IOSR Journal of Business and Management (IOSR-JBM)
e-ISSN:2278-487X, p-ISSN: 2319-7668. Volume 25, Issue 7. Ser. 4 (July. 2023), PP 36-41
www.iosrjournals.org

Data Mining and Machine Learning: technologies for data handling and storage in the stock market

Francisco Antonio Nascimento, Adelcio Machado dos Santos [1], Larissa Torres Ferreira[2], Eduardo Mauch Palmeira [3], Gabriela Halfeld Barros Duarte[4], Luiz Carlos Roncaglione[5], Lucas Alves de Oliveira Lima [6], Luiz Eduardo Takenouchi Goulart[7]. Elaine Ribeiro de Oliveira[8],

Universidade Federal do Ceará (UFC) – Brasil)
[1](Universidade Alto do Vale do Rio Verde (UNIARP) – Brasil)
[2](Faculdade Integrada do Ceará - FIC – Brasil)
[3](Universidade Federal de Pelotas – UFPel- Brasil)
[4](Universidade Federal Rural do Rio de Janeiro - UFRRJ, Brasil)
[5](Universidade Federal de São Carlos UFSCar- Brasil)
[6] (Universidade Federal Rural do Rio de Janeiro - UFRRJ, Brasil)
[7](Universidade de Sao Caetano do Sul – USCS, Brasil)
[8](Universidade do Estado Minas Gerais – UEMG, Brasil)

Abstract:
Data mining and machine learning have proven to be indispensable technologies for the safe and potential development of the stock market. Through the analysis of large volumes of data, these techniques allow the identification of patterns, trends, and hidden correlations, providing valuable information to investors and financial institutions. This study conducts a literature review on the application of data mining and machine learning in the stock market. The methodology used consists of a critical analysis of the existing literature on the topic, covering scientific articles, books, and relevant publications. The objective is to investigate the importance of these mechanisms for the security, storage, and permanence of data in the financial market. The literature review performed in this research allows us to deduce that data mining and machine learning are indispensable technologies for the stock market to develop in a safe and potential way. These techniques allow the efficient analysis and interpretation of large volumes of data, ensuring transaction security, adequate information processing, and quality in the management of financial assets.

Key Word*: Data security. Technology. Analysis.*

Date of Submission: 08-07-2023 Date of Acceptance: 18-07-2023

I. Introduction

In the stock market of major nations, a scenario of intense activity is continually developing. With a considerably high volume of operations, this market becomes an attractive environment for investors and financial institutions seeking profit opportunities. Given this complexity, data mining techniques prove essential to correctly analyze the available information and extract valuable knowledge (RAMINELLI; SANTOS, 2019).

Data mining is a process that uses algorithms and tools to explore extensive data sets, seeking to identify patterns, trends and hidden relationships. In the case of the stock market, these techniques allow investors and financial institutions to perform an in-depth analysis of the available information, assisting them in making decisions for better control and observation of future trends (RAMINELLI; SANTOS, 2019).

By extracting knowledge from data, it is possible to identify correlations between different variables, such as biological factors, politics, and global events, and their impacts on stock prices. This enables investors to anticipate market movements and adjust their investment strategies accordingly. In addition, data mining also helps in identifying arbitrage opportunities, which consist of profiting from price discrepancies between different markets or assets (ASHTIANI; RAAHMEI, 2023).

Another important aspect of data mining in the stock market is the ability to predict future stock prices. Through the analysis of historical patterns, technical and behavioral indicators of the market, it is possible to develop predictive models that estimate the direction and volatility of prices. These predictions can be used to guide investment strategies, such as buying or selling certain stocks at specific times (KUMBURE, 2022).

Thus, the main goal of data mining techniques in the stock market is to obtain the highest possible profits. By correctly analyzing data and extracting valuable knowledge, investors and financial firms can identify investment opportunities, minimize risks, and maximize their financial returns. The use of these techniques

therefore becomes an indispensable tool for those seeking success in the major nations' stock market (KUMAR; SARANGI; VERMA, 2022).

Machine learning is a field of study that aims to provide computers with the ability to learn and make decisions based on data, without the need to be explicitly programmed. Instead of following followed instructions and specific rules, machine learning algorithms are expected to analyze big data and identify patterns and trends in order to make predictions or decisions (KHAN, 2020).

This research is justified to the extent that the traditional analysis based on financial indicators and company fundamentals is not always sufficient to fully understand the behavior of the stock market, which is why data mining represents an important aspect in this scenario. This mechanism complements this traditional approach by allowing the consideration of additional information, such as social media data, market trends, and other unrecorded data sources. This information can reveal trends and patterns that are essential for making exciting and timely decisions (CHOWDHURY et al, 2020).

