International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-2, July 2019

Deep Neural Network for Multi-Class Prediction of Student Performance in Educational Data

V. Vijayalakshmi, K. Venkatachalapathy

Abstract*: *Prediction of student performance is the significant part in processing the educational data. Machine learning algorithms are leading the role in this process. Deep learning is one of the important concepts of machine learning algorithm. In this paper, we applied the deep learning technique for prediction of the academic excellence of the students using R Programming. Keras and Tensorflow libraries utilized for making the model using neural network on the Kaggle dataset. The data is separated into testing data training data set. Plot the neural network model using neuralnet method and created the Deep Learning model using two hidden layers using ReLu activation function and one output layer using softmax activation function. After fine tuning process until the stable changes; this model produced accuracy as 85%.

Keywords*: *Deep Neural Network, Educational data, Machine Learning, Performance Prediction, R Programming.

I. INTRODUCTION

To anticipate how students may perform in the midst of their learning technique is a staggering errand paying little respect to ceaseless increment of information in the data storage connecting to student's scholastics in foundations of higher learning. As demonstrated by [1], the academic administration systems are not organized truly to assist enlightening directors with examining which students are in peril of dropping out of school or college. The educational field has huge amount of instructive data. This data contains the information about the instructor, student, graduated class details, materials, etc. Educational information mining is utilized to discover the examples in this data of decision-making process. There are two sorts of Educational framework: Traditional Education structure and Web based learning system. For higher educational establishments their ambition is to add to the enhancement of nature of advanced education, the accomplishment of generation of human assets is the subject of a subsequent analysis. Along these lines, the desire for students' success is significant for advanced education organizations, in light of the way that the idea of arranging the work is the ability to address learners' issues.

Data Mining (DM) is the path toward finding intriguing patterns and knowledge from a great deal of data [2]. EDM (Educational Data Mining) is one of the uses of Data Mining frameworks on instructive data. The purpose of the EDM is to separate the information and to decide research issues of educational concepts. It is also oversees extending new methodologies to research the instructive information, and make use of data mining systems to all the more promptly grasp understudy learning condition [3, 4]. The EDM system changes over natural data starting from informative structures into important information that could enormously influence the practical and research issues in educational data. EDM authorities think on an arrangement of regions that comprises of a single learner learning from educational programming, PC reinforced network learning, PC versatile testing and also the components that are connected with understudy disappointment or without support in courses [5, 6]. Educational Data Mining applies various methods, for instance, K-Nearest Neighbor (KNN), Naïve Bayes (NB), Neural Networks (NN), Decision Trees (DT) and numerous others. Forecast and investigation of learner execution is a basic perspective in instructive condition. Their scholastic execution is a basic feature in constructing the future. Scholastic execution of learner isn't a outcome of just a single choosing highlight other than it vigorously depends on different components like environmental variables, psychological factors, socio-economic, personal information and others. Machine learning algorithms for instance, DT and NB is especially used in Educational Data Mining. There is a confinement to such calculations, as communicated by Havan Agrawal [7] when input is given in a continuous range to Bayesian classification the accuracy of the models decreases. Such gathering works better with discrete information. Furthermore communicated, that a Neural Network beats with the steady information well. The connection between Machine Learning and Deep learning is delineated in Fig.1.

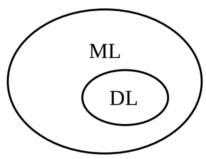

Fig. 1. Machine learning (ML) and Deep learning (DL).

Machine learning is the superset of Deep Learning. It is considered as the best in class instrument for artificial intelligence research which connected in different applications. Deep Learning can be named as: DNN (Deep Neural Network), RNN (Recurrent Neural Network), CNN (Convolutional Neural Network) and Q-learning.

Revised Manuscript Received on 30 July 2019.
***** Correspondence Author
V. Vijayalakshmi*, Computer Science and Engineering, Annamalai University, Chidambaram, India, Email: vivenan09@gmail.com.
K. Venkatachalapathy, Division of Computer and Information Science, Annamalai University, Chidambaram, India.

The deep learning has been of lately utilized for voice/sound recognition, Natural Language Processing, computer vision [8].

The summary of this paper is composed as pursues: Section II presents a review of literature utilized in this study.

Section III gives data collection process and some insights to the structure of data and experimental matrices. Section IV shows architecture diagram of the model. Section V contains data preprocessing steps. Section VI includes the methodology used for implementation work. Section VII includes the implementation and results and Section VIII concludes the study.

II. LITERATURE REVIEW

Neural network is utilized by numerous analysts for prediction of student performance with different dataset and number of factors. Table I plainly portrays the existing system which worked with artificial neural network with all the details such as author name, published year, sample data size they have taken, number of input variables and output variable finally with the result.

Herzog [9] developed neural networks and decision trees, attempted to assess retention rate and time of degree completion of students. As shown by the creator, the capacity for perceiving the students at danger of falling out or those are to set aside an extremely extended effort to modify facilitate to guide interference programs, there they are required for the most part and provides methods for better selection, rate of graduation, and accuracy of estimating educational cost pay. His examination thinks about the prediction of precision of three types of Artificial Neural Networks (ANN) and Decision Trees with the use of Multinomial Logistic Regression. The prediction of retention was relied upon the 2nd year enlistment with 8,018 first year recruits, and the dataset utilized for the TTD (Time to Degree) complete contained 15,457 of records. There are forty indicators were utilized to evaluate the retention rate and seventy nine components were joined into the perplexing forecasts of TTD. In the wake of notwithstanding transfer of students and the C5.0 algorithm were utilized, then the creator achieved an accuracy of 93%.

