A REVIEW ON ADVANCING EDUCATION: MACHINE LEARNING IN TEACHING EVALUATION

[1]Divya, [2]Dr. Shrwan Ram

[1]M.E. Scholar, Department of Computer Science and Engineering,M.B.M. University, Jodhpur
[2]Professor & HOD, Department of Computer Science and Engineering,M.B.M. University, Jodhpur
Email-: Divyavarde1901@gmail.com

* Corresponding Author: Divya

Abstract: *This review paper explores the integration of machine learning and data mining techniques in analyzing Student Evaluations of Teaching (SET) within higher educational institutions. It delves into how SET, an essential tool for assessing instructor competence and course effectiveness, benefits from the advanced analytical capabilities of these technologies. The paper highlights the transformation of raw educational data into insightful analyses, covering aspects like teaching approaches, resource adequacy, and overall course satisfaction. It emphasizes the shift from traditional evaluation methods to data-driven decision-making, underscoring the role of machine learning in predicting trends and data mining in extracting valuable insights. Additionally, the paper discusses the challenges faced in web mining and the potential solutions, providing a comprehensive overview of the current landscape and future possibilities in leveraging technology for educational advancement.*

Keywords: *Student Evaluations of Teaching, Machine Learning, Data Mining, Higher Education, Educational Technology, Instructor Assessment, Course Evaluation, Web Mining.*

I. INTRODUCTION

Student evaluations of teaching (SET) is a widely utilized method on higher educational institutions to gather information about the professional competence of instructors. Student evaluations may contain questions regarding teaching approaches, assessment, resources, and administration, such as, the sufficiency of the course content, competence and qualifications of instructor, and adequacy of handouts. Evaluation forms can provide potentially essential source of information to achieve quality objectives of educational institutions. [1] Several Machine Learning algorithms are applied over the dataset for retrieving results. Generally data mining and machine learning are melded with each other. The point that makes Data Mining and Machine Learning different from each other includes the type of dataset on which the technique is applied along with its interpretations. The technique of Machine learning helps in to prognosticate upcoming incidents which are based on combinations of patterns, whereas data mining proceeds as a informant of information from machine learning to withdraw data [2].

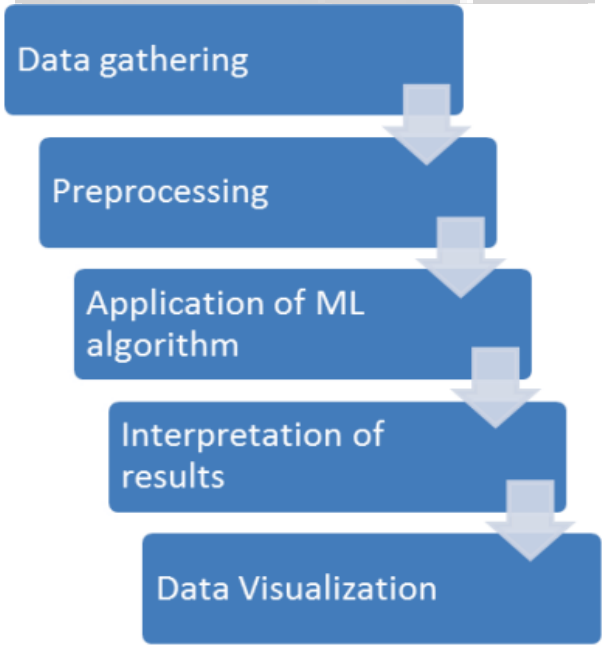

Figure 1 Process of machine learning in data mining

According to the Fig. 1, the process followed to interpret the data obtained from the online reviews. Its initial phase carries the gathering of data and pre-processing which includes collection of data and removal of outliers. The subsequent phase

includes applying Machine Learning algorithm on the new dataset generated from the initial phase. The algorithms used for classifying the overall class from the review parameters are Random Forest Algorithm and Decision Tree. Based on the outcomes generated by the algorithms applied the results are interpreted and visualized for better understanding and decision making.

Web Usage Mining Web usage mining performs mining on web usage data, or web logs. Web logs can be regarded as a sequence of web pages accessed by users. Sometimes it is referred to as clickstream data because each entry corresponds to a mouse click in a client-side browser. Web usage data includes data from web server logs, proxy server logs, browser logs, user profiles, registration data, cookies, user queries, bookmark data, mouse clicks and scroll, and any other data as the results of user interactions on the Web. Web usage mining consists of three phases, namely preprocessing, pattern discovery, and pattern analysis [3].

