ACRN Journal of Finance and Risk Perspectives 8 (2019) Special Issue Digital Accounting, 16-30

BIG DATA, CLOUD COMPUTING AND DATA SCIENCE APPLICATIONS IN FINANCE AND ACCOUNTING

JENNIFER HUTTUNEN, JAANA JAUHIAINEN, LAURA LEHTI AND ANNINA NYLUND[1]
MINNA MARTIKAINEN[1] AND OTHMAR M LEHNER[2] (EDITORS)

[1] Hanken School of Economics, Helsinki
[2] University of Applied Sciences Upper Austria and Hanken School of Economics, Helsinki

Abstract*: The aim of this paper is to get acquainted with big data and how advanced analytical tools like data science and machine learning are helping the financial sector obtain greater insights about their businesses. Further applications also include acquiring a deeper understanding about ways to provide customers the next level experience in banking and/or investing. We have briefly touched on cloud computing as the platform to support the utilization of some of these technologies. The financial sector, including the accounting industry, will experience disruption albeit at an unknown pace. Despite the benefits derived from exploiting these emerging technologies, a slow uptake is still observed. The authors believe that a more receptive attitude towards innovative technologies could pave the way to more application in the field. Consequently, standards and regulations will evolve to keep up with these changes.*

Keywords*: Big Data; Data Science; Machine Learning; Analytics; Accounting; Finance*

Introduction

The digitalization of the world continues and new innovations within accounting and finance will affect every day work tasks. Without insights about what is currently happening on the field and what will most likely change in the foreseeable future, various professions will be put at risk. The gap between information technology and the traditional accounting and finance roles is predicted to rapidly diminish. The understanding of concepts such as Big Data, Cloud Computing, and Data Science Applications will be crucial for succeeding in the coming job market as a decision maker in financial sector. The study focuses on the efficient use of data for business decision-making in the financial industry. The paper examines the use of Big Data analysis and Data Science Applications to improve budgeting and investment decisions within accounting and finance field. As the topic itself is relatively new, the study is done as a literature review. Due to the scarcity of academic research materials, the review will include research articles complemented by various internet sources regarding current developments and ongoing applications in the field. The authors developed a theoretical framework in Figure 1 as a representation of the relationship present in this topic.

We begin by defining the relevant terms for this paper and proceed to discuss their relevance in the field of accounting and finance. We will also evaluate current developments and ongoing research in the field, including a description of commonly used machine learning algorithms in Big Data analysis. Application of these technologies in the field will be presented through some case examples. We then proceed with an investigation of the critical voices from the society and concerns related to the ongoing development. Lastly, we will conclude based on our investigations and include speculations about the outlook.

Figure 1: Theoretical framework, source: authors

Why and for who?

Aside from using Big Data for marketing purposes such as targeted advertising, business sectors including accounting and finance will experience a rapid change brought about by the changing technology. Both accounting and finance deals with large volumes of data, instead of physical products, but the actual use of Big Data and its analytical possibilities is still at a very early stage (Cockcroft & Russell, 2018). To progress, it is important that professionals within accounting and finance field manage to adopt some basic IT knowledge – since understanding and utilizing new techniques will alleviate monotonous manual work leaving more time for strategical thinking and decision making. This paper examines what is currently going on in accounting and finance and presents concepts related to the following topics:

- Efficient use of data for financial decision making in financial institutions
- Why: getting better insights for budgeting and investment decisions
- For who: decision maker in financial institutions, reporting

Key Terms and Concepts

Big Data

The term "Big Data" is defined in Oxford English Dictionary as: "n. Computing (also with capital initials) data of a very large size, typically to the extent that its manipulation and management present significant logistical challenges; (also) the branch of computing involving such data" (Oxford English Dictionary, 2008). Another way to try to understand the term Big Data is to examine the commonly agreed criteria of Big Data which can be described in three V's:

Volume – the total amount of data stored, exabytes and petabytes instead of terabytes
Velocity – the intensity in which new data is created and needs to be stored, i.e. every card transaction around the world recorded in data centers and data warehouses
Variety – the heterogeneity of the data and the mess it creates, for example videos, click streams, comments, numerical information and transactions creates data without any notable structure

Nowadays two more V's of Big Data has been widely recognized; Veracity and Value. These two V's refer to the fact that when more and more data is collected the trustworthiness of the data and the added value it brings come into question (Reca, 2018).

Big Data consist of both structured and unstructured data. Structured data has a pattern and includes often well-organized information which are easier to embrace than the latter one. Unstructured data is in a nutshell everything else, a mess of "loose" information as streams from social media, audio, location services and Internet of Things technologies (King, 2018). A comparison of structured and unstructured data is presented in Table 1.

Table 1: Structured vs. Unstructured Data table. Adapted from Structured vs. Unstructured Data by C. Taylor, 2018. Datamation. Retrieved from https://www.datamation.com/big-data/structured-vs-unstructured-data.html

	Structured Data	Unstructured Data
Characteristics	• Pre-defined data models • Usually text only • Easy to search	• No pre-defined data model data model • Maybe text, images, sound, video or other formats • Difficult to search
Resides in	• Relational databases • Data warehouses	• Applications • No SQL databases • Data warehouses • Data lakes
Generated by	Humans or machines	Humans or machines
Typical applications	• Airline reservation system • Inventory control • CRM systems • ERP systems	• Word processing • Presentation software • Email clients • Tools for viewing or editing media
Examples	• Dates • Phone numbers • Social security numbers • Credit card numbers • Customer names • Addresses • Product names and number • Transaction information	• Text files • Reports • Email messages • Audio files • Video files • Images • Surveillance imagery

To get value out of Big Data, combining both structured and unstructured data, analysis of it need to be made – and this is the real challenge. Information from different sources is useless if there is no way to find any structure in it or answers to specific problems and this cannot be done by humans due to the three V's mentioned before. Big Data can be envisaged as the base, all the information collected need to be stored somewhere which requires cloud computing and Data Science Applications uses Big Data to do different kind of analysis. However, notice that the exist of each of these components are not dependent on one other.

