Knowledge Extraction and Data Visualization: A Proposed Framework for Secure Decision Making using Data Mining

Hazzaa N. Alshareef[1]
College of Computing & Informatics
Saudi Electronic University
Madinah, Saudi Arabia

Ahmed Majrashi[2]
Muhammad Tahir[4]
College of Computing & Informatics
Saudi Electronic University
Riyadh, Saudi Arabia

Maha Helal[3]
College of Computing & Informatics
Saudi Electronic University
Jeddah, Saudi Arabia

***Abstract*—The decision-making process, promptly on time, is a crucial success factor in large organizations. Generally, the data warehouses of these organizations grow rapidly with the data generated from various business activities. This huge volume of data needs to be analyzed and decisions must be made quickly to meet the market challenges. Accurate knowledge extraction and its visualization from big data can guide decision-makers to conduct key analysis and make correct predictions. This paper proposes a decision-making framework that not only takes into account knowledge extraction and visualization but also considers the security of the data. The proposed framework uses data mining techniques to extract useful patterns, then, visualizes those patterns for further analysis and decision making. The significance of the proposed framework lies in the mechanism through which it protects the data from intruders. The data is first processed and then stored in an encrypted format on the cloud. When the data is needed for analysis and decision making, a temporary copy of the data is first decrypted, and then important patterns are visualized. The proposed framework will assist managers and other decision-makers to analyze and visualize the data in real-time with an enhanced security mechanism.**

***Keywords*—Big data; data mining; data visualization; classification; cloud computing; security**

I. Introduction

Digital transformation has impacted many aspects of everyday life. Advancements in technology have resulted in an overall transformation in many industries changing the traditional ways of performing and managing tasks. As a result, the volume of available data has increased tremendously, which led to what is known as the era of big data [1] and [2]. Nowadays, organizations have the advantage of being equipped with large subsets of data that could potentially provide them with a competitive advantage. This is made possible by extracting insightful information that could aid organizations in decision making [2].

However, big data includes a mixture of structured and unstructured data that is diverse in nature and collected from many different sources such as smart devices, IoT sensors, social media applications, websites, emails, medical records and different types of documents [3] and [4]. Therefore, the data collected in its many forms does not directly help organizations. These large unstructured datasets need to be analyzed using specific methods such as data mining, machine learning, and artificial neural networks [5]. Once this is achieved, only then can big data serve as a driving force for organizations and provide value with advantages in many ways. These advantages include discovering trends and hidden patterns that could serve as a foundation for making future decisions in different aspects such as resource allocation, guiding production, and exploiting new opportunities [1] and [6].

The main contribution of this paper is to introduce novelty in classifying big data for decision making. The paper provides a proposed framework that helps organizations to make decisions based on information derived from classifying raw unstructured data using data mining algorithms. This information is then visualized in appropriate ways to present the extracted information in the best possible manner to enable managers to make informed decisions. Furthermore, a security mechanism is provided to ensure a high-level of security and protection while extracting rich insights of the data. Though, encrypting the data prior to storing in the database consumes additional resources, it is highly beneficial to the reliability of the data.

The remaining of this paper is organized as follows: Section II provides previous work found in the literature relating to big data and data mining. Subsequently, Section III explains the proposed framework while Section IV provides details regarding the experimental work. This is followed by Section V that discusses the results and findings of this research. Finally, Section VI concludes the work and provides future directions.

II. Related Work

While it remains ambiguous on what constitutes big data, there resides a consensus on at least three dimensions that prevail in the literature; volume, variety and velocity [7]. In brief, volume refers to the size of the dataset; variety refers to the different forms and sources of data that construct the dataset; and velocity refers to the speed of data generation and analysis [5], [7], and [8].

The concept of big data forced organizations to revolutionize the way they manage their data. Rather than just focusing on adopting effective methods to collect and store data, the challenge has shifted to finding effective mechanisms to extract valuable knowledge from this data. It also provides them with gaining meaningful insights that were hidden otherwise, which in turn offers many valuable opportunities. In doing so, organizations have the potential of gaining a competitive advantage over their competitors [6].

