



ORIGINAL

Using Data Mining Principles in Implementing Predictive Analytics to Different Areas

Utilización de los principios de la minería de datos en la aplicación del análisis predictivo a distintos ámbitos

Bahar Asgarova¹, Elvin Jafarov¹, Nicat Babayev¹, Allahshukur Ahmadzada¹

¹Azerbaijan State Oil and Industry University, Baku, Azerbaijan.

Cite as: Asgarova B, Jafarov E, Babayev N, Ahmadzada A. Using Data Mining Principles in Implementing Predictive Analytics to Different Areas. Data and Metadata. 2024; 3:.380. <https://doi.org/10.56294/dm2024.380>

Submitted: 02-02-2024

Revised: 04-05-2024

Accepted: 23-08-2024

Published: 24-08-2024

Editor: Adrián Alejandro Vitón-Castillo 

ABSTRACT

This study delves into the realm of information-based knowledge discovery technologies and underscores the growing necessity for extensive data representation to enhance the management of care and mitigate the financial costs associated with promoting long-term care. The proliferation of information collected and disseminated through the Internet has reached unprecedented levels in the context of long-term financial health statistics, posing a challenge for businesses to effectively leverage this wealth of data for research purposes. The explicit specification of costs becomes paramount when dealing with substantial volumes of data. Consequently, the literature on the application of big data in logistics is categorized based on the nature of methods employed, such as explanatory, predictive, regulatory, strategic, and operational approaches. This includes a comprehensive examination of how big data analysis is applied within large corporations. In the healthcare domain, the study contributes to the evaluation of usability by providing a framework to analyze the maturity of structures at four distinct levels. The emphasis is particularly on the pivotal role played by predictive analytics in the healthcare industry through big data methodologies. Furthermore, the study advocates for a paradigm shift in management's perception of large business data sets, urging them to view these as strategic resources that must be seamlessly integrated into the company. This integration is seen as imperative for achieving comprehensive business analysis and staying competitive in the ever-evolving landscape of healthcare. The study concludes by shedding light on the limitations inherent in the research and delineating the specific focus areas that have been addressed.

Keywords: Data Mining Technology; Big Data; Predictive Analysis; Data Processing.

RESUMEN

Este estudio se adentra en el ámbito de las tecnologías de descubrimiento de conocimientos basados en la información y subraya la creciente necesidad de una amplia representación de datos para mejorar la gestión de los cuidados y mitigar los costes financieros asociados a la promoción de los cuidados de larga duración. La proliferación de información recopilada y difundida a través de Internet ha alcanzado niveles sin precedentes en el contexto de las estadísticas de salud financiera a largo plazo, lo que plantea a las empresas el reto de aprovechar eficazmente este caudal de datos con fines de investigación. La especificación explícita de los costes adquiere una importancia capital cuando se trata de volúmenes considerables de datos. En consecuencia, la bibliografía sobre la aplicación de big data en logística se clasifica en función de la naturaleza de los métodos empleados, como enfoques explicativos, predictivos, normativos, estratégicos y operativos. Esto incluye un examen exhaustivo de cómo se aplica el análisis de macrodatos en las grandes empresas. En el ámbito sanitario, el estudio contribuye a la evaluación de la usabilidad proporcionando un marco para analizar la madurez de las estructuras en cuatro niveles distintos. Se hace especial hincapié en el papel fundamental que desempeña el análisis predictivo en la industria sanitaria a través de las metodologías

de big data. Además, el estudio aboga por un cambio de paradigma en la percepción por parte de los directivos de los grandes conjuntos de datos empresariales, instándoles a considerarlos como recursos estratégicos que deben integrarse perfectamente en la empresa. Esta integración se considera imprescindible para lograr un análisis empresarial exhaustivo y seguir siendo competitivos en el panorama en constante evolución de la atención sanitaria. El estudio concluye arrojando luz sobre las limitaciones inherentes a la investigación y delineando las áreas de interés específicas que se han abordado.

Palabras clave: Tecnología de Minería de Datos; Big Data; Análisis Predictivo; Procesamiento de Datos.

INTRODUCTION

Predictive analysis includes statistical models and other empirical methods used to make empirical predictions, not only theoretically possible predictions, but also to assess the quality of these predictions. In addition to training, prediction plays an important role in developing, such as theoretical, analytical tests, and performance evaluation (Mahmoud, 2017). The bottom line is that companies can use predictive analytics to find and use models in their data to identify risks and opportunities. For forecasting purposes, Amazon's success is primarily due to its review of innovative ideas, such as the latest technologies implemented from the start, the possibility of small order processing, predictive logistics such as data processing, and large data sets (Spiegel et al., 2013). Amazon regularly sets the address of the item to be purchased and ships it to the public before purchase. Large data sets are important in determining consumer market trends and acquisition patterns, which can better match supply and demand relationships and therefore reduce supply chain costs (Wang et al., 2016; Hazen et al, 2014) acknowledge that organizations must employ massive data sets to be active and competitive and that predictive analytics and data mining provide creative approaches to enhance supply chain procedures.

In today's era where data extensively generated and gathered the importance of predictive analytics should not be underestimated. Use of predictive analysis spread over different areas ranging from forecasting patient health trends in healthcare, to understanding consumer habits in retail, and predicting weather changes in environmental research. These examples not just show the scope of predictive analysis but also show its potential in industries development, decision making and planning.

Literature review and problem statement

In response to the growth rate of massive data, many domestic and international stakeholders in massive data mining and knowledge discovery have conducted in-depth research. The storage and processing capacity of massive data itself puts forward high requirements for data mining or machine learning, and the work done by Google in this regard is significant. Map Reduce, proposed by Google, is a framework model for concurrently processing massive data on large computer clusters (Wei et al, 2009; Shi et al, 2013). It first transforms the input data into corresponding key-value pairs by setting a Map function, then aggregates the values with the same key through a custom reduce function and outputs the result. Most of the natural world can use this model to represent the processing of large amounts of data. In addition, the parallel database is a product of the combination of database technology and parallel technology. It is regarded as a high-performance database system that can significantly improve the efficiency of processing massive data in relational databases.

