**Student’s Grade Analysis using Machine Learning Techniques of Bachelor of Engineering Student of Pokhara University, Fall 2019**

****

**Faculty Research Report (Grant No: …./ )**

**Submitted to**

**Pokhara University Faculty Research Center (PUFRC)**

**Pokhara University**

**Pokhara-Lekhnath Metropolitan City- Ward 30, Kaski**

**Submitted by**

**Yagyanath Rimal, Lecturer**

**(2021)**

## **Declaration**

I hereby declare that this faculty research report is my own work and that it contains no materials previously published. I have not used its materials for the award of any kind and any other degree. Where other authors' sources of information have been used, they have been acknowledged. Ideas expressed in this report not necessarily represent the views of PUFRC. However, this will be the joint property of the research and PUFRC.

Signature:

Name of Principal Investigator: Yagyanath Rimal

Date: 2021/04/16

Appendix III: Expert Evaluation Sheet

**Student’s Grade Analysis using Machine Learning Techniques of Bachelor of Engineering Student of Pokhara University, Fall 2019**

Submitted by

Yagyanath Rimal. Lecturer

Has been evaluated by

Signature:

Name of Expert:

Designation:

Date:

Appendix IV: Format for Approval of Research Report by Research Evaluation Committee (REC)

**Student’s Grade Analysis using Machine Learning Techniques of Bachelor of Engineering Student of Pokhara University, Fall 2019**

Submitted by

Yagyanath Rimal, Lecture

Has been approved by REC

1. Signature Chairman REC:

Name:

1. Signature Member:

Name:

1. Signature Member:

Name:

1. Signature Member:

Name:

1. Signature Member:

Name:

## **Acknowledgement**

I would like to express my sincere thanks and appreciation to the director school of engineering, for his constant encouragement, mentoring, and faculty research grant support. This work would not have been possible without the constructive and insightful discussions with him and the faculty of science and technology years. I gratefully thank to research grant committee RMC, SOE for being always accessible and allowing me to interrupt him numerous times, regardless of busy scheduled to get his valuable feedback and new directions that I might explore in midterm research defense.

My earnest thanks go to program coordinator, student, and enumerator for serving constructive suggestions. My sincere thanks and appreciation go to Prof. Dr. Binaya Kumar Mishra for encouraging me to complete research problems, being always available to organized the contents, and for his collaborations and invaluable comments and suggestions in most of my written manuscripts. I would like to thank Associate prof Madhusudan Kayatta for his useful comments and his interesting contribution for helping me in sample selection at Pokhara University and for his constant encouragement. I would also like to express my gratitude to all the student respondent, data enumerators, Ashesh, Anish, Rita’s madam for collecting large sets of data encouragement and support.

Finally, I would like to express my gratitude to my advisor and reviewer for pointing me toward the right path to starting a career in academia and for his encouragement and support. I would also like to take this opportunity to thank his constant encouragement and support during my analysis few years at Pokhara University.

I cannot end without thanking my family Sakuntala Pageni, Aashraya Rimla, and Aaradhya Rimal, whose constant encouragement and love I have relied on during my thesis report design and complication of this research work.

## **Abstract**

Data analysis and its statists prediction of student performance before the final examination can help the student, faculty, and ultimately parents to make timely decisions. Here the researcher would like to predict student grade before the final examination and then grade comparison between internal rank and final scored rank of each sample subjects. Finally proposed the best network model to predict his/her coming student grade based on training and test data sets of 580 sampled students of Pokhara University Nepal 2019. The internal assessment, the high school, plus two, total, and subjects like physics, math’s, chemistry and bio subjects scored and education from either private and public schools and parent educational status ultimately affects to score final grade were significantly analyzed 18 neural network model.

Machine learning offers one of the most cost-effective and widely used approaches to developing determinant prediction tools. In the last few years, several advances in machine learning research have emerged. The researcher utilizes recent advances in neural network machine learning research to provide GPA prediction tools with improved predictive performance. First, the researcher used the Boruta algorithm to separate large independent variables for the prediction of final GPA, and second, the random forest methodology is used for predicting GPA epitopes then single, multiple hidden learned models for prediction of GPA of student model with their respective accuracy will be analyzed we used matrix, histogram, confusion matrix, and regression analysis were being highly used for the of machine learning neural network model in predicting final grade accuracy. Finally, we formulate the problems of qualitatively and quantitatively predicting the flexible length major histocompatibility complex of student grade using multiple instance regression.

The development of reliable machine learning neural network model prediction tools is not feasible in the absence of high-quality data sets with test and training includes both passed student and final student grades and their association between their past performance. However, the neural network model for GPA prediction demonstrates the pitfalls of these commonly used data sets for evaluating the performance of machine learning approaches to GPA prediction. Finally, the researcher proposes a similarity reduction procedure accurate model of the neural network model that is more stringent than currently used similarity reduction methods for student grade prediction using forward and backward neural networks.

## **Contents**

[**Declaration** 2](#_Toc69454768)

[**Acknowledgement** i](#_Toc69454769)

[**Abstract** ii](#_Toc69454770)

[**Contents** iii](#_Toc69454771)

[**List of Tables** v](#_Toc69454772)

[**List of Figures** v](#_Toc69454773)

[**Keywords** vi](#_Toc69454774)

[Chapter 1: Introduction 1](#_Toc69454775)

[**Background** 1](#_Toc69454776)

[**How a Neural Network Model Works** 6](#_Toc69454777)

[Chapter 2: Literature Review 13](#_Toc69454778)

[**Literature Review** 13](#_Toc69454779)

[Chapter 3: Methodology 24](#_Toc69454780)

[**The Study Area** 24](#_Toc69454781)

[**Methodology** 25](#_Toc69454782)

[Chapter 4: Findings and Discussions 27](#_Toc69454783)

[**Data Acquisition and Preprocessing** 27](#_Toc69454784)

[**Model Design and Setup** 28](#_Toc69454785)

[**Data Correction and Management using Boruta Algorithm** 29](#_Toc69454786)

[**Random Forest Methodology for Large Feature Selection** 34](#_Toc69454787)

[**Before Examination Grade Prediction** 44](#_Toc69454788)

[**Plus, two Back Propagation Grade Prediction** 49](#_Toc69454789)

[**Neural Network of Pass/Fail Prediction using Single Layer** 54](#_Toc69454790)

[**Neural Network Correction with Single Layer with Five Neurons (N=5)** 58](#_Toc69454791)

[**Neural Network of Two Hidden Layer (N=7, 3)** 61](#_Toc69454792)

[**Neural Network with Three Hidden Layers (N= 7, 3, 2)** 63](#_Toc69454793)

[**Neural Network with Original Datasets (N=7 ,3 ,2)** 67](#_Toc69454794)

[**Back Propagation of One Single Layer** 71](#_Toc69454795)

[**Back Propagation of Five Hidden Layers (N=5)** 75](#_Toc69454796)

[**Back Propagation of Two Hidden Layers (N = 7, 3)** 77](#_Toc69454797)

[**Back Propagation of Three Hidden Layers (N= 7,3,2)** 79](#_Toc69454798)

[**Neural Network with Four Hidden Layers (N= 25, 12, 7, 3)** 81](#_Toc69454799)

[**Hypothesis Testing** 83](#_Toc69454800)

[**Multi-nominal Logistic Regression** 85](#_Toc69454801)

[**Correlation and Linear Regression** 87](#_Toc69454802)

[Chapter 5: Conclusion and Recommendation 94](#_Toc69454803)

[**Performance Measures** 94](#_Toc69454804)

[**Conclusion and Recommendation** 95](#_Toc69454805)

[**References** 98](#_Toc69454806)

[**Annex1: Research Questionnaire** 102](#_Toc69454807)

[**Annex 3: Summary Statists of Sample Data** 103](#_Toc69454808)

[**Annex: Photographs** 104](#_Toc69454809)

## **List of Tables**

[Table 1: University Result Summary Fall 2019 Examination 25](#_Toc69480817)

[Table 2: Final Grade Conversion Table 28](#_Toc69480818)

[Table 3: Data Sets in Normal Sample 29](#_Toc69480819)

[Table 4: Data Sets in Normalized Sample 30](#_Toc69480820)

[Table 5: Summary Table by Program 92](#_Toc69480821)

[Table 6: Neural Network Model Accuracy 95](#_Toc69480822)

## **List of Figures**

[Figure 1: Conceptual Framework of Research Design 2](#_Toc69480825)

[Figure 2: First Before and After Histogram of Same Variable 7](#_Toc69480826)

[Figure 3: Single Hidden Layer Neural Network 8](#_Toc69480827)

[Figure 4: Histogram of Output Result 9](file:///D:\dataanalysisfinal\SOE\rcode\SOE%20Rimal%20sir%20Menuscript%20finallllllFF.docx#_Toc69480828)

[Figure 5: Neural Network with Passed Student 10](#_Toc69480829)

[Figure 6: Test and Training Conversed Output Histogram 11](#_Toc69480830)

[Figure 7: Human Cell of Neuron Structure 20](#_Toc69480831)

[Figure 8: Data Organization and Management 28](#_Toc69480832)

[Figure 9: Output of Boruta Algorithm 33](#_Toc69480833)

[Figure 10: Important, Tentative Important and Non important Plot 33](#_Toc69480834)

[Figure 11: OOB Plot of Random Forest 36](#_Toc69480835)

[Figure 12: Error Vs Trees and Mtry Plot 38](#_Toc69480836)

[Figure 13: Nodes and Number of Trees Histogram 40](#_Toc69480837)

[Figure 14: Mean Decrease Accuracy and Ranking Plots 41](#_Toc69480838)

[Figure 15: Partial Dependency Plots 42](#_Toc69480839)

[Figure 16: Partial Dependency Plot of Internal Grade 43](#_Toc69480840)

[Figure 17: MDS Plot of All GPA Points 44](#_Toc69480841)

[Figure 18: Mean Decrease Gini Plot 45](#_Toc69480842)

[Figure 19: Partial Plot of Math’s and Plus2 Grade 45](#_Toc69480843)

[Figure 20: Mean Accuracy and Gini Decrease Plot 46](#_Toc69480844)

[Figure 21: Partial Plot of Internal Marks 47](#_Toc69480845)

[Figure 22 : Partial Plot of Previous Grade 47](#_Toc69480846)

[Figure 23: MDS Plot of Grades 48](#_Toc69480847)

[Figure 24: Mean Accuracy and GINI Decrease Plot 48](#_Toc69480848)

[Figure 25: Partial Plot of Internal Grades 49](#_Toc69480849)

[Figure 26: Histogram of Plus to and Bachelor Grade Distribution 50](#_Toc69480850)

[Figure 27: Trees and Mtry Plot 51](#_Toc69480851)

[Figure 28: Histogram Nodes and Gini Decrease Plot 51](#_Toc69480852)

[Figure 29: Mean Decrease Accuracy and Gini Decrease Plot 52](#_Toc69480853)

[Figure 30: Partial Dependency Plot of Plus2 Grade 52](#_Toc69480854)

[Figure 31: Neural Network of 21 Independent Variables 53](#_Toc69480855)

[Figure 32: Histogram of Conversed Output Values 53](#_Toc69480856)

[Figure 33: Histogram of Half Model Passed Student 54](#_Toc69480857)

[Figure 34: Neural Network Singe Layer of All Result 55](#_Toc69480858)

[Figure 35: Histogram of Training and Test Conversed Output 56](#_Toc69480859)

[Figure 36 : Histogram of Half Neural Network Model 57](#_Toc69480860)

[Figure 37: Neural Network with Single Layer 5 Neurons 58](#_Toc69480861)

[Figure 38: Histogram of Conversed Output of Single Layer 5 Neurons 59](#_Toc69480862)

[Figure 39: Histogram of Single Layer Conversed Output 5 Neurons 60](#_Toc69480863)

[Figure 40: Neural Network of Two Hidden Layers ( 7,3) 61](#_Toc69480864)

[Figure 41: Histogram Output of Test and Training Data 62](#_Toc69480865)

[Figure 42: Histogram Output of Training and Test 62](#_Toc69480866)

[Figure 43: Neural Network Model of Three Hidden Layers 64](#_Toc69480867)

[Figure 44: Histogram of Test and Training Output 65](#_Toc69480868)

[Figure 45: Histogram of Training and Test Data 66](#_Toc69480869)

[Figure 46: Neural Network Model of (7,3,2) with Original Data 68](#_Toc69480870)

[Figure 47: Histogram of Training and Test of Original Data 68](#_Toc69480871)

[Figure 48: Histogram of Three Hidden Layers with Original Datasets 69](#_Toc69480872)

[Figure 49: Single Layer Backpropagation Neural Network 73](#_Toc69480873)

[Figure 50: Histogram of Single Hidden Layer Backpropagation 73](#_Toc69480874)

[Figure 51: Neural Network of Single Layer Backpropagation 74](#_Toc69480875)

[Figure 52: Histogram of Single Layer Conversed Value 74](#_Toc69480876)

[Figure 53: Full and Half Model Neural Network of Single Layer 5 Neuron 75](#_Toc69480877)

[Figure 54: Full Model Histograms of 5 Neuron Single Layer 76](#_Toc69480878)

[Figure 55: Half Model Histogram of Training and Test data 77](#_Toc69480879)

[Figure 56 : Neural Network with Two Hidden Layer N = 7,3 78](#_Toc69480880)

[Figure 57: Histogram of Conversed Using Two Layers 78](#_Toc69480881)

[Figure 58: Full and Half Neural Network Model using N= 7,3,2 79](#_Toc69480882)

[Figure 59: Full and Half Model Histogram of Neural Network 80](#_Toc69480883)

[Figure 60: Full and Half Model Neural Network of (N= 25,12,7,3) 82](#_Toc69480884)

[Figure 61: Histogram Output Neuron of Four Layer Network 82](#_Toc69480885)

[Figure 62: Correlation Plots 89](#_Toc69480886)

[Figure 63: Residual and QQ Plots 89](#_Toc69480887)

[Figure 64: Residual and QQ Plot 91](#_Toc69480888)

## **Keywords**

Correlation; Assignments; Final examination; lowest marks; highest marks; Grade Marks prediction; Educational Data Mining (EDM), Neural Network Analysis.