However, with all the benefits that could be achieved from big data, several challenges are imposed. Having large amounts of data makes it more difficult for organizations to extract valuable information [5]. Datasets are derived from many different sources such as databases, social media applications, emails, videos, documents, and IoT devices. In addition, the nature of these datasets is different where some could be structured, semi-structured or unstructured [9]. Therefore, traditional methods of analyzing data which organizations have been using are no longer effective in handling the large volume and diversity of data residing in large datasets [2]. Hence, new methods have been introduced to deal with the complex task of analyzing big data such as data mining, machine learning and artificial neural networks [5].

Data mining can be defined as the systematic process of extracting useful knowledge by examining large datasets from different sources and discovering hidden patterns. The knowledge that is discovered as a result of data mining provides valuable insights and aids organizations in decision making [10]. Data mining algorithms can be classified into two categories: descriptive models and predictive models. Descriptive models, referred to as unsupervised learning, are used to search for different patterns in the dataset and recognize any associations between them after applying revision techniques. In contrast, predictive models, referred to as supervised learning, are mostly used to predict and forecast outcomes from present behavior [11] and [12].

There are also different data mining techniques for analyzing descriptive and predictive models such as clustering, association mining, and classification. Clustering is the process of classifying similar objects into the same cluster depending on the specific characteristics of different objects in the dataset. Subsequently, the objects are grouped into different classes. By doing so, different inherent relationships become apparent, which provide valuable information that assists managers in decision making [13]. Alternatively, association mining involves discovering relationships between two items or concepts by first identifying the frequent itemset and then generating rules for them [14].

Moreover, classification in data mining techniques consists of two important phases namely, training and testing [11]. The training phase is concerned with building the classification model based on collecting training data in order to generate and create the classification rules. Subsequently, during the testing phase, the classification model is tested by applying classification rules and the accuracy of the result is determined by evaluating the true results of the classification rules [11]. Classification can also be categorized as supervised or unsupervised depending on whether the objects or cases are known in advance or not [10].

Many different classification algorithms can be used to analyze big data such as K-Nearest Neighbor, Naïve Bayes, Support Vector Machines (SVM), and Decision Tree [15] and [10]. K-Nearest Neighbor is a non-parametric simple classification method that is based on distance measurement. The algorithm classifies and stores any new cases depending on the distance function [15]. Alternatively, Naïve Bayesian algorithm is a probabilistic classifier, which deals with classification problems as probabilistic problems [16]. It fits very well with text data and requires a small amount of training data. However, the output or probability value should be assessed to ensure a high level of accuracy. Another example of a well-known classification algorithm is SVM where training data is represented as points in space separated into categories. Subsequently, new data is mapped to space, which belongs to such a category in that space. It is a memory-efficient algorithm suitable for high dimensional spaces and has the benefit of having a fast computational process [17] and [18].

In addition, the Decision Tree algorithm is also considered to be a widely used classification algorithm. It classifies data by generating a sequence of rules after assigning attributes to the data together with its classes. This algorithm is simple and easy to implement since it requires less data preparation and can work with numerical and categorical data. However, generating complex trees could be an issue. Moreover, it could be unsuitable where small variations exist in the data, which might result in a completely different tree [19]. In the current paper, Naïve Bayesian algorithm is used due to its fast and high scalability characteristic for the classification process.

As stated earlier, big data analysis provides many benefits and valuable opportunities for organizations with huge amounts of data [6]. These large complex datasets need to be normalized and analyzed in order to produce insightful knowledge for the organization to make important decisions and predictions for future situations [5]. Therefore, organizations and decision-makers need to exploit the data that is being generated and collected on a daily basis by analyzing it thoroughly. This will help them make informed decisions to run their business operations efficiently [20].

For this purpose, the concept of big data has become one of the important topics that organizations should invest in. This is because it provides effective ways of creating insights and knowledge from large amounts of unstructured raw data that would neither be obvious nor understandable unless some form of analysis is performed [20]. However, organizations and decision-makers first need to identify the data that has possible benefits. They also need to consider and prioritize their business needs and subsequently initiate the process of data collection and analysis. This is to reduce wasting valuable time in collecting and analyzing irrelevant data that will not lead to generating insightful knowledge [9].