Data Science

Data science is an umbrella term for an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from data in various forms like data mining and machine learning (Hayashi, 1998). Data mining is a data analysis technique that aims to discover patterns in big data involving machine learning with a goal to extract information from a big data and transform the information into a comprehensible structure (Cliffon, 2018). Machine learning means studying of models that computer systems use to improve their performance on a specific task. Machine learning is based on algorithms that build a model of sample data in order to make predictions or decisions without being explicitly programmed to perform the task. Machine learning can be supervised or unsupervised learning. Supervised learning is something that human knows at teaches for the application like e.g. how to create invoices. Unsupervised learning means that the machine is given certain big data, but it is not told what to look for. The application studies the data, builds clusters and learns on its own. It can detect e.g. that there are some similarities in customers that tend to pay late or not pay at all (Koza, Bennet, Andre, & Keane, 1996).

Artificial intelligence (AI) is defined in Oxford Dictionary as "the theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages" (Oxford Living Dictionaries, n.d.). Meaning that AI is needed for machines to be able to learn to perform above mentioned tasks. AI can be described also as a science that describes how computers imitate human intelligence.

Analytics

Big data analytics is a form of advanced analytics. It allows the massive volumes of big data gathered to be processed. The process of extracting information can be categorised into five stages and two sub-groups, as described on Figure 2. Data Management is concerned with the technologies to acquire, store, and prepare data. Analytics is then concerned with extracting insights from the data gathered (Gandomi & Haider, 2015).

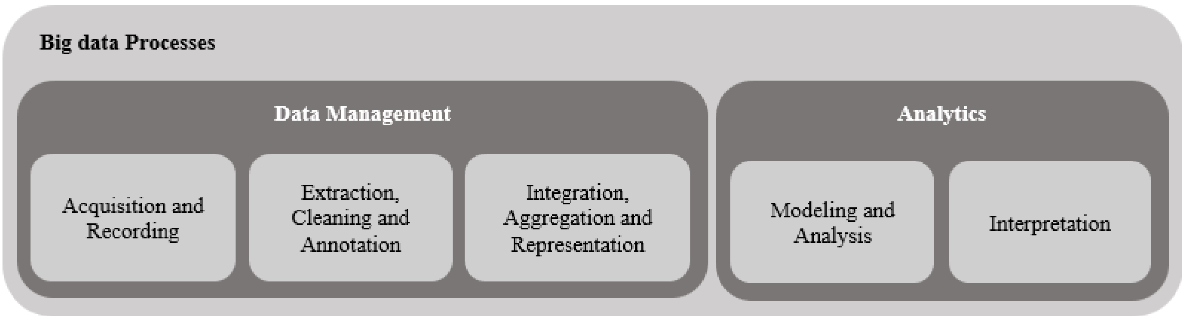

Figure 2: Big Data Processes. Adapted from Beyond the hype: Big data concepts, methods and analytics by A. Gandomi & M. Haider, 2015, International Journal of Information Management, 35(2), page 141.

An abundance of analytical techniques exists for structure and unstructured data. Text, audio, video, social media analytics are all examples of these categories. Analytics can be further classified into descriptive, diagnostic, predictive and prescriptive (Decide Soluciones, 2017).

Descriptive analytics describes what has happened, and diagnostic analytics answers why something has happened. This is generally considered the traditional level of analytics used in business intelligence. The further levels are known as advanced analytics. Prescriptive analytics describes what we should do. Prescriptive analytics anticipates what will happen in the future and suggest options for future courses of action. Predictive analytics, as the name states, predicts future events based on historical and current data. Using machine learning and AI, which are described in the following sections, predictive analytics seeks to capture patterns and relationships in data. These can be, for example, historical patterns or interdependencies between factors. These techniques are mainly based on statistical methods. (Gandomi & Haider, 2015) Predictive analytics can be seen as a subset of data science in general.

Cloud Computing

Cloud Computing is about using flexible infrastructure in todays' business. IBM defined cloud computing as the delivery of on-demand computing resources. For companies, this means the use of offsite shared servers hosted on the internet to store, manage, and process data instead of capitalizing on servers and computers locally which requires regular maintenance by the company. The flexibility of using the cloud endeared corporate consumers for the following reasons (EcourseReview, 2017; Jain, 2018):

On-Demand Service – use it when needed. This provides some degree of freedom for the customers.
Network Access – utilizes the internet and can be accessed using laptops, workstations, and smart phones.
Pooling of Resources – resources are pooled to provide customers customizable variable costs based on business size.
Scalability – scale up or down based on your current needs.

Cloud Computing Service providers offer different service models according to the customer's needs. The service models are called SaaS, PaaS, and IaaS. These models are often depicted in a pyramid-like structure as shown in Figure 3.

