IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 5.731
Volume 2, Issue 6, January 2022

Mobile Phone Data as a New Big Data

Mr. Atul Yadav
Shri G. P. M. Degree College, Vile Parle (E), Mumbai, Maharashtra, India

Abstract: *This paper introduces mobile phone data as a new Big Data source for credit scoring and shows that while it is a powerful source of information, it should be used strictly in a positive framework to increase the access to financing to borrowers who would otherwise be out of options until a much later stage. To motivate the use of this information in financial institutions, its potential is studied in both statistical and profit terms*

Keywords: Big Data

I. INTRODUCTION

This paper introduces mobile phone data as a new Big Data source for credit scoring and shows that while it is a powerful source of information, it should be used strictly in a positive framework to increase the access to financing to borrowers who would otherwise be out of options until a much later stage. To motivate the use of this information in financial institutions, its potential is studied in both statistical and profit terms.

Big data analysis in credit scoring refers to the comprehensive examination of extensive and diverse datasets using advanced analytical techniques to evaluate an individual's creditworthiness. This approach involves leveraging large volumes of data, including traditional financial information, transaction histories, social media activities, and other unconventional sources. By employing sophisticated algorithms and statistical methods, financial institutions aim to extract meaningful insights, patterns, and correlations from the data. The goal is to enhance the accuracy of credit scoring models, allowing for more informed and nuanced lending decisions based on a holistic understanding of an individual's financial behavior and risk profile. Data which are very large in size is called Big Data. Normally we work on data of size MB(WordDoc ,Excel) or maximum GB(Movies, Codes) but data in Peta bytes i.e. 10^15 byte size is called Big Data. It is stated that almost 90% of today's data has been generated in the past 3 years. Big Data is a collection of data that is huge in volume, yet growing exponentially with time. It is a data with so large size and complexity that none of traditional data management tools can store it or process it efficiently. Big data is also a data but with huge size. We produce a massive amount of data each day, whether we know about it or not. Every click on the internet, every bank transaction, every video we watch on YouTube, every email we send, every like on our Instagram post makes up data for tech companies. With such a massive amount of data being collected, it only makes sense for companies to use this data to understand their customers and their behavior better. This is the reason why the popularity of Data Science has grown manifold over the last few years. Let’s try to understand what is big data and its benefits and uses!

Definition

Big data analytics describes the process of uncovering trends, patterns, and correlations in large amounts of raw data to help make data-informed decisions. These processes use familiar statistical analysis techniques—like clustering and regression—and apply them to more extensive datasets with the help of newer tools. Big data has been a buzz word since the early 2000s, when software and hardware capabilities made it possible for organizations to handle large amounts of unstructured data. Since then, new technologies—from Amazon to smartphones—have contributed even more to the substantial amounts of data available to organizations. With the explosion of data, early innovation projects like Hadoop, Spark, and NoSQL databases were created for the storage and processing of big data. This field continues to evolve as data engineers look for ways to integrate the vast amounts of complex information created by sensors, networks, transactions, smart devices, web usage, and more. Even now, big data analytics methods are being used with emerging technologies, like machine learning, to discover and scale more complex insights.Big data analytics tools and technology

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 5.731
Volume 2, Issue 6, January 2022

Big data analytics cannot be narrowed down to a single tool or technology. Instead, several types of tools work together to help you collect, process, cleanse, and analyze big data. Some of the major players in big data ecosystems are listed below.

Big data analytics tools and technology

Big data analytics cannot be narrowed down to a single tool or technology. Instead, several types of tools work together to help you collect, process, cleanse, and analyze big data. Some of the major players in big data ecosystems are listed below.Hadoop is an open- source framework that efficiently stores and processes big datasets on clusters of commodity hardware. This framework is free and can handle large amounts of structured and unstructured data, making it a valuable mainstay for any big data operation.

