JOURNAL OF CRITICAL REVIEWS

ISSN- 2394-5125VOL 7, ISSUE 15, 2020

Applications of Analytics in BFSI industry and the Challenges/Limitations in their implementation

Anurag Kochupurakal[1]

[1] Symbiosis Centre for Information Technology, Symbiosis International University, Pune, India.

ABSTRACT: The idea behind starting this dissertation would be to identify the areas in the financial sector where data analytics and in particular predictive analytics is being used to improve the businesses in terms of revenue, operational efficiency, customer-base, customer satisfaction, tailored business models to meet specific requirements of individual customers, etc. while exploring further possibilities and also understanding the crucial challenges and limitations faced by the financial organisations in implementing these disruptive technologies. The methodology would require an extensive literature review ofwith a qualitative research approach that includes research papers, articles and blogs published from authentic sources.

Introduction

In the current scenario of the finance world, the changes that are brought about by the advancements in the technology has become a game changer in almost every domain and financial industry is not an exception. Several financial organisationsfacea lot of challenges in order to build competitiveness for winning over the competitors. Inclusion of Data Analytics in decision-making has been one of the most crucial competence lately for banks, insurance companies, trading companies and several other organisations providing financial services, that not only results in better performance and predominant customer service but also gives much neededuplift to innovation and thereby giving them strategic and competitive edge[1]. On top of that, the ability to analyse the data and provide crucial feedback that could have been overlooked or even missed due to human error on decision making has been addressed by use of Data Analytics. The expansion of financial industry with rapid growth can be accreditedto the continuous transformation in the expectations of consumers, that are rising from the upcoming technologies and availability of various products and services[2]. This disruption caused by the technology in financial sector has given rise to a new class of organisations which call themselves the “Fin-Tech” organisations.

With all this being said, one must also understand that generating business value using data-analytics requires these financial institutions to successfullydesign as well as deploy these analytical solutions. The design part ideally involves the procedure of assembling, developing and customizing, suitable analytical solutions, while deployment would require the adaptation of the said solutions to particular environments. The environments usually involve different aspects for integratingand hostinga solution and are very likely to evolve with the ever-changing government guidelines related to finance, thus solutions need to be continuously modified. These processes require a remarkable effort that could be different from each context to another, where it would also require distinct capabilities and thorough understanding of the analytics domain.

Literature Review

It is the need of the hour to be able to utilise the available and upcoming technologies, in such a way as to transform the vast amount of data available in hand to more valuable insights helpful to make critical decisions related to the business. It is believed that this generation of data processing technologies will be greatly dominated by the Big data, Internet of Things(IOT) along with Prescriptive and Predictive Analytics. Predictive Analytics and Prescriptive Analytics together are considered to be the frontier in helping overcoming this gap[3,10]. Business performance in any field can be improved significantly by using data driven decision making, especially when there is a need to make automatic decisions on large scale[10,19]. Organisations around the world are functioning in a complex as well as competitive environment and thus need to consider investment in analytics as a strategic move to improve their competence. But this should be done keeping in mind how inclusion of analytics if going to affect their existing technologies, businesses, applications, websites and above all how its going to affect their customer and how the customers are going to react to the change, how the existing employees can be trained to cater the needs of additional technologies, recruitment of experts, etc.,[8].

The voluminous growth of data everyday due to various activities being performed brought about a question regarding the storage of such gigantic volumes of data that was very well answered by the advent of Big data technology. This created an opportunity to help fight the fraudulent activities happening in the financial industries by helping the investigators sift through volumes of data collected by integrating various heterogeneous sources[8,9]. According to [5] deep learning techniques could be used to create a subsystem for Robo-Advisors that helps customers build a custom portfolio based on their previous transactional patterns in investor profile, stock prices and other related alternative data. As per [14] a conventional data warehouse usually has separate buyer-side, seller-side transactions to store the respective data due to difference in business purposes but by using aggregation financial model architecture we can have cumulative information stored by defining relationship between different entities and their behavioural aspects.