Vandamme et al. [10] drove investigate went for first year students performance of academic data prediction. This paper's purpose was to arrange learners into 3 events: the first is a high probability of accomplishment for low risk students, second is a student with medium risk they thrive if the university takes appropriate measures, and the third is high risk students who have a high probability of getting fail or dropping out. This work attempted to describe the learners into the 3 groups, before writing the test for first year; this will be made easier to support them. The exploration test included the count of 533 students. The Classification algorithm Iterative Dichotomiser 3 (ID3) was picked with five input factors. The accurate rate of classification was 40.63%, unequivocally, high risk had 48.65%, medium risk had 18.46% and the low risk students had 60.34%. The weekly course cooperation was the most important component by the learners and they had a good choice of selecting the specific and better university.

Ibrahim and Rusli [11] took a gander at an (ANN) Artificial Neural Network system, Linear Regression and Decision Tree for predicting the students of their performance of academic details. The Cumulative Grade Point Average (CGPA) was utilized to assess the graduation of academic performance of learners. The CGPA and the profile of demographic details of students for the essential semester considered as factor variable of undergraduate students for the academic performance. This investigation of research showed that all the three models produced more than 80% as outcome. Nghe et al. [12] investigated 2 important information mining strategies, the Bayesian Network and the Decision Tree system. They utilized Grade Point Average (GPA) and the academic records of second year students as far as possible for predicting the third year performance. The Bayesian Network produced a precision of 90.27% and the Decision Trees produced a precision of 94.03%.

Cortez et al. [13] introduced work that means the learner accomplishment in the secondary education utilizing Data Mining procedures. The currently used information like statistic, grades of students, school and social associated highlights were accumulated by utilizing questionnaires and records. They utilized 4 data mining models, for example, Support Vector Machines, Neural Networks, Random Forest and Decision Trees with and also without past grades were tested. The aftereffect of this investigation, most effective tool for prediction of learners was created; the nature of education was improved and upgrading the resource organization of schools.

Zekić-Sušac et al. [14], displayed the accomplishment of student prediction were planned using Neural Networks and Decision Trees. This work was guided for the students from the second year to the final year. The data collected from 165 learners. They considered input variables as test grade, attendance of students, number of test taken by students, study materials, time spending to learn, scholarship, gender, and participation during lecture and the output variables as past academic year's average grade communicated as 2 classes, that is not actually or proportionate to three and more prominent than three. The Neural Network system produced the low precision rate as 66.26% and the measurable T test demonstrated an authentically colossal differentiation. The Decision Tree produced better classification precision rate as 88.36%. The investigation of the importance of input factors exhibited that the time focused on thinking about was the mainly noteworthy.

Delen [15] made a systematic model for predicting and explaining the motivation of the wearing out of learners, utilized educational information of five years with a couple of information mining strategies, for instance, Logistic Regression, Support Vector Machines, Decision Trees and the Neural Network system. The affectability examination of the system discovered that cash related factors in educational data were along with the mainly basic forecaster. In perspective on results of sample, the accurate precision rate of 81.18% produced by support vector, which was trailed by Logistic Regression, Neural Network, and the Decision Trees. In light of the referenced analysis, it ends up clear that different creators utilized many data mining techniques to anticipate the achievement of students. When in doubt, Decision Trees produced the better rate for grouping the students effectively. In for all intents and purposes all examinations, output variable were grade averages, communicated in any event of two classes.

Retrieval Number: B2155078219/19©BEIESP
DOI: 10.35940/ijrte.B2155.078219
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

In like manner, CART (Classification and Regression Tree) and the C4.5 (J48) algorithms was more generally utilized decision tree algorithm in past work. Consequently, the authors utilized decision trees, in the data mining tool called Weka by looking at the accomplishment of certain algorithms accessible.

Osmanbegović and Suljić [16] took at various information mining techniques and methodologies for predicting the achievement of learners, by giving the overall data gathered from students of first year and the information about enlistment. The 257 records considered as sample dataset. The achievement of student was the output variable which was purely relied upon the evaluations in the course of Business Informatics. The Input variable of the model comprised of 12 factors. The NB (Naive Bayes) produced a predominant prediction with better precision of 76.65% than other algorithms such as, MLP (Multilayer Perceptron) and Decision Tree (J48).

Dorina Kabakchieva [17] concentrated on the progression of data mining methods for anticipating the performance of students, in light of their own, before university and during university performance. This work utilized the dataset which gathered and consolidated the data about the students for consecutive three years joined in the university. A few of without a doubt comprehended classification algorithms of data mining, including Nearest Neighbor classifier, Neural Network system, Decision Tree classifier and Rule Learner are associated on the sample. The presentations of the calculations are explored.

Kovačić [18] attempted to anticipate achievement of students by extracting enrolment information. The 450 students joined in the Information System course was gathered as the dataset. He utilized logistic regression and decision trees. The highest percentage of classification rate 60.5% produced by CART which is amongst the decision tree developing techniques.

Ramesh V, Parkavi P, Ramar K [19] displayed examination, a review with trial method was grasped to make a data storage and it was worked from a fundamental and an auxiliary data. They got results from hypothesis testing expose that sort of school isn't impact performance of student and guardians' job expect a vital activity in anticipating grade. This research may assist the education related fields with recognizing the learners who are in danger and to give best extra preparation to the slow learners.

