

Big Data Analytics in Finance Industry

Gagan Arora
Indiana University
gkarora@iu.edu

ABSTRACT

Discusses the importance of Big data, data classification, cross industry comparison and analytic and its impact on finance industry and how customer satisfaction can be improved to achieve competitive advantage

KEYWORDS

Big data, finance, data classification, data type, big data impact, cross-industry comparison, banking, customer personalization, pattern recognition, multi document summarization, structured and unstructured data

1 INTRODUCTION: DISCUSS IMPORTANCE, HOW DOMINATING

By its nature of the business, the finance industry is always driven and dominated by data. The existence of Big data in the finance industry has exposed the big opportunity of growth and value extraction but at the same time imposed the various new challenges, which demand new skill set. [4] suggests that finance experts believe there is a huge potential in terms of value extraction from the financial big data. They also believe that finance industry can benefit more than any other industry. Historically, data was always there in some format either non-digital or digital. However, with digitalization, this data has fallen into the prevalence of high volume of information, which we call as Big Data. Dominant drivers for the actuality of big data in the finance industry are mainly customer call logs, social media, news feed, regulatory data etc. Call logs, news feed and etc. fall into the category of unstructured data which is identified as an area where we can extract vast amount of business value. We will discuss the various types of data, which is being generated at a phenomenal rate and what business value can be extracted out of big data. We will also discuss what all challenges this big data imposes on the finance industry. [5] talks about the effectiveness of extracting value out of big data in the finance industry and utilizing it to improve business operation.

2 MARKET IMPACT: PACE AT WHICH MARKET IS ADDING DATA

[2] talks about the three V of big data in finance industry: volume, velocity and variety. We will also discuss the fourth V aspect of it in a later section, which is a vulnerability. Fig 1 [4] clearly depicts the amount of financial data pouring in the daily basis. TechNavio's forecast (Technavio 2016) predicts data will grow at a CAGR [compound annual growth rate] of 61 percent over the period of 2017-2021. According to the IDC financial insight 2016, every second there is around 10,000-payment card transaction and this number is expected to double by the end of this decade. The Capgemini/RBS Global payments study for 2012 suggests there was about 260 billion transactions in 2012 and is expected to grow

between 15 and 22 percent for developing countries. Main drivers contributing to the big data in the finance industry are Data growth, increasing scrutiny from regulators, digitalization of financial products, changing the business model and increased customer insight platforms such as customer service. Fig 2 [[2]] shows 76 percent of banks say the business driver for embracing big data is to enhance customer engagement, retention, and loyalty and 71 percent of banks say that to increase their revenue, they need to better understand customers and big data will help them to do so.

3 COMPETITION/PROBLEM: HOW PEER INDUSTRY ADDING DATA AND THEIR INITIATIVE

Thinking about the data strategy, the financial industry has taken the business-driven approach to a big data. According to the IBM report, all financial organizations are not keeping the same pace as peer industry is keeping. Today because of increased competition, customers always expect more personalized banking service and at the same time, there is increased regulatory surveillance which in result creates big pressure on finance industry to better utilize the value of Big data. To achieve better-personalized experience, many banks have started the initiative to utilize the information gained from the vast ocean of data to offer better-personalized products and gain competitive advantage. Despite the fact that financial industry is data-driven, there is a gap in the amount of initiative financial industry has taken to extract the value out of big financial data. Technavio 2016 report has shown only 26 percent of financial organizations has focused on understanding the principal notation of Big data and most of those 26 percent are still struggling to define the clear roadmap. This clearly concludes that finance industry lag behind their cross-industry peers in using more varied data types. A good example to support this fact is that there are very less research and domain knowledge in extracting value out of retail bank call logs.

