INTERNATIONAL JOURNAL OF INTEGRATED ENGINEERING VOL. 14 NO. 7 (2022) 108-121

IJIE

http://penerbit.uthm.edu.my/ojs/index.php/ijie
ISSN : 2229-838X e-ISSN : 2600-7916

The International Journal of Integrated Engineering

Improved Lion Optimization based Enhanced Computation Analysis and Prediction Strategy for Dropout and Placement Performance Using Big Data

Kiran Kumar[1*], Kavitha K S[2]

[1]Data Science & Data Engineer
International School of Engineering (INSOFE), Bengaluru, Karnataka, INDIA

[2]Department of CSE
Global Academy of Technology, Bangalore 560098, INDIA

*Corresponding Author

DOI: https://doi.org/10.30880/ijie.2022.14.07.009
Received 26 April 2022; Accepted 1 July 2022; Available online 31 December 2022

Abstract: Background: Predicting the undergraduate's placement performance is vital as it impacts the credibility of educational institutions. Hence, it is significant to predict their performance based on placement in the early days of degree program. Objectives: The study intends to predict the undergraduate's placement performance through the introduced ANN-R (Artificial Neural Network based Regression) as it is able to handle fault tolerance. For efficient prediction, relevant feature selection is needed that is performed by the proposed ILO (Improved Lion Optimization) algorithm as it has the ability to find nearest probable optimal solution. Methodology: Initially, the parameters and population are initialized. Subsequently, first best-agent is stated in accordance with fitness function. Subsequently, position of present search agent is updated. This iteration continues until all the features are selected and optimized result is attained. Here best score is computed using the proposed ILO for feature selection. Finally, the dropout analysis and placement performance of students is predicted using the introduced ANN-R through a train and test split. Results/Conclusion: Performance of the proposed system is analyzed in accordance with loss metrics. Additionally, internal comparison is performed to find the extent to which the actual and predicted values correlate with one another during prediction using the existing and proposed system. The outcomes revealed that the proposed system has the ability to predict the student's placement performance along with domain of interest with minimum errors than the traditional system. This makes the proposed system to be highly suitable for predicting student's performance.

Keywords: Placement performance, dropout analysis, artificial neural network-based regression, improved lion optimization, machine learning and educational data

1. Introduction

Educational data has a vital role in predicting the performance of undergraduate students. Various existing studies have attempted to predict the placement performance of students through various data mining approaches. Accordingly, the conventional study [1] takes into account the influence of engineering student's Grade Point Average (GPA) through use of a Covenant University, Nigeria as the case study. Effort has been made in the research to present the degree to which the final year student's CGPA (Cumulative Grade Point Average) could be predicted through the use of GPA from the initial three years of the degree program. The dataset has been analyzed that comprised of thousand eight hundred and forty one students admitted and then graduated from 2002-2014 through seven engineering

*Corresponding author: kiran.kumar@insofe.edu.in

disciplines of the Cohort university. Analysis has been carried out using regression analysis in Matrix Laboratory (MATLAB) and Konstanz Information Miner model that rely in data mining. Dataset also comprised of GPA of students attained from the initial three years of their study and final CGPA. The prediction outcomes has shown an accuracy of 89% [2, 3]. Few existing studies aimed to predict the student placement percentage through the data associated with the academic reputation of the institution, city opportunities where the institution is situated, institution's cultural opportunities and facilities [4]. When the accuracy of the model has been assessed with respect to the performance metrics, the employed Extreme Gradient Boosting (XGBoost) explores more prediction accuracy in comparison to ML strategies. Further research has to incorporate extra experimentation with ML algorithms, attain many training records through the inclusion of supplementary majors or engineering departments for generalizing the model and enhance its performance. Various traditional systems have used classification algorithms for predicting the chances of undergraduate students in getting placed in companies. Random Forest (RF) and Decision Tree (DT) has been applied by existing work [5] for predicting the student's placement probability. Recommended model has intended to predict if the student gets placed or nor placed during the recruitment process. For this prediction, the student's academic history has been taken as a data like backlogs, overall percentage and credits. Algorithms have been employed on student's previous years. Accuracy attained after examining the proposed algorithms is found to 86% for RF and 84% for DT. Thus, the analytical outcomes explore the efficacy of RF than DT. Other than experimental analysis, few studies attempt to analyses through questionnaires that enquires the students regarding their attitudes in class and the programming task outcomes in the class [6]. Then, this information is utilized to make Machine Learning (ML) model for predicting the placement outcomes and ranking outcomes from the programming competition. Three descriptive variables have also been utilized that include programming task, psychological scale and student answered questionnaire. Effective prediction results have been accomplished. Yet, the suggested method's prediction performance has to be enhanced by improving the algorithms as well as incorporating other significant variables [7]. The existing studies have already made few advancements. Yet, further enhancement has to be accomplished to accurately predict the placement performance of the students as existing methods like Konstanz Information Miner model have shown 89%, RF have shown 86% and DT have shown 84%. Though effective outcomes have been attained by existing methods, they lack in terms of accuracy in predicting the placement performance of students. To further improve accuracy, the present study intends to employ ML models to predict the placement probability of the students and provide mentoring system for students based on their academic performance and dropout analysis by considering all the significant parameters for efficient prediction in real-time.
The main objectives of this research are listed below.

- To select relevant features from the dataset based on best score through the proposed ILO (Improved Lion Optimization) algorithm.
- To predict undergraduate student's placement performance in the early days of their degree program by considering aggregate mark, number of certificate courses, arrear not more than one, more interest in learning new things and core competent of the students and analyzing it using the introduced Artificial Neural Network based Regression (ANN-R).
- To develop a mentoring system for counselling the students based on the dropout analysis under certain threshold of student's aggregate mark, fees structure and attendance using fuzzy clustering.
- To perform internal comparison of the introduced system with the traditional system during the implementation with respect to Mean Square Error (MSE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE).

