, they started collecting customer feedback in order to understand and fix issues with the functioning of bank.

In 2013, they experienced a dip in their customer satisfaction measurement, along with which their customer retention also dropped. We have been given the project by the Bank to perform the following:

- Determining the root cause of drop in customer satisfaction measurement.
- Analyze the spending patterns of their card holders (4 cardholders as a subset)
- Channel usage analysis – debit/ credit descriptions, as well as payment modes – ATM, cards
- Behavior of a customer and product cross selling

For our case, the following points have been taken into consideration:

- Transactional data for 4 cardholders (set of around 5000 records), for the time series January 2011 – June 2014.
- Access available to 20,000 records stored with a third party responsible for collecting feedback for XYZ Bank.

2.1. Methodology

We begin with analyzing the customer satisfaction measurement data provided to us. This will also help us understand if the issues XYZ Bank was facing were due to poor services or some other issue.

After segmenting the issue with help of feedback analysis, we will try and figure the reason why issue happened and propose improvements.

We will also do customer segmentation and propose suitable products which can be sold to a customer, based on their type.

3. Analysis and Inferences

3.1. Feedback Analysis

Feedback processes are important for any organization to help and understand the potential areas of improvement and if done on a regular basis, they help to identify gaps in services rendered. XYZ Bank also started to collect feedback from their customers; from those who visited bank branches as well as from those who used online services.

3.1.1. Data Collection and Sample Size

Following data was gathered and accumulated over a period of 3 years 6 months. Customers visiting any branch of XYZ Bank were asked to rate the bank anonymously on a scale of 1 – 5 on the following parameters:

- Is the customer happy with the quality of service?
- Is the customer happy with the speed of service?
- Are customer queries addressed effectively?

The analysis below is performed using the pertaining subset of the total data collected, comprising of feedback from around 20,000 customers.

When we plot the data, there are some curious findings

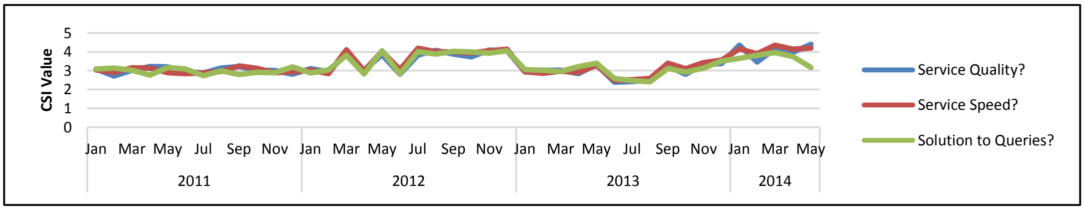
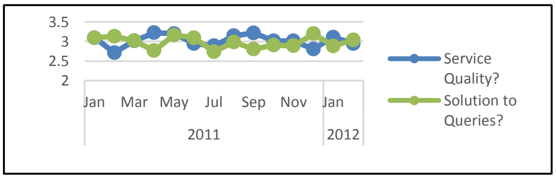


Fig. 1. Overall Customer Feedback for provided parameters

3.1.2. Feedback Analysis and Inference

- The ratings received prior to February 2012 are fairly stable and low. Service quality, service speed and effective addressing of queries were all ranked with equal weightage.
  Inference – The customers rated bank services as average and the bank did not take up any corrective measures during this period to improve its customer ratings.



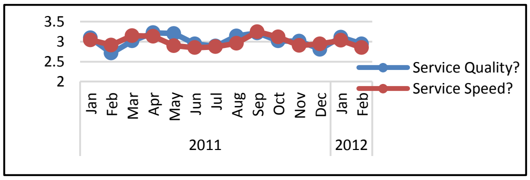


Fig. 2.(a) “Service quality” vs “Solution to queries”; (b) “Service quality” vs. “Service speed”

