Journal of Theoretical and Applied Information Technology
15th April 2025. Vol.103. No.7

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

STUDENT ACADEMIC PERFORMANCE PREDICTION USING PROBIT-GENERATIVE RESTRICTED DEEP BOLTZMANN MACHINE

C. VIVEK[1] , P.TAMILSELVAN[2]

[1]Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore-641 021

[2]Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore-641 021

vivekcmca@gmail.com[1], tamilselvancs@kahedu.edu.in[2]

ABSTRACT

Big data mining also known as Knowledge Discovery in Databases (KDD) is a process that uses statistical and computational techniques to identify significant hidden patterns and insights in massive datasets. The prediction of students' academic achievement is a key component of modern educational systems that aim to raise educational standards. Even though predicting students' performance has been the focus of numerous studies, there are still problems with early dropout prediction, learning pattern recognition, and student success prediction, which cause delays and reduced efficiency. To overcome these issues, Probit Generative Restricted Deep Boltzmann Machine (PGRDBM) method utilizes Deep Boltzmann Machine to include four different layers such as one input layer, two hidden layers, and one output layer. At first, the data collection followed by efficient performance of two significant stages such as feature selection and data classification to enable more accurate and time, space-efficient student's academic performance prediction in educational institutions. At the input layer, the process of analysing educational performance begins with the collected student data. The Concept Drift-based Feature Selection (CD-FS) model is designed at the hidden layer to choose most pertinent features that are stable and relevant even in drift for precise prediction. The Probit Regression-based Data Classification (PR-DC) model is applied for robust student academic performance prediction by student data and classifying student performance into different classes. Then, hidden layer 2 sends the prediction results to the output layer. It is feasible to predict academic performance effectively. The performance analysis is effectively proposed in student academic performance prediction and two existing methods are conducted on different specific metrics such as accuracy, time, error rate, and space complexity

Keywords: *Big data Mining, Educational Performance Analysis, Prediction Achievement, Concept Drift, Feature Selection, Probit Regression and Data Classification.*

1. INTRODUCTION

A method called big data mining assists in finding patterns, trends, and connections in enormous datasets which helps educational institutions make well-informed decision making. The goal of educational institution performance analytics is to better understand and enhance several aspects of an educational institution's operations by gathering, evaluating, and interpreting data. This data-driven strategy aids organizations in making well-informed decisions, improving student achievement, and allocating resources as efficiently as possible. The evaluation of student performance involves assessing grades, attendance, engagement, and graduation rates to identify areas requiring improvement and provide specialized assistance to challenging students. There are many machine learning algorithms that are intended to forecast students' academic achievement.

In [1], the Attention-Based Artificial Neural Network (Attn-ANN) was developed for performing the student performance prediction while taking into account the learning activities. However, it failed to perform early identification of at-risk students. In [2], the Multi-Model Ensemble Approach was developed for performing the prediction of students in higher education by discovering students at-risk, differences in assessment across various environments, and estimating the relationship between teacher employment status and student achievement. However, the time for performing prediction process was not minimized effectively. In [3], the

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

study implemented the DeepFM-based prediction model for student dropout by performing significant processes such as preparation, feature engineering, model construction, training, assessment, and deployment. The study in [4] developed the system for determining the quality of education in relation to sustainability by analysing the abundance of data generated by the system for effective planning and future development. However, it was ineffective as a way to increase forecast accuracy. In [5], the study employed 'black box' machine learning models expanded with Pennsylvania's socioeconomic and educational data for the purpose of forecasting academic achievement in the state by removing the influence of logical relationships on such accuracy. In [6], the study focused on clustering and classification methods to determine how early student performance affects GPA. However, the prediction performance was not improved because there was insufficient combination of academic and non-academic features for efficient feature selection.

The study in [7] developed the supervised machine learning algorithms with the aim of investigating the variables that had a detrimental effect on the academic achievement of college students on probation, or underperforming students. For the purpose of enhancing student learning effects in a collaborative learning context, the study in [8] developed the methodology that integrates AI performance prediction model with learning analytics approaches. But, the student academic accomplishment analysis was not conducted in an efficient manner. In [9], the study focused on influence of academic self-efficacy on university students' academic performance by examining the relationships among academic self-efficacy, academic engagement, and academic performance utilizing Pearson correlation coefficients. In [10], the study focused on implementing machine learning techniques like random forest and feature to kenizer transformer for performing academic attrition prediction. However, there was no effective enhancement in the prediction accuracy due to the lack in the feature selection performance. In [11], the study focused on student performance prediction by the implementation of Adaptive Feature Selection Algorithm (AFSA) is a novel method that combines an ensemble approach for initial feature ranking with normalized mean ranking. In [12], the discrete feature oversampling technique called GLoW SMOTE-D was implemented with the goal of increasing the prediction model performance of student's failure in courses. Nevertheless, the time spent on prediction was not efficiently reduced. In order to forecast and analyze student performance, the study created Ensemble Models and Shapley Additive Explanations in [13]. In [14], the study focused on academic performance of deaf scholars by developing a novel ML-based system with eXplainable AI (XAI) utilizing Local Interpretable Model-Agnostic Explainer (LIME) and Shapley Additive Explainer (SHAP).With this system, two goals such as identifying DHH students who are at risk and elucidating risk variables, were significantly accomplished. However, the time and space complexity remained unaddressed. The research in [15] investigated the effectiveness of Long-Short Term Memory Networks (LSTM) in Educational Data Mining (EDM) with the objective of predicting student academic performance during the seventh, eighth, twelfth, and sixteenth weeks of the semester. Nevertheless, the feature selection process was not successfully implemented to increase the data prediction's accuracy.

1.1Research Contributions of the Paper

The research contributions of the proposed technique are stated below.

- The proposed PGRDBM method is developed for achieving student academic prediction performance in educational institutions.
- Proposed PGRDBM method is designed with feature selection, and data classification using the Deep Boltzmann Machine.
- Deep Boltzmann Machine includes one input layer, two hidden layers, and one output layer.
- To minimize the time for precise classification, Concept Drift-based Feature Selection (CD-FS) model is employed in proposed PGRDBM method with data segmentation, feature probability distribution estimation, drift detection, and feature selection. Most significant features are identified and selected via the novelty of the Chi-Squared Test.
- To obtain accurate prediction results, Probit Regressive Data Classification is developed for increasing the accuracy.
- Novelty of Jaccard Index Estimation is employed to analyze testing and training student data with the selected relevant features.
- Student data into numerous classes are determined by using the innovation of Probit Regression during Maximum Likelihood Estimates.
- To decrease the error, gradient descent is utilized.

- The performance of proposed PGRDBM method is thoroughly experimentally contrasted to the two existing methodologies in terms of taking into account four crucial metrics to show how well the proposed research effort predicts students' academic success.

1.2 Outline of the Paper

This paper's outline is given below. The literature review takes up the following part as section 2, while section 3 presents a succinct explanation of the proposed PGRDBM method. The results and analysis of the experiment are presented in Sections 4 and 5. Section 6 describes the discussion. Section 7 concludes with some final thoughts.

2. RELATED WORKS

Summary of literature review: The study focuses on computer-supported predictive analytics in related works. ML and DL approaches are used for predicting and explaining student performance in education.

The data mining techniques have garnered a lot of attention in the subject of education in recent years. These techniques use various classification algorithms to extract new trends and patterns from massive datasets. The proposed model was implemented in [16] for analysing the historical academic grades among university students to uncover relationships between courses, students, and courses, as well as between different students. The lag in grade prediction was avoided via attention method. But, the accuracy was not enhanced. In order for teachers to offer each student personalized guidance, the student performance prediction approach was implemented in [17]. This was accomplished by developing the hybrid neural network model based on a relationship matrix (RMHNN) and the relationship matrix-based bipartite network approach (RMBN) with Louvain clustering. However, it was not able to minimize the space complexity. In [18], the study implemented the multi-graph spatial-temporal synchronous network (MGSTSN) with the goal of enhancing the accuracy of student performance forecasts. Gated graph convolution is employed for capturing and examining spatial-temporal dependencies. The prediction of student performance in this network was not executed efficiently with a low error rate. In [19], the study utilizes trajectory-based computerized adaptive assessment for determining Learning Coefficients. In order to help students concentrate better on their studies and enhance their performance going forward, learning coefficients also give them quantifiable measurements. However, it was not able to increase the accuracy effectively. In [20], the Augmented Education (AugmentED) model was implemented for performing the academic performance prediction while taking into account the multisource, multi featured behavioural data. LSTM were employed for students' behavioral time series analysis. However, it failed to enhance the forecast performance.