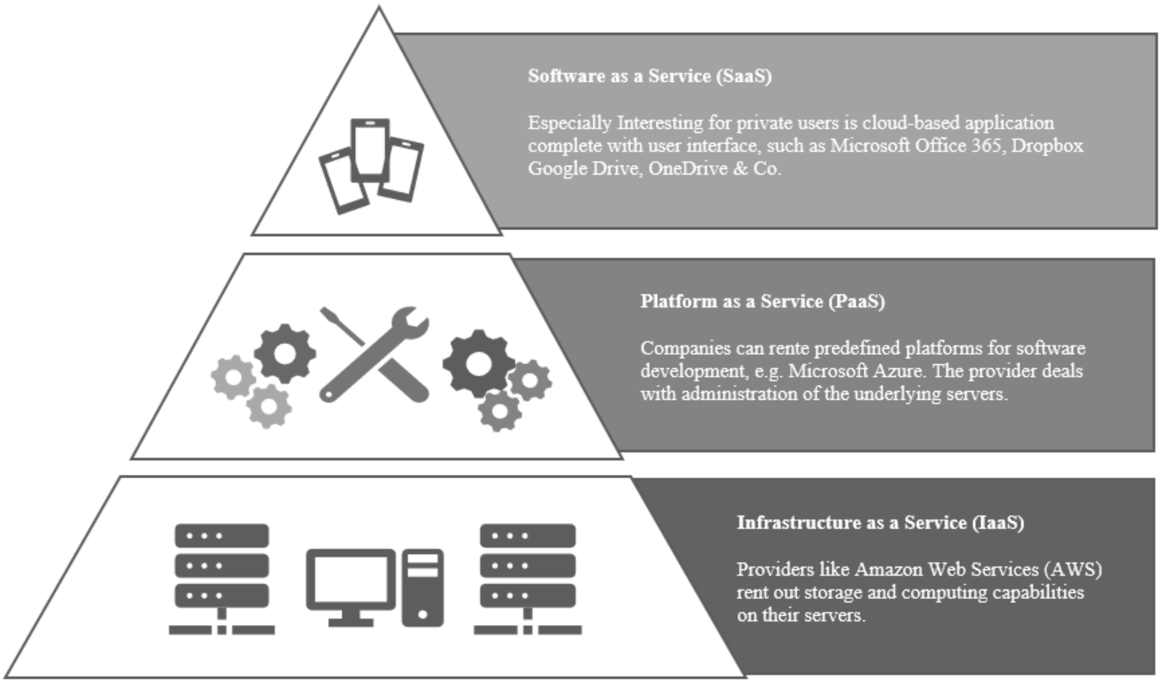

Figure 3: SaaS, PaaS, IaaS Pyramid Model. Adapted from "Cloud Definition" by Boxcryptor, 2017, Boxcryptor. Retrieved from https://www.boxcryptor.com/en/blog/post/what-is-the-cloud-a-beginner-s-guide/

Below are more examples of applications that utilizes these models:

SaaS – Banking applications, Social media apps, Slack, App-based games
PaaS – SAP, AWS Elastic Beanstalk, Google App Engine, Apache Stratos (Hou, n.d.)
IaaS – these are more for system administrators and include Rackspace, Google Compute Engine (GCE)

Machine Learning Algorithms

As mentioned, utilizing Big Data is impossible without data mining and machine learning. Following analytics algorithms are frequently used to discover patterns in Big Data:

Linear regression – One of the most basic and most used algorithms. This algorithm uses the relationship between two sets of continuous quantitative measures. Linear regression examines the relationship between the independent variable(s) and the dependent variable. The goal is to describe the dependent variable in terms of the independent variable(s), i.e. how time effects revenues, is there a relationship between age and income or what will the price of X be in 6 months.

Logistic regression – In difference from linear regression which examines two quantitative variables, logistic regression is used to solve problems involving categorization. The output variable values are distinct and finite instead of continuous and infinite as in linear regression. This algorithm can be used to answer clearly defined yes or no questions, i.e. credit worthiness or if it is likely that a customer will buy again. (Hiltbrand, 2018) Different regression analysis is most used for determine the strength of predictors, forecasting an effect and trend forecasting (Statistics Solutions, 2013).

Classification and Regression Trees – These algorithms use a decision, based on a question related to an input variable, to categorize data. The data is processed through a set of questions and for each corresponding decision it will move towards being categorized in a specific way. The way the data is processed creates a tree-like structure, where each set of question lines ends in a category called the leaf node of the tree. The classification and regression trees may easy become very large and complex, since a single tree can include several branches of logic. Hence, there is a variant version called random forest which consist of many small and simple trees instead of one big, which are then composited together to a final prediction. Classification and regression trees can be used to envisage multi-value categorizations (Hiltbrand, 2018).

K-Nearest Neighbors – A simple classification algorithm which allows for multivalued categorizations of the data. It is based on its samples in the training set and determines the distance of new samples to each data point, i.e. evaluates the nearest neighbours in each category. If the size and scope of the training set is very large, the process of this algorithm may be computationally expensive since the new samples needs to be compared to the whole training data set. Also, if the training data contains errors, the classifications will be misrepresented. KNN is common since it is easy to use and train, and the results that are simple to interpret. This algorithm is used i.e. in different search applications to suggest similar products (Hiltbrand, 2018).

K-Means Clustering – An algorithm commonly used in data exploration, which creates groups of instances with similar characteristics, and the different groups are called clusters. When the clusters are determined the new samples can be evaluated against the them to see which cluster is most suitable. The number of clusters are pre-defined where after the K-Means clustering process are dividing the data accordingly. However, the clusters are not yet categories, just closely related instances of input variables. The relevant clusters need to be discovered, analysed and renamed into business categories to bring value. It is also worth to notice that K-Means Clustering are sensitive to outliers which may distort the whole analysis (Hiltbrand, 2018).

