Indonesian Journal of Electrical Engineering and Computer Science
Vol. 36, No. 1, October 2024, pp. 703~714
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v36.i1.pp703-714

Development process of decision support systems using data mining technology

Bahar Asgarova[1], Elvin Jafarov[2], Nicat Babayev[2], Allahshukur Ahmadzada[2], Vugar Abdullayev[1], Triwiyanto[3]
[1]Department of Computer Engineering, Azerbaijan State Oil and Industry University, Baku, Azerbaijan
[2]Azerbaijan State Oil and Industry University, Baku, Azerbaijan
[3]Department of Medical Electronics Technology, Health Polytechnic Ministry of Health Surabaya, Surabaya, Indonesia

Article Info

Article history:

Received Mar 17, 2024
Revised May 12, 2024
Accepted Jun 5, 2024

Keywords:

Data mining
Database
Decision support system
ETL
Machine learning

ABSTRACT

Decision support systems (DSS) play a pivotal role as computerized tools, guiding and enhancing decision-making processes vital for organizational progress. This research focuses on developing a system tailored for dynamic decision-making, particularly emphasizing the integration of data mining technology. Decision algorithms and neural networks are discussed in depth, providing a comprehensive understanding of the analytical tools crucial for effective decision support. Additionally, the research sheds light on potential risks, ensuring a nuanced view of challenges that may impact the development of DSS. A significant portion of the study is dedicated to the design of DSS architecture and the strategic integration of data mining within the database. The proposed development stages for a business information system, ranging from feasibility study to release, serve as a structured framework for practical implementation. Details within each stage, including data analysis, cleaning, and module development, are meticulously examined. Emphasis is placed on critical steps such as system design, database design, and extract, transform, load (ETL) process design, elucidating their importance in the holistic development of DSS. The conclusion reinforces the paramount importance of leveraging data mining technology in the process of developing decision support systems.

Corresponding Author:

Vugar Abdullayev
Department of Computer Engineering, Azerbaijan State Oil and Industry University
Baku, Azerbaijan
Email: abdulvugar@mail.ru

1. INTRODUCTION

The notion of decision support system (DSS) has been around for a long time, and as a concept, it should be emphasized that it evolves with the advancement of information technology [1]–[5]. Data mining technologies started to evolve by the end of 20th century as development of database information systems created base for it. First DSS systems were rule based and evaluated results according to predefined rules by experts but today with the advance of machine learning, artificial intelligence and predictive analysis they act more independently and are more complex [6]–[8]. Advanced use of machine learning, artificial intelligence helps DSS to provide more accurate analysis, advices and enlarges it is possible use in many different industries.

In healthcare DSS applied for analyzing the patient symptoms, data to diagnose the illness and plan required treatments, in finance it can help with budget planning, risk analysis and fraud expose [9]–[12]. Another use of it could be for consumer analysis in retail market for different purposes like consumer segmentation, consumer behavior analysis and targeted marketing which can significantly improve the sales and customer satisfaction [13]–[17]. Despite the advancement of DSS systems there are still some challenges and limitations like even artificial intelligence has significantly evolved we still can fully rely on its accuracy which requires further verification of results, another point of concern is data quality for accurate analysis DSS systems requires quality data, it is also not easy to maintain the complex algorithms and data privacy, security [18]–[20].

Many investigations show that the inventors Moore and Chang characterized this system as a "scalable system." "In scheduled and unforeseen moments, this system's analytical and decision modeling capacity drew our focus positively toward the future," they said of their findings [21], [22]. Carlson and Sprague, on the other hand, took a different approach and proposed the notion of decision support systems: "structured and interactive systems that assist producers in addressing semi-structured economic issues utilizing data and models." Turban, the DSS developer, characterizes it as "an interactive, versatile, and adaptive system" in 1998, and it is regarded as a procedure for addressing management challenges. The system uses data (internal and external) and models to give a simple and easy-to-use interface. As a result, it has power over the decision-making process of the decision-maker. DSS gives assistance at all phases of the decision-making process [14], [23], [24]. According to the findings of this study, the process of creating a DSS began with a notion of how the DSS's objectives may be met. A DSS's components include the end-view users of the qualities and what such a system can perform (to support the decision-making process, to solve structured and unstructured problems). Several contributions of this study are as follows:

- This study presents a systematic approach to developing decision support systems that integrate data mining techniques, and discusses the challenges and opportunities of applying data mining in various domains and applications.
- This study proposes a flexible architecture for decision support systems that combines data storage, OLAP, data mining, and business intelligence tools, and illustrates the development stages and methods of data mining integration within the database.
- This study explores the evolution and current state of decision support systems and data mining, and provides a comprehensive overview of the data mining methods, algorithms, and tools used for decision support, as well as their benefits and limitations.

2. LITERATURE REVIEW

In data mining, decision algorithms describe a form of tree view, also known as connectivity in a DSS Figure 1 [25]–[27]. The mechanism of creating trees is to gather all the variables that the analyst perceives and plays a major role in making decisions [28]. It analyses the results in terms of impact to evaluate them. As an application, decision algorithms determine which variables are most important based on sorting data. Decision algorithms in data mining are applied in areas such as sorting of loan requests in business areas, ranking of applicants for different positions [29]–[31].

Many institutions and companies achieve high results in the development process by enriching their databases with valuable information (mainly about customers), because this method helps them to identify the needs of customers [16], [32], [33]. However, the effective extraction of information from this data and the discovery of hidden patterns requires the use of machine learning algorithms. The decision tree algorithm is most used for purposes such as classification and probability. As for the areas of application, it is used in areas such as customer relationship management, fraud detection in bank accounts, energy consumption, and medicine [25], [34]–[36]. For example, energy consumption is about how much electricity people use [29], [37]. Research in this direction helps companies determine the amount of energy they need at the right time. The hierarchical structure of the decision tree method ensures a clear and accurate display of information visually. It is possible to build this algorithm by considering characteristics such as average temperature, humidity, and pressure to predict whether the weather will be foggy, rainy or stormy.

