Role of Big Data Analysis in Predicting Financial Market

[1]P. Deepa,
Ph. D Research Scholar,
Department of Computer Science,
Dr. SNS Rajalakshmi College of Arts and Science (Autonomous), Coimbatore-641049.

[2] Dr. B. Murugesakumar,
Head, Department of Computer Science,
Dr. SNS Rajalakshmi College of Arts and Science (Autonomous), Coimbatore-641049.

Abstract:

In the ever-changing world of financial markets, the infusion of big data analytics has become a game-changer, reshaping how we traditionally analyze and forecast market trends. Big data is the term used to characterize the vast amounts of data that are gathered, saved, and examined in order to obtain knowledge and improve decision-making. A variety of sources, including social media, online analytics, user behaviour, and machine-generated data, are frequently used to gather big data. In the field of finance, big data refers to vast, varied, complex (both organized and unstructured) data sets that can be leveraged to address persistent business problems faced by global banking and financial services organizations. The phrase is increasingly understood to be a business necessity rather than only being used in the context of technology. Financial services businesses are utilizing it more and more to revolutionize their organizations, workflows, and sectors as a whole. This study offers a comprehensive examination of big data finance, including how it functions and whether predictive trends are present.
The Association Rule Mining Algorithm, Linear Regression Algorithms, Logistic Regression Algorithms, and Support Vector Machine (SVM) are the approaches used to predict the big data in this article.

Keywords: Big data, Dataset, Financial services, Linear Regression Algorithms, Logistic Regression Algorithms, Machine learning, SVM.

Introduction:
Despite a protracted conversion process involving both behavioural and technological changes, financial institutions and financial services are not new to the digital realm. Big data in finance has played a significant role in recent years' significant technological advancements that have enabled the sector to provide straightforward, safe, customised, data-driven solutions. Thus, big data analytics has been successful in changing not just the operations of individual businesses but also the entire banking industry. The banking sector has seen a significant transformation as a result of data science; among other things, machine learning algorithms are being utilised to forecast stock values and enhance risk assessment while making loan decisions.

Five V's of Big data
Objectives of the study:

Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 6 (2023)

- To investigate the (large-scale data) big data.
- To research big data's predictive power in finance.
- To research big data challenges in Finance industry.

Literature Review:

1. Fergusson and Seow (2011) and Gaunt (2014) an exploratory search was carried out using the term "big data" and the ISSN numbers of all the journals for the field of research (FOR) codes relating to IS, Accounting and Finance within the Australian Business Deans Council (ABDC) list. Whilst this list is aimed at Australian academic audiences, it is comparable to other business centred journal lists worldwide such as the chartered association of business schools ABS list. Four main business reference sources were used: Scopus, ABI/INFORM, Web of Science and EBSCOhost. The search was limited to peer-reviewed and scholarly journals, and limited to papers published in the ten years between 2007 and 2016, this resulted in 3082 results. Then a filter was applied to only retrieve journals on the ABDC list, resulting in 529 journal articles.
2. Misra et al. [21] discussed linear classification with support vector machine (SVM) and artificial neural networks to accurately forecast the market's daily movements. An artificial neural network (ANN) and SVM are commonly applied in handling heterogeneous data and finance data. Rechenthin, Street, and Srinivasan proposed a system to predict the U.S. oil market (USO) stock chatter, which can examine the USO index to predict the true value of the stocks. Yahoo! Finance has information of all opening and closing prices of the USO. Machine learning Electronic copy available at: https://ssrn.com/abstract=3827106 2572 CMC, 2021, vol.67, no.2 approaches to Spark, Hadoop, and regression methods have been applied in the literature to predict the movements of stock prices. Spark was used to predict real-time data because Hadoop with MapReduce cannot work correctly in a real-time system
3. Seif et al proposed a model to forecast markets using news and other external factors. Machine algorithms and other neural network techniques were applied to collect data to predict accurate values. The naive Bayes technique, SVM, and text mining were used to finalize the dataset to predict the stock market index . This system collects positive and negative views, and historical stocks' opening and closing price information to increase the forecasting of stock market values. The proposed system predicts heterogeneous data such as news, social media, and historical movements of share prices. The researchers stated that results of the k-nearest neighbors (KNN) algorithm used to examine the relationship between social media news and the movement of the stock prices are significant .
4.Chordia, Green, and Kottimukkalur (2018) address the issue that there is a huge developing market for obtaining news feeds split seconds before others. Does getting such macro news seconds ahead of its release predict returns? Using intradaily data, they find that the returns to trading on news around macroeconomic announcements (such as inflation, GDP, unemployment, etc.) are fairly small amongst various combinations of entry/exit times, attain a maximum of 8bp for entering within 0.1 s and exiting by 5 s. Such modest levels of profit imply that the returns to attaining information seconds before others are fairly modest, indicating a good level of market efficiency at the intraday horizon. However, as technology improves and information becomes available to trade upon microseconds sooner, it remains to be seen whether the profits continue to remain at low levels or increase.
5. Dang et al. [19] proposed a two-stream GRU model that incorporated the financial news sentiments with stock features as inputs to forecast S&P 500 index trends and prices. Results showed that the two-stream GRU model outperformed other models, including both the LSTM model and the original GRU model. The authors also pointed out that the two-stream GRU model requires long time for training and huge computational resources because of the complexity of the enlarged GRU model.