Nifty and Sensex are two major indicators for predicting stock market conditions in India. By using the technique of sentiment analysis, live server data of Nifty and Sensex can be fetched which can be used for Stock Market prediction which can help investors make a decision as to which stocks to invest on[16]. Stock market portfolio Recommender Systems can be built using Association rule mining by applying soft computing techniques like ARM fuzzy classification[21]. Different classifiers were compared for varying degrees of imbalance dataset for credit scoring and it was found that Gradient boosting as well as Random Forest classifiers yield the best performances for extreme degrees of data imbalances[22]. According to [23] use of Support Vector Machines(SVM) in bankruptcy prediction model gives better generalisation as well as higher level of accuracy than back-propagation neural network model. Thus in the current paper we try to understand different kinds of application of analytics in the finance world.

Traditional Business Intelligence and Predictive Analytics

Let us first understanding the basic difference between Traditional business intelligence methodologies and use of Predictive analytics in the business. Traditional BI has been used bythe industry since decades for reporting purposes, descriptive models, visualisations for better decision making etc.[10].Traditional BI is more descriptive in nature and helps understand the business problems and other opportunities. The volume of data gathered by the financial is huge and is also rapidly rising with each passing day. But conventional BI cannot meet the demand for faster processing and bigger data storage units, that's where the Big Data technologies have come in demand. The difficulty of voluminous-data has been handled pretty well by use of Big Data in the recent years[3]. Predictive Analytics is a collection of technologies that helps us identify patterns and relationships hidden inside the giant volumes of data being stored and collected. The united force of Predictive analytics along with Big Data technologiesand several data-modelling techniques brings about a wealth of prospects for the BFSI industry to create predictions in various aspects of the industry.

Predictive-Modelling Algorithms

Predictive-modelling is the action of using familiar results to process, validate and build a model which can be used to predict future results. It is a tool used in predictive analytics, a mining technique that strives to answer the question "what might possibly happen in the future?" Organisations use predictive modelling to forecast events, customer behaviour as well as risks related to finance, market and economic conditions. The following are the most commonly used predictions algorithms in the industry.

i. **Classification**: Classification is a technique used to predict the value of a categorical variable by creating a model based on group of training data in order to classify the new inputs into a set of category defined by the classification prediction model[22]. The different types of classification algorithms used are as Knn, Logistic Regression, Naïve Bayes, Decision Tree, Linear Discriminant Analysis, QDA, Support Vector Machines, Artificial Neural Networks, etc. to name a few.

ii. **Regression**: Regression is a technique used to predict the value of a numerical target variable by creating a model based one or more than one numerical as well as categorical variables generally known as predictors. Usually regression techniques are usedto predict the outputs that are continuous in nature. It means that the outcome can be flexibly determined by changing the conditions (predictor variables) rather than being confined to only a set of outcomes[12]. The common regression techniques used are Multiple Linear Regression, Decision Tree, Knn, Support Vector Machines, Artifical Neural Networks, etc. to name a few.

iii. **Clustering**:Clustering is a process of categorizing the given data-set into groups in such a way that the items belonging to one group have similar characteristics while also being dissimilar to members of the other categories at the same time[15]. The most common methods of Clustering are k-means, agglomerative and divisive.

iv. **Association Rule mining**: Association Rule mining is a rulebasedtechnique of machine learning usedto find out insightful relationships among variables in large datasets containing good number of variables[20]. An association rule is defined as an expression $X \rightarrow Y$, where X and Y are group of items. The cognitive meaning of such a rule is that transactions of the database that contain the item X also tend to contain the item Y. Market Basket analysis is where the Association Rule mining generally finds its applications.

Applications of Predictive-Analytics in the Finance Industry

Investments banks and financial organisations all over the world are dealing with the problem of growing data processing needs. These issues not only originate from an expanding variety of data sources and increasing regulatory requirements, but also from on-going demands inreduction of cost without having to compromise on system flexibility and scalability. In this regard, the capacity to apply propitious state-of-the-art big data technologies to generate the highest value from the large amounts of data being generated is giving rise to a lot of buzz in the financial sector. In this paper, we have tried to identify the PA algorithms and the areas where PA has been used to achieve better performance. Use of predictive analytics in financial institutions can be categorised into revenue growth, prevention of risk and operational efficiency.

