Machine Learning Techniques for Analysing Students Feedback Towards Quality Management in Higher Education

Shaifali Garg[1,*][a], Malik Jawarneh[2][b], Meenakshi[3,†][c] and Sammy F.[4,†][d]
[1]*Amity Business School, Amity University Madhya Pradesh, India*
[2]*Faculty of Computing Sciences, Gulf College, Muscat, Oman*
[3]*Apeejay Stya University Sohna, Haryana, India*
[4]*Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India*

Keywords: Educational Data Mining, Student Feedback Analysis, Machine Learning, SVM, Accuracy, Classification.

Abstract: The area of research referred to as educational data mining is one that makes use of data mining, machine learning, and statistics in order to investigate material that has been especially obtained from educational settings. The goal of the learning and teaching process is to provide pupils the best possible experience they can have in terms of learning and comprehending the material being taught. Educational data mining can be used for a variety of purposes, including predicting student performance and identifying students who are at risk, determining important concerns in the learning patterns of various groups of students, increasing pass-out rates, accurately assessing the performance of the institution, making the most of campus resources, and optimising the renewal of subject curriculum. This article provides machine learning techniques for analysing students feedback towards quality management in higher education. Student feedback data set is preprocessed to remove noise. Then student feedback data is analysed using SVM, ANN and random Forest algorithm. Performance of SVM algorithm is found better for analyzing student feedback data for overall quality improvement in higher educational institutions.

1 INTRODUCTION

Data mining finds patterns and connections across multiple data categories to extract meaningful information from large databases. This is done through pattern detection and exploration. Predictive data mining analyzes existing data to forecast the future. Machine learning is an area of artificial intelligence that studies ways to teach machines new skills. The field of "educational data mining" uses data mining, machine learning, and statistics to study educational data. This area offers great data mining possibilities. The "teaching-learning process" refers to the system that uses factual data and scientific criteria to evaluate student education (Veluri et al, 2022). Educational data mining involves researching and analyzing growing volumes of data from educational institutions and settings. The data may include school administrative or online education data. Both may be included. Educational data mining is expanding rapidly as new data mining and machine learning techniques and methods are developed. Data mining has enabled the development of unique methods for extracting creative, interesting, interpretable, and relevant information that can improve our understanding of students and their learning environments (Zhang et al, 2020).

The anomalies in the data may reveal important patterns in class efficiency or student academic progress. Multiple interdependencies between variables may reveal substantial correlations and regressions. When analyzing educational data,

[a] https://orcid.org/0000-0002-5647-3347
[b] https://orcid.org/0000-0001-6894-2756
[c] https://orcid.org/0000-0002-4175-0508
[d] https://orcid.org/0000-0001-5756-8439
[*] Associate Professor
[†] Assistant Professor

Garg, S., Jawarneh, M., Meenakshi, . and F., S.
Machine Learning Techniques for Analysing Students Feedback Towards Quality Management in Higher Education.
DOI: 10.5220/0012615200003739

In *Proceedings of the 1st International Conference on Artificial Intelligence for Internet of Things: Accelerating Innovation in Industry and Consumer Electronics (AI4IoT 2023)*, pages 309-313
ISBN: 978-989-758-661-3

context, time, and sequencing are crucial. Educational data mining can predict student performance, identify at-risk students, determine important concerns in the learning patterns of various student groups, increase pass-out rates, accurately assess the institution's performance, maximize campus resources, and optimize subject curriculum renewal (Hicham et al, 2020).

A thorough quality management system helps universities make necessary changes to combat process entropy and provides critical feedback for continuous progress. This ensures high-quality university services. Keeping note of and investigating the myriad ideas and emotions that arise throughout teaching and learning is crucial. These feed-forward mechanisms control deviations and introduce appropriate interventions at the right times to ensure smooth and effective teaching and learning growth. Make-up classes, bridge courses, and extra homework may be needed to maintain and improve university education (Kovalev et al, 2020).

Learning and teaching aim to give students the best experience possible in learning and understanding the topic. Data and information for each activity, such as teaching and learning, is crucial to comprehensive quality management in higher education. Empirical evidence should support TQM rather than subjective assessments. Intelligent data analysis, classification, and prediction offered by machine learning may improve this issue (Khodeir et al, 2019).