In [21], the study focused on investigation of the effectiveness of deep learning in EDM, particularly in predicting academic performance and identifying students at risk of failure. Data pre-processing as well as different ML methods were employed to attain enhanced outcomes. However, there was no significant improvement in the accuracy. The purpose of this study in [22] was to evaluate a methodical approach to applying artificial neural networks to forecast academic achievement in higher education and examine the importance of different variables. However, the error rate was not minimized efficiently. During the COVID-19 epidemic, the study in [23] investigated an effective data mining method for evaluating satisfaction levels with online learning among higher education students. A pertinent subset of features was chosen by Feature Selection. However, the computational complexity was higher. In order to improve the performance of underperforming students, the study in [24] focused on creating a system that can forecast student performance and assist teachers in implementing remedial interventions on time. Collaborative learning was employed with higher performance of students. However, the space complexity remained unaddressed. In [25], the study analyses machine learning techniques for improving the predictive accuracy of the final student grades in first semester courses focusing on two modules such as six well-known machine learning techniques includes Decision Tree (J48), Support Vector Machine (SVM), Naïve Bayes (NB), K-Nearest Neighbor (kNN), Logistic Regression (LR) and Random Forest (RF) and multiclass prediction model. But it was unable to significantly improve the academic achievement of the students. In [26], the study focused on developing the prediction model that uses recently extracted statistical features to forecast students' performance based on their online activity. Statistical features were classified into three broad categories namely activity type, timing statistics, and peripheral activity count. These features were minimized via through entropy-based selection method. However, the performance of the classification was inefficient. In [27], the study

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

focused on Interpretable Prediction Method to proactively identify students experiencing academic difficulties before their situation escalates into a full-blown crisis (e.g., academic probation, withdrawal, and dropout). Data fusion, filtering, missing value, and transformation were designed in educational accomplishment forecast. However, the accuracy was not enhanced to an efficient level. In [28], the study was focused to explore the relationship between medical students' motivation, self-efficacy, learning engagement, and academic performance while considering demographic and sociocultural factors. But, the amount of time needed to research the academic performance and motivation of medical students was not adequately reduced.

The study in [29] implemented the hybrid regression model and multi-label classification model with the aim of forecasting the student performance and their key influential factors for early intervention and support. Capacity to create competitive performance was increased by hybrid approach. However, the relevant features were not selected with lesser time. In [30], an innovative model based on machine learning algorithms such as random forests, nearest neighbour, support vector machines, logistic regression, Naive Bayes, and k-nearest neighbour algorithms were implemented. Through this model, the early identification of at-risk students was performed by predicting the final exam grades of undergraduate students regarding midterm exam grades as the source data. However, this model was ineffective in examining students' learning behaviours, addressing their issues, optimizing the educational environment for effective data-driven decision making. Comparison of existing works described in table 1.

Table 1 Comparison of Existing Works

Method & Ref No	**Contribution**	**Merits**	**Demerits**
Attn-ANN [1]	Attn-ANN was developed for student performance prediction	Accuracy was enhanced	Early identification of at-risk students were not carried out
Multi-Model Ensemble Approach [2]	Multi-Model Ensemble Approach was employed in higher education	Precision was improved	Time was not minimized
DeepFM-based prediction model [3]	DeepFM-based prediction model developed for student dropout	Time was lesser	Features were not considered
ML approaches [4]	Quality of education was discovered with ML	Space complexity was reduced	Failed to increase forecast accuracy
'black box' machine learning models [5]	'black box' machine learning models were used to academic accomplishment	Accuracy was improved	Classification performance was not sufficient
Clustering and classification methods [6]	Early student performance was attained by clustering and classification methods	Time was reduced	Prediction performance was not increased
Supervised machine learning algorithms [7]	The academic achievement of college students on probation was determined	F1-score was enhanced	Time was higher
AI performance prediction model [8]	Learning analytics approaches were used to predict student performance	Precision was increased	Student academic accomplishment analysis was not conducted
Influence of academic self-efficacy on university students' academic performance [9]	Academic self-efficacy and academic achievement were compared	Academic performance was improved	Feature selection was not performed
Machine learning techniques [10]	The academic attrition forecast was carried out by ML	Time was minimized	Prediction accuracy was not increased

The above table literature study states that in order to improve the quality of education, it is important to predict students' academic performance with more precision. The accuracy of student performance forecasts were improved by goal of synchronous network with low error rate and failed to execute efficiently. The space complexity was not addressed in performance of student academic. The performance of the classification was failed to efficient the online activity. Failed to reduce the amount of time required by explore the academic performance and medical students. In order to achieve this, this study makes use of efficient machine learning classification methods that were the most successful in forecasting students' academic achievement.

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

3. PROPOSED METHODOLOGY

Big data mining is a technique used to extract valuable insights from large and complex datasets by identifying patterns, trends, and relationships that helps in making well-informed decisions and problem-solving. With the help of big data mining, educational institutions may now better understand student performance and enhance overall results. The performance analytics of educational institutions' primary areas of focus are student performance, faculty performance, curriculum effectiveness, and resource allocation.

The proposed PGRDBM method is made with two efficient processes such as feature selection and data classification that are used in the different layers of the Deep Boltzmann Machine to improve the prediction performance of student academic achievement in educational institutions. The figure 1 illustrates the architecture diagram of proposed PGRDBM method for student academic performance prediction.

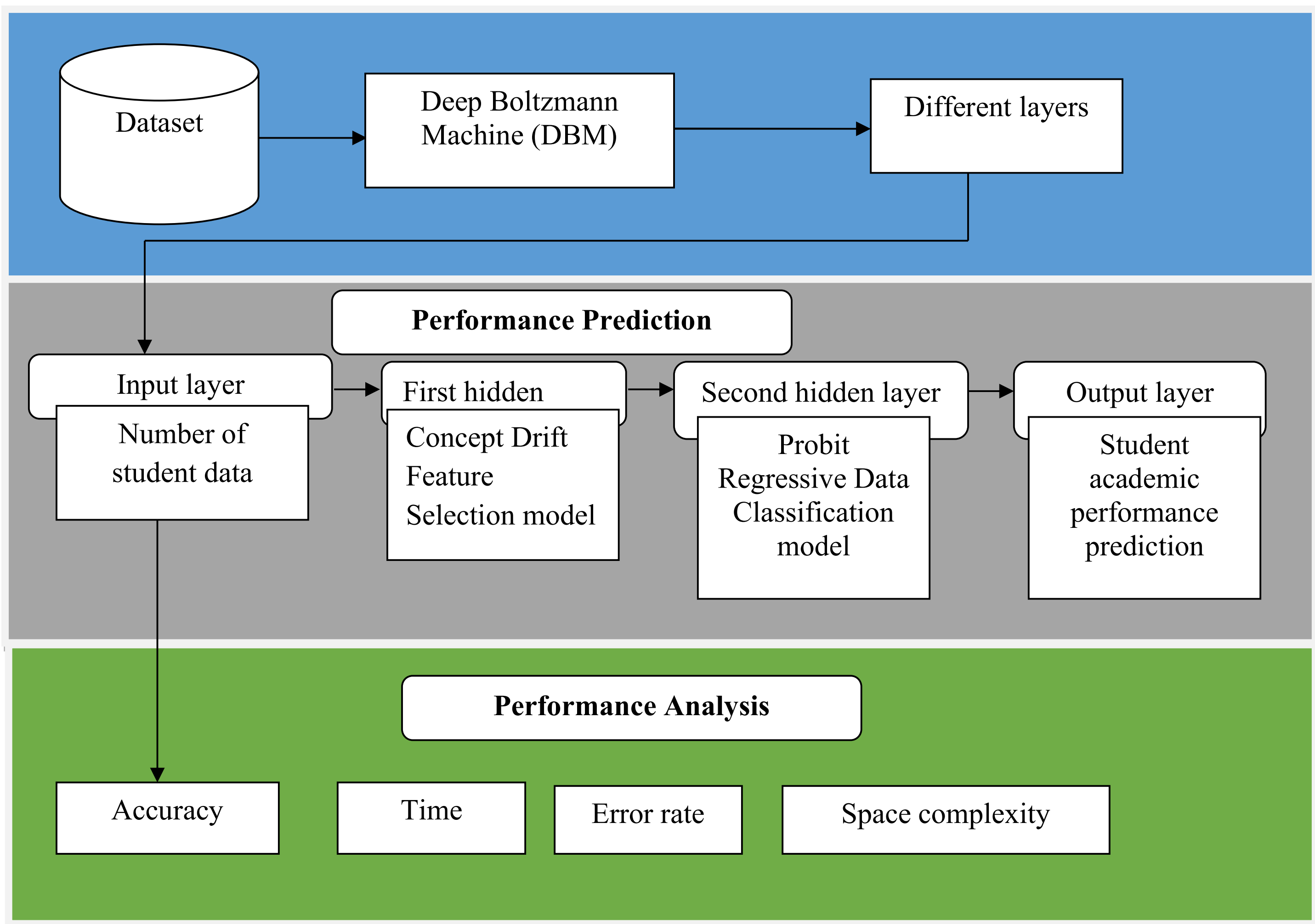

Figure 1 : Architecture diagram of proposed PGRDBM method for student academic performance prediction

