Convolution Neural Network to Uncover Bank's Hidden Patterns and unknown Correlation

Mr. RUKUNDO Prince
Business Information Technology
University of Rwanda
rukuprince@gmail.com

Dr. MBANZABUGABO Jean Baptiste
Business Information Technology,
University of Tourism Technology and Business Studies
s2baptiste@gmail.com

NSABUMUHIRE Innocent
Business Information Technology Dept.
University of Rwanda
nsabumuhirei@gmail.com

KAGEMANA Gaston
University of lay Adventists of Kigali (UNILAK)
gastonkagemana@gmail.com

ABSTRACT: ***This paper sought to study the relationship among financial institutions information with high degree of organization (structured data) such as customer's available balances, their loans, mortgages, investment, spending etc... and Unstructured data such as customer's tweets, Google search, YouTube watch videos, face book likes, received and sent e-mails etc...Most of these structured data are being kept in bank's warehouses and servers with purpose of retrieving, sharing with different bank's cite and sometime forecasting their interest in the future. Few banks will take advantage of customer's unstructured and structured data as unit to uplift their business values based on customer's preference and satisfaction to enhance business solutions as well..***

KEYWORDS: ***Big data, data mining, data analytic, unstructured data, structured data, hybrid algorithms, patterns, correlation, customer preferences, business solutions***

I. INTRODUCTION

Big Data is being known as extremely large data sets that may be analyzed computationally to reveal patterns, trends, and associations especially relating to human behavior and interactions, whereas Data Mining is the technology to extract the knowledge from the data. It is used to explore and analyze the same. The data to be mined varies from a small data set to a large data set which is being known as big data. The research conducted in 2012 by the world's information shown that the world's data is doubling every 1.2 years. It is known that 5.1 billion out of 7 billion people owns cell phones on planet, we send over 11 billion text messages, watch over 2 billion YouTube videos, Almost 5 billion Google search per day, the interesting part of it we are not consuming this data to finish them one day rather we create them, we are data creator with more than 2.5 quintillions bytes every day from consumer transactions, communication devices and streaming service etc.. 80% of these data are unstructured data (include text and multimedia content. Examples include e-mail messages, word processing documents, and videos, photos, audio files, presentations, Web-pages and many other kinds of business document), where Financial services such as Banks are one of the Big data producers through online services and apps, online payment and social media etc... Most of these generated data are mostly stored into physical Bank' storage devices and deployed on servers for being shared with different banks cite, retrieving, and sometime forecasting their interest in the future. They are rarely few banksResNe which take advantage of big data analytics and data mining algorithm to transform these resources into tangible information based business solutions.

Many algorithms have been used such as Genetic algorithm, support vector machines, decision trees, cluster analysis algorithms, HADOOP, SPSS, MATLAB etc… to handle both structured and unstructured data to disclose the hidden pattern inside this large set of data. Still it have been proved that such algorithms have weakness of analyzing structured and unstructured data as a single unit, processing speed capacity and extracting tangible knowledge based information for better decision making has been tested poor.

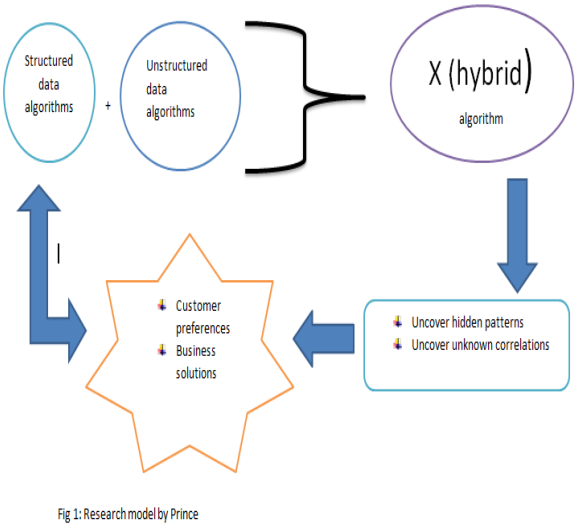


Figure 1: Work outflow using X hybrid algorithm.

II. REVIEW OF RELATED LITERATURE

many researcher have highlights the concern of working under a big data environment where bellow challenge have been putted in place ;

A. VOLUME

The main characteristic that makes data "big" is the sheer volume. It makes no sense to focus on minimum storage units because the total amount of information is growing exponentially every year. (Thomson, 2010) estimated in its annual report that it believed the world was "awash with over 800 Exabyte of data and growing."