Another important justification is the need to predict future stock trends and prices. Data mining enables the construction of complex predictive models that consider a variety of variables and indicators. These models can help anticipate market movements, identify specific investment opportunities, and manage risk more effectively (KUMBURE, 2022).

Thus, this research is composed of an introduction, in which the research elements are presented with a simple and objective contextualization to direct the reader to the theme; the methodology, in which the selection process of the selected literature will be explained; the results analysis section, in which the aspects observed during the research will be presented; and the final considerations, in which the conclusions to which it was possible to reach after the data collection and analysis will be presented.

II. Material And Methods

In this research, a literature review will be conducted, which consists of a systematic process of research and analysis of existing studies and academic papers on a particular topic or research problem. It is an essential step in the preparation of academic research, dissertations, theses or scientific articles, because it allows the researcher to obtain a comprehensive overview of the knowledge already produced on the subject in question (LIMA; MIOTO, 2007).

A literature review involves the search and selection of relevant studies, the reading and critical analysis of these studies, and the summary of the information found. The main objective is to identify gaps, trends, controversies or consensus in the existing literature, as well as to contribute to the construction of a solid and grounded argument for the developing research (DE SOUSA; DE OLIVEIRA, 2021).

In this context, within the databases responsible for storing research and studies conducted by scientists, specialists and researchers, it is possible to conduct searches so that one can have access to investigations and literature that provide knowledge about the theme to be researched. Furthermore, this process must be accomplished by means of criteria inherent to the purpose of the research and appropriate to the method chosen by the researcher, making it necessary to use filters that work from logical timelines that enable the bibliometric review (PIZZANI et al., 2012).

The present research was developed, first, through the selection of the Science Direct, Gloogle Academic, CAPES and IEEE Digital Library storage banks. For the selection of articles, the following keywords were entered: " ‘‘Técnicas de mineração de dados no mercado de ações"; "Mineração de dados e aprendizado de máquina no mercado de ações";"Data mining techniques in the stock market"; "Data mining and machine learning in the stock market".

In order to present the research questions, a visualization was developed to guide the search conducted according to the research objectives:

Chart 1 - Research questions

QP1	What benefits do data mining and machine learning techniques bring to the stock market?
QP2	In contemporary times, what are the needs regarding the use of these mechanisms for the stock market?
QP3	How can data mining and machine learning techniques boost the fields of production and development in the stock market?

Source: Prepared by the author, 2023.

Inclusion and exclusion criteria were conducted as follows:

Inclusion criteria:

- CI1: Research published in the last 5 years.
- CI2 Titles that have the same or similar terms to the keywords used.
- CI3: Scientific articles, dissertations or thesis.
- CI4: Publications in English or Portuguese language.

Exclusion criteria:

- CE1: Research published before the year 2019.
- CE2: Research that is not directly related to the subject of this research from the key words used and from reading the titles and abstracts.
- CE3: Publications in languages other than English or Portuguese.

2.1 Applying the criteria in each selected database

2.1.1 Science Direct

After entering the keyword, 20,960 results were found. After applying inclusion and exclusion criteria, the search was reduced to 16 publications. After reading the abstracts, 5 articles considered to be directly related to the theme of this research were selected:

Chart 2 - Selected searches Science Direct

YEAR	AUTHOR	TITLE
2023	Matin N. AshtianiBijan Raahemi	News-based intelligent prediction of financial markets using text mining and machine learning: A systematic literature review
2022	Mahinda Mailagaha Kumbure, Christoph Lohrmann, Pasi Luukka, Jari Porras	Machine learning techniques and data for stock market forecasting: A literature review
2021	Deepak KumarPradeepta Kumar SarangiRajit Verma	A systematic review of stock market prediction using machine learning and statistical techniques
2021	Abdulhamit Subasi,Faria Amir,Kholoud Bagedo,Asmaa Shams,Akila Sarirete	Stock Market Prediction Using Machine Learning
2020	Reaz Chowdhury a,MRC Mahdy a,Tanisha Nourin Alam a,Golam Dastegir Al Quaderi b,M. Arifur Rahman	Predicting the stock price of frontier markets using machine learning and modified Black-Scholes Option pricing model

Source: Prepared by the author, 2023.

No articles fitting the inclusion and exclusion criteria in Portuguese were found in this search platform.

2.1.2 CAPES Journals

After entering the keyword, 410 results were found. After applying the inclusion and exclusion criteria, the search was reduced to 3 selected searches, considering that searches found in duplicity with the other searched platforms were also excluded.

