Journal of Theoretical and Applied Information Technology
31st January 2025. Vol.103. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

PREDICTING INSTRUCTOR PERFORMANCE IN HIGHER EDUCATION USING STACKING AND VOTING ENSEMBLE TECHNIQUES

ABDELBASET R. ALMASRI[1], NOR ADNAN YAHAYA[2], SAMY S. ABU-NASER[3]

[1,2] University Malaysia of Computer Science & Engineering (UNIMY), Malaysia.
[3]Faculty of Engineering and Information Technology, Al-Azhar University, Palestine.
E-mail: [1]P02220005@student.unimy.edu.my, [2]adnan.yahaya@unimy.edu.my, [3]abunaser@alazhar.edu.ps

ABSTRACT

Instructors' evaluation is crucial to maintaining educational quality and meeting student needs. It is done through a Student Evaluation of Teaching (SET) survey in higher education to provide constructive student opinions to their instructors and help them improve their courses and teaching practices. This study used extensive mining analysis to analyze the students' responses. A public dataset of 5820 SET survey records from UCI was analyzed to reveal insights into the students' perceptions and expectations of how the courses prepare and help them solve real-world issues. In this analysis, the study used six different machine learning methods: K-nearest neighbor (KNN), Support Vector Machine (SVM), Decision Tree (DT), Gradient Boosting (GB), Random Forest (RF), and Extra Trees (ET). The study validated each of these methods individually and in various combinations using two ensemble methods: stacking and voting. The study goal was to identify the best-performing individual methods and determine the best combinations of methods for predicting outcomes. Based on the study, it was found that an ensemble classifier, comprising the four best-performing classifiers (ET, RF, DT, GB) with stacking, performed better compared to other classifiers. This ensemble achieved an accuracy of 91.616%, which was 0.791% higher than the accuracy of the best single-based classifier (ET), which was 90.825%. The results obtained suggest that the use of ensemble learning can effectively enhance instructor performance predictability.

Keywords: *Machine Learning, Instructor, Performance, Prediction, High Education*

1. INTRODUCTION

Higher education (HE) plays a vital role in economic and societal progress. As the number of students pursuing higher education increases, universities are faced with the challenge of enhancing the quality of education to meet student expectations [1]-[5]. One of the key concerns for HE institutions is evaluating the performance of instructors and determining the satisfaction levels of students regarding their instruction [6]-[10]. To achieve this, HE managers should identify the crucial factors influencing student satisfaction and instructor performance [11]-[15]. This includes determining and predicting factors that lead to satisfaction or dissatisfaction and how these factors vary based on the instructor and course aspects[16]-[20]. In higher education, institutions typically administer Student Evaluation of Teaching (SET) surveys to gather feedback from students at the end of each term. The collected data is then analyzed using statistical analysis techniques[21]-[25]. However, such statistical analysis is insufficient for providing broad knowledge and identifying the complex interplay between the instructor performance factors. To address this, data mining (DM) and machine learning technologies can be applied to enable decision-making in HE institutions. These technologies can analyze detailed data and apply techniques such as classification model induction, association rules, evolution and deviation analysis, and clustering for related data items. Using ensemble machine learning methods is considered one of the most effective ways for many data mining approaches and classification methods in particular, as training a set of classification models and combining their outputs can improve the prediction performance of a single model [26]-[30] A study by Fernández-Delgado et al.(2014) [31] analyzed 179 classifiers across various datasets and real-world problems, concluding that ensembles are the most effective approach for solving machine learning problems. Despite its potential, ensemble methods are often overlooked in data mining studies in higher education. Thus, the primary objective of this study is to propose and build suitable

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

classification models teamed up and boosted with stacking and voting ensemble methods for accurately predicting teaching constructs that strongly impact student satisfaction and expectations.
This paper compares six different machine algorithms for classification with proposed ensemble approaches. These individual algorithms are used and fused to develop highly accurate predictive ensemble-based models classifying students' feedback in teaching evaluations. The six algorithms used are K-nearest neighbor, Support Vector Machine, Decision Tree, Gradient Boosting, Random Forest, and Extra Trees. The study utilizes two ensemble methods, stacking, and voting, to create various ensembles in particular forms of these individual algorithms. Furthermore, the paper discusses the performance evaluation measures for each classification model and presents the key features that affect the prediction outcomes of these models.

2. PROBLEM STATEMENT

The quality of education in higher education institutions (HEIs) hinges significantly on the performance of instructors, as reflected in student satisfaction and their learning experiences. Traditional approaches to analyzing SET data—statistical methods and isolated machine learning models—are often limited in capturing the complex relationships between instructor performance, student expectations, and satisfaction. Additionally, these approaches struggle with challenges such as imbalanced datasets and model accuracy.

Although ensemble learning has demonstrated success in addressing similar issues in other domains, its application to predicting instructor performance in higher education remains underexplored. This research identifies this gap and aims to design an ensemble-based predictive framework that accurately evaluates instructor performance by integrating the strengths of individual machine learning classifiers. Using stacking and voting ensemble techniques, the study seeks to overcome the limitations of single-model approaches and provide a robust solution for higher education management to make data-driven decisions for quality improvement.