For that same year, IBM release a study that in banking sector makes data storage devices, thought it was closer to 900 Exabyte and would grow by 50 percent every year. No one really knows how much new data is being generated, but the amount of information being collected is huge.

B. VARIETY

Variety is one the most interesting developments in technology as more and more information is digitized. Banking sector has adopted this approach where Traditional data types (structured data) include things on a bank statement like balances, loans, savings... These are things that fit neatly in a relational database.

Banking sector is now experiencing additional times of data which are not structured such as which doesn't fit neatly in a relation database such as Twitter feeds, audio files, MRI images, web pages, and web.

C. VERACITY

Through years trustworthy of data have been critical, banking sector has been also being victim of not taking enough attention to this approach, here Veracity refers to the trustworthiness of the data. Can the manager rely on the fact that the data is representative? Every good manager knows that there are inherent discrepancies in all the data collected.

D. VELOCITY

Velocity is the frequency of incoming data that needs to be processed. Banking sector is one of the most generated feedback SMS messages of deposited and withdraws, Facebook status updates, or credit card swipes are being sent on a particular telecom carrier every minute of every day. A streaming application like Amazon Web Services Kinesis is an example of an application that handles the velocity of data.

Data analysis and data mining are a subset of business intelligence (BI), which also incorporates data warehousing, database management systems, and Online Analytical Processing (OLAP).

The technologies are frequently used in customer relationship management (CRM) to analyze patterns and query customer databases. Large quantities of data are searched and analyzed to discover useful patterns or relationships, which are then used to predict future behavior. Meanwhile this algorithm works better to uncover hidden pattern and unknown correlation in some of the banking sector. Such of those algorithms are classified as Supervised,semi-supervised and unsupervised data mining algorithm.

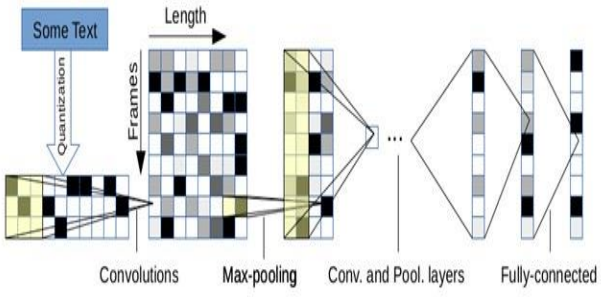
Most of these algorithms analyze data a s a single unit very few of them can handle both structure and unstructured data as a single units and if some of them does, handling the 4 V's (velocity, volume, veracity and variety) seems to be unmanageable . Hybrid algorithm will have a double advantage to tackle both structure data and match it with an unstructured data so that the velocity issues may increase at spot regardless of the data' variety.

III. PROPOSED METHODOLOGY

Text classification using Convolution neural network

The proposed model for the X-hybrid algorithm using convolution neural network to classifier both text and Image at the same time as their input layer data sets.

As the built model mention above bank's text data are being injected as input through a quantification



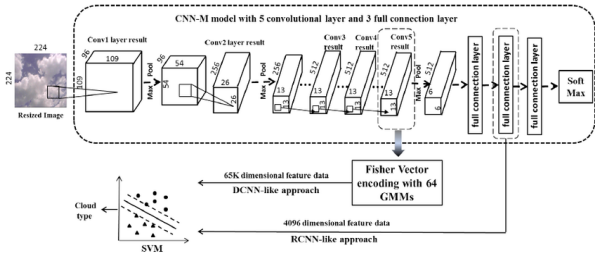
technique this in return convent the text into a matrix format where a max-pooling layer is used to reduce the amount of parameters and computation in the network, after the max-pooling layer, these text data are the being sent to a fully-connected layer which includes the forward and back-propagation process. This extract features and classifies text into similarities this assist the bank management understand peoples' overview of the required services so that the bank may act accordingly

Deep-learning convolution neural network Image classification (recurrent Neural network)

we adopted a Deep-learning methodology to classify images which have been posted by different customers on social media, either wanting or buying a particular product, so to extract some features and classify them. This with a unique purpose of learning how the bank can acquire more information of its customers image data

and link them with the services which should be deployed to them. Bellow highlights the model adopted.

after applying above two model of deep learning we



finally combine both these models to produce the X-hybrid algorithm which will both analyze both Structure and unstructured data by extracting features and classifying so to learn their property and predict the customer sentiments towards the provided data.