# **Chapter 1: Introduction**

## **Background**

Teaching learning process include various methodology storytelling, discussion, teaching, training, and directed research practices exchange between student and teacher. Learning is the process of getting known from the unknown concept, which requires the more practical and appropriate theoretical concept of brain development which ultimately facilitate learning, acquisition of knowledge, skills, values, beliefs, and habit of human learning. The methodology of teaching is called pedagogy that facilitates the student to learn new knowledge. Similarly, the evaluation is the process of characterizing and appraising some aspect of an educational process of learning achievement outcomes. Educational institutions usually require evaluation student data to demonstrate the effectiveness of stakeholders and to provide a measure of performance for marketing purposes. The students on their class performance and activities to improve their performance is an integral part of every instruction. Although there were many practices are being applied to measure students’ performance in instant duration. Most of the western international quality assurance sets out various criteria to accreditation standards measures are used that directly setout of the various procedure of practical activities of teaching evaluating process in new approach of mechanism for determining their grades. An evaluation like tests, interviews, questionnaires, problem-solving, case studies, quiz are various tools were being used to standardized the learning process globally. The knowledge measurement is another key important milestone of learning institution term as grade the pass/ failure. However, the formative evaluation focus to implementation the program whereas summative evaluation test with various measures (Yahya, Mohamed, & Abdelfatah, 2011) evaluates students grade properly. On the other hand, the teaching and learning process through eastern education philosophy stresses the major outcomes form the teachers which means that teachers are completely responsible for their class to make effective planning and activities inside course for their students grading. The assessment preferences inventory (API) which measured students’ preferences as a pre-test and perceptions as a post-test (Filip, 2008) are always positive for learning. There are two major grading system still available in worldwide, one is percentage system which ultimately grades students as first grade, second grade, third grade and pass / failed graded with equivalent grade points margins. Generally, there were 9 grades, which are A+, A, B+, B, C+, C, D, E, and N.A refers to ‘outstanding’ will be given to students scoring between 100 and 90, A to students scoring between 80 and 89, B+ to those scoring between 70 and 79, B to scores between 60 and 69, C+ to scores between 50 and 59, and C to scores between 40 and 49 worldwide practices which may vary each university system rules. While grade point calculation (GPA) is weighted by credit hours, multiply each numeric grade value by the number of credits the course worth. The sum of these numbers together and divide by the total number of credits each student scored is another grading system also emerging smoothly.

**Theoretical Framework**

Substantial studies based on neural networks have been conducted on data from schools, colleges, and online multiple choices courses, aiming at predict student achievement (Junemann, 2007) . here researcher apply artificial neural networks to estimate future student performance based on students’ family, social, and wealth characteristics as inputs parameter to predict student grade using forward and backward neural network model prediction.

SGPA

Gender, GPA,

parent’s education status

Unit test score

Assignment mark

Internal score

Attendance

Extracurricular activities

SLC, +2 Grades

Machine learning Model

Verification

Actual

Predicted

Figure : Conceptual Framework of Research Design

As the teaching-learning process is a mixture of various dimensions this research tries to find out neural network prediction of various input variables who’s predicated SGPA will again further validated with university final GPA of the sampled student for the model validation reduction of misclassification errors.

**Objectives**

The specific objectives of this research are to analyze the student grade of SOE fall 2019.

To assess whether the performance of continuous assessment determined the final grade.

To find out the correlation between the internal marks and final grades obtained by the students.

To predict the student grade before final examination based on associative previous performance using various machine learning techniques.

**Hypotheses**

H1: there is no significant difference between subject grades with continuous assessment of the student.

H1: there is a significant difference between pass ratios of different batches of students.

**Research Question**

What are the strategies which can be adopted to reduce the consistency of internal and final grade of the student?

Are there similarities among internal grades and final similar among various subjects?

**What is Neural Network**

A neural network used in deep learning consists of different layers connected at work on the structure and function like of the human brain. It learns from huge volumes of data and uses a complex algorithm to train the neuron. When data with multidimensional input are feed to the network for the prediction of two images of dogs, the information of dogs are in vectors of information were being fed through hidden layers for prediction accuracy. While forwarding network the past information did not remember while forming another layer, while weight and output form each node. The feed-forward network used regression and classification problems of predictor variables. Where convolution neural network is used in image recognition. The deep neural network is used for acoustic modeling and deep network is for cancer detection and the recurrent neural network for speech recognition.

According to (Woodford, 2017) neural network works like brain neuron computer programs assembled from hundreds, thousands, or millions of artificial brain cells that learn and behave in a remarkably similar way to human brains of 100 billion neurons. Neurons are so tiny that you could pack about 100 of their cell bodies into a single millimeter transmitted the axon to node. The neural network is to simulate lots of densely interconnected brain cells inside a computer so to learn things, recognized patterns and decisions in human like. The best thing of the neural network is to learns itself like brain is sometimes called artificial neural network which millions of artificial neurons called units arranged in a series of layers, each of which connects to the layers on either side.  Input, hidden and output units are together connected fully connected. Which means each input and output networks were connected one unit to other units and the weight of represents its weights. the higher weights influence to another. A richer structure is called deep neural network for tackling more complex problem however it requires more training speed. neural network works like brain neuron computer programs assembled from hundreds, thousands, or millions of artificial brain cells that learn and behave in a remarkably. A richer structure is called a deep neural network for tackling more complex problems however it requires more training speed.

**Feedforward Neural Network**

The information process in a neural network flows in two ways when data are feed into another layer using training the pattern is feed into the network via input singles which are activated using bios terms for the next layers this unidirectional flow is known as forwarding network. Where each unit received the inputs from, the other units are multiplied by its weights of connection they travel using activation function. Using sigmoid and threshold value triggers and units of its. The neural network learns the things using feedback while doing so it modify weight reducing the between actual and prediction intended output where two exactly coincide. In this type of network, the information is only moving forward using an activation function with constant bios to another layer. The input layers pass to hidden layers to output layers. Network activation flows in one direction only: from the input layer to the output layer, passing through the hidden layer. Each unit in a layer is connected in the forward direction to every unit in the next layer.

**Back Propagation**

Backpropagation described by Arthur E. Bryson and Yu-Chi Ho in 1969, that it gained recognition, and it led to a “renaissance” in the field of artificial neural network research.

As the algorithm's name implies, the errors propagate backward from the output nodes to the inner nodes. so technically speaking, backpropagation is used to calculate the gradient of the error of the network concerning the network's modifiable weights This gradient is almost always used in a simple stochastic gradient descent algorithm to find weights that minimize the error. Backpropagation is used in a more general sense, to refer entire procedure encompassing both the calculation of the gradient and its use in stochastic gradient descent. Backpropagation usually allows quick convergence on satisfactory local minima for error in the kind of networks to which it is suited.

Backpropagation is a multilayer feed-forward network with one layer of z-hidden units. The y output units have b(i) bias and the z-hidden unit has b(h) as bias. It is found that both the output and hidden units have a bias. The bias acts like weights on the connection from units whose output is always 1. The input layer is connected to the hidden layer and the output layer is connected to the output layer utilizing interconnection weights. The architecture of backpropagation resembles a multi-layered feed-forward network. The increasing number of hidden layers results in the computational complexity of the network. As a result, the time taken for convergence and to minimize the error. The bias is provided for both the hidden and the output layer, to act upon the net input to be calculated.

The training algorithm of backpropagation involves four stages. Initialization of weights some small random values are assigned. Feedforward each input unit (X) receives an input signal and transmits this signal to each of the hidden units Z1, Z2, Zn. Each hidden unit then calculates the activation function and sends its signal Zi to each output unit. The output unit calculates the activation function to form the response of the given input pattern. Backpropagation of errors output unit compares activation Yk with its target value Tk to determine the associated error for each unit. Based on the error, the factor δ K (k=1, m) is computed to distribute the error at output unit Yk back to all units in previous layer. Similarly, the factor δj (j=1, p) is compared for each hidden unit Zj updating of the weights and biases respectively.

The standard method and generally works well used in the algorithm can be applied to any network. It does not require any special features of the function to be learned. Computing time is reduced if the weights chosen are small at the beginning. A batch update of weights exists, which provides a soothing effect on the weight correction terms.

The problem appears to have overwhelming complexity when large data sets, but there is a solution. It is easy to create several examples of the correct behavior. The solution to the problem may change over time, within the bounds of the given input and output parameters can be or non-numeric.

Feature’s extraction method limits the nature and output of the preprocessing step and the decision to use gray-scale versus binary image, filled representation or contour, thinned skeletons versus full-stroke images depends on the nature of the features to be extracted. For training the neural network, backpropagation with momentum training method is followed. This method was selected because of its simplicity and because it has been previously used on several pattern recognition problems. The recognition of the isolated handwritten digits, but also to the recognition of the handwritten numeric chains constituted of a variable number of digits.

In 1979, Kunihiko Fukushima invented an artificial neural network which has a hierarchical multilayered architecture and acquires the ability to recognize visual patterns through learning.

## **How a Neural Network Model Works**

The neural network is the best tool for large noisy data however it has less interpretability as compared to other models of machine learning. Here the researcher tries to explain clearly elaborate its working modality and hyperparameter designed of sample research data. After loading the neural network library, the data sets of student records loaded using read.csv command as the input data variable in r console. The str command is used to describe the database description as however, the database value needs to be normalized using min and max procedures in machine learning model. Here researcher just selects 16 independent variables for prediction student marks of 580 observations with their normalized values. Here researcher tries to explain with all students’ model with GPA as explanatory variables having "0" "1" "2" "3" "4" "5" "6" "7" levels of final scored grades indicates 0 for fail,1 for A,2 for A-,3 for B+ and so on order the structure of data looks like.

$ GPA: int 2 3 1 3 2 5 5 2 2 2 ...

$ SLC: num 0.69 0.792 0.583 0.833 0.861 ...

$ School: num 1 1 0 0 0 1 0 1 1 1 ...

$ Plus: num 0.833 0.722 0.694 0.778 0.778 ...

$ Physics: num 0.811 0.568 0.541 0.676 0.676 ...

$ Math: num 0.519 0.75 0.577 0.865 0.865 ...

$ Chemistry: num 0.789 0.553 0.763 0.395 0.395 ...

$ Bio: num 0.924 0.815 0.761 0.761 0.761 ...

$ Parent1: num 0 1 0.667 0.333 0.333 ...

$ Parent2: num 0.667 1 1 1 1 ...

$ X1: num 0.88 0.94 0.98 0.8 1 0.84 1 0.88 0.9 0.92 ...

$ X2: num 0.78 0.78 0.88 0.82 0.82 0.72 0.82 0.88 0.88 0.9 ...

$ X3: num 0.542 0.375 0.542 0.917 0.917 ...

$ X4: num 0.88 0.82 0.86 0.72 0.76 0.72 0.76 0.92 0.84 0.9 ...

$ X5: num 0.84 0.96 0.78 0.72 0.9 0.92 0.9 0.78 0.84 0.78 ...

$ X6: num 0.84 0.8 0.82 0.64 0.84 0.68 0.84 0.64 0.94 0.94 ...

The data has a dependent and independent association of variables whose values in integer type of all independent variables in GPA. The GPA has 0 if the student has failed otherwise passed in final examination results 2019. The student marks SGPA could be converted with a respective categorical factor with 10 levels of A, A-, B+ B, B-, C+, C, C- D+, D, D- and Fail grade. Here the research problem is supervised machine learning needs classification of neural network with numerical representation.

The normalization data as presented table 3 plot histogram with all database or each column will be normalized the concept of substation of minimum and divides by max and subtraction of minimum of each record of observation. The summary statists present in the figure 1. The dependent variable GPA produce10 categories with factor type numeric type as.

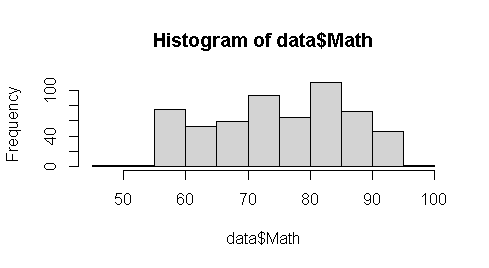
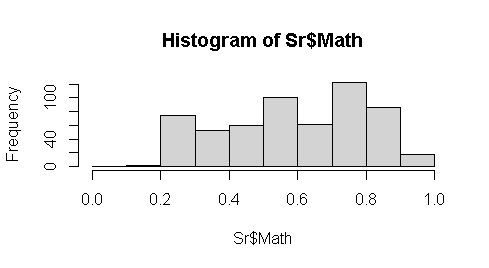
 

Figure 2: First Before and After Histogram of Same Variable

After setting out seed value to get same output after next execution from the model. The tidyverse package is used to filter the selected datasets make new research split subset. The data partition splits sample function divides two samples with 70% and 30% for training and test for model design validation randomly. After loading neural network package of neural network using n=neuralnet (p ~ SLC + School + Plus + Physics +Math +Chemistry + Bio + Parent1 + Parent2+ X1+ X2 + X3 +X4 +X5 + X6, data=training, hidden=c (1), err.fct="sse, linear. Output = FALSE) model. here the hyperparameters like “sse” is chosen for error calculation purpose and with single hidden layer were chosen whose linear output become set as false. The differential function for calculation of errors "ce" cross entropy or "sse" sum of the square could be selected. The plot “plot(n)” displays the output depends on the dependent variable GPA having numeric or factor category too after seeing repetition with minimum errors in plot (n, rep=1) if we set more repetition in design may select other execution. The hidden hyper parameter 1 indicates a single hidden layer for mathematical explanation here.