- However, between Jan 2011 and Mar 2012, “Service Quality” and “Effective addressing of queries” have a negative correlation = -0.38922.
  Inference – This means that service quality was not being reflected in the resolution to queries being provided in the interaction with customers, as is evident from following graph below -
- During the same period “Service Quality” and “Service Speed” have a positive correlation = 0.66965, while the overall rating remained average for both the factors.
  Inference – It signifies that the quality was a direct reflection of speed of quality delivery. Whenever the speed of delivery improved, the customers perceived corresponding improvement in the quality of service as well. This implies that customers perceive speed as a quality parameter.
- There is some improvement in the service from January 2012 to April 2013; however, there is a significant drop in the ratings from May 2013 onwards.
  Inference – Occasional spikes indicate the bank does take some measures to improve the customer satisfaction index, and it is evident as the average ratings improve during this period. If we have a closer look at the 3 months Apr 2013 – June 2013, it is observed that the average customer ratings drop by a factor of 3

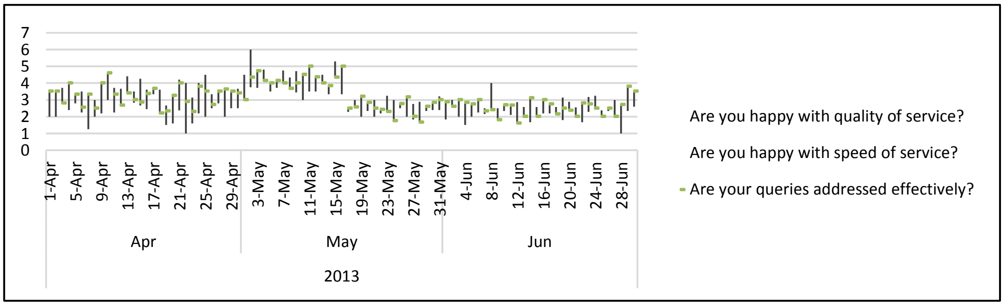


Fig. 3. Customer Satisfaction Measurement – High low indicators

- From Aug 2013 onwards, the ratings recover and gradually improve.
  Inference – Drop in ratings signified some breach of trust of customers, as customers sentiments are a direct measurement of the perception of service. Hence the concerned bank took some measures to correct the issue, and this resulted in ratings improvements.

3.2. Transactional Analysis

The next section of the study will try and isolate the root cause of the drop in customer satisfaction ratings for bank, as well as evaluate and look at various strategies used in analytics. As mentioned above, the following will be the basis of this part of the study:
The XYZ Bank Dataset comprises of the Transactional History of 4 cardholders from January 2011 – May 2014, and will be evaluated as per the heads given below.

3.2.1. Analysis of spending pattern

Let's have a look at the general card usage per month per year over the time series from 2011 – 2014. We will analyze the net transaction value for cards, and proceed to inspect the trends in credit patterns and debit patterns of our cardholders.

3.2.1.1. Total Account credit amount per month per year

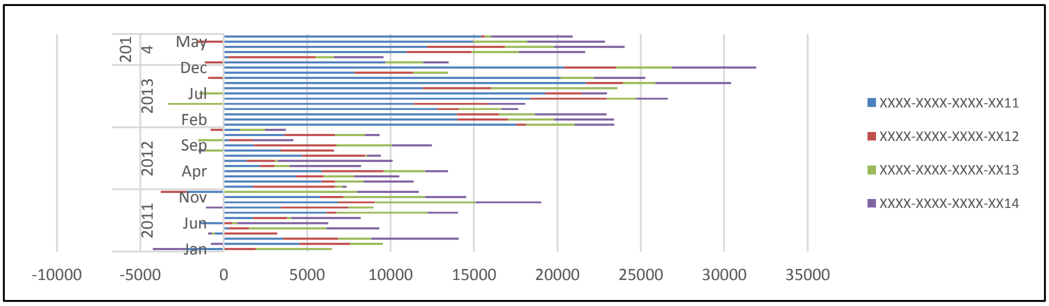


Fig. 4. Net Transaction per card

Observations -

- As we can see, from 2011 to 2014, there has been a gradual increase in volume of money being circulated in the system. We also observe that for some cards, while the money transaction volume has increased, for others it has remained fairly consistent.