This article uses machine learning to analyze student comments on higher education quality management. Preprocessed student feedback data removes noise. SVM, ANN, and random Forest algorithms analyze student feedback. SVM algorithm performs better for assessing student feedback data to improve higher education quality.

2 LITERATURE SURVEY

The best decision tree algorithm, SVM, C4.5, Naive Bayesian, and RIPPER prediction algorithms were compared (Eswara et al, 2017). When FP rate, Precision, F-M, Recall, and MCC are compared, Naive Bayes wins. It is unclear if or how these algorithms can improve college instruction.

Researcher employed decision tree categorization on student evaluation results to improve teaching and learning quality. They sought to identify students at risk of poor performance. The technique fails to identify student strengths and weaknesses, creating a knowledge gap (Mesaric et al, 2016). The feedback will be excellent or bad based on how well the person can identify the lesson or understanding that needs improvement. A study on the reliability of student feedback ratings or quantitative characteristics used linguistic qualities of the accompanying free text in feature space (Kannan et al, 2011). A stronger awareness for textual evidence leads to higher marks. The quantitative ratings and qualitative remarks regarding the feature are compared. Naive Bayes was used to classify Gujarati texts into a few main groups (Rakholia et al, 2017). The classifier performs better on randomly partitioned 10 times test data than 2 times. This shows that the classifier may have had insufficient training data for the latter instance. Since feature selection enhances prediction accuracy, not employing a classifier is better. Additionally, it works on small data sets. K-Nearest Neighbor (K-NN) and Naive Bayes were tested for movie and hotel evaluation accuracy, precision, and recall. The Naive Bayes approach outperformed the K-NN method in movie rating prediction, but both systems performed similarly in hotel rating prediction.

3 METHODOLOGY

This section provides machine learning techniques for analysing students feedback towards quality management in higher education. Student feedback data set is preprocessed to remove noise. Then student feedback data is analysed using SVM, ANN and random Forest algorithm. An excellent example of supervised learning is the SVM classifier, which processes enormous volumes of data in order to recognize patterns that were not previously visible. The categorization and analysis of multivariate data are two examples of common uses. The support vector machine (SVM) classifier assigns a single label to newly collected data. This classifier is built on the basis of the probabilistic binary classifier. Nevertheless, in spite of the fact that it belongs to the non-linear category, it is sometimes referred to as the kernel. Support vector machines, often known as SVMs, are frequently hailed as the most efficient margin classifiers. This is generally due to the fact that SVMs are able to efficiently split n-numbers of records that belong to the same category. It is better to have a non-linear relationship between the margin value and each of the categories. The SVM idea is used rather often when it comes to the process of data classification Arifin et al, 2021).

ANNs are not a novel concept by any means. Because of the interdependence of its input and output data, this system may replicate a sophisticated

one. This location houses the systems responsible for input and output. Because of the training assignment, ANN is now able to acquire and retain information on the compound system without the need to reference any more data from outside sources. ANN is unrivalled when it comes to solving problems, making predictions, recognising patterns, and classifying data. When applied to more complicated systems, ANN is able to develop non-linear correlations between identifiers that are more accurate and exact than those generated by traditional methods. In addition to this, it is able to take into account a wide range of data types, including those that are inaccurate, insufficient, or noisy.

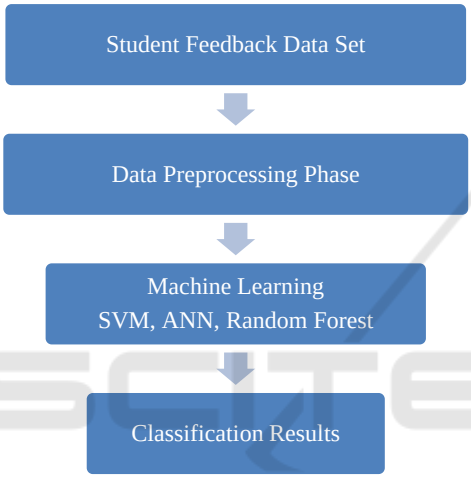

Figure 1: Machine Learning Techniques for Analyzing Students Feedback Towards Quality Management in Higher Education.

In recent years, ANN has been used to solve challenges connected to decision-making because of its dependability and superior abilities in gathering non-linear correlations among the identifiers of essential systems. It has recently gained popularity as an option for use in the development of medical diagnostic models. These models may be helpful to medical professionals in aiding them in developing a diagnosis based on the patient's reported symptoms. They are based on the information that was supplied by the patients themselves (Kour et al, 2021).