NoSQL databases are non-relational data management systems that do not require a fixed scheme, making them a great option for big, raw, unstructured data. NoSQL stands for "not only SQL," and these databases can handle a variety of data models.MapReduce is an essential component to the Hadoop framework serving two functions. The first is mapping, which filters data to various nodes within the cluster. The second is reducing, which organizes and reduces the results from each node to answer a query.YARN stands for "Yet Another Resource Negotiator." It is another component of second-generation Hadoop. The cluster management technology helps with job scheduling and resource management in the cluster.Spark is an open source cluster computing framework that uses implicit data parallelism and fault tolerance to provide an interface for programming entire clusters. Spark can handle both batch and stream processing for fast computation.Tableau is an end- to-end data analytics platform that allows you to prep, analyze, collaborate, and share your big data insights. Tableau excels in self- service visual analysis, allowing people to ask new questions of governed big data and easily share those insights across the organization.The ability to analyze more data at a faster rate can provide big benefits to an organization, allowing it to more efficiently use data to answer important questions. Big data analytics is important because it lets organizations use colossal amounts of data in multiple formats from multiple sources to identify opportunities and risks, helping organizations move quickly and improve their bottom lines. Some benefits of big data analytics include:

Cost savings. Helping organizations identify ways to do business more efficientlyProduct development. Providing a better understanding of customer needsMarket insights. Tracking purchase behavior and market trends Big data brings big benefits, but it also brings big challenges such new privacy and security concerns, accessibility for business users, and choosing the right solutions for your business needs. To capitalize on incoming data, organizations will have to address the following:

Making big data accessible. Collecting and processing data becomes more difficult as the amount of data grows. Organizations must make data easy and convenient for data owners of all skill levels to use.

Maintaining quality data. With so much data to maintain, organizations are spending more time than ever before scrubbing for duplicates, errors, absences, conflicts, and inconsistencies.

Keeping data secure. As the amount of data grows, so do privacy and security concerns. Organizations will need to strive for compliance and put tight data processes in place before they take advantage of big data.

Finding the right tools and platforms. New technologies for processing and analyzing big data are developed all the time. Organizations must find the right technology to work within their established ecosystems and address their particular needs. Often, the right solution is also a flexible solution that can accommodate future infrastructure changes.

Big Data Scoring is an online, cloud-based service that helps consumer lenders like banks and leasing companies to improve their loan quality, acceptance rates and credit losses using big data.Big Credit Scoring collects data from a variety of public sources - including social media, Google searches and IP addresses - to build up a picture of a potential client's payment behaviour. It's aimed at helping underwriters to do their jobs, especially with more difficult clients like millennials, or in emerging markets where credit information is scarce.

The company promises better credit quality and smaller credit losses based on this client profile, because it helps the consumer lender make better decisions.Big data analytics refers to collecting, processing, cleaning, and analyzing large datasets to help organizations operationalize their big data.

1. Collect Data

Data collection looks different for every organization. With today's technology, organizations can gather both structured and unstructured data from a variety of sources — from cloud storage to mobile applications to in-store IoT

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 5.731
Volume 2, Issue 6, January 2022

sensors and beyond. Some data will be stored in data warehouses where business intelligence tools and solutions can access it easily. Raw or unstructured data that is too diverse or complex for a warehouse may be assigned metadata and stored in a data lake.

2. Process Data

Once data is collected and stored, it must be organized properly to get accurate results on analytical queries, especially when it's large and unstructured. Available data is growing exponentially, making data processing a challenge for organizations. One processing option is batch processing, which looks at large data blocks over time. Batch processing is useful when there is a longer turnaround time between collecting and analyzing data. Stream processing looks at small batches of data at once, shortening the delay time between collection and analysis for quicker decision-making. Stream processing is more complex and often more expensive.

3. Clean Data

Data big or small requires scrubbing to improve data quality and get stronger results; all data must be formatted correctly, and any duplicative or irrelevant data must be eliminated or accounted for. Dirty data can obscure and mislead, creating flawed insights.

4. Analyze Data

Getting big data into a usable state takes time. Once it's ready, advanced analytics processes can turn big data into big insights. Some of these big data analysis methods include:

Data mining sorts through large datasets to identify patterns and relationships by identifying anomalies and creating data clusters.

Predictive analytics uses an organization's historical data to make predictions about the future, identifying upcoming risks and opportunities.

Deep learning imitates human learning patterns by using artificial intelligence and machine learning to layer algorithms and find patterns in the most complex and abstract data.Big data brings big benefits, but it also brings big challenges such new privacy and security concerns, accessibility for business users, and choosing the right solutions for your business needs. To capitalize on incoming data, organizations will have to address the following:

Making big data accessible. Collecting and processing data becomes more difficult as the amount of data grows. Organizations must make data easy and convenient for data owners of all skill levels to use.