Chart 3 - Selected searches CAPES journals

YEAR	AUTHOR	TITLE
2023	Sakhare, Nitin Nandkumar ; Shaik, Imambi S. ; Saha, Suman	Prediction of stock market movement via technical analysis of stock data stored on blockchain using novel History Bits based machine learning algorithm
2020	Nabipour, Mojtaba; Nayyeri, Pooyan; Jabani, Hamed; S., Shahab; Mosavi, Amir	Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; a Comparative Analysis
2022	Raubitzek, Sebastian ; Neubauer, Thomas	An Exploratory Study on the Complexity and Machine Learning Predictability of Stock Market Data

Source: Prepared by the author, 2023

No Portuguese language research was found on this platform that fit all the search criteria.

2.1.3 Google Scholar

According to the selected keywords, 80,500 results were found on this search platform. After applying all inclusion and exclusion criteria, the search was reduced to 30 results. After reading the abstracts to check for compatibility with the topic, and excluding the searches already selected in the previous search platforms, 5 searches were selected:

Chart 4 - Selected Google Academic Searches

YEAR	AUTHOR	TITLE
2019	Daniele Gonçalves de Toledo Luchetta Raminelli, Bruno Samways dos Santos	Applying Data Mining and Machine Learning Techniques Machine in the Stock Market: A Systematic Review
2021	Leonardo Leon Dias, Kelly Bezerra da Mota, Gabriela de Vasconcelos, Diogo Tavares Cavalcanti de Moraes, Guilherme Teixeira Ribeiro de Albuquerque	Application of Data Mining for Evaluating Financial Indicators
2021	Ana Carolina de Oliveira	Artificial intelligence applied to the financial market for decision making
2020	Almeida, Lucas Gomes Rocha, Rafael da Silva	Data mining of stock data with fundamentalist indicators
2020	Piloneto, Julia, Vieira, Mateus Hecht, Renan	Predictive analysis of the stock market with natural language processing

Source: Prepared by the author, 2023

2.1.4 IEEE Digital Library

After entering the keywords, 19,823 results were found. After applying the exclusion criteria and eliminating the searches that had already been contemplated in the previous platforms, 3 searches were selected:

Box 5 - IEEE Digital Library

YEAR	AUTHOR	TITLE
2019	Zhihao Peng	Stocks Analysis and Prediction Using Big Data Analytics
2020	Hind Bourezk; Amine Raji; Nawfal Acha; Hafid Barka	Analyzing Moroccan Stock Market using Machine Learning and Sentiment Analysis

2021	Golshid Ranibaran; Mohammad-Shahram Moin; Sasan H Alizadeh; Abbas Koochari	Analyzing effect of news polarity on stock market prediction: a machine learning approach

Source: Prepared by the author, 2023

After observing the collected works in these 4 search platforms, it is possible to observe a higher incidence of research in English language. In addition, it could also be observed that there is not a large number of researches related to the theme, since most of the files found referred to data mining in educational fields.

III. Result

During the analysis of the files selected for the literature review, it was possible to observe that data mining and machine learning has been a highly explored field in the most varied market spheres. In the stock market, these mechanisms prove to be of extreme importance for data storage, information security, data control and handling, and maintaining management quality.

In addition, data mining contributes to the identification of vulnerabilities and security gaps in data storage systems. Through data analysis, it is possible to identify insufficient points in the processing infrastructure and adopt corrective measures to ensure data security (RAMINELLI; SANTOS, 2019).

In the context of data warehousing, data mining helps organize and structure information efficiently. By analyzing stored data, it is possible to identify usage patterns, relationships between different data sets, and access trends. This information can be used to optimize storage, improve data distribution, and ensure that the most relevant information is available for query and analysis (PENG, 2019).

In relation to management control and quality, data mining allows valuable insights to be identified from the available data. By analyzing the data, it is possible to discover market behavior patterns, influenced by customers, consumption trends, and other relevant aspects for strategic decision-making. These insights help organizations adapt their strategies, improve operational efficiency, and offer products and services better suited to customer needs (DIAS et al, 2021).

Furthermore, data mining also contributes to the control and monitoring of the quality of the processes and operations of organizations. Through the analysis of the collected data, it is possible to identify deviations, errors or problems in the processes, allowing the implementation of corrective actions and the continuous improvement of quality (OLIVEIRA, 2021).

When it comes to information security and control, machine learning techniques play a central role in detecting and preventing cyber threats and fraudulent activities. These algorithms are able to analyze large volumes of data in real time, identifying suspicious patterns, anomalous behavior, and signatures of known attacks. With this, it is possible to significantly improve the ability to detect and respond to security incidents, protecting information effectively and avoiding financial and reputational damage (ALMEIDA; ROCHA, 2019).