In addition, how to deal with massive amounts of data has been a bottleneck to be solved by data mining. Many algorithms for handling massive data, such as parallel or serial algorithms, generally need help to solve the contradiction between speed and correctness. However, distributed computing has apparent advantages in data processing. In 2004, Tan Zhengrin, Wu Yu, and others split the original massive data set into thousands of small data sets and then processed them in a distributed way; based on this idea, a massive data segmentation algorithm based on Rough Set was given. It is proved through experiments that the scheme combining the distributed processing of the data partitioning algorithm can process the massive data quickly and maintain the correctness of the algorithm compared with the algorithm that processes the whole dataset. In 2004, Zhang Zhaogong, Li Jianzhong, et al. (Zhang et al, 2004), in response to the current problem that the running time of the association rule mining algorithm is too long when the database is enormous, put forward an effective way to solve the problem that is parallel computing In order to solve the problem of the long running time of association rule mining algorithms on a large amount of data, a new method of parallel random sampling is proposed by using a parallel algorithm that can ignore the local frequent itemsets generated only on less than 1/4 of the nodes and combining with the characteristics of the I/O of the cluster parallel machine which is highly parallel and its processing capability, the efficiency and capability of processing large amount of data is improved. Simulation results show that. The acceleration ratio of this algorithm is close to the number of processors p , the communication complexity is the logarithm of the number of processors p , and it has good

scalability, high accuracy, and the ability to process massive data. Wei Ting, (Wu et al, 2009; Cao, 2013) studied data mining applications based on granular computing. Granular Computing (GrC) is a computational paradigm and a new concept in information processing, which covers the entire research on theories, techniques, tools, and methods related to granularity and is mainly used to deal with fuzzy, massive, uncertain, and incomplete information. Sun Zhichang, Feng Zuhong, et al. proposed a new hybrid compression algorithm (HC-DM algorithm) in response to the fact that association rule mining algorithms based on vertical data formats must keep an extensive list of transaction flags in memory during the frequent item-finding process. The limited memory capacity will become the biggest bottleneck for such algorithms. This efficient hybrid compression data mining algorithm combines the HC-DM algorithm with the dEclat algorithm and a sorting step, which can effectively reduce memory usage in mining frequent item sets. In January 2011, (Zhang, 2011) analyzed and summarized the massive data processing techniques for dealing with the query efficiency of massive data.

Eckerson (2007) claims that predictive analytics works in many ways, including induction, and does not provide data assumptions, machine learning techniques, neural computing, robotics or computational mathematics, and artificial intelligence techniques. Companies that find predictive analytics can find millions of inventions, but smart robots will continue to review each other until they find something interesting in the data (Eckerson, 2007). A few years ago, the firm was mocked for owning a little portion of the shoe and built a prediction algorithm to identify clients who could terminate cellular phone service. The model reduces the number of cancellations and provides a competitive advantage in providing company information on customer behaviour, such as suppliers, to review business and marketing processes (Eckerson, 2007). The author used production models to offer special offers to abandoned customers and work with our managers to modify licensing policies that affect tax-free rates.

The need for predictive analytics varies by industry, as it is used to identify relationships and patterns in the data and analyse the past to make better security decisions in the future (Mahmood, 2017). For example, suppliers can use predictive analytics to predict the response to consumer advertising campaigns, low cost, and reduced product costs. Mahmood (2017) believes that predictive analytics can help credit card customers predict trends by identifying or informing potential customers. Firms and markets in many countries often use predictive analytics in new businesses. For example, a study describing the characteristics of data in a memory database and its impact on future memory problems uses the Singapore Markov Chain Model to control future purchases by government agencies (Mahmood, 2017). We reach this goal for many candidates by evaluating methods such as probability distribution analysis, simple random sampling, linear analysis, and the Markov chain (Mahmood, 2017). This not only reduces the effectiveness of the cluster layout, but also helps predict more candidates. Use variable correlation methods to find relationships.

Therefore, for other purposes, Business Predictive Analytics is widely used to improve the productivity of the organization and reduce the time required to provide services (Pandey, Nepal & Chen, 2011). The organization complies with the signed service level agreement. Customers must predict event logs and processes, build models and predict workflow time.

The healthcare industry, like other businesses, now has access to a vast amount of data from medical information systems and provides informed medical and administrative decision-making, its goal is to combine several years of medical data based on medical data from medical providers, but traditionally most medical data is kept in paper files, images and generic radiographs (Reddy & Kumar, 2016). Using rapid data for rapid line detection, future use for detection of dangerous diseases and infections, these predictions can enable hospital outbreaks, patient morbidity, real-time ice surveillance data, and other tools to quickly eliminate fatal infections and analyze large amounts of health data (Bates et al. 2014). Health information and decision-making play an important role in the healthcare industry. Large data sets can analyze these health data to predict a patient's disease, to prescribe appropriate drugs for analysis, and to manage the large data needed. There are many problems with predicting too much data to help innocent companies. IT innovations and new model packages can help solve various problems in hospitals, banks, social networking sites, and construction sites. The big data management solutions available can assist businesses in making the correct decisions at the right time.

Data Mining Technology

Data Mining (DM) is extracting information and knowledge from large amounts of incomplete, noisy, and ambiguous data implied and unknown to people in advance but potentially helpful (Zhu, 2008).

Generally speaking, knowledge is a form of expression, and both data and information belong to the category of forms of expression. In addition, concepts, rules, models, rules, and constraints also belong to the category of knowledge.

Knowledge discovery can be performed in the areas of management, optimization, decision-making, control, querying, and maintenance of one's own data. It is also a data mining process, and as such, it requires the support of multiple disciplines and cannot be solved by a single discipline. This brings experts and scholars from

various fields together for the same purpose.

Data can be classified into three types: structured data, semi-structured data, and heterogeneous data (Sagiroglu & Sinanc, 2013). Structured data is data that has been identified and easily stored in a database; semi-structured data may have been labeled but is not sure which class it belongs to; and heterogeneous data is data of random type, which can be text, images, or even sound, and is difficult to recognize by a database.

