Ensemble Feature Selection for Student Performance and Activity-Based Behaviour Analysis

Varsha Ganesh[1], S Umarani[2]*
Research Scholar, Department of Computer Science, SRMIST, Ramapuram, Chennai, India[1]
Professor, Department of Computer Science and Application (BCA), SRMIST, Ramapuram, Chennai, India[2]

***Abstract*—Analyzing students' behaviour during online classes is vital for teachers to identify the strengths and weaknesses of online classes. This analysis, based on observing academic performance and student activity data, helps teachers to understand the teaching outcomes. Most Educational Data Mining (EDM) processes analyze students' academic or behavioural data; in this case, the accurate prediction of student behaviours could not be achieved. This study addresses these issues by considering student's activity and academic performance datasets to evaluate teaching and learner outcomes efficiently. It is necessary to utilize a suitable method to handle the high dimensional data while analyzing Educational Data (ED), because academic data is growing daily and exponentially. This study uses two kinds of data for student behaviour analysis. It is essential to use feature reduction and selection methods to extract only important features to improve the student's behaviour analysis performance. By utilizing a hybrid ensemble method to get the most relevant features to predict students' performance and activity levels, this approach helps to reduce the complexity of the feature-learning model and improve the prediction performance of the classification model. This study uses Improved Principal Component Analysis (IPCA) to select the most relevant feature. The resultant features of the IPCA are given as input to an ensemble method to select the most relevant feature sets to improve the prediction accuracy. The prediction is done with the help of Residual Network-50 (ResNet50) is combined with a Support Vector Machine (SVM) to classify students' performance and activity during online classes. This performance analysis evaluates the students' behaviour analysis model. The proposed approach could predict the performance and activity of students with a maximum of 98.03% accuracy for online classes, and 98.06% accuracy for exams.**

Keywords*—Behaviour analysis; deep learning; educational data mining; student performance prediction; students activity monitoring; machine learning

I. Introduction

Educational Data Mining (EDM) [1] techniques help in understanding students' learning situations and improve the teaching support for better decision-making in the educational system. The modern education system, which is evaluated from offline learning to online teaching [2] and learning mode, assesses the outcomes [3] and teaching effects [4]. Online learning, which has been increasing significantly during the last decade, enables students to learn in a comfortable environment. Students are the core resources of any educational institution. The academic sectors must deal with many changing factors to offer quality education using offline and online systems. Management must implement innovative [5] and effective teaching and learner outcome evaluation methods [6] to improve the quality of their graduates. This also helps teachers to evaluate the learners' effects to understand their condition easily by analyzing students' online data such as activity and behaviours [7], concentration levels, and academic performance [8]. These behavioural changes also affect learners' academic performance. Hence it is necessary to keep track of students' learning patterns by monitoring their activities and analyzing academic performance [9]. EDM techniques [10], which help in performing this monitoring and analyzing task, use Machine Learning (ML) [11], Deep Learning [12], Statistics, and other data mining techniques to analyze student behaviours and predict their performance. In addition, the COVID-19 pandemic forced the education system [13] to continue regular learning and teaching actions online. These changes made EDM an emerging research field to make the teaching and learning process more effective for online learning environments.

A. Research Objective

Researchers have utilized both quantitative and qualitative methodologies, revealing that students frequently exhibit unforeseen behaviours throughout online class sessions. So, the management implements some preventive actions by using many online data analytics tools to control the students. These online teaching platform-based devices produce many student activity and academic performance-related data. Proper utilization of analytics techniques in these ED gives better analysis results to predict student behaviours. Analyzing students' behaviours during the online platform is a vital part of teachers identifying their strengths and weaknesses.

- This analysis of observed academic performance and student activity data helps teachers to understand the teaching outcomes.
- Many Educational Data Mining (EDM) studies focus on either academic or behavioural data, yet in this instance, accurate prediction of student behaviours remained elusive. This study addresses these issues by considering both datasets to evaluate teaching and learner outcomes efficiently.
- Employing an appropriate technique to manage the expanding high-dimensional educational data is essential due to its daily growth.
- A prediction model is designed to investigate the same data types simultaneously. But it does not correlate students' behaviours with academic performance. So, this study uses two different kinds of datasets for

*Corresponding Author.

performance analysis. Alternative methods of analyzing student behaviour and performance necessitate the implementation of an efficient model.

The Hybrid DL model is developed by combining the Convolutional neural network with ResNet50 to perform this analysis for the prediction task.

B. Paper Organization

The remainder of the article is organized as: Section II discusses the various author's research opinion on the student's performance prediction-based methods. Section III explains the student's activity and performance prediction methods adopted in this research. Section IV gives details about the hybrid ensemble feature selection model. Results and analysis and discussion are given in Section V and Section VI respectively. Finally, Section VII concludes the paper.

II. Literature Review

Erick Odhiambo Omuvan [14] et al. (2021) described irrelevant and redundant information negatively influencing and creating complexity during selection operations in the classification algorithm. Principal Component Analysis (PCA) is utilized to support the ML-based classification models to improve the performance by avoiding irrelevant and redundant features. It selects the best feature combinations from the original data to support the ML classifiers.

A. Jenul et al., 2021[15] presented a feature selection approach using the Repeated Elastic Net Technique (RENT) which uses an ensemble model with elastic net regularization. Each model is trained with different feature sets of data. It follows three strategies to evaluate the weightage distribution of features among all the elementary models, which leads to relevant feature selection with higher stability that improves the robustness of the final model. It also provides valuable data for the model analysis concerning identifying objects in the data that are difficult to predict during the training. The performance of the RENT, analyzed with different healthcare data, shows that RENT achieves better performance than other methods.

Wen Xiao et al. (2021) [16] developed a hybrid feature selection method for student performance prediction. It uses score-based feature ranking and a heuristic approach to build the RnkHUE algorithm. The heuristic search strategy and forward ranking from the Genetic Algorithm (GA) help to select the significant features from the students' dataset. Initially, it identifies the evaluation criteria based on considering student performance factors such as distance, information metrics, dependency, and consistency. The heuristic method finds the best subsets among all the features using the search strategy. Further, the selected candidate feature sets are used for feature selection to improve the prediction performance of the proposed approach.

Ali AI Zawqari et al. (2022) [17] developed a flexible feature selection model for student performance prediction in four categories of student performance data. This prediction framework uses two concepts: improving the prediction performance with feature selection, and skipping feature engineering. Initially, features are embedded continuously and applied directly on an Artificial Neural Network (ANN) to perform prediction. The second approach uses all the embedded features to perform feature reduction with the help of Random Forest (RF) before performing the prediction. The evaluation results show that the feature selection-based model helps the prediction model to obtain a better accuracy of 93% for dropout prediction. This model also obtained 86% accuracy prediction for students' pass grade and 88% prediction for distinction grade data.

Sing R et al., (2021) [18] prepared a comprehensive study on the performance of various feature selection methods on students' academic data. It discusses the different contemporary approaches broadly used to foresee the educational outcome of the under study. It brings forth the fact that the academic performance of the enrolled students in any course has some patterns. Moreover, the feature choice predicts student performance to obtain significant results.

R. Singh et al., 2020 [19], developed a Machine Learning (ML) based ensemble model to predict students' performance. This model utilizes the ensemble of Decision Tree (DT), K-Nearest Neighbour (K-NN), extra tree, and Naive Bayesian (NB) methods. It uses bagging-based boosting methods for performance prediction. The ensemble model accuracy is improved to 86.83% for the students' performance dataset. The results show that the NB performs well compared to other models. However, the complex structure of ensemble models failed to obtain a reliable accuracy level with NB.