1.1 Paper Organization

The paper is organized as follows. Initially, the significance of predicting the student's performance is discussed. Followed by this, various methods used by existing systems in this content are explored. Then, section III presents the overall proposed study. The results obtained from the implementation of the proposed system are discussed in section IV. Finally, the overall study is concluded in section V.

2. Review of Existing Study

Various existing studies have used several methods to predict the undergraduate student's performance in accordance with numerous factors. Those methodologies are discussed in this section along with its drawbacks.

A student must be graduated on time which is vital to create colleges highly affordable. If not, the student's performance will not be able to satisfy the graduation norms fixed for degree program. Thus, it is essential to create a system that can persistently track the academic performance of the students so as to predict their succeeding performance like the probability of their graduation and estimated final Grade Point Average (GPA) [8]. To predict this, the study introduced a bi-layer structure consisting of an ensemble-predictor layer and base predictor layer. A Data Driven Course Clustering (DDCC) technique has also been developed that rely on the Probabilistic matrix Factorization (PMF) which provides automatic outputs of course clusters [9] relying on grade data of large, heterogeneous and sparse course of the student. Numerous simulations have been undertaken on a dataset consisting of 1169 undergraduate student's data that is congregated over a period of three-years at UCLA (University of California, Los Angeles). The results reveal that the introduced methodology has the ability to perform better than the existing techniques thereby conserving education interpretability. Yet, this study has to prolong the proposed system's prediction performance to elective courses as well as utilizing the prediction outcomes to suggest the students with courses. Several studies focused on various methods and factors to predict the undergraduate student's performance.

Accordingly, the paper [10] intended to explores the data mining capabilities in higher educational organizations to study the educational data. It exhibits the way in which data mining might assist in enhancing the process for decision making in universities. This paper aims to predict the academic performance of students by the end of 4 year bachelor's degree program for finding the efficient measures of students those are at risk in the initial days of their education [11]. It affords the organization with the essential evidence through which it could afford sufficient necessities for enhancing quality. Various classifiers have been implemented like Decision Trees with Information Gain (DT-IG), Naïve Bayes (NB), Random Forest with Information Gain (RF-IG), Random Forest with Accuracy (RF-ACC), Random Forest with Gini Index (RF-GI),Neural Networks (NN) and Decision-Tree with Gini-Index (DT-GI) [12]. The implementation results explore that DTs could be employed to find the courses acting as low performance indicator. Finding these courses will enable to give students early warning during their course of study. This study has to consider all other factors as well to enhance the prediction. Educational data mining (EDM) could be accomplished through numerous data mining methods like DT, Support Vector Machine (SVM),NB classifier and K-Nearest Neighbour (K-NN) [13]. Nevertheless, all these methods fall only under two main methods like clustering and classification. Though, there are exists various algorithms that assist in predicting undergraduate student's performance in academics by the classification strategy [14], numerous other algorithms achieve the same aim through the adaptation of the clustering strategy.

The study introduces a hybrid strategy that integrates classification and clustering algorithms for accomplishing high accuracy in prediction in comparison to the traditional methods. The accuracy of this system is found to be 0.75 when employed to extra [15], behavior and academic student dataset features. This is also found to outperform the traditional systems. However, the model has to be further extended for assisting huge kinds of student dataset features. The existing studies have also analyzed and compared the traditional Machine Learning (ML) techniques to examine the different input data types afforded to optimize the specific ML method [16] impacted the precision of predicting the student's exam performance [17]. For this reason, various student associated features have been taken into account to find the ideal combination of data types taken as input, for instance-past performance of the student, demographic data and engagement of the student. In addition, analysis has been undertaken to find the computational time required for predictions of individual methods. To assess and test, the Open University Learning Analytics (OULAD) dataset has been used as one of the many wide-ranging publicly accessible dataset with respect to data diversity associated with students and learning data. For classification purpose, SVM, K-NN, NB, DT, Logistic Regression (LR) and Artificial Neural Network (ANN) have been deployed. On the other hand, for regression, SVM, DT, LR, ANN, K-NN and Bayesian Regression (BR) have been analyzed. The maximum precision has been attained for ANN in classification. This also exhibited that the demographic data unexplored significant influence on precision.

Various classification models have been used in the article [18] to predict the performance of the student using the data gathered from an university in Australia. Data involves the enrolment details of the student and the generation of activity data from Learning Management System (LMS) of the university. Enrolment data comprise of the information of the student like attendance type (part time vs. full time), admission basis and socio-demographic features. Classification has been performed through four algorithms namely NB, J48 (a decision tree is generated comprising of three varied kinds of nodes – leaf nodes, internal nodes and root nodes), SMO (it utilizes an optimization algorithm to train the SVM) and JRip (a classification technique that relies on rule and produces clear rules through IF-THEN structure. Results showed no single technique exhibited high performance in all the considered aspects. Yet, the trees based and rule based techniques generate models of high interpretability which makes them highly valuable to design efficient student support. Moreover, the study planned for generating the student profiles through the use of unlabelled data for exhibiting the clusters of the interesting student as well as their characteristics which is yet to be implemented.