- There are some cards, for which the net debit exceeded net credit in the system (this does not mean that the account balance was sub-zero, rather, was negative, and as a result, their net spending was more than net earnings during that period, as is evident from (1)

$$NetTransactionValue = TotalCredit - TotalDebit \tag{1}$$

We will now break this up and analyze the net credits and net debits per card per month.

3.2.1.2. Total Account credit amount per month per year

We have plotted the total account credits per month for the card holder accounts.

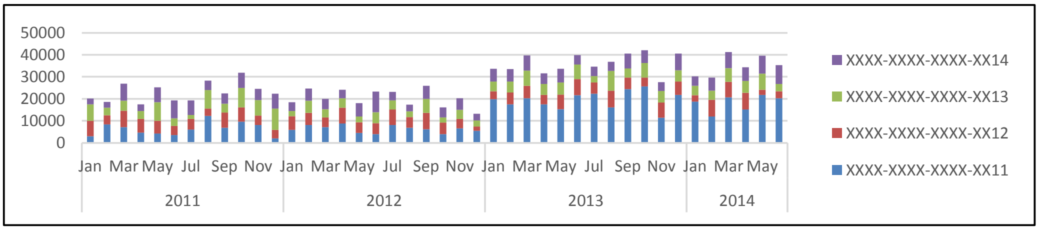


Fig. 5. Net credits per account

Observations:

- There has been a gradual increase in net money flowing into accounts. This goes in sync with the way monetary assets increase with time for any financial institution in general.
- For one card holder account, there is noticeable increase in money being credited on monthly basis.

Inference:

- This card holder could be a regular salaried person, with an annual increase in his take home. The fluctuations in the credit may be due to a variable component in the salary, however, the basic salary would have increased in this case. Identification of stable regular salaried card holders can further benefit the banks as these card holders can then be approached with better savings and liability products and schemes.
- Rest of the card holders could be contractual employees, or have staggered sources of income. Identification of non-salaried card holders can help bank create and deliver products such as small savings plan and fixed deposit plans with attractive returns.

Let's also look at the net debit amount per card per month per year:

3.2.1.3. Total Account debit amount per month per year

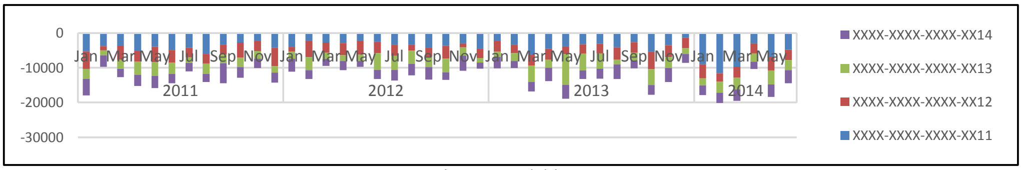


Fig. 6. Net debits per account

Observations –

- There is a gradual increase in net spending per account per month per year.

- There is a possible cyclicality and/or seasonality in the spending patterns. This can be dependent on a lot of factors, like
    - macro-economic conditions,
    - festive seasons,
    - income sources of the entities under observation
    - spending habits of the entities under observation

3.2.1.4. Consumer Behavior Analysis based on Channel usage analysis

Transactions can occur via many modes and channels, like ATM, online transactions (also called CARD NOT PRESENT, or CNP Transactions), swipe/ Chip and Pin (also called CARD PRESENT or CP transactions). The nature of transaction is also considered an important parameter for understanding the needs, and habits of a customer. Following transaction types have been taken into consideration for the case:

- Credit Transaction
    - Allowance/ Regular Salary
    - Account transfer
- Debit Transaction
    - ATM Withdrawal
    - Card debit (card swipe/ chip and pin)

