Prediction of Student Score Performance of Sentiment Analysis Using Hybrid Cross Validation Machine Learning Techniques

Neha Singh
Department of Information Technology and Computer Application, Madan Mohan Malaviya University of Technology, Gorakhpur, India
Email: nehaps2703@gmail.com
Umesh Chandra Jaiswal
Department of Information Technology and Computer Application, Madan Mohan Malaviya University of Technology, Gorakhpur, India
Email: ucjitca@mmmut.ac.in

ABSTRACT

As a result of advancements in technology, one of the most beneficial and significant study areas nowadays is the forecast of educational outcomes on test results. In the subject of education, machine learning is incredibly beneficial, especially for evaluating student performance. This work suggests a hybrid cross validation machine learning technique to forecast student performance on sentiment analysis in order to enhance results. Additionally, alternative cross validation machine learning techniques already in use are contrasted with the suggested model. The study's findings demonstrate that the suggested hybrid cross validation technique yields the best compression outcomes.

Keywords - **Cross Validation, Education, Machine Learning, Sentiment Analysis, Student Performance.**

I. INTRODUCTION

The computational analysis of people's attitudes, opinions, and feelings regarding an object is called sentiment analysis, also referred to as opinion mining. The thing can stand in for specific occasions or subjects in the following fields: politics, e-commerce, health, and many more [1]. It is challenging to forecast students' success because learning databases contain a wealth of data. In institutes of higher education, predicting student success is essential. This is so because a high-quality college's stellar track record of academic success is one of its key tenets. The majority of Indian higher education institutions often utilize students' final grades to assess their performance. The syllabus, internal assessment results, and final test results are considered for determining grades. Examining student performance enables educators to carefully create a strategic course of action for each student's academic career. There are numerous techniques final scores are based on a variety of factors, including internal and external evaluations, laboratory file work, vivas, and sessional exams. A student's performance is based on how many grades they obtain on their final exam [5].

The performance of its pupils is one of the most important requirements for every institute. Based on their prior academic success, teachers can predict how well children would succeed. The findings imply that a connection between students' interests and success may exist. Depending on the performance, feedback, and other factors of their pupils, teachers might be evaluated. This kind of examination helps a university raise the calibre of its education. To gauge the degree of difficulty, question pa- pers can be graded. Such data helps an institution normalize the scores of all students taking multiple-session exams [6]. available at the moment for rating student performance [2]. The education sector has access to a lot of data in the form of student information. On this data, computer programmes can be used to extract information that will be beneficial for high-quality instruction. An educational institution has to have an ap-proximate previous understanding of the student body to forecast a student's performance in future academics. This gives them the chance to recognize and improve kids who are likely to receive worse grades while also assisting them in identifying promising individuals. Although there isn't a precise scale for gauging knowledge, exam scores are one way to gauge students' success [3]. Numerous factors, such as personal, social, and technological ignorance, among others, have an impact on students' academic performance and results [4]. The majority of institutions and universities in India use students' final exam grades as a measure of their academic performance. Any student's

In order to predict the student score performance of sentiment analysis, we used one hybrid model and four cross validation machine learning techniques in this study. In this work, we'll give a comparison of all these techniques after each technique's pre-diction.

The study is set up as follows: The literature survey is summarized in section 2, and the selection processes and features are explained in section 3. Section 4 explains the findings. The conclusion is presented in Section 5.

II. LITERATURE SURVEY

Sultana et al. [7] suggests for predicting student performance methodology. According to the findings, deep learning and support vector machines outperformed other classifiers in terms of general performance.

Mohamed et al. [8] improve educational systems through

an examination of student, teacher, and course comments. The language-based approach is used by the proposed SASCM model to determine how to remove as much information as feasible from each comment in the dataset.

Moreno-Marcos et al. [9] compares various machine learning algorithms using a real-world case research to see if the findings may reveal anything about the feelings or trends of MOOC students. The sentiment analysis employed both supervised and unsupervised techniques.

Aggarwal et al. [10] compares two models. Eight alternative classification strategies are used to build the models, which are then contrasted to determine the characteristics that will contribute to the creation of the most accurate model for classifying a student in accordance with his performance.