Neural networks are the most used technique in data mining [30], [38], [39]. This method is placed in a system connected through of arc nodes called observation group. The idea is analogical to how neurons work in human brain. Neural networks have a stable complex structure that reflects at least three layers of nonlinear connections. Each input data to a node in the first layer and output data to the last layer places and reflects the result. Last layer of the neural networks has nodes for every category which has outputs. This layer is used to classify the neural networks. Usually, these neural networks keep hidden intermediate nodes which make the model more complex. The difference between obtained and target results are gathered, adjustments are made to the node's price and re-entered into the system. The process stores the complete cycle in memory till the inputted data are sort correctly by network.

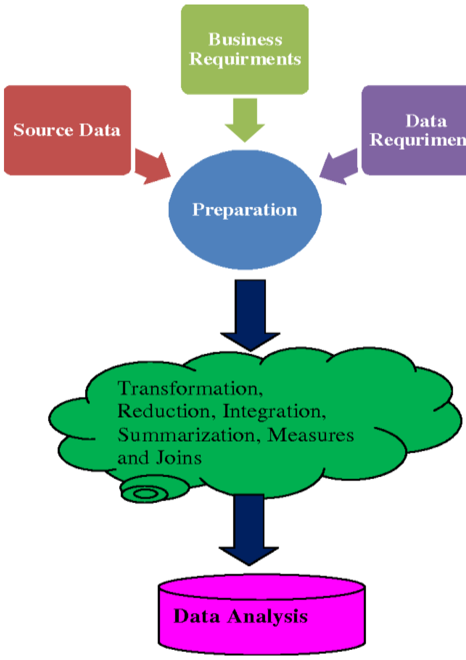

Figure 1. Data mining in DSS

The below stated risks can affect the efficiency and success rate of the development of DSS, high returns, human resource. Reason why DSS are developed is to help managers in making decisions during budgeting and financial planning. It is especially useful for the investors who are in public funds [30], [32]. Currently, many companies are investing heavily in building a database. The analyst intends to improve the efficiency and performance of reporting activities. So, there are expensive programs with the ability to analyze proposals, some future plans and predict the evolution of work. Some of these can be analyzed statistically or using neural networks. Based on the research, in order to effectively build a DSS, several ways and methods that allow improving the accuracy of the analysis should be combined, which should be approached from two main perspectives - accurate data and predictions. In order to realize this requirement, it is necessary to have a flexible architecture that combines data storage, online analytical processing (OLAP), data mining and business intelligence tools, where reports can be included: At data model level, an extract, transform, load (ETL) process should be applied to clean and load data into the database. At the application level, OLAP and data recognition methods are used to apply analytical models which is combined for predictive and historical analysis. At interface, based on the assessment of business opportunities, reports and tables are applied [16], [29], [40]. This article will discuss the design of DSS architecture and describe the ways and methods of data mining integration within the database. In this research, we will outline several processes for developing a business information system, including project planning, feasibility study, development, analysis, and release. While these stages can be modified and used to decision support systems, they must be handled individually to apply the distinctions between general system modelling and decision support systems modelling during development [41], [42].

Step 1. The goals of a feasibility study are to create better the decision-making process by defining the needs and business potential. The predicted costs and benefits for each of the suggested solutions should be used to support them.

Step 2. Planning a project entail determining future needs, current infrastructure components, and prospects for sustainability. The project plan serves as the final output of this activity. The project can begin in earnest after receiving approval.

Step 3. Prioritizing and in-depth examination of the initial organizational management team requirements are done during the job opportunity analysis stage. Typically, managers and project personnel undertake interviews to determine the needs. These specifications could result in modest project changes, but they should make the development team leaders aware of DSS's potential and constraints. This in turn lowers the possibility of business requirements being unfeasible [42].

The most significant part of a decision support system development project is data analysis, where it is decided what data is required, what it contains, and how it links to other data. Compared to systems analysis carried out using conventional approaches, data analysis is more closely tied to business analysis.

Data cleansing is done first [14], [39]. Data cleansing involves filtering and transforming sources for developing an analysis module. This process occurs as follows [43]. Data cleansing is a crucial step in preparing data for analysis. It involves filtering and transforming data from various sources. First, we identify the essential information required from functional modules. Next, we thoroughly analyze the content of the selected information sources. We choose relevant data that aligns with the project's objectives. Requirements related to data filtration are applied to remove noise, inconsistencies, and irrelevant data. Finally, we select the tools and techniques needed for the cleaning and filtration process. In summary, data cleansing ensures that our analysis module works with accurate and reliable data, enhancing the quality of subsequent analyses.

Several key aspects should be considered during the source selection process: data integrity, sensitivity, data format and accuracy. The processes are vital for the successful ETL procedure. Analysis of big data is a crucial activity where every defined requirement will change based on the structure and variation of big data and is stored in the big data dictionary. A big data dictionary offers contextual information about the project's data. The system analysis phase can be completed by creating a prototype that will be submitted to managers and project employees for functional attributes evaluation. Rapid development tools are readily available, enabling the development of new interfaces based on the analytical model.

Selecting the technologies to be employed in the creation of the prototype the system will create in the end is a crucial step at this point. Using a comparative evaluation of each technology's benefits and drawbacks, different approaches should be considered for the project: the use of a database, the inclusion of OLAP functions, the use of data extraction algorithms, data source integration tools or, at the last stage and considering that there is a parallel approach to the construction of the system considering the use of application integration tools [24], [44].

