Real-time Data Mining Algorithm Analysis and Application: A Review

Lana Lateef Nahmatulla[1]

[1]Department of Computer Science, Salahaddin University-Erbil, Kurdistan Region, Iraq

***Abstract*— The goal of real-time data mining is to find patterns and insights in massive volumes of data as they are created, rather than having to wait for data to be gathered and analyzed. Applications in government, business, and other fields where accurate and fast information is critical for reaction and decision-making need this technique. Rather of waiting for data to come or processing it in batches, algorithms and methods used in real-time data mining examine data in real-time. Problems arise when trying to account for the ever-changing nature, rapid pace, many data types, and large amounts of data involved in real-time data mining. In order to gain insights and enhance system performance, real-time data mining may analyze multi-dimensional data, including power grid data in the power business. Building intrusion detection systems (IDSs) based on real-time data mining allows for the processing of audit data and the real-time detection of intrusions, all while addressing concerns of usability, efficiency, and accuracy. Because of the critical nature of real-time data processing and the need for systems to adapt to dynamic situations, there is a pressing need for research into methods to guarantee the accuracy and dependability of these systems.**

***Index Terms*—Real-time data mining, Timely information, Algorithms and techniques, Multidimensional data analysis.**

I. Introduction

Data gathering, organizing, and cognitive analysis are the cornerstone methods for discovering patterns and correlations in every human endeavor. The most up-to-date and useful resources for examining and analyzing data are data mining methods. Data mining techniques are used in a wide range of operations, from applications on mobile phones to supercomputers. Collecting and analyzing such large datasets presents a number of challenges, including, but not limited to, the complexity and variable quality of the data itself (McKinsey and Company, 2016). Predictive skills cannot be provided by data-driven decision-making processes. This may be circumvented by making use of real-time data mining tools, which can analyze data and provide patterns in both short and long periods. Tools for data-driven decision-making pale in comparison to real-time data mining and analysis in many respects. Considering the following comparisons made by Gartner in 2023:

- ➢ Real-time data mining analysis techniques center on analyzing data as it happens, which is ideal for circumstances that need a prompt reaction or intervention. However, other data-driven decision-making methods are used for the purpose of identifying the variables and reasons behind events that have already occurred.
- ➢ Quick decision-making and prompt actions are accomplished via the use of data processing technologies that are either real-time or near real-time. We use diagnostic analytics to look at past data. Determining meaning from previously obtained data often takes a considerable amount of time.
- ➢ Objective: Real-time data mining analysis aims to swiftly monitor, identify, and respond. Conversely, data-driven methods are used for the purpose of analyzing the sources and causes of certain events.
- ➢ For real-time data processing, data sources might include sensors, Internet of Things devices, social media feeds, or live transactional data. In contrast, non-real-time data analysis relies on information retrieved from systems, data warehouses, or databases.
- ➢ Methods for evaluating Data: Data steaming and real-time data visualization are methods for evaluating data in real-time data mining. Examining past data for patterns, correlations, and causal linkages is an important part of diagnostic analytics. Statistical analysis, root cause analysis, and hypothesis testing are tools for this process.
- ➢ Valid instances of real-time data mining methods include fraud detection and real-time dashboards. Contrarily, patient outcome analysis aims to determine the reasons for a drop in sales.

According to (Taylor, 2022), in order to define data mining clearly, Data mining is the process of identifying important patterns in huge datasets by sifting through the data and using learning algorithms to analyze it. In addition, a condition exists that denotes two periods of as is common knowledge in

DOI:http://doi.org/10.24086/cocos2024/paper.1091

database management, data mining and knowledge discovery are used synonymously. Data mining methods enable the examination of massive amounts of data, which would be very difficult for humans to do manually. The underlying linkages in the massive amounts of data with varied forms and dimensions become more complicated. This data emerged from a few key sources. Greater difficulty in discovering relationships is caused by diverse dimensionalities and representations of big data sets. (Sanders et al. 2016) include a variety of sources as sources of data.Internet data, stock prices, and other forms of electronic trade are all considered business data. Scientific data, including that gathered via remote sensing and modeling, and societal data, including that gathered through the news, pictures, videos, and so on (Firth et al., 2017).Traditional data analysis does not cover the extra requirements that end users have for the data that is already accessible. On the other hand, trends, similar to business-driven demands, are elements that cause the data collected over time to alter. As a result, it is critical to create novel approaches to data analysis based on DM (data mining). Researchers are finding DM to be an increasingly intriguing field because to the growing demand for it in both social media and the real world (Gongora et al., 2018). In the data mining application sector, each kind of approach serves a unique function. Two major categories of data analysis techniques are statistics and machine learning (ML). We use statistical tools to generate conclusions and test our established assumptions. Additionally, ML algorithms are taught to generalize seen data with fresh data (Hartmann, 2019). When deciding. With DM, decision-makers may optimize, segment, associate, and categorize data, in addition to discovering rules concealed in it. DM summarizes the results of an analysis that takes into account several dimensions and points of view. The previous approach could not properly extract valuable information in real-time (Iorfino et al., 2021). Consequently, the need for the automated extraction of valuable information arose. When it comes to handling massive volumes of data, DM has a significant impact in limiting human decision-making constraints like subjectivity and mistake. The comparison between traditional methods and data mining methods are presented in the mentioned researcher.