Neural Network – In comparison to other algorithms, Neural Network (NN) is next level. Basically, NN is a framework for different machine learning algorithms to work together on complex data inputs. It mimics the human brain with its artificial neurons and enables the computer to learn, e.g. deep learning (AI) through a supervised or unsupervised learning process (DeMuro, 2018). Neural network can adapt to changing input without redesign the output criteria. The "neurons", or nodes, in NN are mathematical functions that collects and classifies information, i.e. in similarity to multiple linear regression. A neural network contains layers of these interconnected nodes, and each layer combined with the learning curve minimize the marginal of errors in the outputs. Neural network is used in applications as forecasting, trading, enterprise planning, risk assessment, fraud detection and Natural Language Processing (NLP) (Chen, 2019).

Current Research and Developments

The use of Big Data in accounting and finance is most researched within four areas: Financial distress modelling, Financial fraud modelling, Stock market prediction and Quantitative modelling, and Auditing. There has been progress with adapting research results into practice in the other areas, but auditors have been slow with implementing the results. One explanation could be that auditors are reluctant to implement technology which is far ahead of those used in their client firms. Although, some leading auditing firms have started to adapt Big Data techniques in practice – which can be a valuable addition to the profession especially combined with the expertise (Gepp, Linnenluecke, O'Neill & Smith, 2018).

What is the direction of development currently?

Financial technology applications vary from simple automation to complex decision-making tools. Many of these applications rely on big data and require investments in cloud infrastructure and analytics tools (Das, 2018). In the past years, financial companies have been looking at opportunities to shape themselves into lean and customer-centric organizations. In many cases, assuring access to viable financial and banking analytics is the key in this transformation. Analytics allows companies to make intelligent and data-driven decisions.

Many of today's intelligent tools answer the question "what is happening", but the more important question of, "why is this happening" is trickier to answer. Good analytics tools can help to shape this answer. Answering these important questions help financial institutions to figure out what their customers are looking for, how to shape the financial offerings to fit the customer needs, and how to best communicate to customers via marketing. Answering these questions require discovering new insights within the possessed data with data analytics.

However, it must be remembered that even the most sophisticated analytics tools are not a turnkey solution to success. Companies must educate their workforce to use these tools and make the best out of them. Some critical pitfalls must be avoided with caution, one classic example being the attempt to prove causation with correlation. Perhaps the most dangerous danger is the risk that users of data analytics are searching for answers with their own – sometimes unrecognized – bias. This causes them to find the answers they want to find, instead of the real ones that could be the stepping stone to the next innovation.

An example to the above-mentioned challenges is IBM's so called "augmented analytics", which infuses traditional business intelligence (BI) solutions with more analytical intelligence. Augmented intelligence refers to data analytics that can learn and adapt to different needs and guide users to new insights while removing the human bias from the equation. On the technical level, IBM's analytics technology includes machine learning, data discovery, natural language queries, automated pattern detection and sophisticated data storytelling capabilities (Walker, 2018).

As a practical example, IBM's analytics technology uses data patterns to identify clients with a high probability of accepting marketing offers and then adjust the campaigns accordingly. Some banking companies have seen increase of revenue by 60% in some customer segments. Some banks are now using analytics to do a fully automatized, yet highly personalized loan offers while automating the actual loan-approval process (Walker, 2018).

Forecasting

Data science plays a key role in many aspects of accounting and finance. It can boost competitiveness and profitability by allowing accounting professionals to make more accurate and detailed forecasts. Better forecasts are possible when the company can anticipate market trends better (Goh, C. 2017). Big data and predictive analytics have a key role in the development of forecast accuracy from the increasingly growing amounts of data available (Hassani, 2015). Current limitations to forecasting with big data are that traditional tools do not have the processing capabilities for the size, speed and complexity of the data, which poses challenges for organizations.

There is also a research made that suggest that business decision makers prefer to use complex forecast methods, including big data analytics when possible, instead of more simple methods. Surprisingly, the forecast accuracy did not improve when using complex methods and the errors even increased when using too much complicity in the forecast procedure. This only added unnecessary complication only because of a user's preference. The chosen forecast method should always be understandable for the user to avoid errors (Green, 2015). Though, the veracity and value of the Big Data used in forecasts will probably play a big role in the accuracy.

However, improvement in forecasting will provide immense opportunities. Financial institutions can forecast the bankruptcy of a firm better and predict defaulters of loans with more accuracy. This can be extended to predict the reliability of a credit card applicants' tendency to delay payments or risk of defaulting. Stocks can be ranked more accurately based on their risk, allowing investors to create portfolios based on their risk tolerance level. More accurate forecasting of interest rates can affect investment decisions.

Budgeting

A common challenge in budgeting is that setting goals and strategy, going through scorecards, reports and forecasts is time consuming. Often also the technology is designed for individual productivity rather than company-wide collaboration. Data science can optimize budgeting since data analytics tools allow companies to combine diverse financial and non-financial data and produce more comprehensive reporting (Goh, 2017).

The above-mentioned IBM planning analytics tool is giving a good example of how to automate budgeting and forecasting process. It creates reports and analyses from data it has collected from numerous sources and aggregated based on the models that were updated there. It is reducing the time spent on data analysis up to 70 % by same time improving accountability and reliability since it's using more information than what human could use (Anderson, 2010).