Step 4. System architecture. database architecture. The required information will be kept in accordance with the system's needs. It means they will be kept both at the low level and at the general level. So, it is important to build a coherent, objective or multidimensional design. In this section, an elegant and detailed new system should be developed to meet the managers' expectations for reporting and analysis based on a logical data model.

3. MATERIALS AND METHODS

In data analysis operation, the process is focused on the sources of data (input data or output data) from the modules. The objectives or data are focused on reports, analyses, and surveys at the current stage. So, the following list of best practices ought to be considered. Due to the above aspects, a centralized database aimed at solving management and data processing is maintained at the organization stage. Using logical and physical standards, the database is divided into pieces of information at a uniform level and is facilitated to be maintained, developed with a different team using the exact characteristics [45].

The design of the ETL in Figure 2 activity it is the most complicated part of life cycle of project and mainly depends on the quality of the data. We provide the creation of an ETL process and the integration of target databases into a single environment. In this particular case, the risk of contrasting data is decreased by avoiding the separation of each target module. In same environment, there is also a strategy for building data, but these are already enriched. What is important here is the fact that ETL process should be the exact same between all layers (one coherent principle) [40].

The architecture of the ETL process involves several essential prerequisites. First, we perform preliminary data source processing to ensure data quality and consistency. Next, adherence to a specific format is crucial for seamless data integration. Additionally, data reconciliation ensures alignment between different data sets. Lastly, we focus on data exclusion and the elimination of deviations to maintain accuracy and reliability throughout the ETL pipeline. The following actions are conducted when developing an ETL process:

a. Mapping of sources with characteristics of change connected to specific directions. This can be drawn as a matrix or variation chart [46].
b. Selection and testing will be used in ETL tools. Currently, there are several types of ETL process modelling and application tools, but selecting one depends based on the features they offer and their support for integrating data sources throughout the same modification process.
c. ETL procedure development - depending on the data model, different data extraction and transmission operators are utilized (sorts, concatenation, joins, and allocations). This procedure is divided into distinct separate subprocesses that will run to reduce processing time. Processing time will be modelled with timing diagrams.
d. Four steps of data loading are used, depending on the program from which the data is loaded in an ETL procedure [46]:

– The initial load refers to the initial loading of destinations with operational data.
– History load refers preliminary loading of locations with historical data that has been archived.
– Growth load refers to the routine loading of destinations with the most recent information from operating systems.
– Deciding on a management environment for the ETL process, which is partitioned and centralized utilizing a dedicated server. The choice is made based on the resources at hand, the processing time, and the anticipated completion time of the process.

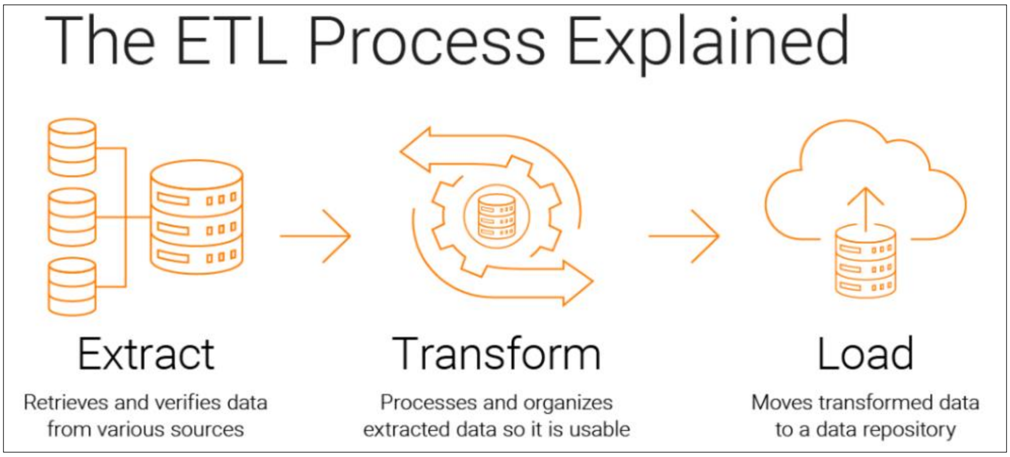

Figure 2. Extract/transform/load (ETL)

The outcome of the activities is implemented in data map documents, ETL process diagrams, data transformation and execution of these processes. Big data storage design when the choice is to build a dedicated storage location, a logical big data model will be implemented for the new system based on the data storage options and a relational, objective, or multidimensional model will be applied. This sub-step may have minor changes depending on the requirements defined in the big data analysis sub-step. When the option is to build dedicated storage, a logical and physical model of big data will apply [46].

e. System initial settings up. Currently, the technologies utilized to create decision support systems constitute a unique subset of the category of intelligent technologies and include technologies for organizing database data. Include ETL, OLAP analysis systems, data mining algorithms, CASE (Computer Assisted Software Engineering) modelling processes, and web systems solutions. (Power D.J., http://dssresources.com/dssbook/).
f. Launching the systems to the marketplace. Manager and business training sessions are held at this phase, along with the implementation of data loading procedures, the installation of applications, and performance monitoring. At the end of the stage, the production history of the system is reviewed, and it ends with signing of the finished documentation. Instructions for users are provided for its application.

One of the most prevalent data mining approaches is association rule analysis. It is used to construct a system model from which information may be retrieved by detecting patterns between connected things [47]. Examples of application of association rules can be the following tasks:
– identification of sets of services that are often ordered together or never ordered together.
– determining the proportion of customers who are positive about changes in tariff plans.
– determination of the profile of electricity consumers; determination of the proportion of cases in which the client's debt leads to the creditor's losses.