How Finance Is Being Revolutionised by Big Data

In the past, humans handled the number-crunching and made decisions based on conclusions from patterns and estimated risks. However, computers have recently supplanted such capacity. Consequently, the finance sector presents a highly promising industry with immense potential for big data technology.

1.**Real-time stock market insights**

Big data is radically changing how investors make judgements about their investments and how stock markets operate globally. When computers are given data, machine learning—the process of utilising computer algorithms to identify patterns in vast volumes of data—allows them to make predictions and conclusions that are accurate and human-like, allowing them to execute trades quickly and frequently. By including the best prices available, it enables analysts to avoid human errors caused by biases and behavioural factors and make informed conclusions. Thus, algorithmic trading in conjunction with big data is producing highly optimised insights for traders to maximise returns on their portfolios.

2**. Big data analytics in financial models**

Enhancing predictive modelling to more accurately predict rates of return and investment outcomes is made possible by big data analytics.

3. **Customer analytics**

In order to predict future behaviour, provide sales leads, capitalise on new channels and technology, improve their products, and raise customer happiness, businesses are attempting to understand the wants and preferences of their customers.
For instance, in order to create an event-based marketing strategy, the Oversea-Chinese Banking Corporation (OCBC) examined enormous volumes of previous customer data to ascertain the preferences of each individual consumer.

Methodologies:

Big data prediction in the finance business appears to be a challenging subject because there are a lot of unanswered aspects and it doesn't appear statistical at first. However, by using machine learning techniques correctly, one may teach the system to learn from the data and make appropriate assumptions by relating the past data to the present. Although there are many models for machine learning as a whole, this study uses the two most significant models to make its predictions.

1.**Classification Methods:** Among the techniques for classification are Support Vector Machine (SVM), Decision Tree, Random Forest, K-Nearest Neighbours, Näıve Bayes, Logistic Regression, and Stochastic Gradient Descent. All of these are typical in machine learning. One method for predicting discrete (usually binary) outcomes is classification. One can use logistic regression to forecast a future event that is either certain to occur or not. The decision tree is the most widely used classification technique. This is made up of nodes that make up a rooted tree, which is a directed tree that describes the relationships between the categorization and includes binary judgements. All Random Forest does is take the decision tree method and apply it to several trees. Naive Bayes is another well-liked method that determines if a given data point falls into a particular category. Numerous aspects of finance, such as portfolio creation, risk management, option pricing, and strategic hedging, are impacted by predictions made using classification matrices.