Hussain et al., 2021 [20] prepared an automatic students' marks and grade forecasting framework using ML models. A Genetic Algorithm (GA) selects features from the students' dataset. The GA-selected parts are classified by Regression and DT classifier. The regression model achieved a dependable accuracy rate of 96.64%. However, with the escalating volume of data, scalability becomes a significant concern. The ML-based model requires further refinement to enhance its performance. So a deep learning-based regression model needs to be integrated.

Tarik A et al., 2021 [21], designed an ML model to predict Moroccan students' performance in the region of Guelumim Qued Noun through a recommendation system using artificial intelligence. The prediction model presented in their study indicates the baccalaureate mean as a function of many exploratory variables, such as grades and core subjects. The performance of linear regression, regression-based DT, and regression-based Random Forest (RF) models is evaluated. Among these three, DT with RF method obtained a maximum of 61.08% accuracy. However, poor model fitting led this combination to perform poorly for students' datasets.

Abellan-Abenza J et al., 2017 [22] introduced a surveillance system based on the human behaviour analysis technique. The current behaviour of a person is identified while crossing a surveillance camera. Various human behaviour expression image datasets are utilized for training the classifier. The behaviour identification is performed by combining the Convolutional Neural Network (CNN) with the Recurrent Neural Network (RNN).

Rastrollo Guerrero JL et al., 2020 [23] prepared a deep preview for predicting students' performance. This review focuses on identifying the students' classroom behaviour-based dropout prediction model. This study utilizes the image datasets for the analysis. Its review describes the various stages of the prediction processes to perform the dropout prediction.

Chowanda et al., 2021 [24], the performance of multiple machine learning models was evaluated on sentiment-related text datasets derived from students. Emotions of students were detected using Naive Bayes (NB), Generalized Linear Model (GLM), Support Vector Machine (SVM), Decision Tree (DT), Fast Large Margin (FLM), and Artificial Neural Network (ANN). Among these models, GLM achieved the highest accuracy rate of 0.902. While the emotions anger and joy were consistently identified with high accuracy, the classification of the emotions fear and sadness posed challenges for the classifier in emotion recognition tasks.

J Zhao et al., 2020 [25] used educational data analytics containing text enhancement phase, Synonyms Replacement (SR), Random Insertion (RI) of words, Random Swap (RS), and Random Delete methods were performed while extracting the text emotion reorganization. This reorganization has been achieved with the help of a Directed Acyclic Graph (DAG) with an SVM model to train the various textual sentiment data.

D Selvapandian et al., 2020 [26] introduced an Efficient Fusion based Neural Network (EF-NN) model for sentiment analysis from feedback documents of students. This hybrid model integrates the SVM classifier with CNN. Students' feedback data set is extracted based on attribute features like the interaction between the student, examination, and notes given.

BHKT H.M. Perera et al., 2021 [27], an innovative e-learning surveillance system was introduced to assist instructors in online exam monitoring. This system is capable of identifying low engagement levels, detecting suspicious activities, and flagging instances of multiple logins at the onset of online exam sessions. What sets this approach apart is its ability to not only predict academic performance but also forecast learning behaviours. Consequently, it enhances the accuracy of performance prediction among students, thereby contributing to improved assessment quality.

Saba T [13] et al., 2021 [28] developed an automatic exam monitoring system to assist instructors in monitoring students without being present in the exam centres. It builds a deep model to form a 46-layered CNN model. The extracted features are used for selecting significant features using Atom Search Optimization (ASO) to improve the prediction performance of variants of SVM and KNN models; among these, KNN model obtained the best accuracy rate (93.88%).

A. Problems Identified

This review identifies that ML models suffer from fitting issues while handling different kinds of educational data. This has been overcome by adopting suitable feature selection, and reduction approaches to manage high or low volumes of student data.

- Moreover, educational data is of different types based on the kind of analysis. However, most ML models can perform well on similar types of educational data.
- It is necessary to develop a hybrid model to analyze multiple types of educational data. This study has developed a mixed method to lessen the fitting issues.
- Generally, hybrid methods take longer for data processing. Because it combines the features of two methods, this study utilizes the ResNet-50 method to improve the time complexity during the prediction process.
- The linear SVM method generally performs well for educational data in low dimensions. Nevertheless, achieving a better balance in the model is necessary as it is currently influenced by irrelevant features within the dataset.
- So it is necessary to develop an effective feature selection method to avoid fitting issues which would also help to reduce the loss rate and improve the prediction performance by using two levels of the feature selection approach.

B. Research Contribution

- The first stage performs the feature reduction using the IPCA method to remove the irrelevance of the students' behaviour-related academic performance and online activity data.
- The second stage uses the reduced features for the most relevant feature, which supports improving the students' performance and activity prediction performance using the ML-based ensemble feature selection method.
- The ML models used in the ensemble methods are chosen based on their performance analysis on student datasets in recent studies.
- The ensemble method strengthens the weak ML methods used in this approach in more potent ways by using ensemble stacking. The ensemble stacking method identifies the most relevant feature combination for behaviour analysis.

The functionality of the proposed students' performance and online activity-based behaviour prediction approach is described in a subsequent section.

III. Student Performance and Activity Based Behaviour Analysis Approach

This section discusses the functionalities of various methods used in behaviour analysis approaches. This Student data contains four phases: data collection, preprocessing, feature selection, and prediction analysis. Initially, the students' performance and online activity datasets utilized in this section are taken from two publicly available datasets. The preprocessing stage utilizes the one-hot encoding method to normalize categorical data. The third phase develops a feature selection method using a hybrid ensemble method; it combines IPCA with the Ensemble feature method. Finally, the prediction phase uses ResNet-50 to train the model, the SVM classifier to test the data and classify the students' behavioural data. The general flow of the four phases is depicted in Fig. 1.

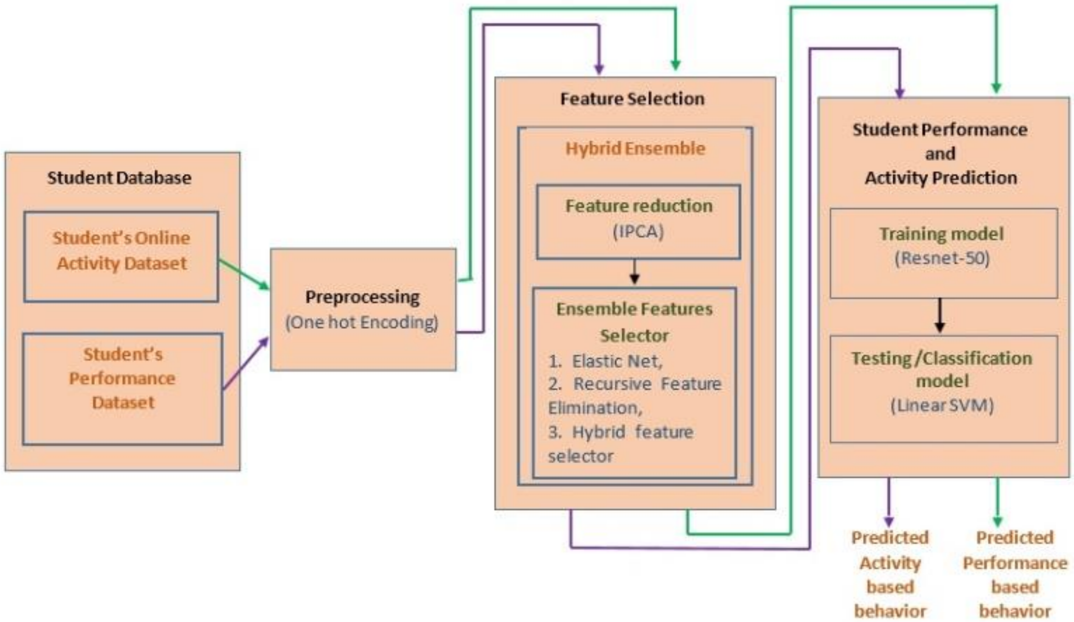


Fig. 1. Workflow of the student's performance and activity-based behaviour prediction approach.