An association rule consists of two subject sets, called a condition and a consequence, written as X → Y, which reads as follows: “From X follows Y.” As a result, the associative rule is written as follows: “If <condition>, then <consequence>”. Rules are usually displayed with arrows pointing from condition to consequence, for example, service1 → service3. The condition and consequence are sometimes called the right-hand and left-hand components of an association rule, respectively [47]. Association rules have characteristics that reflect the features of the connection between the condition and the consequence. Basically, they operate with two characteristics-support (written as supp or S) and conformity (written as conf or C). The ratio of the number of transactions having both the condition and the consequence to the total number of transactions is known as support in (1).

$$S = (A \rightarrow B) = P(A \cup B) = \frac{the\ quantity\ of\ transactions\ containing\ A\ and\ B}{the\ total\ amount\ of\ transactions} \quad (1)$$

Conformity is a measure of an association rule's correctness; it is defined as the ratio of transactions having a condition and a consequence to transactions simply containing a condition in (2):

$$C = (A \rightarrow B) = P(A/B) \ = P(A \cup B)/P(A) = \frac{the\ quantity\ of\ transactions\ containing\ A\ and\ B}{the\ quantity\ of\ transactions\ containing\ only\ A} \quad (2)$$

Analytical conclusions can be drawn by calculating the properties of the link. If there is enough support and conformity, it may be claimed that each future transaction that includes a condition will likewise include a consequence. Consider the following example. Take, for example, the association power supply → mobile notification. Given that the number of transactions having both power supply and mobile notification is 4, and the total number of transactions is 9, the following association will be supported:

$$S(power\ supply \rightarrow mobile\ notification) = 4/9 = 0.44$$

since the number of transactions containing only power supply (condition) is 4, then the validity of this association will be:

$$C\ (power\ supply \rightarrow mobile\ notification) = 4/4 = 1$$

In other words, all transactions containing power supply also contain a mobile notification, from which we conclude that this association can be considered as a rule. All customers ordering electricity also use the mobile notification service. Interest is the ratio of the frequency of occurrence of a condition in transactions that also contain a consequence to the frequency of occurrence of the consequence. Interest values greater than one indicate that the condition occurs more frequently in transactions that contain a consequence than in others. We can say that interest is a generalized measure of the relationship between two subject sets: for lift values > 1, the relationship is positive, for 1 it is absent, and for values < 1 it is negative. Interest is calculated as follows:

$$L(A \rightarrow B) = C(A \rightarrow B)/S(B) \quad (3)$$

$$S(power\ supply) = 4/9 = 0.44; C(mobile\ notification \rightarrow power\ supply) = 4/7 = 0.57$$

therefore,

$$L(mobile\ notification \rightarrow power\ supply) = 0.57/0.44 = 1.295$$

now consider the association $mobile\ notification \rightarrow internet$.

$$S(Internet) = 6/9 = 0.67; C(mobile\ notification \rightarrow internet) = 4/7 = 0.57$$

Then,

$$L(mobile\ notification \rightarrow internet) = 0.57/0.67 = 0.85$$

A higher value of interest for the first rule shows that the connection of mobile notification has more influence on the purchase of electricity than the Internet. While interest is widely used, it is not always a good measure of the importance of a rule. A rule with less support and more interest may be less significant than an alternative rule with more support and less interest because the latter applies to more transactions. The importance of the link between the condition and the consequence grows as the number of transactions grows. To take this significance into account, the concept of a level is introduced. The level is the difference between the observed frequency with which the condition and effect appear together (that is, the support of the association), and the product of the frequencies of occurrence (supports) of the condition and effect separately (4).

$$T(A \rightarrow B) = S(A \rightarrow B) - S(A)S(B) \quad (4)$$

Let’s consider the associations cost visualization → mobile notification and power supply → mobile notification, which have the same support C = 1, since cost visualization and power supply are always sold together with mobile notification. The interests for these associations will also be the same, since in both associations the support of the consequence S (mobile notification) = 7/9 = 0.77. Then,

$$L(cost\ visualization \rightarrow mobile\ notification) = L(power\ supply \rightarrow mobile\ notification)$$
$$= 1/0.77 = 1.3$$
$$S(cost\ visualization \rightarrow mobile\ notification) = 3/9 = 0.33$$
$$S(cost\ visualization) = 0.33; S(mobile\ notification) = 0.77$$

thus,

$$T(cost\ visualization \rightarrow mobile\ notification) = 0.33 - 0.33\ 0.77 = 0.08$$
$$S(power\ supply \rightarrow mobile\ notification) = 0.44;\ S(power\ supply)$$
$$= 0.44; S(mobile\ notification) = 0.77$$

therefore,

$$T(power\ supply \rightarrow mobile\ notification) = 0.44 - 0.44\ 0.77 = 1$$

If there are k objects in the database of transactions and all associations are binary (that is, they contain one object each in the condition and consequence), then it will be necessary to analyze k2k−1 associations. Because real transactional databases considered in data mining typically contain thousands of records, the computational cost of finding association rules is enormous. To ensure acceptable efficiency, algorithms are widely used in the process of generating association rules to reduce the number of associations that need to be analyzed. One of the most common is a technique based on the detection of so-called frequent sets, when only those associations that occur quite often are analyzed. This technique is called Apriori (Figure 3) [20].