3. RELATED WORKS

Data mining is extensively utilized to analyze the behavior of instructors and the performance of students, aiming to improve teaching quality and support professional growth. This approach is particularly beneficial in guiding decision-making processes related to instructor assignments, assessments, and professional training programs. The adoption of data-driven methods in educational settings allows institutions to uncover patterns and trends that can drive improvements in both teaching strategies and student outcomes. In this section, we present examples of related work that demonstrate the application of data mining techniques in analyzing instructor and student performance.

Lalata et al. (2019) [32] conducted a comprehensive analysis comparing numerical ratings to sentiment analysis derived from student feedback entries. By employing a voter ensemble classifier comprising five individual classifiers, the study established connections between quantitative ratings and sentiments expressed in evaluations of faculty members. Their work highlighted the significance of leveraging ensemble methods to capture nuanced insights from feedback data. Similarly, Ajibade et al. (2020) [33] utilized various classification and ensemble methods to enhance the predictive accuracy of student performance models. Their findings underscored the importance of behavioral features in improving academic performance prediction, emphasizing that ensemble classifiers often outperform single-model approaches.

Ravinder Ahuja and S. C. Sharma (2020) [34] explored instructor performance evaluation by applying twelve classification methods on a labeled dataset obtained through agglomerative clustering and k-means algorithms. Among these methods, the Support Vector Machine (SVM) with Principal Component Analysis (PCA) for feature selection demonstrated the highest accuracy, reaffirming the utility of feature engineering in enhancing model performance. In another notable study, Abunasser et al. (2022) [35] examined instructor performance using 18 machine learning and deep learning algorithms. The study revealed that the Extra Trees Classifier achieved the best accuracy, outperforming other algorithms, and underscored the role of advanced ensemble methods in handling complex datasets.

In the domain of physical education, J. Zhao (2022) [36] utilized a decision tree algorithm to evaluate instructional quality, demonstrating how domain-specific analysis can reveal critical insights about teaching efficacy. Hou (2022) [37] adopted logistic regression and decision tree models, as well as their combined application, to predict student success in

English examinations. Their research integrated both student learning behaviors and instructor teaching methodologies, highlighting the multifaceted nature of factors influencing academic performance.

A recent study by Almasri et al. (2023) [1] showcased the effectiveness of data mining techniques in analyzing educational data, particularly the behavior of instructors and its impact on student satisfaction. By employing K-NN clustering and the C4.5 classification algorithm, the study achieved remarkable accuracy in predicting student satisfaction, further validating the potential of data-driven methodologies in improving the educational experience.

Despite the extensive focus on student performance modeling in higher education, instructor performance modeling remains comparatively underexplored. This gap in the literature raises critical questions about how instructors can effectively evaluate and enhance their performance, identify areas for professional development, and determine the specific characteristics that contribute to improved student satisfaction, motivation, and achievement. Moreover, while individual models have been widely studied, the potential of ensemble models—leveraging multiple algorithms to collectively enhance accuracy—remains underutilized. Future research should address these gaps by integrating diverse data sources and exploring ensemble modeling techniques to provide a more holistic understanding of the factors driving educational success.

4. METHODS AND MODELING APPROACH

In this section, we will explore the dataset utilized in our study, outline the experimental design methodology, and detail the evaluation metrics employed.

4.1 Dataset

The study dataset is collected through a survey completed by students to evaluate their instructors across various courses [9]. The dataset comprises 5820 records with 33 features, consisting of details such as instructor and course codes, attendance, course difficulty, and student responses to 28 survey questions. These questions are of the 5-point Likert-type, ranging from 1 (poor) to 5 (excellent), and cover different aspects of the course structure, student satisfaction, and the instructor's educational practice. The questions are categorized from Q1 to Q7 for the course's characteristics, Q8 to Q12 for student satisfaction with course activities, class participation, meeting initial expectations, professional growth, and the course's relevance to real-world issues, and Q13 to Q28 for instructor performance. The study focuses on Q12 to predict student preparedness and expectations, as it is assumed that students' satisfaction with the course's relevance to real-world issues is a key indicator of their overall satisfaction. The dataset shows that Q12 has possible values of "poor," "fair," "good," "very good," and "excellent," with 1052, 822, 1696, 1367, and 883 instances, respectively.

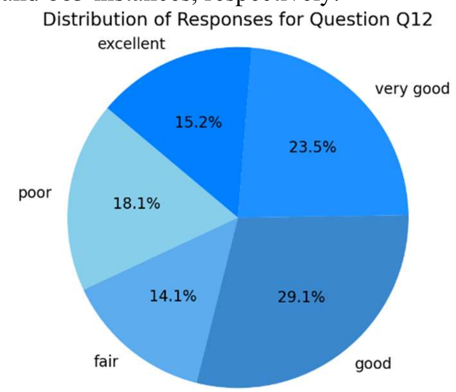


Figure 1. Class distribution of the target feature

As shown in Figure 1, the target feature displays an imbalanced distribution across its five classes. To avoid any bias towards the majority class, the study employed a random oversampling technique to over-present the minority [27]-[42] classes to ensure that the number of instances for each class is equal. This approach helps to achieve a balanced dataset and allows for a fair evaluation of instructors by ensuring that the developed models are not skewed towards any particular class [42]-[45].