There exist various challenges for predicting the performance of undergraduate students in academics. A study [19] introduced Multi-Adaptive Neuro Fuzzy Inference System (MANFIS) and is employed to the Multi Input Multi Output (MIMO) student dataset for solving the issues associated with the prediction of student's academic performance. The introduced model has been empirically validated against MANFIS, Random Tree (RT), OneR and ANFIS (Adaptive Network based Fuzzy Inference System) in a benchmark dataset comprising of student performance from UCI (UC Irvine). From the analysis, it has been explored that the proposed system could be employed to MISO (Multiple Input Single Output) and MIMO datasets. The empirical study revealed the outstanding performance of proposed system than other related existing methods with respect to accuracy. This study intends to examine dynamic neuro fuzzy model as well as to analyses the convergence rate. Employing MANFIS-S to real issues in economy, education etc must also be undertaken. Few studies like [20] aimed to make use of the real internet utility data of students as a new measure to explore the probable association amongst the academic performance and internet behaviors [21]. Various ways will be found in which students are considerably varying amongst different performer's group. Subsequently, the study employed commonly employed ML algorithms to confirm the efficiency in predicting the student's academic performance from their internet data utility. Lastly, DT, SVM and Neural Network (NN) has been used as three general ML predictive algorithms to perform classification for making new features as well as predictive framework possess generalized value for predicting the performance of undergraduate students in academics. Though effective outcomes have been attained, the studies have to consider more feature to afford the educators with a comprehensive reference to develop the learning management of students.

Existing research [22] introduced a RTV-SVM (Reduced Training Vector-SVM) that has the ability to predict marginal and at risk students. The study provided more accuracy. Yet, other learning applications have to be found that is needed to differentiate the study status of students or students who fail to complete their degree on time, predicting which particular students have more drop out chance and finding the students those will have the ability to graduate through a narrow pass. Hence, research has to be carried out by considering all these parameters to improve the prediction accuracy. ANN and SVM has been utilized in the study [23] for finding the effective algorithm for predicting the student's academic performance. It has been found that SVM showed accurate outcomes. However, there exists no significant variation among these two algorithms and combination of them. Accuracy is also explored to be 82%. These studies have predicted the graduation rate of the students by relying only on the academic performance instead of socio-economic factors. Various other factors corresponding to success have to be taken into account in future research [24]. Predictive modelling through ML possess more potential for the development of early warning to find at risk students who have high chances of dropping and help those students in advance [25]. Random Forests (RFs) have been used in this study for predicting the at risk dropout chance students. Data utilized in this paper include 165, 715 samples from students of high school in the year 2014. Outcomes reveals that schools were able to find the students with high dropout risks with more accuracy by analyzing, recording as well as creating observation data or background information on students [26]. This study has a drawback. It considered only limited predictors or features. There exists numerous factors associated with the dropout rate of students that have to be taken into account [27]. Various factors were not present in the dataset that the study opted for [28]. Managing training and placement records in big organizations are complex as the count of students are more. In these situations, classification and differentiation on varied categories turn out to be tedious [29]. Hence the study aimed to analyses many datasets of students for classes that are pre-defined through the use of fuzzy logic and easily predict their performance. It will be well-being asset to institutions for information analysis of datasets and student information. Further research are needed to enhance the prediction system's performance [30].

3. Proposed System

The study predicts the placement performance of the student and affords mentoring system to students who are at risk of dropout analysis based on certain criteria. For this purpose, the study uses machine learning based techniques. Various processes are undertaken to achieve this as shown in fig. 1.

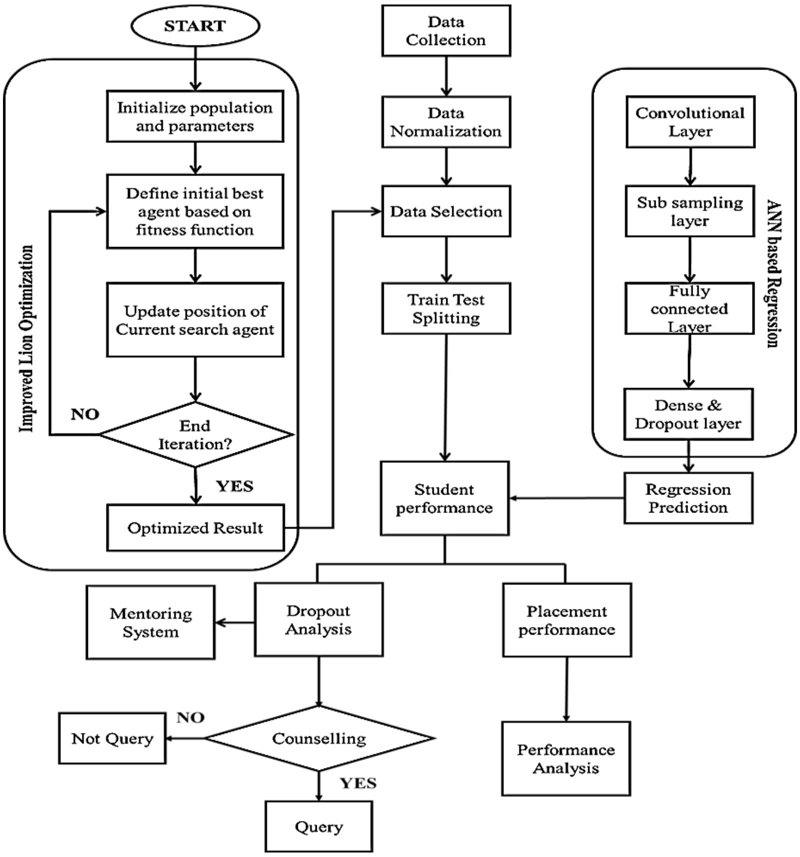

Fig. 1 - Overall view of proposed work

The process starts with population and parameter initialization. Subsequently, the first best agent is stated in accordance with fitness function to update the current search agent position. This iteration continues until an optimized result is achieved. Here, data selection is the significant factor needed to improve the prediction accuracy. The study proposed Improved Lion Optimization (ILO) algorithm to determine the best score so that only the relevant features can be selected. This is fed to the train and test split for student performance prediction. Placement performance along with domain of interest and mentoring system are the two main specificities focussed in the study. Prediction is achieved through Artificial Neural Network based Regression (ANN-R). Here the students who are at risk of dropout are considered for mentoring system to provide counselling. Not all students are provided counselling. Only students who have an acceptable aggregate mark, fees structure and attendance will be allowed to get counselling. Finally, the performance of the introduced system is analysed in terms of loss metrics. This analysis helps in finding the effective algorithm that has the ability to provide accurate prediction with minimum loss.