2.**Regression Methods:**
Predictive finance makes heavy use of regression techniques, which are also utilised in explanatory modelling. They are, in fact, the most prevalent type of forecast model in DSS.The capital asset pricing model (CAPM) is the most commonly utilised model in finance.
$E(Ri) = Rf + B\ (E(Rm)\ Rf)$
is essentially a single factor linear regression model. Its foundation is the linear relationship between the market return (E(Rm) explanatory variable and the expected return (E(Ri) dependent variable. Because of this, the slope that results from regressing individual security returns against the market indicates how volatile a security's return is in relation to the volatility of the market's returns (Beta). The relationship between systematic risk, or the component of returns that cannot be diversified, and expected return for assets.

3.**Clustering and association rule methods:**
When different groups, classifications, and subclassifications of data in IS are unknown, descriptive unsupervised classification algorithms such as clustering and association rules can be used to handle the data. Machine Learning (ML) models are used by association rules (AR) to search databases for patterns or connected occurrences. Events or transactions that are typically linked to fraud, like repeatedly requesting refunds, can be highlighted after data mining. For instance, Shah & Murtaza predicted financial insolvency by combining clustering with a neural network-based model.

4.**K-Means Clutering:**
K-Means An unsupervised machine learning technique called clustering divides the unlabeled dataset into various clusters.Teaching a computer to use unlabeled, unclassified material and allowing the algorithm to work on it unsupervised is known as unsupervised machine learning. In this scenario, the machine's task is to arrange unsorted data based on parallels, patterns, and variances without any prior data training.

5.**Anomally Detection Method:**
The process of spotting unusual occurrences or observations that can cause suspicion because they differ statistically from the majority of observations is known as anomaly detection. Such "anomalous" behaviour usually indicates the presence of an issue, such as a cyberattack, a failed server, credit card fraud, etc.

Big Data Challenges in the Finance Industry:

1.**Quality of Data**: Financial services firms want to make use of their data, not merely hold it. Data management is made more difficult by the fact that data originates from so many distinct sources. These data records' security, dependability, and integrity are guaranteed by information processing systems.Simultaneously, real-time analytics solutions give massive data repositories visibility, accuracy, and speed to assist businesses in deriving high-quality real-time insights and to launch new products, services, and capabilities.

2. **Complying with regulations:** The Basel Committee on Banking Supervision (BCBS) created the strict regulatory criteria known as the Fundamental Review of the Trading Book (FRTB), which control access to vital data and necessitate faster reporting. These standards must be met by financial organisations.

3. **Data privacy**: The use of cloud computing technologies raises yet another important privacy risk. Businesses are concerned about storing confidential data in the cloud. While some have built private cloud networks, these initiatives can be expensive.

4.**Data Silos**: One of the biggest business intelligence challenges today is the inability to link data across departmental and organisational silos. This can result in complex analytics and impede big data initiatives.

5.**Cyber Security**: A data breach that exposed the personal data of more than half of Americans was announced by Equifax in September 2017. A major large data breach in the twenty-first century revealed the identities and highly sensitive personal data of 147 million people. Since then, the financial services sector has made cybersecurity one of its top big data concerns. Hadoop, Spark, Casandra are just a couple of examples of big data technologies used in this industry.

Current Financial Industry Trends using Big Data:

