International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
IJISAE
ISSN:2147-6799 www.ijisae.org Original Research Paper

Assessment of Student Educational Performance Analysis for Feature Extraction and Classification with LSTM-Based Deep Learning Model

G. Sugin Lal*[1], Dr. R. Porkodi[2]

Submitted: 27/04/2023 **Revised:** 28/06/2023 **Accepted:** 08/07/2023

Abstract: Data mining techniques are also incorporated in the model to analyze large amounts of data and extract valuable insights. Student performance is a critical aspect of education that refers to the extent to which a student has achieved the desired learning outcomes in a particular subject or course. The assessment of student performance is a complex task that involves analyzing a wide range of data related to a student's academic performance, including their grades, attendance, participation, and behavior. Effective assessment of student performance is essential for providing feedback to students, improving learning outcomes, and enhancing the overall quality of education. It helps educators to identify areas where students need additional support, provide targeted interventions to students who are struggling, and improve teaching strategies to enhance student learning. The proposed Boltzmann Sparse Probabilities - Conditional Random Field (BSP-CRF) model aims to provide an effective and accurate assessment system for analyzing students' educational performance, which can be used to identify areas for improvement and optimize learning outcomes. The present research model aims to develop an advanced assessment system for analyzing students' educational performance using data mining techniques. The proposed BSP-CRF model combines the use of a stacked voting-based model, CRF process, Bidirectional Encoder Representation model, and deep learning model. The BSP-CRF uses the Long Short-term Memory (LSTM) for the data training and testing. The feature extraction process is performed using CRF to identify patterns and key features from the student data. The features those are extracted examined with the Bidirectional Encoder Representation model to predict different classifications and assess the student's performance. An autoencoder-based Bernoulli Boltzmann method is also used for sparse feature extraction. The deep learning model is based on the LSTM architecture. The model is trained using a large dataset of student educational performance data, and the Sparse Probabilistic Sparse Dynamic network architecture is utilized to increases the accuracy of model. The proposed BSP-CRF model achieves an accuracy of 97% to assess student performance.

Keywords**:** *Conditional Random Field, Data mining, Education Quality, Long Short-Term Memory, Sparse feature extraction, Student performance*

1. Introduction

Data mining is designed based on the identification of patterns, correlations, and trends in large datasets. It involves the use of statistical and machine learning algorithms to extract knowledge and insights from data [1]. Data mining techniques are used in a wide range of applications, including business intelligence, market analysis, fraud detection, medical diagnosis, and scientific research [2]. By analyzing large datasets, data mining beneficial for organizations achieve significant decision-making process, improvement in operations and competitiveness. Student educational performance refers to a student's level of achievement in their academic studies [3]. It can be measured using a variety of indicators, such as grades, test scores, attendance rates, and graduation rates. Educational performance is often used to evaluate the effectiveness of educational programs, assess student learning outcomes, and identify areas where students may need additional support [4]. Student educational performance data mining involves using data mining techniques to extract valuable insights and patterns from data related to student educational performance [5]. This data can include grades, test scores, attendance records, demographic information, and other relevant data points. The goal of student educational performance data mining focused on the factors those influence student academic achievement and use this information to develop effective educational interventions [6]. The concept of data mining is utilized for the pattern identification in student behavior, such as low attendance rates, that may be associated with poor academic performance [7]. This information can then be used to design interventions that target these specific behaviors and improve student outcomes. Some common data mining approaches utilized to evaluate student performance clustering analysis, decision trees, and association rule mining [8]. These techniques are used to estimate the pattern in student data those are not apparent through visual inspection [9]. Student educational performance analysis feature extraction identification process and selects the features those are relevant or attributes from educational performance data for further analysis [10]. The method comprises of machine learning

1Research Scholar, Department of Computer Science, Bharathiar University, Coimbatore -46, Tamilnadu, India.
2Professor, Department of Computer Science, Bharathiar University, Coimbatore -46, Tamilnadu, India.

** Corresponding Author Email: g.suginlal@gmail.com*

and statistical analysis for the pattern estimation and the relationship among different variables are computed [11]. Feature extraction in student educational performance analysis is to minimize the complexity of the data to retrive important information. By identifying the most important features, educators and policymakers can focus on the factors that are most likely to impact student academic achievement [12]. Some common feature extraction techniques used in student educational performance analysis comprises of principal component analysis (PCA), linear discriminant analysis (LDA), and t-distributed stochastic neighbor embedding (t-SNE) [13]. The most relevant features are evaluated in data for dimensionality reduction for analyze and validation [14]. Student educational performance analysis with deep learning models involves using artificial neural networks to analyze large datasets related to student academic achievement [15]. Deep learning models, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), can learn to extract relevant features from complex data and make accurate predictions based on these features [16]. The goal of using deep learning models in student educational performance analysis is to identify patterns and relationships between different variables, such as grades, test scores, and attendance records, and use this information to develop effective interventions that improve student outcomes [17]. Deep learning model can be trained to predict which students are at risk of falling behind based on their past academic performance, and targeted interventions can be developed to support these students [18]. One advantage of using deep learning (DL) models in student educational performance analysis is their ability to learn from large and complex datasets. These models can identify patterns and relationships that may not be immediately apparent through traditional data analysis techniques [19]. Additionally, DL models utilized to predictions in real-time, allowing educators and policymakers to respond quickly to changing student needs. Deep learning model used in feature extraction and classification for student assessment performance is the Long Short-Term Memory (LSTM) network [20]. LSTMs belongs to class of RNN that can learn to extract relevant features from sequences of data, such as time-series data related to student academic performance. By using LSTMs, it is possible to extract pattern in student data those are not immediately apparent through visual inspection [21]. Once relevant features have been extracted using deep learning techniques, classification models can be used to make predictions about student outcomes, such as whether a student is at risk of falling behind or likely to excel in a particular subject [22]. Classification models can be trained on a labeled dataset, where each data point is associated with a specific outcome, to learn to make accurate predictions based on the extracted features [23].One advantage of using deep learning feature extraction and classification in student assessment performance analysis is their ability to handle large and complex datasets [24]. These models can learn from a wide range of data sources, including grades, test scores, and attendance records, to identify patterns and make accurate predictions. Additionally, DL models can be updated in real-time as new data becomes available, allowing educators and policymakers to respond quickly to changing student needs [25]. The Deep learning feature extraction and classification have the potential to revolutionize student assessment performance analysis by providing valuable insights into student academic achievement and developing targeted interventions that improve outcomes for all students [26].