The literature on big data analysis covers a wide spectrum of studies that have used classification algorithms for decision making in different fields [21], [22], [23], and [24]. In their study [21], they reported the successful use of Naïve Bayesian

algorithm to classify and evaluate relevant alternatives for decision making in human-machine systems of critical applications. They indicated that although Naïve Bayesian is a simple classifier, trials have revealed that it is as effective as other complicated algorithms. In addition, the authors in [22] also used different classification algorithms to classify and predict solutions to help in making decisions with regards to heart diseases in patients. This helped the organization gain a competitive advantage as practitioners were able to make faster decisions based on the output of the algorithms.

However, these studies did not consider security measures. When unstructured data gets classified and structured to inform decisions, this poses security concerns that organizations need to consider. As mentioned earlier, the volume and type of data being created in recent times are much greater than before. Different types of data are being generated through different types of applications such as smart devices, IoT sensors, social media applications, websites, emails, and different types of documents [1]. Therefore, traditional security and privacy mechanisms used by organizations to protect their data are not fully capable of providing the same level of protection to big data as it holds different types of characteristics [3].

In [25], the authors reported that traditional security mechanisms such as access control, encryption, authorization, and multi-factor authentication are not considered to be effective methods for providing a protective environment for big data. Securing the network used for big data access, transformation, and storage is an essential part to prevent different types of intrusions and attack activities such as DoS attacks, unauthorized access, spoofing and spamming [25]. Credibility, availability, data privacy, confidentiality, authentication and integrity are a few of the security and privacy issues associated with big data [3], [4], and [26]. For instance, the authors in [27] proposed a 3D security model for big data that is based on user roles, data processes as well as security requirements. According to the authors, these requirements define the security objectives and goals that should be achieved to preserve data security and privacy.

Visualizing the patterns and output results of data mining could provide a comprehensive overview of the results of any process. This helps in understanding and providing insightful discovery since the data is presented in an attractive manner. Moreover, when knowledge or patterns are visualized, finding the relationship amongst data that is tested becomes an easier task than presenting it as normal data without any visual means. This is because the amount of tested data is enormous in big data and visualizing it in an attractive and understandable manner is essential for decision making [28]. In [29], the authors described several techniques that need to be considered in order to create meaningful visual data. They stated that data size and structure play a major role.

However, the visualization process might face some challenges. The authors in [30] stated that visualization tools should be able to provide an interactive output with minimum latency to meet or achieve user satisfaction. There are some techniques discussed in the literature that could be used to reduce the latency issue such as pre-computed data, parallelization for the processed and rendered data, and applying predictive middleware [31]. All these techniques could be used to reduce latency and avoid such drawback.

Visualization tools should be able to deal with and process semi-structured and unstructured data as most of the big data is in such a format [3]. Moreover, visualization tools should be capable of optimizing the performance in terms of scalability, functionality, and response time. Another challenge that could be faced with regards to visualization is information loss in order to scale the size of data for better performance which leads to data loss. In addition, the noise of visualization is a challenge because of irrelevant data or elements in the dataset. Generally, there is a need for high-performance tools to meet and achieve the desirable scalability, functionality, and response time [30].

III. Proposed Framework

In this section, we present the details of our proposed framework.

A. System Overview and Main Components

The system overview design describes the broad view of the system where such design is expressed by three major components that include the organizations, the cloud, and the decision-makers. The organizations generate the data, the cloud offers processing facilities, and the decision-makers have authoritative access to the processed data.

Several components formulate the framework presented in the above figure (Fig. 1) which are:

- The organization where the data is generated.
- Managers or decision-makers who access the cloud to retrieve the discovered knowledge.
- The database where all the generated data is stored.
- The classifier that categorizes the stored data.
- The processor that further processes the classified data.
- The pattern where the classified data relationship is discovered.
- The visualization process where the data becomes readable by managers and decision-makers.

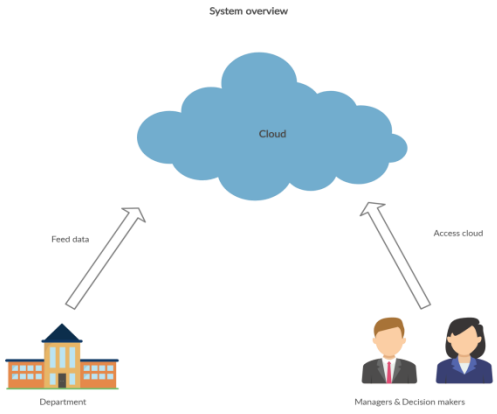

Fig. 1. Framework Overview.