IV. METHOD OF INVESTIGATION AND SUMMARY OF FINDINGS

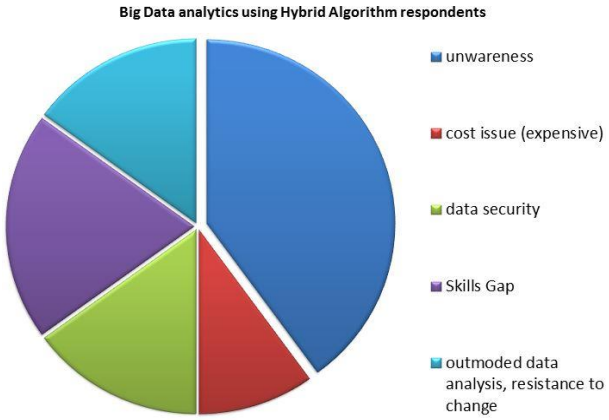
Systematic and proactive methods have been adopted in collecting information from a population of interest. It tends to be quantitative in nature and aims to collect information from a sample of the population such that the results are representative of the population within a certain degree of error free.

The data collection procedure was an In-depth interviews based with IT partners of financial institutions especially banking sectors in Rwanda, Universities and Business Companies, specific institution that implement Data analysis, IT consultants and ERP users, government and private organs in Rwanda that considers the adoption of Big Data analysis as a prior sector in today technological world, professionals comments from social media blogs is another source of data to be used in this research.

The other data sources are to be collected from the scholar studies and recent articles that focus on Big Data and analytics.

The interviews based on open questions and built objectively without interfering or guiding the respondents to specific answers was conducted in various target whereabouts.

#	Interviewers	Respondents
1	Analytic Manager/Director	2
2	Data analysis Consultants	7
3	Big Data Programmer	2
4	Business Intelligence Consultant	5
5	Banks loyal customer	15



The body of organization and financial institutions banks in particular their knowledge about big data analysis is still very diminutive in terms of understanding the advantages of using algorithm to analysis their data with the purpose of getting understand exactly what their customer prefer in their daily based services. Bank's Analytics Manager/Director and ICT responsible are unaware of using advanced Hybrid algorithm to analyze their customers data, almost 40% of responded have shown unawareness of big data and analytics in their respective banks. 10% of responded raised the cost issue expressing how adopting big data and analytics in Banking sector may be expensive, some of them understand how useful it is to use big data and analytics with new Algorithm (Hybrid) in banking but still Skills gap become their outmost problem which took almost of 20 % of the respondent, security of customer's data as well as banks data have been an issue at the extend of 15 % nearly 15% of responded have presented the fact of resistance to change, basically these are using Outmoded data analysis mode and never want any updated in terms of data analysis using an Hybrid algorithm.

V. CONCLUSION AND RECOMMENDATION

This paper has used an Hybrid algorithm made of a Convolution neural network model with a recurrent Neural network both to classify structure and unstructured data as a single, doing so it was possible to uncover customers hidden patterns and unknown correlation by increase the speed of processing and provides bellow results in banking sector;

1. Obtain a 360-degree view of their customers through data integration and a multidisciplinary approach;
2. Highlights new ways to make smarter and faster decisions, and obtain superior insights into customer needs, behaviors, mindset, purchase preferences, sentiments and profitability.
3. Illustrate ways Bank can harness unstructured data such as call center logs, blogs and social media posts—in order to overcome modern business challenges and increase revenues.

4. Create new product and services to help customer to manage their finances and save the money

This research has been done with limited data-sets of 2G including Images and texts, we recommend the next research to adopt a heavy data-sets. We recommend the next researcher to train and produce accuracy which will be compared to different ac-curacies using deep-learning technology and improve the existing ones. This research did not focus on Image and text segmentation, we recommend the next research to focus on Image segmentation and provide accurate features of posted images.

REFERENCES

[1] "The World's Technological Capacity to Store, Communicate, and Compute Information". Martin Hilbert.net. Retrieved 13 April 2016.

[2] New Horizons for a Data-Driven Economy – Springer .doi:10.1007/978-3-319-21569-3.

[3] boyd, dana; Crawford, Kate (September 21, 2011). "Six Provocations for Big Data".Social Science Research Network: A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society. doi:10.2139/ssrn.1926431.

[4] "Data, data everywhere".The Economist. 25 February 2010. Retrieved 9 December 2012.

[5] "Community cleverness required".Nature 455(7209): 1. 4 September 2008.doi:10.1038/455001a.