A method referred to as Random Forest was first developed for use in order to categorise and predict data. During the development of this system, an approach known as ensemble learning was used. During this stage of the process, a forest of decision trees is constructed, and regression methods are used to make predictions on which branches will produce fruit. It has a low standard deviation and is effective at integrating the many components of the incoming data, which are both requirements for producing accurate predictions. The use of random forest classification is first met with reluctance from a great number of individuals due to the misconception that it is difficult to implement. This is one of the explanations that may be given (Chowdhery et al, 2021).

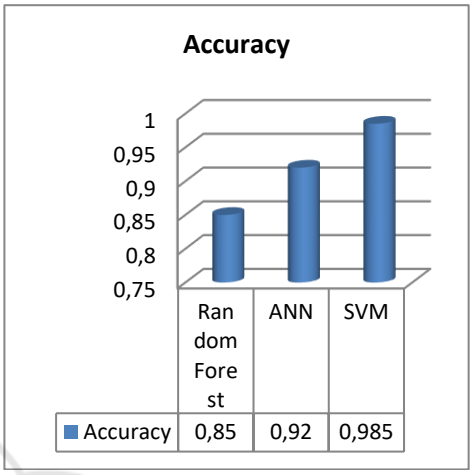

Figure 2: Accuracy of Machine Learning Classifiers for Student Feedback Data.

4 RESULTS ANALYSIS

Student feedback dataset was contributed to by the students of a renowned educational institution located in northern India. It is highly recommended that the Institutional Report be created with this data set serving as the basis for student comments. This database primarily contains the following categories of information: classroom teaching, course materials, examinations, laboratory exercises, library resources, and extracurricular activities. Each group of information is organised into two columns, and each column may be labelled with either a 0 (indicating a neutral value), a 1 (indicating a positive value), or a -1 (indicating a negative value) (negative). There are 185 entries in all that may be found there. Results are shown in figure 2, figure 3 and figure 4.

Accuracy= (TP + TN) / (TP + TN + FP + FN)
Sensitivity = TP/ (TP + FN)
Specificity = TN/ (TN + FP)
Where
TP= True Positive
TN= True Negative
FP= False Positive
FN= False Negative

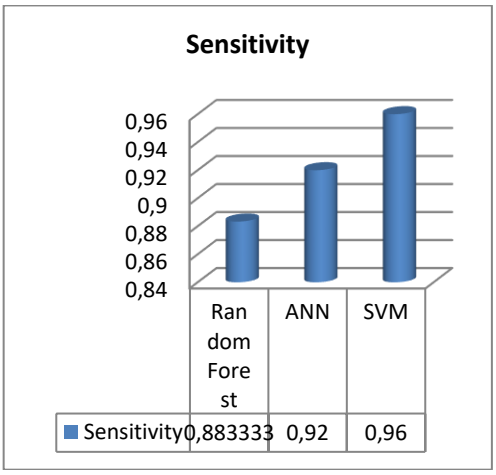

Figure 3: Sensitivity of Machine Learning Classifiers for Student Feedback Data.

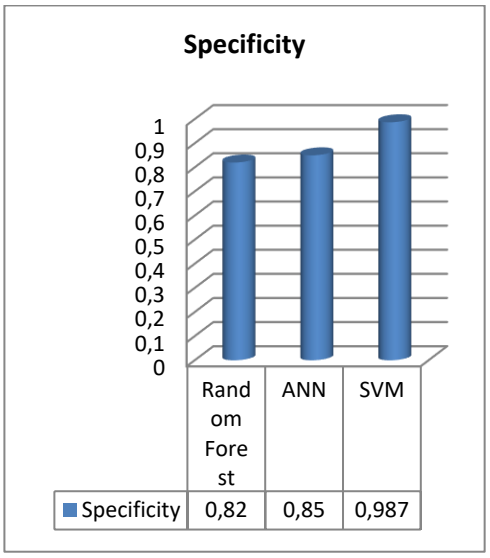

Figure 4: Specificity of Machine Learning Classifiers for Student Feedback Data.