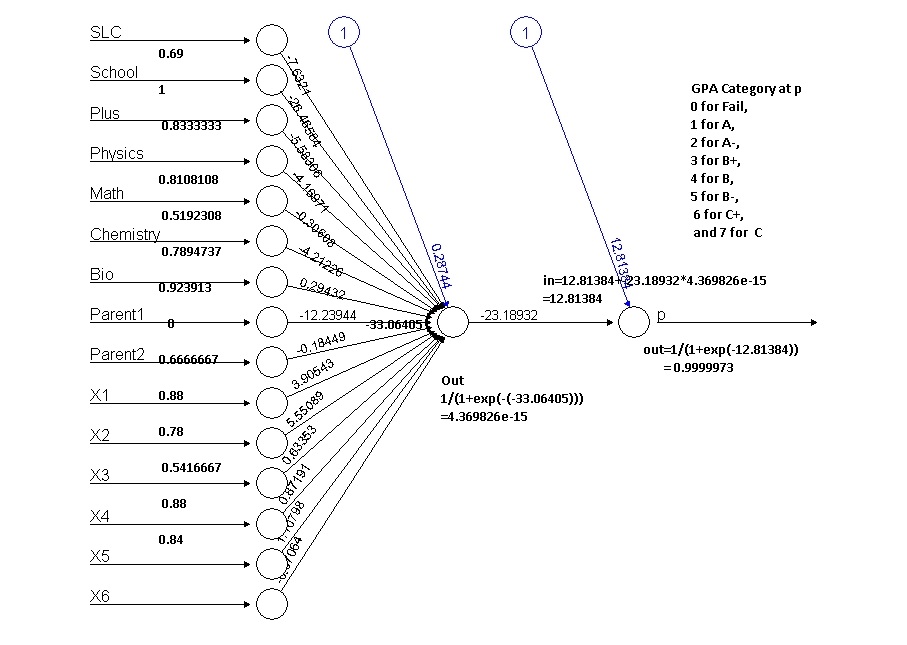


Figure 3: Single Hidden Layer Neural Network

The model output with original value is calculated using compute function output=neural net: compute (n, training [, -1]) command display the output is 0.4812963, 0.4812963 0.4812963

580 0.4812963 values of all 580 sample variables of student database. The head (training [1,]) describes first records is GPAX, SLC, Plus, Physics, Math, Chemistry, Bio, Parent1, X1,

X2, X3, X4, X5 and X61 respectively their weights are 0.3907611 0.7916667, 0.7222222, 0.5675676, 0.75, 0.5526316, 0.8152174, 1 0.94, 0.78, 0.375, 0.82 ,0.96 and 2 0.8 describes (y-) respectively the output conversed value product at last layer after traversed though hidden layer in between. The hidden parameter computes more adjustment if multiple layers need to converse to the GPAX. Another hyper parameter err.fct="see" is chosen for only when the data sets with having more than two categories either true or false type if so err.fct=” ce” dependent outcome is applicable for design the model. The first figure received first 13 input nodes and hidden layer 14 node is calculated with the sum of first bios and sum of each weight of previous input neuron as in14 = 0.28744+ (-7.6321\* 0.69)+ (-26.465584\*1)+ (-5.58306 \* 0.8333333) + (-4.16971 \*0.8108108) + (-0.30608 \* 0.5192308) + (-4.211226 \* 0.7894737) + (0.29432 \* 0.923913)+(-12.23944\*0) + (18449\* 0.6666667)+ (3.90543\*0.88)+(5.55089\*0.78)+(0.63353\* 0.5416667 )+ (0.87191\*0.88)+ (1.10798 \* 0.84)+ (-0.07064\*0.84) which produce the value of (-33.06405) is input at node 15th node and produced the output from using sigmoid function is 1/(1+exp(-in14)) is 4.369826e-15 which always lines in between 1 and 0 because which used sigmoid function that indicates association and the 15th node input become to calculated its weight using 12.81384+(-23.18932\*4.369826e-15) produces the value -12.81384 ultimately the output from 16th is again calculated using sigmoid function 1/(1+exp(-(-12.81384))) produces each input signals 0.9999973 which is shown in figure conversed outcomes matched with head first records using result command whose output is display as.

1 0.99996886

8 0.99996885

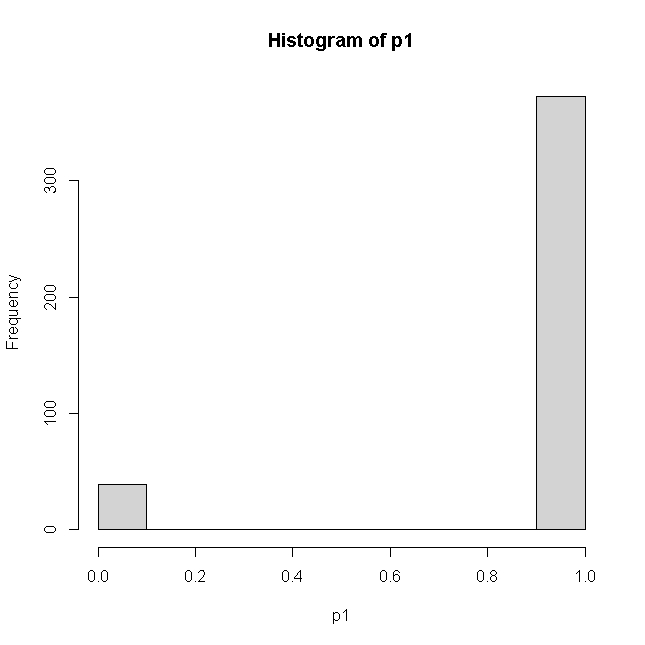


Figure : Histogram of Output Result

9 0.99996864

11 0.99996871

12 0.99996874

13 0.99996881

576 0.01189466

578 0.99996886

580 0.99996886

This is our inquiry of prediction of grade of eight scale graded whose histogram predicated on seven scale grades model (0 for Fail, 1 for A, 2 for A-, 3 for B+ 4 for B 5 for B- 5 for C+ and 7 for C-. The model testing and validation using confusion matrix of model could be calculated using compute function compute (n, training [, -1]) function with excluding its dependent variable GPA. The head result plots the histogram further plotted for the proper prediction of accurate grade based on input signals. Whose prediction and accuracy could be calculated based on predicted function after grouped up again using the logic pred1= ifelse (p1>=0.88,1, ifelse(p1>=.75, 2, ifelse (p1>=.65,3, ifelse(p1>=0.55,4, ifelse(p1>=0.30,5, ifelse(p1>=0.2,6, ifelse(p1>=.15,7,0))))))) and table of confusion matrix predicated and training data describes.

pred1 0 1 2 3 4 5 6

0 38 0 1 0 0 0 0

1 177 30 66 42 32 22 4

The accuracy is calculated using formula is 1-sum (diag (tab1))/sum(tab1) become 0.8349515. This accuracy implies that the total grade prediction exactly matched with 84 percent of student final grade were validate by model. Similarly, the test prediction produces confusion matrix as.

pred2 0 1 2 3 4 5 6 7

0 4 0 1 1 1 0 0 0

1 70 19 28 22 6 13 1 1

3 1 0 0 0 0 0 0 0

Which accuracy using formula 1-sum (diag (tab2))/sum(tab2) is 0.8630952. The above accuracy indicates these data need to be make prepare more correction in the model to get more accuracy although there were more than 84 percent accuracy in both training and test data sets of 580 student records. If incase used 22 predictor variables using final grade the accuracy goes around 50 percent and when the dependent variable only has two categories this model produces 97 percent accuracy. This process indicates based on dependent variables type and margin of accuracy with categorial type model production largely differs and needs some data preparation too. However, due to many dependent variables with eight prediction category this model did accuracy more than 85 percent is good enough.

Similarly, the prediction of neural network with only passed student model after selection of same data sets of above 13 input data sets with the only passed student whose scored final grade of categorized A for 1, A- for 2, B+ for 3, B for 4, B- for 5, C+ for 6, and C- for 7 seven category result of engineering students. After normalized of predictor variable to explanatory variables the training and testing data separation of 70 and 30 percent of only 290 records (Half Model). The model n=neuralnet (GPA~SLC+ School+ Plus+Physics +Math Chemistry+ Bio+Parent1+ Parent2 +X1+ X2+ X3+ X4+ X5+ X6, data=training, hidden=c (1), err.fct="sse”, linear. output=TRUE). Here error factor is used as the sum of the square with a single hidden neural network conversed at.

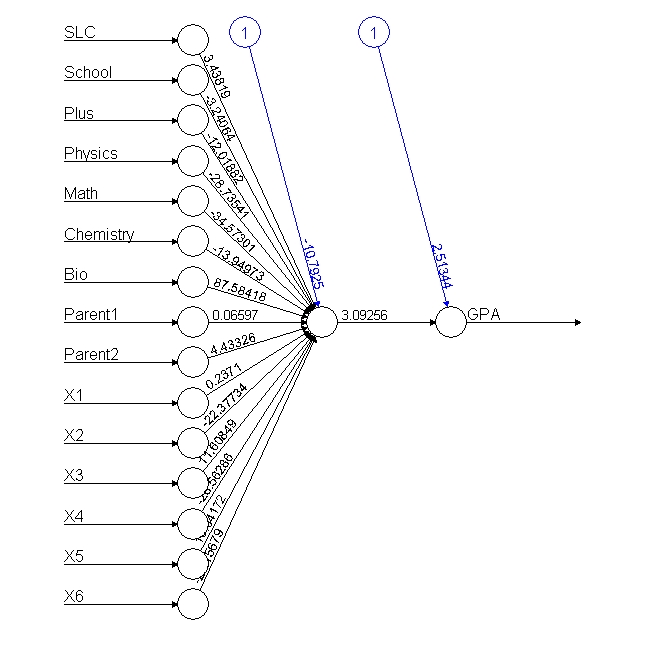


Figure 5: Neural Network with Passed Student

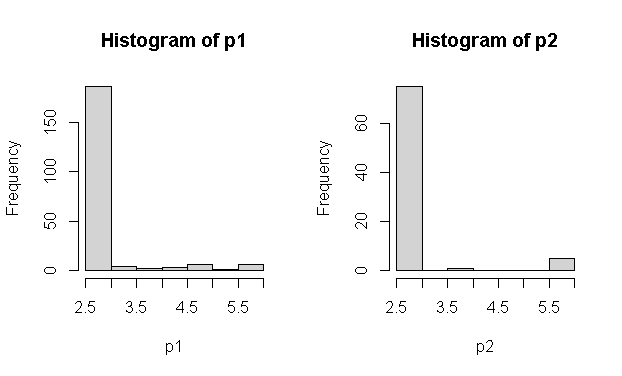
From the above output conversion of neuron, the input to in14th node is calculated 2.39998+ (2.08578\*0.69) + (4.64526\*1) + (0.25413\*0.8285714) + (2.83408\*0.8108108) + (1.0983\* 0.4047619) + (2.69845\*0.7894737) + (1.2807\*0.923913) + (0.65258\*0) + (4.74991\*0.6666667) + (0.31636\*.7777778) + (0.93483\*0.56) +(0.29306\*0.5217391) +(0.63501\*0.7777778) +(0.59444\*0.7575758) +(1.36278\*0.7647059). Which produce the input to 14th node is 20.82648 and is again passed to sigmoid function input to another node 1/(1+exp(-in14)) is 1 and again input to 15th node is 1 must always lines between 1 and 0 because of sigmoid function then output from that node is again calculated with 0.21029 +(2.43648\*out14) is 2.64677 wight to another node. Thus, the final node SGPAO=1/(1+exp(-SGPI)) output from the last neuron is 0.9338116 implies the accuracy is 93 percent matched. This matched with head first records of net$result value of model. The after formation of output with net. neuron result calculated as data of 290 records of passed students. 

Figure 6: Test and Training Conversed Output Histogram

The prediction after calculating compute function the respective values of each grade is further categorized assuming ifelse(p1>=5.5,1, ifelse(p1>=5,2, ifelse(p1>=4.5,3, ifelse(p1>=4.0,4, ifelse(p1>=3.5,5, ifelse (p1>=3,6,7)))))) for the purpose of total predication formed table (pred1, training$GPA). The confusion matrix of test and training data as bellow indicates the accuracy of model.

pred1 1 2 3 4 5 6 7

pred2 1 2 3 4 5 6

1 0 1 2 0 2 0

5 0 1 0 0 0 0

7 14 25 13 12 9 2

1 0 0 0 1 3 0 2

2 0 0 0 0 0 1 0

3 0 0 0 1 5 0 0

4 0 0 0 2 1 0 0

5 0 0 1 1 0 0 0

6 0 0 3 1 0 0 0

7 38 66 48 19 15 1 0

Whose prediction is again calculated using 1-sum(diag(tab1))/sum(tab1) is 0.9904306% in training data sets of 209 records of the test set and the testing result accuracy is 0.8271605 percent of model design is more than 82 presents is quite a good model prediction although the random samples of splitting differ validation. However, the model predicts more accurately than full model although needs some improvement of selecting dependent variable before design final model of GPA prediction.

These above explanations summarized that data preparation needs to get more accurate result. The average of both tests of training and test accuracy implies more than 85 percent accuracy is good model without data management.

# **Chapter 2: Literature Review**

## **Literature Review**

Although a similar study had been carried out department of education, the University of Maastricht (David, Rijt, Filip, & Janine, 2008) concluded that there are no relationships were found between students’ perceptions of assessment and their assessment scores. Although students preferred written tests, including home exams and papers, in which they are allowed to use supporting materials such as notes and books, as well as papers or projects. The oral tests and the other modes of alternative assessment mentioned in the question computerized tests and portfolios, are not amongst the students’ preferences procedures. Similarly, type of grade prediction research had been carried out in The Islamic University of Bahawalpur (Muhammad & Aijaz Ahmed, 2007) concluded 16 students got more than 80 marks in internal assignments but they failed in the final examination but this study covers master in education student only. Similarly, for those students, who scored highest 88.07 marks in assignments, while the least mean 49.57 shows that students are getting the lowest marks in the final examination.

Similarly, (Brijesh Kumar & Saurabha, 2011) apply the ID3 algorithm use sample of 50 students to identify various attributes to perform best scores. (Naqui, 2006) Used linear regression on a sample of 300 students from affiliated colleges of Panjab University. The study (John & Manabete, 2015) surveyed 1847 students in Nigerian University concluded that the internal consistency reliability of instrument which was 0.86 used heterogeneous classes in university. They applied the Naive Bayes method to predict student performance from the data warehouse of the student data. The research collected sample of 220 students out of which 175 records are considered as training, 45 records are considered as test data to achieve 86.66% accuracy. Similarly, (Patel, 2017) suggest that the internal examination pattern must be in university standard of three tests of any subject should be conducted properly recommended to use one algorithm to have good predictive accuracy among a large class data. (Sorour, Mine, & Hirokawa, 2014) claims to achieve 82 percent accuracy using artificial neural network using decision tree algorithm based on J48 and multiple linear regression to predict student performance using regression analysis of 181 students’ marks.