As shown in Figure 1, the proposed PGRDBM method efficiently performs student educational performance prediction with greater accuracy, less error rate, time, and space complexity. The performance of the proposed PGRDBM method is carried out with the implementation of four layers such as one input layer, two hidden layers, and one output layer. In the beginning, the proposed PGRDBM method collects different number of student data from Students Dropout and Academic Success Dataset. As the first step, various student data is taken as input. The input data is forwarded to the first hidden layer. In the hidden layer 1, the Concept Drift Feature Selection model selects the most relevant features. Followed by, the Probit Regressive Data Classification model applies in the hidden layer 2 for classifying the student performance data into different classes according to the relevant features selected. The hidden layer 2 sends the classified results to the output layer. Then, the output layer obtains the prediction results for student educational performance analysis.

3.1 Data Collection

During experimental conduction, the proposed PGRDBM method utilizes Students Dropout and Academic Success Dataset for performing the student academic performance prediction. This dataset is taken from https://www.kaggle.com/datasets/mahwiz/students-dropout-and-academic-success-dataset. There are 4424 instances and 37 features or attributes in the dataset. The features are Marital status, Application mode, Application order, Course, Daytime/evening attendance, Previous qualification, Previous qualification (grade), Nationality, Mother's qualification, Father's qualification, Mother's occupation, Father's occupation, Admission grade, Displaced, Educational special needs, Gender, Scholarship holder, Age at enrollment, International, Curricular units 1st sem (credited), Inflation rate, GDP, and Target. The proposed research work collects student data from the provided dataset as input and carries out two important procedures including feature selection and data classification using the Deep Boltzmann Machine in order to predict students' academic success.

3.2 Deep Boltzmann Machine (DBM)

The Deep Boltzmann Machine (DBM) is a particular kind of generative artificial neural network that is notable for its unique design. The DBM learns a probabilistic representation of its input data which is made up of multiple layers of nodes i.e., units including one visible (input) layer, several hidden layers, and one output layer with connections between nodes pointing in every direction. This implies that nodes in any layer have the ability to influence nodes below and above them. Each of the nodes, which are neurons, represents a binary state, meaning that it can be in either of the two states commonly represented by the numbers 0 and 1. The weighted edges or a connection, which connects the nodes. The weights show how strongly the nodes are connected to one other. There are connections between nodes that are visible to visible, hidden to hidden, and visible to hidden. The Figure 2 illustrates the structure of the Deep Boltzmann Machine.

As shown in Figure 2, the proposed PGRDBM method makes use of DBM with two hidden layers to accurately forecast student academic achievement. This architecture allows the model to learn complex, hierarchical representations of student data, and capturing intricate relationships between various factors that influence academic success.

- **Visible layer:** The bottom layer takes the student's data with features including socioeconomic characteristics, attendance records, extracurricular activities, demographic data, and previous academic success input.
- **Hidden Layers:** These layers capture intricate patterns and relationships by extracting progressively features from the incoming data.
- **Output layer:** The output layer indicates the expected level of academic achievement, including GPA, graduation rate, and dropout risk.

Initialize the DBM with random weights and biases.

The process of proposed PGRDBM method is initialized with the assumption of Students Dropout and Academic Success Dataset as 'DS'. In the visible layer of DBM, the number of data '$sd = sd_1, sd_2, sd_3 \dots, sd_n$' with number of features '$\beta_1, \beta_2, \beta_3, \beta_4 \dots \dots, \beta_m$' from theinput dataset 'DS' are taken as input. The visible layer forwards the collected student input data to the first hidden layer of DBM. The weight matrix and additional bias "B" are multiplied by the "n" input that the neuron gets in the hidden layers. In order to illustrate the connection between the visible (input) and hidden units, the belief network uses a random weight during the training phase. The neuron activity 'Act_n' is mathematically expressed as given below.

$$Act_n = \left(\sum_{i=1}^{n} sd_{i_h} * w_{ij}\right) + b (1)$$

Here, 'w_{ij} ' denotes the weight between the i^{th} input layer and j^{th} hidden layerand the bias is symbolized as 'b'

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

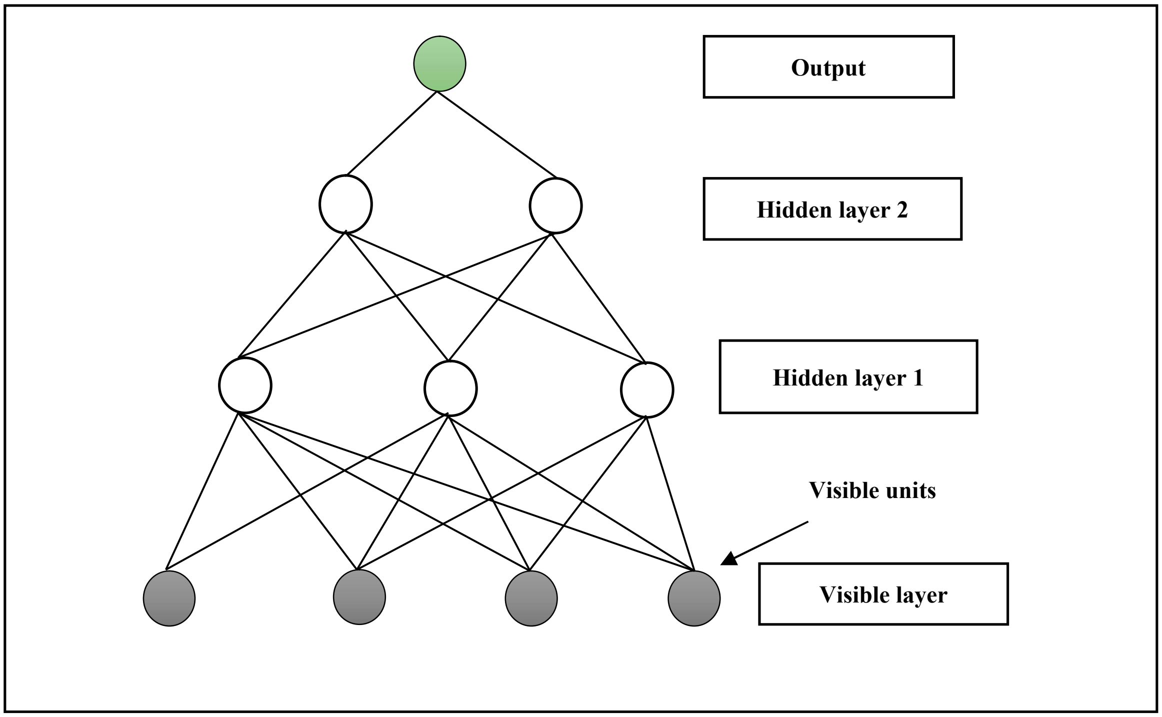

Figure 2 Structure of the Deep Boltzmann Machine

3.2.1 Concept drift-based feature selection model (At the first hidden layer):

When it comes to handling the input data, the features are essential. It is necessary to select the key features for effective data classification. The main objective of feature selection is to decrease the dimensionality of the dataset by finding the most informative feature. In the machine learning, the feature selection is a crucial process of identifying, selecting the most relevant features and removing the irrelevant features from the input. This leads to reduce the noise for minimizing the computational complexity that results in enhancing the accuracy for data classification. In the student performance prediction process, the feature selection is indeed a critical phase. For that reason, the proposed PGRDBM method employs feature selection phase as initial after getting the student data from provided dataset as input at the input layer. The proposed PGRDBM method utilizes Concept Drift-based Feature Selection (CD-FS) model in the hidden layer 1 with the aim of selecting the most significant features even when the underlying data distribution changes from historical student input data for a given prediction task.

The phenomenon known as concept drift describes how the distribution of underlying data evolves over time. The effective drift detection is crucial to trigger the feature selection process by identifying changes in the data distribution. In the context of student performance prediction, concept drift arises due to changes in curriculum, teaching methods, or student demographics. The feature selection helps in identifying the most relevant factors (e.g., attendance, study habits, socio-economic factors) that consistently influence student outcomes despite these changes.