Simeunović and Preradović [20] conceived a model for anticipating the learner excellence by utilizing information extraction. This model made use of the learner behavioral information, socio demographic information, frames of mind towards learning, character attributes, and also the whole teaching process concern, will in general group learners into two classifications of achievement. The performance was assessed utilizing the learner Grade Point Average (GPA) attained through the range of students. They attempted three information mining techniques; Logistic Regression, Neural Network and Decision Tree. The decision tree showed a precision of 71.25%, logistic regression produced 74.8%, and the neural network systems produced with high precision rate as76.4%. Cheewaprakobkit [21] developed a model, with decision trees and neural network to group students rely upon their achievement of academic data. The sample included 1,600 learners' data with 22 variables of learners chosen in the scope from 2001 and 2011 at a Thailand University. The prediction accuracy was calculated using Cross Validation

Table- I: Literature Survey

S. No.	Author Name	Sample Size	No of Input variables	Output variable	Result
1	Herzog	8,018	40	Student retention	85%
2	Vandamme et al.	533	25	Average mark	51.88%
3	Ibrahim and Rusli	206	4	CGPA	Average squared error 0.1714
4	Nghe et al.	20,492	14	GPA	90.27%
5	Cortez and Silva	788	29	Student Class	91%
6	Zekić Sušac et al.	165	8	Grade Average	66.26%
7	Delen	7,018	39	Second Fall Registered	79.85%
8	Osmanbegović and Suljić	257	12	Success in the course	71.20%
9	Dorina Kabakchieva	10067	14	Student Class	73.59%
10	Kovačić	453	9	Study outcome	59.40%
11	V.Ramesh, P.Parkavi, K.Ramar	500	10	HSCGRADE	72.38%
12	Simeunović and Preradović	354	17	GPA	76.40%
13	Cheewaprakobkit	1,600	20	CGPA	83.88%
14	Parneet Kaura et al.	152	14	Student Class	75%

technique with 10. The decision tree classifier produced with high precision of 85.188% which shown that decision tree generate better accuracy than the neural network system. Parneet Kaura [22] centered perceiving the moderate students among learners and appearing by a judicious data mining technique which utilizing algorithms based on grouping concept. The high school student's information's were collected and filtered using one of the best data mining tools called WEKA. The academic information about the students were checked and passed to numerous characterization techniques for instance, REPTree, J48, NB, SMO, and MLP. Thusly, bits of knowledge are delivered reliant on every classification techniques and relationship of all 5 classifiers were done to calculate the exactness and to identify the better performing classification technique from all methods. This work showed the criticalness of Classification and Prediction support of information mining algorithms in educational sectors. Hongsuk Hongsuk, et al. [23] proposed a DNN that is supervised method to assess interface based stream of conditions of traffic. A TPI (Traffic Performance Index) was utilized to observe a blocked traffic to a nonstop the progress of traffic condition. With a 3 layer it had show the capacity to evaluate the blockage with a 99% of precision.

III. DATASET

The wellspring of data for building the proposed deep neural system to anticipate the students' performance is gained from https://www.kaggle.com/aljarah/xAPI-Edu-Data. It is an educational dataset accumulated from learning the board structure called kalboard 360. The instructive accumulation expands into 500 students with 17 features. The highlights are assembled into three essential classes: the first is Demographic features, for example, nationality and sexual orientation.

The second is Academic background features; for example, grade Level, educational Stage, and section. The third is Behavioral highlights, for example, parent answering Survey, raised hand on class, visited resources and Parent School Satisfaction. The dataset subtleties, for example, Name of the features, category, data type, number of values and description are in Table II. The dataset comprises 480 records.

Table- II: Data set

Features Category	Name	No of Values	Description
Demographical Features	Gender	2	female or male
	Nationality	14	Nationality of student
	Place of Birth	14	Birth place for the student
	ParentResponsible	2	father or mom
Academic Background Features	Stages	3	primary, middle and high school levels
	Grades	12	Grade of the student
	SectionID	3	A, B, C
	Topic	12	Course topic
	Student Absent day	2	Above-7, Under-7
	Semester	2	First or second
Behavioral Features	Raised hand	0-100	Student Behavior during communication with e-learning system
	Visited Resource	0-100	
	Viewing Announcement	0-100	
	Discussion Group	0-100	
Parents Participation on learning	Parent Answering	2	Yes, No
	Parent Satisfaction	2	Good, bad

The 16 features are independent variables and one feature is dependent variable with three classes such as low, middle and high.

IV. ARCHITECTURE

The architecture of out proposed system comprised of three phases as follows:

i) Preparing the data (data cleaning, data transformation, feature selection and data visualization).

ii) Building the neural network model (Installing Keras library, Create the DNN model, Compile and fit the model, Fine tune the model).

iii) Measure the performance (accuracy) of the system.

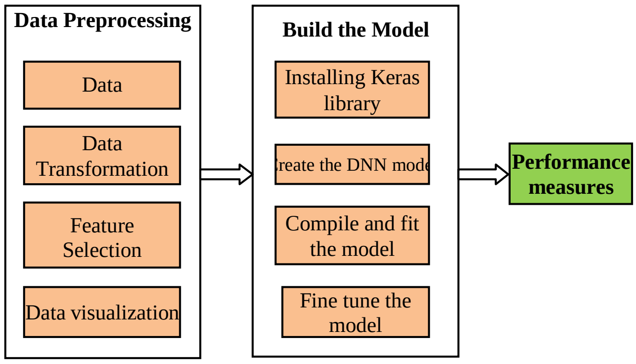

Fig. 2.Architecture of proposed system.