In the context of data warehousing, machine learning techniques play an important role in organizing and structuring data efficiently. These techniques enable the analysis of stored data by identifying hidden correlations, relationships and patterns in datasets. This enables a better understanding of the data, improving the indexing, retrieval, and distribution of stored information, as well as optimizing the performance of storage systems (SAKHARE; SHAIK; SAHA, 2023).

As for management control and quality, machine learning techniques provide valuable insights and data-secure information to aid strategic decision making. By analyzing the available data, these techniques identify patterns and trends, providing important information about market behavior, customer influences, and the internal dynamics of the organization. These insights help managers make controlled and informed decisions, improving operational efficiency, optimizing resource allocation, and anticipating trends (KUMAR; SARANGI; VERMA, 2022).

In addition, machine learning techniques also work for controlling and monitoring the quality of processes and operations. By analyzing the data, it is possible to identify deviations, inconsistencies, and problems in processes, allowing corrective actions to be implemented proactively. This results in improved quality of products and services, reduced costs, and increased customer satisfaction..

IV. Conclusion

From this research, it was possible to conclude that data mining represents a fundamental mechanism to ensure the security and control of information, in data storage and in the control and quality of management in the stock market. This technique makes it possible to identify suspicious patterns, prevent fraudulent activities, optimize data storage, identify valuable insights, and monitor the quality of processes. By using data mining effectively, organizations can make more informed decisions, ensure information security, and improve operational efficiency.

Furthermore, it was possible to observe that machine learning techniques represent an essential tool in guaranteeing the security and control of information, in data storage, and in the control and quality of management in organizations. These techniques allow the detection and prevention of cyber threats, an efficient organization of data, obtaining valuable insights for strategic decision making, and monitoring the quality of processes. By effectively applying machine learning techniques, organizations can strengthen their security, improve operational efficiency, and maintain a competitive advantage in the stock market..

References

[1]. ALMEIDA, Lucas Gomes; ROCHA, Rafael da Silva. Mining stock data with fundamentalist indicators. 2019.

[2]. ASHTIANI, Matin N.; RAAHMEI, Bijan. News-based intelligent prediction of financial markets using text mining and machine learning: A systematic literature review. Expert Systems with Applications, p. 119509, 2023.

[3]. CHOWDHURY, Reaz et al. Predicting the stock price of frontier markets using machine learning and modified Black-Scholes Option pricing model. Physica A: Statistical Mechanics and its Applications, v. 555, p. 124444, 2020.

[4]. DE SOUSA, Angélica Silva; DE OLIVEIRA, Guilherme Saramago; ALVES, Laís Hilário. Bibliographic research: principles and fundamentals. Cadernos da FUCAMP, v. 20, n. 43, 2021.

[5]. DIAS, Leonardo Leon et al. Application of Data Mining for Evaluation of Financial Indicators. Journal of Engineering and Applied Research, v. 6, n. 5, p. 37-46, 2021.

[6]. KHAN, Wasiat et al. Stock market prediction using machine learning classifiers and social media, news. Journal of Ambient Intelligence and Humanized Computing, p. 1-24, 2020.

[7]. KUMAR, Deepak; SARANGI, Pradeepta Kumar; VERMA, Rajit. A systematic review of stock market prediction using machine learning and statistical techniques. Materials Today: Proceedings, v. 49, p. 3187-3191, 2022.

[8]. KUMBURE, Mahinda Mailagaha et al. Machine learning techniques and data for stock market forecasting: A literature review. Expert Systems with Applications, p. 116659, 2022.

[9]. LIMA, Telma Cristiane Sasso de; MIOTO, Regina Célia Tamaso. Methodological procedures in the construction of scientific knowledge: the bibliographic research. Revista katálysis, v. 10, p. 37-45, 2007.

[10]. OLIVEIRA, Ana Carolina de. ARTIFICIAL INTELLIGENCE APPLIED TO FINANCIAL MARKETS FOR DECISION MAKING. 2021.

[11]. PENG, Zhihao. Stocks analysis and prediction using big data analytics. In: 2019 international conference on intelligent transportation, big data & smart city (ICITBS). IEEE, 2019. p. 309-312.

[12]. PIZZANI, Luciana et al. The art of bibliographic research in the search for knowledge. RDBCI: Digital Journal of Library and Information Science, v. 10, n. 2, p. 53-66, 2012.

[13]. RAMINELLI, DGDETL; SANTOS, B. DOS S. Application of Data Mining and Machine Learning Techniques to the Stock Market: A Systematic Review. In: Brazilian Congress of Production Engineering. 2019.

[14]. SAKHARE, Nitin Nandkumar; SHAIK, Imambi S.; SAHA, Suman. Prediction of stock market movement via technical analysis of stock data stored on blockchain using novel History Bits based machine learning algorithm. IET Software, 2023.