[6] Reichman, O.J.; Jones, M.B.; Schildhauer, M.P. (2011). "Challenges and Opportunities of Open Data in Ecology".Science.331(6018): 703–5.doi:10.1126/science.1197962.PMID21311007.

[7] Kohavi, Rothleder and Simoudis (2002). "Emerging Trends in Business Analytics". Communications of the ACM.45(8): 45–48.doi:10.1145/545151.545177.

[8] Phillips, Judah "Building a Digital Analytics Organization" Financial Times Press, 2013, pp 7–8.

[9] "Security analytics shores up hope for breach detection". Enterprise Innovation. Retrieved April 27,2015.

[10] Naone, Erica."The New Big Data". Technology Review, MIT. RetrievedAugust 222011.

[11] Inmon, Bill; Nesavich, Anthony (2007).Tapping Into Unstructured Data.

[12] Moore, M.A. Little, P.E. McSharry. J.R. Geddes, G.M. Goodwin (2012). Forecasting depression in bipolar disorder. IEEE Trans Biomed Eng., 59(10):2801-7

[13] akeo Kanade (6 December 2012). tree dimensional machine vision . Springer Science & Business Media.

[14] DAVID (1982). Vision. W. H. Freeman and Company.

[15] William Freeman; Pietro Perona; Bernhard Scholkopf (2008). "Guest Editorial: Machine Learning for Computer Vision". International Journal of Computer Vision. 77 (1).

[16] Milan Sonka; Vaclav Hlavac; Roger Boyle (2008). Image Processing, Analysis, and Machine Vision. Thomson.

AUTHORS' BIOGRAPHIES

RUKUNDO Prince
Position: Assistant Lecturer
E-mail: rukuprince@gmail.com
Career: Education
+ (250) 787752493

EDUCATION BACKGROUND

- 2008-20013: **M.sc Integrated (Information Technology),** India- Annamalai University.
- 2011-2013: **MBA,** India- Annamalai University (Distance learning).

SERVED INSTITUTION

- 2015 to date: University of Rwanda full time employment.
- 2014-2015 : University of Tourism Technology and business studies

ROLES AND RESPONSIBILITY

- Programming language developer in (JAVA, PHP, and MySQL).
- Researcher in Data science and Data analysis.
- ICT consultancy.

AREA OF INTEREST

- Big Data Technology and Data analysis.
- Data mining and warehouse.
- Cloud computing and Data Center.
- Deep-Learning with Python

"Information is the oil of the 21st century, and analytics is the combustion engine"
Gartner.

MBANZABUGABO Jean Baptiste, PhD., A highly organized IT professional who has a proven track record and innovative at every Academic move. Possessing a sound knowledge to modern software development Technologies; equally successful in both team and self-directed settings; and proficient in a range of computer systems, Programming technologies, Tools and a track record of achieving the desired outcomes, on-time.

His greatest strengths are being able to handle as the situations demand, patience, sincere, committed to hard work, self-learning, research and motivated to innovate.

Having completed his Bachelor of Science in Engineering from the former Kigali Instituted of Science and Technology (KIST), Masters of Computer Applications from Bangalore University and Annamalai University for Masters of Science in Software Engineering, he started his Doctorate studies in Computer Science at Atlantic International University (AIU) concentrating on use and issues related to Cloud Computing Technology.

Mbanzabugabo Jean Baptiste has a passion for research in the area of computer related technologies and Trends. His current affiliations include being Dean of the Faculty of Applied Science and Technology at the University of Tourism, Technology and Business Studies (UTB), a Lecturer of IT related Modules in various Rwandan Universities and colleges. He can be reached at i_engineer@netzero.com.

Innocent NSANZUMUHIRE
Position: Assistant Lecturer
E-mail: nsabumuhirei@gmail.com
Career: Education

EDUCATION BACKGROUND

- **M.sc (computer Engineering)**

SERVED INSTITUTION

- University of Rwanda full time employment.

ROLES AND RESPONSIBILITY

- Programming language developer in (JAVA, PHP, and MySQL).
- Researcher in Telecommunication Engineering ,ICT consultancy

Gaston KAGEMANA
Position : Director of Academics Affairs and Registrar
E-mail: gastonkagemana@gmail.com
Career: Education

EDUCATION BACKGROUND

- **M.sc IT**

SERVED INSTITUTION

- University of lay Adventists of Kigali (UNILAK)

ROLES AND RESPONSIBILITY

- Director of Academics Affairs and Registrar.
- Assistant lecturer