4.2 Feature Selection Technique

The study employed a feature selection technique[46] that used Gini index analysis to identify the most relevant features for the classification process, to lower the risk of overfitting, and to enhance the models' performance and interpretability of the results. The Gini Index is a criterion used in decision trees to measure node purity, which calculates the probability of a randomly selected element in the node being incorrectly classified based on class distribution. A lower Gini Index indicates a purer node, with most samples belonging to one class. Decision trees split nodes based on the feature that minimizes the Gini Index, producing child nodes with better class purity. When constructing decision trees, it is recommended to use features with higher Gini Decrease scores to split nodes. To ensure that only the most influential variables are considered in developing the predictive models, the study selected 75% of the dataset features with higher Gini decrease scores. Based on

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure 2, which ranks features by Gini decrease score, the study determined that the top five important features for predicting student preparedness and expectations are "professional growth" (Q11), "fulfillment of course expectations" (Q10), "learning activities" (Q8), "educational methods" (Q7), and "class participation" (Q9).

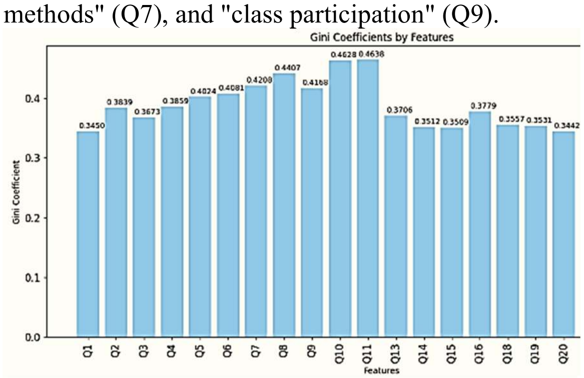


Figure 2: Features ranked by Gini Decrease

4.3 Experimental Design

Our study aims to evaluate and compare the effectiveness of ensemble approaches in building accurate predictive models. Therefore, the study uses six traditional classification methods, K-nearest neighbor (KNN), Support Vector Machine, Decision Tree (SVM), Gradient Boosting (GB), Random Forest (RF), and Extra Trees (ET), along with ensemble techniques like stacking and voting. The goal is to determine the optimal ensemble way for accurate predictions. The experimental design approach in this study is demonstrated in Figure 3, which includes producing diverse predictive models fused using stacking and voting methods. Firstly, the approach involves collecting and preparing the dataset. Second, the six classification methods are employed to develop individual predictive models for instructor performance. All individual models are trained and tested using 10-fold cross-validation. In the next step, these individuals are evaluated, and the top performers (Top 2, Top 3, Top 4, Top 5, and All classifiers) are chosen to construct diverse ensemble models using stacking and voting techniques. To ensure accuracy, the study validated and tested all individual models and ensembles using the same K-fold cross-validation (K=5, 10, 20).

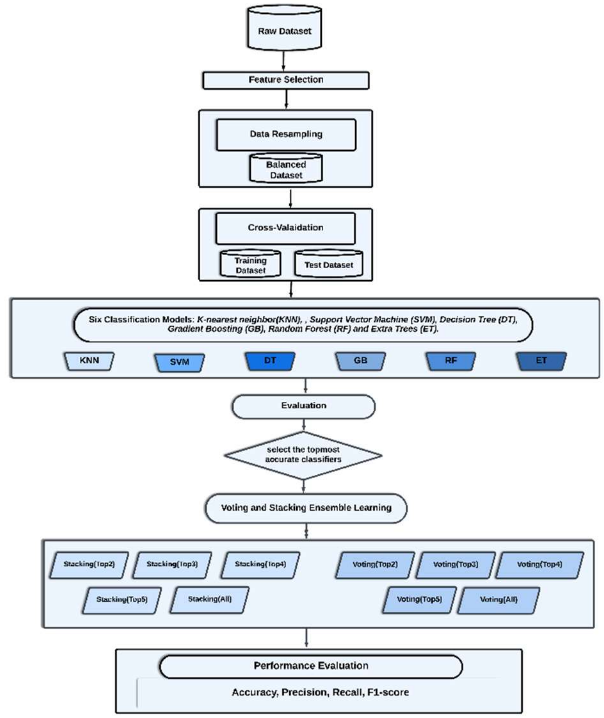


Figure 3: The Experimental Design for Predictive Model Fusion

4.4 Classification Validation

Cross-validation is a crucial method used to evaluate the effectiveness of a classification model. This procedure involves dividing a dataset into different subsets, where one subset is used to train the model, and the other subset is reserved for evaluating it. The most common approach is called "k-fold Cross-Validation" (k-fold CV). In this approach, the training dataset is divided into k smaller sets that do not overlap. The model is then iteratively trained on k-1 of these folds and evaluated on the remaining fold. Finally, the performance metric is calculated by averaging the values obtained from all iterations. This metric is reported as the outcome of the k-fold cross-validation process. In our experiments, the study varies the number of folds used (specifically, k = 5, 10, 20) to determine which k value results in the best model performance. Although the 10-fold validation method is popular in data mining research, our objective is to find the optimal k value that meets the specific requirements of our study.

