International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

How to Improve Efficiency of Banking System with Big Data (A Case Study of Nigeria Banks)

Sanni Hafiz Oluwasola

Abstract: ***In banking industry today which their data has now turn to what we call Big data, some banks has now started making advantages of these big data to reach the main objectives of marketing. The banking industry can use the data to increase their efficiency by identifying the key customer, improving the customer feedback system, detect when they are about to lose a customer, to enhance the active and passive security system and efficiently evaluating of the system. This paper focus on different analysis and algorithms the banking industry can use to achieve all the advantages of these big data especially Nigeria banking industry. Analysis such as Link analysis, survival analysis, neural analysis, text analytics, clustering analysis, decision tree, sentiment analysis, social network analysis and datammer for predicting the security threat.***

1. Introduction

What is big data?

According to Wikipedia, Big data is a term that describe large volume of data both structure, semi structure and unstructured data. The bank data has turn to a big data because of it volume(terabyte Record, Transaction, tables and files) , Velocity(Batch, Real time and stream), variety(structure, semi structure, un structure data and all of the above) has now exceed the abilities of what IT system can ingest, store, analyze and process it in time.

Big data does not always fit into tables of columns, there are many new data types both structure, unstructured, semi structure that can be processed to yield insight into a business or a condition like data from twitter feed, blogs, third party data, call detail report (audio), network data video from surveillance cameras and equipment sensor often isn't store in a data warehouse. In 2012 Gartner updated the definition of 3V model (variety, volume and velocity) to (high variety, high volume and high velocity) due to tedious data in a big sector such as Bank industry. Gathering large amount of data from different source makes big data very powerful for bank industry to make effective decision faster and better than traditional business tools. The technology behind big data is Hadoop ().

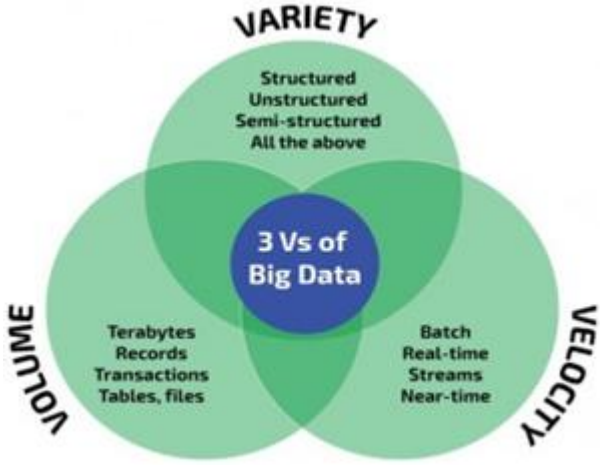


Figure 1: 3V Model

Advantages of big data in banking system

- Customer segmentation: is the way of grouping a customer that share the same interest together
- Risk management: it prevent fraud because no unauthorized transaction will be made and massive information play a big role desegregation the bank's needs into a centralized, practical platform.
- Efficient evaluation of the system: by using big data, banking industry will provide exact evaluating report.
- Improving the Customer feedback: banks will able to identify the exactly problem of a customer in time and provide the accurate services promptly.
- Identify the key customer. They will be able to identify the key or important customer to the bank.
- Security enhancement: it easily detect fraud and also improve the active and passive security.
- Detect when customer about to leave: it will notify if customer is about to leave the banking hall.
- Sentiment analysis: it help to monitor the customer satisfaction.

2. Methodology

How to implement the above advantages in to the banking system, Nigeria banks as a case of study.

Monitoring the customer satisfaction with sentiment analysis

With aid of **sentiment analysis** it is possible to understand the customer better and know what the customer is saying about the bank e.g. services or product. Taking NED BANK LTD in South Africa as example, they realize a great advantage of what the customers is saying about the bank with aid of social media analytics But Nigeria Banks mainly make use of impractical method which is not highly efficient because they have to read all the comments on the blog, review site and direct report which is impractical for them to read it all but with aid of the sentiment analysis (social media analytic) since the data are in text format, text analytical algorithm is suitable for these problem such as Naïve Bayes Algorithm. These algorithm will just analyze the text and produce the highly negative to neutral and to highly positive. The scoring system will assist the bank to know if the customer is happy without reading all there comment on the social media such as blog, review site, twitter, Facebook etc.fig 2.0 shows how text analytical works.