Architecture diagram is depicted in fig. 2. Data Processing module contains data leaning, data transformation, feature selection and data visualization. Building the model has four steps that is installing the keras packages, creating the DNN model, compiling the model and finally fine tune the model. The last stage is performance measures.

V. DATA PREPROCESSING

Information preprocessing is the critical advance before applying information mining algorithm; it changes the unique information into a proper data which is to be utilized by specific mining technique. The data preprocessing incorporates diverse assignments like data cleaning, feature selection and data transformation [24].

Data visualization is an important preprocessing errand, which used graphical representation to disentangle and comprehend complex information. Visualization strategies are used to picture web based learning perspectives. This research envisions the collected sample data with Rstudio tool. The values of attribute gender are visualized with 305 males and 175 females, which is in Fig. 3.

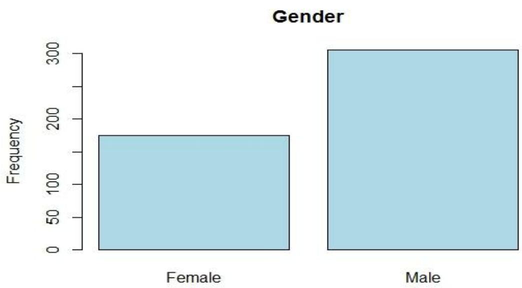

Fig. 3. Gender Feature Visualization.

The data is separated into 70% as training data (337 instances) and 30% as testing data (143 instances). Data cleaning is a standout amongst the most essential task of preprocessing, which is connected on the sample data to evacuate insignificant things and omitted values.

The sample data includes twenty missing values in different features from five hundred records; the missing qualities records are expelled from the informational collection, and the data set subsequent to cleaning winds up 480 records. There are two packages imported such as Caret and mlbench then we positioned the features of Importance by utilizing linear model [25].

The best rank feature is StudentAbsenceDays. The most exceedingly terrible feature is GradeID.

There are 12 features are high position and 4 features are least position and the variable importance is appeared in Table III. The correlation matrix is determined to obtain the redundant attributes. At that point the profoundly related qualities are recognized and evacuated. For the most part, we should expel features with a complete correlation with 0.75 or the value is high. The plot of the variable importance is in fig. 4 which obviously demonstrates that the most astounding position feature is StudentAbsenceDays and the least rank position include is GradeId.

Table- III: Variable importance

Variable	Importance
StudentAbsenceDays	11.90394
VisITedResources	5.35671
Raisedhands	4.60107
Relation	3.80151
ParentAnsweringSurvey	3.67388
Topic	2.90231
Grade	2.85927
PlaceofBirth	1.67164
AnnouncementsView	1.46002
NationalITy	1.25713
Discussion	1.23744
ParentschoolSatisfaction	1.11010
Semester	0.67332
StageID	0.57633
SectionID	0.21072
GradeID	0.02735

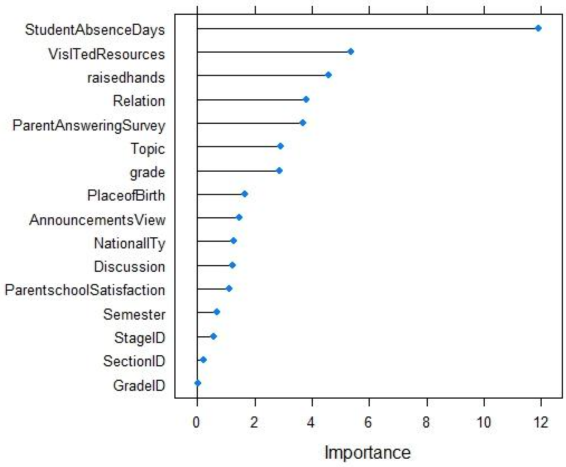

Fig. 4.Plot of Variable importance.

Feature selection is a basic assignment in data preprocessing territory. The reason of feature determination process is to pick a proper division of features which can intensely clarify the input data, diminishes the dimensionality of feature space, and take out repetitive and immaterial data [26]. In the proposed framework we utilized Recursive Feature Elimination. The routine method of feature selection was utilized for assemble numerous models with various parts of a data and it is used to recognize the data needed or not to form the exact model. A well known programmed technique for feature determination given as caret package of R is known as Recursive Feature Elimination (RFE). Feature choice techniques are ordered into wrapper-based and filter-based strategies. Recursive Feature elimination is wrapper based strategies. It is a greedy optimization algorithm which means to locate the best performing feature subset. It over and again makes models and keeps aside the best or the most noticeably bad performing feature at every iteration. It builds the next model with the left features until the point that every one of the features is depleted. It at that point positions the features dependent on the order of their elimination. First we stacked the libraries mlbench, caret and randomForest. At that point characterized the control using a random forest selection function and ran the RFE algorithm. The plot of chosen features is in fig. 5. Recursive feature selection with Cross Validated (10 fold) using Outer resampling method was applied. The main 5 variables (out of 14 chosen attributes) are StudentAbsenceDays, VisITedResources, raisedhands, AnnouncementsView, Relation.