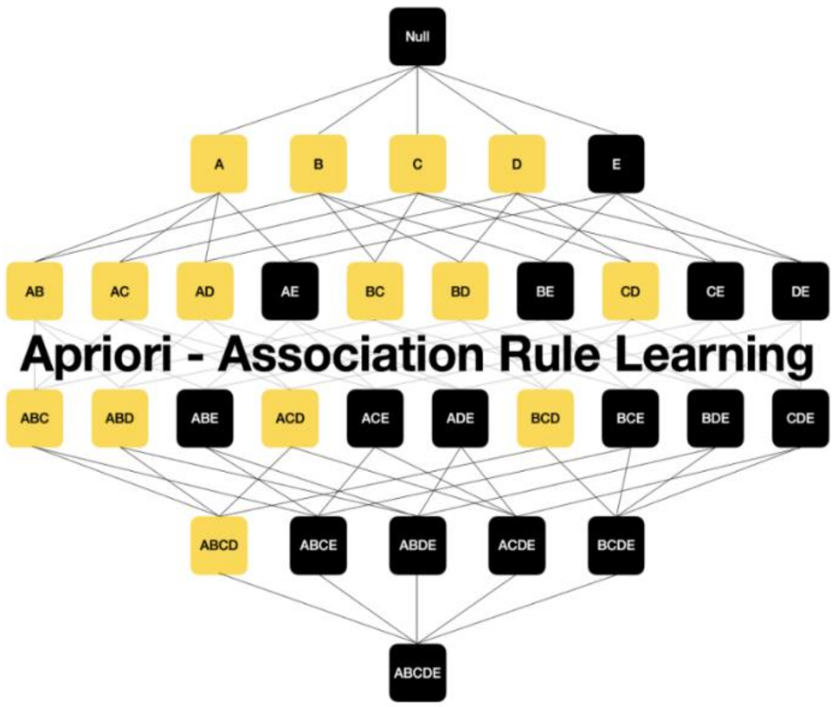

Figure 3. Apriori algorithm for association rule learning

The Apriori algorithm is a powerful technique used in association rule mining [20]. It involves several essential steps. First, data preprocessing occurs, where all data is transformed into binary form, and the data type is adjusted. Next, the algorithm detects frequently occurring sets of elements, with a focus on 1-element frequently occurring sets. Subsequently, item set generation takes place, creating potentially frequent item sets (referred to as candidates) and calculating their supports. Candidate verification follows, ensuring that candidate support values meet a minimum threshold. Finally, the algorithm extracts rules. To do so, it identifies all non-empty subsets of a frequently occurring set F. For each subset M, a rule of the form M → (F – M) is formulated if the rule's conformity (C (M → (F – M))) is not less than the established threshold. In summary, the Apriori algorithm systematically uncovers valuable associations within data, making it a fundamental tool in data mining and decision-making processes. The analysis of association rules is constantly evolving. New techniques and algorithms are emerging. The importance of this data analysis technique for solving problems dealing with subject sets remains unchanged.

4. RESULTS

The research illustrates how DSS have evolved with advances in information technology. Initially rule-based, modern DSS have adopted machine learning and artificial intelligence, which has resulted in increased autonomy and complexity. DSS applications have spread across a wide range of sectors, helping with healthcare diagnostics, financial planning, fraud detection, and retail customer analysis. Despite technical breakthroughs, obstacles remain. Because of concerns about the correctness of growing artificial intelligence models, outcome verification techniques must be implemented. Furthermore, the report emphasizes data quality as a vital aspect in assuring exact analysis, noting issues in maintaining complicated algorithms as well as protecting data privacy and security. The article introduces decision algorithms, which are especially common in the domain of data mining within DSS.

These methods, which are depicted in a tree-view structure, are useful in applications such as loan request sorting. The report does, however, address problems such as technological obsolescence and the importance of maintaining data quality for successful decision-making. Neural networks, which are widely used in the data mining industry, are marketed as complex systems that mimic the functions of the human brain. Their utility extends to areas like client relationship management. Nonetheless, the study emphasizes the dangers of technological obsolescence and the crucial need of preserving data quality for trustworthy results. The research describes a methodical approach to developing DSS. A feasibility study, project planning, and data analysis are all critical stages. Data analysis necessitates critical stages such as data purification, emphasizing the importance of a precise ETL procedure. DSS development relies heavily on centralized databases. The integration of data storage, OLAP, data mining, and business intelligence tools is emphasized in the study. For efficient decision support, big data storage architecture entails developing a logical model and establishing an ETL process, with data quality staying key.

5. DISCUSSION

The research thoroughly examined the growing landscape of DSS and data mining technologies, emphasizing their critical significance in modern information technology. The essay emphasized DSS's transformation from rule-based systems to the present era of machine learning, artificial intelligence, and predictive analytics. DSS application domains were emphasized in particular, spanning from healthcare, banking, and retail markets [48]. Notable uses include healthcare diagnosis and treatment planning, financial budgeting and risk analysis, and retail customer analysis for segmentation and targeted marketing. Despite progress, difficulties like as data quality, algorithm maintenance, and data privacy/security were identified. While artificial intelligence has advanced substantially, the research correctly stated that there is a need for continual verification of outcomes. Furthermore, the need of quality data for proper analysis and the difficulties involved with maintaining complicated algorithms were emphasized.

The paper delves into essential data mining techniques, with a special emphasis on decision algorithms, neural networks, and association rule analysis. Decision algorithms were characterized as a critical component of data mining, used for sorting data and generating judgements based on specified criteria. Neural networks, which simulate the activity of the human brain, have emerged as a dominating approach with applications in disciplines as diverse as customer relationship management and fraud detection [21].

The Apriori method, especially association rule analysis, was investigated as an effective tool for discovering patterns between related elements [20]. The necessity of frequent sets and support values in extracting meaningful rules from huge datasets was emphasized during the debate. The study not only provided light on the current status of DSS and data mining, but it also paved the way for future research areas. The necessity for a flexible architecture that combines data storage, OLAP, data mining, and business intelligence tools was stressed during the conversation. It emphasized the need of addressing issues such as technology obsolescence, data quality, and system architecture in order to establish a successful DSS. The article finished by emphasizing the field's continual progress, with new approaches and algorithms always developing. It acknowledged the continuing relevance of data analysis techniques, particularly in tackling challenges involving topic sets, and hinted at future improvements in the discipline. In conclusion, the complete examination of DSS, data mining techniques, and their applications gave significant insights into the present status of decision support and data analysis technology. The problems mentioned are critical considerations for scholars and practitioners, paving the way for further breakthroughs in this dynamic subject.