Similarly, (Ramaswami & Rasthnasbapathy, 2010) applied the Bayesian and CHID algorithm used 5560 students to predict the performance of higher secondary students based on 35 contributing attributing. Similar findings were reported by (Brijesh Kumar & Saurabha, 2011) using the Naïve Bayesian classification of 300 students. (Pandey & Pal, 2012) Investigated a sample of 600 students used Bayes classification algorithm. Similar (Cohen, 1995) research conducted to predict student performance of 778 Portuguese students using machine learning algorithm concluded that tree algorithm is the best, to predict whether the student would pass or fail in mathematics and Portuguese language courses.

There are many studies that investigated the ways of applying machine learning techniques for various educational purposes. (Havan & Harshil, 2015) studies that investigated the application of machine learning techniques in distance learning for dropout used Bayesian classification for their dropout student. Similarly, (Nikolopoulos, 2009) research conducted national institute of Athens conducted a survey on student drop out survey use of Naive Bayes indeed performed better than any other machine learning algorithm. However, the vital contribution of this study was that time-invariant features may be detrimental to the machine learning process, are better left out of the study entirely. He also concluded that instead of demographic characteristics of students, using initial attendance and homework grades produces a better prediction rate at earlier stages. (Brijesh Kumar & Saurabh, 2012) conducted a study on the student performance-based by selecting 300 students from 5 different degree college conducting BCA in Dr. R. M. L. Awadh University, Faizabad, India use Bayesian classification method on 17 attributes, it was found that the factors like students' grade in secondary exam, living location, medium of teaching, mother's qualification, students another habit, family annual income, and student's family status were highly correlated with the student academic performance. In the present study, those variables whose probability values were greater than 0.70 were the highly influencing variables with high probability values. These features were used for prediction model construction they used both variable selection and prediction, model construction using MATLAB.

The origin of the neural network in 1940s when two researchers Warren McCulloch and Walter Pitts, tried to build a model to replicate how biological neurons work. Though the focus of this research was on the anatomy of the brain, it turns out that this model introduced a new approach for solving technical problems outside biological science. Although there were two types of machine learning techniques i.e., supervised and unsupervised which could be selected based on problem definition and prediction. In general, there were four categories; classification techniques are carried out when there is a new category will be predicted after having various attributes being contributed in different attributes of input data. For example, after having some temperature, wind, humidity, wind speed, seasons, monsoon data sets attribute it can easily predict today or tomorrow’s weather, in various categories sunny, windy, rainy, cloudy, etc. Similarly determine if the loan application is a high, medium, or low risk of IPO. Whereas regression analysis always predicts fixed numeric values after setting out various input data attributes. The price of the stock is always in numeric increment or decrement, not a category.

The cluster analysis techniques are applying when organizing similar items into various groups, the resulting grade of the student in a single class will be A, A-, B+, B, B- C+, C- D category. Data analysis is the rule to capture the association between items and events which may be applied to test the association with non-similar products like when most of the Friday supermarkets sold beer and diapers. Therefore, supervised learning only applies in different labels whereas unsupervised techniques apply on unlabeled data items where the target is not available.

Similarly, (Rushika, Juilee, Pooja, Sachee, & Priya, 2018) farmers' crop prediction in Maharashtra crop suits their soil quality, soil nutrients, and soil composition used Multiple Linear Regression (MLR) technique for crop analysis prediction. The decision tree algorithm and classification are used to perform analysis of over 362 datasets and provide result. The training dataset here is classified into organic, inorganic, and real estate for predicting the type of soil. The backpropagation network uses a hidden layer which helps in better performance in predicting soil properties. The two-regression supervised machine learning methods are used Support Vector Machine (SVM) and Relevance Vector Machine (RVM) to show effectiveness in soil quality prediction used smart wireless device for sensing soil moisture and meteorological data. The wireless device gives an error rate of 15% and 95% accuracy. However, it has not been tested for real-time data. Which used three methods includes Decision tree, Naive Bayes classifier, and KNN classifier which analyses soil and predicts crop yield, farmers' crop prediction. (El & Yasser, 2008) use Support Vector machine to identify the B cell antigenic sequences is critical for understanding disease pathogenesis, for identifying potential autoantigens, and for designing vaccines and immune-based cancer therapies for predicting linear B-cell epitopes and flexible length linear B-cell epitopes, respectively, using string kernel-based support vector machine (SVM) classifiers. Likewise, (Verdes, Granitto, Navone, & Ceccatto, 2010) predict frost prediction for horticultural with considering various methodological variables temperatures frost protection of fruits and vegetables of machine learning techniques usually employed in regression and classification problems, including artificial neural networks, used Bayes classifiers, and k-Nearest Neighbors. The results obtained in this preliminary study reveal a very noisy structure of the data (Shoon Lei, Zaw Zaw, Faridah, & Ibrahim, 2014) entropy-based gene selection approach to select relevant prognostic genes that are directly responsible for recurrence prediction. This proposed system has achieved an average accuracy of 98.9% in predicting cancer recurrence over 3 cancer datasets is one of the deadliest diseases in the world and is responsible for around 13% of all deaths worldwide. The cancer incidence rate is growing at an alarming rate in the world Naive Bayes, which is fundamentally built on the strong independence assumption, is a very popular classifier in machine learning due to its simplicity, efficiency, and efficacy (Fisher, 2014).

(Kalejaye, Folorunso, & Usman, 2015) at the Federal University of Agriculture, Abeokuta, Nigeria used of artificial neural network (ANN) model for predicting students’ academic performance in a University System, based on the previous datasets consists of 60 students in the department of computer and information science, Tai Solaria University of education in Ogun state, who have completed four academic sessions from the university. The codes were written and executed using MATLAB format.

The test data evaluations showed that the ANN model can predict correctly, the final grade of students with 91.7% accuracy for CGPA prediction. In this work, various demographic features (age, gender, and race), as well as college entrance examination results, were used to train a neural network to predict student final grade of students in e-learning courses using multiple feed-forward neural networks and multiple-choice test data administered to the students of National Technical University, Athens, Greece as input. Engineering course of the University of Ibadan using various influencing factors such as unlevel scores, matriculation exam scores, age on admission, parental educational background, and others as input variables. The results indicated that the neural network was able to correctly estimate the performance of more than 70% of prospective students at University Malaysia, Pohang using a fuzzy logic system and predictive factors such as secondary school results from strength, number of sittings, mode of entry, parent literacy and others. This study suggested that AI techniques are capable to handle the uncertainty associated with students’ performance, to determine their strengths and weaknesses used four machine learning models learning tree, bagging, random forest, and boosting have been used to predict the academic performance of students by Hudson and Christian in 2014 results showing high prediction accuracy. (Lykourentzou, Giannoukos, Mpardis, Nikolopoulos, & Loumos, 2009).

A feed-forward neural network (FFNN), is a backpropagation network that allows the signal to flow one way, from input to output layer. No feedback mechanism tends to be the straightforward network that connects inputs with outputs. Besides, FFNN consists of one or more hidden layers of neurons in which neural connections, called synapses, do not form a directed cycle. The information moves only forward, from the input to the final output node. During its learning phase, the network is presented with a set of examples called the training set. Each example consists of an input vector and the corresponding output vector. This type of learning is known as supervised learning. The goal of the FFNN training is to minimize the sum of square error (SSE) and more recently, mean square error (MSE) between its actual and target outputs, by adjusting the network synaptic weights and neuron biases more particularly, these network parameters are adjusted based on the backpropagation algorithm. Inspired by the structure of the brain, a neural network consists of a set of highly interconnected artificial neural units, called processing elements. Each unit is designed to mimic its biological counterpart and accepts a weighted set of inputs and responds with an output. The neural networks address the problem that is often not easy for traditional computers to solve, typically, speech and pattern recognition, time-series forecasting, scheduling, regression, and function approximation. An artificial neural network is the area of science that deals with methods and systems for information processing using a neural network is called neuro computation (Adepoju, Ogunjuyigbe, & Alawode, 2007). To address these constraints, (Karamouzis, 2000) reported the development of a prototype electronic system called (Final Grade Model). This study evaluated FIG’s predictive power model. (Miranda & Martin, 2013) conducted a study to identify the various impacting factors of student performance using 120 students of mathematical and tutorial classes.

Similarly, the (Mubarak & Umar, 2019) conduct a higher student grade prediction of 61 computer network data set using WEKA software predicts 73.68 percent of student prediction whose 66.6 percent were likely to drop out before graduating. However higher education institute's goal is to tards improving higher education quality, the success of human capital creation is a continuous issue (Osmanbegović & Suljić, 2012). Thus, knowledge of data mining to extract hidden knowledge useful not only for the student but also for teachers, administrations, and academic institutions (Romero & Ventura, 2010). The data mining, machine learning and statistical methods used decision tree and genetic algorithm to develop the performance of each subject predict secondary school student failure by applying and comparing four data mining algorithms decision tree, random forest, neural network, and support vector model. This model has to make early score prediction even before the semester starts for the final examination of engineering courses (Saleem, Kathiry, Al-Osimi, & Badr, 2015) which use logistic regression and discriminant analysis for student failure. The most of analysis use the neural network method to discover potential data pattern of student marks. However, this study tries to used past student grades association with current internal grades to predict and classification student marks. Similarly, (Yang & Brinton, 2017) conduct research on massive online courses (MOOS) of pre-student performance prediction found 60 percent accuracy of average performance using neural networks. The predictive learning networks that predict learning outcomes in different points ultimately informed poor student attention from an instructor and cause dropout in early attention (Brinton & Chiang, 2016). However, the greatest challenges of grade prediction are many students may choose a single alternative answer in multiple type questions. And learning process may depend on many personalized motivations which largely increased due to online materials increases 61% more RMSE and on average more than 10 percent importance in grade performance.

Similarly, according to (Zacharis, 2016) research was conducted with a multilayer perception neural network to predict student performance on blended learning courses. Which predicts 93% accuracy of classification of the potential student of the first year. And (Wongkhamdi & Seresangtakul, 2010) achieve 86 percent of student prediction accuracy. (Shahiri, Husain, & Rashid, 2015) study the effectiveness of artificial neural network in forecasting graduate student outcome with 93.3 as compared discriminant analysis 81.3 percent. The research found that 98% of accuracy in neural network and 91% accuracy on decision tree prediction whereas 76 percent on k nearest and 76 percent accuracy using naïve biased prediction matches of educational data mining.

Similarly, (Sidath, 2018) defines that a classification model as a model or classifier that is constructed to predict categorical labels whereas the prediction model a continuous-valued function of ordered value. In recent years education data management has significant attraction has emerged as a core research area due to its large potential for prediction. Therefore, it is very necessary to analyzed internal marks its association with the final grade of the student in SOE Pokhara university. Besides the subject prediction, this study tries to find out factor’s hindrances of internal examination management of SOE in early stage. In many cases, human behavior of previous experiences makes decisions for future design. This concept will be predicted for the improvement of higher education in the school of engineering their class performance and motivating them to continue to improve their performance is an integral part of every instruction.

We know that the teaching-learning process always requires continuous assessment of all-day classroom activities which finally meets curriculum is a difficult task when grading categorization among students. The curriculum is the formal activities offered to the learner through the educational institution. When Oxford University first held an examination in 1958, assumed that examination would give stimulus to scholars and teachers (Ranjita, Sushobhan, Ananya, Nina, & Patralekha, 2014). The continuous internal assessment and evaluation system is strong a type of teaching learning activity (Nnodium, 1992). Similarly, study conducted (Lima, Mohanan, & Harichandran, 2017) in medical student with 40 sample study concluded that there was a positive correlation between attendance and marks in all the internal theory and practical examination. The examination always measures the knowledge level of a student whose progress is always required for university management. The effective management depends on the ability to assess the quality of services provider in the education sector. According to (Lykourentzou, Mpardis, & Nikolopoulos, 2009), learners’ performance prediction can help to identify the weak learners at early stages and properly assist them to cope with their study.

Although similarly study on students scored mean 88.07 in internal marks in assignment were getting lowest marks in the final examination indicates there were large missed matched internal and final university results. Based on the literature empirical study, on importance-performance analysis, the study identifies the strengths and weaknesses of neural networks (Silva, 2018). Similarly, (Daguang & Xie, 2017) suggested to measure student performance in the workplace is assessed in terms of written and verbal communication skills, ability to innovate, analytical problem-solving skills, group management ability, self-study skills, adaptability, performance under pressure, workplace comportment, ability to work with others, and sense of work responsibility. Preparedness for work is assessed in terms of professional knowledge, fundamental knowledge, extra-curricular knowledge, English proficiency, computer proficiency, and research ability which only measure the total quality of the student. If student performance predicted before the final examination, could help all stakeholders, a student, to make better timely decisions. Before the semester-end examination, the office of advisable council has to decide which courses to offers visiting scholars’ precautions. So that students can prepare better teaching materials which ultimately leads to the dropout ratio of classes. This creates an atmosphere of uncertainty and frustration for management and student life in national development. Timely counseling can help the student to make extra effort to avoid unwilling drop-out in the university examination.

There are various applications of neural networks in artificial intelligence and solving various business problems like classification includes pattern recognition, face recognition feature extraction image processing matching signature in the bank, prediction of historical pattern and largely applicable in noise reduction of large input data. The supervised learning is the machine learning task for label data class can also be called dependent variables whereas age gender occupation is termed as independent variables to predict dependent component. Likewise, unsupervised learning task of exploring the data of deriving some inference from data sets. When the prediction of data was not used to design the model used unlabeled type data structure is known as unsupervised learning. Which tries to find out the patterned of relationship from the relationship defied its prediction based on training and test. When the target variables are continuous variables apply to the regression model. Whereas the prediction variables with more categories are called classification. When data without header fall under unsupervised techniques supports dimension reduction techniques PCA and factor analysis for association analysis when using e-commerce prediction automatically referred another product after customer bias products. Therefore, unsupervised predicts could not develop the model for future analysis just clustering.

The neural network model has to mimic the neuron simulation like human brain is known as an artificial neural network. In our brain, we get signals to from external receptors cell which transmits an impulse from sensors to the receptors that enable human to receive mechanical, thermal chemical, and stimuli, to other receiver neuron signals that process of signals and send them to corresponding executive organs so-called effectors.