B. System Flows

The organization creates and sends the data to the cloud through the web application that interacts with the cloud. The sent data is stored in a database that is located within the cloud. The stored data is then passed through a classifier that categorizes the data, which is then processed by the mining algorithm to generate insights or knowledge from such data. Subsequently, the produced result is visualized by visualization techniques which helps decision-makers take the right actions based on the knowledge that is produced. Managers and decision-makers access the cloud to retrieve the results using a web browser, which allows them to use a web application for granting access to the cloud. Fig. 2 illustrates the system flows graphically.

C. Security Mechanism

The system should implement security methods that protect and prevent data from breaches. According to the proposed framework, the data is first encrypted by the data owner or sender before storing it on the cloud where the data will be processed and stored in its encrypted form. A secure key is then sent to the receiver through a secure channel other than the cloud. Subsequently, the receiver accesses the cloud and fetches the encrypted data in an encrypted form, which is decrypted with a secure token. Fig. 3 shows the proposed security method.

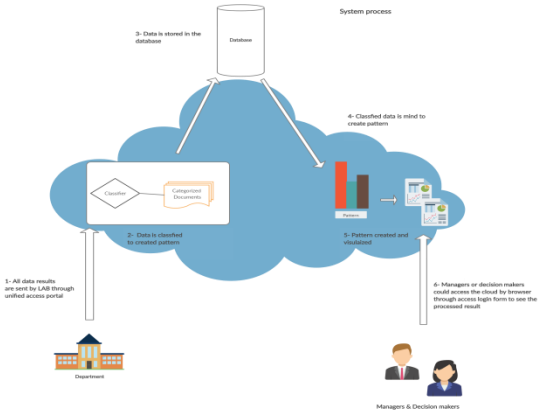

Fig. 2. Framework Process.

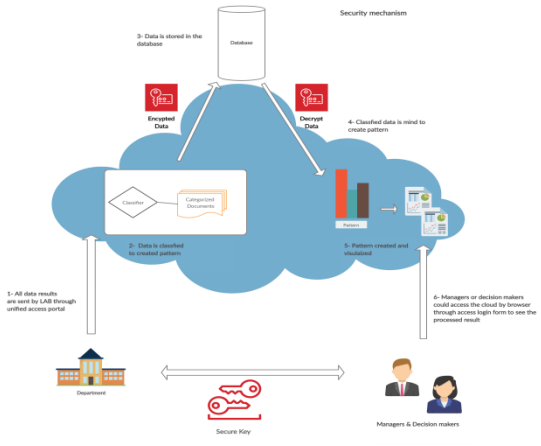

Fig. 3. Security Mechanism.

IV. Experimental Work

This section details the steps followed in performing the experiment which includes the setup and how the data is classified and how it is visualized.

A. Experiment Setup

Localhost for Apache server and MySQL database server were used during the implementation process with XAMMP software which offers all the necessary services for this experiment. In addition, a web application has been developed using ASP.NET and PHP to execute the functionality needed for the proposed framework. This includes processes for categorization, visualization, as well as data storage management. PHP was selected due to the fact it is supported by a range of libraries, functions, and modules that make the development of PHP based web applications swifter without any complications [32]. Similarly, ASP.NET is an open-source server-side programming language that is introduced by Microsoft to facilitate the development of dynamic web pages.

B. Classifier

The classifier functionality is based on organizing unrelated raw data into a categorized form that is based on predefined rules. In this work, we have utilized Naïve Bayesian classifier due to its speed and scalability features. The Naïve Bayesian classifier is based on Bayes' theorem. It assumes that each feature of a class is highly independent where the appearance of such feature in a specific class or category is not associated with any of the other features [33]. The classification process is achieved through the earlier probability and likelihood of a sample to a class. For this experiment, the classifier is built and trained using the above-mentioned web application to calculate the probability of each category.

C. Visualization

In practice, data visualization is used to convert plaintext data into figures and graphs that make such data more understandable for the managers and decision-makers. It provides a clear and comprehensive overview of the data in an attractive and interesting manner. In this work, Google Charts (Tableau) was used due to its capability of producing impactful and easy to read visual reports. Additionally, it allows the creation of customized dashboards based on the user's needs and provides the option of integrating them into the above-mentioned web application.