Khan et al. [11] offers a comprehensive examination of studies on the effectiveness of learners in conventional
classroom settings. The result demonstrates that during the training, the researcher prediction accuracy rises noticeably. *Sultana et al.* [12] determine the categorizing method that yields the best outcomes while also learning about student performance. The educational dataset was assembled using a database from a Saudi institution.

Yaacob et al. [13] forecast student performance at a certain Malaysian university by developing prediction models using a classification method. The most crucial data properties can be found using the integrated prediction model.

Kiu et al. [14] determine the importance and bearing on academic performance of student background, extracurricular activities, and coursework accomplishment. According to the experimental findings, a student's social and intellectual background were important predictors of their performance on a two-level classification.

Punlumjeak et al. [15] strategies for categorization The study's findings revealed that utilizing a neural network classifier and a feature selection technique for student data produced the best overall accuracy result.

Zohair et al. [16] demonstrate the viability of designing and training a prediction model with a respectable accuracy rate using a little dataset. The results showed that the clustering technique, out of the ones used, is capable of locating important indications in tiny datasets.

Punlumjeak et al. [17] In this study, it is suggested that four feature selection techniques be compared. The outcome of this study demonstrates the benefit of future selection in locating a minimal and substantial.

The results of all the aforementioned surveys explain numerous student enhancement, prediction, and algorithmic strategies, and they also attain the highest degree of accuracy. It's employed to forecast student outcomes using a variety of methodologies.

III. METHODOLOGY

We use a hybrid model of cross validation machine learning approaches in this study to predict student score prediction of sentiment analysis. The major goal of the re-search is to compare each technique performance in terms of student score prediction. Our evaluation of the sentiment analysis score prediction by student use of Jupyter Notebook. Fig. 1. illustrates the proposed methodology.