Basically, anything that can be read (text, pdf, html) can be read automatically nowadays which removes human working hours and makes the time-to-market faster. Natural language processing (NLP) that combines neural networks with other algorithms can read financial statements. Machine learning and data science is used for forecasting algorithms for budgeting and it is possible to compare multiple future outcomes and customize the budgeting model to various assets and portfolios under review.

Risk Management

Analytics helps companies manage risks, as it enables continuous auditing and monitoring in a wide range of areas within the company. These areas can include for example the regulatory environment, the supply chain, or business strategy. Banks and other financial institutions can use these analytics techniques also for fraud prevention. Managing risks in both financial and operational environments is crucial for financial institutions.

Banks store vast amounts of data which can be used to analyse behavioural patterns of both new and existing customers. This can help the bank anticipate customer needs, their response to new products, and it can help the bank price loans better and calculate probability of repayment more carefully (Peric, Kozarevic & Polic, 2016). Risk management is a key area of big data analytics and applications in the financial industry, and it has only grown in importance after the recent financial crisis (Hassani, 2015). Additionally, to the previously mentioned credit risk and performance analysis, compliance and regulatory reporting and insurance evaluation can also be analysed and improved through big data.

Financial risk management refers to use of financial instruments by a firm to manage exposure to various types of risk, such as operational, credit, market, foreign exchange or liquidity risk. The key sources of risk are analysed and measured, and measures are taken to address them. Machine learning can be used to identify, prioritize and monitor these various risks. Algorithms can help develop more accurate risk scoring models, additionally being more cost-effective. An example of an area where machine learning and AI can be applied is evaluating creditworthiness. Machine learning can be used to analyse past behaviours and spending patterns. Zheng et al. (2018) describe that fintech can reach a level of technology described as 'intelligent finance', where financial technology can integrate the internet, finance and big data to achieve more precise and faster calculations and transform the industry through data science technology.

As mentioned earlier, improved forecasting methods can also help improve risk management processes. There are massive amounts of customer data available for financial institutions to use, which vary both in volume and structure. Processing semi-structured or unstructured data can be challenging and time consuming. For instance, natural language processing, data mining and text analytics are techniques that can transform masses of data into more easily usable forms. The financial industry already produces masses of highly structured data which can be utilized when developing AI technologies (Zheng et al. 2018).

The previously described predictive analytics can help cover patterns in data that indicate some specific event happening in the future. This event can then be acted upon already in the present day. For example, predicting price movements or anticipating customer lifetime value. Fraud detection can also be greatly improved through machine learning.

Cases

Case Danske Bank

Danske Bank has recognized early on that the banking industry is shifting to a more technology-driven world. And in order to survive and thrive, the company must adapt. The company realized that their customers' behaviour has been changing, demanding for faster response and more tailored approach as a result of increasing technological advances available at their fingertips. To stay on top of the game, Danske realized the need to address the changing customer needs and wants.

With individuals becoming more active online and creating digital footprints, a wealth of data is generated consequently. Nadeem Gulzar, Senior Development Manager at Danske Bank from [2015 - 2017], said in an interview, these big data sets are like the new oil or gold (Hortonworks, 2017). In the last few years, Danske has used data analytics to provide a more personal approach to enhance customer service.

A few years ago, the company executed a strategy using advanced data analytics to find ways to address business challenges. This strategy entailed putting together a team of data scientist to work with business managers. An example given by Büchmann-Slorup, Head of Sales Development and Analytics at Danske Bank, in the Sternberg article (Sternberg, 2017), was a study done for a certain lending market. The use advance data analytics gave them insights into the type of customers that are most likely to avail of this facility by focusing on precise behaviour parameters. This was a more accurate understanding about their customers compared to classifying them based on a more generic category like income, age, etc. (Sternberg, 2017).

Detection of digital fraud is another solution that Danske addressed using machine learning. Figure 4 shows the challenges encountered by the bank using the traditional approach of an investigator looking at customer data. By deciding to work with Think Big Analytics, a Teradata company, Danske utilized a data driven methodology that works with machine learning and deep learning. The goal was to increase fraud detection rate and decrease false positive results. The result with using machine learning reaped a decrease in false positives by thirty five percent (35%). And by adding deep learning, accuracy of fraud detection rate increase by approximately fifty percent (50%) (Groenfeldt, 2017).