4.5 Evaluation Measures

During the modeling phase, it is important to evaluate the generated models to determine their accuracy and effectiveness. This evaluation process involves testing and analyzing the models against a designated test dataset using cross-validation techniques. Multiple evaluation measures are utilized, including accuracy, precision, recall, and

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the F1-score. These metrics are valuable for assessing the performance of the models[10].
Accuracy is one of the most commonly used metrics. It is calculated by dividing the number of correct predictions by the total number of instances in the dataset. Recall, also known as the true positive rate (TPR), measures the proportion of correct positive predictions in relation to the total number of actual positives. Precision measures the proportion of correctly classified positive predictions to the total number of positive predictions, regardless of whether they were classified correctly or incorrectly. The F1 measure provides a comprehensive evaluation of classification performance by combining recall and precision into a single metric. It offers a balanced assessment of a model's ability to make accurate positive predictions while minimizing false positives and false negatives.

5. RESULT

The experiments in the current study were conducted using various Python libraries on Google Colab. These libraries include pandas, numpy, sklearn, and others. Many of the classifier algorithms used were from the sklearn library, which offers a wide range of machine-learning methods. Some of the default parameter values for a few of these algorithms were manually adjusted. In this study, the performance of various machine-learning classifiers was evaluated using a K-fold cross-validation approach. The classifiers considered included Extra Trees, Random Forest, Decision Tree, Gradient Boosting, Support Vector Machine, and K-Nearest Neighbors. Notably, Extra Trees emerged as the top-performing classifier, demonstrating exceptional accuracy, recall, precision, and F1 score.

Table 1: 10-Fold cross-validation accuracy scores for individual classifiers

Rank	model	accuracy	recall	precision	F1
1	ET	90.554	90.537	90.597	90.524
2	RF	90.118	90.117	90.23	90.079
3	DT	88.797	88.776	88.8	88.704
4	GB	85.554	85.509	85.615	85.519
5	SVM	85.366	85.338	85.53	85.38
6	KNN	83.573	83.546	84.037	83.677

Table 1 displays the performance of individual classifiers during a 10-fold cross-validation. The models are ranked based on their accuracy, with Extra Trees (ET) achieving the highest accuracy of 90.554%, followed by Random Forest (RF) with an accuracy of 90.118%. In terms of recall, Extra Trees (ET) and Random Forest (RF) again perform well, indicating their ability to correctly identify positive instances

Table 2: K-Fold Cross-Validation accuracy scores for individual and voting classifiers

K	Individual classifiers						voting ensemble classifiers				
	KNN	DT	SVM	ET	GB	RF	Top 2	Top 3	Top 4	Top 5	All
5	83.137	88.113	85.094	89.917	85.507	89.811	89.917	89.552	**90.035**	**90.035**	89.316
10	83.573	88.797	85.366	90.554	85.554	90.118	90.366	90.342	**90.731**	90.849	89.976
20	83.785	89.068	85.283	90.825	85.2	90.837	90.566	90.672	**90.943**	90.684	90.13

Table 3: K-fold cross-validation accuracy scores for individual and stacking classifiers

K	Individual classifiers						Stacking ensemble classifiers				
	KNN	DT	SVM	ET	GB	RF	Top 2	Top 3	Top 4	Top 5	All
5	83.137	88.113	85.094	89.917	85.507	89.811	90.507	**90.531**	90.436	90.566	90.472
10	83.573	88.797	85.366	90.554	85.554	90.118	91.392	91.191	**91.521**	91.203	91.38
20	83.785	89.068	85.283	90.825	85.2	90.837	91.333	91.486	**91.616**	91.415	91.403

Our study evaluates individuals and selects top performers to construct diverse ensemble models. These models consist of the top 2 (ET, RF), top 3 (ET, RF, DT), top 4 (ET, RF, DT, GB), top 5 (ET, RF, DT, GB, SVM), and all classifiers, which are created using stacking and voting techniques. To ensure accuracy, the study validated and tested all individual models and ensembles using the same K-fold cross-validation (K=5, 10, 20). Tables 2 and 3 present the experimental results of the classifiers' accuracy measures, both individually and in voting and stacking ensembles, at different K values (5, 10, and 20) in a K-fold cross-validation setup. The study assessed the classifiers' performance at various K values and found that increasing K from 5 to 20 resulted in more dependable and consistent performance estimates. Decision Trees showed consistent performance across different K values. The ensemble models outperformed single classifiers, indicating that combining predictions from different models can enhance accuracy and robustness. At k=20, the classifier with an ensemble of the Top 4 (ET, RF, DT, GB) had the highest accuracy of 90.943% and 91.616% using voting and stacking methods, respectively.