6. CONCLUSION

Data mining DSS experience has also revealed that business users want the flexibility to establish their own library of classification rules, association rules, and cluster descriptions. Users should also be allowed to add their own comments to the archived rules in order to capture concepts implied or derived from the rules. The categorization algorithm will also be applied to new customers in the future to anticipate the category to which a new client may belong. These enhancements are scheduled to be implemented in the future. Association rule search techniques are designed to find all associations that satisfy the support and validity constraints imposed by the analyst. When analyzing big data, this leads to the need to consider tens and hundreds of thousands of associations. Processing so many rules manually is unrealistic. The number of rules needs to be reduced. However, support and conformity characteristics are not enough to narrow down the associations that are important to the system. It is required to provide filtering rules by significance.

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.

REFERENCES

[1] J. Liu, “The impact of data mining on management and digital marketing in the age of big data,” *Computer-Aided Design and Applications*, vol. 21, no. S4, pp. 229–247, 2024, doi: 10.14733/cadaps.2024.S4.229-247.

[2] S. Li and F. Brennan, “Digital twin enabled structural integrity management: critical review and framework development,” *Proceedings of the Institution of Mechanical Engineers Part M: Journal of Engineering for the Maritime Environment*, 2024, doi: 10.1177/14750902241227254.

[3] S. Wang, W. Han, X. Huang, X. Zhang, L. Wang, and J. Li, “Trustworthy remote sensing interpretation: concepts, technologies, and applications,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 209, pp. 150–172, 2024, doi: 10.1016/j.isprsjprs.2024.02.003.

[4] 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.

[5] S. Mishra, “A comparative study for time-to-event analysis and survival prediction for heart failure condition using machine learning techniques,” *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 4, no. 3, pp. 115–134, Jul. 2022, doi: 10.35882/jeeemi.v4i3.225.

[6] S. Lifang, “Design of intelligent environment observation and decision support for regional world economy,” *Mobile Information Systems*, vol. 2021, 2021, doi: 10.1155/2021/9078814.

[7] Y. Wang and Z. Zou, “Spatial decision support system for urban planning: case study of harbin city in China,” *Journal of Urban Planning and Development*, vol. 136, no. 2, pp. 147–153, 2010, doi: 10.1061/(asce)0733-9488(2010)136:2(147).

[8] I. Prasetyaningrum, K. Fathoni, and T. T. J. Priyantoro, “Application of recommendation system with AHP method and sentiment analysis,” *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 18, no. 3, pp. 1343–1353, Jun. 2020, doi: 10.12928/TELKOMNIKA.v18i3.14778.

[9] A. M. Arif, Abubaker M. Hamad, and M. M. Mansour, “Internet of (healthcare) things based monitoring for COVID-19+ quarantine/ isolation subjects using biomedical sensors, a lesson from the recent pandemic, and an approach to the future.,” *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 5, no. 1, pp. 1–12, 2023, doi: 10.35882/jeeemi.v5i1.267.

[10] S. Bhatta, “Empowering rural healthcare: MobileNet-driven deep learning for early diabetic retinopathy detection in Nepal,” *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 5, no. 4, pp. 290–302, Oct. 2023, doi: 10.35882/jeeemi.v5i4.326.

[11] S. Mishra, “Artificial intelligence: a review of progress and prospects in medicine and healthcare,” *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 4, no. 1, pp. 1–23, Jan. 2022, doi: 10.35882/jeeemi.v4i1.1.

[12] Y. Sarkingobir *et al.*, “Biomedical waste management among primary health care workers, bauchi local government area, Bauchi State, Nigeria,” *Jurnal Teknokes*, vol. 15, no. 4, pp. 242-251–242–251, Dec. 2022, doi: 10.35882/teknokes.v15i4.348.

[13] A. Piazza, C. Zagel, S. Huber, M. Hille, and F. Bodendorf, “Outfit browser – an image-data-driven user interface for self-service systems in fashion stores,” *Procedia Manufacturing*, vol. 3, pp. 3521–3528, 2015, doi: 10.1016/j.promfg.2015.07.686.

[14] A. Cogato, M. Brščić, H. Guo, F. Marinello, and A. Pezzuolo, “Challenges and tendencies of automatic milking systems (AMS): A 20-years systematic review of literature and patents,” *Animals*, vol. 11, no. 2, pp. 1–21, 2021, doi: 10.3390/ani11020356.