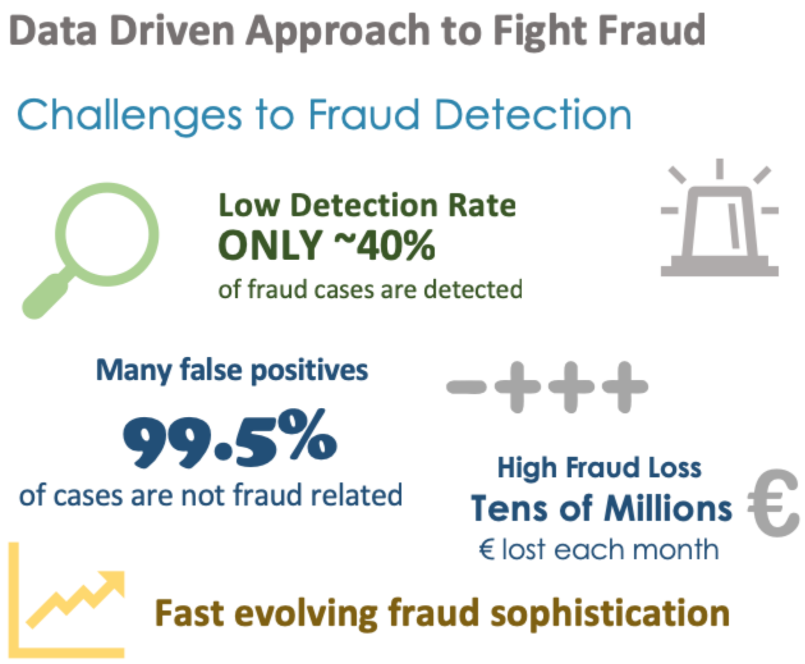

Figure 4: Presentation Slide by Danske Bank and Teradata. Adapted from “Fighting financial fraud at Danske Bank with artificial intelligence,” by R. Bodkin, 2017, Slideshare. Retrieved from https://www.slideshare.net/RonBodkin/fighting-financial-fraud-at-danske-bank-with-artificial-intelligence

In an article in CMO by O’Brien (2018), Danske was heralded as mastering the use of data analytics. They are a success story of assimilating the work of engineers and data scientist with those of the business people. Their Head of Global Analytics has advocated the use of machine learning and even artificial intelligence to help give the customers a whole new experience. “We simply use it [data analytics] to make ourselves, and our products, more relevant for the customer. So, trying to predict the need from the customer and then be there as soon as the customer needs us - that’s the key for us,” Nadeem Gulzar, Senior Development Manager at Danske Bank in 2015, said.

However, even with the success experienced by the bank with these technological tools, they caution the users to be more perceptive and pragmatic with its applications. These technologies are not akin to a magic wand that can fix everything. Sometimes, the basic information that we have is all that is needed to address a simple problem.

Case Bridgewater

Many hedge funds already use algorithmic trading today. An algorithmic trading strategy has three components: entry, exit, and position sizing (Williams, 2017). Data-driven and quantitative funds have experienced tremendous growth. Algorithmic trading utilizes intelligent predictive analytics to arbitrage the advantages of big data. No matter what strategy is used, transaction costs will be reduced for the investor (Oxford Algorithmic Trading Programme, 2010). Renaissance Technologies is a known example, functioning fully with only algorithmic funds. However, in this case, we will focus on Bridgewater, a hedge fund that is not only focusing on automating its trading but its management as well.

Bridgewater is the world’s largest and most profitable hedge fund. Their main fund utilizes algorithms to trade stocks, bonds, currencies and other assets. The fund tracks factors such as interest rates and retail sales, which creates its algorithms. The fund has anticipated many economic up and downturns in the past, including the financial crisis of 2007-2008 a year earlier (Copeland and Hope, 2016). The team in charge of the development of the algorithmic software is headed by David Ferrucci, who was also the lead developer for IBM's Watson, the computer system capable of answering natural-language questions.

Now the fund aims to not only automate trading, but also to develop a system that will automate management decision-making for the whole fund. The company is highly data-driven. Meetings are recorded, and employees have to rank each other throughout the day. These ratings then show each employees' strengths and weaknesses. They can also set goals and track how efficiently they are being achieved (Solon, 2016). The algorithms are now being developed to make decisions such as finding the new staff and ranking opposing arguments in the case of disagreements. The new AI system is called the "Book of the Future". This kind of automation would remove the human emotional aspect from decision-making (Solon, 2016). The algorithms would also plan and organize the employee's entire day, down to whether they should take a specific phone call or not.

The AI technology used by financial services businesses is evolving every day. Even though quantitative trading is becoming more and more sophisticated, the development is moving towards other areas too, such as the automation of management. This can eventually lead to an entire organisation being supported by AI decision-making.

Criticisms

In addition to all potential benefit gained from the use of data science and big data for financial institutions, drawbacks must also be considered. As banks and other institutions are developing new techniques and technologies to handle and analyse the data, they must consider the security aspects at each step. Both security issues as well as the privacy of data are key concerns. Cloud-computing has become popular due the benefits it offers – broad network access, on-demand service, and resource pooling. This presents many issues for privacy and security, as many systems are still quite young, and security protocols are not fully established. Security monitoring must be still developed. As we are still at the very beginning phases of working with big data, financial institutions must develop this ecosystem with care to avoid any possible security and privacy breaches (Sinanc, Sagiroglu & Terzi, 2015).

Regulation is also a concern for the future of financial institutions utilizing these new technologies. The question of consumer protection is highly relevant. By nature, many of these institutions are very global and serve various cross-border transactions. The legal aspect of these transactions should be considered – which country's legal system should be followed? Financial regulation can be fragmented and specific to certain countries and regions, so there is great risk for both consumers and financial institutions if they are not aware of all regulations.

Additionally, when the use of Big Data analysis increases, it is important to understand that not everyone will be evenly benefitted. Information technology helps investors to do less risky investment decisions but on the other side are the firms that need capital financing. One macroeconomic question today is why big firms seem to replace smaller ones and a possible explanation is the cost of capital. Large and incumbent enterprises have longer track records and thus more big data available to reduce the asymmetric information problem in comparison to new and smaller firms. More available information reduces the investors' risks which lead to cheaper access to external capital markets and reduces the cost of capital of big firms - this again accelerates their growth while a small firm's growth is initially idle (Begenau, Farboodi & Veldkamp, 2018). When doing investment decisions based on analyses extracted from Big Data, it is vital to be aware that there might be more profitable investment options than the suggested ones and higher risks can depend on the lack of available data rather than its inferiority.