Licensed Under Creative Commons Attribution CC BY



Figure 2: Extracting information from social media

Identify the key customer of the bank

If the bank can be able to identify their key customer such as those that have been flagged for high number of referrals. Those customer are important to the bank because they are enjoying banking services and products of the bank, they spread the information to other customers and interact with them. Such customer can be identified with the help of link analysis which depend on highly unstructured social media network data and data from the third party such as blog, review site and twitter, decision tree can also be used to determine how customer are relating with each other and scoring the customer from highest to lowest in ranking order.

Improving the customer feedback

Nigeria banking industry examine the customer feedback. They use a research tool such as customer survey and it is not only time costuming but also not accurate because it sample some groups but if sentiment analysis tools is brought into the system it will be able to gather all the customer report on social media platform, log and blogs. Taking Barclays as an example, Barclays was able to gather the report about their mobile app problems from social media with the aid of sentiment analysis tools like social media analytics and they were able to make amendment immediately also dell computer manufacturer gather report from the world wide web about the dell product they launch that have an overheating problems as a result of social media analytic tools that they used were they able to know these problems immediately. If Nigeria banks can make use of the method they will be able to improve the customer feedback.

Identify the customer profile

Nigeria banks sent a general mail to every customer which people are now treating the mail as a spam by not reading the mail any longer for example sending information about kids account to an adult etc. but if the bank can use a recommendation that can determine which product message is suitable for the customer, using *classification algorithm such as neural network or decision tree* which can help us to determine what content customer is interested in. Determine the products or services of the bank the customer is more interested or enjoying, bank can determine what product customer is more engage in by using cookies URL to determine the links the customer clicked most.

Detect when customer is about to leave

The problems banks are facing is retailing their customer taking union bank as an example, these bank is among the banks that have highest number of customer before, but as a result of not knowing when the customer is leaving the business make them loose almost their customers. It will be easier if banks understand their customer in a holistic way. Customer data transaction can be used to know how customer is banking with them, sentiment analysis can be flag if customer is not happy the best method to use to know when customer is leaving, is survival analysis the reason being that survival analysis compare different customer segment across time series

Customer segmentation

Customer segmentations is now beyond the traditional segmentation that group a customer base on the account type, account balance, age, marital status etc. customer segmentation is more than that we can have a prediction system which can group the customer base on life style, life stage and special event like what customer use to buy online with the credit or debit card. *Clustering algorithm* is the best way to do the advance segmentation.