5 CONCLUSION

The field of study known as educational data mining is one that investigates content that has been specifically gathered from educational settings by using data mining, machine learning, and statistics. This kind of research is referred to as "educational data mining." The purpose of the learning and teaching process is to offer students with the greatest experience they can have in terms of learning and understanding the content that is being taught to them. This is the best experience that can be provided to them. It is possible to use educational data mining for a variety of purposes, including predicting student performance and identifying students who are at risk, determining important concerns in the learning patterns of various groups of students, increasing pass-out rates, accurately assessing the performance of the institution, making the most of campus resources, and optimising the renewal of subject curriculum. These are just some of the potential applications of educational data mining. This article presents many methods of machine learning that may be used to analyse the responses of students to questions on quality management in higher education. The data collection including student feedback is preprocessed in order to reduce noise. The data collected from the student feedback surveys are then analysed using SVM, ANN, and the random Forest method. It has been discovered that the performance of the SVM algorithm is superior when it comes to assessing student feedback data for the purpose of improving the overall quality of higher educational institutions.

REFERENCES

R. M. Rakholia and J. R. Saini, “Classification of Gujarati documents using naïve Bayes classifier,” Indian Journal of Science and Technology, vol. 10, no. 5, pp. 1–9, 2017. doi:10.17485/ijst/2017/v10i5/103233

R. Kannan, M. Bielikova, F. Andres, and S. R. Balasundaram, “Understanding honest feedbacks and opinions in academic environments,” Proceedings of the Fourth Annual ACM Bangalore Conference, 2011. doi:10.1145/1980422.1980443

R. Kannan, M. Bielikova, F. Andres, and S. R. Balasundaram, “Understanding honest feedbacks and opinions in academic environments,” Proceedings of the Fourth Annual ACM Bangalore Conference, 2011. doi:10.1145/1980422.1980443

J. Mesarić and D. Šebalj, “Decision trees for predicting the academic success of students,” Croatian Operational Research Review, vol. 7, no. 2, pp. 367–388, 2016. doi:10.17535/crorr.2016.0025

P. V. Eswara Rao and S. K. Sankar, “Survey on educational data mining techniques,” International Journal Of Engineering And Computer Science, 2017. doi:10.18535/ijecs/v6i4.41

N. Khodeir, “Student modeling using educational data mining techniques,” 2019 6th International Conference on Advanced Control Circuits and Systems (ACCS) & 2019 5th International Conference on New Paradigms in Electronics & information Technology (PEIT), 2019. doi:10.1109/accs-peit48329.2019.9062874

S. Kovalev, A. Kolodenkova, and E. Muntyan, “Educational Data Mining: Current Problems and Solutions,” 2020 V International Conference on Information Technologies in Engineering Education (

Inforino), 2020. doi:10.1109/inforino48376.2020.9111699

A. Hicham, A. Jeghal, A. Sabri, and H. Tairi, "A survey on educational data mining [2014-2019]," 2020 International Conference on Intelligent Systems and Computer Vision (ISCV), 2020. doi:10.1109/iscv49265.2020.9204013

L. Ji, X. Zhang, and L. Zhang, "Research on the algorithm of Education data mining based on Big Data," 2020 IEEE 2nd International Conference on Computer Science and Educational Informatization (CSEI), 2020. doi:10.1109/csei50228.2020.9142529

R. K. Veluri et al., "Learning analytics using Deep Learning techniques for efficiently managing educational institutes," Materials Today: Proceedings, vol. 51, pp. 2317–2320, 2022. doi:10.1016/j.matpr.2021.11.416

AjitBrar, "Student feedback dataset," Kaggle, https://www.kaggle.com/datasets/brarajit18/student-feedback-dataset?resource=download (accessed Sep. 24, 2023).

J. Chowdhery, A. Jasmin, A. Jaiswal, and J. A. Jothi, "Automatic student performance prediction system using data mining techniques," 2021 International Conference on Computing and Communications Applications and Technologies (I3CAT), 2021. doi:10.1109/i3cat53310.2021.9629427

S. Kour, R. Kumar, and M. Gupta, "Analysis of student performance using Machine Learning Algorithms," 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), 2021. doi:10.1109/icirca51532.2021.9544935

M. Arifin, Widowati, Farikhin, A. Wibowo, and B. Warsito, "Comparative analysis on educational data mining algorithm to predict academic performance," 2021 International Seminar on Application for Technology of Information and Communication (iSemantic), 2021. doi:10.1109/isemantic52711.2021.9573185