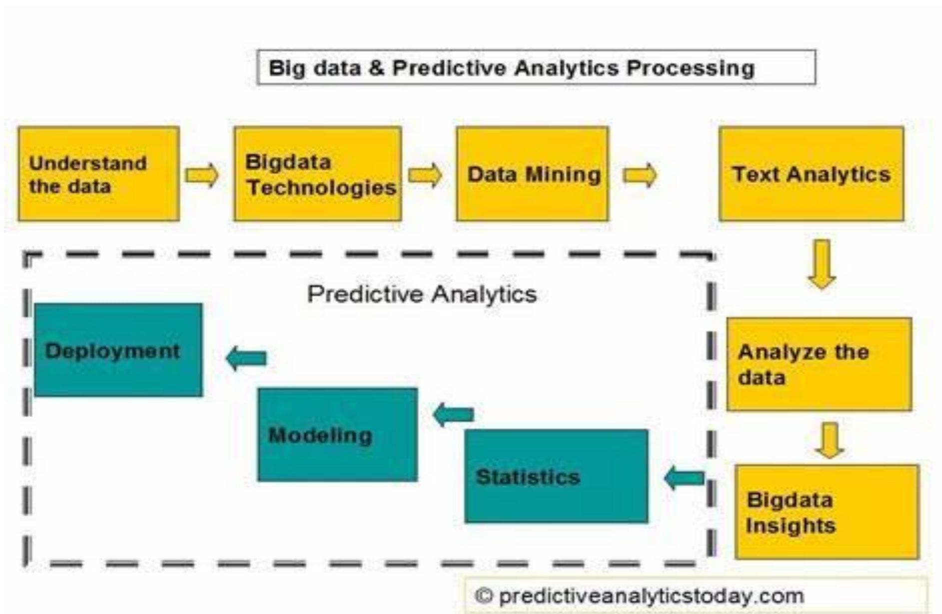
1.**Business Intelligence**: **Qlik**: Qlik is a software company that sells a product called Qlik Analytics Platform. The platform is meant to help banks and other financial institutions run what-if scenarios using big data analytics for events like natural disasters or gain business intelligence insights like which of their products are not selling well.
Financial institutions, according to Qlik, can collaborate with them to first establish a big data repository where all of the company's data is gathered. This covers both structured and unstructured data, including transaction records, social media feeds, website activity and point-of-sale data, server logs, and RFID logs from business computers and equipment.

2.**Cyber Security**: **Versive**: Versive has a product called Versive Security Engine, which it says may assist financial institutions and banks in applying machine learning to analyse massive datasets of transaction and cybersecurity-related data.

Financial institutions can incorporate Versive's software into their conventional infrastructure whether they are operating in cloud, hybrid, or on-premises environments. Clients can feed the Versive Security Engine with their netflow, proxy, and DNS data (computer network data). Banks and other financial organisations collaborate with a Versive development team to integrate their security platform in two stages.

The Versive Security Engine software is first deployed on the firm's on-premises, hybrid, or cloud enterprise network. After that, the software absorbs internal company data. During the second stage, the programme makes use of machine learning methods to find patterns in the network data that represent "normal" network properties. These trends are connected to a "baseline of operations" derived from data points in which there are no cybersecurity incidents at the bank.

3.**Regulatory Compliance**:Ayasdi: Through its software, Ayasdi's Model Accelerator (AMA), the company provides big data analytics and artificial intelligence services. It asserts that AMA can assist financial services firms in predicting and modelling regulatory risk through machine learning. In contrast to conventional rule-based approaches, Ayasdi asserts that their software can assist banks with applications like regulatory compliance for anti-money laundering (AML), automatically monitoring client transaction data to spot anomalies, and lowering the false positive rates in fraud detection. Additionally, according to Ayasdi, its platform makes use of Topological Data Analysis (TDA), which was created for a DARPA-funded study.



Conclusion:

This paper reviewed predictive analytics in finance using a wide range of approaches. It was demonstrated that scholars have employed statistical and computational methods from IS, either internally or outside, to address a variety of financial problems. Using external data to forecast the economy and company earnings is one of these. In this regard, it was demonstrated that predictive models had been applied to historical time series analysis and return forecasting in conjunction with external data. According to an IBM research from 2015, humans produce 2.5 exabytes, or 2.5 quintillion bytes, of data every day. A quintillion has eighteen zeros, to put that into perspective.
Staying ahead of the curve becomes increasingly crucial for financial services executives and C-suites as big data grows. Furthermore, data generation isn't going to stop anytime soon. We will produce 463 exabytes of data per day by 2025, which is the same as 212 million DVDs every day, according to the World Economic Forum. By 2025, it's expected that robo-advisors would function as digital platforms that offer automated financial planning services driven by algorithms in place of human advisors.