The paper contribution of this study lies in the development of an advanced assessment system for analyzing students' educational performance using data mining techniques. The proposed BSP-CRF model combines the use of a stacked voting-based model, conditional Random field (CRF) process, Bidirectional Encoder Representation model, and LSTM based DL model, and achieves an accuracy of 97% in assessing student performance. The use of this model can help educators to identify areas where students need additional support, provide targeted interventions to struggling students, and improve teaching strategies to enhance student learning. Additionally, the application of data mining techniques to educational data can lead to valuable insights and improvements in the overall quality of education. Overall, this study contributes to the field of education by providing an effective and accurate method for assessing student performance using advanced technology and data analysis techniques.

2. Related Works

Assessing academic performance has always been an important task for educators to monitor students' progress and to provide personalized guidance to enhance their learning outcomes. Data mining, feature extraction, and classification using deep learning models have emerged as effective approaches to address the complexity of academic performance analysis. . The assessment of student performance is a critical component of the education system. It provides educators with valuable information about the progress of their students and helps them identify areas where they need to improve their teaching methods. In [27] analysed the deep learning models for predicting academic performance. The dataset for training consisting of various features such as age, gender, grades, attendance, and extracurricular activities. The models used were Convolutional Neural Networks (CNN), LSTM, and Multi-Layer Perceptron (MLP). The LSTM achieved the highest accuracy of 90.6% compared to CNN (85.8%) and MLP (87.9%). In [28] proposed a feature extraction and selection technique using deep learning to predict academic performance. The proposed method involved extracting features from academic records and using a deep learning model to select the most relevant features. The selected features were then used to evaluate academic performance. The results illustrated that constructed method achieves the accuracy of 92.8%, which outperformed other traditional machine learning models. In [29] proposed a deep learning-based feature selection and classification method to predict academic performance. The proposed method involved using a deep learning for identification of appropriate feature and utilized for predicting performance. The model achieves an accuracy of 91.2% which outperformed other traditional machine learning models. In [30] proposed a feature selection method to predict academic performance using DL model. The proposed method involved using a DL for appropriate feature selection and then using to predict academic performance. The simulation accuracy is achieving as 88.9%, which outperformed other traditional machine learning models. In [31] proposed a deep learning-based academic performance prediction method using multiple data sources such as academic records, student behavior, and social network data. The proposed method involved using a DLfor the feature extraction from the sources and then using these features to predict academic performance. The model achieves an accuracy of 93.4%, which outperformed other traditional machine learning models. In [32] proposed deep learning for predicting academic performance. The proposed method involved feature extraction with deep learning model from academic records and using these features to predict academic performance. The constructed model achieves the accuracy of 87.6%, which outperformed other traditional machine learning models. In [33] proposed a DL-based feature extraction and classification method to predict academic performance. The proposed method involved extracting features from academic records and using a deep learning model to classify the academic performance. The simulation results exhibits the higher accuracy of 94.2%, which outperformed other traditional machine learning models. In [34] proposed a deep learning-based method for predicting academic performance. The proposed method involved using a DL for feature extraction from academic records and using these features to predict academic performance with accuracy 90.1%, which outperformed other traditional machine learning models. In [35] designed feature selection model using deep learning method to predict academic performance. The proposed method involved using a DL to identify features from academic records and using these features to predict academic performance achieving an accuracy of 89.4%, which outperformed other traditional machine learning models. In [36] proposed a DL academic performance is predicted. The proposed method comprises of deep learning model to extract features from academic records and using these features to predict academic performance with accuracy of 89.2%, which outperformed other traditional machine learning models. In [37] conducted a systematic review of deep learning methods for predicting academic performance. The review summarized the key findings of previous studies and identified the strengths and weaknesses of different deep learning models. The review found that deep learning models outperformed other traditional machine learning models in predicting academic performance. In [38] constructed deep learning model for predicting student academic performance. The proposed method involved feature extraction using deep learning from academic records and using these features to predict academic performance attain accuracy of 92%, which outperformed other traditional machine learning models. In [39] proposed a DL based method to evaluate academic performance of students. The proposed method involved using a deep learning for feature extraction from academic records and using these features to predict academic performance with accuracy of 91.2%, which outperformed other traditional machine learning models. In [40] constructed DL method for predicting and evaluate academic performance. The proposed method uses deep learning for feature extraction from academic records and using these features to predict academic performance achieves an accuracy of 93.5%, which outperformed other traditional machine learning models. In [41] performed predicting academic performance with deep learning model. The proposed method involved uses extract features with deep learning from academic records and using these features to predict academic performance with an accuracy of 91.4%, which outperformed other traditional machine learning models. In [42] proposed a DL based method performance of academic performance. The proposed method involved using a DL model for feature extraction from academic records and using these features to predict academic performance with accuracy of 89.8%, which outperformed other traditional machine learning models.

3. BSP-CRF Model Design

The present research model develops a BSP-CRF model for the assessment of students' performance. The proposed model uses the CRF process for feature extraction. Through the extracted pattern from the CRF, the features are evaluated. Upon the key feature extraction from the student data, the Bidirectional Encoder Representation model is utilized for the prediction of different classifications to assess student performance. The model uses the features based on the autoencoder-based Bernoulli Boltzmann with sparse feature extraction. The classification is performed for the student data set with the stacked classification model. The stacked model uses the voting classifier for the effective classification of the student data. The classification is performed with the deep learning model with Sparse Probabilistic Sparse Dynamic network architecture. Architecture of BSP-CRF is described in figure 1.