The process of data mining can be summarised as describing the problem, collecting and pre-processing the data, executing the mining of the data, and interpreting and evaluating the results (Huang, 2013) (i.e., parsing the conclusions), as shown in figure 1.

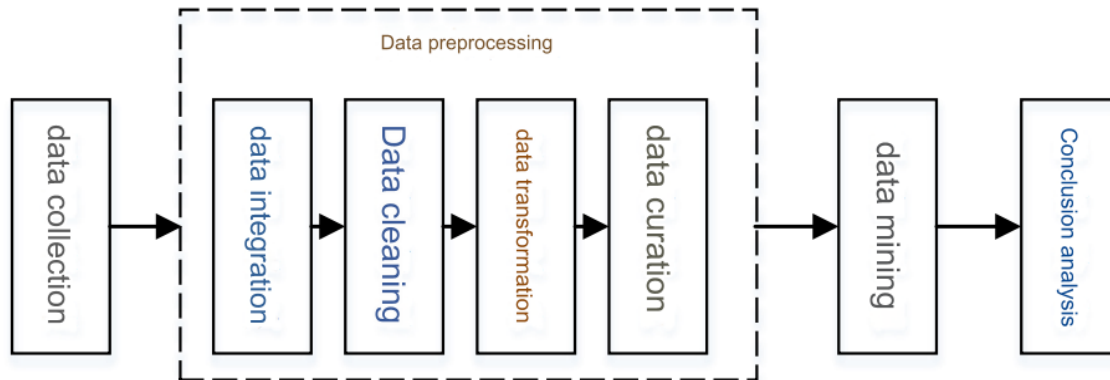


Figure 1. The process of data mining

Describe the problem

Simply put, we are engaged in data mining with a purpose, not unthinkingly looking at a vast dataset aimlessly. Therefore, the first step of data mining is to determine what you want to do in the face of this vast data, i.e., what information you want to get from this data! For this purpose, the first step of data mining will be completed; if there is no this first step, even if the algorithm is perfect and accurate, the mining results are worthless. This first step will make the result worthwhile, even if the algorithm is perfect and precise. After that, we should look for ways to achieve the purpose. The so-called methods can be algorithms, reasoning, or even simple graphical displays, like the data shown in Figure 1, and we instantly understand the trend of data mining after looking at these two figures.

Acquisition and Data Preprocessing

With the first step to do the purpose, we will focus on the purpose of the second step in the collection of data related to the collection of data, and then the collection of data for the initial processing, that is, preprocessing. Then, we will initially process data for standardized processing to carry out calculations or practical comparative analysis.

Data collection: The purpose of data collection is to obtain a certain amount of data for data mining; data collection should be based on the purpose of data mining reasonable choice; for example, if we want to predict the weather tomorrow, then before we should collect the last few days of weather-related data (temperature, wind, rainfall, PM values, humidity, etc.), and not even the GDP and other data should also be collected.

The purpose of data preprocessing is to process the collected data in a simple way because sensors collect the data, and there must be some data missing, noise, etc., due to internal or external factors during the data collection and storage process. Therefore, we need preliminary processing to make it as accurate as possible.

The purpose of normalization is to normalize data that cannot be compared together in different ranges and scales into the same range so that they can be compared. In the meantime, normalization also includes dimensionality reduction of the data, i.e., attribute simplification and other processing.

Implementation of data mining

With the first and second steps as the foundation, the third step integrates comprehensive factors and selects appropriate methods or algorithms to analyze the known massive data. The process of analysis is the process of method implementation or algorithm operation. In the end, we will express the information mined reasonably according to the characteristics of the data and the way we need to express it.

Interpretation and evaluation of results (i.e., conclusion analysis)

After the raw and massive data have been expressed in the third step, we need to analyze the information

expressed in the third step to refer to the next step of decision-making, management, and control. Suppose we are unsatisfied with the results expressed in the third step or feel they could be more trustworthy through experience or other means. In that case, we should go back to the third step and choose the appropriate methods or algorithms to re-explore them to get the needed information.

The data mining process is complete at this point. The following is an introduction to massive (extensive) data mining methods.

Cluster analysis

Clustering is the process of grouping physical or abstract sets. The groups generated by clustering are called clusters. A cluster is a collection of data objects.

In statistics, cluster analysis relies on modeling to simplify the data to be analyzed before analyzing it. It is mainly used in research applications based on distance and similarity. The traditional cluster analysis method is shown in figure 2.

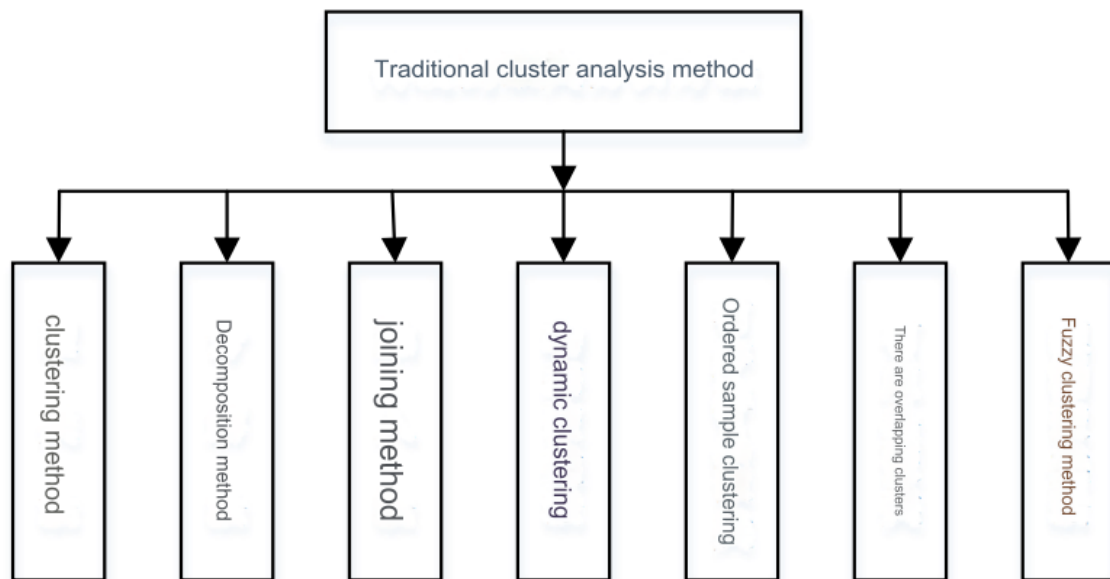


Figure 2. Traditional cluster analysis methods

In practice, cluster analysis is also a task of data mining in most cases. The main application is on efficient and practical cluster analysis algorithms for massive data or big data.