Figure 7: Human Cell of Neuron Structure

The human brain has a massive parallel connection with quick fault-tolerant, the capability of adapting to surroundings, the generalized form is known examples and priorities more than collective behaviors renter than individual behaviors. The final grade prediction based on the limited initial data of students and courses is a challenging task because, at the beginning of undergraduate studies, most of the students are motivated and perform well in the first semester but as time passed there might be a decrease in motivation and performance of the students. Their study showed that in-class exams were better predictors of the overall performance of a student than the homework assignment. The study also demonstrated that timely prediction of the performance of each student would allow instructors accordingly. (Meier, Xu, & Van, 2016) According to their findings, the undergraduate performance of the students could explain 54% of the variance in graduate-level performance.

(Agoritsa & George, 2019) argues that the performance that a student achieved in a subset of the earlier courses can be used to predict how well he/she will perform in future courses. Here researcher query is to developed a grade prediction method, called course-specific regression that predicts the grade that a student will achieve in a specific course linear combination of the grades that the student obtained in past courses using historical student course grade information to accurately estimates how well students rely on the neighborhood-based collaborative iterative model. The researcher proposed low-rank matrix factorization is the best performing scheme, based on course specifics regression, achieves an RMSE across the different courses of 0.632 whereas the best competing method achieves an RMSE of 0.661 percent accuracy. However, the researcher used only three-semester grades of students of 2,949 students, records. Similarly, (Yupei, Yue, Huan, Jiaqi, & Xuequn, 2020) used a majorization minimization algorithm to analyzed 1325 students of higher education data sets. This research used decision tree logic regression and support vector machine methodology. This process used non-negative matrix factorization (NMF) was used to integrate the nonnegativity of student grade matrix is an implicitly low rank, low-rank matrix factorization (LRMF) investigated in data sets from the online learning platform in the work (Lorenzen, Pham, & Alstrup, 2017). The family information background, such as the economic situation and educational level of their parents, influences the scope of student knowledge (Brecko, 2004). (Emaan Abdul, 2016) machine-learning approaches predict student marks and grades before the final examination for 2500 student course records of an institute of higher education suggest that the grade of more than 96% of students can be accurately predicted. Similarly, (Tair & Alaa, 2012) conducted a study to identify the attributes impacting student performance the most. They used only a genetic algorithm on a sample of 120 students. Their analysis model considers quantitative factors such as theoretical, mathematical, practical, and other departmental marks were used as independent components. They applied only the ID3 algorithm of 50 students to identify the attributes that affected the performance of the students (Hijazi & Naqvi, 2006) used linear regression on a sample of 300 students (225 males, 75 females) from a group of colleges affiliated with University of Punjab. They used social and demographic variables alongside the grades of the students found that the mother’s education and student’s family income were highly correlated with the student's academic performance. A similar finding reported by using Naïve Bayes classification on a similar sample of 300 students course records of eight semesters are obtained for the undergrad programs from different semester. For each semester, 25 to 30 courses with an aggregate enrolment of 800 to 1000 students. Therefore, no model could produce hundred percent accuracy on these data set because the weightage of the final exam varies from 25% to 50% of the total marks, which is quite a high weightage cannot model these types of variables, therefore it is less realistic to predict the final grade with 100% accuracy (Maghari, 2017).

(Bhardwaj, 2011) proposed a research work that any pattern could be useful as the strongest prediction for students’ performance at the UNWE university based on pre-university and personal characteristics. The results showed that the decision tree performs the best accuracy, followed by KNN classifier whereas the Bayes classifiers had the lowest accuracy. The authors analyzed the data gathered from colleges student performance prediction by using MATLAB tools. Their study showed that the performance of the students does not always depend on the student effort, but, other factors can affect the student’s performance too.

Recently, other researchers (Hamsa, Indiradevi, & Kizhakkethottam, 2016) conducted a study to predict the performance of academic students in both master's and bachelor's degrees using different classification algorithms and different subjects. However, a few efforts applied classification algorithms on educational datasets of secondary schools in the Gaza Strip. In our study applied major kinds of classification techniques KNN and Naïve Bayes to help the ministry of education to predict the performance of new students, and improve results for the next year. (Agoritsa & Karypis, 2016) presents future course grade prediction methods based on sparse linear and low-rank matrix factorization models that are specific to each course or student–course tuple.

Therefore, it is concluded that the bachelor of engineering student grade prediction using neural network needs to be study for better improvement however it needs more data sets and program wise analysis using background and current student achievement to predict student GPA before final examination.

# **Chapter 3: Methodology**

## **The Study Area**

The semester-end result of engineering students of various programs regularly notified to the student but the analysis on them was not been carried out in higher education engineering student in Nepal. Out of 11 engineering programs the five programs are currently being running in SOE. There are 624 regular students were appeared in the final examination fall 2019 out of them 280 were completed with CGPA scored composed of 47.44 percentage. However, the overall university pass ratio of these programs was 11.25 percentage with considering regular or retake examination result in the fall 2019. When considering program wise pass, the CIVIL VIII students had the highest 52.38 percentage whereas Computer I, CIVIL l VII and Civil V were the second highest and rural students have lowest pass ratio 12.28 percentage. The Electrical III semester students have a better result than Civil III semester students. Civil V and VII have the highest pass ratio and Elex V had least pass ratio is 8 percentage in 2019 result of SOE. Similarly, university wise comparison of program CIVIL V and Elex V has mismatched the university passed ration as compared SOE results. For data analysis using machine learning 280 pass student’s records, as well as same amount of failed student’s records, were collected and verified from the school of engineering administration office records where 580 questionnaire surveys were collected for neural network model design with the help of student enumerators constitute the database.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Program | Total | Pass | Percentage | SOE | PASS | Percentage | Total sample |
| Civil I | 2818 | 232 | 8.232789 | 156 | 24 | 15.38462 | 48 |
| Civil III | 2788 | 162 | 5.810617 | 172 | 20 | 11.62791 | 40 |
| Civil V | 2320 | 154 | 6.637931 | 176 | 46 | 26.13636 | 92 |
| Civil VII | 1743 | 208 | 11.93345 | 148 | 40 | 27.02703 | 80 |
| Civil VIII | 1269 | 458 | 36.09141 | 147 | 77 | 52.38095 | 154 |
| Computer I | 1012 | 136 | 13.43874 | 47 | 17 | 36.17021 | 34 |
| Civil/Rural I | 94 | 12 | 12.76596 | 48 | 6 | 12.5 | 12 |
| Software I | 340 | 59 | 17.35294 | 47 | 9 | 19.14894 | 18 |
| Elex I | 327 | 33 | 10.09174 | 82 | 12 | 14.63415 | 24 |
| Elex III | 328 | 28 | 8.536585 | 96 | 13 | 13.54167 | 26 |
| Elex V | 302 | 20 | 6.622517 | 75 | 6 | 8 | 12 |
| Elex VII | 218 | 20 | 9.174312 | 64 | 8 | 12.5 | 16 |
| Elex VIII | 210 | 28 | 13.33333 | 68 | 12 | 17.64706 | 24 |
|  | 13769 | 1550 |  | 1326 | 290 | 21.87029 | 580 |
| 13\*48 |  | 11.257 |  | 624 | 0.464744 |  |  |

Table 1: University Result Summary Fall 2019 Examination

This table represents the result comparison with percentage as compared national and school of engineering based on result 2019. The overall university campuses results had 25% better results than other campuses located in Nepal.

## **Methodology**

This study design includes a mixed (retrospective and non-interventional) study including quantitative and qualitative aspects of marks prediction.

Study Setting: This study was conducted in the Pokhara University under the School of Engineering where there were 8 groups of civil engineering classes and 4 groups of electrical and electronic, one group of each in civil and rural, computer and software engineering classes are running in constitute college under the director of the school of engineering comprises total study universe. Whose current results of fall 2019 were collected in even semester end program composed of I, III, V, VII and VIII students were the sample population.

Study Population: Regarding internal marks prediction of all groups of civil and electrical and electronics, computer, software and civil and rural student (48\*13~624) composite the study population with CGPA collected as the equal number of pass and fail students of each program.

Duration of the study: May 2019 to May 2020.

Sample Size: Although 13 classes were being currently being running under bachelor program in faculty of science and technology in constituent campus, out of them all sample cluster were selected (4 Civil and 4 Electrical and Electronic, 1 Civil and Rural and 1 from Computer and 1 from Software) with the consideration of each batch representation.

Sampling Techniques: A stratified random sampling procedure selected among 13 groups of SOE where the both successful student in their final examination selected complete in the same ratio (290) of fail student were selected by random order, no repetition were taken as sample composed 580 sample machine learning analysis a primary study.

Data Collection Tools and Techniques: Vermiculated structured questionnaire was being used to collect the data regarding the background previous information like high school, parents’ education, and their plus to grades with physics, math chemistry, and bio total scored were collected and validate with administrative records. Besides, a focus group discussion with CR with guidelines question used as data collection tools. Data were collected through an individual questionnaire of the sampled student of each ratio of passed and failed students. Informed consent was taken from each of the participants before the data collection. The principal investigator and class representative were involved in data collection.

Data Management and Analysis: Collected data were managed carefully by taking the consideration of safety in first Excel program. Editing, coding, and data cleaning were maintained. The data were analyzed using R software application. The appropriate descriptive statistics such as percentage, standard deviation, max, and min were applied to conclude research. Similarly, regression analysis, performance measures, decision tree classification, random forest algorithm, and Boruta algorithm were ruffianly analyzed and explained to predict data internal marks and final grade point. Thematic analysis was used for the qualitative data.

Therefore, this study tries to apply supervised machine techniques i.e., decision tree classifiers and Boruta algorithm were taken as data preparation and neural network classifiers to predict dependent variable GPA. Similarly, the correlation coefficient after multicollinearity test of internal and final marks was calculated and explained using R programming. The relationship between each sample program was analyzed.

Validity and Reliability: The validity and reliability of the study were maintained by the following techniques: A structured and pretested questionnaire was used for the data collection. The expert advice was explored using pretesting of the tools was carried out amongst 10% of samples.

# **Chapter 4: Findings and Discussions**

## **Data Acquisition and Preprocessing**

This primary study used 580 student data sets of Pokhara University engineering students, whose final grade points were collected from an online result sheet published in various months in 2020 that is available under download section of university home page. Similarly, the second part of their respective internal marks of sample student was collected from the school of engineering internal examination sections of all 580 sampled students. The 50/50 (Pass/ Fail) student data were collected from primary datasets from the enumerator to complete datasets for machine learning neural network analysis. The data were carefully reviewed preprocessed and organized in excel data base before loading into r console. The incomplete and irrelevant data were being eliminated. The respective GPA of all semester student records of first, third, fifth, seventh, and final eight semesters of each respective course comprised datasets. This final dataset includes internal marks of their respective six subjects were collected in each row fashion. The student's school-leaving grades and their +2 grade were being collected with primary subjects like Physics, Chemistry, and Biology and their final grade point average GPA. Similarly, the student background of the institution of learned from i.e., government school or private boarding schools were collected through a selected enumerator and codded. The GPA and marks of each student were converted and uniformly organized based on the following table. The parent highest education level was collected in 1 for SLC completion, 2 for Plustwo completion 3 for Bachelor completion, and 4 for Master and above level education.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| SLC and +2 Grading | | | | Bachelors | | |
| Grade | Grade Point | % | Description | Grade | Grade Point | % |
| A+ | 3.6- 4.0 | >=90 | Outstanding | A | 4.0 | >=90 |
| A | 3.2-3.6 | >=80&&<90 | Excellent | A- | 3.7 | >=85 |
| B+ | 2.8-3.2 | >=70&&<80 | Very Good | B+ | 3.3 | >=80 |
| B | 2.4-2.8 | >=60&&<70 | Good | B | 3.0 | >=75 |
| C+ | 2.0-2.4 | >=50&&<60 | Above Average | B- | 2.7 | >=70 |
| C | 1.6-2.0 | >=40&&<50 | Average | C+ | 2.3 | >=65 |
| D+ | 1.2-1.6 | >=30&&<40 | Billow Average | C | 2.0 | >=60 |
| D | 0.8-1.2 | >=20&&<30 | Insufficient | C- | 1.7 | >=55 |
| E | <=0.8 | >=0&&<20 | Very insufficient | D+ | 1.3 | >=50 |
|  |  |  |  | D | 1.0 | >=45 |
| Table 2: Final Grade Conversion Table | | | | F | 0.0 |  |

Note: The cumulative grade point is the average grade point of all the semester which is calculated as CGPA= cumulative total honor points earned/ cumulative total number of credit hours taken.

The high school studied of sampled student were stored in school column with 1 for private school and 2 for government school and the parent education of each sampled student were stored in 1 for SLC, 2 for +2, 3 for Bachelor and 4 for Master degree completion categorical data. Altogether there were 49 columns of records with 580 composed 28,420 cells record of the student information for GPA predication. The percentage and GPA conversion table 2 was based on Higher school grade table of government of Nepal and university grade conversion or vice versa for the purpose of data transformation.

## **Model Design and Setup**

The R powerful opensource tools offer a wide variety of classification and prediction methods was used to build the test the student marks prediction. A neural network algorithm implementation of the bachelor degree grade prediction algorithms. Each classifier is evaluated using the holdout approach i.e., we train the model 70% of the data while we evaluate it on the remaining 30% which is referred to as the testing data or the unseen data. This is done to avoid overfitting, a phenomenon that arises when we evaluate the model on the same set of records that we used to learn it from. Similarly, the Boruta algorithm and random forest model were sufficiently analyzed to separate unwanted columns of research 49 columns.

.