3.1 Improved Lion Optimization (ILO)

The proposed ILO undergoes various processes to find the best score and select only the relevant features. The stepwise process is shown as algorithm.
Algorithm I: ILO

STEP 1. Initialization Every isolated solution is termed lion
N_{var1} indicates the dimensional optimization issue
For LION $= [a_{i1}, a_{i2}, a_{i3} \dots \dots \dots, a_{N_{var1}}]$
IF Cost (fitness value) of each Lion Fitness value of lion $= f$ (lion) $= f(a_{i1}, a_{i2}, a_{i3} \dots \dots \dots, a_{N_{var1}})$
$n_{a_{pop1}}$ Solutions are generated randomly in search space
Remaining population will be randomly partitioned into p_{a1} prides
the formed population in the final step are the females and the remaining indicates males.
End IF
End FOR

STEP 2. Hunting OBL – Opposition Based Learning
For $A_i(a_{i1}, a_{i2}, a_{i3} \dots \dots \dots, a_{N_{var1}})$ be a point
If N_{var1} = dimensional space
Where $(a_{i1}, a_{i2}, a_{i3} \dots \dots \dots, a_{N_{var1}})$ are real number $a_{ii} \sum(f_{ii}, e_{ii})$ $ii = 1, 2, 3 \dots .. a_{N_{var1}}$
End IF
If oppositepart(A_i) is shown as $X_i(x_{i1}, x_{i2}, x_{i3} \dots \dots \dots, x_{N_{var1}})$
// $x_{ii} = c_{ii} + d_{ii} - x_{ii}$, $ii = 1, 2, 3 \dots . N_{var1}$ the principle of Opposition – Based Learning OBL
End IF
IF A dummy Prey $(prey_{11})$ is considered in centre of hunter $(prey_{11}$
$= \sum ht(a_{i1}, a_{i2}, a_{i3} \dots \dots \dots, a_{N_{var1}})$/number of hunter
$prey_{11}' = prey_{11} + rand(0,1) * p_{ai} * (prey_{11} - ht)$
$ht = (rand(2 * prey_{11} - ht) prey_{11}, (2 * prey_{11} - ht) < prey_{11})$
$ht = (rand(2 * prey_{11} - ht) prey_{11}, (2 * prey_{11} - ht) > prey_{11})$
Where $PREY_1$ current position of prey, Hunter is is currentposition hunter and Hunteris new position.
$ht' = (rand(prey_{11}, ht)\ ht < prey_{11})(rand(prey_{11}, ht)\ ht > prey_{11})$
Divide hunters into three sub group randomly

Generate $prey_{11}$

For $x = 1 : na_h$ (na_h denotes the hunter counts)

Move xth hunter in the direction of $prey_{11}$ in accordance with its related group

If ath hunter's new place is effective than its last positionn$prey_{11}$ escape from the corresponding hunter

END

END

END FOR

STEP 3. Moving Toward Safe Place
IF $LF' = LF + 2R * rand(0,1)\{T_i 1\} + U(-1,1) * \tan(\theta) * R_{ii} * \{T_i 2\}$
$\{T_i 1\}.\{T_i 2\} = 0 || \{T_i 2\ || = 0$

$$SB_i(j, k, p_{a0}) = \begin{cases} 1 & Best_{jp_a}^{k} < Best_{jp_a}^{k^{-1}} \\ 0 & Best_{jp_a}^{k} = Best_{jp_a}^{k^{-1}} \end{cases}$$

For $ii = 1$ to R_{np} (R_{np} is number of pride)
Compute tournament size for the ith pride
For $jj = 1$ to P_f (p_f is the number of lasting female in ith pride)
Choose an area among pride territory through tournament selection
Move the jth female towards the chosen place
End
End
END
END
END IF

STEP 4. Roaming
For $i = 1$ to Lr_m (Lr_m is count of resident_male)
Select $\%Lr_m$ of territory for random visit by ith male
For $j = 1$ to SB_j (S_l is count of chosen place in final stage)
Go towards place jth
If new position of ith male better compared to personal and best position visited
Marking which position as territory (best position visited is updated)
End
End
Choose best, visited position through its male current position
End

STEP 5. Mating
It is the parent's linear combination to produce two offsprings
IF
$$\text{offspring } 1 = \beta_i * Lf + \sum \frac{1-\beta_i}{\sum sB_i} * Ml * SB_{ij}$$
Else
$$\text{offspring } 2 = (1-\beta_i) * lf + \sum \frac{1-\beta}{\sum SB_i} * Ml * SB_{ij}$$
END
STEP 6. Defence
IF
Defence against the new matured_resident males.
ELSE
Defence against the nomad_males.
END

Various processes in accordance with the ILO are presented that includes initialization, hunting, moving towards safe place, roaming, mating and defence. The proposed ILO has the ability to handle multi-modal and non-linear issues. It can also find the nearest probable optimal solution. The convergence rate is also found to be fast as a result of roulette wheel selection and random walk mechanisms. Due to these advantages, the proposed ILO performs feature selection by considering only the relevant features.