V. Result and Discussion

For this paper, data from a petroleum testing lab has been used and categorized into three classes based on the type of request from various customers. Those requests were grouped into three categories which are: critical requests (High priority), important requests (Mid priority), and normal requests (Low priority). As mentioned above, the Naïve Bayes classifier has been trained on the existing data which calculates the probability percentages of the submitted requests.

Based on predefined data and labels which are the three categories, the classifier has the ability to classify the submitted requests. For instance, keywords such as “urgently needed”, “quick action”, and “less than 8 hours” are categorized as High. In addition, keywords such as “diesel key test” and “quality

check" are categorized as Mid, whereas keywords such as "more than 8 hours" and "storage facility" are categorized as Low. A condition was built on the body text of the request. If the request's body text percentage for High is greater than the body text percentage for Mid, the request will be categorized as "High". Alternatively, if the request's body text percentage for Mid is greater than the body text percentage for High or Low, the request will be categorized as "Mid". The default value is set to be "Low", therefore if the first two conditions were not met, the request will be categorized as "Low".

The categorized request is submitted to the database for visualization to assist decision-makers and managers. This categorization and visualization process provides an overview and explanation about requests fulfilment and the number of submitted requests.

The data is stored in an encrypted format using the AES algorithm with a 256-bit key that is difficult to break. Once the data is needed to be visualized, a temporary copy from the database is decrypted in order to extract knowledge and identify the important patterns available through a visual format. After the completion of this task, the temporary copy is discarded and only the obtained results are processed further. As a result of the visualization process, several graphs have been generated to fulfil the information needs of different customers. The following graphs illustrate the dataset that was used which includes 89 samples. The classifier gave the following results: high-priority requests (9), mid-priority requests (55), and low-priority requests (25). Fig. 4 and 5 present the visualization of these samples based on the day of one week, types of customers, and type of sample.

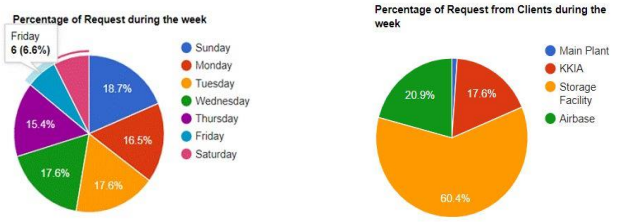

Fig. 4. Percentage of Requests during the Week and Percentage of Requests from Clients.

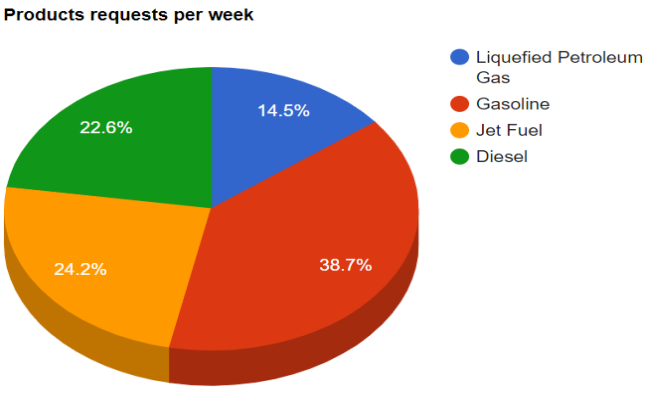

Fig. 5. Percentage of Product Types per Week.

Different types of tests could be performed in the lab based on customer's needs and each type has a specific duration requirement. Some tests could be performed in less than 8 hours, whereas other tests may take longer. For instance, the key test type could be identified in less than 8 hours whereas a full certificate test requires more than 8 hours. In all cases, the determination of the test duration is predefined by the lab unit. Fig. 6 shows the number of key tests and full certificate tests that have been requested during a week.

Additionally, the capacity of the lab in fulfilling the number of requests depends upon the number of employees available on that day. Fig. 7 shows the number of employees and the number of requests for each day of the week. This will help in balancing the workload among employees and avoids accepting any requests that cannot be fulfilled. Once the day is highlighted in red, this indicates the number of requests reached the maximum number of available employees. Thus, any subsequent requests will be denied for that day.

With regards to data security, data is stored in an encrypted format using the AES algorithm with a 256-bit encryption key. In this way, the data is protected against any malicious attacks that may affect the integrity and confidentiality of the stored data. Fig. 8 shows the encrypted data stored in the database.