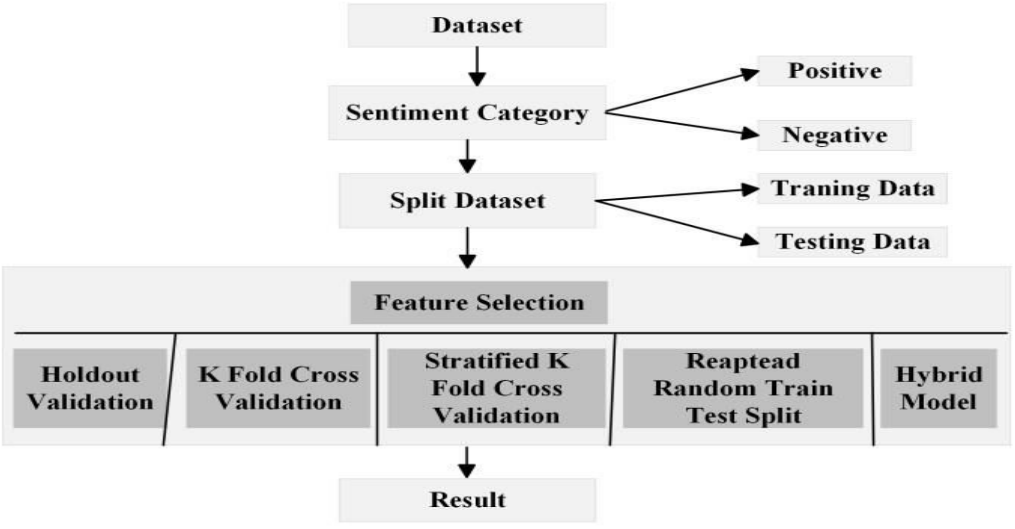


Fig. 1. Proposed Model

3.1. Dataset Description

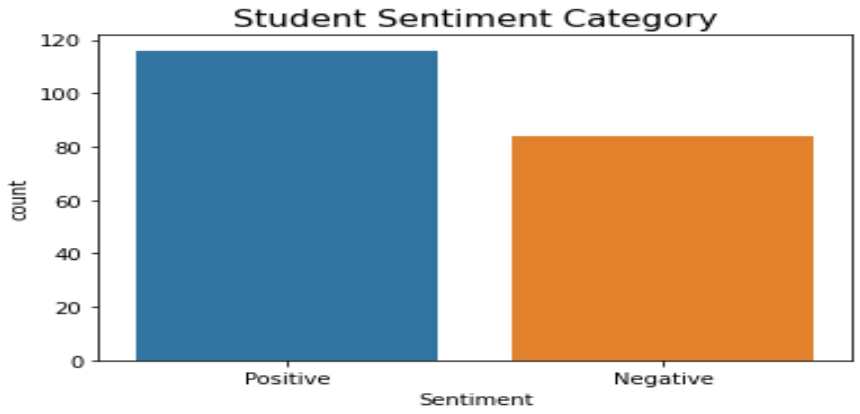
The data set collected for this study is linked at (https://www.kaggle.com/datasets/samarsaeedkhan/scores/c ode) and was collected on July 26, 2022. 200 student records with 6 attributes make up the small dataset collection. We also add an additional property that categorizes the records based on where the students are placed. If a student is placed, it falls under the positive sentiment category; if not, it falls under the negative sentiment category.

Table 1. contain information regarding the dataset

Table 1. Displays the types and attributes of the dataset

Attributes	Type
Python	Numeric
Sql	Numeric
ML	Numeric
Tableau	Numeric
Excel	Numeric
Student Placed	Nominal
Sentiment	Nominal

Fig. 2. displays the basic categories of student placement and the number of positive and negative emotion categories, both of which are 116 and 84 respectively.




```
Positive     116
Negative      84
Name: Sentiment, dtype: int64
```

Fig. 2. Shows student sentiment category

3.2. Feature Selection

There are various cross validation machine learning techniques accessible to fore-cast the performance of sentiment analysis on student test scores. The following techniques for analysis and prediction were used:

3.2.1. Holdout validation approach: It divides the dataset into training and validation sets of data at random. Usually, the distribution of training data is more even than that of test data. The model is built using training data, and validation data are used to evaluate the model's performance [18].

3.2.2. K fold cross validation: In this, the initial dataset is uniformly split into k folds or subparts. One of the k-folds or groups is selected as the validation data for each iteration, and the remaining (k-1) groups are selected as the training data. Each group is used as validation after the technique has been performed k times, and the leftover data is used as training data [18].

3.2.3. Stratified K-Fold cross validation: In this, the dataset is partitioned into k groups or folds so that the validation data contains an equal number of instances of the target class label. By doing this, especially when the dataset is uneven, it is ensured that one particular class is not overrepresented in the validation or train data [18].

3.2.4. Reaptead random train test split: This method randomly divides the data into training and test sets before further randomly splitting the data to evaluate the algorithm [19].

3.2.5. Proposed hybrid model based on Reaptead random train test split and Stratified K-Fold cross validation: A hybrid technique employs a number of strategies to tackle various issues and improve accuracy. Student success in sentiment analysis is predicted using Stratified K-Fold cross validation and Reaptead Random Train Test Split. Both of two validation technique has been created to address classification problems. In the end, our suggested model offers better results when compared to the other evaluated techniques. In terms of improving results, our suggested model per-forms better than previous findings.

3.3. Proposed Algorithm

Prediction of Student Score Performance Algorithm Representation for Hybrid Cross Validation Machine Learning Techniques

Input: Training on dataset using hybrid model

Output: Prediction of student score performance

1) Import required library
2) Load the dataset
3) Perform visualization result of overall dataset
4) Categorized student sentiment into positive and negative category
5) Split dataset into training and testing set
6) Feature selection using cross validation machine learning techniques
7) Performance Result

Fig. 3. Show the performance compression of various techniques and this comparison hybrid model gives best result in comparison to other techniques.

IV. RESULT

In this study, four techniques are utilized to predict student score performance on sentiment analysis. The goal was to create a hybrid cross validation machine learning method that, in comparison to other cross validation machine learning methods, was more accurate.