3.2 Artificial Neural Network based Regression (ANN-R)

The study uses ANN-R for prediction of student performance which is the method of computing the value corresponding to the output variable relying on the values corresponding to input variable. In education, the resultant variables that indicate the undergraduate student's performance could be in the form of numerical values, decisions, marks and categorical values. The linear regression tasks partitions the techniques namely a simple LR (Linear Regression) methods of which aims to discover the correlation amongst one dependent and independent variable. In addition, a Multi-Linear Regression aims to discover the association between several independent variables and one dependant variable. MLP aims for modelling the association amongst two or numerous independent-variables in addition to dependant variable by fitting the Linear Equation (LE) to the actual data. Theory based assumption behind this method is that each unit change in the independent variable results in a uniform alteration in the variable that is dependant. The architecture of ANN is shown in fig. 2.

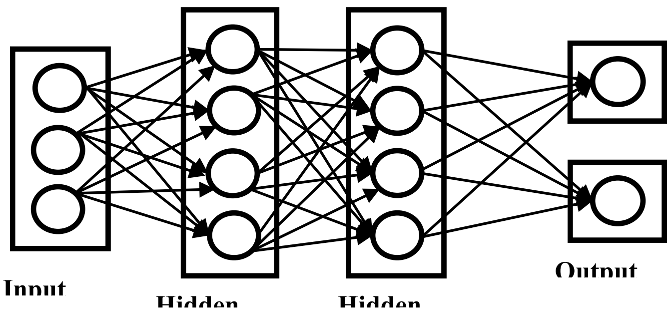

Fig. 2 - ANN-Architecture

In addition, the pseudo code corresponding to the proposed ANN-R is shown in algorithm II. Various process involved in predicting the undergraduate student's performance are explored in algorithm II.

Algorithm II: ANN-R

```
 Step-1:  Initialization (%)
          Initialize ANN-R
          Step-2: Set ω and €
          Step-3: for each Fully Connected (FC) bi-partite later (ANN) do
          Step-4: Alter FC with dropout and dense connected layer possessing
topology
End
Step-5: Initialize parameters (training algorithm)
          Step-6: Training (%)
                    for each (epoch training) do
                     Step-7: Execute standard-training method
                     Step-8: update weight
                     Step-9: for each ANN layer do
                              eliminate a portion ω of the reduced positive-weight
                              eliminate a portion ω of the reduced negative-weight
                     Step-10: if (e)≠ last epoch-training do
                    Step-11: Include new weights in a random way in an identical way
          that are detached
                               randomly
                    Step-12: Analyse the performance of the student
                    End
          End
End
```

The architecture of ANN is presented that comprise of input layer, hidden layers and the output layer. The proposed ANN-R has various advantages like ability to operate with in suffice knowledge, efficient fault tolerance [31], parallel processing, distributed memory etc. These advantages make the proposed ANN-R to be highly suitable for predicting the placement performance and domain of interest of the students from the dataset collected. Moreover, dynamic dataset is used in this study which makes it flexible to be employed to predict the student's performance from other colleges. All these merits enable the proposed system to be efficient and it is proved to be outstanding than the existing studies in the comparative analysis.

4. Results and Discussion

The proposed methodologies are implemented for predicting undergraduate student's placement performance and afford counselling to students in accordance with the dropout analysis. In this section, results attained by this implementation are presented. Analysis of the introduced system is carried out by internal comparison with the traditional methods in accordance with loss metrics namely Mean Absolute Error (MAE), Mean Square Error (MSE) and Root Mean Square Error (RMSE). This analysis is carried out as it is significant in exploring the efficacy of the proposed method.

4.1 Dataset Description

The study uses real-time dataset. Dataset encompassing of computer science student's data is collected from a college in Karnataka named "Global Academy of Technology". It included the passing year and supplementary details like aggregate marks, plans after graduation, count of active backlogs, area of interest, core competencies, fascination in learning new things, number of internships that are completed and attended workshops, hours spent in attending the placement training etc. This study uses dynamic dataset by efficient pre-processing. Hence, flexibility is achieved.

4.2 Performance Analysis

The proposed system is analyzed through comparison with the traditional CNN (Convolutional Neural Network), Google net and SVM in accordance with Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Square Error (MSE). The equations corresponding to these loss metrics are discussed here.

4.2.1 Mean Square Error (MSE)

It indicates the similarity of regression lines to a collection of points and is given by equation (1).

$$\mathbf{MSE} = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2 \qquad (1)$$

In equation (1),

MSE indicates the Mean Square Error

N is data point count

Y_i represents the observed-values

$\hat{Y}_i$ is predicted values

4.2.2 Root Mean Square Error (RMSE)

It is the ideal metric that measures the accuracy of the system. It compares only the varied model's prediction errors or model configurations for a particular variable, however not between variables which is due to scale reliance and is provided in equation (2).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \hat{x}_i)^2}{N}} \qquad (2)$$

In equation (2),

RMSE is the Root Mean Square Error

i represents variable

N denotes the non-missing data-points

x_i is the original time series observation

$\hat{x}_i$ represents the computed time series

4.2.3 Mean Absolute Error (MAE)

It a model's evaluation metric used with regression models. It is the mean of actual values analogous to the errors in individual prediction over each and every test set cases and is given by equation (3).

$$MAE = \frac{\sum_{i=1}^{N}|y_i - x_i|}{n} \qquad (3)$$

In equation (3),

MAE is the Mean Absolute Error

x_i is the true value

y_i is the prediction

N represents the overall data point count

The introduced system is analysed by comparing with conventional system encompassing ANN, Google net and SVM in accordance with the above specified loss metrics. The obtained results are presented in table 1.