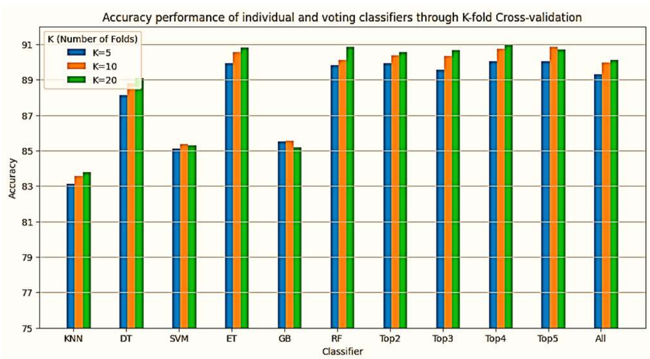


Figure 4. K-fold cross-validation accuracy scores for individual and voting classifiers

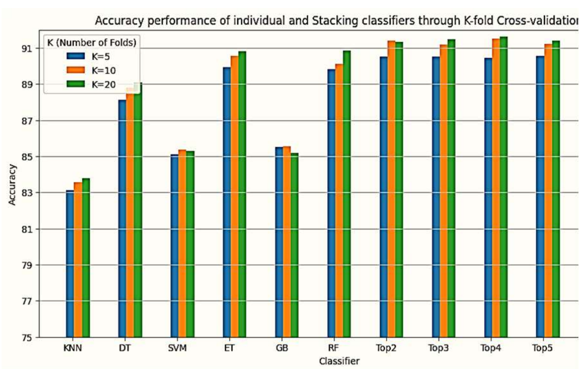
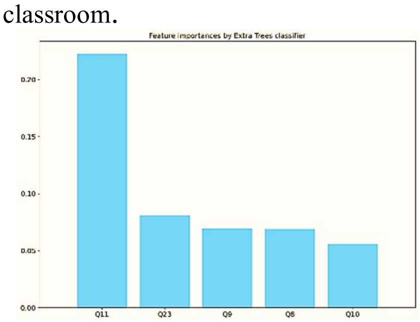


Figure 5. K-fold cross-validation accuracy scores for individual and stacking ensemble classifiers

Comparisons between six individual classifiers and the ensemble classifiers which include them are displayed in Figures 4 and 5. The study focused on classifying the instructor performance using the SET survey data and found that ET was the most effective among the individual classifiers considered. Additionally, the "Top 4" ensemble classifiers generally performed better than the others, whether in the single classifiers or the voting and stacking ensembles. This suggests that combining the classifiers in the "Top 4" model can result in highly accurate predictions in this context.

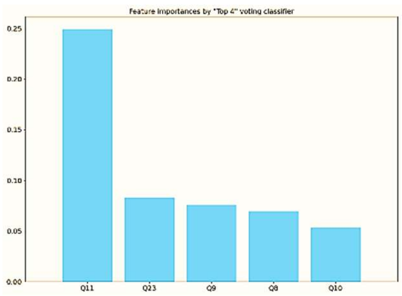
Regarding feature importance, in Figure 6, the top five important features are displayed based on their Mean Decrease in Accuracy (MDA) rankings for three classifiers that performed well: Extra Trees (a), "Top 4" voting (b), and "Top 4" Stacking (c). MDA measures the accuracy decrease of a classifier when feature values are randomly altered. A zero MDA value indicates that the feature was not used in the prediction, whereas a high MDA value indicates heavy reliance on that particular feature. For predicting the target feature "Q12," the top three attributes were "Q11" ("The course was relevant and beneficial to my professional development."), "Q23" ("The instructor encouraged participation in the course"), and "Q9" ("I greatly enjoyed the class and was eager to actively participate during the lectures.") based on the outperformed classifiers. Based on these results, it is clear that the majority of the feedback pertains to two key areas. Firstly, there is a substantial emphasis on the course content and its role in enhancing students' professional skills. Secondly, considerable attention is given to the instructor's strategies for promoting student involvement and active participation in the classroom.



