



## AN EMPIRICAL STUDY OF BIG DATA ANALYTICS IN INDIAN BANKING SECTOR

**Mr. RAMPRAVESH GOND,**  
Research Scholar, JJT University,  
Rajasthan.

**Dr. RASHI GUPTA,**  
Research Guide, JJT University,  
Rajasthan.

### ABSTRACT

*Social media has opened new avenues and opportunities for organizations to connect with their customers easily, but the volume of data by way of communications about brands, products and services on different social platforms can be overwhelming. Big data analytics helps to quickly read all this data, provide an executive summary of what people like and don't like about company brand or products and helps in providing valuable business insights. The Banking industry is a fast growing sector in India. In order to stay competitive, banks in India are taking the data analytics route to attract new customers, retain them, find opportunities to upsell and cross-sell and minimize their own losses. The aim of this paper is to study and understand the Big data Analytics, key characteristics of Big data, how to derive value from big data, the trend of data analytics, rethinking Data Management and capture how big data analytics is being successfully used in Indian banking sector, with respect to different aspects like Spending pattern of customers, Channel usages, Customer Segmentation and Profiling, Product Cross Selling based on the profiling to increase hit rate, Sentiment and feedback analysis, Security and fraud management.*

**Keywords:** Big data Analytics; Banking; Data Management; Customer Behaviour Analysis; Spending Pattern Analysis; Transactional Analysis.

### INTRODUCTION:

Recently, while planning a family trip to nainital, enquired couple of hotels on line through various sites, next day surprisingly each time while visiting to some other sites over the internet, getting popup about best hotel offer availability frequently. Similarly you shouldn't be surprised to see promotional offers from your favourite retailers on specific product or service that you care about or frequently shop for based on your buying interests. Have you ever paused and thought about how these vendors or service providers are able to analyse and communicate to you directly to suit your interests and needs? There are various ways to capture data at the point of occurrence in real time, store and analyse data to help them understand the behavioural patterns, and guide them as to trends. These data are high in volume, get generated at high velocity, come in a wide variety, and are therefore rightly termed as big data.

For current time's Big data is categorised as a major revolution across the globe. It is becoming critical in all kind of business, and it will help in gaining sustainable competitive advantage for the firm. As the technology platform is frequently changing and maturing, all types of organizations need to give strategic importance to big data sources so that they could gain insights about customer behaviour and could offer their products and services based on customer needs.

In 1970s the term big data is first used to describe large amounts of data generated by oceanography and meteorological experiments. Big data can be understood as a natural evolution of database management techniques that has changed the way data is analysed.

Early implementations of big data solutions can be found during the 1980s — the era of the first generation of software-based parallel database architecture. Google developed the first wave of big data tools in the early 2000s, which gave birth to several other frameworks and techniques that make the handling, processing, and interpretation of large data sets more economical. By leveraging big data, companies can extract value and meaningful insights from voluminous data beyond what was previously possible using traditional analytical techniques.

This also deals with new phenomena of the volume, velocity, and variability of massive data coming from social media, web logs, and sensors combined with transactional systems.

Big data is characterized primarily by large and rapidly growing data volumes, varied data structures, and new or newly intensified data analysis requirements. This enables us to deliver our customers in context the right offer, message, recommendation, service, or action, tailored and personalized to deliver unequalled value.

With the evolution of social media, we started seeing the emergence of non-traditional, less structured data such as web logs, social media feeds, e-mail, sensors, photographs, and YouTube videos that can be analysed for useful information. With the reduction of cost in both storage and computing power, it is now feasible to store and analyse this data for meaningful purposes. As a result, it is important for existing businesses and for new businesses to understand and evaluate the relevance of big data for their business intelligence and for decision making.

Big data addresses all types of data coming from various data sources, such as enterprise applications data that generally includes data generated from enterprise resource planning (ERP) systems, customer information from customer relationship management (CRM) systems, supply chain management systems, e-commerce transactions, and human resources (HR) and payroll transactions. It also attributes semantic data that comprise call details records (CDRs) from call centres, web logs, and smart meters, manufacturing sensors, equipment logs, and trading systems data generated by machine and computer systems.

---

## KEY CHARACTERSTICS

**Volume:** As we know social media like YouTube, LinkedIn, Twitter and Facebook, and so forth generate a large volume of data that need to be stored and analysed rapidly in context for the right decision making. The volume of data generated through all is much larger than the traditional data volume. The benefit gained from the ability to process large amounts of information is the main attraction of big data analytics.