Due to the increasing size of data, therefore, in future research, cluster analysis should have the following properties:

1. Able to handle multiple types of data;
2. Scalability, especially with respect to big data;
3. Able to handle high dimensional data;
4. Ability to discover clusters of different shapes;
5. The ability to reduce noise;
6. The ability to process data independent of its order;
7. The ability to rely on user-defined parameters or a priori knowledge;
8. Processing results that are easy to understand and usable;
9. Ability to perform constrained clustering.

Association Rules

An association rule is a rule that the support and confidence in a data set satisfy a given threshold.

Definition 2.1: Let $I = (i_1, i_2, i_m)$ be the set of words (words are called items), D be the set of things, and if $T \in I$, then each of the things is labelled as TID. For a set X of some items in I , T is said to contain X if $X \in I$.

Association rule form:

$X \rightarrow Y$ ($X \in I$ and $X \cap Y = \emptyset$)

A rule $X \rightarrow Y$ is said to be bound to the set of things D with confidence $c\%$ if $c\%$ of the things in D contain X and also contain Y .

If $s\%$ of the things in D contain then $s\%$ is the support of rule $X \rightarrow Y$ in D .

Association rules reflect the dependence and association between things. If things are related, then one of them can be predicted from the other known things. Among them, association rules are classified into Boolean association rules and quantitative association rules.

Classification and Regression Analysis

Classification and regression analyses are commonly used in data analysis methods for discrete data and continuous data, respectively. Sometimes, these two types of analyses can be implemented with the same algorithm, so these two analyses are discussed together in this section.

Classification, as the name suggests, is to divide data according to some rules. Classification analysis is often used to classify (predict) discrete data. For example, classification analysis can be performed in licence plate recognition. As we all know, the licence plate is composed of 26 letters and 10 numbers 0-9. Therefore, we can first divide the collected images of the licence plate into letters and numbers one by one, and then use classification algorithms to compare and identify them, and then compose these identified data into one, so as to achieve the purpose of licence plate recognition.

Regression, on the other hand, is the establishment of a functional relationship between one or more sets of data and another set of data, the purpose of which is to establish the intrinsic correlation of data through mathematical functions. Regression analysis is often used to predict continuous data.

Currently, the following methods are commonly used in classification and regression analysis: decision tree and Bayesian classification analysis, classification and prediction based on genetic algorithms and artificial neural network algorithms. Of these, the latter two algorithms are two of the more common and widely used methods.

METHOD

This document focuses on various big data tools, technologies, and methods. For the predictive analysis of hospital data, this document is divided into five parts, beginning with the introduction, literature review, methodology, results, discussion, and conclusions. In this study, we first focused on the development of the framework. Then the authors created the prediction model of the hidden Markov model to detect the workflow model existing in the historical event log to complete the business process. Pandey et al. (2011) estimate the time required and test the prototype that integrates the proposed architecture with Predictive analytics technology. In fact, in the first stage, scientists implemented four prediction methods. It is a regression model, descriptive statistics, and system (Demir, 2014). Comparing the transitions with annotations and implicit Markov models, but after testing and evaluating the above four models, they found that the secret Markov model is the most predictable compared to other models (Pandey, Nepal & Chen, 2011). Therefore, this study was accepted. In this study, the limitations of this study focused on parameters, without considering different configurations that may affect business processes such as IT capacity and readiness and human resources. Also, the proposed model should be tested on real data to improve its accuracy and performance.

This paper does a systematic literature review using content analysis to investigate the application of data processing and predictive analysis in healthcare. We use the term “predictive analytics” to search the literature, although the term is general, it contains quantitative literature. The IEEE Xplore, SAGE, SpringerLink, ScienceDirect (Elsevier) and Google Scholar databases were searched. We restricted the search space to exclude all content, including magazines, conference books and publications, and gray literature such as white papers and blog posts. Blog posts have different attributes that affect your credibility. It should be emphasized that the above study is not thorough, but the most relevant publications published since 2010 were chosen. After the abstract, 131 articles were identified, and only 41 of these articles were reviewed and analyzed.

For each article, we classified the seven types of predictive analysis into seven types of prediction analysis:

- Drug Analysis Methods
- Probability Theory
- Machine Learning/Data Processing
- Statistical Analysis
- Mathematical Programming
- Evolutionary Calculus
- Simulation, And Logical Modeling

Research articles use a combination of methods instead of a process; we focus on what kind of predictive analysis methods and order the articles by application area. We discussed how to use the problem in context and identified based on belief and determined the appropriate contributions and completed the integration, analyzed possible directions based on existing research projects and literature reviews.

Just two publications in specialist organizations in use and three articles in the International Journal of Information and Operations Research have been published one time in 11 articles relevant to our research

topics, showing that the issue has piqued the attention of many individuals. The identification and selection of documents is carried out in three steps, filing related documents and measurements at each stage, searching the scientific database for the elements discussed in the first phase, the first phase of the study, the full text, especially the full text of the publication, without restrictions, for a total of 131 articles, called “Predictive analytics.” The term has increased significantly, which is a recent trend that has led to increased interest in research and a consistent review of the peer-reviewed literature.

Although quantitative analysis is not necessary in the first stage of the investigation, we searched in the second phase of research activities that included the expression of queries in metadata, abstract, words, or other registry metadata. In the second phase, 99 documents were produced. Still, this field was not included in some areas, but cited the term “predictive analytics.” In most cases, the introduction is used as the basis for business analysis.