Licensed Under Creative Commons Attribution CC BY

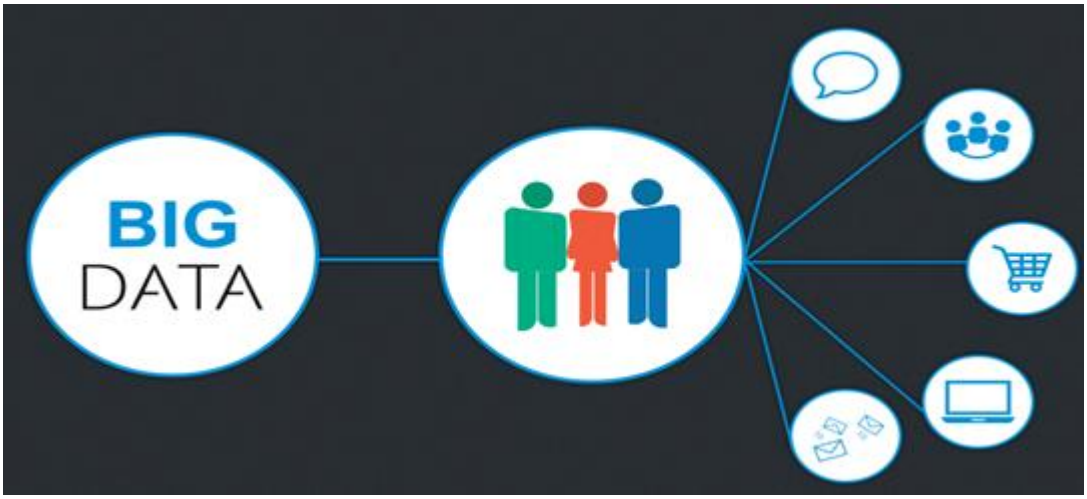


Figure 3: Group of customers with common interest (customer segmentation)

Security enhancement
Security of the data is consider paramount in the banking sector, banks are trying in the security part by using strong security protection for both their data and the network but it will be more better if the bank has already know that threat is coming and its nature for example a certain company make use of Datameer to follow a virus that started in Russia, moved all across the Asia continent then to United states. If the banks can incorporate a Datameer to their system it will flag them when the virus or threat is coming because Datameer will follow the trend of how the virus is moving from particular geographical area to another so it can predict the next move of the threat.

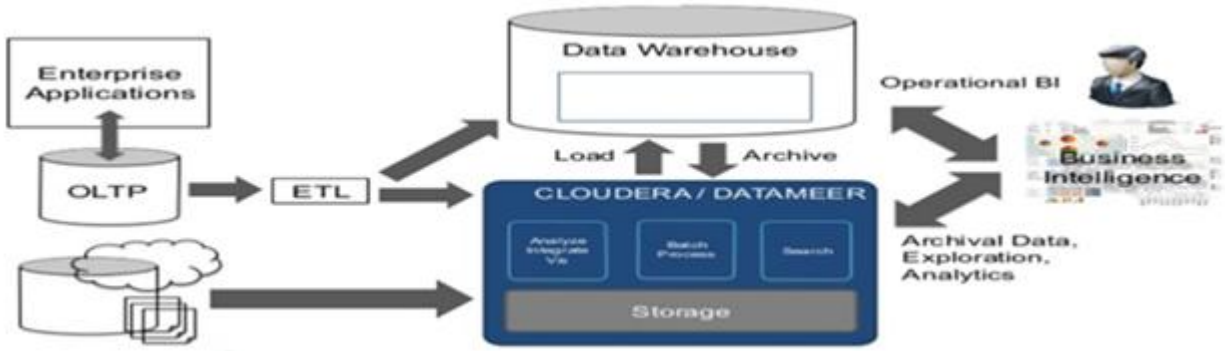


Figure 4: Datameer

3. Analysis and Algorithms

Survival analysis
Things need to consider when working with survival analysis (when customer will leave)

- Customer transaction records
- Customer relation records
- Customer feedback

T denotes the response variable, $T \geq 0$.
The survival function is $S(t) = Pr(T > t) = 1 - F(t)$.

The survival function gives the probability that a subject will survive past time t.
As t ranges from 0 to ∞, the survival function has the following properties
It is non-increasing
At time t = 0, S(t) = 1.
In other words, the probability of surviving past time 0 is 1.
At time $t = \infty$, $S(t) = S(\infty) = 0$. As time goes to infinity, the survival curve goes to 0. – In theory, the survival function is smooth. In practice, we observe events on a discrete time scale (days, weeks, etc.).

The hazard function, h(t), is the instantaneous rate at which events occur, given no previous events.
$h(t) = \lim \Delta t \rightarrow 0$
$Pr(t < T \leq t + \Delta t | T > t) / \Delta t = f(t)\ S(t)$.
The cumulative hazard describes the accumulated risk up to time t, $H(t) = R_t\ 0\ h(u)du$.

Sentiment analysis
Is an algorithm that turned to analyze the sentiment of social media content, like tweet and status updates and return sentiment rating for negative, neutral and positive.

It will extract all the text(tweet, updates etc.) by *text analytics* then
It will check through all the tweet, status updates of customers then it will return the resulting rating inform of positive, neutral and negative.