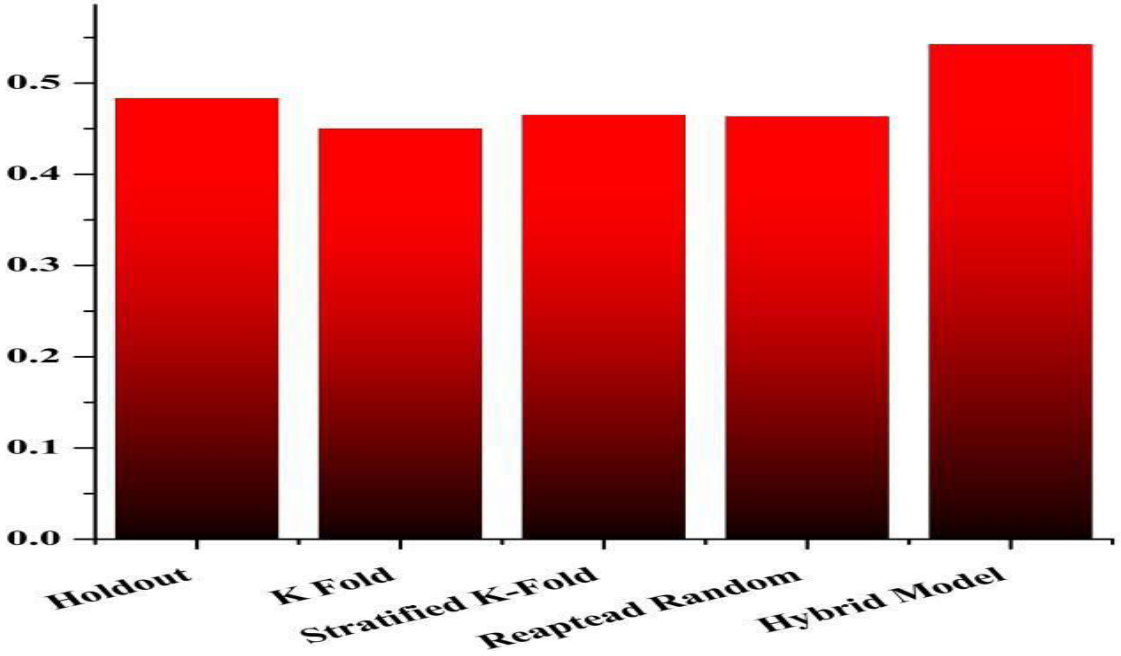


Fig. 3. Performance comparison of various techniques

Table 2. Comparative analysis of proposed model

Techniques	Result
Holdout Validation	0.48333
K Fold Cross Validation	0.44999
Stratified K-Fold Cross Validation	0.46499
Reaptead Random Train Test Split	0.46333
Hybrid Model (Reaptead Random Train Test Split and Stratified K-Fold Cross Validation)	**0.54241**

In Table 2, the hybrid model performs better in terms of correctness than the other four techniques. Results from the stratified K-Fold cross validation and the replicated random train test split are substantially identical. When compared to other procedures, K fold cross validation produces less accurate results.

V. CONCLUSION

There has been possible to gather a sizable amount of data as a result of the recent advancements in several domains. Institutions of higher learning now gather data on their students. Predicting and assessing the performance of their pupils is one of these institutions' biggest concerns. In a significant and relevant way, machine learning, a powerful analytical technique, can be utilized to forecast student performance. The major objective of this research is to use hybrid cross validation techniques to predict student score performance on sentiment analysis. A summary of earlier work is also provided by this research. One hybrid model and four cross validation machine learning approaches were employed in this experiment. The study's findings demonstrate that, when compared to the other strategies, the suggested hybrid cross validation machine learning technique produces the best compression results. Future student score prediction work will make use of cutting-edge machine learning approaches to enhance the findings.

ACKNOWLEDGEMENTS

I sincerely appreciate Prof. Umesh Chandra Jaiswal's unceasing guidance.

REFERENCES

[1] Umar, M., Ahmad, N.B. and Zainal, A., 2020.
Sentiment Analysis of Studentâ€™ s Opinion on Programming Assessment: Evaluation of NaÃ¯ ve Bayes over Support Vector Machines. International Journal of Innovative Computing, 10(2).

[2] Geetha, R., Padmavathy, T. and Anitha, R., 2021.
Prediction of the academic performance of slow learners using efficient machine learning algorithm. Advances in Computational Intelligence, 1(4), pp.1-12.

[3] Mohan, M.M., Augustin, S.K. and Roshni, V.K., 2015, December. A BigData approach for classification and prediction of student result using MapReduce. In 2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS) (pp. 145-150). IEEE.

[4] Ganorkar, S.S., Tiwari, N. and Namdeo, V., 2021.Analysis and prediction of student data using data science: a review. Smart Trends in Computing and Communications: Proceedings of SmartCom 2020, pp.443-448.