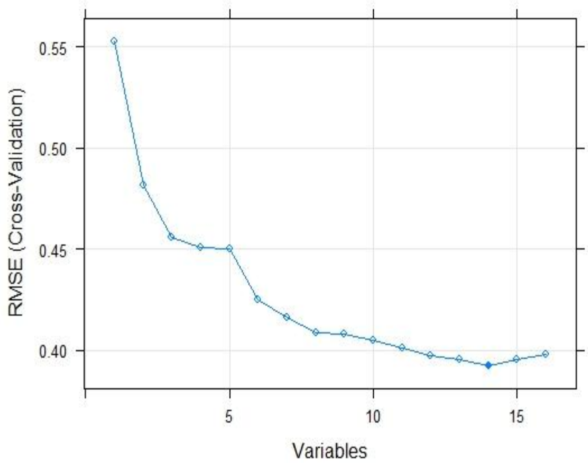

Fig. 5.Plot of feature selection.

VI. METHODOLOGY

Deep Neural Network (DNN) is a class of various models of Neural Network. It includes the input layers, arbitrary number of hidden layers and also an output layer. The layers are comprised of neurons which share similitude to human brain neurons. Network characterizes the neuron as a Central Processing Unit (CPU) which plays out a mathematical function to create a result from a collection of inputs. The neuron's output is produced based on the weighted total function applied on the bias and the input values [27]. Every neuron plays out an exceptionally straightforward task that includes initiating if the aggregate sum of signal received surpasses an activation threshold. The development of neural networks with huge number of hidden layers is called as Deep Neural Network. The fig. 6 shows the architecture of DNN. There are abundant products and packages are accessible in the marketplace for deep learning. Several of these are TensorFlow, h2o, Keras, and many more. Our proposed system used Keras and TensorFlow in R programming.

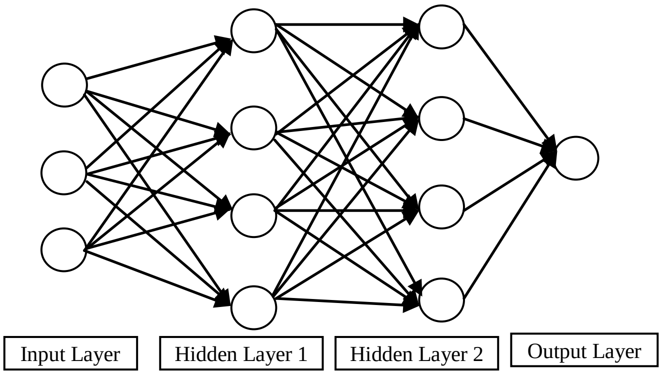

Fig. 6. Architecture of DNN.

Terminologies in DNN are Epoch, Activation functions, Weights, Bias, Output Layer, Hidden Layer, and Input Layer. There are 3 layers. The first layer is Input Layer which getting the input and then pass to the middle layer(s) this plays out the calculations is known as the Hidden Layer(s) and finally the output produces by the Output Layer. Epoch is a pass or iteration throughout the development of giving the input of the network and also updating the weight of the network. An activation function is an operation based on mathematical concepts which translates the input to an output. There are many types of activation function such as Unit Step Activation Function, Linear Function, Hyperbolic Tangent, Sigmoid Function, and Rectified Linear Unit (ReLU).

R was designed by Ross Ihaka and Robert Gentleman. There has dependably been an extreme challenge between R and Python with regards to Data Science and implementing Machine Learning. R has always been a statistician's choice. Because of the ongoing dispatch of Keras library in R with Tensorflow at the backend, it is again back in the challenge. Keras is an advanced API to construct, train and fine tune the deep learning models.

VII. IMPLEMENTATION AND RESULTS

The data comprise of 480 observations with 17 factors and three class target variable stacked by utilizing read.csv() method. Installed and loaded the libraries for example, keras, mlbench, dplyr, magrittr, and neuralnet. We formed the deep neural network model using neuralnet method with two hidden layers. First hidden layer contains ten neurons and second hidden layer contains five neurons. The deep neural network model is in fig 7. In each level the bias is computed with weight and produce to the next layer subsequently.