Figure 1.1: Data mining process (Kassianos et al, 2017).

Supervised, unsupervised and semi-supervised are techniques in DM learning. DM eliminates subjectivity and mistakes from massive data sets, easing human decision-making. In contrast to traditional techniques, DM uses repeating procedures from data purification to knowledge presentation (Ponniah, 2001; Chandra & Gupta, 2018). This image shows data purification, integration, selection, and translation into suitable forms in the DM process. Three DM learning approaches exist: supervised, unsupervised, and semi-supervised. The purpose of this review paper to explore the methodologies, applications, and advancements in real-time data mining. The sections of the papers outline as follows: The purpose of real-time data mining analysis and application is to extract useful patterns and insights from data as it is generated or collected, and then use those insights to make immediate decisions or take immediate actions. This can help organizations quickly respond to changing conditions, trends, and customer needs, and make more informed and timely decisions. Real-time data mining analysis and application can be used in various industries and applications, including financial services, healthcare, retail, manufacturing, and transportation (Kim et al, 2017).

II. Literature review

In this section the most reviewed articles are presented in the various fields with diverse applied method Starting from Twitter where publicly posted messages and geolocations may be accessed using the social networking tool Twitter (Bartels 2019). Twitter was mined for relevant data using data mining tools that looked for certain phrases and keywords. Using sentiment analysis and machine learning techniques, we were able to weed out tweets that did not pertain to our study. Maps of earthquakes and precipitation as well as a dashboard depicting the spread of the virus were made to illustrate the findings. To ensure that users could accurately comprehend and draw conclusions, a short survey-based research, eye and

DOI:http://doi.org/10.24086/cocos2024/paper.1091

mouse trackers, and human factors analysis approaches were used. Using API and keyword filtering, this research retrieved tweets on earthquakes, flu, and rain. Data mining methods uncovered patterns and crucial terms. Contextually abused terms were eliminated from sentiment analysis with the help of machine learning. For speedy decision-making, an application was developed that used Plotly to graphically display outcomes. There included an earthquake and rainfall map, a flu dashboard with symptoms and distribution data, and graphs showing hashtag use. Becker et al. (2018) used human factor approaches such as eye tracking and mouse tracking to measure user attention. Participant questionnaires were used to confirm the results, and the graphs were interpreted accurately.

▪ Data mining in smart city

Using cutting-edge innovation and smart infrastructure, "Smart Cities" have recently arisen as a game-changing answer to the problems associated with urbanization. Congestion in urban areas is a major problem for smart city transportation networks and the lives of city dwellers. Digital twins provide insights into traffic dynamics to help with this difficulty and make educated decisions. To depict city transportation systems, these digital twins combine analytics, simulation models, and real-time data. With an emphasis on data-driven decision-making for efficient congestion management, this study explores the use of Digital Twins, Trajectory Mining, Process Mining, and Decision Making to tackle the problem of urban traffic congestion. By better managing traffic for environmentally friendly urban transportation networks, this study hopes to aid in the development of Smart City transport systems as shown in (Bell et al., 2017).

▪ -Data Mining in industry

Production and quality management in particular are producing massive amounts of data as a result of the widespread use of Industry 4.0 technology. In (Boddington, P., 2017) the need of effective data visualization and analysis, which are fundamental in quality management for achieving a balance between tighter controls and faster delivery are highlighted. For better data visualization and analysis in the quality sector, the research suggests a framework that combines data mining with Augmented Reality (AR). Applying data mining and augmented reality together offers a novel approach to the problems of visualizing and analyzing unstructured data in low-technology, dynamic industries, as shown in a case study of the fashion industry. To compensate for the lack of integrated data mining and AR contributions to quality, the suggested approach improves decision-making via product-specific sampling criteria.

▪ A.I. in Real-Time Data Mining

According (Firth et al., 201) ML and AI have been recognized as powerful tools for data classification and pattern identification. By merging e-mental health strategies with AI and Big Data analytics, Intelligent Health (iHealth) signifies a watershed moment in the evolution of mental healthcare. In iHealth, data mining technologies integrate electronic health records (EHR), physician observations, ecological momentary assessment (EMA), patient self-monitoring, and EHR. Patient self-management is empowered, therapy is tailored to the individual, episodes are predicted and prevented, and risk assessment is improved. Apps and wearable sensors capture data in real-time, reducing the possibility of recall bias in retrospective self-reports. Still, problems with user interest remain, leading to short sessions and low overall utilization of self-monitoring tools.