Predictive-Analytics is used almost in every aspect of the financial industry such as Stock Prediction, Credit Defaulter Prediction, Bankruptcy prediction, Social Media Prediction, Cyber Crime Prediction, Forex rate prediction, Macro Economic prediction, Risk predictions, Customer Buying Habit Prediction, Credit scoring etc. The most widely used area of Predictive-Analytics in the finance industry is the CRM predictions. PA techniques that can be applied to the CRM strategies are as follows, Customer Targeting and Customer Attraction, Customer Segmentation, Customer Retention and Customer Development, etc. Predictive Analytics is also used in building predictive models for forecasting stock prices, detection of fraud, credit card fraud by using the analytical framework to action thehugeamount of data and implementation ofa variety of machine learning algorithms for detection of fraud on real time basis.

Table 1: Applications and Algorithms used

Application	**Algorithms usually used**
Stock Prediction	Sentiment Analysis along with Fuzzy logic module, Support Vector Machine, Multiple Linear Regression, Decision Boosted Tree
Bankruptcy Prediction	Support Vector Machines, Artificial Neural Networks (also known as black box model), Recommender Systems
Cybercrime Prediction	K-Means classifier & Influenced Association Classification with J48 Prediction Tree
Forex rate Prediction	Quantile Regression Random Forest Big Data Analytics Mechanism that uses systematic collection, organisation and analyses of publicly available information on social media
Macro Economic Prediction	Standardised GLM model that uses Logistic Regression.
Customer buying habit Prediction	K- Means clustering and Logistic Regression with Binary Classification model.
Credit Scoring	Random Forest, Logistic Regression and LASSO
Social media Prediction	Natural Language Processing, Semantic Analytics, Text Analytics, Wordclouds.
Credit Defaulter Prediction	Decision Tree Classifier
CRM Prediction	Logistic Regression, Random Forest, Decision Tree, Neural Networks, Monte Carlo Simulations, Support Vector Machine, Markov Chains, Segmentation, Quantile Regression.
Fraud Detection	Auto Associative Neural Network, Bootstrapped Optimistic Algorithm for Tree Construction, Bidirectional Artificial Neural Networks

Challenges and Limitations of implementing Predictive Analytics

A multi-disciplinary method was followed to probe the obstaclesin order togatherthe lessonslearnt from different areas, especially financial institutions. Firstly in the finance industry, compliance controls and the continuously changing regulations make the process of documentation , validations and evaluations, a time consuming as well as adaunting task[8]. Then again, Natural Language Semantics (NLS) techniquesand Text Analytics could be mergedonto software systems for checking regularity. Consequently, in Wealth Management, contemporary behavioural analyticaltechniques may help in complying to terms of customers product suitability. However, the recognition of customers receiving advise without being engaged in a contract helps lessenorganisationalaccountability for reporting and clarity in the advice provided.

Organisational challenges: A Predictive Analytics studylead by Ventana Research for the benefit of IBM identified several hindrances the institutions have faced in their use of predictive analytics.[8] The research also identified the biggest business hurdles to the use of predictive analytics and its productive deployment.

1. Problem of accessing source data through the tools available.

2. Shortage of resources such as budget and skills.

3. Problem of merging analytics into an organization's database architecture.

4. Absence of understanding and awareness of application of predictive analytics to business relatedissues.

5. Dearth of in-house experts to implement the obtained results.

6. Problem in using the results obtained.

7. Insufficiency ofhistoric patterns.

8. Laxity of results obtained.

9. The data is exorbitant to quantify.

Infrastructure: To performstorage, data ingestion, refining and investigation, where every step faces distinct challenges and needsdefinite analytical capabilities. For example, in the analysis of financial documentation, as not even one of the fund-related documents come in a convenient format which is machine-readable, data processing would require techniques such as semantics techniques, natural language processing and machine learning.

Customization: Solutions enabled by analytics would require a considerable amount of customization in order to be deployed as a potential integration and thecompatibility issues will have to be considered. For instance, the methods required to process the documents written in natural language would vary from one document to another in distinct domains, with regardto the difficulty of the final goal in view. Also, there is huge distinction in the format of data that is extracted to that of the data being visualised. Therefore, visualization, data ingestion and storage systems also need to be sufficiently customized to accommodateas per the needs of the entire infrastructure.

