A study: Scope of Big Data Analytics in The Indian Financial Markets

Shiv Nidhi Bhardwaj
Vanchan Tripathi

1. Student, IMS UC, Ghaziabad, INDIA
2. Assistant Professor, IMS UC, Ghaziabad, INDIA

Abstract

Aristotle said, “Man by nature is a social animal”, arguably over the past 2.5 millennia since his death, human interaction has only increased, this interaction produces data which has increased, in tandem. In recent times, over 2.5 quintillion bytes of data is generated on a daily basis. This data is very large in size, updates at high speeds and is of different types and is therefore aptly termed, ‘big data’. While big data has found immense applicability in the financial industry, particularly that in parlance with insurance and consumer finance, it’s usage in the financial markets has been limited. This paper, therefore, aims to ascertain the future of big data analytics within the Indian financial markets. The research for this paper is based on secondary data collected from previously written academic papers and financial journals pertaining to big data and the Indian markets. This paper shall be written beginning with gaining an overview of what is big data and its analysis, whereon we’ll attempt to understand the Indian markets and their data output. This paper shall conclude with an attempt to understand what the future usage of big data in financial markets looks like.

Keywords: Big Data, Indian Financial Markets, Metadata, Data Analytics, Finance

Introduction

Data, when boiled down to its very basics is something that provides us with information about a topic. For this paper, we shall however, limit our understanding of the term to numerical values only, when these numerical values are put in groups, they are called sets of data. For example, A data set would be the total number of individuals within an office who travel to and from work by car and another set would be those who take the bus, etc.

Big data is any set of data that fulfills certain constraints, viz.

- It should be high volume
  The size of data set should be large, while there is no defined lower limit, generally it should use over 1 terabyte of memory.
- It should be variable
  It should be changing, if not at a constant rate then at least at a rate that it stays relevant to the source and is of real use in future/present analysis.
- It should be multivariate in its nature
  The data should be of different variety thereby allowing a multipronged analysis.
Any research paper should be apt in measuring present day reality in context of the material presented, with respect to this, the following facts are presented for reviewal -

The social media site, Facebook creates over 500 terabytes of data every day, the world's largest stock exchange NYSE generates over 1 gigabyte of data during 7 hours (9-4) and a jumbo jet engine generates ~10 gigabytes of the same for 30 minutes of flight time. While these values seem huge, it would be apparent that they pale in comparison to future numbers as predicted by industry leaders, most of which are attributed to the fourth industrial revolution that is said to be currently ongoing. When speaking of big data analysis, we must remember that this analysis does not refer to basic inquiries such as the MAX/MIN, i.e. imputing the maximum or minimum value of a function but rather towards more relational tests that help establish empirical models which are predictive in nature and analytical at core.

**Current Scenario of big data analysis**

The previous perception of data as a tool for representing past results has significantly changed. Let us take a look at the example of Google flu trends, a failed project by Google, wherein the corporate giant used the data it had, searches for symptoms, medicine shopping trends, physician searches, etc. to compute the onset of flu in the USA before the CDC (Centre for Disease Control). It managed to do this successfully the first time around but over time (2008-2011) the project failed, due to miscalculations, such as Google’s own search prediction compromising data quality. While this project may have failed, it successfully showcased the myriad applications that big data can have when used in a correct manner. Another prime example of the executory strength of big data is blockchain technology, which will for most people be recognizable from the crypto-currency bitcoin.

Global data generation by year (*Statista.com*)
The way blockchain works is, no full component of data is stored at any one location, this creates a network of internodal communication relays wherein the data is shared over huge webs, while this can also be scaled down to a few intra-office computers, its true efficiency is not seen unless applied at global levels. The most major factor that prevents big data from occupying the level it should is restrictive computing.

While over the past 2 decades, the cost of computing has gone down tremendously and computational efficiency has increased at an exponential pace, it still fails to be as fast as required at as low costs as needed, this means that only a few industries get to access and utilise big data, of these industries even, only a few major players get to do so such as Google, Facebook, developed economies such as the USA, etc. This prompts us to look towards the upcoming artificial intelligence and machine learning technological booms, in the form of the fourth industrial revolution as these are approximated to increase current global computational capacity by upwards of 3 times that of current capacity which would certainly be helpful to big data analysis.