Classification algorithm (neural)
Neural networks are a classification type algorithm, which means they assign data to a predefined target field. The target field would typically be a scoring function, such as the customer's propensity towards a particular product, or an estimated value for a house

Licensed Under Creative Commons Attribution CC BY

Clustering algorithm

Cluster detection helps penetrate this noise by finding clusters of data that form natural groupings within the data set. For example, one such cluster might contain grouping a customer base on life style, life stage.

The steps of the K-means algorithm are given below.

1) Select randomly k points to be seeds for the centroids of k clusters.
2) Assign each point to the centroids closest to the point.
3) After all points have been assigned, recalculate new centroids of each cluster.
4) Repeat step 2 and step 3 until the centroids no longer move

Datameer

Is an end to end big data platform, which ignores the limitation of ETL and static schemas to empower the bank to integrate data from any source into Hadoop. It has pre built data connector wizards for common structured, semi structured and unstructured data sources, data integration is totally simplified. It ensure that user are usually up to date and provide the link to any other source. It integrate the data fast and easy.

4. Conclusion

The technical key to successful usage of Big Data and digitization of business processes is the ability of the organization to collect and process all the required data, and to inject this data into its business processes in real-time - or more accurately, in right-time. Big knowledge analytics is currently being enforced across numerous spheres of banking sector, and helps them deliver higher services to their customers, each internal and external, alongside that is additionally serving to them improve on their active and passive security systems. If the banking industry can implement all the above analysis and algorithm effectively (sentiment, clustering, link, survival, decision tree or neural analysis and datameer) it will increase the efficiency of the banking system.

References

[1] High-performance big data analytics, computing sysem and approaches by pethuru Raj, Anupama, Dhivya Neganaj, Siddhartha Duggirala.
[2] K-Mean Clustering Algorithm Implemented To E-Banking ,Kanika Bansal Banasthali University, Anjali Bohra Banasthali University, Vol. 2 Issue 4, April – 2013
[3] WHITEPAPER Big data in banking for marketers, Evry INNOVATION LAB
[4] Introduction to Survival Analysis, BIOST 515, February 26, 2004
[5] Lectures on Survival Analysis, Richard D. Gill Mathematical Institute, University Utrecht, Budapestlaan 6, 3584 CD Utrecht, Netherlands.
[6] Stable Algorithms for Link Analysis, Andrew Y . Ng Computer Science Division U.C. Berkeley, Alice X. Zheng Computer Science Division U.C. Berkeley, Michael I. Jordan CS Div. & Dept. of Stat. U.C. Berkeley.
[7] Impact of Big Data analytics on banking sector, Abhinav kathuria, ISSN: 2278 – 7798 International Journal of Science, Engineering and Technology Research (IJSETR) Volume 5, Issue 11, November 2016
[8] OCBC Bank nets profits with interactive, one-to-one marketing and service. IBM Software: Smarter Commerce; July 2012
[9] Big data tutorial, Marko Grobelnik marko.grobelnik@ijs.si Jozef Stefan Institute Ljubljana, Slovenia Stavanger, May 8th 2012
[10] Social media analytics, Mathew Ganis, Avinash Kohirtar, 2015
[11] https://en.m.wikipedia.org/wiki/social_media_analytics
[12] http://en.wikipedia.org/wiki/Data_mining
[13] http://www.slideshare.net/aramshaw/implementing-triggerbasedmarketingtodrivecustomerloyalty
[14] http://www.mckinsey.com/App_Media/Reports/Financial_Services/Retail_Banking2010_Relationship.pdf
[15] http://ias.cs.unb.ca/research.html
[16] http://www.nornir-insight.com/datameer.html
[17] http://sci2s.ugr.es/Bigdata
[18] http://www.qubole.com/blog/big-data-customer-microsegementation

Author Profile

Sanni Hafiz Oluwasola did B.sc Information technology, Kebbi State University of Science and Technology, Aliero, Kebbi State Nigeria. He can be reached at Computer Science Department, Lens polytechnic offa, Kwara State Nigeria

Licensed Under Creative Commons Attribution CC BY