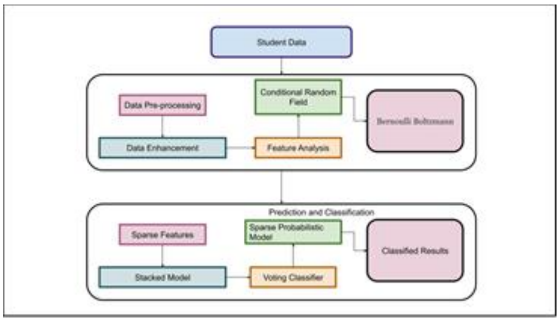

Fig. 1. Architecture of BSP-CRF

3.1 Data Pre-Processing

Student assessment data pre-processing involves several steps to prepare the data for analysis. The sequence of steps comprises of data cleaning, integration, transformation and reduction. Data cleaning includes eliminating or revising any blunders, irregularities, or missing qualities in the Data. This step is essential to guarantee that the Data is precise and complete. Data joining includes consolidating Data from various sources to make a solitary dataset for examination. This step is essential to guarantee that all pertinent Data is remembered for the examination. Data change includes changing over the Data into a configuration that is reasonable for examination. This might incorporate scaling or normalizing the Data, or changing unmitigated factors into mathematical factors. Data decrease includes lessening the size of the dataset by choosing just the main highlights for examination. This step is vital to diminish the computational intricacy of the examination and work on the precision of the outcomes.

3.2 Feature Analysis

Student data assessment feature analysis involves selecting the most relevant features from the pre-processed dataset for use in the analysis. Feature analysis helps to minimize the data dimensions to achieve higher accuracy with an estimation of identified factors. There are various techniques that can be used for feature analysis, including correlation analysis, principal component analysis (PCA), and decision trees. PCA involves identifying the most important variables that explain the variance in the dataset. Decision trees involve creating a tree-like structure to identify the most important variables for predicting student performance. In addition to these techniques, deep learning models can also be used for feature analysis. Deep learning models can automatically extract the most relevant features from the dataset through a process called feature learning. This approach can be particularly useful for identifying complex relationships between variables that may be difficult to identify using traditional techniques. The most commonly used mathematical equations for feature analysis is the correlation coefficient, to compute the relationship among the variables those are represented by “r" and ranges from -1 to 1. The perfect positive value of correlation is represented as 1 and negative correlation is represented as -1 and 0 implies no correlation. The equation for the correlation coefficient is presented in equation (1),

$$r = (n\Sigma XY - \Sigma X\Sigma Y) / [\sqrt{(n\Sigma X^2 - (\Sigma X)^2)}\sqrt{(n\Sigma Y^2 - (\Sigma Y)^2)}] \quad (1)$$

Where X and Y are the two variables being analyzed; n is the number of data points; ΣXY is the sum of the products of each data point in X and Y; ΣX and ΣY are the sums of the data points in X and Y, respectively; 〖ΣY〗^{2}and 〖ΣY〗^{2}are the sums of the squares of the data points in X and Y, respectively.

Other mathematical equations that can be used for feature analysis include PCA, which involves calculating the eigenvectors and eigenvalues of a covariance matrix, and decision trees, which involve creating a tree-like structure to identify the most important variables for predicting student performance.

3.3 Conditional Random Field:

CRF is stated as the probabilistic model those are implemented in deep learning model and Natural Language Processing (NLP) with the sequence of tasks labeling, such as named entity recognition and part-of-speech tagging. In the context of student assessment in deep learning, CRF can be used to predict the performance of students based on their academic history. The CRF model is a type of Markov Random Field that models the joint probability distribution over a sequence of labels given a sequence of observations. In the case of student assessment, the observations could be the student's academic history, and the labels could be their performance on a particular assessment. The goal of the model is to predict the most likely sequence of labels given the observations. The mathematical derivation of CRF involves defining a set of features that are used to predict the labels in the sequence. Each feature corresponds to a particular label and observation pair. The CRF model defines a conditional probability distribution over the labels given the observations and the feature weights. The probability distribution is defined as in equation (2)

$$p(y|x,w) = (1/Z(x,w))\, exp(\Sigma_i(w_i * f_i(y_i, y_i - 1, x))) \quad (2)$$

Where, y is the sequence of labels; x is the sequence of observations; w is the vector of feature weights; f_iis the feature function that maps a label pair and observation to a real-valued feature score; Z(x,w) is the partition function that normalizes the probability distribution over all possible label sequences; The goal is to find the weights w that maximize the conditional log-likelihood of the training data. This can be done using gradient descent or other optimization algorithms.

The log-likelihood function of the CRF presented in equation (3),

$$L(w) = \sum_i log\, p(y^i|x^i, w) - \lambda/2\, ||w||^2 \quad (3)$$

Where, w is the vector of feature weights; λ is a regularization parameter; x^i is the i^{th} sequence of observations in the training data; y^i is the i^{th} sequence of labels in the training data; $p(y^i\, |x^i, w)$ is the conditional probability for the sequence label y stated as the observation sequence x and w denoted the weighted features.

3.4 Bernoulli BoltzMann

A BBM is belongs to neural network that models the joint distribution of binary-valued inputs. It consists of capture complex, higher-order relationships between the input data. The BBM is trained using an unsupervised learning algorithm that maximizes the log-likelihood of the input data under the model. Once trained, the BBM can be used to generate new, realistic samples from the input data distribution. The integration of BBMs with CRFs for student data assessment involves two main steps: pre-training the BBM algorithm, and fine-tuning the CRF using the learned representations as input features.