Maintaining quality data. With so much data to maintain, organizations are spending more time than ever before scrubbing for duplicates, errors, absences, conflicts, and inconsistencies.

Keeping data secure. As the amount of data grows, so do privacy and security concerns. Organizations will need to strive for compliance and put tight data processes in place before they take advantage of big data.

Finding the right tools and platforms. New technologies for processing and analyzing big data are developed all the time. Organizations must find the right technology to work within their established ecosystems and address their particular needs. Often, the right solution is also a flexible solution that can accommodate future infrastructure changes.

Big data and statistical computing empower banks to detect potential fraud before it even occurs. Specialized algorithms track and analyze spending and behavioral patterns, allowing banks to identify individuals who may be at risk of committing fraud. Retail banks, investment banks, and other financial organizations often have dedicated Risk Management departments that can prevent fraud and that heavily rely on big data analysis and Business Intelligence (BI) tools.

Analyzing big data for credit scoring involves processing large datasets to assess creditworthiness. It includes factors like financial history, transaction patterns, and other relevant data to predict a borrower's risk. Advanced analytics, machine learning, and statistical models play a crucial role in making accurate predictions and improving traditional credit scoring methods.Big data analysis of credit scoring refers to the use of extensive and diverse datasets, often including financial transactions, payment history, and various other relevant information, to assess and predict an individual's creditworthiness. This approach leverages advanced analytics, machine learning algorithms, and statistical models to provide more accurate and nuanced evaluations, enhancing traditional credit scoring methods by

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 5.731
Volume 2, Issue 6, January 2022

incorporating a broader range of data sources and factors for decision-making.Big data analysis involves extracting valuable insights from massive and diverse datasets through advanced processing techniques, often utilizing technologies like machine learning and analytics to uncover patterns, trends, and correlations that can inform decision-making and improve understanding in various domains.Big data analysis is the process of examining and interpreting large and complex datasets to uncover meaningful patterns, trends, and insights. It involves three main components:

1. Volume: Big data typically involves a massive amount of information, often beyond the capacity of traditional databases to handle efficiently.
2. Velocity: Data is generated and collected at high speeds in real- time. The analysis must keep up with the rapid influx of new data.
3. Variety: Big data comes in various formats, including structured (like databases), semi-structured (like JSON or XML files), and unstructured (like text documents or social media posts).

Key techniques in big data analysis include:

- Machine Learning: Algorithms and models that enable systems to learn from data and make predictions or decisions without explicit programming.
- Data Mining: Extracting patterns and knowledge from large datasets using statistical methods, machine learning, and artificial intelligence.
- Predictive Analytics: Using statistical algorithms and machine learning techniques to identify the likelihood of future outcomes based on historical data.
- Data Visualization: Presenting complex data in visual formats like charts or graphs to make it more understandable and actionable.
- Distributed Computing: Utilizing parallel processing and distributed storage systems to handle and process large datasets efficiently.

Industries such as finance, healthcare, marketing, and logistics extensively use big data analysis to gain insights, optimize processes, and make data-driven decisions.

Credit scoring is a statistical analysis performed by lenders and financial institutions to determine the creditworthiness of a person or a small, owner-operated business. Credit scoring is used by lenders to help decide whether to extend or deny credit. A credit score can impact your ability to qualify for financial products like mortgages, auto loans, credit cards, and private loans.

KEY TAKEAWAYS

Credit scores determine a person's ability to borrow money for mortgages, auto loans, and personal loans.

FICO and VantageScore are both popular credit scoring models. Lenders use credit scoring in risk-based pricing in which the terms of a loan, including the interest rate, offered to borrowers are based on the probability of repayment.

Credit ratings apply to corporations and governments, while credit scoring applies to individuals and small, owner-operated businesses.A credit score ranges from 300 to 850. The higher your credit score, the better your standing as a borrower is. A respectable credit score is above 670Numerical Evolution of Creditworthiness in Short Notes

1. Credit Scores Range: Typically ranges from 300 to 850.
2. Higher is Better: Higher scores indicate better creditworthiness.
3. Risk Assessment: Lenders use scores to assess the risk of lending.
4. Credit History Impact: Reflects credit history and payment behavior.
5. Common Factors: Includes payment history, credit utilization, and more.
6. Financial Decision Influence: Influences loan approvals and interest rates.
7. Maintenance Importance: Regularly managing credit impacts the score positively.