In the CD-FS model, the statistical Chi-Squared Test based Kullback–Leibler (KL) Divergence method is applied for discovering and selecting stable and relevant features in the presence of concept drift. With the help of KL Divergence, the features are considered as more stable if their distribution varies less over time. By using Chi-Squared Test, the features are considered more relevant if they show a substantial correlation with the objective variable (e.g., student grades).The Figure 3 illustrates the overview of Concept Drift-based Feature Selection model that includes four essential steps such as data segmentation, feature probability distribution estimation, drift detection, and feature selection.

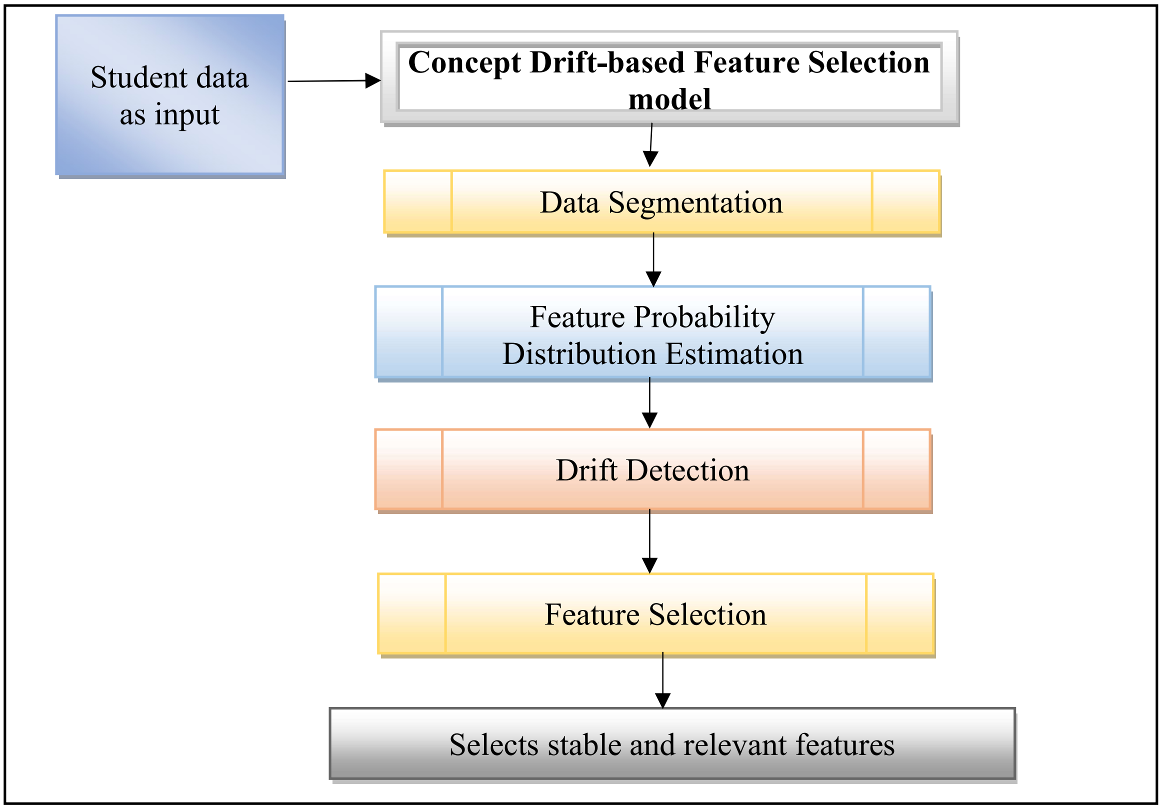

Figure 3 Overview of Concept Drift-based Feature Selection model

As illustrated in Figure 3, the CD-FS model obtains most relevant features for precise data classification by taking the student data from the given dataset as input. In the beginning, the student data is divided into time-ordered windows (e.g., daily, weekly, monthly).This segmentation is essential for identifying changes in the distribution of data across time. After that, the KL Divergence is applied for determining the probability distribution of features in consecutive time windows by estimating KL divergence between these distributions. According to this, the stable feature is identified. After that, the Chi-Squared Test is conducted for discovering the relevant feature. From that, the CD-FS model effectively identifies stable features and selects relevant ones with strong relationships with the target variable by combining the KL Divergence with the Chi-Squared Test.

At first, the historical student data including demographic information, academic performance (grades, attendance, etc.), and other relevant variables are taken as input. Secondly, the input student data is segmented into time-ordered windows for capturing the potential concept drift. For each time window, the KL Divergence is applied for discovering the probability distribution of each feature by estimating the difference between two probability distributions. Here, the significant KL divergence indicates potential drift. The mathematical expression for estimating the KL Divergence between the probability distributions of each feature in consecutive time windows is obtained as follows.

$$KL_D(p||q)_\beta = \sum p\,(\beta) \log\left[\frac{p\,(\beta)}{q(\beta)}\right] \quad (2)$$

As shown in above equation (2), the KL-divergence between the probability distributions of a feature 'β' is represented as '$KL_D(p||q)_\beta$'. Here, 'p' and ' q ' are the probability distributions of a feature in two consecutive time windows such as current time window and previous time window, and '$\sum$' is the summation over all possible values of the feature 'β'.

$$KL_D(p||q)_\beta = \begin{cases} 0, & No\ concept\ drift \\ > 0, & Greater\ degree\ of\ concept\ drift \end{cases} \quad (3)$$

There is a threshold for the KL Divergence. The potential drift in the distribution of the feature is indicated by a value above this threshold. If '$p\,(\beta)$' and '$q(\beta)$'are different, the KL Divergence is positive. The larger difference in the probability distributions is implied by a higher KL Divergence value which suggests a higher level of concept drift. If '$p\ (\beta)$' and '$q(\beta)$'are identical, the KL Divergence is zero. This suggests that the feature's probability distributions in the two successive time frames are identical.

$$Div(p\ (\beta)||q(\beta)) = Min\ Div(p\ (\beta), q(\beta)) \quad (4)$$

As shown in above equation (4), the features that have low KL Divergence values over time periods is selected which suggest that their distribution is largely stable. For each of the selected stable features, the contingency table relating the feature to the target variable is generated and the Chi-Squared Test is employed to determine if there is a statistically significant association between the feature and the target variable e.g., Dropout, Enrolled, and Graduate. The mathematical expression for applying Chi-Squared Test to select features that are relevant to the prediction task is given below.

$$\chi^2{}_\beta = \sum \frac{\left((OF_{i,j}) - (EF_{i,j})\right)^2}{(EF_{i,j}} \quad (5)$$

With the help of above equation (5), the Chi-Squared Test is conducted for each selected stable features '$\chi^2{}_\beta$'. '$OF_{i,j}$'is the observed frequency in cell (i, j) of the contingency table. '$EF_{i,j}$' IS THE expected frequency in cell (i, j) of the contingency table, and '$\sum$' is the summation over all cells in the contingency table. By the combination of KL Divergence with the Chi-Squared Test, the CD-FS model identifies stable and relevant features while effectively addressing concept drift in a historical student data. The algorithmic representation of the Concept Drift-based Feature Selection model is given below algorithm 1.

//Algorithm 1: Concept Drift-based Feature Selection

Input: Input dataset 'DS', Number of student data '$d = d_1, d_2, d_3 \dots, d_n$', Number of features '$\beta_1, \beta_2, \beta_3, \beta_4 \dots \dots, \beta_m$'

Output: Identifies and selects stable and relevant features

Begin

1. **Collect** the number of data '$d = d_1, d_2, d_3 \dots, d_n$' and features '$\beta_1, \beta_2, \beta_3, \beta_4 \dots \dots, \beta_m$' from the given dataset
2. **For** each feature
3. **Determine**KL Divergence between the probability distributions
4. **If**$KL_D(p||q)_\beta = 0$,
5. Feature is considered as stable
6. **Else** $KL_D(p||q)_\beta > 0$,
7. Feature is considered as non-stable
8. **Select** stable feature
9. **For** each selected stable feature
10. **Apply**Chi-Squared Test
11. **Select** feature that is relevant for prediction
12. **Obtain** stable and relevant features for precise data classification
13. **End if**
14. **End else**
15. **End For**
16. **End For**

End

Algorithm 1 Concept Drift-based Feature Selection model

As shown in above algorithm 1, the Concept Drift-based Feature Selection model combines the results of KL Divergence and Chi-Squared Test for discovering and selecting the stable and relevant features for achieving higher accuracy in the data classification.

3.2.2Probit regression-based based data classification (at the second hidden layer):

According to the selected stable and relevant features from given student input data in the hidden layer 1, the proposed PGRDBM method employs the Probit Regression-based Data Classification (PR-DC) model in the hidden layer 2. The PR-DC model aimed at predicting the student performance by categorizing the data into different classes while taking into account the selected features.