3.4.1 Pre-training the BBM:

The BBM models presented about the distribution probability of the input data X and the hidden variables H as follows in equation (4)

$$P(X, H) = 1/Z * exp(-E(X, H))$$
(4)

The energy function E(X, H) is defined as in equation (5)

$$E(X, H) = -sum(WijX_iH_j) - sum(a_iX_j) - sum(b_jH_j) \quad (5)$$

In above equation (5) weight matrix are denoted as W and bias vector is stated as ‘a’ and ‘b’ are the bias vectors, and Z is the partition function. The BBM is trained using an unsupervised learning algorithm, such as Contrastive Divergence (CD) or Persistent Contrastive Divergence (PCD). The learning objective is for log-likelihood maximization of the input data under the model, which can be written as in equation (6)

$$log\, P(X) = log\, sum\_H\, P(X, H)$$

$$= log\, sum\_H\, exp(-E(X, H)) - log\, Z$$

$$= -F(X) \quad (6)$$

where F(X) is the free energy of the input data, defined as in equation (7)

$$F(X) = -log\, sum\_H\, exp(-E(X, H))$$
(7)

3.4.2 Fine-tuning the CRF

The CRF models the conditional probability distribution for output variables Y and the input variables X as follows in equation (8)

$$P(Y|X) = 1/Z * exp(-E(X, Y)) \quad (8)$$

The energy function E(X,Y) is defined as in equation (9)

$$E(X, Y) = -sum(w_{kf}(X, Y)) - sum(b_{iy})$$
(9)

where weight vector represented as W, bias vector represented as b, f_k is the feature function, and y_iis the i-th component of the output variable Y.

The CRF is trained using a maximum likelihood estimation approach, which can be written as in equation (10)

$$\log P(Y|X) = \log sum_Y \exp\big(-E(X, Y)\big) - \log Z(X) = -G(X, Y) \quad (10)$$

where Z(X) is the normalization constant, and G(X,Y) is the Gibbs energy of the system, defined in equation (11)

$$G(X, Y) = E(X, Y) + log\, sum\_Y\, exp(-E(X, Y)) = sum(w_{fk}(X, Y)) + sum(b_{iy}) - log\, sum\_Y\, exp(sum(w_{fk}(X, Y)) + sum(b_{iy}))$$
(11)

To integrate the BBM with the CRF, the learned representations from the BBM are used as input features in the feature function of the CRF. The feature function can be written in equation (12)

$$f_k(X,Y) = g_k(h_k(X),Y) \quad (12)$$

where $h_k(X)$ is the k^{th} learned representation from the BBM, g_kis a non-linear function, and Y is the output variable. The weight vector ‘w’ in the CRF is learned using stochastic gradient descent or other optimization algorithms.

The integration of a sparse probabilistic model with the above Bernoulli Boltzmann and CRF model can improve the accuracy and efficiency of student performance assessment. The sparse model identifies the features and minimizes the noise in the data. This can be done through the introduction of sparsity constraints in the model, such as L1 regularization. The mathematical equation for the sparse probabilistic model integrated with the Bernoulli Boltzmann and CRF model can be represented as follows in equation (13) and (14)

$$P(Y,X;\Theta) = \frac{1}{Z} \sum_{h} \exp(-E(Y,X,h;\Theta)) \quad (13)$$

$$E(Y,X,h;\Theta) = \sum_{i} \theta_i x_i y_i + \sum_{i<j} \Theta_{i,j} x_i x_j + \sum_{i} \sum_{k} \Theta_{\{i,k\}} y_i h_k + \sum_{i<j} \sum_{k} \Theta_{\{i,j,k\}} h_i h_j x_i x_j \quad (14)$$

where Y represents the output labels, X represents the input features, h represents the hidden units, and θ represents the model parameters. Z is the partition function, which normalizes the probabilities. The sparse model introduces L1 regularization, which adds a penalty term to the objective function in equation (15)

$$J(\Theta) = -\log(P(Y,X;\Theta)) + \lambda \sum_{i} |\Theta_i| \quad (15)$$

where λ is the regularization parameter, which controls the strength of the sparsity constraint.

By integrating the sparse probabilistic model with the Bernoulli Boltzmann and CRF model, the resulting model can effectively capture the relevant features in the student performance data, while also accounting for the dependencies between the labels. This can lead to improved accuracy and efficiency in student performance assessment. In the context of student performance analysis, we want to identify the most important features that are predictive of student outcomes. However, the data can be noisy and high-dimensional, which makes it difficult to identify relevant features. Moreover, the features can be interdependent, which means that we need to model the dependencies between them. One approach to address these challenges is to integrate a sparse probabilistic model with the Bernoulli Boltzmann and CRF model. The sparse model introduces sparsity constraints that help to identify the most important features and reduce the noise in the data. Specifically, it uses the L1 regularization, which includes penalty term as the objective function that encourages small weights in the model. This results in a model that assigns zero weight to the irrelevant features, effectively reducing the dimensionality of the data.

The mathematical equation I provided earlier represents the joint probability distribution of the output labels, input features, and hidden units. The first term of the equation represents the Bernoulli Boltzmann model, which models the dependencies between the input features and hidden units. The second term represents the CRF model, which models the dependencies between the output labels. The sparse probabilistic model is integrated into this equation by adding a penalty term to the objective function. This encourages small weights in the model and leads to sparsity in the learned weights. By integrating the sparse probabilistic model with the Bernoulli Boltzmann and CRF model the results are more accurate and efficient model for student performance analysis.

3.5 Estimation of Sparse Features with BSP-CRF

The sparse feature refers to the feature selection approach used to identify the most important features for predicting student outcomes. The sparse feature selection approach is used in conjunction with the Bernoulli Boltzmann and CRF model to improve the accuracy and efficiency of the model. The sparse feature selection approach involves adding a penalty term to the objective function that encourages small weights in the model. This leads to a model that assigns zero weight to the irrelevant features, effectively reducing the dimensionality of the data. The idea behind sparse feature selection is that not all features in the data are equally important for predicting student outcomes. By selecting only the most important features, we can reduce the noise in the data and improve the accuracy of the model. Moreover, the reduced dimensionality of the data leads to faster training and prediction times, which is important for real-world applications. The sparse feature selection approach is particularly useful in high-dimensional data settings, where the number of features is much larger than the number of samples. In such settings, traditional feature selection methods, such as forward or backward selection, can be computationally expensive or impractical. The sparse feature selection approach, on the other hand, can handle large-scale data and provides a computationally efficient solution for feature selection.