Based on Traded History and Financial Habits Short Notes

1. Traded History Impact: Reflects one's history of financial transactions.
2. Payment Timeliness: Timely payments positively influence creditworthiness.
3. Credit Utilization: Managing credit balances impacts the score.
4. Debt Levels: High levels of debt may negatively affect the score.
5. Diverse Credit Types: A mix of credit types can be beneficial.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 5.731
Volume 2, Issue 6, January 2022

6. Length of Credit History: Longer histories can enhance creditworthiness.Big data analysis in credit scoring refers to the comprehensive examination of extensive and diverse datasets using advanced analytical techniques to evaluate an individual's creditworthiness. This approach involves leveraging large volumes of data, including traditional financial information, transaction histories, social media activities, and other unconventional sources. By employing sophisticated algorithms and statistical methods, financial institutions aim to extract meaningful insights, patterns, and correlations from the data. The goal is to enhance the accuracy of credit scoring models, allowing for more informed and nuanced lending decisions based on a holistic understanding of an individual's financial behavior and risk profile.Credit scoring is without a doubt one of the oldest applications of analytics. In recent years, a multitude of sophisticated classification techniques have been developed to improve the statistical performance of credit scoring models. Instead of focusing on the techniques themselves, this paper leverages alternative data sources to enhance both statistical and economic model performance. The study demonstrates how including call networks, in the context of positive credit information, as a new Big Data source has added value in terms of profit by applying a profit measure and profit-based feature selection. A unique combination of datasets, including call-detail records, credit and debit account information of customers is used to create scorecards for credit card applicants. Call-detail records are used to build call networks and advanced social network analytics techniques are applied to propagate influence from prior defaulters throughout the network to produce influence scores. The results show that combining call-detail records with traditional data in credit scoring models significantly increases their performance when measured in AUC. In terms of profit, the best model is the one built with only calling behavior features. In addition, the calling behavior features are the most predictive in other models, both in terms of statistical and economic performance. The results have an impact in terms of ethical use of call-detail records, regulatory implications, financial inclusion, as well as data sharing and privacy.

Credit scoring is without a doubt one of the oldest applications of analytics. In recent years, a multitude of sophisticated classification techniques have been developed to improve the statistical performance of credit scoring models. Instead of focusing on the techniques themselves, this paper leverages alternative data sources to enhance both statistical and economic model performance. The study demonstrates how including call networks, in the context of positive credit information, as a new Big Data source has added value in terms of profit by applying a profit measure and profit-based feature selection. A unique combination of datasets, including call-detail records, credit and debit account information of customers is used to create scorecards for credit card applicants. Call-detail records are used to build call networks and advanced social network analytics techniques are applied to propagate influence from prior defaulters throughout the network to produce influence scores. The results show that combining call-detail records with traditional data in credit scoring models significantly increases their performance when measured in AUC. In terms of profit, the best model is the one built with only calling behavior features. In addition, the calling behavior features are the most predictive in other models, both in terms of statistical and economic performance. The results have an impact in terms of ethical use of call-detail records, regulatory implications, financial inclusion, as well as data sharing and privacy. Data analysts, data scientists, predictive modelers, statisticians and other analytics professionals collect, process, clean and analyze growing volumes of structured transaction data, as well as other forms of data not used by conventional BI and analytics programs.

The following is an overview of the four steps of the big data analytics process:

1. Data professionals collect data from a variety of different sources. Often, it's a mix of semistructured and unstructured data. While each organization uses different data streams, some common sources include the following:

o Internet clickstream data.
o Web server logs.
o Cloud applications.
o Mobile applications.
o Social media content.
o Text from customer emails and survey responses.
o Mobile phone records.
o Machine data captured by sensors connected to the internet of things.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 5.731
Volume 2, Issue 6, January 2022