Analysis of big data

Simple availability of big data does not create any value for its collectors, it is only useful if it is able to give value higher than the amounts spent to collect and process it. This is done through the analysis of raw datasets in different forms further, this value depends upon several factors such as the exactitude of the data collected, its pertinence to the subject of the study and the mode of analysis used to derive this value. Keeping this in mind we examine what are the major methods of analysis of big data

1. Descriptive analysis
   This is a backward-looking form of analysis and is generally considered to be the most basic and simple form of data analysis, it involves usage of basic statistical tools such as mean, median, mode, standard deviation, variance etc. in order to create models for what has already occurred in the past.

2. Predictive analysis
   This aims to create statistical models of data based upon techniques such as regression analysis, and machine learning. Not only does predictive analysis attempt to create models based upon historical behaviour and extrapolate them to future conditions but also to establish relations between variables in future conditions and base models on these.

3. Prescriptive analysis
   This attempts to establish a relation between the results of predictive/descriptive analysis and business policies established for optimizing efficiency and increasing total output. While this is relatively difficult to perform due to a lack of varied data types that can be found in relational data sets and business optimization processes, it is of extreme import due to its result oriented nature that establishes a cause and effect relationship between big data analysis and the policy changes made due to the same.
4. Pre-emptive analysis

This refers to the actions that an organization is likely to take in order to prepare for eventualities that have been established through predictive analysis, have been faced during descriptive analytics and has not yet been seen in the prescriptive (real time scenario).

Diagram 1. Flow chart of method of big data analysis (Source: Critical analysis of big data analysis methods)

The Indian financial markets

The Indian economic system is that of a mixed economy, i.e. the economy performs with traits of both a socialistic and a capitalistic outlook, where control of the economy is not given in the hands of private enterprises alone nor does it depend upon the government’s decisions either. This complicates the economy and creates a very complex financial market.

Further we shall attempt to understand these markets in the perspective of the Indian economy and the data they process and operate on the basis of.

First let us look at the Indian financial market system

- The common public is allowed to invest in companies through broker houses such as India Infoline services, Karvy etc.
- Only those companies can take public money which are listed on stock exchanges.
- There are over 5000 public companies in India.
- India’s major stock exchanges are the National Stock Exchange and the Bombay Stock Exchange
- The Bombay Stock Exchange has a transaction latency of 6/1000000 seconds, i.e. it has the capacity to execute 166666 transactions in 1 second.

There are many different securities that are traded on these exchanges besides company stock (equity) some of these are short term such as treasury bills issued by the government, commercial paper issued by corporations, futures and options, while others are long term securities such as mutual funds, bonds, etc. All of these and many more securities are traded and invested in by people on a daily basis, this results in the creation of high volume data that is available to the exchanges, regulators such as SEBI and very often to the general public too through portals such as the NSE data & analytics pvt. ltd. for a nominal fee.
In order to understand the data generated by the activities conducted in these markets, we must first differentiate the different types of big data as it exists in broad spectrums:

- **Structured Data**
  Data that exists within terms that allow for it to be categorized, indexed and operated upon is called structured data. Compared to unstructured data it is much easier to process this data. An example of structured data is the price history of a commodity or security as viewed by people. According to Gartner, structured data is approximately 20% of all enterprise data.

- **Unstructured Data**
  Data that is characteristically not bound to any format is known as unstructured data. It is generally extremely difficult to search through unstructured data as it does not fit into database structures such as SQL, etc. An example of unstructured data is webpage code, algo imaging, etc. Gartner approximates that ~80% of all enterprise data is in the form of unstructured data.

- **Semi-structured Data**
  This can be considered as a subset of unstructured data. It is different in the sense that it can be parsed to a certain extent with the help of advanced methods such as lexical analysis. E-mails and such data forms are an example of semi-structured data.

![Chart 1. Growth of structured vs. unstructured data (source:guru99.com)](image)

**Data in the financial markets**

The financial markets revolve around the concept of buyers, sellers, brokers and regulators trading and performing activities relating to the same in a controlled environment. These activities lead to the creation of tremendous amounts of data, the following paragraphs establish the forms of data seen in the financial markets.

Structured data such as price, volume of a security exists on a very basic level. When explored further we can see data such as price history, moving average history, bids, asks, open/float interest rates etc. also form part of this. These data forms fall under the aegis of structured data as they are often seen in the form of tabular/chart-based representations (relational databases).