Therefore, the third phase of the study aims to study the content of these documents in more detail according to the following criteria: the date of publication such as January 2010, the type of scientific publication, conference, or book publication. In the third stage, 56 articles were selected, including 19 journal articles, 36 conference articles, and book chapters. It should be noted that as the trend develops, as the number of publications that contribute to the ethical analysis of article number distribution increases, and there are no journals or conferences to collect scientific articles.

In terms of measurement tools, data processing to “improve performance and extraction process” only focuses on improving the quality of care. We use the method of Ben Gore et al. (2006) to assess the quality of nursing measures to determine the application of process improvement. Seven articles discuss “performance” and “nursing” in two studies and seek acceptable / patient-centered care, evidence-based medicine. Best health outcomes help effectively treat patients with respiratory distress syndrome based on a diagnosis of Acute pulmonary edema (AMI), maternal death/hemodialysis, and the results of cardiovascular disease (Zheng et al., 2015). A special system is designed to distribute the appropriate quantity of medicines to obtain information from text mining information (Rubrichi & Quaglini, 2012). Quality and safety are attributed to low-risk and unsuitable users of medical services, as well as their preference for an acceptable personal quality of service based on patient statistics and medical records and expectations. Therefore, in predicting security-related events, your work is divided into high-quality safety and care factors, and this study includes a comprehensive data-mining program that uses customer experience and feedback to understand the best care practices (Spruit, Vroon & Batenburg, 2014). Prediction route of medical route on-site care, clinical or parental process innovation, providing medical services and detailed clinical instructions or referral process

James and Savitz (2011) drew attention to analyzing the data mining workflow to determine the clinical data path. In one case, the difference between the recommended route and the actual health route. The relationship between the logo of the medical course and the journey between the hospital and the hospital care network. Because the nursing process has improved significantly, I created a supporting article that identified and analyzed several treatment methods, outlined the diversity of methods and other clinical care goals, and predicted the length of hospital stay for treated patients in research, especially radiology. In summary, data processing is an interdisciplinary path.

We discovered that these approaches may be utilized as retrospective reviews under the “Person, Job, and Organization” category. Spruit et al. (2014) forecast several economic indices using historical data from Dutch long-term care firms. The results of each activity organized by the school, the employment situation of the company, and the services provided by the company, these projections allow managers to control costs and generate income.

RESULTS

Data sources, sample sizes, data descriptions, and observation durations are all collected and stored using predictive analytics. Most research employs Electronic Health Records and a database of health and public services in hospitals and medical facilities. Zhang et al. (2012) obtained particular references from the Iowa Hospital (SID) dataset. Rubrichi and Quaglini (2012) conducted text analysis studies at an Italian pharmacy. In both studies, the researchers collected actual data, including the Senior database. Numerous examples and modeling tools have been considered in the literature, and the technology commonly reported in the data mining environment, the same problem in data mining, confirms the results (Lepenioti, Bousdekis, Apostolou & Mentzas, 2020). Although this is the criterion, the project uses four different systems, methods to compare performance, a possible explanation of the modeling trend and helps to select the best model. The company seems to understand a modeling system, Path Clinics is primarily concerned with determining dependent columns, uses different process mining methods to model time and events, correlates with determining the preferred model and nursing path, tool trend of locally encoded software such as R, spss, weka, pro Sites that seem to be distributed.

DISCUSSION

We conduct integrated research using predictive analytics and data mining to deliver healthcare services. The collection, analysis, and periodic interpretation of all relevant empirical evidence on research questions are essential components of good scientific practice in systematic reviews. To analyze the collected data, we use systematic research to collect data and the types of analysis suggested by Chering and Gold (2012). This section reflects Sense and Kofteros (2015), highlighting gaps in knowledge system change and suggesting new recommendations for future research and health practices. General concepts include all the steps necessary to provide patient care, but: Anal data is only a small part of the monitoring process, such as quality of care, route identification, and planning. Most data mining applications show that you are focused. For patients and staff, these applications are important in keeping up with the increasing demands for patient health and safety functions and help increase the power of large data sets and improve clinical productivity for other business functions, by example, physical design and process analysis. This seems to be the limit of literary criticism. The same applies to content.

Further analysis of the collected data may include a properly designed clinical organization and resource planning systems. In the analysis process, the nursing process focuses on determining and analyzing the flow of common patients and the characterization of the guidelines. James & Savitz (2011) focus on introducing others who discovered that while there is a large routing cost structure, these concepts are associated with business processes and their design. It can be rehabilitated and used for better health results (Lepenioti, Bousdekis, Apostolou & Mentzas, 2020). One possible explanation for this apparent limitation is that the healthcare provider route restricts the use of evidence-based methods to shape the entire healthcare network, and the data collected is transparent because it does not improve health but is accurately integrate with the health information system.

Future models aim to expand the model through a holistic care delivery process, which allows patients to use captured actions to make decisions (Lepenioti, Bousdekis, Apostolou & Mentzas, 2020). This appears to be a promising area of health research. This study looked at data management, modeling, and a refined body of knowledge in addition to data processing in the healthcare business. The modeling trend is to construct prediction models using basic classifiers (support vector machines (SVM) and neural networks). However, random forest collisions and multidimensional adaptive regression outperform, according to some studies (Demir, 2014).The process of improving model performance can be evaluated internally, compared to actual computer behavior or other modeling methods, for example, by comparing different programming models with external models, including studies involving internal evaluation, costs of patient management, and innovations based on evaluation tools during this process (Pandey, Nepal & Chen, 2011). Treatment period Improving and externally implementing other similar scenarios will increase the generalization of the growth model and finally, as mentioned above, two-thirds of the investigations have not yet been carried out, limiting the benefits of using large data sets, such as medical care, concept implementation, and predictive analytics.