A. Data Sources

This analysis uses two different datasets for students' behaviour and performance prediction. The student activity and academic data are collected from publicly available open databases. The academic dataset is collected from the Kaggle [29] database, which covers 480 instances and 16 attributes. These features are categorized into three groups: (1) Demographic features such as gender and nationality; (2) Academic background features such as educational stage, grade level and section; (3) Behavioural features such as raised hands-on class, opening resources, answering survey by parents and school satisfaction.

Student activity datasets are taken from the UCI repository [30], which is publicly available for educational research and contains log information for each student. Generally, these data, captured using various LMS tools, is given on a per session basis, per student basis, and exercise basis. It is comprised of six sessions of data. Each exercise file contains the session's start, end, and learning activity. The dataset consists of 230318 records and 13 attributes, recorded and taken for analysis from 115 subjects.

The analysis divides both datasets for training and testing the model. The ResNet50 network's training and SVM model's testing phases use 70% and 30% of students' activity data respectively. The student's academic performance data is divided into 75% for training and 25% for testing.

B. Preprocessing

These datasets contain both numerical and categorical data. So, it is necessary to use proper preprocessing steps to normalize the datasets. This study uses one hot encoding method to normalize the raw datasets. It represents the categorical data as numerical data to train the ML models and improve the model performance by providing more information about the unlimited data.

Every categorical data in the datasets is part of a given categorical feature written in vectors, consisting only of 0 and 1. It converts into a vector whose elements are only 0's or 1's. Each word is encoded uniquely in this method. It allows the term can be identified uniquely by its one-hot vector. Table I shows the uniquely converted code for student genders for the labels. The male, female, and transgender data are converted as 100, 010, and 001, respectively.

However, this one hot encoding method increases the dimensionality of the dataset and may lead to overfitting and sparse data issues. So it is essential to use proper feature reduction approaches to reduce the dataset's dimensionality and identify the most significant students' academic and activity-related feature information.

TABLE I. SAMPLE DATA NORMALIZATION USING ONE HOT ENCODING

S. No.	Male	Female	Trans
1	1	0	0
2	0	1	0
3	0	0	1

IV. HYBRID ENSEMBLE FEATURE SELECTION METHOD

Normalization of students' records avoids the overfitting by adopting a suitable feature reduction technique. Since ML-based classification models must be more balanced due to irrelevant dataset features, it is necessary to develop an effective feature selection method to avoid fitting problems. This also helps reduce the loss rate and improve the prediction performance. Two levels of the feature selection approach can achieve it. The first stage is performing the feature reduction using the IPCA method to remove the irrelevant behaviour-related academic performance and online activity data. The second stage uses the reduced features for the most relevant feature, which supports improving the students' performance and activity prediction performance using the ML-based ensemble feature selection method. The ML models used in the ensemble methods are chosen based on their performance analysis on student datasets in recent studies.

The ensemble method strengthens the weak ML methods used in this approach in more potent ways by using ensemble stacking, which identifies the most relevant feature combination for behaviour analysis. This study uses the hybrid ensemble method to perform the feature selection, which combines the Improved Principal Component Analysis (IPCA) with the ensemble feature selection method to reduce the dimensionality of the students' record at the initial level. The ensemble method is designed to select the most relevant features to predict academic performance and activity datasets.

Fig. 2 illustrates the hybrid feature selection method using IPCA and ensemble method. The IPCA is utilized to identify the reduced set. Then the resultant sets are used in the ensemble method to select the relevant feature set, which influences the classification model to improve the prediction accuracy.

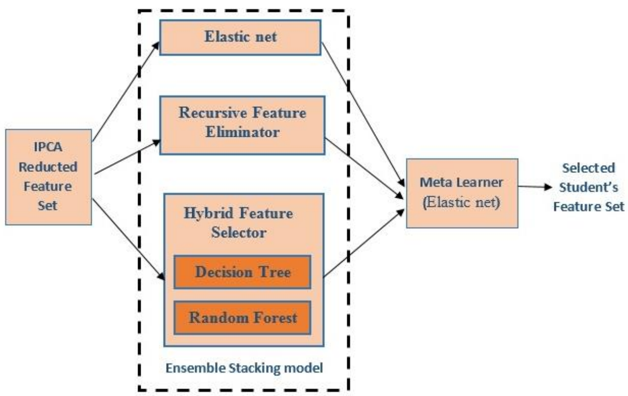


Fig. 2. Ensemble feature selection.

A. Principal Component Analysis

Any high-dimensional dataset can use this PCA to reduce the dimensionality. It rotates the cordiality system to convert a large dataset of possible interrelated indicators into a smaller set of linear correlated indicators. Each feature's interrelationship is ensured using the PCA to examine the correlation between indicators. The standardization process in traditional PCA leads to loss of dispersion degree information of the original dataset. These issues can be avoided by utilizing the IPCA approach for feature reduction, which performs the following six steps for measuring students' academic and activity features: standardizing the input matrix, computing correlation coefficient, computing eigenvalues and eigenvector, defining principal component, identifying the indicators belonging to the determined PCs, and calculating the component score.

1) Standardization of the input matrix: The number of input data samples is n and m indicators are considered for the feature reduction. The input feature vector is represented as X_{nxm}. The standardization assures perfect comparability between indicators. The original feature matrix X_{nXm} is transformed as Y_{nxm} with zero mean and unit variance.

$$y_{ij} = (x_{ij} - \bar{x}_j)/S_{x_j} \tag{1}$$

The value of i and j in eq(1) is initiated as i=1,2,..n and j=1,2,..m. The representation x_{ij} is the jth indicator value of the ith sample in the feature matrix of student records X_{nxm}, $and\ x_j$ is the jth indicator of X_{nxm}. Then, the mean and standard deviation of x_j is represented as $\bar{x}_j$ and S_{x_j} respectively. The standardized value of x_{ij} is y_{ij}.

2) Computing the correlation coefficient: Correlation information between the indicators is computed using the correlation coefficient (ϕ).

$$\phi = (\rho_{y_j,y_k})_{mXm} = \frac{1}{n-1} Y^T Y \tag{2}$$

In eq(2), $y_j\ and\ y_k$ are the jth and kth column vectors of Y_{nxm}, respectively. The expression ρ_{y_j,y_k} denotes the correlation coefficient between ($y_j\ and\ y_k$), which are the jth and kth indicators.

3) Computation of eigenvalues and eigenvector:

$$|\phi - \lambda E| = 0 \tag{3}$$

The eigenvalues and eigenvectors of ϕ are obtained using eq(3).

All the eigenvalues are arranged in descending order as $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_j, \dots, \lambda_m$. Each eigenvalue has its corresponding eigenvector.

$$\phi b_j = \lambda_j b_j \tag{4}$$

According to eq(4), the unit vector(b_j) corresponds to λ_j and assigned $\sum_{i=1}^{m} {b_{ij}}^2 = 1$. Therefore, $B = (b_1, b_2, \dots, b_m)$ is a unit orthogonal matrix consisting of all the units of the eigenvector.

4) Defining principal component: The number of principal components is generally determined according to the criterion of eigenvalues > 1, which is along with the screen plot, or cumulative percentage variance < 80%, which is constructed using the values of a_j and β_p in Eq. (5) and Eq. (6).

$$a_j = \frac{\lambda_j}{\sum_{j=1}^{m} \lambda_j} \tag{5}$$

The a_j in Eq. (5) is the percentage variance of ith Principal Component (PC).

$$\beta_p = \sum_{k=1}^{m} \lambda_k / \sum_{j=1}^{m} \lambda_j \tag{6}$$

Eq. (6) is used to calculate the cumulative percentage variance (β_p) of p PC and the $p \leq m$. Whenever $\beta_p \geq 80\%$ first appears, the PC (p) is selected.