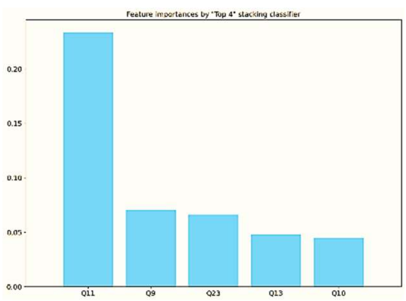
(a) (b)

Journal of Theoretical and Applied Information Technology
31st January 2025. Vol.103. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195



(b)



©

Figure 6: Top five most important features for outperformed models (a) Extra Trees; (b) "Top 4" voting ; (c) "Top 4" Stacking

6. DISCUSSION

In order to ensure the credibility of predictive models, it is essential to conduct verification. This study aimed to compare our results to previous research that utilized ensemble methods to predict instructor performance. Lalata et al. (2019) [4] developed an ensemble classification system to predict faculty evaluation and it resulted in improved machine-learning outcomes with an accuracy of 90.26%. They further increased accuracy by 0.06% using the voting method with DT, LR, RF, SVM, and NB. In Ahuja & Sharma's (2021) [11] study, experimental results demonstrated a 2% increase in the accuracy of the proposed model compared to existing literature when ensemble models (stacking and voting) were utilized. Our "Top 4" predictive model (ET, RF, DT, and GB with Stacking) had a 91.616% accuracy when compared to state-of-the-art studies using the same dataset and targeting the same feature. This is a significant improvement compared to the accuracies reported in the previous studies by Almasri et al. 2022 [12] which had an accuracy of 86.545% using ET with the SMOOT method and Afrin et al. (2020) [13] which had an accuracy of 79.86% using SVM method. Moreover, the experimental results indicate a 1.06% increase in accuracy when combining the individual models with the ensemble approach that incorporates ET, RF, DT, and GB, and stacking techniques. This suggests that the predictive performance has been effectively enhanced. Verification is crucial in creating predictive models as it determines their credibility.

7. CONCLUSION

The study employed a new approach to predicting instructor performance in higher education through stacking and voting ensemble techniques. The approach uses data mining techniques like feature selection, data resampling, classification, ensemble modeling, and important feature extraction. The models are tested using different k values of cross-validation. The study found that the "Top 4" stacking model, which uses ET, RF, DT, and GB, had a higher accuracy of 91.616% than previous studies conducted by [9] [10], which had accuracies of 86.545% and 79.86%, respectively. Furthermore, the experimental results show a 0.791% increase in accuracy compared to other the best single-learner models used in the current study. Therefore, the study's findings contribute to the existing literature on the effectiveness of ensemble techniques in predicting instructor performance. Although the study provides valuable insights, it has some limitations since it only focuses on specific variables and techniques. As such, further research should explore additional factors and other ensemble techniques to more accurately capture the complexity of instructor performance. Nevertheless, educational institutions seeking to predict and evaluate instructor performance can still benefit from the study's valuable insights.

REFERENCES:

[1]. Almasri, T. Obaid, M. S. S. Abumandil, B. Eneizan, A. Y. Mahmoud, and S. S. Abu-Naser, “Mining Educational Data to Improve Teachers’ Performance,” 2023, pp. 243–255. doi: 10.1007/978-3-031-16865-9_20.

[2]. Ragab M., Abdel Aal A. M. K., Jifri A. O., and Omran N. F. , “Enhancement of Predicting Students Performance Model Using Ensemble Approaches and Educational Data Mining Techniques,” Wirel Commun Mob Comput, Vol. 2021, 2021, doi: 10.1155/2021/6241676.

[3]. Abu-Naser, S. S., Abunasser, B. S.: The Miracle Of Deep Learning In The Holy Quran. Journal Of Theoretical And Applied

Information Technology, Vol. 101, No. 17, 2023.

[4]. Abu-Naser, S. S., El-Hissi, H., Abu-Rass, M., El-Khozondar, N.: An expert system for endocrine diagnosis and treatments using JESS. Journal of Artificial Intelligence Scialert Vol. 3, No. 4, pp.239-51,2010.

[5]. Abunasser, B., Daud, S., Zaqout, I., Abu-Naser, S.: Abunaser-A Novel Data Augmentation Algorithm For Datasets With Numerical Features. Journal of Theoretical and Applied Information Technology Vol. 101, No. 11, 2023.

[6]. Taha, A., Ariffin, D., Abu-Naser, S. S.: Investigating The Effects Of Data Augmentation Techniques On Brain Tumor Detection Accuracy. Journal Of Theoretical And Applied Information Technology, Vol. 101, No. 11, 2023.

[7]. Abunasser, B. S., AL-Hiealy, M. R. J., Zaqout, I. S., Abu-Naser, S. S.: Literature review of breast cancer detection using machine learning algorithms. AIP Conference Proceedings; 2023: AIP Publishing, 2023.

[8]. Abunasser, B. S., AL-Hiealy, M. R. J., Zaqout, I. S., Abu-Naser, S. S.: Breast Cancer Detection and Classification using Deep Learning Xception Algorithm. International Journal of Advanced Computer Science and Applications, Vol. 13, No. 7, 2022.

[9]. Abunasser, B. S., Al-Hiealy MRJ, Zaqout IS, Abu-Naser, S. S.: Convolution neural network for breast cancer detection and classification using deep learning. Asian Pacific journal of cancer prevention: APJCP, Vol. 24, No. 2, pp. 531, 2023.