**Velocity:** The data from all the sources comes at faster pace into the data management system and it requires very quick analysis for decision making. Here the importance lies in the speed of data received, quickly arranged the unstructured data and analyse the data using different analytics tools and share the analysis to the organisation for quick decision making to gain the competitive advantages over rival firm. Social media bring a large amount of unstructured data containing the customers taste and preferences and their behaviour across the channels, all these data are valuable for customer relationship management. Most of these data received may be in unstructured form, therefore analytical processing may be required in order to transform the data into a usable form or derive meaningful information.

**Variety:** There are different types of varied data involve in the collection of Big data. It includes different type of image or video data, raw data from varied sources, types of access and web logs generated by web and application servers, text from social networks. As all these data are unstructured or raw data, this cannot easily integrated in any applications for analysis. A big data analytics tools have been used for big data processing so that we can extract meaningful information and same can be used by application for further analysis. Big data brings a lot of data that has patterns, sentiments, and behavioural information that need analysis.

**Value:** The economic value of data collected is having significant value for the organisation. Normally, there is quality information hidden in quantitative unstructured data. Big data offers greater value to businesses in bringing real-time market and customer insights, enabling improvement in new products and services. Big data analytics can reveal insights such as peer influence among customers, revealed by analysing shoppers' transactions and social and geographical data.

---

## DEFINING DATA ANALYTICS

In simple words Data analytics (DA) is the science of evaluating and examining the unstructured data with the aim of deriving meaningful information which helps in decision making. In today's world, Data analytics is used in many sector like manufacturing, retail, banking and so forth. This allow companies and organization to make better business decisions.

Big data analytics is the process of examining big data to uncover hidden patterns, unknown correlations and other useful information that can be used to make better decisions.

---

## TRENDS IN BIG DATA ANALYTICS

**1. Big data analytics in the cloud:** Today, Small and midsize companies and even large enterprises are mapping strategies that take more of their applications to the cloud and out of the data centre. Most of the organisation using cloud technology to store all their unstructured data. This is because it is cost effective and easy to analyse data with no risk. There are

number of technologies are available for processing data in the cloud for examples Amazon's Redshift hosted BI data warehouse, Google's BigQuery data analytics service, IBM's Bluemix cloud platform and Amazon's Kinesis data processing service. In coming future big data will be a hybrid of on-premises and cloud.

**2. Hadoop: The new enterprise data operating system:** Hadoop is a general-purpose data operating system. We can perform many different data manipulations and analytics operations by plugging them into Hadoop as the distributed file storage system. The ability to run many different kinds of queries and data operations against data in Hadoop will make it a low-cost, general-purpose place to put data that you want to be able to analyse.

**3. Big data lakes:** A data lake is like a man-made reservoir. At first you draw a dam that is build a cluster, then you let it fill up with water that is data. Once you establish the lake, you start using the water (data) for various purposes like generating electricity, drinking, and recreating. A data lake is a method of storing data within a system or repository, in its natural format, that facilitates the data in various schemata and structural forms. The functionalities of the data lake and the data warehouse are complementary to each other.

**4. More predictive analytics:** With big data, analysts have not only more data to work with, but also the processing power to handle large numbers of records with many attributes. Traditional machine learning uses statistical analysis based on a sample of a total data set, where as in big data analytics, we now have the ability to do very large numbers of records and very large numbers of attributes per record and that increases predictability.

**5. SQL on Hadoop: Faster, better:** This is a tools that support SQL-like querying. SQL on Hadoop opens the door to Hadoop in the enterprise because businesses don't need to make an investment in high-end data scientists and business analysts.If you're a smart coder and mathematician, you can drop data in and do an analysis on anything in Hadoop. These tools are nothing new. Apache Hive has offered a structured, SQL-like query language for Hadoop for some time. But commercial alternatives from Cloud era, Pivotal Software, IBM and other vendors not only offer much higher performance, but also are getting faster all the time.

**6. More, better NoSQL:** This is an alternatives to traditional SQL-based relational databases, called NoSQL short for "Not Only SQL" databases. This is rapidly gaining popularity as tools of analytic applications. There are 15 to 20 open-source NoSQL databases are available each with its own specialization. For example, a NoSQL product with graph database capability, such as ArangoDB, offers a faster, more direct way to analyse the network of relationships between customers or salespeople than does a relational database.

**7. Deep learning:** Deep learning, a set of machine-learning techniques based on neural networking, is still evolving but shows great potential for solving business problems. Deep learning enables computers to recognize items of interest in large quantities of unstructured and binary data, and to deduce relationships without needing specific models or programming instructions.Big data will do things with lots of diverse and unstructured text using advanced analytic techniques like deep learning to help in ways that we only now are beginning to understand.