CONCLUSION

The information provided in this document helps explain how data processing and predictive analytics can contribute to healthcare delivery. Although the literature indicates that data processing is a new energy sector in healthcare, it is a specific component of healthcare. The document focuses on the nature of clinical courses, quality of care, and resource allocation, but research in this area focuses on the use of great tools to improve operational efficiency, team design, analysis of processes, and patient satisfaction. The first module provides the architecture and concept of the current architecture, and the second module provides the most representative functions/processes related to: providing steps to test data mining applications. Third, this research uses a unique method of inductive integration, which is divided into two parts, one of which integrates and classifies field research in specific fields at various levels. In addition, the final framework describes the three main uses of data processing in healthcare: identifying medical resources, improving capacity planning, improving quality of care, ensuring optimal design, design analysis, and effectiveness of clinical care for health professionals and important functional meanings Complete data can only be used by referral agencies, and reflective data researchers and experts can use reflective models and relationships in data management, modeling, and evaluation. In particular, the implementation can be used as another research platform.

REFERENCES

1. Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., & Escobar, G. (2014). Big data in health care: Using analytics to identify and manage high-risk and high-cost patients. *Health Affairs*, 33, 1123-1131.
2. Cao Yiding. (2013). Research on data mining algorithm based on granular computing. (Doctoral dissertation, Xidian University).

3. Demir, E. (2014). A decision support tool for predicting patients at risk of readmission: A comparison of classification trees, logistic regression, generalized additive models, and multivariate adaptive regression splines. *Decision Sciences*, 45, 849-880.
4. Eckerson, W. W. (2007). Predictive analytics. Extending the Value of Your Data Warehousing Investment. TDWI Best Practices Report, 1, 1-36.
5. Hazen, B. T., Boone, C. A., Ezell, J. D., & Jones-Farmer, L. A. (2014). Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *International Journal of Production Economics*, 154, 72-80
6. Huang Wen. Research on data mining algorithms and their applications. (Doctoral dissertation, Nanjing University of Posts and Telecommunications).
7. James, B. C., & Savitz, L. A. (2011). How Intermountain trimmed health care costs through robust quality improvement efforts. *Health Affairs*, 30, 1185-1191
8. Lepenioti, K., Bousdekis, A., Apostolou, D., & Mentzas, G. (2020). Prescriptive analytics: Literature review and research challenges. *International Journal of Information Management*, 50, 57-70.
9. Mahmoud, F. Z. M. (2017). The application of predictive analytics: Benefits, challenges and how it can be improved. *International Journal of Scientific and Research Publications*, 7(5), 549-566.
10. Pandey, S., Nepal, S., & Chen, S. (2011, October). A testbed for the evaluation of business process prediction techniques. In *7th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom)* (pp. 382-391). IEEE.
11. Reddy, A. R., & Kumar, P. S. (2016, February). Predictive big data analytics in healthcare. In *2016 Second International Conference on Computational Intelligence & Communication Technology (CICT)* (pp. 623-626). IEEE.
12. Rubrichi, S., & Quaglini, S. (2012). Summary of Product Characteristics content extraction for a safe drugs usage. *Journal of Biomedical Informatics*, 45, 231-239
13. Shi, Libao, Shen, & Li. (2013). Implementation of an intelligent grid computing architecture for transient stability constrained ttc evaluation. *Journal of Electrical Engineering & Technology*, 8(1), 20-30.
14. Sagioglu, S. , & Sinanc, D. . (2013). Big data: A review. *2013 International Conference on Collaboration Technologies and Systems (CTS)*. IEEE.
15. Spruit, M., Vroon, R., & Batenburg, R. (2014). Towards healthcare business intelligence in long-term care: An explorative case study in the Netherlands. *Computers in Human Behavior*, 30, 698-707.
16. Wang, G., Gunasekaran, A., Ngai, E. W., & Papadopoulos, T. (2016). Big data analytics in logistics and supply chain management: Certain investigations for research and applications. *International Journal of Production Economics*, 176, 98-110.
17. Wei Wei, Wu Xiaowei, Lü Fan, Xiao Yunfeng, Fu Shaojun, & Pei Yuanji et al. (2009). Dry etching of SiO₂/Si by Xef₂. *Journal of University of Science and Technology of China* (6), 5.
18. Wu Jun. (2009). Application and research of data mining based on granular computing. (Doctoral dissertation, Wuhan University of Technology).
19. Zhang Zhaogong, Li Jianzhong, & Zhang Yanqiu. (2004). Parallel algorithm for mining association rules on massive data. *Journal of Harbin Institute of Technology*, 36(5), 5.
20. Zheng, B., Zhang, J., Yoon, S. W., Lam, S. S., Khasawneh, M., & Poranki, S. (2015). Predictive modeling of hospital readmissions using metaheuristics and data mining. *Expert Systems with Applications*, 42, 7110-7120.