The probit regression is a statistical mode utilized for performing data classification. This means it predicts the probability of an observation belonging to one of two possible categories. In the PR-DC model analyses the student data (features like academic performance, demographics, etc.) and estimates the probability of the student belonging to each of the predefined classes (e.g., dropout, continuing enrolment, graduation) utilizing Maximum Likelihood Estimates. The student is then assigned to the class with the highest probability predicted by the probit regression. The figure 4 shows the flow process of Probit Regression-based Data Classification model.

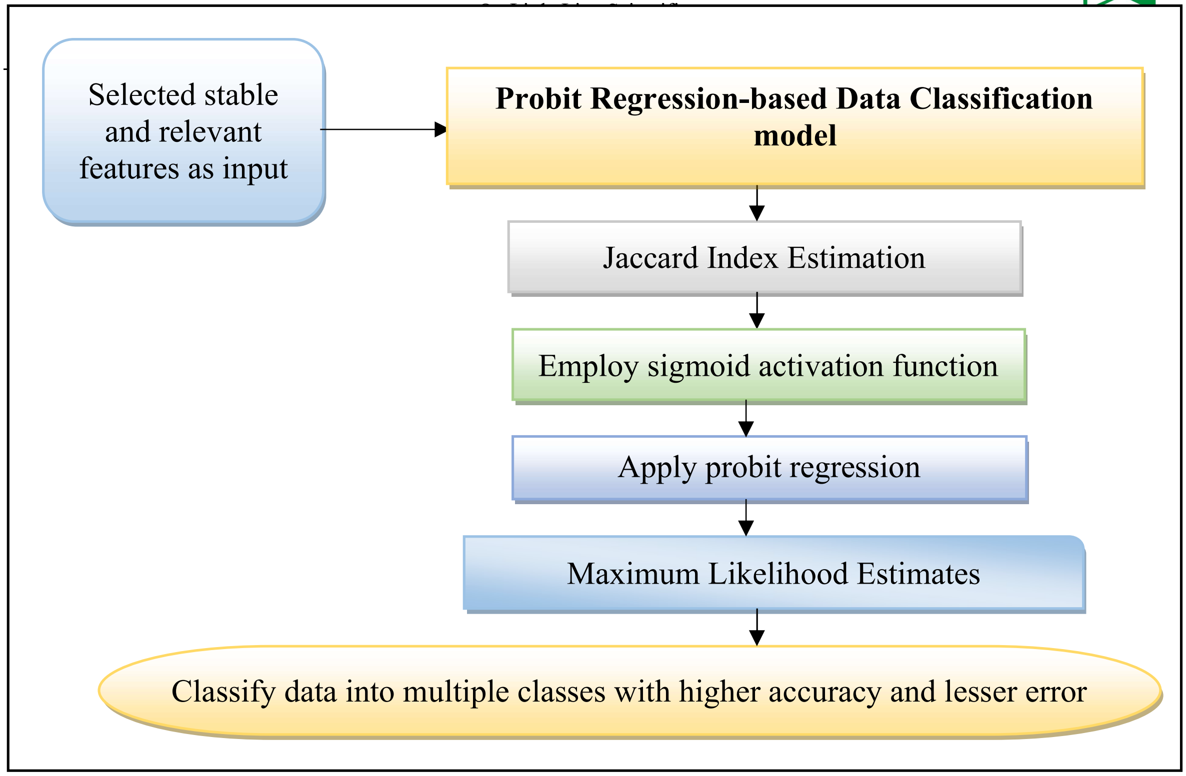

Figure 4 Flow process of Probit Regression-based Data Classification model

As shown in Figure 4, the Probit Regression-based Data Classification model to obtain the classification output 'Y' at the hidden layer 2 of DBM. In the proposed CDPGSRDBMC method, the probit regression predicts probability of belonging to each class by assigning the student to the class with the highest probability.

Here, the set of input vector is represented as'$\{sd_i\}_{i=1}^{n}$' and the target output is indicated as '$\{Y_i\}_{j=1}^{m}$'.At first, the PR-DC model takes the selected features number of selected features '$\beta_1, \beta_2, \beta_3, \beta_4 \ldots \ldots, \beta_b$ with their data '$sd = sd_1, sd_2, sd_3 \ldots, sd_n$'. Secondly, the Jaccard Index Estimation for examining the testing and training data with the selected relevant features is applied as given below.

$$\mu_{(sd_i,sd_j)} = \frac{n\beta_{ij}}{n\beta_i + n_{\ j} - n\beta_{ij}} \quad (6)$$

By using above equation (6), the Jaccard Similarity Coefficient is calculated as '$\mu_{(sd_i,sd_j)}$'.$n_i$represent number of features which appear in data$i$ whereas n_j denotes the number of features which appear in data j. Here, n_{ij} indicates the number of features which appear in both (i.e. input testing and training data in a data set) i and j. According to the consideration of similarity measure between the training and testing data samples, the input student data is classified into different classes. Following that, the sigmoid activation function is utilised for examining the coefficient results that is mathematically expressed as given below.

$$Y = \frac{1}{1+e^{-\mu}} \quad (7)$$

Here, 'Y' denotes the sigmoid activation function which obtains the results aimed at categorizing the data into multiple classes.

$$Y = \sum_{i=1}^{n} w_i \, JSE\ (sd_i, sd_j) + b \quad (8)$$

After this, the probit regression is applied for estimating the relevance of classified outcome and data output through the Maximum Likelihood Estimates.

$$P(Y \mid X) = \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k) \quad (9)$$

Here, 'Y'denotes outcomes, '$P(Y \mid X)$' is the probability of the outcomes (e.g., dropout) occurring given the predictor variables 'X', 'X'is the vector of independent variables (features),'β_0'is the intercept,'β_1' is the vector of coefficients for the independent variables and 'Φ'is the cumulative distribution function of the standard normal distribution.β_1 (coefficient for GPA), β_2 (coefficient for Attendance), β_3 (coefficient for HoursStudied). Here, the maximum likelihood estimation is used to find the values of β_0 and β_1 that maximize the likelihood of observing the data given the model.

$$L(\beta\ |Y,X) = \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k) \qquad (10)$$

Here, β is the vector of model parameters $(\beta_0, \beta_1, \ldots, \beta)$. Finally, the student is assigned to the class with the highest predicted probability.

//Algorithm 2: Probit Regression-based Data Classification

Input: Input dataset 'DS', Number of student data '$sd = sd_1, sd_2, sd_3 \ldots, sd_n$', Number of selected stable and relevant features '$\beta_1, \beta_2, \beta_3, \beta_4 \ldots \ldots, \beta_m$'

Output: Perform data classification

Begin

1. **Take** the number of data '$sd = sd_1, sd_2, sd_3 \ldots, sd_n$' and selected stable and relevant features '$\beta_1, \beta_2, \beta_3, \beta_4 \ldots \ldots, \beta_m$'
2. **For** each selected features 'β' with data 'sd'
3. **Apply**Jaccard Index Estimation between testing data 'sd_T' and training data 'sd_R'
4. **Determine** sigmoid activation function
5. **Apply** probit regression
6. Determine maximum likelihood estimation
7. **Provide** the classification results
8. **End For**
9. **End**

Algorithm 2 Probit Regression-based Data Classification model

As shown in algorithm 2, the Probit Regression-based Data Classification model successfully classifies the student data into multiple classes such as high-achieving, average, underperforming with higher accuracy and less error rate.

3.2.3 Student academic performance prediction (at the output layer):

The output layer receives the classification results from the hidden layer 2. After the data classification, the output layer utilizes gradient descent in order to reduce the error by updating the weights between the layers which is mathematically expressed as given below.

$$w'_{ij} = w_{ij} + \left[LR * \frac{\partial\, l \quad p\,(Y)}{\partial\, \alpha_t}\right] (11)$$

Here, the updated weight is denoted as 'w'_{ij}', the current weight is referred to as 'w_{ij}', the learning rate is denoted as 'LR', the probability of results at the output layer 'Y' is represented as $p\,(Y)$, and '$\frac{\partial \log p\,(Y)}{\partial\, \alpha_t}$' denotes the gradient descent function.It finds the local minimum of a function (i.e., error rate) by changing the current weight. This procedure is repeated until the least amount of error is found. The error rate is then calculated as the square of the discrepancy between the actual and anticipated outcomes. The following is the mathematical expression for this.

$$E = \frac{1}{2}\left(Y_{act} - Y_{pre}\right)^2 (12)$$

Here, 'E' represents an error rate, 'Y_{act}' denotes the actual results, and 'Y_{pre}'indicates the predicted classification results.Then, the output layer obtains final predicted class probabilities for each class for student academic performance. Finally, the proposed CDPGSRDBMC method effectively forecast the students' outcomes with an emphasis on dropout, continuing enrolment, and graduation by the end of the typical course period. Therefore, the proposed research work's objective of improving educational planning and student help activities is accomplished by approaching it as a multi-class classification problem.