▪ Data Mining in predicting the health of equipment

Various methods, including model-based, expert system, and data-driven techniques, are used in prognostics to predict the health of equipment. Because of the closely connected mathematical models, model-based techniques encounter difficulties when dealing with complex systems. In addition to being domain and equipment specific, expert systems depend on expert knowledge. Effective data-driven techniques for complex systems use both historical and real-time sensor data. By avoiding breakdowns and unplanned downtime and only replacing worn out components when absolutely required, predictive maintenance helps keep costs down. Reliable cost reduction requires careful sensor selection using data-driven entropy-based algorithms, expert knowledge, or both. Thirdly, there is application, followed by middleware, and finally observation, which make up the IoT architecture. A vehicle node (VN) uses a gateway and a J1939 network to represent the vehicle in the observation layer. According to Foley and Wollard (2019), the protocol that allows VNs to communicate with fleets is known as MQTT.

▪ Recent Research Findings

In order to be successful and make informed decisions, the data mining knowledge is significantly dependent on expert analysis. The literature emphasizes the need of include human knowledge into decision-making processes, which led to the suggestion of a field termed 'knowledge Mining.' The goal of these algorithms is to extract knowledge with less dependence on experts, and they are similar to data mining techniques (Abbe et al., 2016). This area of computer science is emerging in response to the petabytes of data being generated every day.

DOI:http://doi.org/10.24086/cocos2024/paper.1091

Its primary goal is to sift through past data in search of insights that might have been unavailable before. Data mining is great at finding unstructured information, but it might be much better if it took into account elements that weren't inherent in the discovery process. To fill this need, wisdom mining reduces the need for domain expertise by factoring in external influences. Working with relevant data and restricting the search field speeds up the process, which might lead to a decrease in the requirement for domain specialists in decision-making. One clear difference between data mining and wisdom mining is the latter's emphasis on closing the gap between uncovered information and its practical use (Bakker and Rickard, 2018). Data mining relies on expert analysis for effective decision-making, but the emergent domain of 'Wisdom Mining' proposes a parallel set of algorithms to extract wisdom with reduced reliance on experts. As data volumes reach petabytes globally, Wisdom Mining aims to analyze historical data for previously unknown knowledge. While data mining uncovers raw knowledge, Wisdom Mining addresses gaps by incorporating external influencing factors, potentially reducing the necessity for domain experts in decision-making. The relevant data is helpful in the process, emphasizing the boundary between data mining and the novel process of wisdom mining (Gianfrancesco et al, 2018).

III. Core Components

This emerging field of computer science involves delving into historical data to identify previously unknown yet critical decision-making data, according to Gheware et al. (2014) and Kiranmai and Damodaram (2014). It discovers hidden patterns, and more data enhances its results. Sharma (2014) and Venkatadri and Reddy (2011) define data mining (DM) as a broad area that combines machine learning, data visualization, database technology, and AI. According to Chen et al. (1996) and Gupta and Chandra (2019), organizations may profit from in-depth data studies to understand customer expectations and predict behavior and purchasing trends. Data is optimized, divided, associated, and classified by DM, which also finds latent data rules. In its multifaceted function, DM gathers and summarizes data. We switched from outdated technology to automated knowledge extraction to effectively extract essential information in real time. For class prediction during training, supervised learning employs labeled data and input-output pairings, according to Weiss and Indurkhya (1998) and Fu (1997) Unsupervised learning finds data patterns without labeled data. Many business fields have benefited from predictive and descriptive DM models. Predetermined buckets divide data into classes in the classification approach of supervised learning. DM research relies on recommendation and categorization to enhance real-world problem-solving solutions. Unsupervised clustering organizes data by similarities using partitioning, hierarchical, density-based, and grid-based methods. Association rule mining identifies common traits and produces rules using attribute-value criteria (Ponniah, 2001; Gupta & Chandra, 2019). Outlier analysis identifies dataset outliers. DM's ever-changing computer science trajectory provides interesting knowledge, but tackling practical problems is tough. Although DM results may not apply to the business world, engaging with domain professionals is vital for identifying practical solutions. 'Actionable knowledge' initially arose in business and the social sciences, but computer science swiftly adopted it. DM data must be polished by experts before it can be utilized to make educated judgments and conduct meaningful actions.

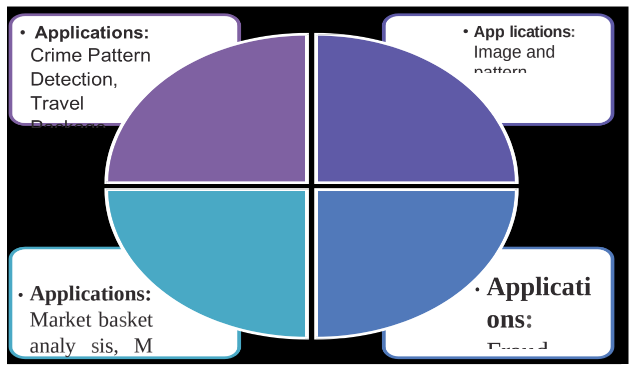

Figure 2.1: The resultant actionable knowledge is referred to as intelligent or actionable knowledge in the literature (Kassianos et al, 2017).