References:

[1] M. Usmani, S. H. Adil, K. Raza and S. S. A. Ali, "Stock market prediction using machine learning techniques," 2016 3rd International Conference on Computer and Information Sciences (ICCOINS), Kuala Lumpur, 2016, pp. 322-327.

[2] K. Raza, "Prediction of Stock Market performance by using machine learning techniques," 2017 International Conference on Innovations in Electrical Engineering and Computational Technologies (ICIEECT), Karachi, 2017, pp. 1-1.

[3] H. L. Siew and M. J. Nordin, "Regression techniques for the prediction of stock price trend," 2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE), Langkawi, 2012, pp. 1-5.

[4] S. Liu, G. Liao and Y. Ding, "Stock transaction prediction modeling and analysis based on LSTM," 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), Wuhan, 2018, pp. 2787-2790.

[5] T. Gao, Y. Chai and Y. Liu, "Applying long short term momory neural networks for predicting stock closing price," 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, 2017, pp. 575-578.

[6] Malkiel, B.G.; Fama, E.F. Efficient capital markets: A review of theory and empirical work. *J. Financ.* **1970**, *25*, 383–417. [**Google Scholar**] [**CrossRef**]

[7] Fama, E.F. Random walks in stock market prices. *Financ. Anal. J.* **1995**, *51*, 75–80. [**Google Scholar**] [**CrossRef**][**Green Version**]

[8] Pai, P.-F.; Lin, C.-S. A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega* **2005**, *33*, 497–505. [**Google Scholar**] [**CrossRef**]

[9] Araújo, R.d.A. A quantum-inspired evolutionary hybrid intelligent approach for stock market prediction. *Int. J. Intell. Comput. Cybern.* **2010**, *3*, 24–54. [**Google Scholar**] [**CrossRef**]

[10] Vui, C.S.; Soon, G.K.; On, C.K.; Alfred, R.; Anthony, P. A review of stock market prediction with Artificial neural network (ANN). In Proceedings of the 2013 IEEE international conference on control system, computing and engineering, Mindeb, Malaysia, 29 November–1 December 2013; pp. 477–482. [**Google Scholar**]

[11] Shi, C.; Zhuang, X. A study concerning soft computing approaches for stock price forecasting. *Axioms* **2019**, *8*, 116. [**Google Scholar**] [**CrossRef**][**Green Version**]

[12] Nti, I.K.; Adekoya, A.F.; Weyori, B.A. A systematic review of fundamental and technical analysis of stock market predictions. *Artif. Intell. Rev.* **2019**, *53*, 3007–3057. [**Google Scholar**] [**CrossRef**]

[13] Lv, D.; Huang, Z.; Li, M.; Xiang, Y. Selection of the optimal trading model for stock investment in different industries. *PLoS ONE* **2019**, *14*, e0212137. [**Google Scholar**] [**CrossRef**] [**PubMed**][**Green Version**]

[14] Atsalakis, G.S.; Valavanis, K.P. Surveying stock market forecasting techniques–Part II: Soft computing methods. *Expert Syst. Appl.* **2009**, *36*, 5932–5941. [**Google Scholar**] [**CrossRef**]

[15] Seng, J.-L.; Yang, H.-F. The association between stock price volatility and financial news–a sentiment analysis approach. *Kybernetes* **2017**, *46*, 1341–1365. [**Google Scholar**] [**CrossRef**]

[16] Chen, K.; Zhou, Y.; Dai, F. A LSTM-based method for stock returns prediction: A case study of China stock market. In Proceedings of the 2015 IEEE International Conference on Big Data, Santa Clara, CA, USA, 29 October–1 November 2015; pp. 2823–2824. [**Google Scholar**].