4. EXPERIMENTAL SETUP

The experimental evaluation of the full proposed PGRDBM method is carried out in Python high level programming language. In order to predict the student academic success, the proposed technique takes into account the Students Dropout and Academic Success Dataset that was taken from https://www.kaggle.com/datasets/mahwiz/students-dropout-and-academic-success-dataset.To verify the effectiveness of the proposed research effort in predicting student academic success, the performance of the proposed technique is contrasted with the two distinct existing methodologies while taking into account the following important metrics.

- Prediction accuracy,
- Prediction time,
- Error rate,
- Space complexity.

5. RESULTS AND DISCUSSION

This section compares the trial outcomes of proposed PGRDBM method and two existing methods to show how successful the proposed research for student academic performance prediction. In order to evaluate proposed PGRDBM method, the essential metrics such as prediction accuracy, prediction time, and error rate and space complexity are used. The performance analysis of proposed method is evaluated based on following metrics with the help of table values and graph given below.

5.1Performance Analysis of Prediction Accuracy

The prediction accuracy is a metric utilized for evaluating the performance of a classification method. The prediction accuracy is defined as the proportion of correctly predicted data out of the total number of data taken for simulation. The mathematical expression for estimating the prediction accuracy is obtained as given below.

$$Acc_P = \sum_{i=1}^{n} \frac{D_{cs}}{D_i} * 100(13)$$

With the help of above equation (13), the prediction accuracy 'Acc_P' is determined in terms of percentage (%). Here, 'D_i' denotes the total number of data considered from dataset and 'D_{cs}' indicates the number of data that are correctly predicted. The technique performs better to guarantee better results for student academic performance prediction due to the high accuracy value.

Table 2 Prediction accuracy of the three different techniques

Number of student data	Prediction Accuracy (%)		
	Proposed PGRDBM method	Existing [1]	Existing [2]
400	98.5	93	94.5
800	97.5	92.375	93.5
1200	96.0833333	91.75	92.75
1600	95.3125	92.125	91.25
2000	94.9	93.45	93.35
2400	97.875	90.3333	91.2083
2800	96.4642857	91.0357	93.2143
3200	95.78125	90.5625	92.75
3600	97.1111111	92.0556	93.4444
4000	96.05	90.45	91.35

Table 2 provides the performance of the prediction accuracy based on the number of student data taken for three different methods while utilizing the Students Dropout and Academic Success Dataset. The simulation is carried out by making a comparative analysis between the PGRDBM method and two existing methods in terms of considering prediction accuracy. The graph in Figure 5 which is plotted with the values in Table 1, shows the measure of the prediction accuracy.

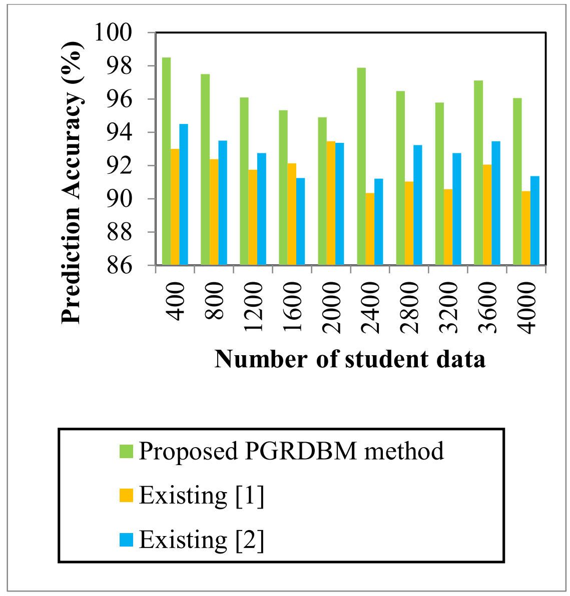

Figure 5 Graphical illustration of Prediction accuracy

Figure 5 depicts graphical representation of prediction accuracy with input student data as input utilizing three various methods for performing student educational performance analysis. The result shown in above graph demonstrates that the proposed PGRDBM method increases the accuracy for performing student educational performance prediction than the other two existing methods. By performing Concept Drift-based Feature Selection (CD-FS) model, the proposed PGRDBM method significantly identifies and selects the stable and relevant features for precise data classification. According to the selected features, the proposed PGRDBM method correctly classifies the data on dropout, continuing enrolment, and graduating students through the implementation of Probit Regression-based Data Classification (PR-DC) model in the hidden layer 2. From that, the output layer obtains the final predicted results for with greater accuracy. Hence, the prediction accuracy of the proposed PGRDBM method is enhanced up to 5% and 4% when compared to existing [1] and existing [2] respectively.

5.2 Investigation of Prediction Time

The prediction time is defined as the amount of time required for performing student educational performance prediction. The following is the mathematical formulation for determining the prediction time.

$$Time_P = D_i * Time\ (to\ predict\ single\ data)(14)$$

By using above equation (14), the prediction time '$Time_P$' is estimated in terms of milliseconds (ms). Here, the total number of data considered from dataset is indicated as 'D_i' and the amount of time consumed for predicting single student data is represented as'$Time\ (to\ predict\ single\ data)$'. The technique performs better to guarantee better results for student academic performance prediction due to the high accuracy value.

Table 3 Prediction time of the three different techniques

Number of student data	Prediction Time (ms)		
	Proposed PGRDBM method	Existing [1]	Existing [2]
400	10.12	22.12	23.12
800	13.352	25.32	26.344
1200	16.272	28.236	29.556
1600	20.32	32.336	34.224
2000	23.28	35.46	37.46
2400	26.136	38.112	40.344
2800	28.616	39.228	42.336
3200	31.136	43.456	45.664
3600	33.156	45.576	47.232
4000	36.8	49.64	52.36

The above Table 3 demonstrates the performance outcomes of the prediction time by the utilization of three various methods regarding the diverse quantity of student data samples considered as input. The experiment shows that the three methods require less amount of time for performing student educational performance analysis according to the number of data taken from the Students Dropout and Academic Success Dataset. However, the proposed PGRDBM method completes prediction process within less amount of time than the other two existing methods. Depends on the values in Table 2, the graphs are plotted as shown in Figure 6.

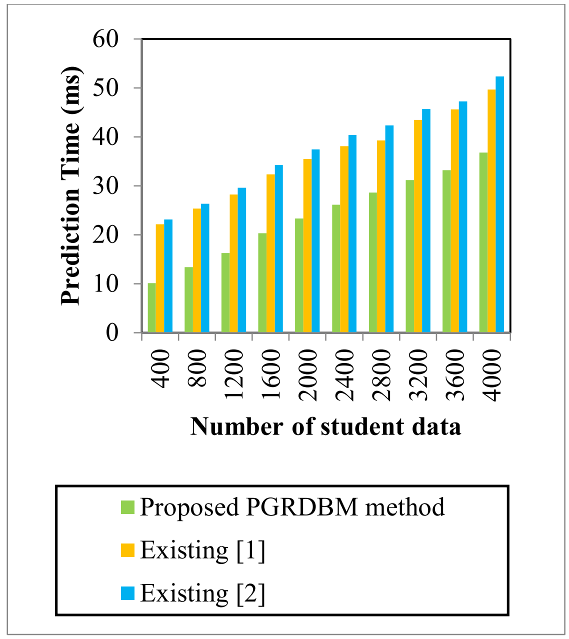

Figure 6 Graphical illustration of Prediction time

The comparison results of the prediction time using input student data ranges from 400 to 4000 for three methods which are shown in Figure 6. As shown in Figure 6, even if three methods significantly enhance the sensitivity, but the proposed PGRDBM method provides better results in the student academic performance prediction within less amount of time more than the other two existing methods. This is due to the implementation of two significant procedures such as feature selection and classification utilizing Concept Drift-based Feature Selection (CD-FS) model and Probit Regression-based Data Classification (PR-DC) model after getting the student data as input. Therefore, the prediction time of the proposed PGRDBM method is minimized by 36% and 39% when compared to existing [1] and existing [2] respectively.