IV. Methods and Techniques in Data Mining

Data mining is an emerging field that uses massive amounts of data to uncover previously unknown insights. In order to make decisions, mining makes use of and processes publicly available data. It entails exploring models in large datasets using methods that combine machine learning, statistics, and database administration. Anomaly detection, data clustering, and the identification of out-of-the-ordinary records are all made easier with its help, as are related rules or dependencies. Data mining is starting to show up in more and more places where people live. This section delves into the several ways techniques if data mining are used to reveal intriguing shapes, which in turn assist educational, health and medical, and trade sectors. Data mining methods are used in several fields to uncover previously undisclosed information inside massive datasets. The approach most often used is support vector machines. A large amount of research has focused on cardiovascular illness as an application domain. The majority of the research that were analyzed found that data mining tasks or operational modes of solutions may be improved. While some researches have shown encouraging outcomes from testing algorithms, it's important to note that

DOI:http://doi.org/10.24086/cocos2024/paper.1091

these findings are specific to each experiment (Martinez and Kreitmair, 2018).

A. Applications

Data mining methods are applied in different aspects in real life. The most obvious ones are:

1. Healthcare

Through the analysis of unique health markers and the integration of scientific data, data mining in mental health aims to provide individualized treatment plans. The goal of researchers is to use patient answers and data aggregation to forecast results (Luxton et al., 2016). Clinical decision-making is improved by its extension to the prediction of the efficacy of surgical treatments, medical testing, and pharmaceuticals. Predictive results were obtained by using a dataset in physiotherapy and using approaches such as naïve Bayesian classifier and decision trees (Massoudi et al., 2019). Forkan et al. presented the CoCaMAAL Model for smart healthcare systems, which uses sensors to gather information linked to context. Using data mining findings in decision-making systems, they suggested a healthcare context-aware monitoring framework (Mittelstadt and Floridi, 2016). The many uses and benefits of data mining in healthcare are shown in this research.

2. Agriculture

Decisions involving the complicated web of variables that affect ultimate fish output, productivity, and profit margins are faced daily by animal agricultural sectors, fish farmers, and businesses. For example, in aquaculture, which is a complex system with interconnected physical, biological, and economic processes, aquatic creatures like fish, crabs, and mollusks are bred, reared, produced, and harvested. Despite the inherent dangers of aquaculture (Moore et al., 2016), smart data mining and ML solutions can improve facility management across the board, from feeding fish to controlling water quality to tracking fish welfare and biomass to detecting and preventing diseases and analyzing the quality of the final product. Machine learning algorithms can be integrated with fishing vessel identification systems and gear deployment systems to monitor and control fishing in real-time by tracing vessel movements, predicting fishing activity type, and quantifying fishing effort intensity (Moran et al., 2017).

3. Higher Education

A growing number of universities are using data mining to sift through mountains of student records. It helps educators make sense of student data by highlighting patterns in things like transfer credits, total credit hours, and skill sets across different groups. In an effort to improve the quality of user engagement, data mining tools in academic libraries often utilize regression and classification models. In order to monitor trends in library systems, "bibliomining" integrates data warehousing, data mining, and bibliometrics. Libraries are the center of attention because of their critical role in decision-making, but this is not new. Bibliomining, however, requires repeated application alongside other assessment techniques in order to provide a thorough report on library systems. When it comes to cardiovascular research, support vector machines rule the data mining world. Research shows that algorithms work better, but the results differ from trial to experiment, therefore it's important to evaluate the results with caution (Morch et al, 2020).

4. Cybersecurity

The massive data transmission that has resulted from the proliferation of digitalization has made it an easy target for cybercriminals who want to steal sensitive information. Databases and security logs may be effectively mined for data to reveal malware, breaches into systems and networks, insider assaults, and other security concerns. Even more impressively, some of these techniques can foresee assaults and reveal hidden dangers. Finding weak points or security holes in computer networks is the main goal of intrusion detection systems. This makes use of a security-agnostic strategy based on constant monitoring that triggers notifications in the event of a breach (Abbe et al., 2016). Another area that uses data mining to find suspicious or unusual activity is risk assessment and fraud. Cyberattack detection relies on cryptography and other methods that render data unintelligible to everyone except those in possession of the secret key. In order to detect intrusions, it is necessary to examine the gathered data. In order to improve upon more conventional forms of cybersecurity, such as firewalls, data mining is essential for detecting viruses, intruders, and fraud (Bakker and Rickard, 2018). Both descriptive and predictive database mining involve organizing information or making predictions based on patterns that have been seen. Figure 3 provides visual representations of six critical data mining methodologies for cybersecurity.