Table 1 - Analysis of the proposed system for student's placement prediction

Algorithm	Root Mean Squared Error	Mean Absolute Error	Mean Squared Error
ANN-R	1.075675758	0.699713875	1.007078336
CNN	1.119261567	0.731323047	1.252746454
SVM	1.159129382	0.719221897	1.343580925
GoogleNet	1.126078527	0.771612016	1.268052848

From table 1, it is clear that the introduced ANN-R shows minimum error rate (MAE-0.699, RMSE-1.075 and MSE-1.007) than the conventional CNN, Google net and SVM. The existing studies have used algorithms like CNN, Google net and SVM for predicting undergraduate student's placement performance. Though minimum error rate is achieved by all these methods, the proposed system shows minimum error in comparison to the conventional algorithms proving its outstanding efficacy to predict the undergraduate student's placement performance. It is graphically shown in fig. 3.

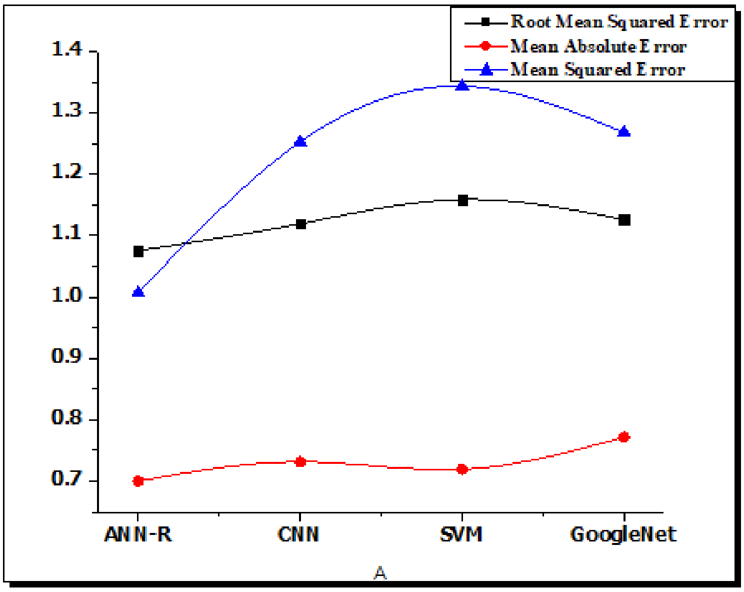

Fig. 3 - Comparative analysis of the traditional and proposed methods in predicting student's placement performance

From fig. 3, it can be clearly seen that the introduced system explores minimum error than existing methods. Proposed ANN based regression has the ability for fault tolerance and thus it minimizes the error in predicting student's placement performance thereby improving its accuracy in this prediction. In addition, proposed system is analysed by comparison with the traditional methods to find its efficacy in predicting the domain of interest of the students. The obtained results are presented in table 2.

Table 2 - Analysis of the proposed system for student's domain of interest prediction

Algorithm	Mean Squared Error	Mean Absolute Error	Root Mean Squared Error
ANN-R	0.003597626	0.043381634	0.059980214
CNN	0.038344631	0.14633753	0.195817852
SVM	0.013742101	0.054971786	0.117226706
GoogleNet	0.025684016	0.511016683	0.160262335

This analysis also shows that the minimum error rate (MAE, RMSE and MSE) of the introduced ANN-R than the existing SVM, Google net and CNN. The existing studies have used algorithms like CNN, Google net and SVM for predicting the student's domain of interest. Though minimum error rate is achieved by all existing methods, the proposed system shows minimum error rate than the traditional algorithms proving its outstanding performance in predicting the domain of interest of the students. It is graphically shown in fig. 4.

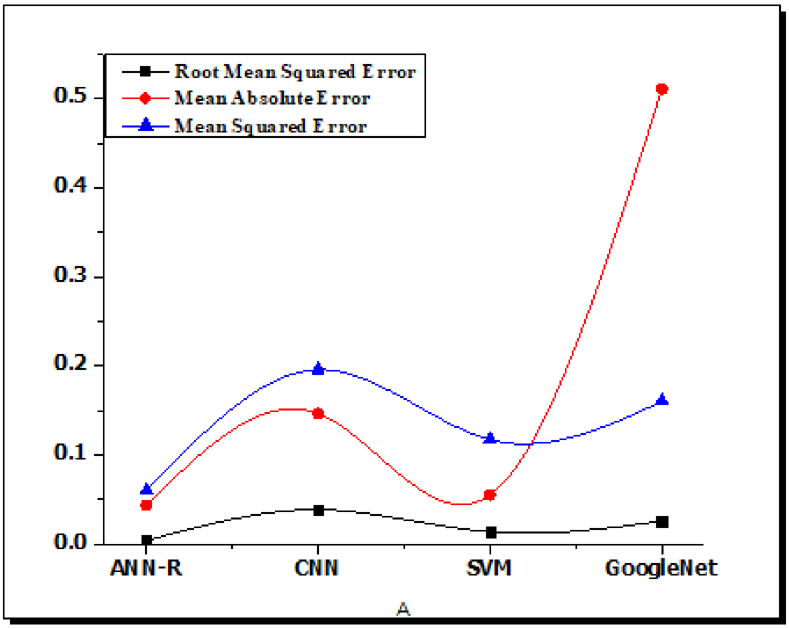

Fig. 4 - Comparative analysis of the traditional and introduced methods for predicting the domain of interest of the students

From fig. 4, it can be clearly seen that the introduced system explores minimum error compared to existing methods. The proposed ANN based regression is advantageous as it is able to tolerate faults and thus it minimizes the error in predicting the student's domain of interest thereby improving its accuracy in this prediction. Thus, from the above two analysis it can be concluded and confirmed that the proposed system is effective than the existing methods in predicting the student's placement performance and domain of interest.

4.3 Comparative Analysis

The proposed system is internally compared with the existing system during the implementation. The traditional SVM, Google net and CNN are considered for the analysis with the introduced ANN-R. The obtained outcomes are given in this section. At first the results of the proposed ANN based regression is shown in fig. 5. It can be clearly seen that the actual and predicted values correlate with one another showing the degree to which correct prediction is made.

