https://doi.org/10.31871/IJNTR.5.10.37
International Journal of New Technology and Research (IJNTR)
ISSN: 2454-4116, Volume-5, Issue-10, October 2019 Pages 69-72

Big Data Analysis in Banking Sector

Shweta Sharma, Tushar Sharma, Bhaarat Kotak, Amogh Hasotkar

***Abstract*— Big data analytics is the often-complex process of examining large and varied data sets or big data to uncover information including hidden patterns, unknown correlations, market trends and customer preferences that can help organizations make informed business decisions. On a broad scale, data analytics technologies and techniques provide a means to analyse data sets and draw conclusions about them to help organizations make informed business decisions. BI queries answer basic questions about business operations and performance. Big data analytics is a form of advanced analytics, which involves complex applications with elements such as predictive models, statistical algorithms and what-if analysis powered by high-performance analytics systems.**

***Index Terms*— Big data, Banking Sector.**

I. INTRODUCTION

Highlight Driven by specialized analytics systems and software, as well as high-powered computing systems, big data analytics offers various business benefits, including new revenue opportunities, more effective marketing, better customer service, improved operational efficiency and competitive advantages over rivals. Big data analytics applications enable big data analysts, data scientists, predictive modelers, statisticians and other analytics professionals to analyse growing volumes of structured transaction data, plus other forms of data that are often left untapped by conventional business intelligence (BI) and analytics programs. That encompasses a mix of semi-structured and unstructured data -- for example, Internet, clickstream data, web server logs, social media content, text from customer emails and survey responses, mobile phone records, and machine data captured by sensors connected to the internet of things. Emergence and growth of big data analytics: The term big data was first used to refer to increasing data volumes in the mid-1990s. In 2001, Doug Laney, then an analyst at consultancy Meta Group Inc., expanded the notion of big data to also include increases in the variety of data being generated by organizations and the velocity at which that data was being created and updated.

Those three factors volume, velocity and variety became known as the 3Vs of big data, a concept Gartner popularized after acquiring Meta Group and hiring Laney in 2005. Adopting the Big Data analytics and imbuing it into the existing banking sector workflows is one of the key elements of surviving and prevailing in the rapidly evolving business environment of the digital millennium. Separately, the Hadoop distributed processing framework was launched as an Apache open source project in 2006, planting the seeds for a clustered platform built on top of commodity hardware and geared to run big data applications. By 2011, big data analytics began to take a firm hold in organizations and the public eye, along with Hadoop and various related big data technologies that had sprung up around it. Initially, as the Hadoop ecosystem took shape and started to mature, big data applications were primarily the province of large internet and e-commerce companies such as Yahoo, Google and Facebook, as well as analytics and marketing services providers. In the ensuing years, though, big data analytics has increasingly been embraced by retailers, financial services firms, pirate bay insurers, healthcare organizations, manufacturers, energy companies and other enterprises. [8]

Shweta Sharma, Information Technology, Thakur college of engineering and technology, Mumbai, India
Tushar Sharma, Information Technology,Thakur college of engineering and technology, Mumbai, India
Bhaarat Kotak, Information Technology,Thakur college of engineering and technology, Mumbai, India
Amogh Hasotkar, Information Technology,Thakur college of engineering and technology, Mumbai, India

II. HOW IT WORKS.

In some cases, Hadoop clusters and NoSQL systems are used primarily as landing pads and staging areas for data before it gets loaded into a data warehouse or analytical database for analysis -- usually in a summarized form that is more conducive to relational structures.

More frequently, however, big data analytics users are adopting the concept of a Hadoop data lake that serves as the primary repository for incoming streams of raw data. In such architectures, data can be analysed directly in a Hadoop cluster or run through a processing engine like Spark. As in data warehousing, sound data management is a crucial first step in the big data analytics process. Data being stored in the Hadoop Distributed File System must be organized, configured and partitioned properly to get good performance out of both extracts, transform and load (ETL) integration jobs and analytical queries.

Once the data is ready, it can be analysed with the software commonly used for advanced analytics processes. That includes tools for data mining, which sift through data sets in search of patterns and relationships; predictive analytics, which build models to forecast customer behaviour and other future developments; machine learning, which taps algorithms to analyse large data sets; and deep learning, a more advanced offshoot of machine learning.