Conclusions and Future Outlook

The disruption of fintech to the financial industry has been relatively low to date. This is partly driven due to the fact that fintech companies provide complementary services to the traditional financial sector. Although alternative lending services are provided, the users will still need bank accounts to make use of these services. There is also a need for the safety of households and firms accounts which new fintech services does not provide, traditional banks are still the best source for this (Lorente & Schmukler, 2018). However, more and more small companies are able to use big data and data science to provide services that are outside the scope of regulatory bodies, or that do not require licensing. New data analysis methods are being constantly developed, many of them being created by the small players instead of established banks (Dapp, 2014).

Customer expectations from financial services are likely to rise in the future as the various technologies develop. As more and more users gain access to fintech, and thus to financials services and credit, they will again fuel the need for the emergence of new and developing financial technologies that are faster, address customer needs better, are more convenient, and provide new solutions. Banks will likely find it difficult to operate completely in the traditional manners and business models (Lorente & Schmukler, 2018; Dapp, 2014).

Overall, the use of Big Data in accounting seems to drag behind, there is yet no academic empirical work apart from anecdotal evidence to turn to. At this stage, the focus is on determining what possible consequences Big Data will have on management accounting, financial accounting and auditing. In general, using Big Data in accounting context will be a disruptive force since it will require significant change both in skill set and the way traditional accountants work. Traditional tasks such as data registration will become less important and management accounting techniques will become obsolete. Big Data technology will soon provide alternatives for asset valuation, cost analysis, forecasting and budgeting. It will also impact the role of bookkeeping as the creator of business knowledge for managers to base their decisions on since they will have access to endless external information provided from Big Data (Rikhardsson, 2018). The principles for accounting has been generally unchangeable for ages, which possible can be an explanation why adaption of innovations is slow and why it get resistance. But Big Data techniques are best used to complement and not replace human experts (Gepp et al. 2018). Furthermore, accounting standards as IFRS and GAAP will probably need to be updated accordingly to technical advancements, if it will change i.e. the way analysis, reports and valuations are assembled and done.

At this stage, finance seems well ahead of accounting what comes to utilizing data science applications and big data to bring added value to the decision-making process. However, these concepts still seek establishment among the field – which may be hard to attain during a phase of new innovations constantly lurking around the corner. Still, the ongoing change should be embraced and not rejected, if it brings new possibilities not achievable otherwise. Resistance and issues can be won by widening the knowledge within the emerging technology and by cooperation between involved fields.

References

Anderson, N. (2010). IBM Business Analytics – Banking Solutions for Financial Performance Management. Available at: https://www.ibm.com/developerworks/community/blogs/5b3fcdff-c109-435b-bb73-e6a34fa4b66f/resource/IBM%20Business%20Analytics/IBM_BudgetingPlanning_Cognos_NickAnderson_OndrejBothe_01dec.pdf?lang=en [Accessed 5 March 2019]

Begenau, J., Farboodi, M. & Veldkamp, L. (2018). Big data in finance and the growth of large firms. *Journal of Monetary Economics*, 97(C), 71-87.

Chen, J. (2019). Neural Network. *Investopedia*. Retrieved from https://www.investopedia.com/terms/n/neuralnetwork.asp

Clifton, C. (2017). Data mining. Encyclopædia Britannica. Retrieved from https://www.britannica.com/technology/data-mining

Cockcroft, S. & Russell, M. (2018). Big Data Opportunities for Accounting and Finance Practice and Research. Australian Accounting Review, 28(3), 323-333. doi:10.1111/auar.12218

Copeland, R. & Hope, B. (2016). Bridgewater builds algorithmic model from employees' brains. Financial News. Retrieved from https://www.fnlondon.com/articles/bridgewater-builds-algorithmic-model-from-employees-brains-20161223

Dapp, F. (2014). Fintech – The digital (r)evolution in the financial sector. *Deutsche Bank Research*. Retrieved from http://www.dbresearch.com/PROD/RPS_EN-PROD/PROD0000000000451941/Fintech_%E2%80%93_The_digital_%28r%29evolution_in_the_financia.pdf

Das, S. R. (2018). The Future of FinTech. California: Santa Clara University∗. Retrieved from https://pbs.twimg.com/media/CYOHdfgW8AERjdO.png

Decide Soluciones. (2017). Let's talk about Advanced Analytics: A brief look at Artificial Intelligence. *Becoming Human: Artificial Intelligence Magazine.* Retrieved from https://becominghuman.ai/lets-talk-about-advanced-analytics-a-brief-look-at-artificial-intelligence-bf1c7a7d3f96

DeMuro, J. (2018). What is neural network? Techradar.pro. Retrieved from https://www.techradar.com/news/what-is-a-neural-network

EcourseReview. (2017). Cloud Computing Pros and Cons for Business and IT Professionals. Retrieved from https://ecoursereview.com/cloud-computing-pros-and-cons-for-business-it/

Gandomi, A. & Haider, M. (2015). Beyond the hype: Big data concepts, methods and analytics. International Journal of Information Management, 35(2), 137-144.

Gepp, A., Linnenluecke, M., O'Neill, T. & Smith, T. (2018). Big Data in Accounting and Finance: A Review of Influential Publications and a Research Agenda. 4th Forensic Accounting Teaching and Research Symposium.

Goh, C. (2017). Are You Ready? Data Analytics Is Reshaping the Work of Accountants. *CFO Innovations*. Retrieved from https://www.cfoinnovation.com/management/are-you-ready-data-analytics-reshaping-work-accountants.