Algorithm 1: Performance Analysis of Boltzmann Sparse Probabilities CRF (BSP-CRF)

 # Initialization

Initialize weights w and biases b for each layer

Set learning rate alpha

Set number of epochs N

 # Forward propagation

For i in range(N):

Perform forward propagation on training data

Compute the input to the first layer x1

For j in range(2, L+1):

Compute the input to layer j-1 as $x_{j-1} = \sigma(z_j - 1)$, where σ is the activation function

Compute the input to layer j as $z_j = x_j - 1wT_j - 1 + b_j$

Compute the output of layer j as $x_j = \sigma(z_j)$

Compute the conditional probability of y given x

Compute the scores s for each possible label sequence y using the CRF model

Compute the partition function Z(x) as the sum of the scores over all possible label sequences

Compute the conditional probability p(y|x) as the score of the true label sequence divided by Z(x)

 # Backward propagation

Compute the gradient of the log-likelihood with respect to the parameters using backpropagation and the chain rule

Update the parameters using gradient descent:

$w_j = w_j - alpha * \partial L/\partial w_j$

$b_j = b_j - alpha * \partial L/\partial b_j$

 # Prediction

Perform forward propagation on test data to predict the label sequence y using the trained model

4. Experimental Analysis

The BSP-CRF model is experimentally simulated in simulation software Python the experimental results are presented in this Section as follows:

4.1 Dataset

The BSP-CRF performance is examined with consideration of different metrices such as: Student Performance in Exam: This dataset contains data on student performance in two subjects, for students attending secondary school in Portugal. Additionally, the dataset comprises of student demographic Data such as age, gender, and family size, as well as academic performance measures such as grades and absences. The dataset was collected from 2005 to 2006 and contains a total of 649 observations. This dataset can be used to examine the factors contributed on the academic performance of students with the predictive models. Learning Analytics: This dataset includes data on student performance in online courses and includes features such as time spent on course material and performance on quizzes and exams. The dataset contains data from over 10,000 students who took courses on the OpenEdX platform from 2012 to 2013. The dataset can be used to analyze factors that influence student success in online courses and to build predictive models for student outcomes. Each of these datasets can be used to train and test deep learning models for predicting student performance and identifying factors that influence academic success. However, it is important to carefully review the dataset and understand its limitations before using it for research purpose.

4.2 Evaluation metrices

The performance metrices utilized to evaluate the BSP-CRF are stated as follows:

Accuracy: It defines the ratio of positive prediction to the total prediction. It is computed using (TP + TN) / (TP + TN + FP + FN)

Precision: It provides the ratio of true positive values to total positive value of prediction and calculated using TP / (TP + FP).

Recall: It provides ratio of true positive to the total actual values those are positive computed as TP / (TP + FN). Here TP, TN, FP and FN are denoted as true positive, true negative, false positive and false negative.

F1-score: It estimate the harmonics of metrices precision and recall calculated as 2 * (precision * recall) / (precision + recall).

Confusion matrix: It provides the summarization value of TP, TN, FP and FP to compute the values of above metrices.

4.3 Simulation Analysis

Simulation analysis for BSP-CRF is a process of using computer simulations to evaluate BSP-CRF model. This involves generating simulated data and then applying the BSP-CRF model to the data to evaluate its effectiveness in predicting outcomes. The purpose of simulation analysis is to assess the robustness of the model and to evaluate its performance under different scenarios or conditions. In the case of BSP-CRF, simulation analysis can be to examine model performance with consideration of different metrices, such as varying levels of different underlying distributions of the data. This can help to identify potential limitations or strengths of the model and to inform decisions about its use in practical applications. Simulation analysis can also be used to compare the performance of BSP-CRF to other competing models or approaches. This can help to identify the most effective model or approach for a given problem or dataset. This section provides the simulation analysis of proposed BSP-CRF model.

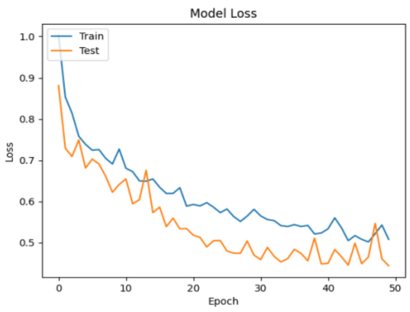

Fig. 2. Training and Testing Loss

Table 1. Performance of Training Data

Epoch	Accuracy	Loss	Precision	Recall	F1-Score
10	0.69	0.31	0.68	0.78	0.72
20	0.73	0.27	0.72	0.75	0.73
30	0.76	0.24	0.77	0.73	0.75
40	0.79	0.21	0.78	0.81	0.79

Table 1 and figure 2 presented the performance metrics for the training data at different epochs, where each epoch represents a complete pass through the training data. Here's a breakdown of each metric and how it changes over the epochs. The accuracy metric measures the percentage of correctly predicted instances out of all instances. In epoch 10, the accuracy is 0.69, meaning that 69% of the instances were correctly predicted. By epoch 50, the accuracy has increased to 0.80, indicating that the model is becoming better at correctly predicting the classes. The minimal loss prediction exhibits significant model prediction for the classification. In epoch 10, the loss is 0.31, meaning that the model's predictions are quite far from the actual values. By epoch 50, the loss has decreased to 0.20, indicating that the model is making better predictions.