Classification

Model

Dataset

Testing

Training Data

Data mining

Algorithm

Results

Validation Verification

Figure 8: Data Organization and Management

## **Data Correction and Management using Boruta Algorithm**

Here researcher used many ways of feature selection and finding out the best important features for further data analysis. The first the Boruta algorithm was used for feature engineering from large variables to get accurate output due to many reasons and associativity of relationship with many independent variables. The Boruta algorithm apply when there were a large number of features i.e., 49 predictor independent variables which delimitates time and constraints to form the best model accuracy. However, it is always necessary to find out some shadow attributes do not require for the model best fit which shuffled all values then generate the model including very insignificant importance for the rebuild final model. The reason behind using Boruta algorithm for the preparation of data before model development was to reduced unimportant and tentative insignificant independent variables omission for best prediction.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |  | 49 |
| SN | Prog | SLC | School | Plus | Physics | Math | Chemistry | FinalGPA |
| 1 | Civil I | 79.84 | 1 | 78 | 85 | 72 | 85 | 80 |
| 2 | Civil I | 83.5 | 1 | 74.2 | 76 | 84 | 76 | 75 |
| 3 | Civil I | 76 | 2 | 73.35 | 75 | 75 | 84 | 85 |
| 4 | Civil I | 85 | 2 | 76 | 80 | 90 | 70 | 75 |
| 5 | Civil I | 86 | 2 | 76 | 80 | 90 | 70 | 80 |
| 6 | Civil I | 87 | 1 | 74 | 76 | 78 | 70 | 65 |
| …………… | | | | | | | | | |
| 578 | ELEX VIII | 55 | 2 | 65 | 55 | 58 | 75 |  | 0 |
| 579 | ELEX VIII | 55 | 1 | 55 | 75 | 80 | 75 | 0 |
| 580 | ELEX VIII | 60 | 1 | 55 | 85 | 90 | 85 | 0 |

Table 3: Data Sets in Normal Sample

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SN | 2 | 33 | 4 | 5 | 6 | 7 | 8 |  | 49 |
| 1 | Prog | SLC | School | Plus | Physics | Math | Chemistry |  | FinalGPA |
| 2 | Civil I | 0.69 | 0 | 0.833 | 0.810 | 0.519 | 0.789 |  | 1 |
| 3 | Civil I | 0.791 | 0 | 0.722 | 0.567 | 0.75 | 0.552 |  | 1 |
| 4 | Civil I | 0.583 | 1 | 0.694 | 0.540 | 0.576 | 0.763 |  | 1 |
| 5 | Civil I | 0.833 | 1 | 0.777 | 0.675 | 0.865 | 0.394 |  | 1 |
| 6 | Civil I | 0.861 | 1 | 0.777 | 0.675 | 0.865 | 0.394 |  | 1 |
|  | Civil I | 0.888 | 0 | 0.722 | 0.567 | 0.634 | 0.394 |  | 1 |
|  |  |  |  |  |  |  |  |  |  |
| 578 | ELEX VIII | 0 | 1 | 0.472 | 0 | 0.25 | 0.526 |  | 0 |
| 579 | ELEX VIII | 0 | 0 | 0.194 | 0.540 | 0.673 | 0.526 |  | 0 |
| 580 | ELEX VIII | 0.13889 | 0 | 0.194 | 0.810 | 0.865 | 0.789 |  | 0 |

Table 4: Data Sets in Normalized Sample

The first table describes the database structure and second table describes the normalized data of designed database of student records. After loading Boruta, mlbench, caret dplyr and tidverse library data management packages the database loaded using read, csv command into the working directory as data=read.csv("D:/dataanalysisfinal/SOE/rcode/Dataanalysiscsv.csv") then database converted into data frame format for data in 2d matrix structure for analysis. The first the null value and missing attributes were omitted if present to the database using na. omit (S) function or sometimes researcher used mutate using mean substation to the database in case there were missing values to be treated in database. The data having various information with data frame describes 580 observation of student records of 49 independent variables needs to select only significant to formed the model before features selection and model development.

$ SNN: int 1 2 3 4 5 6 7 8 9 10 ...

$ SN: int 1 2 3 4 5 6 7 8 9 10 ...

$ Prog: char “Civil I" "Civil I" "Civil I" "Civil I" ...

$ Roll: char “2018-1-04-0836" "2018-1-04-0840" "2018-1-04-0846" "2018-1-04-0849" ...

$ Registration: int 19040243 19040247 19040253 19040256 19040257 19040259 19040263

$ Name: char “Anish Shrestha" "Ashok Ban stola" "Baikal Boral" "Bimal Jaisi" ...

$ SLC: num 79.8 83.5 76 85 86 ...

$ School: int 1 1 0 0 0 1 0 1 1 1 ...

$ Plus: int 78 74 73 76 76 74 66 66 66 69 ...

$ Physics: int 85 76 75 80 80 76 83 83 83 56 ...

$ Math: int 72 84 75 90 90 78 62 62 62 68 ...

$ Chemistry: int 85 76 84 70 70 70 56 56 56 61 ...

$ Bio: int 85 75 70 70 70 80 48 48 48 0 ...

$ Parent1: int 1 4 3 2 2 4 1 1 1 2 ...

$ Parent2: int 3 4 4 4 4 4 4 4 4 3 ...

$ X1: int 44 47 49 40 50 42 36 44 45 46 ...

$ X2: int 39 39 44 41 41 36 40 44 44 45 ...

$ X3: int 39 35 39 48 48 44 32 33 40 40 ...

$ X4: int 44 41 43 36 38 36 38 46 42 45 ...

$ X5: num 42 48 39 36 45 46 40 39 42 39 ...

$ X6: int 42 40 41 32 42 34 42 32 47 47 ...

$ Roll: int 19040243 19040247 19040253 19040256 19040257 19040259 19040263

$ X11: char “A-" "B" "A-" "A" ...

$ X22: char “B" "A-" "A" "A" ...

$ X33: char “B+" "C+" "A-" "B+" ...

$ X44: char “A" "B+" "A-" "C+" ...

$ X55: char “A" "B-" "A" "C-" ...

$ X66: char “B" "B-" "A-" "A-" ...

$ School.1: char “Private" "Private" "Govt" "Govt" ...

$ Parent1.1: char “SLC" "Master" "Bachelor" "2" ...

$ Parent2.1: char “Bachelor" "Master" "Master" "Master" ...

$ Gap: char “A-" "B+" "A" "B+" ...

$ GPA: num 3.45 3.01 3.82 3.26 3.48 2.48 2.53 3.36 3.38 3.59 ...

$ Rill: int 19040243 19040247 19040253 19040256 19040257 19040259 19040263

$ X111: int 85 75 85 90 85 65 60 80 85 90 ...

$ X222: int 75 85 90 90 85 75 70 90 90 90 ...

$ X333: int 80 65 85 80 75 60 70 75 80 70 ...

$ X444: int 90 80 85 65 90 70 65 75 65 65 ...

$ X555: int 90 70 90 55 85 65 70 70 85 85 ...

$ X666: int 75 70 85 85 70 65 70 85 70 90 ...

The above structure of data dimension described the total row and columns (580\*49) the field data structure and their properties with various pattern of research data. The following points need to be considered while doing so.

**Sub Setting Data**

After loading tidverse and deplyr package the real independent variables which have many associative variables to predict student grade could be easily grouped up using S % > % select (GPA, SLC, School, Plus, Physics, Math’s, Chemistry, Bio, Parent1, Parent2, X1, X2, X3, X4, X5, X6, X111, X222, X333, X444, X555, X666) while doing so we may use verities of category like whose GPA is grater then 0 indicates only passed students for model development %>% filter (GPA>0).

The machine learning model sometimes always necessary to convert selected values needs to be in normalization because this column-wise normalization makes the values ranges six intervals using as.data. frame (apply (ss [, 1:22], 2, function (x) (x – min (x)) / (max (x)- min (x)))) structure formed the normalization table with 22 independent variables using sub setting. Whose summary values like min, max, median,1stquatrail 3rd quartile were displayed using summary (Sr) function as.

GPA SLC School Plus Physics Math Chemistry

Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000 1st Qu.:0.0000 1st Qu.:0.4444 1st Qu.:0.0000 1st Qu.:0.4722 1st Qu.:0.2973 1st Qu.:0.4423 1st Qu.:0.3882 Median :0.2487 Median :0.6694 Median :1.0000 Median :0.6667 Median :0.5405 Median :0.5962 Median :0.5263 Mean :0.4073 Mean :0.6072 Mean :0.5862 Mean :0.6031 Mean :0.5161 Mean :0.6002 Mean :0.5190

3rd Qu.:0.8244 3rd Qu.:0.7694 3rd Qu.:1.0000 3rd Qu.:0.7778 3rd Qu.:0.6757 3rd Qu.:0.7692 3rd Qu.:0.7171 Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000

The write.csv (Sr, file ="D:/dataanalysisfinal/SOE/rcode/SSeeeee.csv") command is used to again rewrite the normalized data specific location for future purposes. Then the Boruta algorithm applied to all 22 selected independent variables for the prediction of student GPA score with eight categories stored in Boruta=Boruta (GPA~., data = Sr, doTrace = 2, maxRun = 400). This model runs 400 iterations to converse the output for the finding out important, tentative important, and non-important categories to produces the best output with 1. run of importance source... 2. run of importance source...12. run of importance source... After 12 iterations, + 5.8 secs: confirmed 14 attributes were Bio, Chemistry, Physics, Plus, X1 and 9 more; rejected 1 attribute: X3; still have 6 attributes left.398. run of importance source... 399. run of importance source. This process print (Boruta) out of 21 predictor variables declared 17 conformed important, 2 unimportant, and 1 tentative in with 2.61 minutes after runs on 400 iterations (Boruta performed 399 iterations in 2.757619 mins). Therefore, out of 22 independent variables the algorithm declared with similarities value to formed the model 17 attributes confirmed important were SLC, Plus, Physics, Math’s, Chemistry, Bio, Parent1, X1, X2, X4, X5, X6, X111, X222, X333, X444, X555, and X666. Similarly, 3 attributes confirmed unimportant were Parent2, School, X3, finally, 1 tentative attribute left is X5 variables.

The plot described the three categories of variable the red is tentative whereas the blue is not significant and the rest 17 plots were the very significant association to predict model using plot (Boruta, las=2, cex. axis=0.7) command.

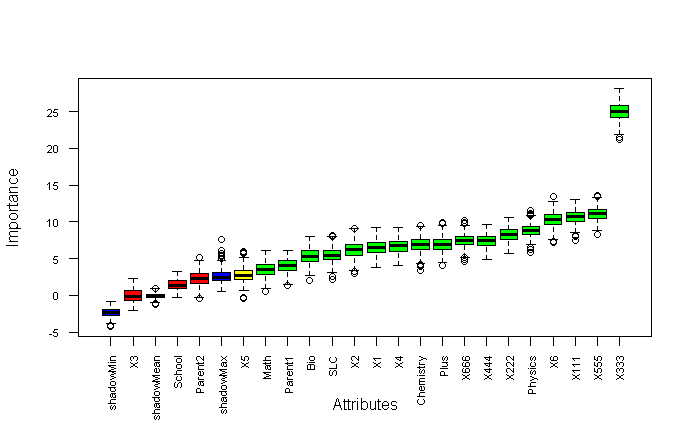


Figure 9: Output of Boruta Algorithm

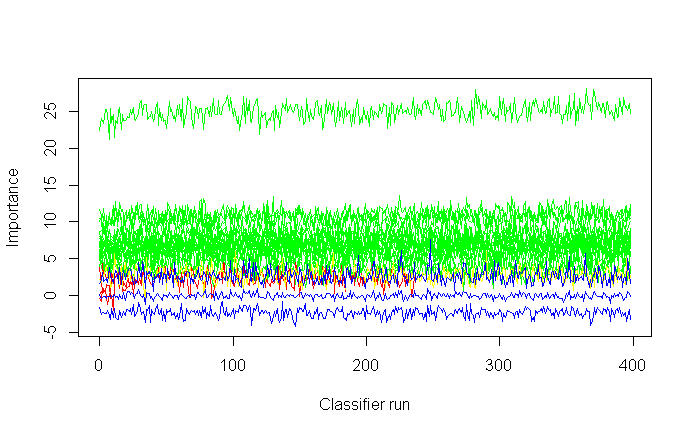


Figure 10: Important, Tentative Important and Non important Plot

From the above graphs, it is concluded that there were 17 variables significant for the model development plotImpHistory (Boruta). When using tentative fixed test further reject this bor=TentativeRoughFix (Boruta) rejected four variables unimportant Parent2, School, X3, X5. This type of rejection of input variables describes statists of all variables using variables using attStats (Boruta) command as.

meanImp medianImp minImp maxImp normHits

SLC 5.523187753 5.4931231 2.1647344 8.133079 0.97243108

School 1.424280858 1.3000044 -0.3183249 3.328378 0.01002506

Plus 6.956021117 6.9350844 4.1107703 9.965286 0.99749373

X555 11.066508610 11.1027908 8.2696448 13.640935 1.00000000

X666 7.456300397 7.4291433 4.6689392 10.213021 0.99498747

Which finally tells to concludes the decision that the variable SLC is confirmed to required whereas School information of either private or government had least predictability to GPA become rejected, X5 tentative and plus confirmed to all independent variables. Finally, the getConfirmedFormula (Boruta) finalized 17conformed GPA ~ SLC + Plus + Physics + Math + Chemistry + Bio + Parent1 + X1 + X2 + X4 + X6 + X111 + X222 + X333 + X444 + X555 + X666 seventeen variables out of 22 independent variables to the model.

## **Random Forest Methodology for Large Feature Selection**

Random forest methodology aggregate trees which can be applied in both classification as well as regression for avoiding overfitting large features selection which can deal with large numbers of features helps to reduced feature selection based on importance of features and user friendly because only two parameters would be given tree and mtry variable as hyperparameter. The default tree is 500 and mtry is 4 initially. The hyperparameter mtry is sampled as a parameter at each split default sqrt (p) features for classification and p/3 for regression based on predictor variables. Where p is features of ntree bootstrap sample where each bootstrap sample grows on the unpruned tree by choosing the best split based on a random sample of mtry predictor at each node and predict new data using majority votes for classification and average for regression based on ntree for final prediction of the model. After loading library caret, random forest in r console the database is again selected using data=read.csv ("D:/dataanalysisfinal/SOE/rcode/ Dataanalysiscsv.csv") command and split its data-based into the working table using tidyverse packages. While doing the ss=S %>% select (GPA, GPAA, GPAAA, SLC, School, Plus, Physics, Math’s, Chemistry, Bio, Parent1, Parent2, X1, X2, X4, X5, X6, X111, X222, X333, X444, X555, X666) % > % filter (GPA>0) is used for filtering only those records of passed student 290 final result of engineering student. Then after omitting na values and converting response variable with factor category as table command describes as table(S$GPAA) produced as A (49) A-(96)B+(65)B(39)B-(35)C+(5) and C-(1). Similarly, the table(S$GPAAA) a real factor data of student grade of 2019 result describes with their groups as.