5) Identifying the indicators belonging to the determined PCS: The factor loading of each indicator on each persistent PC is

$$\theta_{jk} = b_{jk}\sqrt{\lambda_k} \tag{7}$$

The correlation coefficient θ_{jk} between the jth indicator and kth PC is calculated by Eq. (7). λ_k indicates the eigenvalue corresponding to kth PC, and b_{jk} is the jth value of b_k. It considers the indicator with $|\theta_{jk}| \geq 0.5$, indicating that the jth indicator belongs to kth PC.

6) Calculating component score: Every PC is a weighted linear combination of all indicators, and the PC scores $(f_1, f_2, f_3, \ldots, f_p)$ are obtained using $F = YB$.

$$w_k = \frac{\lambda_k}{\sum_{j=1}^{m} \lambda_j} \tag{8}$$

Besides, the percentage of variation explained by each PC is used as a weight CF, calculated using Eq. (8).

$$CF = \sum_{k=1}^{p} w_k f_k \tag{9}$$

The total component score CF is obtained based on Eq. (9).

$$z_{ij} = (x_{ij} - \bar{x}_j)/\varepsilon_{x_i} \tag{10}$$

Eq. (10) in the traditional PCA set the variance of each indicator to 1. This reduces the influence of dispersion degree difference on PCs. So, improved standardization is used in this study. Eq. (10) is used to compute the enhanced standardization, where $\varepsilon_{x_i} = \max(x_j) - \min(x_j)$ and the $\varepsilon_{x_i} > 0$, z_{ij} is the standardized value of x_i.

$$\bar{z}_j = \sum_{i=1}^{n} \frac{x_{ij} - \bar{x}_j}{\varepsilon_{x_i}} / n = \frac{\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)}{n \varepsilon x_i} \tag{11}$$

$$S_{z_j} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (z_{ij} - \bar{z}_j)^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{x_{ij} - \bar{x}_j}{\varepsilon_{x_i}}\right)^2} = \frac{S_{x_i}}{\varepsilon_{x_i}} \tag{12}$$

The mean and the standard deviations of *the* jth indicator (z_j) in Z_{nxm} are obtained using Eq. (11) and Eq. (12) respectively.

$$\rho_{z_j, z_k} \left. \begin{aligned} &= \frac{C_{z_j, z_k}}{S_{z_j}, S_{z_k}} = \frac{1}{n-1} \sum_{i=1}^{n} (z_{ij}, \bar{z}_j)(z_{ik}, \bar{z}_k)/S_{z_j}, S_{z_k} \\ &= \frac{\frac{1}{n-1} \sum_{i=1}^{n} \frac{(x_{ij}, \bar{x}_j)}{\varepsilon_{x_j}} \frac{(x_{ik}, \bar{x}_k)}{\varepsilon_{x_k}}}{\left(\frac{S_{x_j} S_{x_k}}{\varepsilon_{x_j} \varepsilon_{x_k}}\right)} \\ &= \frac{1}{n-1} \sum_{i=1}^{n} (x_{ij}, \bar{x}_j) \frac{(x_{ik}, \bar{x}_k)}{(S_{x_j}, S_{x_k})} = \rho_{x_j, x_k} \end{aligned} \right\} \tag{13}$$

According to Eq. (13), the Z_{nxm} and X_{nxm} are the same correlation coefficient matrix. It indicates that the improved standardization methods retain correlation information of all indicators. Notably, the dispersion degree difference of all hands is partly retained according to the standard deviation in Eq. (12).

B. Ensemble Features Selection

1) Elastic net: Elastic net is a combination of Ridge and Lesso regression methods which are popular for regularizing variants of linear regression. Lesso used the penalty L1, and Ridge used the penalty L2 method. The specialty of the elastic net is that it uses both L1 and L2 for penalty regularization.

$$ElasticNet = MSE(y, y_{pred}) + a_1 \sum_{i=1}^{m} |\theta_i| + a_2 \sum_{i=1}^{m} |\theta_i| \tag{14}$$

The elastic net function is expressed as in Eq. (14) to compute the loss value between actual (y) and predicted output class with the loss value of ridge regression $(a_1 \sum_{i=1}^{m} |\theta_i|)$ and loss value of Lesso regression $(a_2 \sum_{i=1}^{m} |\theta_i|)$. The control parameters are a_1 and a_2 to control the L1 and L2 penalty respectively. The number of optimal parameters is represented as.

2) Recursive Feature Eliminator (RFE): RFE is a wrapper-type feature selection method. In contrast with filter-based feature selection that scores each feature and selects those features with the most significant score, RFE searches for a subset of features by starting with all features in the training dataset and successfully removing features until the desired number remains. It has been used to fit the ML algorithm. Rank features by importance. It gives an external estimator that assigns weights to features. The estimator is trained on the initial set of features, and the features' importance is obtained through any specific attribute. Discard the less critical features and re-fit the model. These steps are repeated until the preferred number of features is eventually reached.

3) Hybrid method: The hybrid method combines the (i) Decision Tree (DT) and (ii) Random Forest (RF) methods. The single DT method is unsuitable for high dimensional data, so the RF method is combined with the DT to improve the performance of the feature selection model.

a) Decision tree (DT): DT is a graphical representation for all possible solutions to a problem based on given conditions. DT is a tree-structured method where internal nodes indicate the dataset's features, branches show the decision rules, and the leaf node indicates the prediction outcome. The decision nodes contain multiple units and make any decisions. It does not have any additional nodes. It asks questions to split the tree into subtrees based on the answers. The main issue in the DT algorithm is the best attribute selection for root and sub-nodes. It uses two popular methods to perform the best attribute.

$$IG = Entropy(S) - (Weighted\ average * Entropy(each\ features)) \tag{15}$$

The DT algorithm improves an attribute's Information Gain (IG) using Eq. (15). The attribute or node having the highest IG is split first. The total number of student records is represented as S.

$$Entropy(S) = -P(yes) log2\, P(no) log2\, P(no) \tag{16}$$

The impurity of an attribute is specified randomly in data by estimating the entropy (Entropy(S)) in eq(16). Probability of yes and no is represented as P(yes) and P(no).

DT Algorithm
Step 1: Begin the tree with the root node; it contains the complete dataset.
Step 2: Select the best attributes in the dataset using the attribute selection method.
Step 3: Spilt the S into subsets which contain possible values for the best attributes.
Step 4: Create a DT node which contains the best attributes.
Step 5: Repeatedly make new decision trees using the subsets of the dataset created.

The DT algorithm is applied to the dataset to select the best feature sets. However, the DT's performance falls with greater number of samples.

b) Random Forest: RF is simply a collection of DTs whose results are aggregated into one final result. RF is a strong modelling technique and much more potent than a single DT. It aggregates many DTs to limit overfitting and errors due to bias. It can restrict overfitting without significantly increasing error due to bias. It reduces variance by training on different samples of the data. Another method is by using a random subset of features. Each tree can utilize a specified number of random features. More trees in the RF include many or all features. The presence of many features helps in limiting the errors due to bias and due to variance. If features are not selected randomly, base trees in the forest correlate highly. Since some features are partially predictive, many base trees can choose the same features. Many of these trees contain the same features; it cannot be combined error due to variance. The proposed hybrid ensemble method uses academic performance and online activity datasets to evaluate the performance of the hybrid ensemble method.

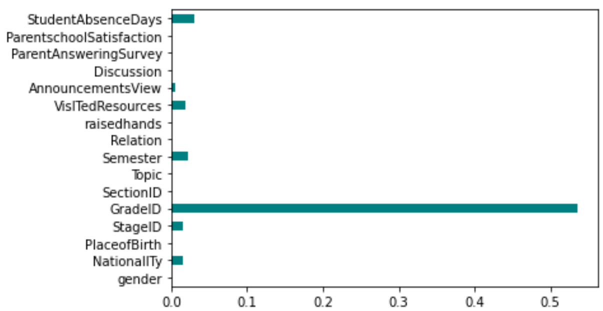


Fig. 3. Feature importance graph for students' academic dataset using the hybrid ensemble method.