[10]. Abunasser, B. S., Daud, S. M., Abu-Naser, S. S.: Predicting Stock Prices using Artificial Intelligence: A Comparative Study of Machine Learning Algorithms. International Journal of Advances in Soft Computing & Its Applications, Vol. 15, No. 3, 2023.

[11]. Abunasser, B. S., Daud, S. M., Zaqout, I., Abu-Naser, S. S.: Convolution Neural Network For Breast Cancer Detection And Classification-Final Results. Journal of Theoretical and Applied Information Technology, Vol. 101, No. 1, pp. 315-29, 2023.

[12]. Almasri R., Yahaya A., and Abu-Naser S. S., "INSTRUCTOR PERFORMANCE MODELING FOR PREDICTING STUDENT SATISFACTION USING MACHINE LEARNING-PRELIMINARY RESULTS," *J Theor Appl Inf Technol*, Vol. 15, pp. 19, 2022, [Online]. Available: www.jatit.org

[13]. Alkayyali, Z., Idris, S. A. B., Abu-Naser, S. S.: A systematic literature review of deep and machine learning algorithms in cardiovascular diseases diagnosis. Journal of Theoretical and Applied Information Technology, Vol. 101, No. 4, pp. 1353-65, 2023.

[14]. Alkayyali, Z. K., Idris S., Abu-Naser, S. S.: A new algorithm for audio files augmentation. Journal of Theoretical and Applied Information Technology, Vol. 101, No. 12, 2023.

[15]. Alrakhawi, H., Jamiat, N., Umar, I., Abu-Naser, S.: Improvement of Students Achievement by Using Intelligent Tutoring Systems-A Bibliometric Analysis and Reviews. Journal of Theoretical and Applied Information Technology, Vol. 101, No. 11, 2023.

[16]. Alrakhawi, H. A., Jamiat, N., Abu-Naser, S. S.: Intelligent tutoring systems in education: a systematic review of usage, tools, effects and evaluation. Journal of Theoretical and Applied Information Technology, Vol. 101, No. 4,pp. 1205-26, 2023.

[17]. Al-Zamily, J. Y. I., Ariffin, S. B., Abu Naser, S. S. M.: A survey of cryptographic algorithms with deep learning. AIP Conference Proceedings; 2023: AIP Publishing.

[18]. Alzamily, J. Y. I., Ariffin, S. B., Abu-Naser, S. S.: Classification of Encrypted Images Using Deep Learning–Resnet50. Journal of Theoretical and Applied Information Technology, Vol. 100, No. 21, pp. 6610-20, 2022.

[19]. Arqawi, S. M., Rumman, M. A., Zitawi, E.A, Abunasser, B., Abu-Naser, S.: Predicting Employee Attrition and Performance Using Deep Learning. Journal of Theoretical and Applied Information Technology, Vol. 100, No. 21, pp. 6526-36, 2022.

[20]. Arqawi, S. M., Zitawi, E. A., Rabaya, A. H., Abunasser, B. S., Abu-Naser, S. S.: Predicting university student retention using artificial intelligence. International Journal of Advanced Computer Science and Applications, Vol. 13, No. 9, 2022.

[21]. Barhoom, A., Jubair, M. R., Abu-Naser, S. S.: A survey of bone abnormalities detection using machine learning algorithms. AIP Conference Proceedings; 2023: AIP Publishing.

[22]. Barhoom, A. M., Al-Hiealy, M. R. J., Abu-Naser, S. S.: Bone abnormalities detection and classification using deep learning-vgg16 algorithm. Journal of Theoretical and Applied Information Technology, Vol. 100, No. 20, pp. 6173-84, 2022.

[23]. Barhoom, A.M., Al-Hiealy, M. R. J, Abu-Naser, S. S.: Deep Learning-Xception Algorithm for Upper Bone Abnormalities Classification. Journal of Theoretical and Applied Information Technology 100(23), pp. 6986-97, 2022.

[24]. El-Habil, B. Y., Abu-Naser, S. S.: Global climate prediction using deep learning. Journal of Theoretical and Applied Information Technology, Vol. 100, No. 24, pp. 4824-38, 2022.

[25]. Elzamly, A., Abu Naser, S. S., Hussin, B., Doheir, M.: Predicting Software Analysis Process Risks Using Linear Stepwise Discriminant Analysis: Statistical Methods. Int J Adv Inf Sci Technol, Vol. 38, No. 38, pp. 108-15, 2015.

[26]. Elzamly, A., Hussin, B., Abu Naser, S. S., Doheir, M.: Classification of Software Risks with Discriminant Analysis Techniques in Software planning Development Process. International Journal of Advanced Science and Technology, Vol. No. 81, Pp. 35-48, 2015.

[27]. Elzamly, A., Messabia, N., Doheir M, Abu Naser, S., Elbaz, H. A.: Critical cloud computing risks for banking organizations: Issues and challenges. Religación Revista de Ciencias Sociales y Humanidades, Vol. 4, No. 18, pp.673-82, 2019.