21. Zhu Ming. (2008). *Data Mining - 2nd Edition*. University of Science and Technology of China Press.
22. Zhang Zhanjie. (2011). A brief discussion on the techniques of massive data processing. *Science and Technology Communication* (2), 2.
23. S. Qazi, M. Usman, and A. Mahmood, "A data-driven framework for introducing predictive analytics into expanded program on immunization in Pakistan," *Wiener Klinische Wochenschrift*, vol. 133, no. 13-14, pp. 695-702, 2021, doi: 10.1007/s00508-020-01737-3.
24. S. Ayesha, M. K. Hanif, and R. Talib, "Performance enhancement of predictive analytics for health informatics using dimensionality reduction techniques and fusion frameworks," *IEEE Access*, vol. 10, pp. 753-769, 2022, doi: 10.1109/ACCESS.2021.3139123.
25. B. K. Reddy, D. Delen, and R. K. Agrawal, "Predicting and explaining inflammation in Crohn's disease patients using predictive analytics methods and electronic medical record data," *Health Informatics Journal*, vol. 25, no. 4, pp. 1201-1218, 2019, doi: 10.1177/1460458217751015.
26. N. Sghir, A. Adadi, and M. Lahmer, "Recent advances in predictive learning analytics: a decade systematic review (2012-2022)," *Education and Information Technologies*, vol. 28, no. 7, pp. 8299-8333, 2023, doi: 10.1007/s10639-022-11536-0.
27. S. Gocheva-Ilieva and A. Ivanov, "Assaying SARIMA and generalised regularised regression for particulate matter PM10 modelling and forecasting," *International Journal of Environment and Pollution*, vol. 66, no. 1-3, pp. 41-62, 2019, doi: 10.1504/IJEP.2019.104520.
28. J. Linghu, J. Chen, and Z. Yan, "Research on forecasting coal bed methane demand and resource allocation system based on time series," *Energy Exploration and Exploitation*, vol. 38, no. 5, pp. 1467-1483, 2020, doi: 10.1177/0144598720953505.
29. M. Borowski, P. Życzkowski, K. Zwolińska, R. Łuczak, and Z. Kuczera, "The security of energy supply from internal combustion engines using coal mine methane—forecasting of the electrical energy generation," *Energies*, vol. 14, no. 11, 2021, doi: 10.3390/en14113049.
30. S. Piri, D. Delen, and T. Liu, "A synthetic informative minority over-sampling (SIMO) algorithm leveraging support vector machine to enhance learning from imbalanced datasets," *Decision Support Systems*, vol. 106, pp. 15-29, 2018, doi: 10.1016/j.dss.2017.11.006.
31. C. M. Olszak and E. Ziemia, "Business intelligence systems in the holistic infrastructure development supporting decision-making in organisations," *Interdisciplinary Journal of Information, Knowledge, and Management*, vol. 1, pp. 47-58, 2006, doi: 10.28945/3011.
32. M. Mandorino, A. J. Figueiredo, G. Cima, and A. Tessitore, "Predictive analytic techniques to identify hidden relationships between training load, fatigue and muscle strains in young soccer players," *Sports*, vol. 10, no. 1, 2022, doi: 10.3390/sports10010003.
33. Y. Cui, F. Chen, A. Shiri, and Y. Fan, "Predictive analytic models of student success in higher education: a review of methodology," *Information and Learning Science*, vol. 120, no. 3-4, pp. 208-227, 2019, doi: 10.1108/ILS-10-2018-0104.
34. M. F. M. Marçal, Z. M. de Souza, R. L. M. Tavares, C. V. V Farhate, S. R. M. Oliveira, and F. S. Galindo, "Predictive models to estimate carbon stocks in agroforestry systems," *Forests*, vol. 12, no. 9, 2021, doi: 10.3390/f12091240.
35. A. Saravanou, C. Noelke, N. Huntington, D. Acevedo-Garcia, and D. Gunopulos, "Predictive modeling of infant mortality," *Data Mining and Knowledge Discovery*, vol. 35, no. 4, pp. 1785-1807, 2021, doi: 10.1007/s10618-020-00728-2.
36. S. Asaduzzaman, M. R. Ahmed, H. Rehana, S. Chakraborty, M. S. Islam, and T. Bhuiyan, "Machine learning to reveal an astute risk predictive framework for Gynecologic Cancer and its impact on women psychology:

Bangladeshi perspective,” *BMC Bioinformatics*, vol. 22, no. 1, 2021, doi: 10.1186/s12859-021-04131-6.

37. Y. Pang, S. Gong, Q. Liu, H. Wang, and J. Lou, “Overlying strata fracture and instability process and support loading prediction in deep working face,” *Caikuang yu Anquan Gongcheng Xuebao/Journal of Mining and Safety Engineering*, vol. 38, no. 2, pp. 304-316, 2021, doi: 10.13545/j.cnki.jmse.2019.0585.

38. M.-J. Liu, W. Yue, L.-Z. Qiu, J.-X. Li, and Z.-G. Qin, “Research progress of real-time bidding for display advertising,” *Jisuanji Xuebao/Chinese Journal of Computers*, vol. 43, no. 10, pp. 1810-1841, 2020, doi: 10.11897/SP.J.1016.2020.01810.

39. A. M. Koli and M. Ahmed, “Machine learning based parametric estimation approach for poll prediction,” *Recent Advances in Computer Science and Communications*, vol. 14, no. 4, pp. 1287-1299, 2021, doi: 10.2174/2666255813666191204112601.

40. A. A. Alharbi, I. Petrunin, and D. Panagiotakopoulos, “Modeling and Characterization of Traffic Flow Patterns and Identification of Airspace Density for UTM Application,” *IEEE Access*, vol. 10, pp. 130110-130134, 2022, doi: 10.1109/ACCESS.2022.3228828.

41. O. Illiashenko, V. Mygal, G. Mygal, and O. Protasenko, “A convergent approach to the viability of the dynamical systems: the cognitive value of complexity,” *International Journal of Safety and Security Engineering*, vol. 11, no. 6, pp. 713-719, 2021, doi: 10.18280/ijssse.110612.

42. P. Kamal and S. Ahuja, “An ensemble-based model for prediction of academic performance of students in undergrad professional course,” *Journal of Engineering, Design and Technology*, vol. 17, no. 4, pp. 769-781, 2019, doi: 10.1108/JEDT-11-2018-0204.

43. J. Linghu, J. Chen, Z. Yan, and C. Yao, “Demand forecast and allocation system of coalbed methane of different grades in mining area,” *Energy Sources, Part A: Recovery, Utilization and Environmental Effects*, 2020, doi: 10.1080/15567036.2020.1859017.

44. S. S. R. Moustafa, M. S. Abdalzaher, M. H. Yassien, T. Wang, M. Elwekeil, and H. E. A. Hafiez, “Development of an optimized regression model to predict blast-driven ground vibrations,” *IEEE Access*, vol. 9, pp. 31826-31841, 2021, doi: 10.1109/ACCESS.2021.3059018.

45. E. Hou, Q. Wen, Z. Ye, W. Chen, and J. Wei, “Height prediction of water-flowing fracture zone with a genetic-algorithm support-vector-machine method,” *International Journal of Coal Science and Technology*, vol. 7, no. 4, pp. 740-751, 2020, doi: 10.1007/s40789-020-00363-8.

46. M. Momenzadeh, M. Sehhati, and H. Rabbani, “Using hidden Markov model to predict recurrence of breast cancer based on sequential patterns in gene expression profiles,” *Journal of Biomedical Informatics*, vol. 111, 2020, doi: 10.1016/j.jbi.2020.103570.