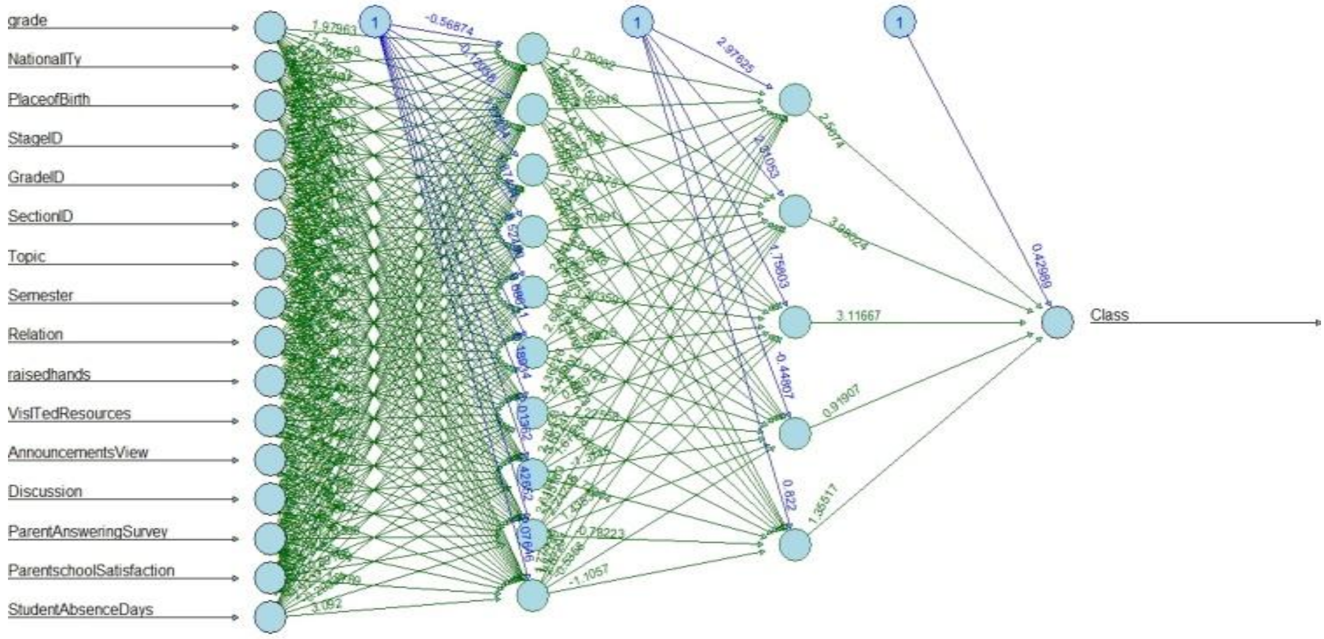

Fig. 7.Deep Neural Network Model.

The data structure is changed into a mathematical matrix utilizing the as_matrix() function before evacuating the variable names. The dataset, which presently exists as a matrix, that is part into training set as 70% (337 instances) and testing set as 30% (143 instances). Set the seed an incentive as 1234. At that point normalized the training data by applying colMeans() function. One-hot encoding is the method of changing categorical data to sparse data, which has segments of only ones and zeros this is also called creating design matrix. All non-numeric information should be changed over to dummy variables. This is basic for twofold Yes / No data since we can essentially change over to ones and zeros. It turns out to be marginally more difficult with various categories, which have need of making new columns of 0's and 1`s for every classification. The one hot encoding position is in Table IV. Created the model by Initialized with keras_model_sequential(). Then the sequential model is made out of a linear heap of layers. At that point connected the layers to the sequential model in which Layers comprise of the input layer, many hidden layers and an output layer. Hidden layers frame the neural network nodes that allow non-linear activation using weights. The hidden layers are made utilizing layer_dense(). Included two hidden layers; the first hidden layer contains ten number of nodes and second hidden layer contains five number of nodes at that point chose kernel_initializer = "uniform" and activation = "relu" (Rectified Linear Unit) for both layers. The primary layer needs to have the input_shape = 36, which is the quantity of sections in the training set. Dropout layers are utilized to control overfitting. This dispenses with weights beneath a cutoff threshold value to avoid low weights from overfitting the layers. We utilize the layer_dropout() function include two drop out layers with rate of 0.10 to expel loads underneath 10%.

Table- IV: One Hot Encoding format

Classes	One-Hot Encoding format
Low	[1, 0 , 0]
Medium	[0, 1 , 0]
High	[0, 0 , 1]

The output layer indicates the state of the output and the method for acclimatizing the educated data. The output layer is connected using the layer_dense() method. The synopsis of the model is in fig 7. For binary values classification, the shape should to be units = 1. For multi values classification, the units must compare to the quantity of classes that is in this case 3. We set the kernel_initializer is uniform and the activation is sigmoid. Then we printed the summary of the DNN model using summary(model) method in fig. 8. The DNN model contains information about the layer (type), output shape and the number of parameters. The input layer is passing 16 independent variables; the first hidden layer is 10 neurons, the number of parameters for this layer is 170 that is (16*10) +10. For second hidden layer are 55 that is (10*5) +5. Likewise For the output layer is 18 that is (5*3) +3. Finally the total number of parameters and trainable parameters for the Deep Neural Network model is 243 produced by summing all the layers parameters (170 + 55 + 18).

```
> summary(model)
Layer (type)                 Output Shape              Param #
================================================================
dense_1 (Dense)              (None, 10)                170
dropout_1 (Dropout)          (None, 10)                0
dense_2 (Dense)              (None, 5)                 55
dropout_2 (Dropout)          (None, 5)                 0
dense_3 (Dense)              (None, 3)                 18
================================================================
Total params: 243
Trainable params: 243
Non-trainable params: 0
```

Fig. 8.Summary of the model.

The subsequent stage is to assemble the model with the utilization of compile() method. We utilized analyzer as adam, which is a champion among the most well known optimization algorithms. We selected loss as categorical_crossentropy since this is a multi class classification. The measurements is c("accuracy") to be evaluated in the midst of training and testing. We utilized the fit() method to run the ANN on our training information. The batch_size is 32 which sets the number samples per slope update inside every epoch. We set the epoch as 100 to control the number preparing cycles. Typically we have to keep the batch size high since this reduction the blunder inside each training cycle that is epoch. We also need ages to be enormous which is basic in imagining the preparation history. We set validation_split as 0.20 to fuse 20% of the information for model approval, which deflects overfitting.