Classification: In cybersecurity, classification is employed to identify fraudulent and spam communications, distinguishing them from legitimate ones. This method aids in enhancing the accuracy of identifying and mitigating potential security threats.

Regression Analysis: This strategy involves creating a relationship model in the database between dependent and independent variables. It helps analyze variable changes, identify causes for fluctuations, and predict trends. In the realm of cybersecurity, regression analysis is commonly used to foresee potential cyber-attacks by assessing the relationships between various factors.

DOI:http://doi.org/10.24086/cocos2024/paper.1091

Time Series Analysis: Focused on mining datasets spanning multiple years, time series analysis provides insights into periodic activities. It is particularly useful for predicting potential security flaws and anticipating assaults during specific events, seasons, or times of the day in cybersecurity.

Association Rules Examination: Widely applied in data mining, this method uncovers hidden patterns and identifies correlations between variables frequently occurring together in databases. In cybersecurity, it aids in evaluating and predicting user behavior, monitoring network traffic, and detecting patterns of attacks.

Clustering: Clustering helps identify data items with similar characteristics, offering insights into the similarities and differences between variables. Unlike classification, clustering doesn't immediately arrange variables but assists in organizing and analyzing existing databases. It allows adjustments to models and generates subclusters without rewriting algorithms, facilitating database analysis in cybersecurity.

Summarization: The goal of summarization is to compile clear explanations of datasets, classifications, and clusters. In cybersecurity, summarization aids in understanding the core of data, minimizing manual searches, and generating reports. This method is crucial for presenting logs and findings in a comprehensible manner.

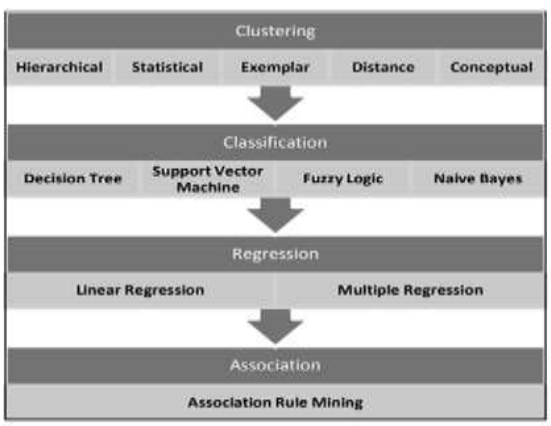

Figure 3.1: Data Mining Methods (Bakker and Rickard, 2018)

B. Challenges

Through data mining analysis and application some challenges arise such as:

1. **Data Volume and Velocity:**

Due to the large and varied nature of the data, which makes accuracy and completeness problematic, data quality assurance is a big issue in data mining. Inaccurate or skewed outcomes, caused by low-quality data, may affect both opportunities and decision-making. An example would be how skewed statistics from one company would not be representative of the whole. Organizations may address this by using strong data quality guidelines and adhering to data governance approaches. Addressing the data quality problem before analysis requires ethical and legal data gathering, as well as complete cleaning and verification (Berrouiguet et al, 2018).

2. **Data Quality**

Problems with privacy, security, and qualified personnel are issues with big data mining. Hackers target the huge amounts of personal or sensitive data acquired, which raises worries about privacy. An organization's credibility and legal position might take a hit if data protection standards are violated. Finding useful patterns in data that comes from a variety of sources isn't easy. Advanced analytics technologies are essential for organizations to conduct successful analyses. The scarcity of qualified big data analytics workers is another obstacle, as it has created a tight labor market and necessitated more funding for education and certification. Companies have a challenge in finding workers who possess the necessary combination of technical and business abilities to successfully manage big data. As a result, wage costs are high and training programs are resource expensive (Berrouiguet et al., 2019).

V. DISCUSSION AND COMPARISON

To discuss between data mining algorithms seven kinds of neural network are chosen by the researchers. Finding cancer is selected for predicting the performance for each algorithm. As inputs, seventeen critical variables are obtained from twenty-three. Data is divided into a test set and a train set, and the algorithm then compares the two sets' performance and chooses the best features. For the purpose of identifying unknown data types, model correctness, as measured by the effectiveness of test sample classification, is crucial. The research under-samples real-world breast cancer patient data in order to guarantee the development of essential features using a nested 5-fold cross-validation strategy for training and testing. For each round of cross-validation, 80 training datasets are generated by randomly selecting 20 control sets that match the case numbers. Each training phase is fine-tuned using five-fold inner cross-validation, which prevents machine learning analysis from going overboard. When evaluating performance, a confusion matrix is used to determine the efficiency, accuracy, sensitivity, and specificity according to predetermined standards. Recurrence prediction accuracy, true positive detection, true negative diagnosis, and overall accuracy are all very important in evaluating model performance, according to the research. The models are evaluated through applying the Receiver Operation Characteristic curve (ROC). As stated in Fiske et al(2019) ROC and Area Under the

DOI:http://doi.org/10.24086/cocos2024/paper.1091

Curve(AUC) are two curves to measure the performance of overall algorithm. Additionally, the F1 score, considering both Precision (p) and Recall (r), assesses the accuracy of binary classification tests. Moreover, F1 score is a symmetric mean of precision and recall, ranges from 0 to 1, where 1 indicating desirable precision and recall. In addition, the proportion between correct positive results to all positive results is a precision, in other hand the ratio of accurate positive results to all relevant samples is considered as recall.