5.3 Impact of Error Rate

The error rate is defined as the proportion of incorrectly classified data out of the total number of data taken for experimental setup. Below is the mathematical expression for determining the error rate.

$$ER = \sum_{i=1}^{n} \frac{D_{ics}}{D_i} * 100(15)$$

As shown in above equation (15), the error rate 'ER' is evaluated in terms of percentage (%). Here, 'D_i' represents the total number of data taken from dataset for simulation purpose and 'D_{ics}' indicates the number of data that are incorrectly classified.

The approach is considered to perform better for student educational performance analysis if the error rate is less.

Table 4 Error rate of the three different techniques

Number of student data	Error Rate (%)		
	Proposed PGRDBM method	Existing [1]	Existing [2]
400	18	22	25
800	21	27	31
1200	24	31	35
1600	27	35	39
2000	29	38	42
2400	31	41	47
2800	35	45	51
3200	37	47	55
3600	39	52	59
4000	42	56	63

Table 4 measures the error rate of three methods includes existing [1], existing [2], and proposed PGRDBM method when predicting educational performance using varying numbers of student data samples from the Students Dropout and Academic Success Dataset. The results of an error rate of the proposed CDPGSRDBMC method are compared with the two existing methods for proving the effectiveness of the research work in the educational performance prediction. From the Table 3, it is clearly evident that the proposed PGRDBM method obtains effective prediction outcomes for student's academic achievement analysis with reduced error rate than the other two existing methods. Plotting of the graph is depicted in Figure 7 and is dependent on the experimental values in Table 3.

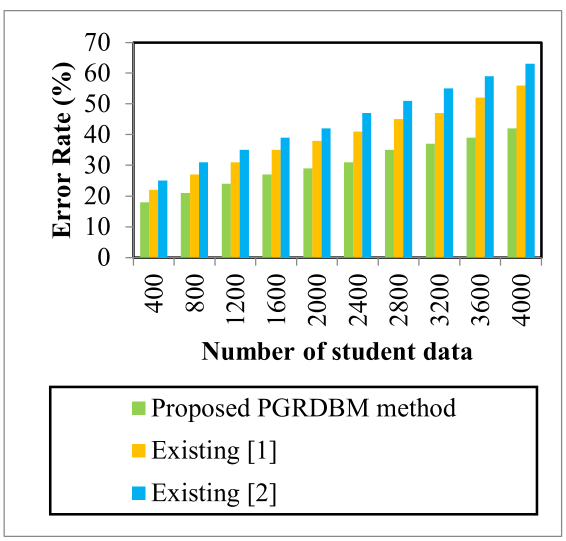

Figure 7 Graphical illustration of Error rate

Figure 7shows the measurement of the error rate based on the number of student data taken while utilizing three distinct approaches for predicting the student's academic performance. In Figure 7, the Students Dropout and Academic Success Dataset contain different quantities of student data taken in the horizontal axis whereas the error rate of the three various methods is observed in the vertical direction. According to the graphical plot, the proposed PGRDBM method efficiently performs student educational performance analysis with less error rate than the other two existing methods. The proposed PGRDBM method uses Concept Drift-based Feature Selection to identify stable features for precise data classification. It then implements Probit Regression-based Data Classification in the hidden layer for greater accuracy. Consequently, the error rate of the proposed PGRDBM method is minimized by 23% and 32% when compared to existing [1] and existing [2] respectively.

5.4 Case Scenario of Space Complexity

The space complexity is defined as the amount of memory space needed to carry out the student's academic performance prediction. The mathematical formula for determining the space complexity is given below.

$$S_{com} = \sum_{i=1}^{n} D_i * memory(single\ data) \quad (16)$$

In the above equation (16), the space complexity is represented as 'S_{com}' which is computed in megabytes (MB). If the space complexity is less, then the performance of the proposed technique for predicting the student's academic performance is said to be more efficient.

Table 5 Space complexity of the three different techniques

Number of student data	Space Complexity (MB)		
	Proposed PGRDBM method	Existing [1]	Existing [2]
400	18	27	29
800	25	30	33
1200	27	32.004	35.004
1600	29.008	37.008	39.008
2000	32	39	42
2400	34.992	42	45
2800	36.988	44.996	49
3200	40.992	48.992	52.992
3600	42.984	51.984	56.988
4000	45	55	60

Table 5 describes the comparative analysis of the space complexity while utilizing the three diverse methods such as existing [1], existing [2], and proposed PGRDBM method with number of student data samples taken as input. By comparing the performances of three methods taking into account the space complexity, the effectiveness in the prediction of student's academic performance for the proposed research work is verified. According to the values in Table 4, a graph between the various quantities of student data and space complexity is drawn as seen in Figure 8.

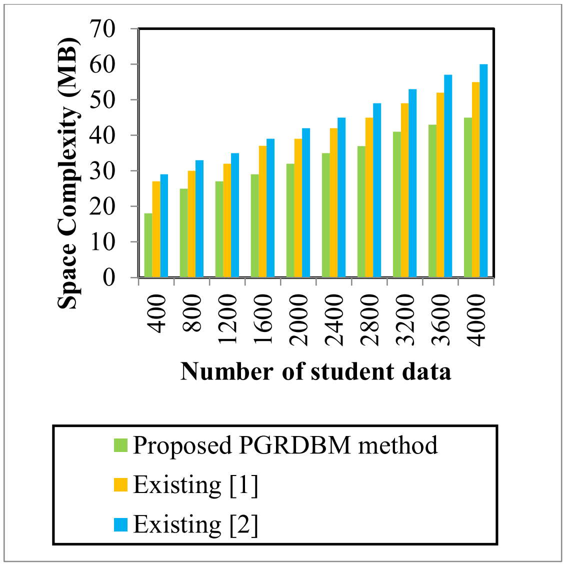

Figure 8 Graphical Illustration Of Space Complexity

Figure 8 shows the space complexity of three various method when predicting student's academic performance using varying numbers of student data samples extracted from the Students Dropout and Academic Success Dataset. Among the three different methods, the proposed PGRDBM method consumes less space to classify the input data samples into different classes for student's academic performance prediction. Due to the performance of the Concept Drift-based Feature Selection (CD-FS) model, the proposed PGRDBM method selects only stable and relevant features for data classification. This leads to minimize the amount of space consumed for performing student academic performance prediction. Hence, the space complexity of the proposed PGRDBM method is reduced by 19% and 25% when compared to existing [1] and existing [2] respectively.

6. DISCUSSION

In this section, existing [1] [1] was utilized with higher accuracy for student academic performance prediction. The limitation of existing methods was higher time, failure to consider early detection of vulnerable students, error was not reduced. To address this issue, the proposed PGRDBM method. This study compares the proposed PGRDBM method with the existing [1], [2] using Students Dropout and Academic Success Dataset based on various parameters. The objective of the proposed PGRDBM method is to enhance accuracy and minimize error and time. The experimental results revealed the superiority of this design; as shown in Table 3, 4 and 5 and Figure 6, 7 and 8, this improvement is achieved by using CD-FS.The reason for lesser time, error rate and space complexity accuracy is to apply CD-FS in DBM. Applicable features are chosen with minimum time and space complexity.

As shown in the tabulated results in Table 2 and Figure 5, the accuracy of the PGRDBM method is higher than the two conventional methods. The reason for improvement is achieved by applying PR-DC in DBM. With select significant features, student data is classified as multiple classes. In this way, the accuracy increased and error minimized by proposed PGRDBM method than conventional methods.

7. CONCLUSION

Student performance prediction was introduced to predict learning performance and recognize vulnerable students. Accurate student performance prediction is challenging issue. Several ML methods developed for student performance prediction. But, the accuracy was not improved and time was not decreased. For addressing this issue, educational data mining has developed into a powerful technique for forecasting students' academic success as well as enhancing the teaching and learning environment by identifying hidden relationships in educational data. The research objective of this study is to predict students' academic success with maximum accuracy and minimum time and space complexity by using feature selection and data classification in the DBM.

The research contribution of the proposed PGRDBM technique chooses only stable and pertinent features for data categorization because of the Concept Drift-based Feature Selection (CD-FS) model's performance. By doing this, the space needed to predict students' academic success is reduced, and precision for data classification is increased in a significant manner. With the selected features, the Probit Regression-based Data Classification (PR-DC) model effectively performs data classification for predicting students' academic performance.

The experiment analysis is conducted by comparing the full proposed PGRDBM method with the two existing methods in terms of various metrics like prediction accuracy, prediction time, error rate, and space complexity based on the different quantities of input student data taken from the provided dataset. The major findings are described as follows,

- The proposed PGRDBM method achieved higher prediction accuracy by 5% when compared to existing [1] and [2].
- Lesser prediction time by 38% than the state-of-art works [1] and [2].
- PGRDBM method also minimizes the error rate by 28% when compared to existing [1], [2] methods.
- Reduce space complexity by 22% than the state-of-art works [1] and [2].