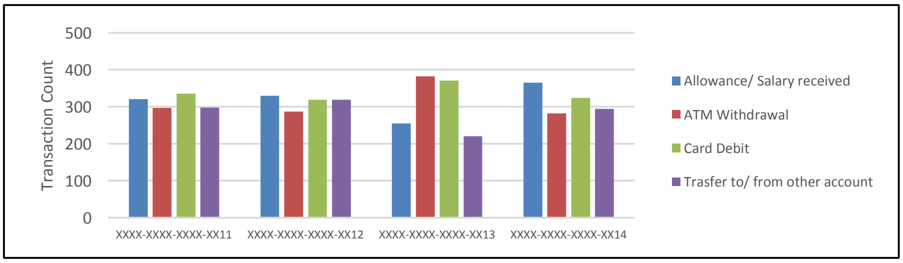


Fig. 7. Channel Usage and Transactions Types

Observations:

- Allowance received and transfers received amount to total incoming (credit) transactions and ATM withdrawals and card debits amount to total outgoing (debit) transactions
- Customer behavior analysis –
    - Card numbers ending with numerals 11 and 12 are someone for whom net credit transactions count is slightly more than the net debit transactions count. He or she is most likely to be a regular/contractual employee, and is expected to spend money majorly only as and when received.
      Inference: This person can be sold a savings plan, or a micro-investment plan, which would give him/her decent returns on maturity.
    - Card number ending numeral 13 is a person for whom net credit transactions count is almost half of net debit transactions count. This person could be a regular income earning person, since he/she is expected to receive salary credit at the beginning of month (usually monthly, or bi-monthly or weekly basis, since the data is for middle east, and employees are also paid in bi-monthly or weekly frequencies), and may receive infrequent transfers from others.
      Inference: This person can be sold an investment plan, since his income is regular, and depending upon his debit frequency, can be sold a credit card offering high credit capabilities.

3.2.2. Consumer Behavior Analysis based on consumption patterns for Cross Selling

We can also use the transactional data to estimate which consumers can be sold what types of financial products, and is used very frequently by banks to segment and target potential clients. Below is the compilation of data from our dataset, for purpose of understanding consumer behavior for cross selling and upselling financial products to customers.

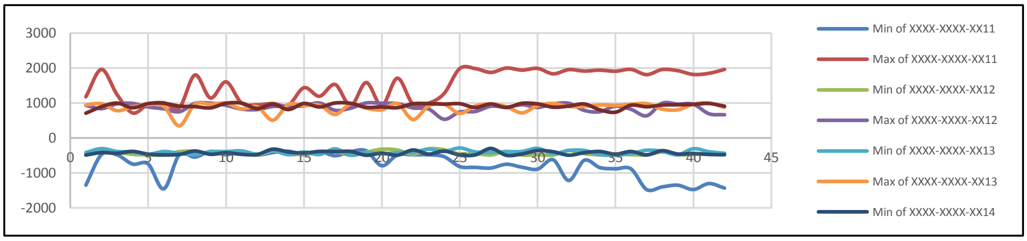


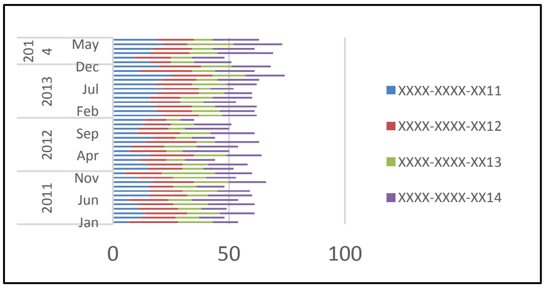
Fig. 8. Spending and credit pattern – Consumer Behaviour

Observations:

- The graph above shows the maximum debit and maximum credit patterns for the cardholders.
- Consumer behavior Analysis:
    - Card ending 11: As is observed, with shift in timeline, the net credit into the account of these cardholder increases, which improves his spending capacity; and, as his spending capacity increases, the debit activity also increases.
    - Not only his spending amount increases but also his spending frequency increases.
- Based on consumer behavior analysis, we can infer the following –
    - This person has a capacity to spend, and though infrequently, he has spent more than what is credited to his account during certain periods. So this person is ideal candidate for potential loan applicant.
    - This person also shows that as his capacity to spend increases, the net debit also increases. Hence this person is also ideal candidate for using credit card. He can be offered a credit card or depending upon if he is already using one, his credit limit can be increased.
    - Credit card linked offers can be extended to this person since he is more likely to use his card.