 A A- B+ B B- C+ C

 49 96 65 39 35 5 1

After making seed and data frame of selected data items of 70 and 30 percent split of data sets with test and training data sets split of 211 and 79 data sets of passed students using sample (2, nrow (S), replace=T, prob=c (.7,.3)), train=S [ind==1,] 211 for a pass only and test = S [ind == 2,] 79 out of 290 passed student for the model validation and testing accuracy purposes.

The random forest provides only two free parameters for the segregation of a large number of features of 22 variables. The rf22=randomForest (GPAAA~., data=train, nodesize=5, ntree= 500). Here the GPAA a categorical variable with numerical representation of all student grades 1 for A,2 for A-,3 for B+,4 for B,5 for B-,6 for C+,7 for C- and 8 for Fail categories in GPAA and vice versa in GPAA variable. All variables model prediction and confusion matrix were based on OOB (Out of Bag) indicates error rate is 2.37% indicates 97 % accuracy of the model to predict student grade. The print (rf22) commands plot in mtry chart which indicates the errors has drastically decreased after 3.2 steps.

A A- B+ B B- C+ C

49 96 65 39 35 5 1

After making seed and data frame of selected data items of 70 and 30 percent split of data sets with test and training data sets split into 211 and 79 data sets of passed students using sample (2, nrow(S), replace=T, prob=c (.7,.3)), train=S[ind==1,]211 for a pass only and test=S[ind==2,] 79 out of 290 passed student. For model validation and testing accuracy purposes.

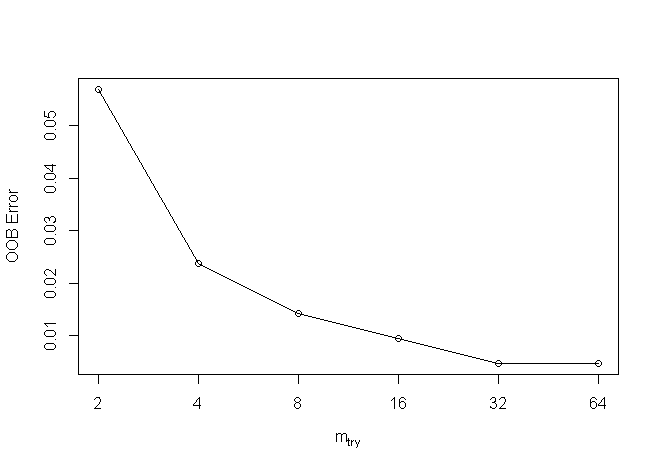


Figure 11: OOB Plot of Random Forest

The random forest is a classification model with the number of trees is 500, of variables, tried at each split is 4 and OOB estimate of error rate is 2.37% indicates confusion matrix with errors as.

Confusion matrix:

A A- B B- B+ C C+ class. error

A 34 0 0 0 0 0 0 0.00

A- 0 70 0 0 0 0 0 0.00

B 0 0 30 0 0 0 0 0.00

B- 0 0 1 24 0 0 0 0.04

B+ 0 0 0 0 48 0 0 0.00

C 0 0 0 1 0 0 0 1.00

C+ 0 0 0 3 0 0 0 1.00

From the above confusion table, the prediction of result in Grade A A-, B, C, and C- has no missing values whereas C+, B- and C has only five-grade prediction values were found miss classified from the model. After loading the caret package, the prediction of the model with train data produced p= predict (rf22, train) the head(p) 99% accurately when matching using first values.

1 3 6 7 9 10

A- A B- B- A- A-

Therefore, this model predicts more accurately in eight levels A A- B B- B+ C C+ when applying confusion matrix with train data confusion matrix and statistics as.

Reference

Prediction A A- B B- B+ C C+

A 34 0 0 0 0 0 0

A- 0 70 0 0 0 0 0

B 0 0 30 0 0 0 0

B- 0 0 0 25 0 1 0

B+ 0 0 0 0 48 0 0

C 0 0 0 0 0 0 0

C+ 0 0 0 0 0 0 3

This output describes the overall statistics accuracy is 0.9953 percent when 95% confidence interval ranged from (0.9739 to 0.9999) percent accuracy is the very strong model. The information rate is 0.3318 is p-value is less than 2.2e-16(0.05) indicates the model has significance. This is similarly explained by using the Kappa test is 0.9939 percent too.

The sensitivity and specificity values in each grade column were calculated based on the confusion matrix as the accurate prediction divided by total misclassification in each grade cluster. However, the grade C prediction has some errors in the model.

Class: A Class: A- Class: B Class: B- Class: B+ Class: C Class: C+

Sensitivity 1.0000 1.0000 1.0000 1.0000 1.0000 0.000000 1.00000

Specificity 1.0000 1.0000 1.0000 0.9946 1.0000 1.000000 1.00000

Pos Pred Value 1.0000 1.0000 1.0000 0.9615 1.0000 NaN 1.00000

Neg Pred Value 1.0000 1.0000 1.0000 1.0000 1.0000 0.995261 1.00000

Prevalence 0.1611 0.3318 0.1422 0.1185 0.2275 0.004739 0.01422

Detection Rate 0.1611 0.3318 0.1422 0.1185 0.2275 0.000000 0.01422

Detection Prevalence 0.1611 0.3318 0.1422 0.1232 0.2275 0.000000 0.01422

Balanced Accuracy 1.0000 1.0000 1.0000 0.9973 1.0000 0.500000 1.00000

As compared OOB errors for each bootstrap iteration and related tree prediction error using data not in bootstrap sample is called OOB is estimated classification accuracy and regression RMSE were selected. Similarly, the confusion matrix and prediction of test data with grade calculated as confusionMatrix (pp, test$GPAAA) function as.

Reference

Prediction A A- B B- B+ C C+

A 15 0 0 0 0 0 0

A- 0 26 0 0 0 0 0

B 0 0 9 0 0 0 1

B- 0 0 0 10 0 0 1

B+ 0 0 0 0 17 0 0

C 0 0 0 0 0 0 0

C+ 0 0 0 0 0 0 0

This implies similar type results with accuracy is 0.9747 percent due to two testing grades were miss classification in C grade with 95% confidence interval in between (0.9115 to 0.9969) percent accuracy for prediction GPA. The information rate is 0.3291 is good with a significant p-value is less than 0.05(< 2.2e-16) of model performance as.

Class: A Class: A- Class: B Class: B- Class: B+ Class: C Class: C+

Sensitivity 1.0000 1.0000 1.0000 1.0000 1.0000 NA 0.00000

Specificity 1.0000 1.0000 0.9857 0.9855 1.0000 1 1.00000

The plot model describes the overall model normalization when it goes more than 200 trees it will normalized constant then after.

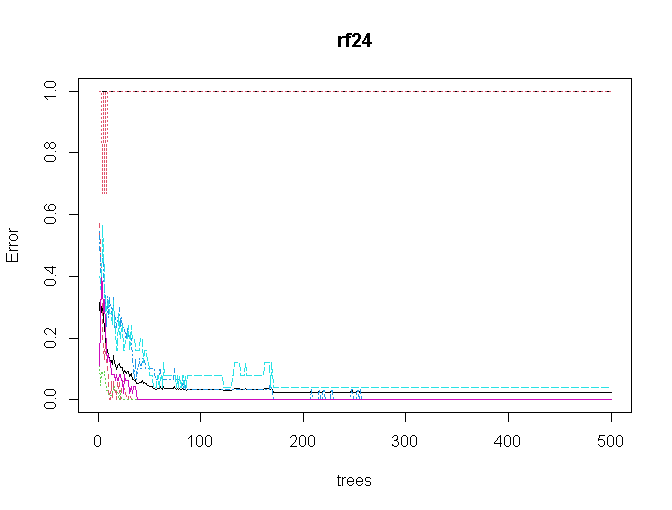
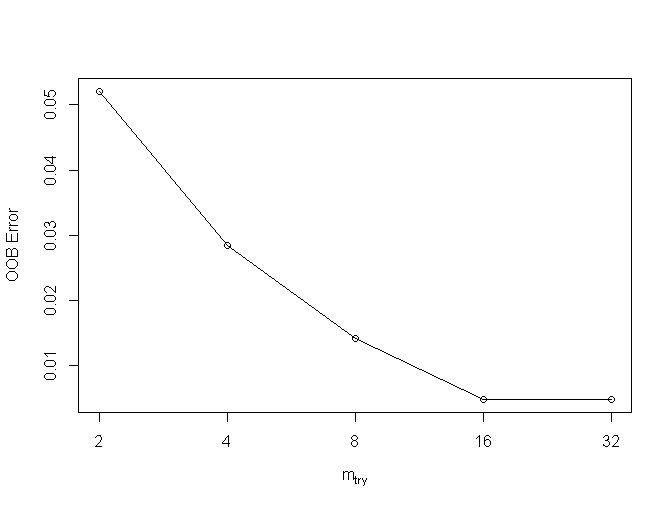
 

Figure 12: Error Vs Trees and Mtry Plot

From the above error and mtry figures the mtry value using finetune using tuneRF (train [, -1, -2], train [,3], step Factor = 0.5 plot=TRUE, ntreeTry = 300, trace = TRUE, improve = 0.05). The improve factor improves in each tree to out-of-bag errors and the mtry inflicted either positive or negative to the model. The second tune plot determine the final value of mtry should be after 16 steps for normalization than after. Therefore, from the above random forest model becomes more accurate for final prediction using ntree is more than 200 and mtry become normalized after 16 with proximity true for the final model as.

rff=randomForest (GPAAA~., data =train, ntree = 200, mtry=16, importance = TRUE, proximity = TRUE). This is finally printed using as print(rff) produces the OOB estimate of error rates 0.47% has largely improvement than earlier whose confusion matrix using confusionMatrix (p, train$GPAAA)

Reference

Prediction A A- B B- B+ C C+

A 34 0 0 0 0 0 0

A- 0 70 0 0 0 0 0

B 0 0 30 0 0 0 0

B- 0 0 0 25 0 0 0

B+ 0 0 0 0 48 0 0

C 0 0 0 0 0 1 0

C+ 0 0 0 0 0 0 3

This confusion matrix predicts 100% accuracy on training data with 98% confidence level with test p value were more less than 0.05 significant statists as.

Class: A Class: A- Class: B Class: B- Class: B+ Class: C Class: C+

Sensitivity 1.0000 1.0000 1.0000 1.0000 1.0000 1.000000 1.00000

Specificity 1.0000 1.0000 1.0000 1.0000 1.0000 1.000000 1.00000

Pos Pred Value 1.0000 1.0000 1.0000 1.0000 1.0000 1.000000 1.00000

Neg Pred Value 1.0000 1.0000 1.0000 1.0000 1.0000 1.000000 1.00000

Prevalence 0.1611 0.3318 0.1422 0.1185 0.2275 0.004739 0.01422

Detection Rate 0.1611 0.3318 0.1422 0.1185 0.2275 0.004739 0.01422

Detection Prevalence 0.1611 0.3318 0.1422 0.1185 0.2275 0.004739 0.01422

Balanced Accuracy 1.0000 1.0000 1.0000 1.0000 1.0000 1.000000 1.00000

Similarly, the test confusion matrix is calculated using confusionMatrix (pp, test$GPAAA) function produces the confusion matrix with more accuracy as.

Reference

Prediction A A- B B- B+ C C+

A 15 0 0 0 0 0 0

A- 0 26 0 0 0 0 0

B 0 0 9 0 0 0 0

B- 0 0 0 10 0 0 0

B+ 0 0 0 0 17 0 0

C 0 0 0 0 0 0 0

C+ 0 0 0 0 0 0 2

Then again produces more than 95 accurate prediction with 95% with confidence interval is (0.9544, 1) which no information rate is 0.3291 when p is less than 0.5 percent < 2.2e-16 significant. However, in grade C prediction there were some errors as.

Class: A Class: A- Class: B Class: B- Class: B+ Class: C Class: C+

Sensitivity 1.0000 1.0000 1.0000 1.0000 1.0000 NA 1.00000

Specificity 1.0000 1.0000 1.0000 1.0000 1.0000 1 1.00000

Pos Pred Value 1.0000 1.0000 1.0000 1.0000 1.0000 NA 1.00000

Neg Pred Value 1.0000 1.0000 1.0000 1.0000 1.0000 NA 1.00000

Prevalence 0.1899 0.3291 0.1139 0.1266 0.2152 0 0.02532

Detection Rate 0.1899 0.3291 0.1139 0.1266 0.2152 0 0.02532

Detection Prevalence0.1899 0.3291 0.1139 0.1266 0.2152 0 0.02532

Balanced Accuracy 1.0000 1.0000 1.0000 1.0000 1.0000 NA 1.00000

After complecting the final model, the histogram of number of nodes could plotted using hist (tree size(rff), main="Number of Todisco="green") function describes that data of nodes of 290 students.

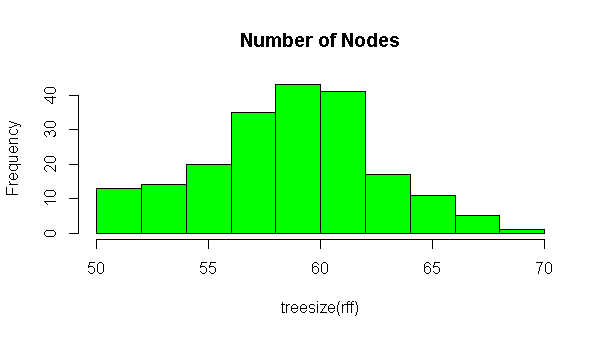
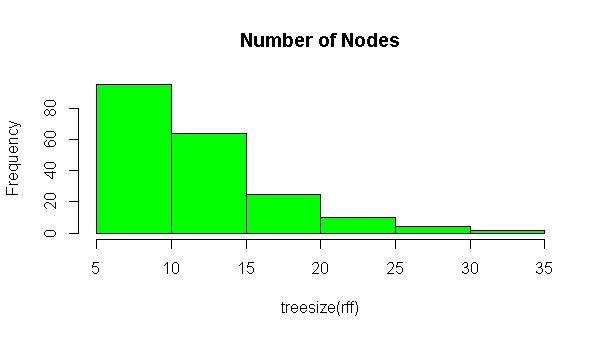
 

Figure 13: Nodes and Number of Trees Histogram

The first histogram describes the number of trees in 6 to 8 trees with 80 nodes whereas there are 4 to 6 trees with 40 nodes and few trees were in14 to 16 nodes there are overall 290 nodes of tree distributed in between 4 to 18 nodes of passed students in 2019 with their mark’s weightage.