4 WHY BIG DATA IN FINANCE AND ITS IMPACT ON THE CUSTOMER PERSONALIZATION AND THEIR SATISFACTION

Big data technologies not only help in extracting the effective business value but analysis of unstructured data in conjunction with a wide variety of data set also helps in extracting commercial value. Big data in finance industry does not necessarily decode to valuable or actionable information. The real benefit lies in developing the technologies, which can be used to extract business and commercial value. [1] talks about what all advantage we can extract from the big data in the finance industry. Few examples are: Detection of false rumors that try to manipulate the finance market, Assessment of exposure to a reputational risk connected to consulting service

offered by banks to their customer and Discover topic trends, detect events, or support the portfolio optimization or asset allocation. Big data based pattern recognition can also help in enhanced fraud detection systems and prevention capability systems. Other benefits of utilizing big data include building a machine learning based algorithm to achieve higher performance and accuracy in the trading algorithm and Enhanced market trading analysis. There has been proven research [7] which states more data increases accuracy and precision of simulations which is the backbone of financial modeling based analytics. This research [7] states Modern modeling techniques are data hungry

5 BIG DATA CLASSIFICATION IN FINANCE INDUSTRY: STRUCTURED UN-STRUCTURED AND SEMI-STRUCTURED

Financial service system has a varied variety of data pools that are held by various stakeholders. At a high level of abstraction, we can classify them into three major categories: Structured data, Unstructured data, and Semi-structured data. With the emergence of too much data supply, there has been operational intelligence initiative such as firms like SPLUNK which uses data mining approaches to fetch valuable information out of any type of log. There have been studies [3] which shows utilizing structured data for analyzing event logs. Another advantage of structured data is that it makes the concept of Data Virtualization easy as data can easily be virtualized if we have structured data, which in turn make easy to extract patterns. Extracting patterns from the customer banking activities gives banks competitive advantage as they can make better personalized financial products for customers.

5.1 Structured data

This reflects the data which has a higher degree of an organization such as a relational database where information/data is easily searchable and we can easily apply standard algorithm to extract patterns out of it. Examples of such data set include Trading applications, Enterprise finance resource planner, Retail banking systems, Credit history database systems and other financial applications that use legacy application systems. Structured data always has a big advantage of being easily entered, stored, queried and analyzed. Most of the personal banking financial statements are stored in a structured way. Structured dataset combined with the distributed systems can be leveraged to achieve Structured big data set on which we can run optimized SQL queries to retrieve patterns. [6] discusses various SQL based ways to specify information quality in data which can be used to filter out the noise.

5.2 Unstructured data

With the emergence of social media, blogging and mobile usage there has been a phenomenal amount of data which we can classify as unstructured data. Example of unstructured data includes Daily stock feeds, Company announcements, Finance news, Articles, Blogs, Customer feedback/reviews and etc. There have been researches such as multi-document summarization and machine learning algorithms to utilize the unstructured data to extract value.

There is a big advantage with the unstructured data that it is a platform, programming language, technology compatible ie. Two or more machines which different platform can interact with each other using unstructured data. This means financial big unstructured data can be stored on distributed systems and pattern recognition application can be used to extract value. There are also other ways such as transforming unstructured data to structured and then fetching intelligence out of it since structured data is akin to machine language.

5.3 Semi-structured data

As the name suggests this data type includes the aspect of both structured as well as an unstructured data type. Examples of semi-structured data includes: Financial products markup language[FpML], Financial Information eXchange[FIX], Interactive Financial eXchange(IFX), Open Financial eXchange(FEDI) , Market data definition language (MDDL) and etc. [5] suggests nowadays semi-structured data dominates the data in finance industry which contributes to around 80-85 percent of finance data

6 VARIOUS CHALLENGES UTILIZING BIG DATA VALUE IN FINANCE INDUSTRY

There are multiple challenges and constraints in extracting value out of big financial data. The biggest challenge is old IT culture and infrastructure. The Much financial organization still uses old IT infrastructure which is not compatible with the big data application thus fail to take advantage of big data. Other challenges include lack of skill set and data privacy and security. With the emergence of digitalization, customer data is saved persistently because of which there has been continued concern regarding the customer privacy. Regulatory bodies guidelines on customer data are always ill-defined because of which is there is always a concern regarding the use of customer data.