[5] Kumar, M., Singh, A.J. and Handa, D., 2017. Literature survey on student's performance prediction in education using data mining techniques. International Journal of Education and Management Engineering, 7(6), pp.40-49.

[6] Pallathadka, H., Wenda, A., Ramirez-Asís, E., Asís-López, M., Flores-Albornoz, J. and Phasinam, K., 2021. Classification and prediction of student performance data using various machine learning algorithms. Materials today: proceedings.

[7] Sultana, J., Sultana, N., Yadav, K. and AlFayez, F., 2018, April. Prediction of sentiment analysis on educational data based on deep learning approach. In 2018 21st Saudi computer society national computer conference (NCC) (pp. 1-5). IEEE.

[8] Mohamed, M.H. and SAYED ABDELGABER, L.A.E., 2022. A Proposed Intelligent Model For predicting Student Performance Using Sentiment Analysis Techniques. Journal of Positive School Psychology, pp.7593-7606.

[9] Moreno-Marcos, P.M., Alario-Hoyos, C., Muñoz Merino, P.J., Estévez-Ayres, I. and Kloos, C.D., 2018, April. Sentiment analysis in MOOCs: A case study. In 2018 IEEE Global Engineering Education Conference (EDUCON) (pp. 1489-1496). IEEE.

[10] Aggarwal, D., Mittal, S. and Bali, V., 2021. Significance of non-academic parameters for predicting student performance using ensemble learning techniques. International Journal of System Dynamics Applications (IJSDA), 10(3), pp.38-49.

[11] Khan, A. and Ghosh, S.K., 2021. Student performance analysis and prediction in classroom learning: A
review of educational data mining studies. Education and information technologies, 26(1), pp.205-240.

[12] Sultana, J., Rani, M.U. and Farquad, M.A.H., 2019. Student's performance prediction using deep learning and data mining methods. Int. J. Recent Technol. Eng, 8(1S4), pp.1018-1021.

[13] Yaacob, W.F.W., Nasir, S.A.M., Yaacob, W.F.W. and Sobri, N.M., 2019. Supervised data mining approach for predicting student performance. Indones. J. Electr. Eng. Comput. Sci, 16(3), pp.1584-1592.

[14] Kiu, C.C., 2018, October. Data mining analysis on student's academic performance through exploration of student's background and social activities. In 2018 Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA) (pp. 1-5). IEEE.

[15] Punlumjeak, W., Rachburee, N. and Arunrerk, J., 2017.Big data analytics: Student performance prediction using feature selection and machine learning on microsoft azure platform. Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 9(1-4), pp.113-117.

[16] Zohair, A. and Mahmoud, L., 2019. Prediction of Student's performance by modelling small dataset size. International Journal of Educational Technology in Higher Education, 16(1), pp.1-18.

[17] Punlumjeak, W. and Rachburee, N., 2015, October. A comparative study of feature selection techniques for classify student performance. In 2015 7th International Conference on Information Technology and Electrical Engineering (ICITEE) (pp. 425-429). IEEE.

[18] https://towardsdatascience.com/understanding-8-types- of-cross-validation-80c935a4976d.

[19]https://www.pluralsight.com/guides/validating- machine-learning-models-scikit-learn.

AUTHOR DETAILS

Neha Singh was born in Gorakhpur, India, in 1995. She received the B. Tech. degree in Computer Science Engineering from Institute of Technology and Management Gida, Gorakhpur, in 2018 and M. Tech. degree in Information Technology from the Madan Mohan Malaviya University of Technology, Gorakhpur, in 2020. She is currently pursuing Ph.D. degree in Information Technology from Madan Mohan Malaviya University of Technology, Gorakhpur. Her research interests include natural language processing and machine learning.

Prof. Umesh Chandra Jaiswal was born in 1967, India. He received the B. E. degree in Computer Science and Engineering from Madan Mohan Malaviya Engineering College, Gorakhpur, in 1988 and M. Tech. degree in Computer Science and Engineering from the Indian Institute of Technology, Delhi, in 1994. He has completed Ph.D. degree in Computer Science and Engineering, from Madan Mohan Malaviya engineering collage, Gorakhpur, in 2011. His research interests include natural language processing, design and analysis of algorithms, parallel Algorithms, machine learning.