[15] Q. Li, "Design of customer churn early warning system based on mobile communication technology based on data mining," *Journal of Electrical and Computer Engineering*, vol. 2022, 2022, doi: 10.1155/2022/9701349.
[16] Y. Yuan, "Decision support system for evaluating English media analytics and the role of visualization in online business development using CRITIC and TOPSIS approaches," *Soft Computing*, 2023, doi: 10.1007/s00500-023-08094-z.
[17] R. R. D. Satya, Eriyatno, A. Ismayana, and Marimin, "Design of traceability system models for potato chips agro-industry based on fuzzy system approach," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 20, no. 4, pp. 797–807, Aug. 2022, doi: 10.12928/TELKOMNIKA.v20i4.23316.
[18] F. A. Bernardi, D. Alves, N. Crepaldi, D. B. Yamada, V. C. Lima, and R. Rijo, "Data Quality in health research: integrative literature review," *Journal of Medical Internet Research*, vol. 25, 2023, doi: 10.2196/41446.
[19] B. Peng and X. Pei, "A decision support system model for middle school education management based on sparse clustering algorithm," *Mobile Information Systems*, vol. 2022, 2022, doi: 10.1155/2022/4807395.
[20] S.-L. Wang, R. Maskey, A. Jafari, and T.-P. Hong, "Efficient sanitization of informative association rules," *Expert Systems with Applications*, vol. 35, no. 1–2, pp. 442–450, 2008, doi: 10.1016/j.eswa.2007.07.039.
[21] Z. Gao, S. Hu, G. Yu, and Y. Li, "A study on optimizing error detection and correction strategies in physical education and sport teaching using data mining algorithms," *Scalable Computing*, vol. 24, no. 4, pp. 1223–1229, 2023, doi: 10.12694/scpe.v24i4.2569.
[22] M. Jansevskis and K. Osis, "Knowledge discovery frameworks and characteristics," *Baltic Journal of Modern Computing*, vol. 11, no. 4, pp. 686–702, 2023, doi: 10.22364/bjmc.2023.11.4.08.
[23] W. A. Teniwut, C. L. Hasyim, and D. Arifin, "A Web-based DSS: information system for sustainable fisheries supply chain in coastal communities of small islands Indonesia," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 11, no. 3, pp. 1186–1192, 2021, doi: 10.18517/ijaseit.11.3.12462.
[24] R. Hou, X. Ye, H. B. O. Zaki, and N. A. B. Omar, "Marketing decision support system based on data mining technology," *Applied Sciences (Switzerland)*, vol. 13, no. 7, 2023, doi: 10.3390/app13074315.
[25] C. H. Fontes and O. Pereira, "Pattern recognition in multivariate time series - A case study applied to fault detection in a gas turbine," *Engineering Applications of Artificial Intelligence*, vol. 49, pp. 10–18, 2016, doi: 10.1016/j.engappai.2015.11.005.
[26] A. Sivanathan, J. M. Ritchie, and T. Lim, "A novel design engineering review system with searchable content: knowledge engineering via real-time multimodal recording," *Journal of Engineering Design*, vol. 28, no. 10–12, pp. 681–708, 2017, doi: 10.1080/09544828.2017.1393655.
[27] Y. Chung, L. Salvador-Carulla, J. A. Salinas-Pérez, J. J. Uriarte-Uriarte, A. Iruin-Sanz, and C. R. García-Alonso, "Use of the self-organising map network (SOMNet) as a decision support system for regional mental health planning," *Health Research Policy and Systems*, vol. 16, no. 1, 2018, doi: 10.1186/s12961-018-0308-y.
[28] D. Gil, J. L. Fernández-Alemán, J. Trujillo, G. García-Mateos, S. Luján-Mora, and A. Toval, "The effect of green software: a study of impact factors on the correctness of software," *Sustainability (Switzerland)*, vol. 10, no. 10, 2018, doi: 10.3390/su10103471.
[29] V. Ahmed, Z. Aziz, A. Tezel, and Z. Riaz, "Challenges and drivers for data mining in the AEC sector," *Engineering, Construction and Architectural Management*, vol. 25, no. 11, pp. 1436–1453, 2018, doi: 10.1108/ECAM-01-2018-0035.
[30] A. Wang and Y. Liu, "Intelligent financial management of company based on neural network and fuzzy volatility evaluation," *Journal of Intelligent and Fuzzy Systems*, vol. 38, no. 6, pp. 7215–7228, 2020, doi: 10.3233/JIFS-179798.
[31] M. L. H. Souza, C. A. da Costa, G. de O. Ramos, and R. da R. Righi, "A survey on decision-making based on system reliability in the context of Industry 4.0," *Journal of Manufacturing Systems*, vol. 56, pp. 133–156, 2020, doi: 10.1016/j.jmsy.2020.05.016.
[32] N. Yang, "Financial big data management and control and artificial intelligence analysis method based on data mining technology," *Wireless Communications and Mobile Computing*, vol. 2022, 2022, doi: 10.1155/2022/7596094.
[33] C. T. Su, Y. H. Chen, and D. Y. Sha, "Linking innovative product development with customer knowledge: a data-mining approach," *Technovation*, vol. 26, no. 7, pp. 784–795, 2006, doi: 10.1016/j.technovation.2005.05.005.
[34] 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.
[35] 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.
[36] A. Shillabeer, "An automated data pattern translation process for medical data mining," *Studies in Health Technology and Informatics*, vol. 129, no. Pt 1, pp. 586–590, 2007.
[37] K. Zhou, C. Yang, and J. Shen, "Discovering residential electricity consumption patterns through smart-meter data mining: a case study from China," *Utilities Policy*, vol. 44, pp. 73–84, 2017, doi: 10.1016/j.jup.2017.01.004.
[38] J. Liu, "Construction of emergency services platform for coal mining accidents by integrating multi source datasets," *International Journal of Simulation: Systems, Science and Technology*, vol. 17, no. 46, pp. 36.1-36.5, 2016, doi: 10.5013/IJSSST.a.17.46.36.
[39] J. Li, Y. Zhang, D. Du, and Z. Liu, "Improvements in the decision making for cleaner production by data mining: case study of vanadium extraction industry using weak acid leaching process," *Journal of Cleaner Production*, vol. 143, pp. 582–597, 2017, doi: 10.1016/j.jclepro.2016.12.071.
[40] C. M. Olszak and E. Ziemba, "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.
[41] F. M. M. de Barros, S. R. de M. Oliveira, and L. H. M. de Oliveira, "Desenvolvimento e validação de um sistema de recomendação de informações tecnológicas sobre cana-de-açúcar," *Bragantia*, vol. 72, no. 4, pp. 387–395, 2013, doi: 10.1590/brag.2013.049.
[42] O. Joldasbayev, D. Rakhmatullaeva, D. Polenov, and S. Joldasbayev, "Analysis of the applicability of existing methods and technologies of project-oriented management for government agencies in the Republic of Kazakhstan," *Public Policy and Administration*, vol. 19, no. 2, pp. 285–297, 2020, doi: 10.13165/VPA-20-19-2-10.
[43] J. C. Kim and K. Chung, "Depression index service using knowledge based crowdsourcing in smart health," *Wireless Personal Communications*, vol. 93, no. 1, pp. 255–268, 2017, doi: 10.1007/s11277-016-3923-3.