Unstructured data also exists within the markets but at a deeper level and is generally not available to the general public, the most obvious example of the same would be market chatter, i.e. all of
the communication that various entities have over different platforms in relation to the markets. For example, if a large enough number of brokers recommend the purchase of the same security to a large enough number of buyers, the forces of demand and supply are bound to raise the price of the same, while this may seem unrealistic, one must remember that it is due to the existence of such relations that chaos theory finds immense application in the markets. Further unstructured data may be found in the form of margins set up on individual accounts and the utilization of the same to calculate the difference between the real and apparent amounts that were traded on a particular exchange within a particular day. These data examples are but a small representation of the mammoth amounts of data that flows through the financial markets of a nation, besides these another data type that exists and can find tremendous application in the future is metadata.

Metadata refers to the background data, i.e. the data that provides information about other data such as the source of creation of price-volume charts for commodities, the metadata has tremendous potential for usage albeit less so currently with conceptual technologies such as machine learning and artificial intelligence still in a nascent point of their life cycles. When we talk about data, particularly in the case of the Indian financial markets, all of the above points stand true and absolutely valid, in fact due to the existence of immense forces of demand and supply these data structures become much larger in their applicability.

Conclusion

As has been established repetitively over the course of this research paper big data is clearly extremely important. In the niche case of the Indian financial markets, the usage of big data is not limited to the future, as we can see ample examples of it in application even today such as the pattern graph tracking based stock price prediction which uses the collective repetitive behavior of multiple stocks on multiple stock exchanges to predict future price behavior.

The findings of this paper are simplistic in their nature and easy to understand. Big data is extremely useful especially in the case of numerical predictive behavior within the financial markets.

Graph: showcasing the forecasted demand for big data professionals (Statista.com)
While it is clear that big data is not as efficient as it could be, it can also be stated, with the same clarity that this difference in the current level of big data use and it’s optimum usage is only due to outside factors such as the limited progress in the field of machine learning and artificial intelligence. The upward demand forecast for the need of professionals proficient in big data and its application is proof of the pace at which it is growing as is the amount of data generated globally over the past years. These facts combined with the research presented in this paper is conclusive proof of the major growth potential in the field of big data and its analytical applications.

Managerial Implications

An article in the Economic times promptly mentioned how a mall in Delhi computed the spending habits of consumers through the use of an algorithm that used body heat generated by crowds seen through CCTV, this allowed the mall’s management to realise a growing need for a new sports apparel store and thereby increase their earnings considerably.

Big data is not centered around economics and mathematical fields alone, anymore. Now big data is used in every aspect of a successful organization. For example, multinational conglomerates such as Amazon, Pepsi track the market behaviour from various sources such as their employees company issued bank A/c etc. and on the basis of the same determine the most apropos hampers to be constructed for the employees.

Besides such myriad applications, big data is also useful for a manager as it allows him to predict future conditions, such as a supplier defaulting on delivery of goods, market demand shifting, employee behaviour etc. Big data also allows a savvy manager to predict and correctly utilize a trend that has not occurred yet in the market.

References

Papers -

Big Data analytic techniques : 6 techniques - Hong Shu

Big Data Analytics Literature review paper : Nada Elgendy & Ahmed Elragal

Big Data & Big Money : Jennifer Q Telewicz - Deutsche Bank technology centre

Big Data : Viktor mayer-Schonberger & Kenneth Cukier - Houghton Mifflin Harcourt

Critical analysis of Big Data challenges and analytical methods - Uthyasahnkar Shivrajah, Muhammad Mustafa Kamal, Zahir Irani & Vishanth Weerakkodylytical methods
Articles -

Introduction to BIG DATA: What is, Types, Characteristics & Example

Big Data Analysis Techniques

7 Big Data Techniques That Create Business Value

How Is Big Data Collected - A Detailed Guide

The Most Effective Data Collection Methods that Companies Are Using

The future of big data: 5 predictions from experts

What We Can Learn From the Epic Failure of Google Flu Trends

Big Data Analytics – The current scenarioDATAQUEST

Structured vs. Unstructured Data

Other sources -

Now, malls mine customers data to offer better deals

- Data created worldwide 2010-2025

- Global Big Data market size 2011-2027

Pattern graph tracking-based stock price prediction using big data