The students' Academic dataset shown in Fig. 3 contains 16 features; the ensemble method selects seven features to predict students' academic performance with a higher accuracy level.

The feature importance graph in Fig. 4 depicts the feature selection results of the hybrid ensemble method. The activity dataset contains a total of 12 features. The hybrid ensemble method selects seven features to predict the students' activity-based behaviours during the online sessions.

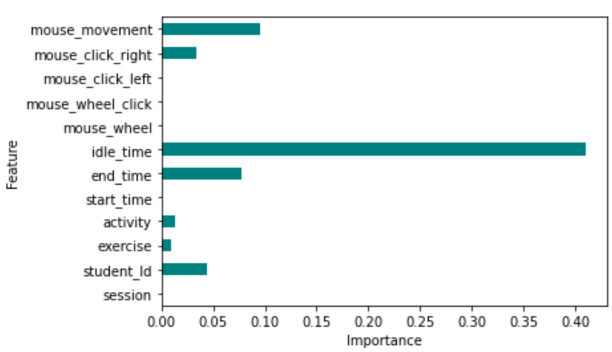


Fig. 4. Feature importance graph for students' online activity dataset using the ensemble method.

C. RESNET50 Trained SVM Model for Prediction

The students' academic performance-based behaviour prediction and online activity-based behaviour prediction are performed using the CNN-trained SVM model. Residual Network-50 (ResNet 50) is a kind of CNN. The 50-layer network model contains 48 Convolutional Layers (CL), 1 max pooling, and one average pooling layer to perform the prediction. The architecture of the ResNet 50-trained SVM model is depicted in Fig. 5. It follows two main rules to process the data. Such amount of filters in each layer is the same contingent on the size of the output feature map; if the feature map's size is split, it has twice the number of filters to preserve the time complexity of each layer. The 50 layers' network utilizing the 1x1 CL helps to reduce the number of parameters and matrix multiplication operation. This feature enables the model to train faster at each layer. A stack of three layers is used in this model. It has one 7x7 kernel convolutional alongside 64 other kernels with 2-sized strides and one 2-sized stride in the max pooling layer. More 9 layers are 3 3x3, 64 kernel convolution and 3 1x1, 256 kernels, and 1x1, 256 kernels. These three kernels are repeated thrice consequently. They are succeeded by 12 layers with 1x1, 128 kernels, 3x3, 128 kernels, and 1x1, 512 kernels. These three kernels are consequently repeated four times. Then 18 more layers with 1x1, 256 cores, 3x3, 256 cores, and 1x1, 1024 cores repeated 6 times. Final 9 more layers with 1x1, 512 cores, 3x3, 512 cores, and 1x1,2048 cores iterated thrice. Followed by this, 50 layers of average pooling and fully connected layers with 1000 nodes using SoftMax activation are incorporated.

The deep model performs better with larger training sample sizes, but the amount of data utilized in this study could be more extensive in dimension and size. So the ResNet-50 model-trained results are used by a machine learning model to improve the performance by utilizing the transfer learning concept to improve the analysis model of students' behaviour. It uses the ResNet 50 network to train the students' behavioural features, and the SVM classifier is utilized to predict students' behaviours. This concept also helps to achieve higher performance even if the model is trained with a small sample of student data.

Support Vector Machine, a popular ML model for classification and regression problems, assigns the newly entered samples to one of the trained categories. So, it is called a non-probabilistic binary linear classifier. It efficiently performs the classification task by applying the proper kernel tricks. SVM classifier separates data points with different class labels using a hyperplane with the maximum amount of margin. The hyperplane acts as a decision boundary. Sample data points are called Support Vectors (SV). This data defines the hyperplane by estimating the margin. Separation gap between the two lines on the closest data points is estimated as a perpendicular distance from the line to data points or SV. The SVM tries to improve the separation gap to get the maximum margin. Sometimes, the sample data points are so discrete that it is not conceivable to distinguish using the hyperplane. In such a situation, kernel tricks transform the input space to a higher dimension space by using a mapping function to transform the input space. The linear separation method is applied to the data points to separate them. This student behaviour analysis model uses the linear kernel to map the students' data to higher dimensional data.

$$K(\bar{x}) = \begin{cases} 1 \; if \; \|\bar{x}\| \leq 1 \\ 0 \; otherwise \end{cases} \tag{17}$$

Eq. (17) is the linear kernel (K(x̄)), which is used to map the students' behavioural data (x̄). The mapping range used by the linear kernel is [0, 1]. Data points which are ⩽1 are mapped as 1; others are considered as 0. This linear kernel is used whenever the input data are need to be separated linearly. It is mainly used for text data classification problems and whenever many data features are in a dataset. The students' behavioural dataset contains categorical data points. So, this study uses linear kernel tricks to map the students' behavioural features.

Moreover, it is also utilized to speed up the classification, since it is required to optimize the regularization parameter. The performance of the selected features by hybrid ensemble method based on the students' performance and online activity are predicted to identify the students' behaviour using the CNN-trained SVM model. The performance analysis is discussed in the consequent section.

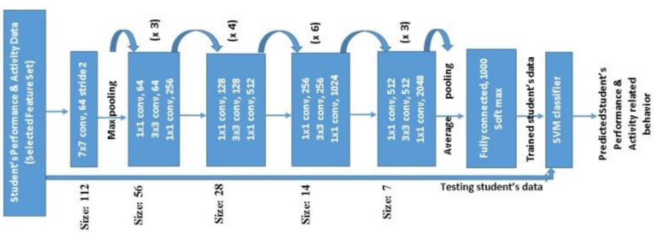


Fig. 5. The general structure of the RESNET50 trained SVM model.

V. Result Analysis

This section analyzes the performance analysis of the proposed hybrid ensemble method's Feature Selection (FS) accuracy for student behaviour analysis using a CNN-trained SVM model. The competence of the hybrid ensemble method is evaluated by comparing the various ML methods and hybrid feature selection methods such as PCA [14], Elastic Net [15], Genetic PSO-ACO based RNKHEU [16], and ANN with RF[17], RF[19], and DT[21]. These comparison methods are considered for analysis based on their superior performance on student datasets in recent times. Different evaluation metrics such as accuracy, precision, recall, f-score, specificity, and sensitivity are utilized to analyze the influence of the FS method on improving the performance of the prediction model.

Table II compares the FS outcomes of the hybrid ensemble method with other FS methods. It contains information on the number of features used by all the FS methods for the analysis and the number of selected features. It shows that the Hybrid Ensemble Feature selection method chose 7 as the most relevant informative feature to predict the students' performance with a higher accuracy rate.

Table III shows the FS results for students' activity data using the hybrid ensemble method compared with other FS methods. It comprises the evidence of the number of features selected by all the FS methods considered. Totally seven relevant features give a higher accuracy rate, thus significantly improving the prediction performance of students' online activity. Thus, the reduction of figures by the IPCA algorithm help the Ensemble FS method to identify the most relevant features to improve the performance.

Table IV displays the prediction outcome of students' academic and online activity datasets before and after applying the ensemble method. The IPCA method reduces the number of required features from the students' datasets. The reduced feature information helps the ensemble FS method to improve the prediction performance by selecting the most relevant features from the reduced feature sets.