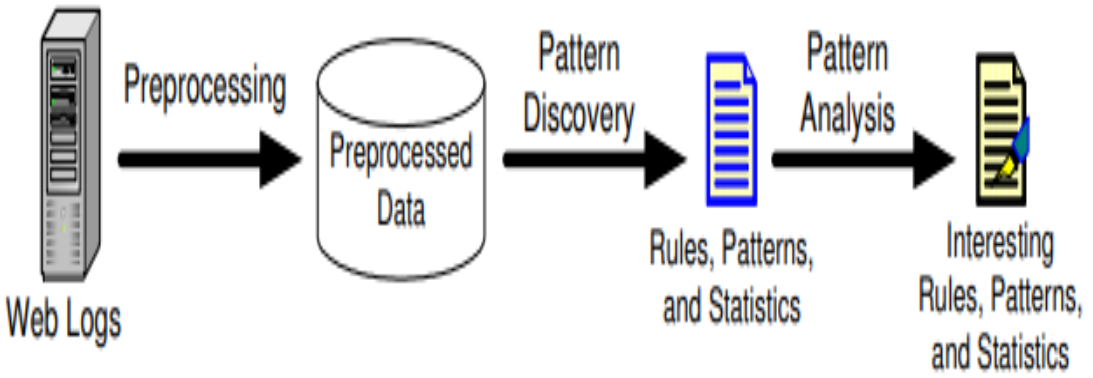

Figure 2 Web usage mining

Data Mining techniques and algorithms have been used on a large scale in almost all the sectors which range from computer science, manufacturing industry, and the healthcare industry. Just as environmental changes witnessed every year, every business and service sector change now and then. Considering the traditional and modern practices of the education, it is evident that an enormous change has happened to the teaching learning process, nature of students coming into education, expected outcomes of education, the urgency of education, the lifestyle of teacher and student, industrial relevance of education and many more [4].

Machine learning, a subset of artificial intelligence, encompasses a variety of types and techniques. Each type is suited for different purposes and data sets. Here's an overview of the main types of machine learning.

Types of Machine learning

- Supervised,
- Unsupervised and Reinforcement
- Semi-Supervised Learning
- Reinforcement Learning
- Deep Learning
- Transfer Learning
- Ensemble Learning

II. DATA MINING TECHNIQUES

Data mining techniques are used to operate on large volumes of data to discover hidden patterns and relationships helpful for decision-making. Some popular algorithms and techniques such as Classification, Clustering, Prediction, Decision Trees, Nearest Neighbor method, social Network Analysis, etc. are explained here [4].