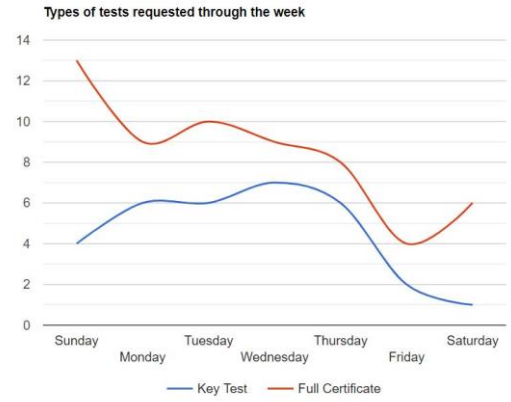

Fig. 6. Number of Requests for each Type of Test.

Fig. 7. Employees and Requests.

Fig. 8. Encrypted Data.

VI. Conclusion and Future Work

Storing, analyzing and the visualization of big data are challenging tasks for any organization around the world. The data generated due to routine activities of an organization leads to an avalanche of raw and unstructured data. Data mining techniques can be used effectively to extract intrinsic patterns from data in the data warehouse and present them in an organized manner to assist managers and decision-makers to set future directions.

The proposed framework is capable of prioritizing requests received from customers with data visualization charts that allow decision-makers to optimize the available resources and produce the laboratory results according to the importance of the requests. The data visualization process can help decision-makers to either stop, increase, or decrease the working hours of laboratory staff for fulfilling various tasks. Furthermore, the proposed framework offers features that visualize the stored data without risking its security. This enables decision-makers and managers to monitor and track all the processes from request generation to the final fulfilment. Many business processes can be improved by identifying important patterns in the data that will lead to enhanced services. The proposed framework can be further enhanced by improving the quality of the classification process by taking into account more features of the input data. Further, the additional time consumed by the encryption technique will also be considered in our future endeavor. Additionally, the proposed framework will be equipped with a greater number of charts to assist decision-makers to visualize the data from different perspectives.

References

[1] Al-Mekhlal, M., Khwaja, A.A., 2019. A Synthesis of Big Data Definition and Characteristics, in: 2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC). pp. 314–322.

[2] Wang, Y., Li, Y., Sui, J., Gao, Y., 2020. Data Factory: An Efficient Data Analysis Solution in the Era of Big Data, in: 2020 5th IEEE International Conference on Big Data Analytics (ICBDA). pp. 28–32.

[3] Bertino, E., 2017. Big data security and privacy, in: 2017 Innovations in Power and Advanced Computing Technologies (i-PACT). pp. 3–3. https://doi.org/10.1109/bigdata.2016.7840581

[4] Gaurav, D., Yadav, J.K.P.S., Kaliyar, R.K., Goyal, A., 2018. An outline on big data and big data analytics, in: 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN). pp. 74–79.

[5] Grable, J.E., Lyons, A.C., 2018. An Introduction to Big Data. J. Financ. Serv. Prof. 72.

[6] Raguseo, E., 2018. Big data technologies: An empirical investigation on their adoption, benefits and risks for companies. Int. J. Inf. Manage. 38, 187–195.

[7] Simsek, Z., Vaara, E., Paruchuri, S., Nadkarni, S., Shaw, J.D., 2019. New ways of seeing big data.

[8] Tonidandel, S., King, E.B., Cortina, J.M., 2018. Big data methods: Leveraging modern data analytic techniques to build organizational science. Organ. Res. Methods 21, 525–547.

[9] Gupta, B., Kumar, A., Dwivedi, R.K., 2018. Big Data and Its Applications--A Review, in: 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN). pp. 146–149.

[10] Job, M.A., 2018. Data Mining Techniques Applying on Educational Dataset to Evaluate Learner Performance Using Cluster Analysis. Eur. J. Eng. Technol. Res. 3, 25–31.

[11] Patel, S., Patel, H., 2016. Survey of data mining techniques used in healthcare domain. Int. J. Inf. 6, 53–60.

[12] Ţăranu, I., 2016. Data mining in healthcare: decision making and precision. Database Syst. J. 6, 33–40.

[13] Bin, N., 2018. Research on Methods and Techniques for IoT Big Data Cluster Analysis, in: 2018 International Conference on Information Systems and Computer Aided Education (ICISCAE). pp. 184–188.