3.2.3. Security and Fraud Analysis

Based on historical transactions and consumption capacity of customers, coupled with the behavioral analysis can help us reveal a potential threat to the system, as well as uncover frauds that might have happened in the past.



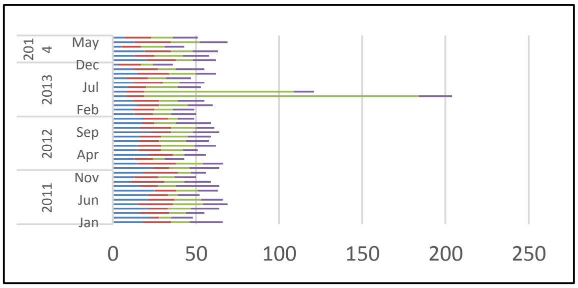


Fig. 9. (a) Net credit transactions count; (b) Net debit transactions count

3.2.3.1. Net Credit Transaction counts per month per year

Observations:

- Net count of credit transactions increase per card with timeline; although the increase is subtle.
- The net ratio of credit transactions with respect to previous month remains the same.

3.2.3.2. Net Debit Transaction counts per month per year

If we look at the debit transactions that happened during the same time, it leads to some interesting outcomes. Following is the plot of all the net debit transactions for respective card holders.

Observations:

- Net transactions count grows with time and subtly.
- In the month of May and June 2013, card ending 13 shows spike in number of transaction count. The number of transactions more than doubled during the period for this card holder. Normally this is where the analyst should sound an alarm. If we are to upscale our small dataset to include millions of card holders, such spikes are dangerous and can mean a potential compromise of the system. It clearly indicates a misuse of the Card by miscreants and unauthorized access of funds by unscrupulous agents.

We will now have a look at transaction time trends for all the transactions in 2013.

3.2.3.3. Transaction Time Trend Analysis for the year 2013

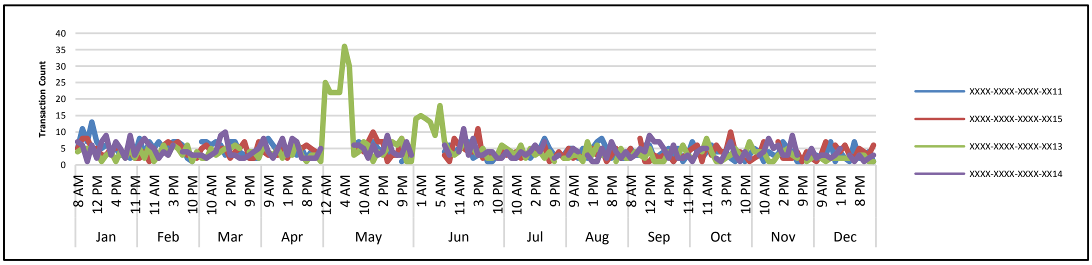


Fig. 10. Transaction Time trend per month for the year 2013

Observations:

- In general, card transactions are expected to occur between 0800 hours to 2300 hours, since that is the period within which normal businesses operate. The graph shows similar observations.
- However, as is evident from the graph, for card ending 13, during the month of May and June, there is a spike in transaction count. Not only this, but the transactions happen during the interval of 2300 hours at night to early 0500 hours in morning, which is abnormal. This could be a hint of unauthorized online transactions by card holder impersonation. In other words, there is a very high probability that this is a case of hacking.
- Given the small size of dataset, it could also mean that if this dataset were to be expanded, there is a high probability that this could be on a mass scale.