[44] D. Dzemydiene and R. Dzindzalieta, "Development of Architecture of embedded decision support systems for risk Evaluation of Transportation of Dangerous Goods," *Technological and Economic Development of Economy*, vol. 16, no. 4, pp. 654–671, 2010, doi: 10.3846/tede.2010.40.
[45] 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.
[46] L. F. G. Morales, P. Valdiviezo-Diaz, R. Reátegui, and L. Barba-Guaman, "Drug recommendation system for diabetes using a collaborative filtering and clustering approach: development and performance evaluation," *Journal of Medical Internet Research*, vol. 24, no. 7, 2022, doi: 10.2196/37233.
[47] H. Ji and X. Luo, "Implementation of ensemble deep learning coupled with remote sensing for the quantitative analysis of changes in arable land use in a mining area," *Journal of the Indian Society of Remote Sensing*, vol. 49, no. 11, pp. 2875–2890, 2021, doi: 10.1007/s12524-021-01430-6.
[48] E. D. Ananta, S. Syaifudin, L. D. Soetjiatie, and B. Utomo, "Development IoT-based infant monitoring system for preventing sudden infant death syndrome (SIDS) with abnormal condition notifications and lost data analysis," *Jurnal Teknokes*, vol. 16, no. 2, pp. 73–79, Jul. 2023, doi: 10.35882/teknokes.v16i2.485.

BIOGRAPHIES OF AUTHORS

Bahar Asgarova working as associate professor in Computer Engineering department of Azerbaijan State Oil and Industry University. Currently researches with her students on topics like extracting knowledge from web information sources using data mining, approaches of information cleaning from big data, and identifying cybersecurity threats from big data with interactive visualization. She can be contacted at email: bahar.askarova@asoiu.edu.az.

Elvin Jafarov studying post-graduate in Information Technologies and Management at Azerbaijan State Oil and Industry University. Current research investigates development of a method of cleaning information from big data. In this dissertation, the existing data cleaning models were reviewed and their shortcomings were identified, and new methods were created. And as a result, this system will show us the information we need, and most importantly, save time. He can be contacted at email: elvincafarov95@gmail.com.

Nicat Babayev studying post-graduate in Information Technologies and Management at Azerbaijan State Oil and Industry University. Current research investigates architecture for knowledge extraction platform from web information sources using data mining. The aim of the research to develop a infrastructure architecture which can host knowledge extraction application from web information sources which utilizes data mining approaches. The platform based on the architecture is deployed to the cloud. He can be contacted at email: nicatbabayev@gmail.com.

Allahshukur Ahmadzada studying post-graduate in Information Technologies and Management at Azerbaijan State Oil and Industry University. Current research explores the utilization of interactive visualization techniques to identify cybersecurity threats within large data sets. This study assesses the effectiveness of existing cybersecurity threat detection methods applied to big data and identifies their limitations. By developing advanced interactive visualization tools, the research aims to enhance the ability to swiftly and accurately pinpoint potential security vulnerabilities within vast amounts of data. Consequently, this approach is anticipated to significantly improve the efficiency and accuracy of cybersecurity threat detection, aiding in the prompt identification and mitigation of security risks. He can be contacted at email: a.ahmadzada1998@gmail.com.

Vugar Abdullayev, Ph.D. was born in Azerbaijan. received the B.S. degree in Automatics and control of technical systems specialty from the Azerbaijan State Oil and İndustry University (ASOİU), Baku, Azerbaijan, M.S. degree in Manufactory Automation and Informatics specialty from the Azerbaijan State Oil and İndustry University (ASOİU), Baku, Azerbaijan in 2000, and a Ph.D. degree from - Institute of Cybernetics of Azerbaijan National Academy of Sciences in 2005. In 2002-2004 – Dr. Vugar Abdullayev has been expert on IT and Payment systems department in the Azerbaijan Central Bank. In 2004-2012 – Dr Vugar Abdullayev has been an Researcher, head researcher in the Institute of Cybernetics of Azerbaijan National Academy of Sciences, Baku, Azerbaijan. Since 2012, he is an doctor of technical sciences, Associate Professor at Azerbaijan State Oil and Industry University, Department of Computer Engineering. He is author of 85 scientific papers. His researchers related to the study of the cyber physical systems, IoT, big data, smart city and information technologies, cloud computing, computational complexity, machine learning (artificial intelligence), and behavioral sciences computing. He has published 20 book chapters and 10 edited books (calling for book chapters - Taylor and Francis) in healthcare ecosystem. He can be contacted at email: abdulvugar@mail.ru.

Dr. Triwiyanto received the B.S. degree in Physics from Airlangga University, Indonesia, M.S. degree in Electronic Engineering from the Institut Teknologi Sepuluh Nopember Surabaya, Indonesia, in 2004, and a Ph.D. degree in Electrical Engineering from Gadjah Mada University, Yogyakarta, Indonesia, in 2018. From 1998 to 2004, he was a Senior Lecturer with the Microcontrollers Laboratory. Since 2005, he has been an Assistant Professor with the Medical Electronics Technology Department, Health Polytechnic Ministry of Health Surabaya, Indonesia. In 2018, Triwiyanto received the best Doctoral Student award from Gadjah Mada University. Additionally, he is Editor-in-chief in several peer review journals and chairman Technical Programme Committee at several International Conferences. His current research interests include a microcontroller, electronics, biomedical signal processing, machine learning, rehabilitation engineering, and surface electromyography (sEMG)-based physical human-robot interactions. He can be contacted at email: triwiyanto123@gmail.com.