Continuous Adaptation: Adaptation is usually defined as an proceeding effort in order to enhance the capabilities that are enabled with analytics such that theycomplyto thechangingbusinessandtechnological requirements in the light of their effectiveness and efficiency. For instance, from 2019 the UCITS KIID was replaced by PRIIPS KID in which the length, the contentas well as the layout of the documentation that was required by law differed, so that regularityin checking the models had to be continuouslyreintegrated.

Multiplicity of Solutions: The solutions that are analytically enabled will have to be deployed depending on the accessibility and shareability,in order to manage and configure the infrastructure. For instance, in Wealth Management, with regards to the General Data Protection Regulation as well as the vulnerability of finance related data, in-house solutions need to be deployed to process and storehuge amounts of transactional data. Edge computing has proven to be a systematic solution to perform the processing of dataat the source of the data, in order to reduce the security issues and the communication bandwidth. Hybrid cloud edge could be another solution that can help simplify the management of locations as well as applications in a safe,synchronous and attainable way.

Cost of training machines: Moore's law, states that the count of the number of components that can be etched onto a microprocessor chip of a given sizeand the quantity of computational power that is available at a given cost becomes twice as much for every two years. Due to the competition and the complexity, the costs are rising sharply. A paper that was published in 2019 by a group of researchers at the University of Massachusetts Amherst

ISSN- 2394-5125VOL 7, ISSUE 15, 2020

evaluated that training a version of "Transformer", big language model, costs around $3m. Jerome Pesenti, Facebook's head of Artificial Intelligence, says that one round of training for these big models could cost around "millions of dollars" in electricity bills.Facebook, estimated to have made around $18.5bn in profit in the year 2019, can easilymanage to paysuch high bills, unfortunately which is not the case with organisations that make lesser profits or no profits at all like any start-ups. Andreessen Horowitz, an influential American venture-capital firm, has brought to the notice that many smaller AI start-ups depend on firms such as Microsoft and Amazon for their processing power requirements. This results in spending 25% or more of their revenue going to pay those bills for processing power, thus making AI start-ups a less attractive option for investments than the old-style software companies.

Beating False-Positives: One of the key challenges faced by the banking and finance industry with regards to the analytics enabled solutions is the rate of false positives. Banks and other financial institutions are latchedonto spending large amounts of money in order to avoid being a victimof fraudulent transactions.Striking the balance between false positives and scrutiny remains the main challenge of fraud detection using data analytics. It must be ensure that the fraud detection platforms are very well calibrated in order to detect fraudulent transactions without blocking the legitimate ones. There could be substantial ripple effects as an impact caused due to the blocking of one genuine transaction.A study reveals that, 2.5 times the amount that is lost to fraud is also lost fighting the fraud. The cost of endorsing a data analytics platform with false positives is very high. Researchers and data analytics enthusiasts are coming up withinnovativeways to reduce false positive rates, of which feature engineering has shown encouraging results.

Summary

Thevast amountsofdataavailable with the organisations isprovidingthemwiththeinformationthat is required tooperatemoresystematically,constructivelyandcautiously,thereforeenablingthemtoovercomesomeof t hedemandingandturbulentobstacles.Over the last five years, the uttercapacities of data has risen exponentially and latest analytics tools have been flourished to turn this surge ofstructured, semi-structured and unstructured data into cognizance.Usage and administration of the Big Data are progressively becoming areas of competitive superiority. Hence, organisations are embracing predictive analytics as one of the competitive advantages to be able to contend in the current market.

Nonetheless, as briefly talked through in this paper, there are technical and organization relateddownsides to predictive analytics. The foremost ones being the amount of up-to-date data it needs and absence of in-house experts to implement the results. This will not be listed as a problem for bigger organizations,but it is not the same case for the smaller ones. The main focus for further research should be to efficiently reduce the cost of deployment, reduce the cost of training machines as well as enable the end user to understand the effectiveness of tailor-made products and create more awareness among them in order to enjoy the fruition of such disruptive technologies.