Text mining and statistical analysis software can also play a role in the big data analytics process, as can mainstream BI software and data visualization tools. For both ETL and analytics applications, queries can be written in MapReduce, with programming languages such as R, Python, Scala, and SQL, the standard languages for relational databases that are supported via SQL-on-Hadoop technologies. [4]

III. ADVANTAGES

BD offers a number of advantages to both banks and their customers. Advantages of BD in terms of functional and business area are given in table I. Some of the world wide accepted advantages of applying BD for banking in India are as follows:

- **Fraud Detection and Prevention**: It is one of the major problems faced by the financial sectors and BD can ensure banks that no unauthorized transactions and access will be made from their systems, providing a level of safety and security that will raise the security standard of the entire financial industry.
- **Customer Segmentation**: customer base into groups of individuals that are similar in particular ways relevant to marketing and business, such as their age, gender, financial conditions, interests and spending habits. This segmentation allows banks to provide or deliver to customers with exactly what they're looking for.
- **Risk Management**: The early detection of fraud is a large and major part of risk management and BD can do as much for risk management, as it does for fraud identification. It locates and presents BD on a single large scale that makes it easier to reduce the number of risks to a manageable number. This would further reduces the chances of losing data or ignoring frauds within transaction in banks.
- **Study of Indian Economy**: Similar to what financial organizations and banks are doing in other countries such as U.S.A., the techniques can be applied in India for studying the Indian economy more efficiently, and can help in improvising it to a better level.
- **Past Data Analysis and Future Predictions**: Banks can also look at the past data, they have already stored, and can plan for the future. BD helps them in spotting patterns in different domains of their services provided to their customers and can use these data patterns to predict their businesses future e.g. where and how to invest their labor, money and time for profitable returns. [7]

IV. BIG DATA ANALYTICS USES AND CHALLENGES.

Big data analytics applications often include data from both internal systems and external sources, such as weather data or demographic data on consumers compiled by third-party information services providers. In addition, streaming analytics applications are becoming common in big data environments as users look to perform real-time analytics on data fed into Hadoop systems through stream processing engines, such as Spark, Flink and Storm.Margaret Rouse.

Early big data systems were mostly deployed on premises, particularly in large organizations that collected, organized and analysed massive amounts of data. But cloud platform vendors, such as Amazon Web Services (AWS) and Microsoft, have made it easier to set up and manage Hadoop clusters in the cloud, as have Hadoop suppliers such as Cloudera and Hortonworks, which support their distributions of the big data framework on the AWS and Microsoft Azure clouds. Users can now spin up clusters in the cloud, run them for as long as they need and then take them offline with usage-based pricing that doesn't require ongoing software licenses.

Potential pitfalls of big data analytics initiatives include a lack of internal analytics skills and the high cost of hiring experienced data scientists and data engineers to fill the gaps.

Recently, the proliferation and advancement of AI and machine learning technologies have enabled vendors to produce software for big data analysis that is easier to use, particularly for the growing citizen data scientist population. Some of the leading vendors in this field include Alteryx, IBM, Microsoft and Knime.

The amount of data that's typically involved, and its variety, can cause data management issues in areas including data quality, consistency and governance. Also, data silos can result from the use of different platforms and data stores in a big data architecture. In addition, integrating Hadoop, Spark and other big data tools into a cohesive architecture that meets an organization's big data analytics needs is a challenging proposition for many IT and analytics teams, which have to identify the right mix of technologies and then put the pieces together.[2]

V. TOOLS AND TECHNOLOGIES

Unstructured and semi-structured data types typically don't fit well in traditional data warehouses that are based on relational databases oriented to structured data sets. Further, data warehouses may not be able to handle the processing demands posed by sets of big data that need to be updated frequently -- or even continually, as in the case of real-time data on stock trading, the online activities of website visitors or the performance of mobile applications. As a result, many of the organizations that collect, process and analyse big data turn to NoSQL databases, as well as Hadoop and its companion tools, including:

- **YARN**: a cluster management technology and one of the key features in second-generation Hadoop.
- **MapReduce**: a software framework that allows developers to write programs that process massive amounts of unstructured data in parallel across a distributed cluster of processors or stand-alone computers.
- **Spark**: an open source, parallel processing framework that enables users to run large-scale data analytics applications across clustered systems.
- **HBase**: a column-oriented key/value data store built to run on top of the Hadoop Distributed File System (HDFS).
- **Hive**: an open source data warehouse system for querying and analysing large data sets stored in Hadoop files.
- **Kafka**: a distributed publish/subscribe messaging system designed to replace traditional message brokers.

- **Pig**: an open source technology that offers a high-level mechanism for the parallel programming of MapReduce jobs executed on Hadoop clusters. [1]

VI. APPLICATION

Fraud Detection

Big data analytics have become an essential part of any strategy to help detect and prevent financial crime, owing to the ever-evolving attack methods used by criminals exploiting multichannel vulnerabilities to compromise technology systems. Big data has enabled banks to implement real time analytics on a large scale to meet the growing threats. Using data mining fraud detection technique that detects both known and novel fraud instances as they occur in real time, with a higher level of accuracy using distributed Ha- doop-based platforms that make it possible to cost-effectively and efficiently store and process large data sets. [3]

VII. DATA MINING TECHNIQUES IN BIG DATA

A. Classification: Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large. Fraud detection are particularly well suited to this type of analysis.