[14] Rojanavasu, P., 2019. Educational data analytics using association rule mining and classification, in: 2019 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT-NCON). pp. 142–145.

[15] Jalota, C., Agrawal, R., 2019. Analysis of educational data mining using classification, in: 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon). pp. 243–247.

[16] Wang, Q., Pan, L., Shen, J., Lu, N., 2019. Power plant data filtering based on Gaussian naive Bayesian classification and prediction error method, in: 2019 Chinese Automation Congress (CAC). pp. 1490–1495.

[17] Mishael, Q., Ayesh, A., 2020. Investigating Classification Techniques with Feature Selection For Intention Mining From Twitter Feed. arXiv Prepr. arXiv2001.10380.

[18] Wardani, S., Sihombing, P., others, 2020. Hybrid of Support Vector Machine Algorithm and K-Nearest Neighbor Algorithm to Optimize the Diagnosis of Eye Disease, in: 2020 3rd International Conference on Mechanical, Electronics, Computer, and Industrial Technology (MECnIT). pp. 321–326.

[19] Kowsari, K., Jafari Meimandi, K., Heidarysafa, M., Mendu, S., Barnes, L., Brown, D., 2019. Text classification algorithms: a survey. Information 10 (4): 150. arXiv Prepr. arXiv1904.08067.

[20] Sahu, S.K., Jacintha, M.M., Singh, A.P., 2017. Comparative study of tools for big data analytics: An analytical study, in: 2017 International Conference on Computing, Communication and Automation (ICCCA). pp. 37–41.

[21] Peredery, V., Voronenko, M., Borchik, E., Krejci, J., 2018. Information Technology of Operative Formation and Evaluation of Relevant Alternatives of Decision-Making in Human-Machine Systems of Critical Application, in: 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT). pp. 33–36.

[22] Saboji, R.G., 2017. A scalable solution for heart disease prediction using classification mining technique, in: 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS). pp. 1780–1785.

[23] Supianto, A.A., Dwitama, A.J., Hafis, M., 2018. Decision Tree Usage for Student Graduation Classification: A Comparative Case Study in Faculty of Computer Science Brawijaya University, in: 2018 International Conference on Sustainable Information Engineering and Technology (SIET). pp. 308–311.

[24] Tai, C.-C., Lin, H.-W., Chie, B.-T., Tung, C.-Y., 2019. Predicting the failures of prediction markets: A procedure of decision making using classification models. Int. J. Forecast. 35, 297–312.

[25] Bhatia, R., Sood, M., 2018. Security of Big Data: A Review, in: 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC). pp. 182–186.

[26] Xu, L., Shi, W., 2016. Security theories and practices for big data, in: Big Data Concepts, Theories, and Applications. Springer, pp. 157–192.

[27] Lv, D., Zhu, S., Xu, H., Liu, R., 2018. A Review of Big Data Security and Privacy Protection Technology, in: 2018 IEEE 18th International Conference on Communication Technology (ICCT). pp. 1082–1091.

[28] Bikakis, N., 2018. Big data visualization tools. arXiv Prepr. arXiv1801.08336.

[29] Toasa, R., Maximiano, M., Reis, C., Guevara, D., 2018. Data visualization techniques for real-time information - A custom and dynamic dashboard for analyzing surveys' results, in: Iberian Conference on Information Systems and Technologies, CISTI. pp. 1–7. https://doi.org/10.23919/CISTI.2018.8398641.

[30] Ali, S.M., Gupta, N., Nayak, G.K., Lenka, R.K., 2016. Big data visualization: Tools and challenges, in: 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I). pp. 656–660.

[31] Agrawal, R., Kadadi, A., Dai, X., Andres, F., 2015. Challenges and opportunities with big data visualization, in: Proceedings of the 7th International Conference on Management of Computational and Collective IntElligence in Digital EcoSystems. pp. 169–173.

[32] Hills, M., 2015. Evolution of dynamic feature usage in PHP, in: 2015 IEEE 22nd International Conference on Software Analysis, Evolution, and Reengineering (SANER). pp. 525–529.

[33] Deeba, K., Amutha, B., 2016. Classification algorithms of data mining. Indian J. Sci. Technol. 9, 1–5.