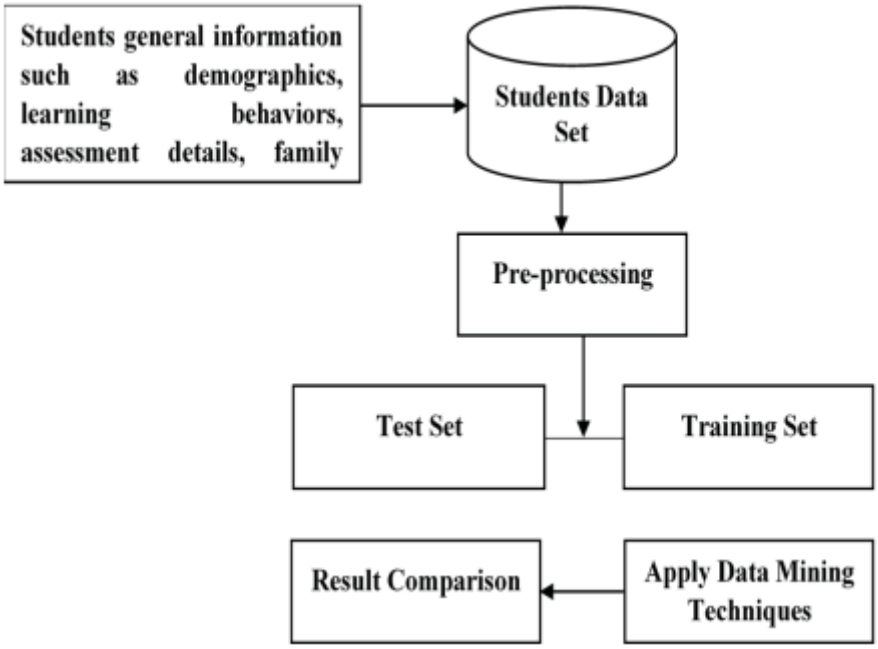

Figure 3 Flow Chart of Student Performance Analysis Model based on Data Mining

As a scientific field, physical education discusses the application of scientific principles and techniques to improve sports performance. Since the relationships between sports results and various data elements are directly affected by several factors such as type of sports, the environment, and the objectives of players, several methods have been suggested to predict the results based on available data. More precisely, while some teams prefer not to use any prediction techniques, others have long depended on either the experience and instincts of the experts (with high error rate) or historical data. Encompasses a broad range of methods and strategies used for extracting meaningful patterns, trends, and insights from large datasets. These techniques are crucial in transforming raw data into valuable information, aiding in decision-making, predicting future trends, and uncovering hidden patterns. Some key techniques in data mining include.

Classification: This is a fundamental technique where data is categorized into predefined classes. It is widely used in applications like email filtering (spam vs. non-spam), loan approval (safe vs. risky), and patient diagnosis (diseased vs. healthy).

Clustering: Unlike classification, clustering groups data points based on their similarity without predefined classes. This is crucial in market segmentation, social network analysis, and organizing large datasets into meaningful categories.

Association Rule Mining: This technique discovers interesting relationships between variables in large databases. A classic example is the market basket analysis in retail, which finds associations between different products bought together, helping in cross-selling and upselling strategies.

Data Visualization: Although it is not strictly a data mining technique, visualization plays a crucial role in the representation of data mining outcomes. By presenting results in a visual format, it significantly enhances the comprehension and interpretation of complex data sets.

III. LITERATURE REVIEW

Umamaheswari and Harikumar [5] proposed an enhanced Classification algorithm, achieving 98.8% accuracy compared to other models and the prevailing SVM-RBF classifier. Social media platforms serve as spaces where individuals share opinions on various topics such as movies, products, fashion, politics, and technology. Platforms like Amazon, Flipkart, and Snapdeal host user reviews. This study introduces a framework for deriving opinion scores and forming insights.

Gopi et al. [6] extracted data from websites, processing it into structured forms using advanced web scraping tools. Their proposed model excelled in accurately categorizing documents from the gathered data. The proposed LR-RFE-BPNN model distinguished the data with an accuracy of 94.63%, outperforming other benchmarked text classification methods.

Karthikeyan et al. [7] proposed a unique weblog-based structure for dynamic online session identification. For a weblog with 1 million log entries, the MAPE was 11%. This framework was superior to traditional models using Artificial Neural Networks (ANN) with a value of 0.0639 and achieved an hourly RMSE of 0.0500.

Sowmya et al. [8] provided a comparative evaluation of various machine learning techniques targeting the prediction of customer buying behaviors. They employed supervised classification machine learning methods in their experiments. Their

proposed classification method achieved an accuracy of 92.402%, marking it as the most effective classifier. Their results demonstrated optimal accuracy with extensive features as the KNN reduced errors when combined with the SGD.

Chaubey et al. [9] assessed their proposed model's performance using multiple performance metrics. Additionally, their web-based recommendation model's accuracy was enhanced using the KNN classification strategy, achieving a recommendation accuracy of 29.5%.

Nigam et al. [10] developed a predictive model using machine learning to forecast the final grades of undergraduate students based on their midterm scores. The data included academic records of 1,854 students enrolled in the Turkish Language-I course at a renowned Turkish university during the 2019-2020 academic year's fall term. The model showcased a classification accuracy between 70% and 75%. The research aimed to identify the best machine learning techniques and detect students at risk of not meeting academic standards at an early stage.

Yağcı et al. [11] examined the relationship between production rules in manufacturing lines using association rules. They applied logical reasoning to address the interplay of rules between automotive production lines and products. The findings revealed that the application of data mining association rules achieved an accuracy rate of over 87%. This approach is valuable for formulating manufacturing guidelines, enhancing management decisions in the realm of IoT, and streamlining the production process.