TABLE II. FEATURE SECTION OUTCOMES FOR STUDENTS' ACADEMIC DATA

Methods	Total number of Features	Selected Features
Random Forest Classifier[19]	17	5
Decision Tree[21]	17	4
Elastic Net[15]	17	4
Genetic PSO ACO RNKHEU[16]	17	6
ANN with RF[17]	17	6
PCA[14]	17	5
Hybrid Ensemble Feature selection(Proposed)	17	7

TABLE III. FEATURE SELECTION OUTCOMES FOR STUDENTS' ONLINE ACTIVITY DATA

Methods	Total number Features	Selected Features
Random Forest Classifier[19]	13	5
Decision Tree[21]	13	4
Elastic Net[15]	13	4
Genetic PSO ACO RNKHEU[16]	13	5
ANN with RF[17]	13	6
PCA[14]	13	6
Hybrid Ensemble Feature selection(Proposed)	13	7

TABLE IV. PREDICTION PERFORMANCE BEFORE AND AFTER HYBRID ENSEMBLE METHOD FOR STUDENTS' ACADEMIC AND ONLINE ACTIVITY DATASETS

Dataset	Accuracy	Precision	Recall	F-Measure
Results obtained before feature selection				
Students' academic dataset	77.22	78.34	73.29	74.87
Students' activity dataset	78.03	78.71	74.18	75.23
Results obtained after feature selection				
Students' academic dataset	97.78	98.13	97.23	97.57
Students' activity dataset	96.34	96.34	96.45	96.35

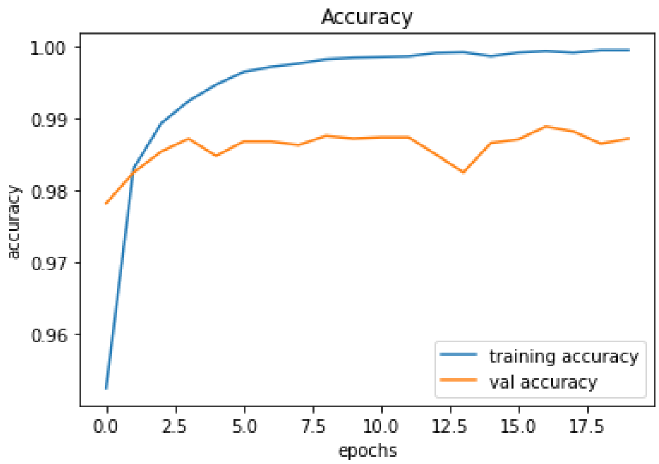


Fig. 6. Training and testing Accuracy rate obtained by the ResNet-50 trained SVM classifier.

Fig. 6 depicts the students' performance accuracy achieved by the ResNet-50 during the training process and the SVM model's testing or prediction accuracy rate. This study uses ResNet-50 network to train the students' behavioural features, and the SVM classifier is utilized to predict the students' behaviours. It reposts another related task for faster prediction. This concept helps achieve higher performance even if the model is trained with a small sample of student data.

Fig. 7 illustrates the students' performance loss rate attained by the ResNet-50 during the training and testing process using the SVM model, which reposts another related task for faster prediction. This concept helps to reduce the loss rate even if the model is trained with a small amount of sample students' data. Moreover, the prediction outcome proves that the IPCA method reduced features supports the ensemble FS method to select more relevant features and hence minimize the prediction loss.

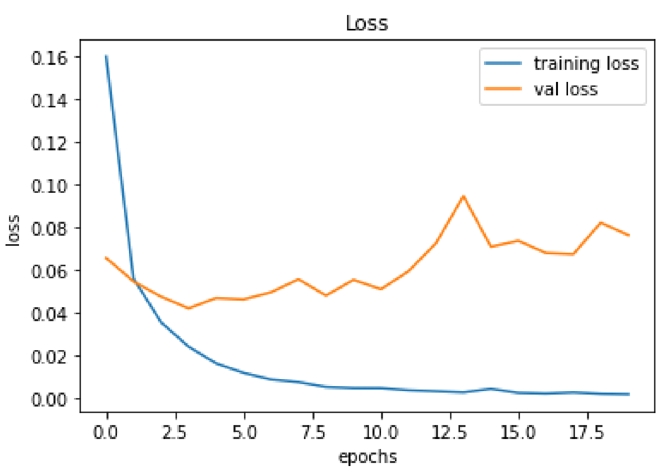
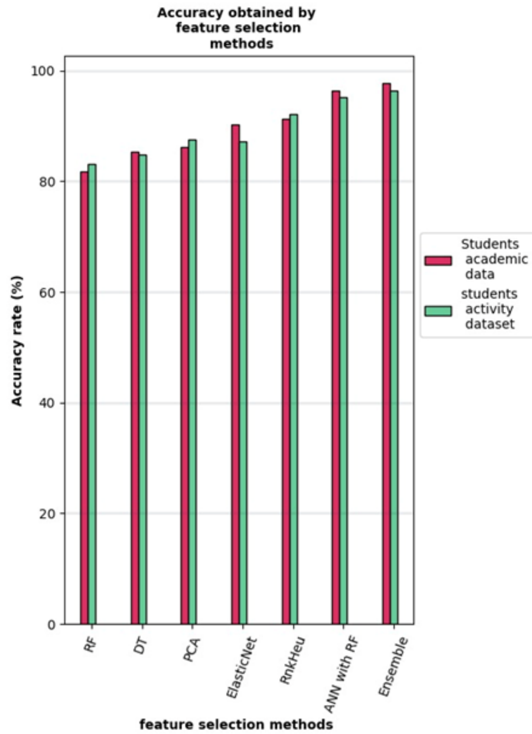
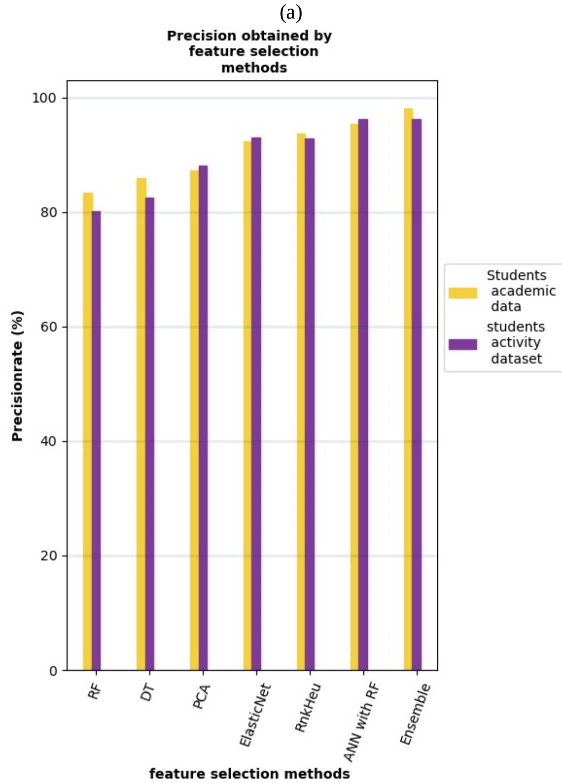


Fig. 7. Training and testing loss rate of RESNET-50 trained SVM classifier.

Fig. 8(a) illustrates the accuracy rate obtained by six FS methods along with the ensemble FS method. The comparison results depict that by introducing the Hybrid ensemble FS methods, the ensemble model achieves a higher level of accuracy on measuring both student performance and online activity datasets. This has been achieved by introducing the Hybrid ensemble FS methods. The ensemble method achieveda maximum accuracy rate of 97.78% for students' academic data and 96.34% for students' activity data. Moreover, the comparison results depict that the ensemble model achieves a level of performance higher than the comparison methods.