3.3. Correlation between observations above and Customer Satisfaction Index analysis

- In early hours of day, during the month of May and June, the bank experienced cyber-attack during which cards were used to make purchases. The hacking activity paused after 16th May 2013 and subsequently ceased to exist around 10th June 2013. Almost immediately, from 17th May 2013, the customer satisfaction measurement index for bank dropped by 3 indices. This means the issue happened at a widespread level and customers were affected in large numbers.
- Bank employed increased protection of its online system to prevent further fraud transactions. This is evident from the fact that no fraud transactions have been observed after the incident.

- Bank also worked towards pacifying the troubled customers, and did activities related to brand damage control. This is also evident from the fact that subsequently after the incident; the customer satisfaction measurement index has been rising steadfast and has exceeded the level pre- May 2013.

4. Conclusion

Big Data analytics is now being implemented across various spheres of banking sector, and is helping them deliver better services to their customers, both internal and external, along with which is also helping them improve on their active and passive security systems. This study analyzed transactional and sentimental analysis for the Banking Sector, and the outcomes of the same are mentioned below:

- We saw one of the ways how customer sentiments are captured and used to assess functioning of the bank. There are many more ways banks and other financial institutions have started to capture customer related data for sentiment analysis, starting from social media websites to various market research channels.
- We observed transactional analysis and observed how banks today use spending patterns of their customers, perform consumer behavior based on channel usage and consumption patterns and segment consumers depending upon the aforementioned attributes, and identify potential customers for selling financial products.
- Most of these indications can be implemented easily into the financial systems used at banks, which can help banks strengthen data security and prevent any type of attack. While some of the checks, like suspicious wire transfers may result in "false positives"; combining several such transactional and sentimental indicators to arrive at a holistic decision making approach and thereby implement sophisticated mechanisms is certainly the need of the hour for the Banking sector.

5. Future Scope

This study can be further extended into trying and quantifying the financial and non-financial benefits that XYZ Bank reaped after their implementation of Big Data Analytics and predict the improvements in financial statements of the bank. This work can also be extended to cover the various data mining techniques that can be used by banks to improve the analysis quality.

References

1. QATCH Regulations, Direct Credit and Debit System Annex No. (153), March 2010
2. OCBC Bank nets profits with interactive, one-to-one marketing and service. IBM Software: Smarter Commerce; July 2012
3. Business Software themes for 2014. Wells Fargo Securities. Equity Research; 2014
4. David Turner, Micheal Shroeck, Rebecca Shockley. Analytics: The real-world use of big data in financial services. IBM Institute of Business Value in collaboration with Said Business School, University of Oxford; 2013.
5. How can financial services industry unlock the value of big data. PricewaterhouseCoopers; 2013.
6. Big Data: The next big thing. Nasscom; 2012.
7. David Floyer. Financial Comparison of Big Data MPP Solution and Data Warehouse Appliance; 2013.
8. Oracle Financial Services. Initial Steps on the Journey through Big Data for Financial Services Institutions; 2012
9. Pivotal Case Study – China CITIC Bank. Driving Revenue and Reducing Risk; 2013.
10. PWC. FS Viewpoint - How the financial services industry can unlock the Big Data; 2013.
11. Emcien. Big Data for Banks - 5 Profitable Use Cases
12. Meta Group. Application Delivery Strategies.; February 2001
13. Sobolevsky et al. 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; 2014.
14. Dhar Vasant. Big Data. 2(2): 65-67. doi:10.1089/big.2014.1524; 2014.
15. Arvid O.I. Hoffmann Cornelia Birnbrich."The impact of fraud prevention on bank-customer relationships", International Journal of Bank Marketing, Vol. 30 Iss 5 pp. 390 – 407; 2012.
16. Katherine J. Barker, Jackie D'Amato, Paul Sheridon. "Credit card fraud: awareness and prevention", Journal of Financial Crime, Vol. 15 Iss: 4, pp.398 – 410; 2008.
17. Cloudera. Identifying Fraud, Managing Risk and Improving Compliance in Financial Services; 2013.
18. Capgemini. Big Data Alchemy: How can Banks Maximize the Value of their Customer Data?; 2014