Table 5.1 . Performance measure (Fiske et al 2019).

Method	Random Forest	LVQ	Bayesian	C5.0	MLP	KPCA-SVM	SVM
TP	1750	1640	1650	2188	1477	2048	1750
TN	2188	2024	2008	2297	1914	2250	2243
FP	985	931	758	657	657	765	985
FN	548	876	1208	329	1423	408	493
Accuracy	0.719	0.669	0.650	0.819	0.619	0.785	0.729
Sensitivity	0.761	0.651	0.577	0.869	0.509	0.833	0.780
Specificity	0.689	0.684	0.725	0.777	0.744	0.746	0.694
The Geometric mean of sensitivity and specificity	0.724	0.668	0.647	0.822	0.615	0.788	0.736
PPV	0.639	0.637	0.685	0.769	0.692	0.728	0.639
NPV	0.799	0.697	0.624	0.874	0.573	0.846	0.819
The Geometric mean of PPV and NPV	0.715	0.667	0.654	0.820	0.630	0.785	0.724
F-measure	0.695	0.644	0.626	0.816	0.586	0.777	0.703
The area under ROC curve	0.729	0.632	0.692	0.763	0.625	0.774	0.742

In the research Fiske et al. (2019) looks at the possibility of using data mining algorithms to help in detecting cancer especially in breast. Detection and preventing recurrence, which might save money and lessen the amount of stress that goes into the process. Out of seven algorithms tested (MLP, LVQ, Bayesian Neural Network, C5.0, KPCA-SVM) as shown in above table. C5.0 and KPCA-SVM both performed very well, outperforming the others in terms of accuracy, specificity, sensitivity, PPV, NPV, F-measure, and AUC, ROC curves. The research highlights the need of high sensitivity, particularly for predicting breast cancer recurrence, whereas MLP, Bayesian, and LVQ have lesser sensitivity. When it comes to improving diagnosis accuracy and early intervention, C5.0 shows the greatest sensitivity result, suggesting it might be a beneficial tool for clinicians. Finally, data mining algorithms, namely C5.0 and KPCA-SVM, play a crucial role in enhancing healthcare efficiency and patient outcomes in the treatment of breast cancer.

VI. CONCLUSION

In conclusion, Real-Time Data Mining analysis has emerged as a pivotal tool, transforming various fields through its immediate and dynamic insights. Its applications span diverse domains, including healthcare, urban management, and predictive modeling. The integration of Artificial Intelligence and Machine Learning has further augmented its capabilities, allowing for rapid decision-making and enhanced precision. The healthcare sector benefits from real-time patient monitoring, enabling early diagnosis and personalized treatment. In smart cities, real-time data analytics facilitates efficient urban management, addressing challenges posed by demographic shifts. The continuous evolution of big data, coupled with advanced analytics, underscores the significance of real-time data mining in uncovering valuable patterns and trends. Despite its transformative potential, user engagement remains a challenge in some applications, necessitating ongoing efforts to enhance usability. The future holds promise for real-time data mining, with its capacity to revolutionize decision-making processes and foster innovation across various industries.

REFERENCES

Abbe, A., Grouin, C., Zweigenbaum, P., et al., 2016. 'Text mining applications in psychiatry: a systematic literature review.' *Int. J. Methods Psychiatr. Res.,* 25(2), 86–100.

Bakker, D., Rickard, N., 2018. 'Engagement in mobile phone app for self-monitoring of emotional well-being predicts changes in mental health: MoodPrism.' *J. Affect. Disord.,* 227, 432–442.

Bartels, S.L., van Knippenberg, R.J.M., Dassen, F.C.M., et al., 2019. 'A narrative synthesis systematic review of digital self-monitoring interventions for middle-aged and older adults.' *Internet Interv.,* 18, 100283.

Becker, D., van Bred, W., Funk, B., et al., 2018. 'Predictive modeling in e-mental health: a common language framework.' *Internet Interv.,* 12, 57–67.

Bell, I.C.H., Lim, M.H., Rossell, S.L., et al., 2017. 'Ecological momentary assessment and intervention in the treatment of psychotic disorders: a systematic review.' *Psychiatr. Serv.,* 68, 1172–1181.