On the other hand, the existing CNN, SVM and Google net is analysed and the results are presented in fig. 6, fig. 7 and fig. 8. From fig. 6, it can be clearly seen that the actual and predicted values have slight variations and few do not correlate with one another. This confirms that there is slight misinterpretation in predicting the performance of the students using existing CNN.

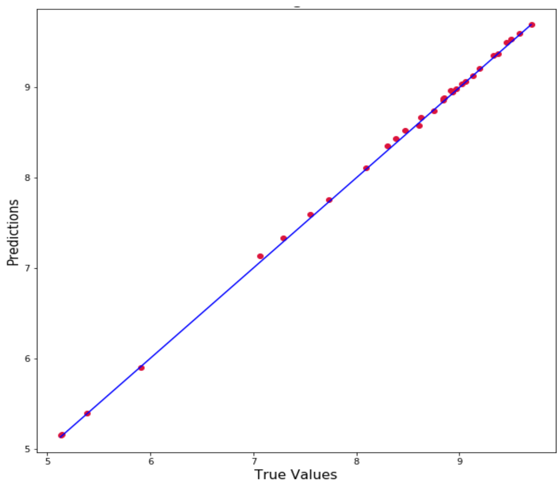

Fig. 5 - Proposed ANN based regression

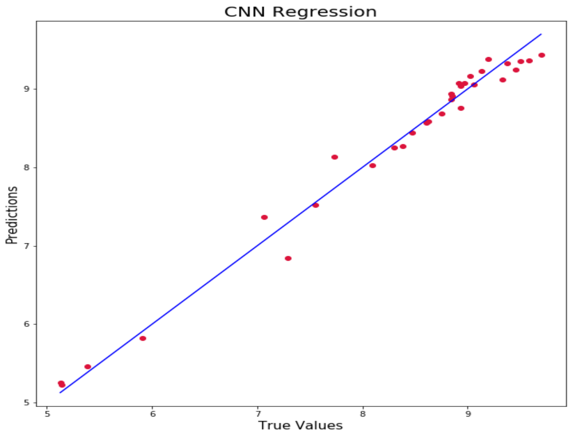

Fig. 6 - CNN regression

The figure 7 shows the regression line of existing SVM. It can be seen that the actual and predicted values do not correlate with one another except few. This reveals its ineffectiveness in predicting the student's performance. Similarly, figure.8 shows the regression line of Google net. The actual and predicted values correlate with one another. Still, there exist few variations that prove the ineffectiveness to some extent.

Fig. 7 - SVM regression

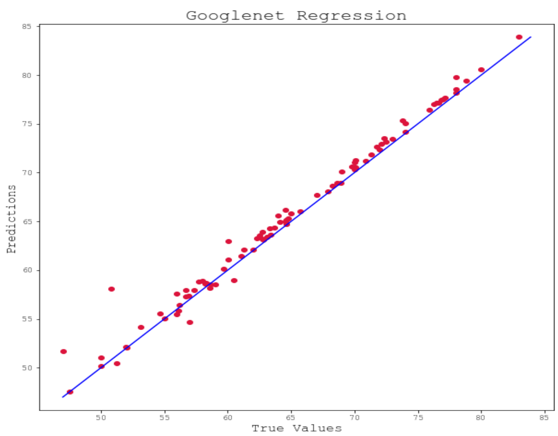

Fig. 8 - Google net regression

Thus, the comparative analysis reveals that the proposed ANN-R shows high degree of correlation with the actual and predicted values. Whereas, the traditional Google net, SVM and CNN shows specific variations in the actual and predicted values. This shows the efficiency and suitability of the proposed system in correct prediction of the placement performance of the students and domain of interest than the existing system.

5. Conclusion

The study uses machine learning based methods for predicting the dropout analysis and placement performance of the undergraduate students in the early years of degree program. Dropout analysis is carried out so that mentoring system can be provided to students who are under specific threshold with acceptable profile (aggregate mark, fees structure and attendance). Various processes are involved to accomplish this. At first, relevant features are selected from the real-time dataset based on best score through the introduced ILO approach. Then, the placement performances of the undergraduate students are predicted in addition to the domain of interest of students encompassing designing, developing, marketing and testing. ANN-R was utilized in predicting the undergraduate student’s performance in academics and analysed with respect to loss metrics. In addition, the proposed system was compared with the existing systems to predict the effective algorithm that predicts the value similar to actual value. Here, comparison was carried out during the implementation phase and the obtained results are presented revealing that the proposed system shows high correlation of actual and predicted values with minimum error than the existing methods. This study used data collected from Global Academy of Technology. The study can be applied to other college data to predict the student performance as it used dynamic dataset.

References

[1] A. A. Saa, M. Al-Emran, and K. Shaalan, "Mining student information system records to predict students’ academic performance," in International conference on advanced machine learning technologies and applications, 2019, pp. 229-239.

[2] A. I. Adekitan and O. Salau, "The impact of engineering students' performance in the first three years on their graduation result using educational data mining," Heliyon, vol. 5, p. e01250, 2019.

[3] V. L. Miguéis, A. Freitas, P. J. Garcia, and A. Silva, "Early segmentation of students according to their academic performance: A predictive modelling approach," Decision Support Systems, vol. 115, pp. 36-51, 2018.

[4] E. Çakıt and M. Dağdeviren, "Predicting the percentage of student placement: A comparative study of machine learning algorithms," Education and Information Technologies, pp. 1-26, 2021.

[5] P. Manvitha and N. Swaroopa, "Campus placement prediction using supervised machine learning techniques," Int J Appl Eng Res, vol. 14, pp. 2188-2191, 2019.