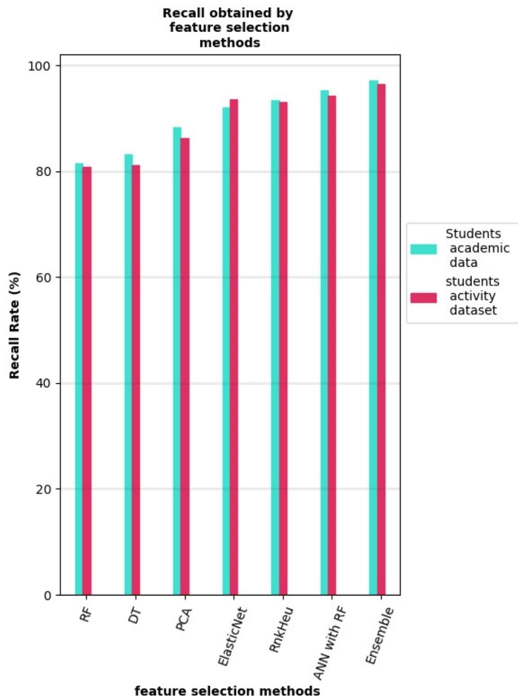
(a)



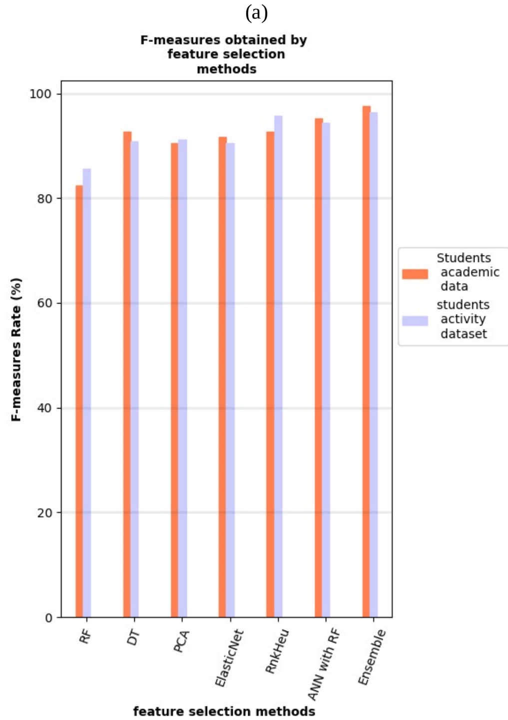
(b)

Fig. 8. (a) Illustrates the accuracy rate comparison, (b) Demonstrates the precision rate comparison.

Fig. 8 (b) demonstrates the precision rate gained by six FS methods and the hybrid ensemble FS method. The ensemble method achieved a maximum of 98.13% precision rate for students' academic data and 96.34% for students' activity. The comparison results depict that the ensemble model reaches a level of performance higher than the comparison methods. This has been achieved by applying the Hybrid ensemble FS method to select relevant features for analysis.



(a)



(b)

Fig. 9. (a) Recall rate comparison, (b) F-Measure rate comparison.

Fig. 9 (a) illustrates the Recall rate obtained by six FS methods along with the ensemble FS method. The comparison results reveal that the ensemble model achieves a higher level of performance both on student performance and online activity datasets. The ensemble method achieved a maximum recall rate of 97.23% for students' academic data and 96.45% for students' activity data. Moreover, the recall rate comparison results reveal that the ensemble model achieves a level of performance higher than the comparison methods.

Fig. 9 (b) demonstrates the F-score rate of six FS methods and the hybrid ensemble FS method for the two students' behavioural-related datasets. The ensemble method achieves a maximum of 98.13% f-measure rate for students' academic data and 96.34% for students' activities. The F-score measure comparison outcomes depict that the ensemble model performs better than the comparison methods.

The comprehensive competence analysis presented in this section reveals that the ensemble method significantly enhances the performance of the behaviour prediction model compared to conventional methods across various educational datasets. This validates that the proposed approach for analyzing students' behaviour effectively fulfills its research objective by enhancing overall performance and managing behavioural data adeptly.

VI. Discussion

This section shows that the overall performance of the prediction model is improved compared to other behaviour and academic performance prediction models. The improved accuracy prediction is observed for academic performance and student activity data. The results can be used by teachers to understand their teaching outcomes. This study used both quantitative and qualitative methodologies, revealing that students frequently exhibit unforeseen behaviours throughout online class sessions. These methodologies support the management in implementing students' performance monitoring actions and taking preventive actions. This approach can integrate with online data analytics tools to control the student's behaviour during online sessions. This prediction model uses online teaching platform-based monitoring devices produced data (student activity and academic performance-related data). Analyzing students' behaviours is one of the vital parts of teachers to identify their strengths and weaknesses.

This improved accuracy, precision, recall and f-score rate for both datasets reveals that the ensemble of different feature learning models efficiently uses the benefits of different models' accuracy to strengthen the weak model feature learning performance. The novel method outcome is reflected in the increased recall rate compared with existing methods. It effectively addresses these issues in predicting teaching and learner outcomes. Employing a suitable dimensionality technique to manage the expanding high-dimensional educational data is essential due to the daily growth of educational data. The previous prediction model (considered from the literature review) is mostly designed to handle single educational data (like behaviour data or academic performance data). However, it does not correlate students' behaviours with academic performance. So, this study uses two different kinds of datasets for performance analysis. Alternative methods of analyzing student behaviour and performance necessitate the implementation of the ensemble feature learning integrated prediction model.

The overall result and discussion section show that the ensemble feature learning integrated DL model is performing effectively on these students' data by combining multiple model features to strengthen the overall outcome.

VII. Conclusion

The study's main objective is to improve the overall performance of the student's behavioural data analysis. This analysis supports the educational sectors to incorporate innovative methods to enhance students' learning outcomes. This study contributes a feature selection method to measure the students' behaviour-related performance and online activity data. The performance evaluation conducted in this study demonstrates that the hybrid ensemble method outperforms the comparison benchmarks. The IPCA truncated features support the different weak ML methods to strengthen the feature selection performance using feature stacking methods. The selected features help the ResNet-50 trained SVM model to achieve higher prediction outcomes for students' academic performance and online activity. The analysis results show that the ensemble method obtained a maximum of 97.78%, 98.13%, 97.23%, and 97.57% as the accuracy rate, precision rate, recall rate, and f-measure rate, respectively, for behaviour-related students' academic performance data. The ensemble method achieves a maximum accuracy rate (96.34%), precision rate (96.34%), f-score rate (96.35%), and recall rate (96.45%) for students' online activity data. The efficiency analysis shows that the ensemble method helps the behaviour prediction model achieve results more accurately than comparison methods for both students' educational datasets. This proves that the proposed students' behaviour analysis approach achieves its research objective of improving the overall performance of the student's behavioural data analysis.

This study suggests the usage of this proposed ensemble FS-based approach for better measurement and prediction of students' behavioural analysis performance as it improves the prediction performance of different behaviour-related educational data. Students' behavioural analysis outcomes of this study do not utilize personalized learning. So, the study is extended to incorporate a customized study material recommendation model based on the prediction outcome of this student's analysis.

References

[1] Batool, S., Rashid, J., Nisar, M.W. et al. Educational data mining to predict students' academic performance: A survey study. Educ Inf Technol 28, 905–971, 2023. https://doi.org/10.1007/s10639-022-11152-y

[2] Dogan, Murat Ertan, Tulay Goru Dogan, and Aras Bozkurt. 2023. "The Use of Artificial Intelligence (AI) in Online Learning and Distance Education Processes: A Systematic Review of Empirical Studies" Applied Sciences 13, no. 5: 3056. https://doi.org/10.3390/app13053056

[3] A. I. Al-Alawi, M. A. A. Alfateh and A. M. Alrayes, "Educational Data Mining Utilization to Support the Admission Process in Higher Education Institutions: A Systematic Literature Review," 2023 International Conference on Cyber Management and Engineering (CyMaEn), Bangkok, Thailand, 2023, pp. 332-339, doi: 10.1109/CyMaEn57228.2023.10051077.