**8. In-memory analytics:** The use of in-memory databases to speed up analytic processing is increasingly popular and highly beneficial in the right setting. In fact, many businesses are already leveraging hybrid transaction/analytical processing (HTAP) — allowing transactions and analytic processing to reside in the same in-memory database. For systems where the user needs to see the same data in the same way many times during the day — and there's no significant change in the data — in-memory is a waste of money.

---

## DATA ANALYTICS - USE CASE FROM INDIAN BANKING

The Banking industry is a fast growing sector in India. According to an IBEF report, the banking industry in India has the potential to become the fifth largest banking industry in the world by 2020 and third largest by 2025 according to a KPMG-CII report.

In order to stay competitive, banks in India are taking the data analytics route to lure new customers, retain them, find opportunities to upsell and cross-sell and minimize their own losses.

- **HDFC Bank- Using Analytics to Get a Complete Picture of the Customer**

According to an article published in Live Mint, one of the first instances of the use of analytics can be traced back to the early 2000s when HDFC Bank Ltd, put in place a data warehouse and started investing in technology that would help it make sense of the massive chunk of unstructured data captured by its information technology (IT) systems. With the analytics engine in place, HDFC Bank can track every aspect of a typical customer's financial habits. For example, whether the customer has an active account or he's just having a salary credited to his account. The analytics tools also gives the bank insights into personal habits of its customers, allowing it to promote offers accordingly. Analytics is also used to reduce chances of money laundering by identifying suspicious activity such moving money to multiple accounts, finding large single-day cash deposits, opening a number of accounts in a short period of time or sudden activity in long-dormant accounts. Using analytics, the bank is also able to keep track of credit histories of customers and can hand out loans accordingly.

- **ICICI Bank- Using BI and Analytics to Reduce Credit Losses**

The 2007 subprime mortgage crisis in US had repercussions in India as well. Banks were required to deal with challenges of reduced liquidity and rising interest rates and still ensure that their customers remained happy and loyal. ICICI bank identified Debt Collection as key process where a friendly approach could improve customer satisfaction. One of the important steps in the debt collection process was choosing the appropriate customer-approach channel for each case. The bank management, with the aim to transform debt collection as a customer retention tool, decided to use technology to achieve the objective. Developed in-house, the BI solution implemented

in ICICI Bank includes components of SAS, Sybase, TRIAD, Posidex, Data Clean, and Blaze Advisor that factored in several parameters such as efficiency of collector, customer profile, risk behaviour, and exposure.

- **Axis Bank – Analytics for Customer Intelligence & Risk Management**

Axis Bank has seen the productivity of the sales staffs increase by five times in the couple of financial year. They are using analytics in almost every sphere. For example, when their sales guys are visiting to a client for some loan, they try and find out what the background of the customer is and what the likelihood of him taking a particular loan is. Axis Bank also uses analytics to increase customer loyalty and reduce loan prepayments due to refinancing with other institutions. Axis Bank uses SAS to provide customer intelligence across the organization. The SAS tool also helps the bank to improve risk management throughout the organization by giving them early warning signals.

- **State Bank of India-Using data analytics**

Though private sector banks are leading the charge in using data analytics for effective decision-making, public sector banks are not far behind. SBI's data warehouse has over 120 TB of data and receives an additional 4 TB of banking data a day. They are using data analytical tools to build their data models to education loans, automotive loans, housing loans, SME loans to try and reduce the percentage of them going bad. For instance, in case of an education loan, they combine data from their bad loans, income tax departments, credit ratings agencies to identify suitable candidates and then send them reminder messages. They also use analytics to determine where ATM branches should be positioned and how much cash should be placed in them.

- **ING Vysya Bank – need for business intelligence (BI) implementation**

ING Vysya (now acquired by Kotak Mahindra Bank) felt the need for business intelligence (BI) implementation when the bank started noticing how different end users attended meetings with inaccurate reports. They required a solution to help users generate accurate and timely reports. With the help of SAP BO, they created a common data repository which helped users get accurate reports and in turn improve efficiency.

---

## OBJECTIVE

- To study and understand the Big data Analytics
- To study the trends in BigDataAnalytics
- To study the Key Characteristics of Big data and used cases in Indian banks.

---

## LITERATURE REVIEW:

The study involve review of different literature. Some of them are as follow.

**Tanvi Ahlawat et al., (2016).** In a Business environment, there are a lot of decisions that are to be taken on the basis of Data & associated analytics and in simple terms, we could define it as Big Data when combined with powered analytics, lot of business related tasks can be accomplished such as Root Cause Analysis can be conducted in real time for associated defects, failures and issues, POS based generated coupons based on Consumer Behaviour, Risk Portfolio – Quick Calculations/Re-Calculations can be conducted in minutes, Conducting Fraud Detection & use of Fraud analytics before hitting organization.