Berrouiguet, S., Billot, R., Larsen, M.E., et al., 2019. 'An approach for data mining of electronic health record data for suicide risk management: database analysis for clinical decision support.' *JMIR Ment. Health,* 6(5), e9766. https://doi.org/10.2196/mental.9766.

Berrouiguet, S., Perez-Rodriguez, M.M., Larsen, M., et al., 2018. 'From eHealth to iHealth: transition to participatory and personalized medicine in mental health.' *JMIR,* 1, e2.

https://doi.org/10.2196/jmir.7412.

Boddington, P., 2017. 'Towards a Code of Ethics for Artificial Intelligence.' Springer International Publishing, Cham.

Bos, F.M., Schoevers, R.B., aan het Rot, M., 2015. 'Experience sampling and ecological momentary assessment studies in psychopharmacology: a systematic review.' *Eur. Neuropsychopharmacol.,* 25, 1853–1864.

Briffault, X., Morgiève, M., Courtet, P., 2018. 'From e-health to i-health: prospective reflections on the use of intelligent systems in mental health care.' *Brain Sci.,* 8, 98. https://doi.org/10.3390/brainsci8060098.

Bulao, J., 2021. How much data is created every day in 2021. *TechJury Blog*.

Carr, S., 2020. ''AI gone mental': engagement and ethics in data-driven technology for mental health.' *J. Ment. Health,* 29(2), 125–130.

Challen, R., Denny, J., Pitt, M., et al., 2019. 'Artificial intelligence, bias and clinical safety.' *BMJ Qual. Saf.,* 28(3), 231–237.

Chandra, P., & Gupta, M.K. (2018) 'Comprehensive survey on data warehousing research', International Journal of Information Technology, 10(2), pp. 217–224.

Chen, M., Han, J., & Yu, P.S. (1996) 'Data mining: an overview from a database perspective', IEEE Transactions on Knowledge and Data Engineering, 8(6), pp. 866–883.

de Laat, P.B., 2019. 'The disciplinary power of predictive algorithms: a Foucauldian perspective.' *Ethics Inf. Technol.,* 21, 319–329.

Dillard-Wright, Jessica, 2019. 'Electronic health record as a panopticon: A disciplinary apparatus in nursing practice.' *Nurs. Philos.,* 20, e12239. https://doi.org/10.1111/nup.12239.

Dipnall, J., Pasco, J.A., Berk, M., et al., 2016. 'Fusing data mining, machine learning and traditional statistics to detect biomarkers associated with depression.' *PLoS ONE,* 11(2), e0148195. https://doi.org/10.1371/journal.pone.0148195.

Dogan, E., Sander, S., Wagner, X., 2017. 'Smartphone-based monitoring of objective and subjective data in affective disorders: where are we and where are we going? Systematic review.' *J. Med. Internet Res.,* 19(7), e262. https://doi.org/10.2196/jmir.7006.

Engel, S.G., Crosby, R.D., Thomas, G., 2016. 'Ecological momentary assessment in eating disorder and obesity research: a review of the recent literature.' *Curr. Psychiatry Rep.,* 18, 37.

Faurholt-Jepsen, M., Munkholm, K., Frost, M., et al., 2016. 'Electronic self-monitoring of mood using IT platforms in adult patients with bipolar disorder: a systematic review of the validity and evidence.' *BMC Psychiatry,* 16, 7. https://doi.org/10.1186/s12888-016-0713-0.

Firth, J., Torous, J., Nicholas, J., et al., 2017. 'The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials.' *World Psychiatry,* 16, 287–298.

Firth, J., Torous, J., Nicholas, J., et al., 2017. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 16, 287–298.

Fiske, A., Henningsen, P., Buyx, A., 2019. 'Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy.' *J. Med. Internet Res.,* 21(5), e13216. https://doi.org/10.2196/13216.

Foley, T., Wollard, J., 2019. 'The Digital Future of Mental Healthcare and Its Workforce: A Report on a Mental Health Stakeholder Engagement to Inform the Topol Review.' *Health Education England, London.*

Fu, Y. (1997) 'Data mining: tasks, techniques, and applications', IEEE Potentials, 16(4), pp. 18–20.

Gartner. (2023). Gartner glossary: Big data. Retrieved from https://www.gartner.com/en/informationtechnology/glossary/big-data

Gheware, S.D., Kejkar, A.S., & Tondare, S.M. (2014) 'Data mining: tasks, tools, techniques and applications', International Journal of Advanced Research in Computer Communications and Engineering, 3(10), pp. 8095–8098.

Gianfrancesco, M.A., Tamang, S., Yazdany, J., et al., 2018. 'Potential biases in machine learning algorithms using electronic health record data.' *JAMA Intern. Med.,* 178(11), 1544–1547.

Gongora´ Alonso, S., de la Torre-Díez, I., Hamrioui, S., et al. (2018) 'Data mining algorithms and techniques in mental health: a systematic review', *Journal of Medical Systems*, 42, 161.