Green, K. C. (2015). Simple versus complex forecasting: the evidence. *Journal of Business Research*, 68(8), 1678-1685.

Groenfeldt, T. (2017). Danske Bank Uses Tech To Prevent Digital Fraud. *Forbes*. Retrieved from <https://www.forbes.com/sites/tomgroenfeldt/2017/10/30/danske-bank-uses-tech-to-prevent-digital-fraud/#2040f7193ae5>

Hassani, H. & Silva, E. S. (2015). Forecasting with Big Data: A Review. *Annals of Science*, 2(1), 5-19.

Hayaschi, C. (1998). What is Data Science? Fundamental Concepts and a Heuristic Example. *Data Science, Classification, and Related Methods.* Studies in Classification, Data Analysis, and Knowledge Organization. Tokyo: Springer

Hiltbrand, T. (2018). 5 Advanced Analytics Algorithms for Your Big Data Initiatives. *TDWI*.Retrieved from https://tdwi.org/articles/2018/07/02/adv-all-5-algorithms-for-big-data.aspx

Hortonworks. (2017). Danske Bank's Mission Critical Data & Vision of Being 100% Data-Driven. *Youtube*. Retrieved 3 March 2019, from https://www.youtube.com/watch?time_continue=3&v=U3zGjDDS9E4>

Hou, T. (n.d.). IaaS vs PaaS vs SaaS Enter the Ecommerce Vernacular: What You Need to Know, Examples & More. *Big Commerce*. Retrieved from https://www.bigcommerce.com/blog/saas-vs-paas-vs-iaas/#the-three-types-of-cloud-computing-service-models-explained

IBM. (2019). IBM Cloud. Retrieved from https://www.ibm.com/cloud/learn/what-is-cloud-computing

Jain, N. (2018). Introduction to Cloud Computing. *Whizlabs.* Retrieved from Whizlabs: https://www.whizlabs.com/blog/cloud-computing/

King, T. (2018). Structured Data vs. Unstructured Data; What's the Difference? *Solutions Review*. Retrieved from https://solutionsreview.com/data-management/key-differences-between-structured-and-unstructured-data/

Koza, J., Bennett, F., Andre, D. & Keane, M. (1996). Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming. *Artificial Intelligence in Design '96*. Springer: Dordrecht. pp. 151–170.

O'Brien, J. (2018). How Danske Bank mastered data analytics. CMO. Retrieved from https://www.cmo.com.au/article/641846/how-danske-bank-embarked-journey-master-data-analytics/>

Oxford Algorithmic Trading Programme. (2010). Algorithmic Trading Hedge Funds: Past, Present, and Future. Retrieved from https://www.getsmarter.com/presentations/uk/oxford-said/algorithmic-trading-hedge-funds-past-present-and-future

Oxford English Dictionary. (2008). Third Edition. Big Data. Retrieved 29 January 2019 from <http://www.oed.com.proxy.shh.fi/view/Entry/18833redirectedFrom=big+data#eid301162177>

Oxford Living Dictionaries. (n.d.). Artificial Intelligence. Retrieved from https://en.oxforddictionaries.com/definition/artificial_intelligence

Peric, A., Polic, N. & Kozarevic E. (2016). Application of Data Science in Financial and Other Industries. Proceedings of the International Scientific Symposium Economy Of Eastern Croatia – Vision And Growth. Osijek: Sveuĉilište Josipa Jurja Strossmayera u Osijeku.

Reca, M. (2018). The 3+ Vs of Big Data: Volume,Velocity, Variety, and a whole lot more. *FlyData*. Retrieved from: https://www.flydata.com/blog/3-vs-of-big-data/

Rikhardsson, P. (2018). Business intelligence & analytics in management accounting research: Status and future focus. *International Journal of Accounting Information Systems*, 29(C), 37-58.

Schmukler, L. & Lorente, J. J. C. (2018). The Fintech Revolution: A Global Threat to Banking? World Bank: Research & Policy Briefs Paper No. 125038

Sinanc, D., Sagiroglu, S. & Terzi, R. (2015). A survey on security and privacy issues in big data. Conference: 2015 10th International Conference for Internet Technology and Secured Transactions (ICITST).

Solon, O. (2016). World's largest hedge fund to replace managers with artificial intelligence. *The Guardian*. Retrieved from: https://www.theguardian.com/technology/2016/dec/22/bridgewater-associates-ai-artificial-intelligence-management

Statistics Solutions. (2013). What is Linear Regression? *Complete Dissertation*. Retrieved from https://www.statisticssolutions.com/what-is-linear-regression/

Sternberg, R. (2017). Nordic financial giant banks its future on digital customer insight. *SAS*. Retrieved from https://www.sas.com/fi_fi/insights/articles/analytics/local/nordic-financial-giant-banks-its-future-on-digital-customer-insight.html

Taylor, C. (2018). Structured vs. Unstructured Data. *Datamation*. Retrieved from https://www.datamation.com/big-data/structured-vs-unstructured-data.html

Walker, T. (2018). Banks make better decisions with smart, unbiased banking analytics. *IBM*. Retrieved from https://www.ibm.com/blogs/business-analytics/banking-analytics/

Williams, G. (2017). Trading places: the rise of the DIY hedge fund. *Wired*. Retrieved from https://www.wired.co.uk/article/trading-places-the-rise-of-the-diy-hedge-fund

Zheng, X., Zhui, M., Li, Q. Chen, C. & Tan, Y. (2018). FinBrain: When Finance Meets AI 2.0*. *Frontiers of Information Technology & Electronic Engineering*.