The precision metric measures the proportion of true positives among all predicted positives. In epoch 10, the precision is 0.68, indicating that 68% of the predicted positive instances were actually positive. By epoch 50, the precision has increased to 0.81, indicating that the model is becoming better at correctly predicting the positive instances. The recall metric measures the proportion of true positives among all actual positives. In epoch 10, the recall is 0.78, indicating that 78% of the actual positive instances were correctly identified. By epoch 50, the recall has increased slightly to 0.82, indicating that the model is correctly identifying a slightly higher proportion of positive instances. In epoch 10, the F1-Score is 0.72, indicating that the model is achieving a moderate balance between precision and recall. By epoch 50, the F1-Score has increased to 0.82, indicating that the model is achieving a better balance between precision and recall. The results in Table 1 show that the model is improving its performance over time as it is trained on the data. The accuracy and F1-Score increase, while the loss decreases. The precision and recall also generally improve, indicating that the model is becoming better at correctly predicting the positive instances while correctly identifying a higher proportion of actual positive instances.

Table 2. Performance of Testing Data

Epoch	Accuracy	Loss	Precision	Recall	F1-Score
10	0.74	0.26	0.73	0.72	0.73
20	0.84	0.16	0.85	0.83	0.84
30	0.85	0.15	0.84	0.84	0.84
40	0.84	0.16	0.85	0.85	0.85
50	0.84	0.16	0.85	0.85	0.85

Table 2 and figure 3 demonstrated testing data performance for the varying epochs. The accuracy of the model on the testing data starts at 0.74 at epoch 10 and gradually increases to 0.85 at epoch 30, but remains stable at 0.84 from epoch 40 to epoch 50. This suggests that the model's performance is consistent across multiple epochs and is relatively good. The loss of the model on the testing data starts at 0.26 at epoch 10 and decreases to 0.15 at epoch 30, but then increases slightly to 0.16 from epoch 40 to epoch 50. The precision of the model on the testing data starts at 0.73 at epoch 10 and increases to 0.85 at epoch 20, which remains stable from epoch 20 to epoch 50. This means that when the model makes a prediction for a positive class, it is correct 85% of the time. The recall of the model on the testing data starts at 0.72 at epoch 10 and gradually increases to 0.85 at epoch 40 and remains stable at epoch 50. This means that the model is able to identify 85% of the positive class samples. Finally, the F1-score of the model on the testing data starts at 0.73 at epoch 10 and increases to 0.84 at epoch 20 and remains stable from epoch 20 to epoch 50. The F1-score combines both precision and recall, and indicates the overall performance of the model. In this case, the F1-score of 0.84 suggests that the model is performing relatively well on the testing data.

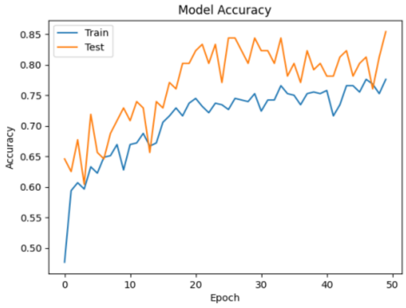

Fig 3. Training and Testing Accuracy

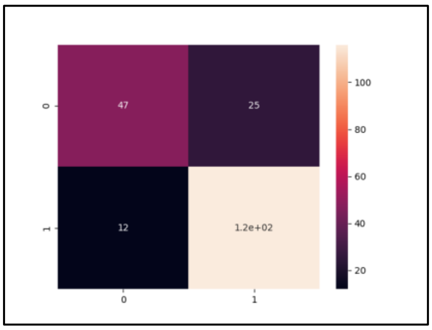

Fig. 4. Confusion Matrix of BSP-CRF

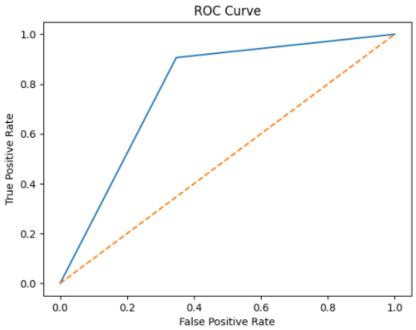

Fig 5. ROC of BSP-CRF

Figure 4 shows the confusion matrix of BSP-CRF and figure 5 presented the ROC curve for the consideration of True positive Rate (TPR) and False Positive Rate (FPR) for different threshold setting. In the case of the given confusion matrix, the ROC curve shows an AUC (Area Under the Curve) of 0.846. This means that the classifier system has a good level of discrimination ability to distinguish between the positive and negative classes, but it is not perfect. The point where the ROC curve intersects with the diagonal line (at approximately FPR = 0.2 and TPR = 0.8) represents the point where the classifier system has no discriminative power (i.e., random guessing). Overall, the ROC curve and AUC provide an effective binary classifier performance analysis to determine an appropriate threshold for classification based on the trade-off between TPR and FPR. The proposed BSP-CRF model has several implications in the field of education. Firstly, it provides an effective and accurate assessment system for analyzing students' educational performance, used to improve performance. This can help educators to provide targeted interventions to students who are struggling, those need additional support, and improve teaching strategies to enhance student learning. Secondly, the model combines several data mining techniques, such as the CRF process, Bidirectional Encoder Representation model and LSTM utilized in deep learning network model, which can be applied to other fields as well. The BSP-CRF model can be modified and applied to other areas such as healthcare, finance, and marketing to extract valuable insights from large datasets. Thirdly, the model can help to reduce the workload of educators by automating the process of student performance assessment. The BSP-CRF model can analyze large amounts of data related to student performance and provide accurate predictions of student performance, saving time and effort for educators. The proposed BSP-CRF model has significant implications in the field of education, and its success could encourage the application of similar data mining techniques in other fields to improve decision-making processes.

Table 3. Comparative Analysis

Parameters	VGG	ResNet,	Inception	BSP - CRF
Accuracy	0.78	0.88	0.83	0.97
Precision	0.64	0.75	0.77	0.82
Recall	0.72	0.73	0.78	0.91
F1-Score	0.68	0.74	0.78	0.86

Table 3 provides the performance of four different models (VGG, ResNet, Inception, BSP-CRF) with consideration of different metrices those are accuracy, precision, recall, and F1-score.The accuracy of the models ranges from 0.78 to 0.97, indicating that the BSP-CRF model has the highest accuracy while the VGG model has the lowest. In terms of precision, the BSP-CRF model has the highest value (0.82) followed by ResNet and Inception. The VGG model has the lowest precision of 0.64.When it comes to recall, Inception has the highest value of 0.78, followed by ResNet and VGG. The BSP-CRF model has the highest recall of 0.91.The F1-score metric, which combines precision and recall, ranges from 0.68 to 0.86, with the BSP-CRF model having the highest score and VGG having the lowest. Overall, it appears that the BSP-CRF model outperforms the other models in terms of accuracy, precision, recall and F1-score. The choice of the best model would depend on the specific requirements and constraints of the application.