7 REQUIREMENTS SOLUTION AND CONCLUSION

In this section, we will discuss various technical requirements needed to achieve value extraction from the big data in the finance industry. There are various technical requirements such as Data Acquisition, Data Quality, Data Extraction, Data Integration, Decision support. In order to fulfill requirements, a hybrid approach combining computer science, algorithms, statistics, data mining, machine learning and pattern recognition study needs to be adopted. To explore the advantage of big data there have been initiatives like data virtualization, multi-document summarization, pattern recognition from LOGS and many start-ups have been emerged. All big companies such as Microsoft, Google, IBM and Amazon are investing heavily in this field to leverage business and commercial value out of it. There has been changed in the industry pattern where financial industry is resorting big data to strategize their business. According to [4] with a very rapid pace, the financial industry is utilizing big data advantage in investment analysis, econometrics, risk assessment, fraud detection, trading, customer interaction analysis and behavior modeling. If we look at the Big promise the Big data holds in the finance industry, progress in this field is still in nascent stage and we expect more growth in upcoming years.

[Figure 1 about here.]

[Figure 2 about here.]

REFERENCES

- [1] Wahlster W. (eds) Cavanillas J., Curry E. 2016. *Big Data-Driven Innovation in Industrial Sectors*.
- [2] Daniel D. Gutierrez. 2014. *Big Data for Finance*. Technical Report. https://whitepapers.em360tech.com/wp-content/files_mf/1427803213insideBIGDATAGuidetoBigDataforFinance.pdf
- [3] M. Hinkka, T. Lehto, and K. Heljanko. 2016. Assessing Big Data SQL Frameworks for Analyzing Event Logs. In *2016 24th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing (PDP)*. 101–108. <https://doi.org/10.1109/PDP.2016.26>
- [4] Kazim Hussain and Elsa Prieto. 2015. *Big Data in Finance*.
- [5] Kazim Hussain and Elsa Prieto. 2016. *Big Data in the Finance and Insurance Sectors*.
- [6] A. Parssian, W. Yeoh, and M. S. Ee. 2015. Quality-Based SQL: Specifying Information Quality in Relational Database Queries. *Computer* 48, 9 (Sept 2015), 69–74. <https://doi.org/10.1109/MC.2015.264>
- [7] Tjeerd van der Ploeg, Peter C. Austin, and Ewout W. Steyerberg. 2014. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. *BMC Medical Research Methodology* 14, 1 (22 Dec 2014), 137. <https://doi.org/10.1186/1471-2288-14-137>

8 BIBTEX ISSUES

Warning—empty publisher in Ref4

Warning—empty address in Ref4

Warning—empty chapter and pages in Ref4

Warning—empty institution in Ref2

Warning—empty publisher in Ref6

Warning—empty address in Ref6

Warning—empty publisher in Ref3

Warning—empty address in Ref3

Warning—empty chapter and pages in Ref3

Warning—empty publisher in Ref1

Warning—empty address in Ref1

Warning—empty chapter and pages in Ref1

(There were 12 warnings)

9 ISSUES

DONE:

Example of done item: Once you fix an item, change TODO to DONE

9.1 Formatting

Incorrect number of keywords and i523 not included in the keywords

9.2 Writing Errors

A few punctuation errors

9.3 Structural Issues

You may want to add a few introductory sentences to the Abstract

9.4 Details about the Figures and Tables

In case you copied a figure from another paper you need to ask for copyright permission. In case of a class paper, you must include a reference to the original in the caption

The figure's caption must include a meaningful description of the figure, along with citation of the reference from which the image is copied

Do use *label* and *ref* to automatically create figure numbers. Refer to them in the text properly

Remove any figure that is not referred to explicitly in the text (As shown in Figure ..)