B. Clustering: Clustering can be said to be the identification of similar classes of objects. In this technique, transactions with similar behaviour are combined into one group. For instance: The customer of a given geographic location and of a particular job profile demand a particular set of services, like in banking sector the customers from the service class always demand for the policy which ensures more security as they are not intending to take risks, likewise the same set of service class people in rural areas have the preferences for some particular brands which may differ from their counterparts in urban areas. This technique will help the management in finding the solution of 80/20 principle of marketing, which says: 20% of your customers will provide you with 80% of your profits, then the problem is to identify those 20% and the techniques of clustering will help in achieving the same.

C. Association Rule: The central task of association rule mining is to find sets of binary variables that co-occur together frequently in a transaction database. Association rule has the several algorithms like: APRIORI, CDA, and DDA. Association rules are if/then statements that help uncover relationships between seemingly unrelated data in a relational database or other information repository.

D. Prediction: The prediction as its name implies is one of the data mining techniques that discover relationship between independent variables and relationship between dependent variables. For instance, prediction analysis technique can be implemented in the banking sector to predict fraud. Money can be seen as the independent variable while the individual (fraudster) could be seen as the dependent variable. Then based on historical data, we can draw a fitted regression curve that is used for attempted fraud prediction. Regression analysis can be used to model the relationship between one or more independent variables and dependent variables. Types of Regression Techniques are as follows:

1.Linear Regression
2.Multivariate Linear Regression
3.Nonlinear Regression [6]

VIII. ANOTHER APPLICATION IN BUSINESS SECTOR

The banking industry has evolved by leaps and bounds over the past decade, when it comes to operations and service delivery. The financial and banking data will be one of the cornerstones of this Big Data flood, and being able to process it means being competitive among the banks and financial institutions.

Big data analytics can improve the extrapolative power of risk models used by banks and financial institutions.

Big data analytics not only brings new insights to the banks, but it also enables them to stay a step ahead of the game with advanced technologies and analytical tools.

It helps the Bank to analyse social media to monitor user sentiment toward a firm, brand or product.

In a highly competitive market, it is driving firms to compete aggressively for customers' wallet: increasing focus on customer acquisition, retention and profitability. While getting a complete view of customer relationship across the enterprise is very important, it is equally essential to use it to offer customized products and service to profitable clients will increase client loyalty and result in increased wallet share & reduce losses by minimizing the risk exposure

1. Efficient Risk Management to Prevent Errors and Frauds

Business intelligence (BI) tools are capable of identifying potential risks associated with money lending processes in banks. With the help of big data analytics, banks can analyze the market trends and decide on lowering or increasing interest rates for different individuals across various regions.

2. Provides Personalized Banking Solutions To Customers

Big data analytics can aid banks in understanding customer behaviour based on the inputs received from their investment patterns, shopping trends, motivation to invest and personal or financial backgrounds. This data plays a crucial role in winning customer loyalty by designing personalized banking solutions for them.

3. Easier Filing of Regulatory Compliances

BI tools can help analyse and keep track of all the regulatory requirements by going through each individual application from the customers for accurate validation.

4. Boosts Overall Performance

With performance analytics, employee performance can be assessed whether or not they have achieved the monthly/quarterly/yearly targets. Based on the figures derived from current sales of employees, big data analytics can determine ways to help them scale better. [5]

IX. CONCLUSION

With respect to the current, highly competitive financial market, big data holds the key to unlocking marketing

potential. Advanced analytics are permitting banks to manage the cumulative cost of compliance and the risk of non-compliance. However, the financial service firms are still lagging behind in implementing big data analytic tools, which indicates an untapped potential for value creation, available for the banking industry. This needs to be evaluated from an IT (Information Technology) or LoB (Line of Business) perspective.

REFERENCES

[1]https://searchbusinessanalytics.techtarget.com/definition/big-data-analytics
[2]http://in.kompass.com/e/en/article-b2b/big-data-analytics/
[3]https://www.academia.edu/14414218/The_Impact_of_Big_Data_Analytics_on_the_Banking_Industry
[4]https://www.simplilearn.com/big-data-applications-in-industries-article
[6] https://acadgild.com/blog/big-data-in-banking
[6] https://bit.ly/2JLeMZb
[7] https://ieeexplore.ieee.org/document/8344732
[8]https://searchbusinessanalytics.techtarget.com/definition/big-data-analytics