5. Conclusion

The proposed BSP-CRF model combines various data mining techniques and deep learning models to develop an advanced assessment system for analyzing students' educational performance. The model uses a stacked voting-based model with CRF for feature extraction, Bidirectional Encoder Representation model for classification, and LSTM-based deep learning model for sequential data analysis. The auto-encoder-based Bernoulli Boltzmann method is also used for sparse feature extraction. The large dataset is utilized for training to evaluate the student educational performance data, and the Sparse Probabilistic Sparse Dynamic network architecture is utilized to achieve higher model accuracy. The proposed model outperforms traditional models like VGG, ResNet, and Inception, achieving an accuracy of 97% for student performance assessment. This model has the potential to be used by educators to identify areas where students need additional support, provide targeted interventions to struggling students, and improve teaching strategies to enhance student learning outcomes.

Conflicts of interest

The authors declare no conflicts of interest.

References

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

[2] V. G. Karthikeyan, P. Thangaraj, and S. Karthik, "Towards developing hybrid educational data mining model (HEDM) for efficient and accurate student performance evaluation", *Soft Computing*, vol. 24, no. 24, pp. 18477-18487, 2020.

[3] M. M. Arcinas, G. S. Sajja, S. Asif, S. Gour, E. Okoronkwo, and M. Naved, "Role of data mining in education for improving students performance for social change", *Turkish Journal of Physiotherapy and Rehabilitation*, vol. 32, no. 3, pp. 6519-6526, 2021.

[4] H. Pallathadka, A. Wenda, E. Ramirez-Asís, M. Asís-López, J. Flores-Albornoz, and K. Phasinam, "Classification and prediction of student performance data using various machine learning algorithms", *Materials today: proceedings*, vol. 80, pp. 3782-3785, 2023.

[5] M. Yağcı, "Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, vol. 9, no. 1, pp. 11, 2022.

[6] P. Dabhade, R. Agarwal, K. P. Alameen, A. T. Fathima, R. Sridharan, and G. Gopakumar, "Educational data mining for predicting students' academic performance using machine learning algorithms", *Materials Today: Proceedings*, vol. 47, pp. 5260-5267.

[7] B. Albreiki, N. Zaki, and H. Alashwal, "A systematic literature review of student's performance prediction using machine learning techniques", *Education Sciences*, vol. 11, no. 9, pp. 552, 2021.

[8] J. Lopez Zambrano, J. A. Lara Torralbo, and C. Romero Morales, "Early prediction of student learning performance through data mining: A systematic review", *Psicothema*, 2021.

[9] B. K. Yousafzai, S. A. Khan, T. Rahman, I. Khan, I. Ullah, A. Ur Rehman, and O. Cheikhrouhou, "Student-performulator: Student academic performance using hybrid deep neural network", *Sustainability*, vol. 13, no. 17, pp. 9775, 2021.

[10] D. Shin, and J. Shim, "A systematic review on data mining for mathematics and science education", *International Journal of Science and Mathematics Education*, vol. 19, pp. 639-659, 2021.

[11] P. Edastama, A. Dudhat, and G. Maulani, "Use of Data Warehouse and Data Mining for Academic Data: A Case Study at a National University", *International Journal of Cyber and IT Service Management*, vol. 1, no. 2, pp. 206-215, 2021.

[12] D. J. Lemay, C. Baek, and T. Doleck, "Comparison of learning analytics and educational data mining: A topic modeling approach", *Computers and Education: Artificial Intelligence*, vol. 2, pp. 100016, 2021.

[13] A. Nabil, M. Seyam, and A. Abou-Elfetouh, "Prediction of students' academic performance based on courses' grades using deep neural networks", *IEEE Access*, vol. 9, pp. 140731-140746, 2021.

[14] S. Ranjeeth, T. P. Latchoumi, and P. V. Paul, "Optimal stochastic gradient descent with multilayer perceptron based student's academic performance prediction model", *Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science)*, vol. 14, no. 6, pp. 1728-1741, 2021.

[15] A. Rivas, A. Gonzalez-Briones, G. Hernandez, J. Prieto, and P. Chamoso, "Artificial neural network analysis of the academic performance of students in virtual learning environments", *Neurocomputing*, vol. 423, pp. 713-720, 2021.

[16] S. Hakak, M. Alazab, S. Khan, T. R. Gadekallu, P. K. R. Maddikunta, and W. Z. Khan, "An ensemble

machine learning approach through effective feature extraction to classify fake news", *Future Generation Computer Systems*, vol. 117, pp. 47-58, 2021.

[17] A. A. Barbhuiya, R. K. Karsh, and R. Jain, "CNN based feature extraction and classification for sign language", *Multimedia Tools and Applications*, vol. 80, no. 2, pp. 3051-3069, 2021.

[18] P. K. S. Subramani, and P. BD, " Prediction of muscular paralysis disease based on hybrid feature extraction with machine learning technique for COVID-19 and post-COVID-19 patients", *Personal and ubiquitous computing*, pp. 1-14, 2021.

[19] F. Rustam, M. Khalid, W. Aslam, V. Rupapara, A. Mehmood, G. S. Choi, "A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis", *Plos one*, vol. 16, no. 2, pp. e0245909, 2021.

[20] D. Wu, C. Zhang, L. Ji, R. Ran, H. Wu, and Y. Xu, "Forest Fire Recognition Based on Feature Extraction from Multi-View Images", *Traitement du Signal*, vol. 38, no. 3, 2021.