[4] K. Okoye, S. D. N. Daruich, J. F. E. De La O, R. Castaño, J. Escamilla and S. Hosseini, "A Text Mining and Statistical Approach for Assessment of Pedagogical Impact of students' Evaluation of Teaching and Learning Outcome in Education," in IEEE Access, vol. 11, pp. 9577-9596, 2023, doi: 10.1109/ACCESS.2023.3239779.

[5] T. -C. Truong and Q. B. Diep, "Technological Spotlights of Digital Transformation in Tertiary Education," in IEEE Access, vol. 11, pp. 40954-40966, 2023, doi: 10.1109/ACCESS.2023.3270340.

[6] Shilpa, K., Adilakshmi, T. (2023). Analysis of SWCET students' Results Using Educational Data Mining Techniques. In: Kumar, A., Ghinea, G., Merugu, S., Hashimoto, T. (eds) Proceedings of the International Conference on Cognitive and Intelligent Computing. Cognitive Science and Technology. Springer, Singapore. https://doi.org/10.1007/978-981-19-2358-6_58

[7] P. K. Bansal and M. Ahmed, "An Expert System for Analyzing the Behaviour of students' in the Higher Education," 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2023, pp. 1569-1574.

[8] E. Alhazmi and A. Sheneamer, "Early Predicting of Students Performance in Higher Education," in IEEE Access, vol. 11, pp. 27579-27589, 2023, doi: 10.1109/ACCESS.2023.3250702.

[9] A. Alsulami, A. S. A. -M. Al-Ghamdi and M. Ragab, "Using Data Mining Techniques to Enhance the Student Performance. A semantic review.," 2023 1st International Conference on Advanced Innovations in Smart Cities (ICAISC), Jeddah, Saudi Arabia, 2023, pp. 1-5, doi: 10.1109/ICAISC56366.2023.10084963.

[10] Ahuja, R., Jha, A., Maurya, R., Srivastava, R. (2019). Analysis of Educational Data Mining. In: Yadav, N., Yadav, A., Bansal, J., Deep, K., Kim, J. (eds) Harmony Search and Nature Inspired Optimization Algorithms. Advances in Intelligent Systems and Computing, vol 741. Springer, Singapore. https://doi.org/10.1007/978-981-13-0761-4_85

[11] Harikumar Pallathadka, Alex Wenda, Edwin Ramirez-Asís, Maximiliano Asís-López, Judith Flores-Albornoz, Khongdet Phasinam, Classification and prediction of student performance data using various machine learning algorithms, Materials Today: Proceedings, Volume 80, Part 3, 2023, Pages 3782-3785, ISSN 2214-7853, https://doi.org/10.1016/j.matpr.2021.07.382.

[12] Y. Pei and G. Lu, "Design of an Intelligent Educational Evaluation System Using Deep Learning," in IEEE Access, vol. 11, pp. 29790-29799, 2023, doi: 10.1109/ACCESS.2023.3260979.

[13] Tan, C., Lin, J. A new QoE-based prediction model for evaluating virtual education systems with COVID-19 side effects using data mining. Soft Comput 27, 1699–1713 (2023). https://doi.org/10.1007/s00500-021-05932-w

[14] Erick Odhiambo Omuya, George Onyango Okeyo, Michael Waema Kimwele, Feature Selection for Classification using Principal Component Analysis and Information Gain, Expert Systems with Applications, Volume 174, 2021, 114765, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2021.114765.

[15] A. Jenul, S. Schrunner, K. H. Liland, U. G. Indahl, C. M. Futsæther and O. Tomic, "RENT—Repeated Elastic Net Technique for Feature Selection," in IEEE Access, vol. 9, pp. 152333-152346, 2021, Doi: 10.1109/ACCESS.2021.3126429.

[16] Wen Xiao, Ping Ji, Juan Hu, "RnkHEU: A Hybrid Feature Selection Method for Predicting students' Performance", Scientific Programming, vol. 2021, Article ID 1670593, 16 pages, 2021. https://doi.org/10.1155/2021/1670593

[17] Ali Al-Zawqari, Dries Peumans, Gerd Vandersteen, A flexible feature selection approach for predicting students' academic performance in online courses, Computers and Education: Artificial Intelligence, Volume 3, 2022, 100103, ISSN 2666-920X, https://doi.org/10.1016/j.caeai.2022.100103.

[18] Singh, R., Pal, S. (2021). A Comprehensive Study of Feature Selection Techniques for Evaluation of Student Performance. In: Sheth, A., Sinhal, A., Shrivastava, A., Pandey, A.K. (eds) Intelligent Systems. Algorithms for Intelligent Systems. Springer, Singapore. https://doi.org/10.1007/978-981-16-2248-9_30

[19] R. Singh, S. Pal, Machine learning algorithms and ensemble technique to improve prediction of students' performance, International Journal of Advanced Trends in Computer Science and Engineering, 2020, 9 (3), pp. 3970-3976.

[20] Hussain, S., Khan, M.Q. Student-Performulator: Predicting students' Academic Performance at Secondary and Intermediate Level Using Machine Learning. Ann. Data. Sci. (2021). https://doi.org/10.1007/s40745-021-00341-0

[21] Tarik, A., Aissa, H., & Yousef, F. (2021). Artificial Intelligence and Machine Learning to Predict Student Performance during the COVID-19. Procedia Computer Science, 184, 835–840. doi:10.1016/j.procs.2021.03.104

[22] Abellan-Abenza, J., Garcia-Garcia, A., Oprea, S., Ivorra-Piqueres, D., & Garcia-Rodriguez, J. (2017). Classifying Behaviours in Videos with Recurrent Neural Networks. International Journal of Computer Vision and Image Processing (IJCVIP), 7(4), 1-15. doi:10.4018/IJCVIP.2017100101

[23] Rastrollo-Guerrero JL, Gómez-Pulido JA, Durán-Domínguez A. Analyzing and Predicting students' Performance by Means of Machine Learning: A Review. Applied Sciences. 2020; 10(3):1042. https://doi.org/10.3390/app10031042

[24] Chowanda, A., Sutoyo, R., Meiliana, & Tanachutiwat, S. (2021). Exploring Text-based Emotions Recognition Machine Learning Techniques on Social Media Conversation. Procedia Computer Science, 179, 821–828. doi:10.1016/j.procs.2021.01.099

[25] J. Zhao, X. Yang, Q. Qiao and L. Chen, "Sentiment Analysis of Course Evaluation Data Based on SVM Model," 2020 IEEE International Conference on Progress in Informatics and Computing (PIC), 2020, pp. 375-379, doi: 10.1109/PIC50277.2020.9350812.

[26] D. Selvapandian, T. Meshach W., K. S. S. Babu, R. Dhanapal and J. Immanuel D., "An Efficient Sentiment Analysis on Feedback Assessment from Student to Provide Better Education," 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2020, pp. 1293-1300, doi: 10.1109/I-SMAC49090.2020.9243594.

[27] B. H. K. T., H. M. Perera, S. Kumarage, P. A. P. Savindri, D. Kasthurirathna and A. Kugathasan, "E-Learn Detector: Smart Behaviour Monitoring System to Analyze Student Behaviours During Online Educational Activities," 2021 3rd International Conference on Advancements in Computing (ICAC), 2021, pp. 19-24, doi: 10.1109/ICAC54203.2021.9671073.

[28] Saba, T., Rehman, A., Jamail, N. S. M., Marie-Sainte, S. L., Raza, M., & Sharif, M. (2021). Categorizing the students' Activities for Automated Exam Proctoring Using Proposed Deep L2-GraftNet CNN Network and ASO Based Feature Selection Approach. IEEE Access, 9, 47639–47656. doi:10.1109/access.2021.3068223

[29] https://www.kaggle.com/aljarah/xAPI-Edu-Data

[30] https://archive.ics.uci.edu/ml/datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+Data+Set