Wang et al. [12] examined the accuracy, precision, and runtime of Naive Bayes, K-Nearest Neighbor (KNN), and Random Forest classifiers for CKD prediction within the healthcare domain. Data processing refers to extracting insights from extensive datasets. The significant role of machine learning, especially in predicting chronic diseases, is emphasized in their study. Consequently, the Random Forest classifier emerged as superior to both Naive Bayes and KNN.

Devika et al. [13] forecasted stock price trends using a combination of a complex network approach and machine learning. They first presented a novel method for constructing pattern networks for multivariate stock time series data. They then applied KNN and SVM algorithms to predict the next day's volatility patterns for specific stocks, using the topological characteristics of each symbolic pattern combination. The results revealed that cross-validation and search strategies can determine the most suitable models for the two algorithms. Regarding the test dataset, the prediction accuracy rates for three indices exceeded 70%. Among the two, the SVM algorithm typically outperformed the KNN in predictions.

IV. CHALLENGES IN IN WEB MINING

Web content mining, while valuable, faces certain challenges and issues. However, for each of these challenges, there are potential solutions some of these include Fig.

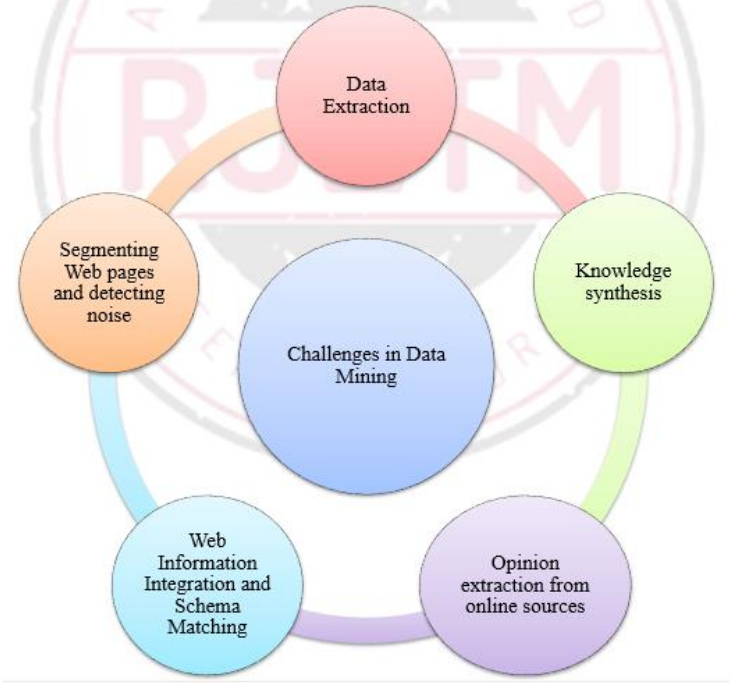

Figure 4 Challenges in Pattern and prediction analysis in web usage data [14]

- Data Extraction: Extracting structured data, such as products and search results, out of web pages. Such data may be extracted to enable service provision. To overcome this issue, machine learning and automated extraction are the two primary kinds of methodologies employed.

- Web Information Integration and Schema Matching: The Web is filled with data, yet each website (or even page) displays the same data differently. Finding or matching data that is semantically similar is a significant issue with several practical applications.
- Extracting opinions from internet sources: There are various places to find opinions online, including forums, blogs, chat rooms, and consumer reviews of goods. For marketing intelligence and product benchmarking, mining opinions is crucial.
- Synthesis of knowledge: Concept hierarchies or ontologies are helpful in a variety of applications. However, manually producing them takes a lot of time. To provide the user with a comprehensive understanding of the issue area, the key application is to organize and synthesize the bits of information found on the web. We'll outline a few current techniques for investigating information redundancy on the web.
- Segmenting Web sites and identifying noise: In many Web applications, one just needs the primary content of the Web page without adverts, navigation links, and copyright notices. It's an intriguing challenge to automatically partition Web pages and extract the key material.