[6] R. Ishizue, K. Sakamoto, H. Washizaki, and Y. Fukazawa, "Student placement and skill ranking predictors for programming classes using class attitude, psychological scales, and code metrics," Research and practice in technology enhanced learning, vol. 13, pp. 1-20, 2018.

[7] A. I. Adekitan and E. Noma-Osaghae, "Data mining approach to predicting the performance of first year student in a university using the admission requirements," Education and Information Technologies, vol. 24, pp. 1527-1543, 2019.

[8] J. Xu, K. H. Moon, and M. Van Der Schaar, "A machine learning approach for tracking and predicting student performance in degree programs," IEEE Journal of Selected Topics in Signal Processing, vol. 11, pp. 742-753, 2017.

[9] M. VeeraManickam, M. Mohanapriya, B. K. Pandey, S. Akhade, S. Kale, R. Patil, et al., "Map-reduce framework based cluster architecture for academic student’s performance prediction using cumulative dragonfly based neural network," Cluster Computing, vol. 22, pp. 1259-1275, 2019.

[10] R. Asif, S. Hina, and S. I. Haque, "Predicting student academic performance using data mining methods," International Journal of computer science and network security, vol. 17, pp. 187-191, 2017.

[11] J.-Y. Wu, Y.-C. Hsiao, and M.-W. Nian, "Using supervised machine learning on large-scale online forums to classify course-related Facebook messages in predicting learning achievement within the personal learning environment," Interactive Learning Environments, vol. 28, pp. 65-80, 2020.

[12] A. Mondal and J. Mukherjee, "An Approach to predict a student’s academic performance using Recurrent Neural Network (RNN)," Int. J. Comput. Appl, vol. 181, pp. 1-5, 2018.

[13] B. K. Francis and S. S. Babu, "Predicting academic performance of students using a hybrid data mining approach," Journal of medical systems, vol. 43, pp. 1-15, 2019.

[14] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," Applied Soft Computing, vol. 97, p. 105524, 2020.

[15] M. Zaffar, K. Savita, M. A. Hashmani, and S. S. H. Rizvi, "A study of feature selection algorithms for predicting students academic performance," Int. J. Adv. Comput. Sci. Appl, vol. 9, pp. 541-549, 2018.

[16] S. Ranjeeth, T. Latchoumi, and P. V. Paul, "Optimal stochastic gradient descent with multilayer perceptron based student's academic performance prediction model," in Recent Advances in Computer Science and Communications. vol. 13, ed, 2020.
[17] N. Tomasevic, N. Gvozdenovic, and S. Vranes, "An overview and comparison of supervised data mining techniques for student exam performance prediction," Computers & education, vol. 143, p. 103676, 2020.
[18] S. Helal, J. Li, L. Liu, E. Ebrahimie, S. Dawson, D. J. Murray, et al., "Predicting academic performance by considering student heterogeneity," Knowledge-Based Systems, vol. 161, pp. 134-146, 2018.
[19] H. Fujita, "Neural-fuzzy with representative sets for prediction of student performance," Applied Intelligence, vol. 49, pp. 172-187, 2019.
[20] X. Xu, J. Wang, H. Peng, and R. Wu, "Prediction of academic performance associated with internet usage behaviors using machine learning algorithms," Computers in Human Behavior, vol. 98, pp. 166-173, 2019.
[21] I. E. Livieris, K. Drakopoulou, V. T. Tampakas, T. A. Mikropoulos, and P. Pintelas, "Predicting secondary school students' performance utilizing a semi-supervised learning approach," Journal of educational computing research, vol. 57, pp. 448-470, 2019.
[22] K. T. Chui, D. C. L. Fung, M. D. Lytras, and T. M. Lam, "Predicting at-risk university students in a virtual learning environment via a machine learning algorithm," Computers in Human Behavior, vol. 107, p. 105584, 2020.
[23] Y. Nieto, V. García-Díaz, C. Montenegro, and R. G. Crespo, "Supporting academic decision making at higher educational institutions using machine learning-based algorithms," Soft Computing, vol. 23, pp. 4145-4153, 2019.
[24] M. V. Ananthi and M. K. Mythili, "Student Performance Prediction in Higher Education Using Lion–Wolf Optimization Algorithm."
[25] J. Y. Chung and S. Lee, "Dropout early warning systems for high school students using machine learning," Children and Youth Services Review, vol. 96, pp. 346-353, 2019.
[26] S. Sultana, S. Khan, and M. A. Abbas, "Predicting performance of electrical engineering students using cognitive and non-cognitive features for identification of potential dropouts," International Journal of Electrical Engineering Education, vol. 54, pp. 105-118, 2017.
[27] A. Cohen, "Analysis of student activity in web-supported courses as a tool for predicting dropout," Educational Technology Research and Development, vol. 65, pp. 1285-1304, 2017.
[28] C. C. Gray and D. Perkins, "Utilizing early engagement and machine learning to predict student outcomes," Computers & Education, vol. 131, pp. 22-32, 2019.
[29] R. K. Rathore and J. Jayanthi, "Student prediction system for placement training using fuzzy inference system," ICTACT Journal on Soft Computing, vol. 7, pp. 1473-1446, 2017.
[30] C. Burgos, M. L. Campanario, D. de la Peña, J. A. Lara, D. Lizcano, and M. A. Martínez, "Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout," Computers & Electrical Engineering, vol. 66, pp. 541-556, 2018.
[31] Karthick Raghunath K.M, Rengarajan N, Response time optimization with enhanced fault-tolerant wireless sensor network design for on-board rapid transit applications, "Cluster Computing: The Journal of Networks, Software Tools, and Applications", Vol. 22, No. July-S4, pp. 9737 – 9753, 2019.