47. S. Ali and N. Bouguila, “A roadmap to hidden markov models and a review of its application in occupancy estimation,” in *Hidden Markov Models and Applications*, N. Bouguila, W. Fan, and M. Amayri, Eds. Cham: Springer International Publishing, 2022, pp. 1-31.

48. A. Shillabeer, “An automated data pattern translation process for medical data mining.,” *Medinfo. MEDINFO*, vol. 12, no. Pt 1, pp. 586-590, 2007.

49. A. Z. Woldaregay et al., “Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes,” *Artificial Intelligence in Medicine*, vol. 98, pp. 109-134, 2019, doi: 10.1016/j.artmed.2019.07.007.

50. M. J. Flores, A. E. Nicholson, A. Brunskill, K. B. Korb, and S. Mascaro, “Incorporating expert knowledge when learning Bayesian network structure: a medical case study,” *Artificial Intelligence in Medicine*, vol. 53, no. 3, pp. 181-204, 2011, doi: 10.1016/j.artmed.2011.08.004.

51. X. Liu, “Design of enterprise economic information management system based on big data integration algorithm,” *Journal of Mathematics*, vol. 2022, 2022, doi: 10.1155/2022/3257748.

52. Y. Jiang, Y. Ye, H. Zhao, S. Zhang, Y. Cao, and J. Gu, "Analysis of smart water conservancy |," *Shuili Xuebao/ Journal of Hydraulic Engineering*, vol. 52, no. 11, pp. 1355-1368, 2021, doi: 10.13243/j.cnki.slx.20210633.
53. C. Laiton-Bonadiez, J. W. Branch-Bedoya, J. Zapata-Cortes, E. Paipa-Sanabria, and M. Arango-Serna, "Industry 4.0 technologies applied to the rail transportation industry: a systematic review," *Sensors*, vol. 22, no. 7, 2022, doi: 10.3390/s22072491.
54. Z. Liu, N. Li, L. Wang, J. Zhu, and F. Qin, "A multi-angle comprehensive solution based on deep learning to extract cultivated land information from high-resolution remote sensing images," *Ecological Indicators*, vol. 141, 2022, doi: 10.1016/j.ecolind.2022.108961.
55. S. Jonnalagadda, T. Cohen, S. Wu, and G. Gonzalez, "Enhancing clinical concept extraction with distributional semantics," *Journal of Biomedical Informatics*, vol. 45, no. 1, pp. 129-140, 2012, doi: 10.1016/j.jbi.2011.10.007.
56. K. B. Kashani, "Automated acute kidney injury alerts," *Kidney International*, vol. 94, no. 3, pp. 484-490, 2018, doi: 10.1016/j.kint.2018.02.014.
57. T. I. Oprea, O. Taboureau, and C. G. Bologa, "Of possible cheminformatics futures," *Journal of Computer-Aided Molecular Design*, vol. 26, no. 1, pp. 107-112, 2012, doi: 10.1007/s10822-011-9535-9.
58. B. Zheng, J. Zhang, S. W. Yoon, S. S. Lam, M. Khasawneh, and S. Poranki, "Predictive modeling of hospital readmissions using metaheuristics and data mining," *Expert Systems with Applications*, vol. 42, no. 20, pp. 7110-7120, Nov. 2015, doi: 10.1016/J.ESWA.2015.04.066.
59. S. Rubrichi and S. Quaglini, "Summary of product characteristics content extraction for a safe drugs usage," *Journal of Biomedical Informatics*, vol. 45, no. 2, pp. 231-239, Apr. 2012, doi: 10.1016/J.JBI.2011.10.012.
60. K. Lepenioti, A. Bousdekis, D. Apostolou, and G. Mentzas, "Prescriptive analytics: Literature review and research challenges," *International Journal of Information Management*, vol. 50, pp. 57-70, Feb. 2020, doi: 10.1016/J.IJINFOMGT.2019.04.003.
61. A. Kofteros, A. Kofteros, and T. Hadzilacos, "Adapt and they shall come: aspects of online teacher-parent collaboration in ...," *Journal of Interactive Learning research*, vol. 30, no. 3, pp. 347-363, 2019.
62. E. Demir, T. Bektaş, and G. Laporte, "A review of recent research on green road freight transportation," *European Journal of Operational Research*, vol. 237, no. 3, pp. 775-793, Sep. 2014, doi: 10.1016/J.EJOR.2013.12.033.
63. X. Zhang, C. Liu, S. Nepal, S. Pandey, and J. Chen, "A privacy leakage upper bound constraint-based approach for cost-effective privacy preserving of intermediate data sets in cloud," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 6, pp. 1192-1202, 2013, doi: 10.1109/TPDS.2012.238.

CONFLICT OF INTEREST

None.

FINANCING

This study was conducted independently without any external funding. The authors declare that no financial support was received from any organization for the submitted work.

DATA AVAILABILITY

Manuscript has no associated data.

USE OF ARTIFICIAL INTELLIGENCE

The authors have used artificial intelligence technologies within acceptable limits to provide their own verified data, which is described in the research methodology section.

ACKNOWLEDGMENTS

The authors wish to express their gratitude to Dr. Latafat Gardashova, Head of Department of Doctoral

Studies at Azerbaijan Oil and Industry University for her invaluable feedback during the article preparation.

AUTHORSHIP CONTRIBUTION

Conceptualization: Bahar Asgarova, Elvin Jafarov, Nicat Babayev, Allahshukur Ahmadzada.

Data curation: Bahar Asgarova, Elvin Jafarov, Nicat Babayev, Allahshukur Ahmadzada.

Formal analysis: Bahar Asgarova, Elvin Jafarov, Nicat Babayev, Allahshukur Ahmadzada.

Drafting - original draft: Bahar Asgarova, Elvin Jafarov, Nicat Babayev, Allahshukur Ahmadzada.

Writing - proofreading and editing: Bahar Asgarova, Elvin Jafarov, Nicat Babayev, Allahshukur Ahmadzada.