We fine tuned the model with various number of hidden layers and different number of neurons. Stop the tuning procedure until the measures are in stable. The accuracy of the model is 0.8766 and validation accuracy is 0.6569 which shown in fig. 9. Up to 5 epochs the accuracy value is increased, at that point after the fifth epoch the accuracy is stable. The fig. 9 and fig. 10 contains accuracy and loss value of the model. The loss of the model is 0.0942 and validation loss is 0.2974 which is shown in figure 10. Up to five epochs the loss value is diminished, at that point after the fifth epoch the loss value is stable. When the model reaches 100th epoch there is no much difference between the loss and validation loss of the data.

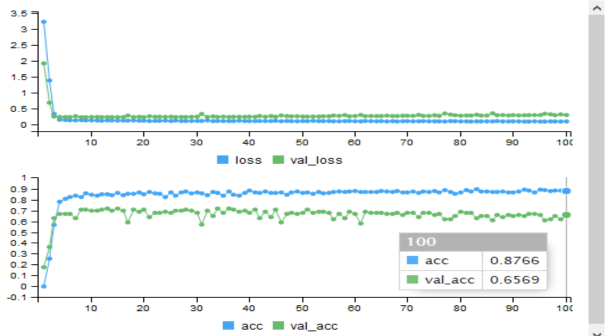

Fig. 9.Accuracy and Validation accuracy of model.

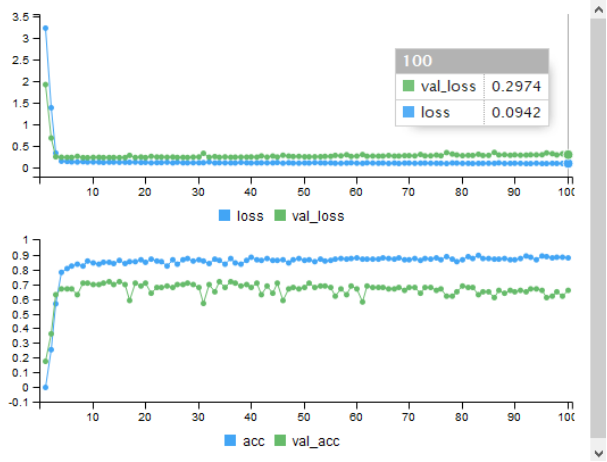

Fig. 10.Loss and Validation loss of the model.

The confusion matrix and the accuracy of the Deep Neural Network model is displayed in table V. Predicted and actual values 1, 2, 3 indicate low, medium and high.

Table- V: confusion Matrix

	Predicted			Accuracy
Actual	**1**	**2**	**3**	85%
1	39	8	0	
2	2	43	7	
3	0	4	40	

The plot of the model is delineated in the fig. 11. When training accuracy is increased the validation accuracy is additionally increased like manner when the training loss value is decreased the validation loss is also diminished. The result is compare with the work [8] which implemented the same dataset using deep neural network in python with keras library.

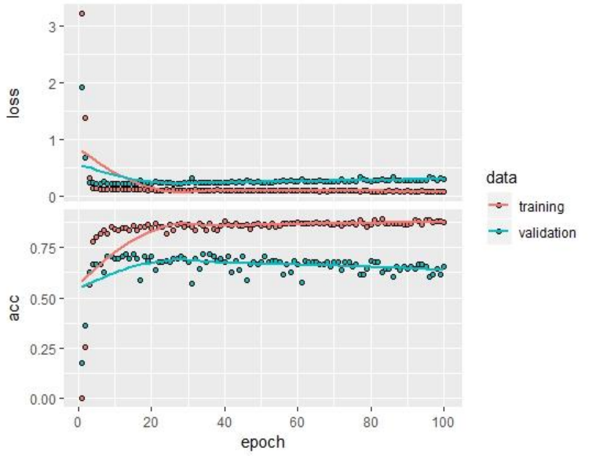

Fig. 11.Plot of the model.

R programming is the best tool for statistical data mining projects and real competition for python. So we executed similar data set using deep neural network in R Programming. Existing system [8] delivered 84.3% as accuracy and our proposed system produced 85% as accuracy.

VIII. CONCLUSION

Prediction of student performance process is achieved by deep neural network model. Machine learning algorithms are leading in which the process of classification is done by deep neural network. Students are classified as lasses such as low, middle and high. R Programming is the toughest competition for python to implement machine learning algorithms. We used R Programming with Keras library and Tensorflow as backend since for statistical concept it is best one. Formed neural network with two hidden layers and produced 85% as accuracy. The future work is to implement the dataset with different machine learning algorithms and compare the results.