**Suriya Begum et al., (2016).** A key part of big data analytics is the need to collect, maintain and analyse enormous amounts of data efficiently. Due to increase in number of sophisticated targeted threats and rapid growth in data, the analysis of data becomes too difficult. Today's attacks are prepared by advanced technologies are not detected until the damage has been occurred. Big Data Security Analytics is important to mitigate the security threats to secure the data more efficiently. In Big Data Analytics, Data Security is a challenging task to implement and calls for strong support in terms of security policy formulation and mechanisms.

**Shilpa et al., (2013).** Big Data has come up because we are living in society that uses the intensive use of increasing data technology. As there exist large amount of data, the various challenges are faced about the management of such extensive data. The challenges include the unstructured data, real time analytics, fault tolerance, processing and storage of the data and many more. The two main problems regarding big data are the storage capacity and the processing of the data. The various techniques and technologies have been highlighted in this paper for manipulating, Analyzing and visualizing the big data.

**Harshawardhan et al., (2014).** The technical challenges must be addressed for efficient and fast processing of Big Data. The challenges include heterogeneity, lack of structure, error-handling, privacy, timeliness, provenance, and visualization, at all stages. Hadoop is a Programming framework used to support the processing of large data sets in a distributed computing environment. Hadoop was developed by Google's MapReduce that is a software framework where an application break down into various parts.

---

## RESEARCH METHODOLOGY

The research for this paper was conducted through review of secondary data, without any empirical work being conducted. A large resource of written material was used, which include books, magazine article, academic journals as well as the websites.

---

## CONCLUSION OF STUDY

The Big Data era is upon us. Organizations need to understand what Big Data is and how it can be used. The benefits and advantages are too big for organizations to ignore. Big Data is often explained using the 3Vs: volume, velocity, and variety. However, to make it more

complete, four additional Vs were added: veracity, variability, visualization, and, of course, value. Big Data consists of many different types of data in different data - sets that are combined to provide valuable insights. The combination of different datasets, such as company data, public data, and social data, will give even more insights. In order to successfully move into the era of Big Data, a culture shift is required. Organizations should become information driven and ensure that all decision makers understand what Big Data is and how it can benefit the organization. Big Data is changing how organizations operate and are managed. It also changes how society works and how consumers live. The impact of Big Data on society will be big, but it remains to be seen how society will impact Big Data.

## REFERENCES:

- Tanvi Ahlawat and Radha Krishna Rambola (2016). Literature Review on Big Data. International Journal of Advanced in Engineering Technology Management & Applied Science. Volume 3, ISSN No: 2349-3224
- Suriya Begum and Kavya Sulegaon (2016). ANALYSIS OF VARIOUS BIG DATA TECHNIQUES FOR SECURITY. International Journal of Computer Science and Mobile Computing. Volume 5, Issue.3, ISSN 2320-088X
- Shilpa and ManjitKaur (2013). BIG Data and Methodology-A review. International Journal of Advanced Research in Computer Science and Software Engineering. Volume 3, Issue 10, October 2013 ISSN: 2277 128X
- Harshawardhan S. Bhosale and Devendra P. Gadekar (2014). A Review Paper on Big Data and Hadoop. International Journal of Scientific and Research Publications, Volume 4, Issue 10, October 2014 ISSN 2250-3153
- [https://en.wikipedia.org/wiki/Data\\_lake](https://en.wikipedia.org/wiki/Data_lake)
- <http://analyticsindiamag.com/analytics-in-indian-banking-sector-on-a-right-track/http://www.ibef.org/industry/banking-india.aspxhttp://www.livemint.com/Industry/F5uNVbogJfsNB7cSt1toBL/Banking-on-Big-Data-analytics.html>
- [http://www.sas.com/en\\_us/customers/hdfc.html](http://www.sas.com/en_us/customers/hdfc.html)
- <http://searchbusinessintelligence.techtarget.in/feature/Improved-debt-collection-with-BI-An-ICICI-Bank-story>
- <http://searchbusinessintelligence.techtarget.in/survey/ICICI-Bank-cuts-credit-losses-with-BI>
- <http://searchbusinessintelligence.techtarget.in/news/2240151280/BI-project-extension-underway-at-ING-Vysya-Bank>
- <http://searchbusinessintelligence.techtarget.in/tip/Follow-Anti-Money-Laundering-AML-guidelines-the-ING-Vysya-way>
- <http://economictimes.indiatimes.com/industry/banking/finance/banking/state-bank-of-india-hires-16-statisticians-and-economists-to-drive-performance/articleshow/46048348.cms>