To obtain the results, the full proposed PGRDBM method works better than the previous ones in terms of accurately predicting students' academic achievement. However, the different parameters are not focused namely precision, recall, specificity, and space complexity. Also, the data pre-processing was not focused on removing the noisy data. The future work research is extended to consider data preprocessing while eradicating the noisy data. Also, several metrics will be estimated for providing student data prediction. In addition, students into four broad categories will be classified namely very weak, weak, average, and good by using novel classification methods at an early stage of student data prediction.

REFERENCES

[1] Sukrit Leelaluk , Cheng Tang, Tsubasa Minematsu ,Yuta Taniguchi , Fumiya Okubo , Takayoshi Yamashita, and Atsushi Shimada, "Attention-Based Artificial Neural Network for Student Performance Prediction Based on Learning Activities", IEEE Access, Volume 12, 2024, Pages 100659-100675

[2] NaveedAnwer Butt, ZafarMahmood, KhawarShakeel, Sultan Alfarhood, MejdlSafran, and Imran Ashraf, "Performance Prediction of Students in Higher Education Using Multi-Model Ensemble Approach", IEEE Access, Volume 11, 2023, Pages 136091 – 136108

[3] Nuha Mohammed Alruwais, "Deep FM-Based Predictive Model for Student Dropout in Online Classes", IEEE Access, Volume 11, 2023, Pages 96954 - 96970

[4] Shah Hussain and Muhammad Qasim Khan, "Student-Performulator: Predicting Students' Academic Performance at Secondary and Intermediate Level Using Machine Learning", Springer, Volume 10, Issue 3, 2023, Pages 637-655

[5] Shan Chen and Yuanzhao Ding, "A Machine Learning Approach to Predicting Academic Performance in Pennsylvania's Schools", Social Sciences, MDPI, Volume 12, Issue 3, 2023, Pages 1-13

[6] EssaAlhazmi and Abdullah Sheneamer, "Early Predicting of Students Performance in Higher Education", IEEE Access, Volume 11, 2023, Pages 27579 – 27589.

[7] Lamees Al-Alawi, Jamil Al Shaqsi, Ali Tarhini, and Adil S. Al-Busaidi, "Using machine learning to predict factors affecting academic performance: the case of college students on academic probation", Springer, Volume 28, 2023, Pages 12407–12432

[8] Fan Ouyang, Mian Wu, LuyiZheng, Liyin Zhang and Pengcheng Jiao, "Integration of artificial intelligence performance prediction and learning analytics to improve student learning in online engineering course", Springer, Volume 20, 2023, Pages 1-23

[9] QianMeng and Qi Zhang, "The Influence of Academic Self-Efficacy on University Students' Academic Performance: The Mediating Effect of Academic Engagement", Sustainability, MDPI, Volume 15, Issue 7, 2023, Pages 1-14.

[10] Andrea Zanellati , Stefano PioZingaro , and Maurizio Gabbrielli, "Balancing Performance and Explainability in Academic Dropout Prediction", IEEE Transactions on Learning Technologies, Volume 17, Pages2140 – 2153

[11] Koushik Roy and Dewan Md. Farid, "An Adaptive Feature Selection Algorithm for Student Performance Prediction", IEEE Access, Volume 12, 2024, Pages 75577- 75598

[12] Susana Limanto, JokoLiantoBuliali, and Ahmad Saikhu, "GLoW SMOTE-D: Oversampling Technique to Improve Prediction Model Performance of Students Failure in Courses", IEEE Access, Volume 12, 2024, Pages 8889- 8901

[13] Hayat Sahlaoui, El ArbiAbdellaouiAlaoui, AnandNayyar , Said Agoujil, and Mustafa Musa Jaber, "Predicting and Interpreting Student Performance Using Ensemble Models and Shapley Additive Explanations", IEEE Access, Volume 9, 2021, Pages 152688 - 152703

[14] N. R. Raji, R. Mathusoothana S. Kumar, and C. L. Biji, "Explainable Machine LearningPredictionforthe Academic Performan

ce of DeafScholars", IEEE Access, Volume 12, 2024, Pages23595 – 23612

[15] Luis Vives, Ivan Cabezas, Juan Carlos Vives, Nilton German Reyes, Janet Aquino, Jose Bautista Condor, and S. Francisco Segura Altamiranom"Prediction of Students' Academic Performance in the Programming Fundamentals Course Using Long Short-Term Memory Neural Networks", IEEE Access, Volume 12, 2024, Pages 5882 – 5898

[16] Daozong Sun, RongxinLuo, Qi Guo, JiaxingXie, Hongshan Liu, ShileiLyu, XiuyunXue, Zhen Li, and Shuran Song, "A University Student Performance Prediction Model and Experiment Based on Multi-Feature Fusion and Attention Mechanism", IEEE Access, Volume 11, 2023, Pages 112307-112319

[17] Ziling Chen, Gang Cen, Ying Wei, and Zifei Li, "Student Performance Prediction Approach Based on Educational Data Mining", IEEE Access, Volume11, Pages 131260-131272

[18] Yifen Zhou and Xian Yu, "Multi-Graph Spatial-Temporal Synchronous Network for Student Performance Prediction", IEEE Access, Volume 12, 2024, Pages 142306 – 142319

[19] PallaviAsthana, Sumita Mishra, Nishu Gupta, Mohammad Derawi, and Anil Kumar, "Prediction of Student's Performance with Learning Coefficients Using Regression Based Machine Learning Models", IEEE Access, Volume 11, Pages 72732 - 72742

[20] Liang Zhao, Kun Chen, Jie Song,XiaoliangZhu,Jianwen Sun, Brian Caulfield, and Brian Mac Namee"Academic Performance Prediction Based on Multisource, MultifeatureBehavioral Data", IEEE Access, Volume 9, 2020, Pages 5453 - 5465

[21] Aya Nabil, Mohammed Seyam, and Ahmed Abou-Elfetouh, "Prediction of Students' Academic Performance Based on Courses' Grades Using Deep Neural Networks", IEEE Access, Volume 9, 2021, Pages 140731 - 140746

[22] Carlos Felipe Rodriguez-Hernandez , Mariel Musso, Eva Kyndt, and Eduardo Cascallar, "Artificial neural networks in academic performance prediction: Systematic implementation and predictor evaluation", Computers and Education: Artificial Intelligence, Elsevier, Volume 2, 2021, Volume 2, Pages 1-14.

[23] Hanan E. Abdelkader, Ahmed G. Gad, Amr A. Abohany, and Shaymaa E. Sorour, "An Efficient Data Mining Technique for Assessing Satisfaction Level with Online Learning for Higher Education Students During the COVID-19", IEEE Access, Volume 10, 2022, Pages 6286 – 6303

[24] Adnan Rafique, Muhammad Salman Khan, Muhammad Hasan Jamal, MamoonaTasadduq, FurqanRustam, Ernesto Lee, Patrick Bernard Washington, and Imran Ashraf, "Integrating Learning Analytics and Collaborative Learning for ImprovingStudent'sAcademicPerformance",IEEEAccess, Volume 9, 2021, Pages **167812** – 167826

[25] SitiDianah Abdul Bujang, Ali Selamat, Roliana Ibrahim, OndrejKrejcar, Enrique Herrera-Viedma, Hamido Fujita, and Nor Azura Md. Ghani, "Multiclass Prediction Model forStudentGradePredictionUsing Machine Learning",IEEE Access, Volume 9, 2021, Pages 95608 - 95621

[26] Ghassen Ben Brahim, "Predicting Student Performance from Online Engagement Activities Using Novel Statistical Features", Arabian Journal for Science and Engineering, Springer, Volume 47, 2022, Pages 10225-10243.

[27] Zhai Mingyu, Wang Sutong, Wang Yanzhang, and Wang Dujuan, "An interpretable prediction method for university student academic crisis warning",Complex& Intelligent Systems, Springer, Volume 8, 2022, Pages 323-336.

[28] Hongbin Wu, Shan Li, Juan Zheng, and JianruGuo, "Medical students' motivation and academic performance: the mediating roles of self-efficacy and learning engagement", Taylor & Francis, Medial Education Online, Volume 25, Issue 1, 2020, Pages 1-9.

[29] Abdullah Alshanqiti and AbdallahNamoun, "Predicting Student Performance and Its Influential Factors Using Hybrid Regression and Multi-Label Classification", IEEE Access, Volume 8, Pages203827 – 203844.

[30] Mustafa Yagci, "Educational data mining: prediction of students' academic performance using machine learning algorithms", Smart Learning Environments, Springer, Volume 9, Issue 11, 2022, Pages 1-19.