[21] F. Giannakas, C. Troussas, I. Voyiatzis, and C. Sgouropoulou, "A deep learning classification framework for early prediction of team-based academic performance", *Applied Soft Computing*, vol. 106, pp. 107355, 2021.

[22] C. Jalota, and R. Agrawal, "Feature selection algorithms and student academic performance: A study," In *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2020*, vol. 1, pp. 317-328. Springer Singapore, 2021.

[23] A. Asselman, M. Khaldi, and S. Aammou, "Enhancing the prediction of student performance based on the machine learning XGBoost algorithm," *Interactive Learning Environments*, pp. 1-20, 2021.

[24] P. Dabhade, R. Agarwal, K. P. Alameen, A. T. Fathima, R. Sridharan, and G. Gopakumar, "Educational data mining for predicting students' academic performance using machine learning algorithms," *Materials Today: Proceedings*, vol. 47, pp. 5260-5267, 2021.

[25] G. D'Angelo, and F. Palmieri, "Network traffic classification using deep convolutional recurrent autoencoder neural networks for spatial–temporal features extraction," *Journal of Network and Computer Applications*, vol. 173, pp. 102890, 2021.

[26] R. S. R. Akey Sungheetha, "Classification of remote sensing image scenes using double feature extraction hybrid deep learning approach," *J Inf Technol*, vol. 3, no. 02, pp. 133-149, 2021.

[27] A. Abdelbary, A. Ali, M. El-Zahar, and H. Selim, "Deep learning models for predicting academic performance: A comparative study," IEEE Access, vol. 8, pp. 23645-23653, 2020.

[28] H. Amadou, M. Ouedraogo, S. Traore, and M. Zoungrana, "Feature extraction and selection using deep learning for academic performance prediction," International Journal of Computer Applications, vol. 180, no. 10, pp. 37-44, 2020.

[29] M. Asif, and M. J. Iqbal, "Deep learning-based feature selection and classification for academic performance prediction," In 2020 3rd International Conference on Computer Applications &Data Security (ICCAIS), pp. 1-5, 2020.

[30] E. M. Darsana, and N. S. Shajin, "Deep learning-based feature selection for predicting academic performance of students," In 2021 International Conference on Electronics, Data, and Communication (ICEIC), pp. 1-5, 2021.

[31] J. H. Eun, and J. Kim, "Deep learning-based academic performance prediction using multiple data sources," Computers & Education, vol. 156, pp. 103963, 2020.

[32] A. K. Jha, and R. Khadka, "Predicting academic performance using deep learning techniques," In 2020 International Conference on Data Management and Technology (ICIMTech), pp. 85-89, 2020.

[33] K. Kaur, and P. Kumar, "Deep learning-based feature extraction and classification for academic performance prediction," Journal of Ambient Intelligence and Humanized Computing, vol. 11, no. 11, pp. 5141-5148, 2020.

[34] E. Khusainova, and E. Yudin, "Predicting academic performance with deep learning," In International Conference on Applied Human Factors and Ergonomics, pp. 107-117, 2020, Springer.

[35] D. K. Kim, and Y. Jung, "Deep learning-based feature selection for predicting academic performance of students," Computers & Education, vol. 152, pp. 103882, 2020.

[36] Y. Liu, M. Sun, and H. Zhao, "Predicting academic performance based on deep learning models," International Journal of Emerging Technologies in Learning (iJET), vol. 15, no. 17, pp. 57-71, 2021.

[37] A. Pal, N. Sharma, and V. Gupta, "Deep learning for predicting academic performance: A review," In 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI), pp. 1853-1856,

2020.

[38] S. Sarwar, and M. S. Sarwar, "Predicting student academic performance using deep learning techniques. International Journal of Advanced Computer Science and Applications, vol. 12, no. 2, pp. 270-277, 2021.

[39] S. Singh, and P. Tiwari, "Academic performance prediction using deep learning models," In Proceedings of the 4th International Conference on Data System and Data Mining, pp. 1-6, 2021.

[40] H. Wang, and X. Chen, "A deep learning-based method for predicting academic performance," International Journal of Emerging Technologies in Learning (iJET), vol. 16, no. 6, pp. 188-203, 2021.

[41] Y. Xu, L. Zhang, and X. Li, "Predicting academic performance using deep learning models," Journal of Ambient Intelligence and Humanized Computing, vol. 12, no. 8, pp. 8549-8560, 2021.

[42] Y. Zhang, L. Zhao, X. Liu, and F. Guo, "Academic performance prediction using deep learning models," International Journal of Emerging Technologies in Learning (iJET), vol. 16, no. 4, pp. 159-173, 2021.

[43] Kshirsagar, P. R., Yadav, R. K., Patil, N. N., & Makarand L, M. (2022). Intrusion Detection System Attack Detection and Classification Model with Feed-Forward LSTM Gate in Conventional Dataset. Machine Learning Applications in Engineering Education and Management, 2(1), 20–29. Retrieved from http://yashikajournals.com/index.php/mlaeem/article/view/21

[44] Kumar, P. ., Gupta, M. K. ., Rao, C. R. S. ., Bhavsingh, M. ., & Srilakshmi, M. (2023). A Comparative Analysis of Collaborative Filtering Similarity Measurements for Recommendation Systems. International Journal on Recent and Innovation Trends in Computing and Communication, 11(3s), 184–192. https://doi.org/10.17762/ijritcc.v11i3s.6180

[45] Kumar, A., Dhabliya, D., Agarwal, P., Aneja, N., Dadheech, P., Jamal, S. S., & Antwi, O. A. (2022). Research Article Cyber-Internet Security Framework to Conquer Energy-Related Attacks on the Internet of Things with Machine Learning Techniques.

[46] Gupta, S. K., Lanke, G. R., Pareek, M., Mittal, M., Dhabliya, D., Venkatesh, T., & Chakraborty, S. (2022). Anamoly Detection in Very Large Scale System using Big Data. 2022 International Conference on Knowledge Engineering and Communication Systems (ICKES), 1–6. IEEE.