Gupta, M.K., & Chandra, P. (2019) 'A comparative study of clustering algorithms', In: Proceedings of the 13th INDIACom-2019; IEEE Conference ID: 461816; 6th International

Conference on Computing for Sustainable Global Development.

Hartmann, R., Sander, S., Lorenz, N. (2019) 'Utilization of patient-generated data collected through mobile devices. Insights from a survey on attitudes toward mobile self-monitoring and self-management apps for depression', *JMIR Mental Health*, 6(4), e11671. DOI: 10.2196/11671.

Internet Live Stats. (2023). Google search statistics. Retrieved from https://www.internetlivestats.com/ google-search-statistics

Iorfino, F., Occhipinti, J.-A., Skinner, A., et al. (2021) 'The impact of technology-enabled care coordination in a complex mental health system: a local system dynamics model', *Journal of Medical Internet Research*, 23(6), e25331. DOI: 10.2196/25331.

Kassianos, A.P., Georgiou, G., Papaconstantinou, E.P., et al. (2017) 'Smartphone applications for educating and helping non-motivating patients adhere to medication that treats mental health conditions: aims and functioning', *Frontiers in Psychology*, 8, 1769. DOI: 10.3389/fpsyg.2017.01769.

Kim, H., Bell, E., Kim, J., et al. (2017) 'iCONCUR: informed consent for clinical data and bio-sample use for research', *Journal of the American Medical Informatics Association*, 24(2), 380–387.

Kiranmai, B., & Damodaram, A. (2014) 'A review on evaluation measures for data mining tasks', International Journal of Engineering and Computer Science, 3(7), pp. 7217–7220.

Li, H., Mukherjee, D., Basappa Krishnamurthy, V., et al. (2019) 'Use of ecological momentary assessment to detect variability in mood, sleep and stress in bipolar disorder', *BMC Research Notes*, 12, 791. DOI: 10.1186/s13104-019-4834-7.

Luxton, D.D., Anderson, S.L., Anderson, M., 2016. 'Ethical issues and artificial intelligence technologies in behavioral and mental health care.' In: Luxton, D.D. (Ed.), Artificial Intelligence in Behavioral and Mental Health Care. Academic Press, London, pp. 255–276.

Martinez-Martin, N., Kreitmair, K., 2018. 'Ethical issues for direct-to-consumer digital psychotherapy apps: addressing accountability, data protection, and consent.' JMIR Ment. Health, 5(2), e32. https://doi.org/10.2196/mental.9423.

Massoudi, B., Holvast, F., Bockting, C.L.H., et al., 2019. 'The effectiveness and cost-effectiveness of e-health interventions for depression and anxiety in primary care: a systematic review and meta-analysis.' J. Affect. Disord., 245, 728–743.

McKinsey & Company. (2016). Getting big impact from big data. Retrieved from https://www.mckinsey. com/capabilities/mckinsey-digital/our-insights/getting-big-impact-from-big-data

Mittelstadt, B.D., Floridi, L., 2016. 'The ethics of big data: current and foreseeable issues.' Sci. Eng. Ethics, 22, 303–341.

Moore, R.C., Depp, C.A., Loebach Wetherell, J., et al., 2016. 'Ecological momentary assessment versus standard assessment instruments for measuring mindfulness, depressed mood, and anxiety among older adults.' J. Psychiatr. Res., 75, 116–123.

Moran, E.K., Culbreth, A.J., Barch, D.M., 2017. 'Ecological momentary assessment of negative symptoms in schizophrenia: relationships to effort based decision making and reinforcement learning.' J. Abnorm. Psychol., 126(1), 96–105.

Morch, ¨ C.M., Gupta, A., Misharad, B.L., 2020. 'Canada protocol: an ethical checklist for the use of artificial intelligence in suicide prevention and mental health.' Artif. Intell. Med., 108, 101934. https://doi.org/10.1016/j.artmed.2020.101934.

Ponniah, P. (2001) Data Warehousing Fundamentals. Wiley, USA.

Sander, L., Rausch, L., Baumeister, H., 2016. Effectiveness of internet-based interventions for the prevention of mental disorders: a systematic review and meta analysis. JMIR Ment. Health 3, e38.

Sharma, M. (2014) 'Data mining: a literature survey', International Journal of Emerging Research in Management & Technology, 3(2), pp. 1–4.

Taylor, P. (2022, October 19). Big data - Statistics & Facts. Statista. Retrieved from https://www.statista. com/topics/1464/big-data/#topicHeader__wrapper

Venkatadri, M., & Reddy, L.C. (2011) 'A review on data mining from past to the future', International Journal of Computer Applications, 15(7), pp. 19–22.

Weiss, S.H., & Indurkhya, N. (1998) Predictive Data Mining: A Practical Guide. Morgan Kaufmann Publishers, San Francisco.

DOI:http://doi.org/10.24086/cocos2024/paper.1091