References

1. Indriasari, E., Soeparno, H., Gaol, F. L., & Matsuo, T. (2019). Application of Predictive Analytics at Financial Institutions: A Systematic Literature Review. 2019 8th International Congress on Advanced Applied Informatics (IIAI-AAI).
2. Deshpande, P. S., Sharma, S. C., &Peddoju, S. K. (2019). Predictive and Prescriptive Analytics in Big-data Era. Security and Data Storage Aspect in Cloud Computing, 71–81.
3. Pathak, A., & Shetty, N. P. (2018). Indian Stock Market Prediction Using Machine Learning and Sentiment Analysis. Computational Intelligence in Data Mining, 595–603.
4. Day, M.-Y., Cheng, T.-K., & Li, J.-G. (2018). AI Robo-Advisor with Big Data Analytics for Financial Services. 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).
5. Brichni, M., Kampas, D., Choory Balaji, A., & Sanz, J. (2018). Some Challenges and Lessons-Learnt From the Practice of Analytics. 2018 IEEE 20th Conference on Business Informatics (CBI).
6. Tsaih, R.-H., Kuo, B.-S., Lin, T.-H., & Hsu, C.-C. (2018). The use of big data analytics to predict the foreign exchange rate based on public media: A machine-learning experiment. IT Professional, 20(2), 34–41.

7. Mohsen, Sharmin, Y.(2018). Opportunities and Challenges for Implementing Predictive Analytics for Competitive Advantage, Vol 9, Issue 2.
8. Makki, S., Haque, R., Taher, Y., Assaghir, Z., Ditzler, G., Hacid, M.-S., &Zeineddine, H. (2017). Fraud Data Analytics Tools and Techniques in Big Data Era. 2017 International Conference on Cloud and Autonomic Computing (ICCAC).
9. Lekha, K. C., &Prakasam, S. (2017). Data mining techniques in detecting and predicting cyber crimes in banking sector. 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS).
10. Ouahilal, M., Mohajir, M. E., Chahhou, M., &Mohajir, B. E. E. (2016). A comparative study of predictive algorithms for business analytics and decision support systems: Finance as a case study. 2016 International Conference on Information Technology for Organizations Development (IT4OD)
11. Abdou, H. A., Tsafack, M. D. D., Ntim, C. G., & Baker, R. D. (2016). Predicting creditworthiness in retail banking with limited scoring data. Knowledge-Based Systems, 103, 89–103.
12. Sudhamathy, G., &Venkateswaran, C. J. (2016). Analytics using R for predicting credit defaulters. 2016 IEEE International Conference on Advances in Computer Applications (ICACA).
13. Kavitha S, Varuna S, & Ramya R. (2016). A comparative analysis on linear regression and support vector regression. 2016 Online International Conference on Green Engineering and Technologies (IC-GET).
14. Nayak, A., Manohara, M.M. Pai,& Radhika, M.Pai. (2016). Prediction Models for Indian Stock Market, Twelfth International Multi-Conference on Information Processing - 2016, (IMCIP).
15. Gollapudi, S. (2015). Aggregating financial services data without assumptions: A semantic data reference architecture. Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015).
16. Xu, D., & Tian, Y. (2015). A Comprehensive Survey of Clustering Algorithms. Annals of Data Science, 2(2), 165–193.
17. Bhardwaj, A., Narayan, Y., Vanraj, Pawan, & Dutta, M. (2015). Sentiment Analysis for Indian Stock Market Prediction Using Sensex and Nifty. Procedia Computer Science, 70, 85–91.
18. Hafiz, A., Lukumon, O., Muhammad, B., Olugbenga, A., Hakeem, O., &Saheed, A. (2015). Bankruptcy Prediction of Construction Businesses: Towards a Big Data Analytics Approach. 2015 IEEE First International Conference on Big Data Computing Service and Applications.
19. Munar, A., Chiner, E., & Sales, I. (2014). A Big Data Financial Information Management Architecture for Global Banking. 2014 International Conference on Future Internet of Things and Cloud.
20. Provost, F., & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. Big Data, 1(1), 51–59.
21. Paranjape-Voditel, P., & Deshpande, U. (2013). A stock market portfolio recommender system based on association rule mining. Applied Soft Computing, 13(2), 1055–1063.
22. Brown, I., &Mues, C. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. Expert Systems with Applications, 39(3), 3446–3453.
23. Shin, K.-S., Lee, T. S., & Kim, H. (2005). An application of support vector machines in bankruptcy prediction model. Expert Systems with Applications, 28(1), 127–135.