V. CONCLUSION

The integration of machine learning and data mining in analyzing Student Evaluations of Teaching marks a significant advancement in educational assessment methods. By harnessing these technologies, higher educational institutions can transform traditional SET data into a rich source of insights, aiding in the enhancement of teaching quality and course design. The ability of machine learning algorithms to predict trends and data mining techniques to uncover hidden patterns offers a more nuanced understanding of student feedback, leading to informed decision-making. However, the application of these technologies comes with challenges, particularly in web mining, necessitating innovative solutions to ensure accuracy and efficiency. As the educational landscape evolves, the continued development and application of these technologies will be crucial in shaping a more data-informed, responsive, and effective educational system.

REFERENCE

[1] Onan, A. “Mining opinions from instructor evaluation reviews: a deep learning approach. Computer Applications in Engineering Education,” vol. 28(1), pp. 117-138, 2018.

[2] Shina, S. Sharma and A. Singla, "A Study of Tree Based Machine Learning Techniques for Restaurant Reviews," 2018 4th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 2018, pp. 1-4, doi: 10.1109/CCAA.2018.8777649.

[3] Baoyao, Z.. “Intelligent Web Usage Mining. Nanyang Technological University, Division of Information Systems,” School of Computer Engineering, vol. 94, 2004.

[4] C. C. Ukwuoma, C. Bo, I. A. Chikwendu and E. Bondzie-Selby, "Performance Analysis of Students Based on Data Mining Techniques: A Literature Review," *2019 4th Technology Innovation Management and Engineering Science International Conference (TIMES-iCON)*, Bangkok, Thailand, 2019, pp. 1-5, doi: 10.1109/TIMES-iCON47539.2019.9024396.

[5] S. Umamaheswari and K. Harikumar, “Analyzing product usage based on twitter users based on datamining process,” Proc. Int. Conf. Comput. Autom. Knowl. Manag. ICCAKM 2020, pp. 426–430, Jan. 2020, doi: 10.1109/ICCAKM46823.2020.9051488.

[6] A. P. Gopi, R. N. S. Jyothi, V. L. Narayana, and K. S. Sandeep, “Classification of tweets data based on polarity using improved RBF kernel of SVM,” Int. J. Inf. Technol., vol. 15, no. 2, pp. 965–980, Feb. 2020, doi: 10.1007/S41870-019-00409-4/METRICS.

[7] Karthikeyan, T., Karthik Sekaran, D. Ranjith, and J. M. Balajee. "Personalized content extraction and text classification using effective web scraping techniques." International Journal of Web Portals (IJWP) 11, no. 2 (2019): 41-52.

[8] H. K. Sowmya and R. J. Anandhi, “An efficient and scalable dynamic session identification framework for web usage mining,” Int. J. Inf. Technol., vol. 14, no. 3, pp. 1515–1523, May 2022, doi: 10.1007/S41870-022-00867-3.

[9] G. Chaubey, P. R. Gavhane, D. Bisen, and S. K. Arjaria, “Customer purchasing behavior prediction using machine learning classification techniques,” J. Ambient Intell. Humaniz. Comput., pp. 1–25, Apr. 2022, doi: 10.1007/S12652-022-03837-6/METRICS.

[10] C. Nigam and A. K. Sharma, “WITHDRAWN: Experimental performance analysis of web recommendation model in web usage mining using KNN page ranking classification approach,” Mater. Today Proc., Oct. 2020, doi: 10.1016/J.MATPR.2020.09.364.

[11] Yağcı, Mustafa. "Educational data mining: prediction of students' academic performance using machine learning algorithms." Smart Learning Environments 9, no. 1 (2022): 11.

[12] L. Wang, B. Lin, R. Chen, and K. H. Lu, “Using data mining methods to develop manufacturing production rule in IoT environment,” J. Supercomput., vol. 78, no. 3, pp. 4526–4549, Feb. 2022, doi: 10.1007/S11227-021-04034-6/METRICS.

[13] R. Devika, S. V. Avilala, and V. Subramaniyaswamy, “Comparative study of classifier for chronic kidney disease prediction using naive bayes, KNN and random forest,” Proc. 3rd Int. Conf. Comput. Methodol. Commun. ICCMC 2019, pp. 679–684, Mar. 2019, doi: 10.1109/ICCMC.2019.8819654.

[14] P. Verma and N. Kesswani, "FEDUS : A comprehensive algorithm for web usage mining," https://doi.org/10.1080/02522667.2019.1616912, vol. 41, no. 3, pp. 835–854, Apr. 2019, doi: 10.1080/02522667.2019.1616912.