[28]. Elzamly, A., Messabia, N., Doheir, M., Abu-Naser, S. S., Yaacob, N. B. M., Al-Aqqad, M., et al.: Assessment Risks for Managing Software Planning Processes in Information Technology Systems. International Journal of Advanced Science and Technology, Vol. 28, No. 1, pp. 327-38, 2019.

[29]. Saleh, A., Sukaik, R., Abu-Naser, S. S.: Brain tumor classification using deep learning. 2020 International Conference on Assistive and Rehabilitation Technologies (iCareTech); 2020: IEEE.

[30]. Taha, A., Ariffin, D., Abu-Naser, S.: Impact of data augmentation on brain tumor detection. Journal of Theoretical and Applied Information Technology, Vol. 101, No. 11, 2023.

[31]. Fernández-Delgado M., Cernadas E., Barro E. and Amorim D., "Do we need hundreds of classifiers to solve real world classification problems?," *Journal of Machine Learning Research*, Vol. 15, 2014.

[32]. Lalata J. A. P., Gerardo B., and Medina R., "A sentiment analysis model for faculty comment evaluation using ensemble machine learning algorithms," *ACM International Conference Proceeding Series*, pp. 68–73, Jun. 2019, doi: 10.1145/3341620.3341638.

[33]. Ajibade S. S. M., Ahmad N. B., and Shamsuddin S. M., "A Data Mining Approach to Predict Academic Performance of Students Using Ensemble Techniques," 2020, pp. 749–760. doi: 10.1007/978-3-030-16657-1_70.

[34]. Ahuja R. and Sharma S. C., "Exploiting Machine Learning and Feature Selection Algorithms to Predict Instructor Performance in Higher Education," *JOURNAL OF INFORMATION SCIENCE AND ENGINEERING XX*, 2020.

[35]. Zhao J., "Application of Decision Tree Algorithm in Teaching Quality Analysis of Physical Education," *Wirel Commun Mob Comput*, Vol. 2022, pp. 1–8, Apr. 2022, doi: 10.1155/2022/3556136.

[36]. Abunasser, B. S., AL-Hiealy, M. R. J., Barhoom, A. M., Almasri, A. R., Abu-Naser, S. S.: Prediction of Instructor Performance using Machine and Deep Learning Techniques. International Journal of Advanced Computer Science and Applications, Vol. 13, No. 7, 2022.

[37]. Hou R., "Application of Decision Tree Algorithm Based on Data Mining in English Teaching Evaluation," *Wirel Commun Mob Comput*, Vol. 2022, pp. 1–10, Jun. 2022, doi: 10.1155/2022/5717895.

[38]. Ernest Fokoue and Necla Gunduz, "{UCI} Machine Learning Repository," *University of California, Irvine, School of Information and Computer Sciences*, 2013, Accessed: Jun. 09, 2022. [Online]. Available: http://archive.ics.uci.edu/ml/index.php

[39]. ParkY. R., Kim Y. J., JuW., Nam K., Kim S., and Kim K. G., "Comparison of machine and

deep learning for the classification of cervical cancer based on cervicography images," *Sci Rep*, Vol. 11, No. 1, pp. 16143, Dec. 2021, doi: 10.1038/s41598-021-95748-3.

[40]. Ahuja R. and Sharma S. C., "Stacking and voting ensemble methods fusion to evaluate instructor performance in higher education," *International Journal of Information Technology (Singapore)*, Vol. 13, No. 5, pp. 1721–1731, Oct. 2021, doi: 10.1007/S41870-021-00729-4

[41]. Albatish I. M., Abu-Naser, S. S.: Modeling and controlling smart traffic light system using a rule based system. 2019 International Conference on Promising Electronic Technologies (ICPET); 2019: IEEE.

[42]. Afrin F., Rahaman M. S., and Hamilton M., "Mining Student Responses to Infer Student Satisfaction Predictors," Jun. 2020, [Online]. Available: http://arxiv.org/abs/2006.07860

[43]. Taha, A. M., Ariffin, D., Abu-Naser, S. S.: A systematic literature review of deep and machine learning algorithms in brain tumor and meta-analysis. Journal of Theoretical and Applied Information Technology, Vol. 101, No. 11, pp. 21-36, 2023.

[44]. Taha, A. M., Ariffin, S. B. B., Abu-Naser, S. S.: Multi-modal MRI-Based Classification of Brain Tumors. A Comprehensive Analysis of 17 Distinct Classes. International Conference of Reliable Information and Communication Technology; 2023: Springer Nature Switzerland Cham.

[45]. Zarandah, Q. M., Daud, S. M., Abu-Naser, S. S.: A Systematic Literature Review Of Machine and Deep Learning-Based Detection And Classification Methods for Diseases Related To the Respiratory System. Journal of Theoretical and Applied Information Technology, Vol. 101, No. 4, pp. 1273-96, 2023.

[46]. Zarandah, Q. M., Daud, S. M., Abu-Naser, S. S.: Spectrogram flipping: a new technique for audio augmentation. Journal of Theoretical and Applied Information Technology, Vol. 101, No. 11, 2023.