2. Data is prepared and processed. After data is collected and stored in a data warehouse or data lake, data professionals must organize, configure and partition the data properly for analytical queries. Thorough data preparation and processing results in higher performance from analytical queries. Sometimes this processing is batch processing, with large data sets analyzed over time after being received; other times it takes the form of stream processing, where small data sets are analyzed in near real time, which can increase the speed of analysis.
3. Data is cleansed to improve its quality. Data professionals scrub the data using scripting tools or data quality software. They look for any errors or inconsistencies, such as duplications or formatting mistakes, and organize and tidy the data.
4. The collected, processed and cleaned data is analyzed using analytics software. This includes tools for the following:
o Data mining, which sifts through data sets in search of patterns and relationships.
o Predictive analytics, which builds models to forecast customer behavior and other future actions, scenarios and trends.
o Machine learning, which taps various algorithms to analyze large data sets.
o Deep learning, which is a more advanced offshoot of machine learning.
o Text mining and statistical analysis software.
o Artificial intelligence.
o Mainstream BI software.
o Data visualization tools.
• Home
• Data science and analytics
What is big data analytics
Big data analytics is the often complex process of examining big data to uncover information -- such as hidden patterns, correlations, market trends and customer preferences -- that can help organizations make informed business decisions.
On a broad scale, data analytics technologies and techniques give organizations a way to analyze data sets and gather new information. Business intelligence (BI) queries answer basic questions about business operations and performance.
Big data analytics is a form of advanced analytics, which involve complex applications with elements such as predictive models, statistical algorithms and what-if analysis powered by analytics systems.
An example of big data analytics can be found in the healthcare industry, where millions of patient records, medical claims, clinical results, care management records and other data must be collected, aggregated, processed and analyzed.
Big data analytics is used for accounting, decision-making, predictive analytics and many other purposes. This data varies greatly in type, quality and accessibility, presenting significant challenges but also offering tremendous benefits.
Organizations can use big data analytics systems and software to make data- driven decisions that can improve their business-related outcomes. The benefits can include more effective marketing, new revenue opportunities, customer personalization and improved operational efficiency. With an effective strategy, these benefits can provide competitive advantages over competitors.
How does big data analytics work?
Data analysts, data scientists, predictive modelers, statisticians and other analytics professionals collect, process, clean and analyze growing volumes of structured transaction data, as well as other forms of data not used by conventional BI and analytics programs.The following is an overview of the four steps of the big data analytics process:
Data professionals collect data from a variety of different sources. Often, it's a mix of semistructured and unstructured data. While each organization uses different data streams, some common sources include the following:
o Internet clickstream data.
o Web server logs.
o Cloud applications.
o Mobile applications.
o Social media content.
o Text from customer emails and survey responses.
o Mobile phone records.
o Machine data captured by sensors connected to the internet of things.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 5.731
Volume 2, Issue 6, January 2022

2. Data is prepared and processed. After data is collected and stored in a data warehouse or data lake, data professionals must organize, configure and partition the data properly for analytical queries. Thorough data preparation and processing results in higher performance from analytical queries. Sometimes this processing is batch processing, with large data sets analyzed over time after being received; other times it takes the form of stream processing, where small data sets are analyzed in near real time, which can increase the speed of analysis.
3. Data is cleansed to improve its quality. Data professionals scrub the data using scripting tools or data quality software. They look for any errors or inconsistencies, such as duplications or formatting mistakes, and organize and tidy the data.
4. The collected, processed and cleaned data is analyzed using analytics software. This includes tools for the following:
o Data mining, which sifts through data sets in search of patterns and relationships.
o Predictive analytics, which builds models to forecast customer behavior and other future actions, scenarios and trends.
o Machine learning, which taps various algorithms to analyze large data sets.
o Deep learning, which is a more advanced offshoot of machine learning.
o Text mining and statistical analysis software.
o Artificial intelligence.
o Mainstream BI software.
o Data visualization tools.

Types of big data analytics

There are several different types of big data analytics, each with their own application within the enterprise.

• Descriptive analytics. This is the simplest form of analytics, where data is analyzed for general assessment and summarization. For example, in sales reporting, an organization can analyze the efficiency of marketing from such data.
• Diagnostic analytics. This refers to analytics that determine why a problem occurred. For example, this could include gathering and studying competitor pricing data to determine when a product's sales fell off because the competitor undercut it with a price drop.
• Predictive analytics. This refers to analysis that predicts what comes next. For example, this could include monitoring the performance of machines in a factory and comparing that data to historical data to determine when a machine is likely to break down or require maintenance or replacement.
• Prescriptive analytics. This form of analysis follows diagnostics and predictions. After an issue has been identified, it provides a recommendation of what can be done about it. For example, this could include addressing inconsistencies in supply chain that are causing pricing problems by identifying suppliers whose performance is unreliable, suggesting their replacement.