REFERENCES

1. Marquec-vera, C. Cano, A. Romero, C., and Ventura, S. “Predicting student failure at school using genetic programming and different data mining approaches with high dimensional and imbalanced data”. In Applied Intelligence, 2013, 1, pp. 1-16.
2. Han, J., Kamber, M., and Pei, J. “Data Mining Concepts and Techniques, third edition. Morgan Kaufmann” 2013.
3. Cristobal Romero, “Educational Data Mining: A Review of the State-of-the-Art”, IEEE Transactions on systems, man and cybernetics. In Applications and Reviews vol. 40 issue 6, 2010, pp. 601 – 618.
4. Baker, R.S., Corbett, A.T., Koedinger, K.R., “Detecting StudentMisuse of Intelligent Tutoring Systems”. In Proceedings of the 7th International Conference on Intelligent Tutoring Systems, 2004, pp. 531-540.
5. Merceron, A., Yacef, K.,”A web-based tutoring tool with mining facilities to improve learning and teaching”. In Proceedings of the 11th International Conference on Artificial Intelligence in Education, 2003, pp. 201–208.
6. Beck, J., & Woolf, B.,”High-level student modelling with machine learning”. In Proceedings of the 5th International Conference on Intelligent Tutoring Systems, 2000, pp. 584–593.
7. Agrawal, H. and Mavani, H., “Student Performance Prediction using Machine Learning” in International Journal of Engineering Research and Technology, 2015.
8. Bendangnuksung and Dr. Prabu P, “Students' Performance Prediction Using Deep Neural Network”, in International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 2, 2018, pp. 1171-1176.
9. Herzog, S. “Estimating student retention and degree-completion time: Decision trees and neural networks vis-à-vis regression”. In New Directions for Institutional Research, 2006, 131, pp. 17-33.
10. Vandamme, J.-P., Meskens, N. and Superby, J.-F. “Predicting Academic Performance by Data Mining Methods”. In Education Economics, 15(4), 2007, pp. 405-419.
11. Ibrahim, Z. and Rusli, D. “Predicting students' academic performance: comparing artificial neural network, decision tree and linear regression”. In 21st Annual SAS Malaysia Forum, 5th September 2007, Kuala Lumpur.
12. Nghe, N. T., Janecek, P. and Haddawy, P. “A comparative analysis of techniques for predicting academic performance”. In 37th ASEE/IEEE Frontiers in Education Conference, 10-13th October 2007, Milwaukee.
13. P. Cortez, and A. Silva, “Using Data Mining To Predict Secondary School Student Performance”, In EUROSIS, A. Brito and J. Teixeira (Eds.), 2008, pp.5-12.
14. Zekić-Sušac, M., Frajman-Jakšić, A. and Drvenkar, N. “Neuron networks and trees of decision-making for prediction of efficiency in studies”, Ekonomski vjesnik (Econviews), 22(2), 2009, pp. 314-327.
15. Delen, D. “A comparative analysis of machine learning techniques for student retention management”. In Decision Support Systems, 49(4), 2010, pp. 498-506.
16. Osmanbegović, E. and Suljić, M. “Data mining approach for predicting student performance”. In Economic Review: Journal of Economics & Business /Ekonomska Revija: Casopis za Ekonomiju i Biznis, 10(1), 2012, pp. 3-12.
17. Kabakchieva, Dorina. "Student performance prediction by using data mining classification algorithms." In International Journal of Computer Science and Management Research 1.4, 2012, pp. 686-690.

18. Kovačić, Z. “Predicting student success by mining enrolment data”. In Research in Higher Education Journal, 15(1), 2012.
19. V. Ramesh,P. Parkavi, and K. Ramar. "Predicting student performance: a statistical and data mining approach". In International journal of computer applications 63.8, 2013, pp. 35-39.
20. Simeunović, V. and Preradović, Lj. “Using data mining to predict success in studying”. In Croatian Journal of Education, 16(2), 2014, pp. 491-523.
21. Cheewaprakobkit, P. “Predicting student academic achievement by using the decision tree and neural network techniques”. In Catalyst, 12(2), 2015, pp. 34-43.
22. Kaur, Parneet, Manpreet Singh, and Gurpreet Singh Josan. "Classification and prediction based data mining algorithms to predict slow learners in education sector". In Procedia Computer Science 57, 2015, pp. 500-508.
23. Yi, Hongsuk, Jung, H. and Bae, S., “Deep Neural Networks for traffic flow prediction”. In Big Data and Smart Computing (BigComp), 2017, IEEE International Conference, pp. 328-331.
24. C. Romero, J. R. Romero and S. Ventura, “A survey on pre-processing educational data”, In Educational Data Mining. Springer International Publishing, 2014, pp. 29-64.
25. A. G. Karegowda1, A. S. Manjunath2 and M. A. Jayaram3, “Comparative study of attribute selection using gain ratio and correlation based feature selection”, International Journal of Information Technology and Knowledge Management, vol. 2, no. 2, 2010, pp. 271-277.
26. Feature Selection: Available at online - https://machinelearningmastery.com/feature-selection-with-the-caret-r-package/
27. Ciaburro, Giuseppe, and Balaji Venkateswaran. “Neural Networks with R: Smart models using CNN, RNN, deep learning, and artificial intelligence principles”. Packt Publishing Ltd, 2017.

AUTHORS PROFILE

V. Vijayalakshmi received the M.Tech. degree in Computer Science and Engineering from Pondicherry University, Puducherry, India in 2011. She is currently pursuing Ph.D. (Computer Science and Engineering) from Annamalai University Chidambaram, India. She has published about 30 papers in National and International journals and presented several papers in National and International conferences. Her current areas of research interest include Data Mining, Machine Learning Algorithms. She is lifetime member of ISTE and IAENG.

Dr. K. Venkatachalapathy is currently working as the Professor in the Division of Computer and Information Science, Annamalai University, Tamilnadu, INDIA. He is a member of CSI and ISTE. He has published more than 45 research papers in reputed national and international journals and conferences. He has guided 9 Ph.D and currently guiding 7 research scholars. His main research work focuses on Digital Image Processing, Computer Networks, and Data Mining. He has 25 years of teaching experience.