Figures should be reasonably sized and often you just need to add `columnwidth`

e.g.

```
/includegraphics[width=\columnwidth]{images/myimage.pdf}  
re
```

LIST OF FIGURES

1	Figure 1	6
2	Figure 2	7

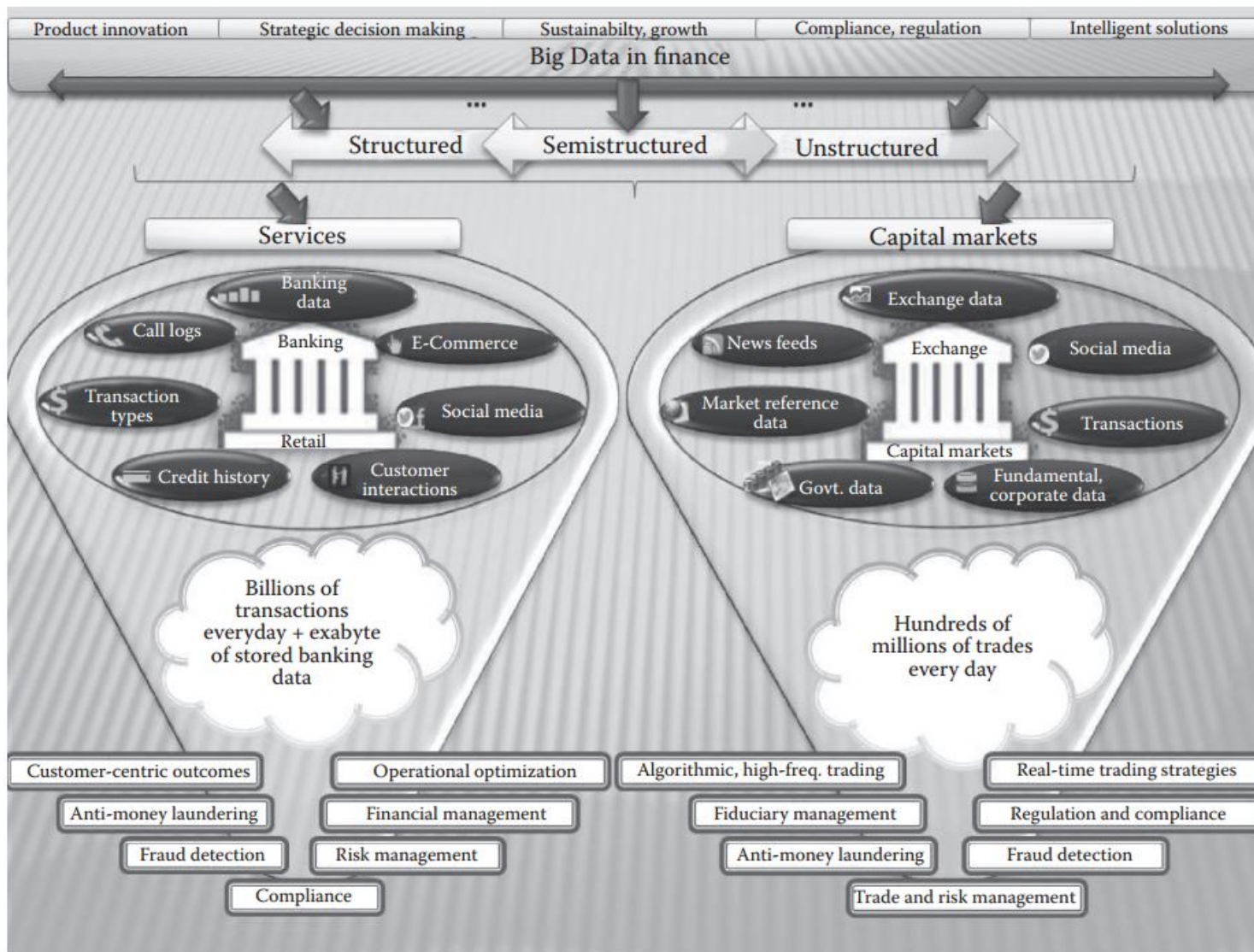


Figure 1: Figure 1

Banking on Big Data

76

Percentage of banks that say the business driver for embracing big data is to enhance customer engagement, retention and loyalty

71

Percentage of banks that say that to increase their revenue, they need to better understand customers and big data will help them do so

55

Percentage of banks that say that having a real-time view of data provides a significant competitive advantage and believe that batch mode data is ineffective

BASED ON A SURVEY OF 183 BANKERS WORLDWIDE; SOURCE: NGDATA IN CONJUNCTION WITH FINEXTRA AND CLEAR2PAY

Figure 2: Figure 2