Key big data analytics technologies and tools

Many different types of tools and technologies are used to support big data analytics processes, including the following:

• Hadoop is an open source framework for storing and processing big data sets. Hadoop can handle large amounts of structured and unstructured data.
• Predictive analytics hardware and software process large amounts of complex data and use machine learning and statistical algorithms to make predictions about future event outcomes. Organizations use predictive analytics tools for fraud detection, marketing, risk assessment and operations.
• Stream analytics tools are used to filter, aggregate and analyze big data that might be stored in different formats or platforms.
• Distributed storage data is replicated, generally on a nonrelational database. This can be as a measure against independent node failures, lost or corrupted big data or to provide low-latency access.
• NoSQL databases are nonrelational data management systems that are useful when working with large sets of distributed data. NoSQL databases don't require a fixed schema, which makes them ideal for raw and unstructured data.
• A data lake is a large storage repository that holds native-format raw data until it's needed. Data lakes use a flat architecture.
• A data warehouse is a repository that stores large amounts of data collected by different sources. Data warehouses typically store data using predefined schemas.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 5.731
Volume 2, Issue 6, January 2022

• Knowledge discovery and big data mining tools help businesses mine large amounts of structured and unstructured big data.
• In-memory data fabric distributes large amounts of data across system memory resources. This helps provide low latency for data access and processing.
• Data virtualization enables data access without technical restrictions.
• Data integration software enables big data to be streamlined across different platforms, including Apache, Hadoop, MongoDB and Amazon EMR.
• Data quality software cleanses and enriches large data sets.
• Data preprocessing software prepares data for further analysis. Data is formatted and unstructured data is cleansed.
• Apache Spark is an open source cluster computing framework used for batch and stream data processing.
• Microsoft Power BI and Tableau end-to-end analytics platforms bring big data analytics to the desktop and back out to dashboards, with full suites of tools for analysis and reporting.
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 more 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.

II. CONCLUSION

This study presents the statistical and economic advantages of exploiting Big Data and social net-work analytics for credit scoring applications. We use phone call logs are used to build call networksand social network analytics applied to enhance the performance of models that predict creditworthi-ness of credit cards applicants. We do this from both a statistical and profit perspective and demon- strate how incorporating telco data has the potential of increasing the Value of credit scoring models.Furthermore, we identify which features are most important for this predictive task, both in terms ofstatistical performance and profit. According to the results, models that are built with features thatrepresent calling behavior perform best, both when performance is measured in AUC and profit. Wealso show that these features dominate other features in terms of importance. This is an interestingresult because it means that how people use their phones can be used as the sole data source whendeciding whether they should be given a loan or not. Thus we propose that the data should be usedin strict positive terms, to facilitate financial inclusion for people that lack enough information forcorrect profiling.The main limitation of our this is the data itself. The scorecards that were built are for the applica-tions of credit cards, and it is unclear how the results would generalize for other types of credits such asmicroloans or mortgages. In industry, numerous applications for granting microloans via smartphonesby analyzing user's behavior exist. According to various reports, behavioral features are important inthese applications as well, but that is difficult to verify without published scientific results. Similardata could be obtained from peer-to-peer lending platforms, or through agreements between telcosand banks/credit bureaus, where there is access to both default status of users as well as behavioral features. Behavioral data similar to the mobile phone data shown in this work could also be gatheredfrom social media platforms such as Twitter. The data in this study originates from a single countrywhere a telco and a bank have a special agreement to share the data. Therefore, an analysis of simi-lar data from other countries or data for other types of credits would strengthen the external validityof the presented results. In practice, lenders use credit bureau variables, such as FICO scores, whenassessing creditworthiness, and unfortunately they were not available for these analyses, but would bean interesting extension of our work.It is already clear that the mobile phone data used in this study is big in the sense of 'Volume','Velocity', 'Veracity' and 'Variety'. Our analysis of the data and the resulting well-performing modelsshow that it also has a positive effect for financial inclusion and on model profit, and as such is alsoimportant for 'Value': the fifth V of Big Data!