The varImpPlot (rff, sort=T, n.var=21, main="Variables Weights ") function plot two plots explain the overall variable importance for the model significant of variables with their contribution percent accuracy so that researcher may deselect the variable in model forming.

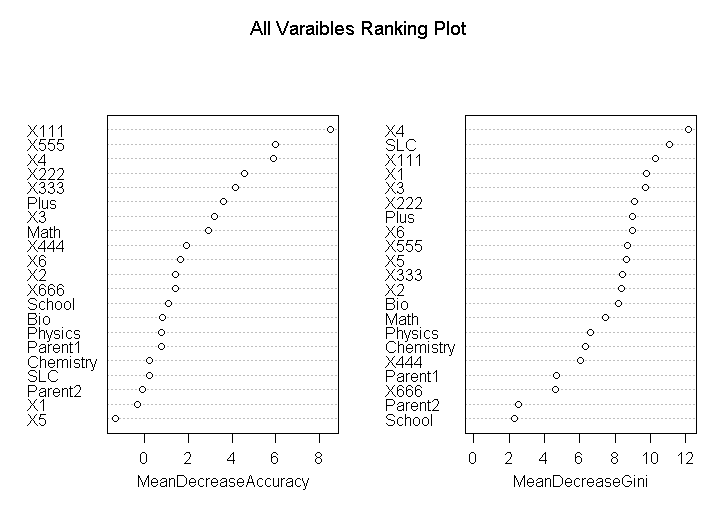


Figure 14: Mean Decrease Accuracy and Ranking Plots

This first plot describes the total accuracy due to the presence of specific variables and the second plot described its meaning decreases. This plot output describes similar type result with Boruta algorithm above that tentatively rejected some independent variables as X5, X1, and Parent2 variables because of very low rank has zero significance results around zero percent. however, the first plot describes its contribution for forming the model at respective weightage. When considering without final GPA of each 6 grade the X4 and Plus2 and X3 results have the significance of having more than 4 percent weighted of a single variable. Therefore, it is concluded that to score a good GPA bachelor's grade based on the final grade of the first subject has very significant implies Mathematics grade of plus two has strong positive results. The grades X111(Math’s) and X555(programming) subjects of final grade have the strong role for scoring the highest Bachelor grade. When considering with internal grade only X4 core engineering subject X3 has the greatest role for scoring Bachelor grade highest grade. The SLC grade has less significant than the Plus grade and the Mathematics scored of plus2 than Chemistry, Bio, and Physic scores were in low contribution for bachelor’s grade prediction. Similarly, the second plot describes the Gini how pure the tree decreased when removing the associative variables. The X4 internal grade and SLC his significant association for decreasing Gini model. Whereas school and parent qualification have no much association for predicting GPA model. Similarly, the importance(rff) model describes its percentage of all associative variables. The importance(rff) produces the table with the respective percentage that could describe and chosen the associative variables.

A A- B B- B+ C C+ MeanDecreaseAccuracy MeanDecreaseGini

SLC -0.222953 1.045989 -0.85129 0.751304 -0.72162 0 0.0000 0.18854 11.0624

School -0.317582 1.803325 2.08448 0.308938 -1.02065 0 0.0000 1.07695 2.3114

Plus -0.351939 4.043537 -0.82929 1.345535 1.02309 0 -1.0025 3.61981 00269

Physics 0.796682 0.889672 1.56792 -0.770135 -1.33302 0 1.0025 0.77283 6.6280

X555 8.907967 1.248824 0.18731 0.23283 1.37419 0 1.0025 5.96817 8.6822

X666 0.617318 0.283206 -0.19839 2.49528 0.1725 0.0000 1.39306 4.6537

Similarly, the function versed (rff) describes first to end times occurred in the model of all variable times in second as 709 187 607 534 545 509 598 361 248 715 667 698 722 654 669 643 597 552 452 521 362 iteration to formed the final model. These highest values indicate its highest associativity to model the second previous school information has least association for marks prediction of higher education scoring with good grade.

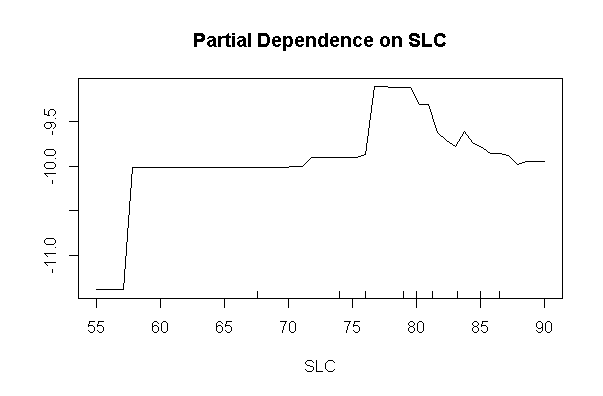
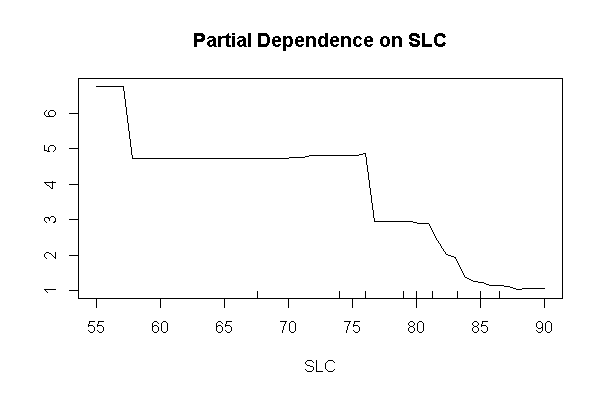
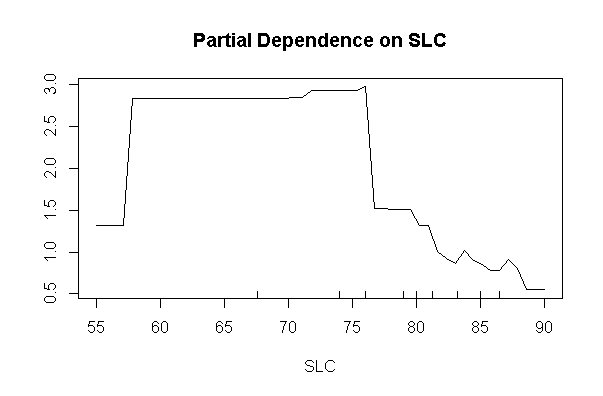


Figure 15: Partial Dependency Plots

The partial dependency plot describes the marginal effect of each category of special class could be easily when partialPlot (rff, train, SLC,"A") partialPlot (rff, train, SLC,"B") partialPlot (rff, train, SLC, "C") predicts the probability value when SLC grade. The first plot implies that it is more chance when SLC score is 75 he/she could score grade A in Bachelor of engineering. Likewise, when SLC grade is below 80 could score B in bachelor degree and when SLC grade is in-between 60 to 75 he/she could score C grade in bachelor degree. Similar type output could be prediction to all 22 independent variables could be easily predicted.

Similarly, the highest importance variable to formed the model is mathematics subject whose association with different variable described in partial plots to indicates bachelor grade plotted partialPlot (rff, train, X111,"A") partialPlot (rff, train, X111,"A-") as.

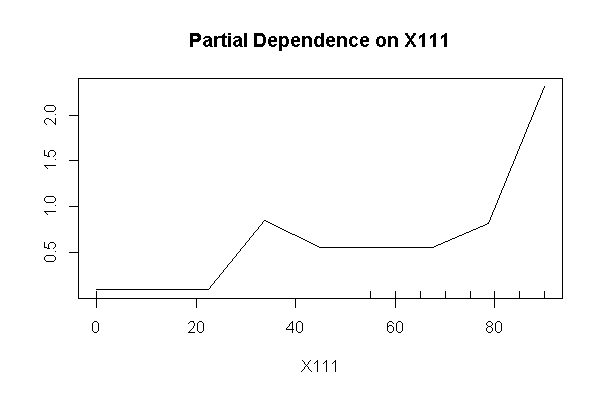
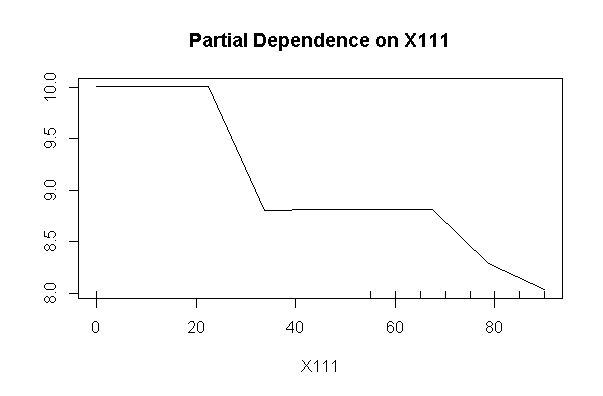
 

Figure 16: Partial Dependency Plot of Internal Grade

The first partical dependency plots when mathematics marks to predict bachelor grade scored to get A grade in final. Which explain that those student whose scored have greater than 80 percent in their mathmatics had higher chance to scored A grade as compared other marks in mathmatics in bachelor final result. Similarly the second plots implies same mathmatics marks however another association to predict B- might less probability of scoring B- bachelor grade. This command plots all independent combination of data all possible grades of students.

Similarly, the random forest terminal node could be visualized as getTree (rff,1,labelVar = TRUE) function. The first tree started at -1 status indicates terminal node whose left daughter and right daughter were absent and so on.

left daughter right daughter split var split point status prediction

1 2 3 GPA 3.710 1 <NA>

2 4 5 GPA 3.305 1 <NA>

3 0 0 <NA> 0.000 -1 A

4 6 7 GPA 3.005 1 <NA>

5 0 0 <NA> 0.000 -1 A-

11 0 0 <NA> 0.000 -1 B

14 0 0 <NA> 0.000 -1 C+

15 0 0 <NA> 0.000 -1 B-

Similarly, the mean density plots MDSplot (rff, train$GPAAA) plots describes the whole trees values in scatter plot with differ colors of differed with clustering GPA grade of student.

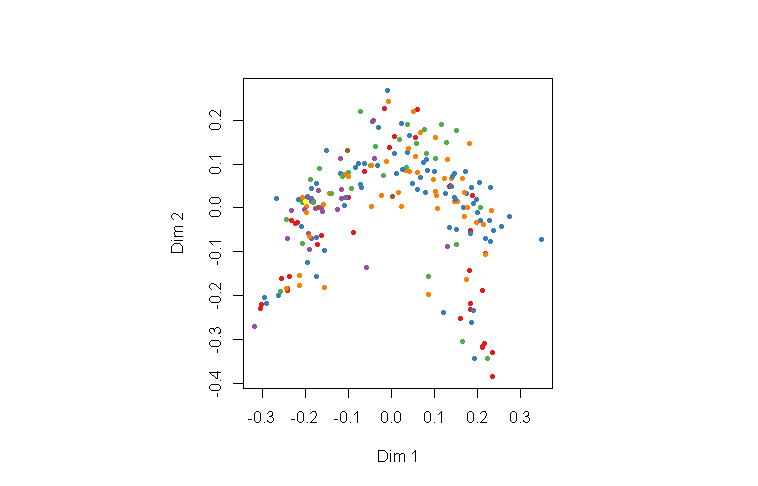


Figure 17: MDS Plot of All GPA Points

The color indicates the GPA grades. Therefore, random forest methodology here is apply for data preparation with their respective explanation to conclude research.

## **Before Examination Grade Prediction**

As we know that the new examination is the compositive mixture of previous grades and courses completed by the student from academic institution. This means that the internal and previous related academic achievement may either play a positive role to score in the upcoming grade. The final grade is the mixture of internal grade and previous scores this study eagerly tries to find out the association of (background information) SLC grade, Plus2 grade, either school from private or government, their major score like physics, bio, chemistry, and mathematics association with internal score grade score of their six ongoing subjects. Therefore, the new random forest model tries to predict with categorical variables. While doing so the sub setting the database training and test splits were designed for before model execution. The label command of dependent variables with (“A" "A-" "B" "B-" "B+" "C" "C+"). The first model final= randomForest (GPAAA ~., data = train) produces the accuracy prediction. After loading the caret package, the training prediction predicts the accuracy is 97 percent on training data sets with significant with 98 percent confidence interval as.

Reference

Prediction A A- B B- B+ C C+

A 37 0 0 0 0 0 0

A- 0 62 0 0 0 0 0

B 0 0 23 0 0 0 0

B- 0 0 0 23 0 0 0

B+ 0 0 0 0 50 0 0

C 0 0 0 0 0 2 0

C+ 0 0 0 0 0 0 2

Similarly, the test data confusion matrix produces 96 percent accuracy with 99 percent confident interval. This implied that model is good enough, however, the grade C prediction with some minimum mismatched prediction. The tunerf function predicts at 10 m try and total trees with 400 in the model design. Similarly, the histogram of tree size indicates there were 55 trees with 80 nodes. The varImpPlot describes all variables associated to form the above model.

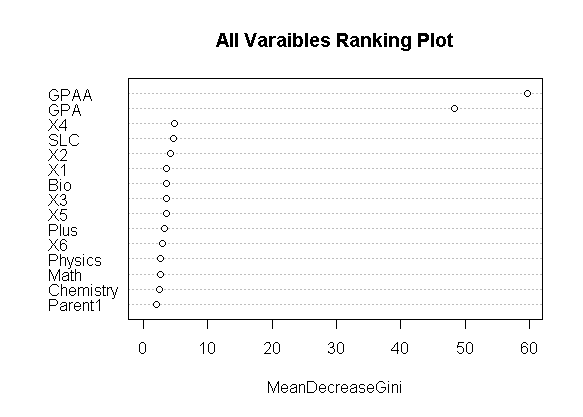


Figure 18: Mean Decrease Gini Plot

The mean decrease Gini plot describes in its percentage of increase probability however this model needs to exclude with GPAA and GPA variable for final prediction this could predict information is exactly matched with var used model. Whose accuracy output is further evaluate using the partial plot describes each prediction to achieve any prescribed final grade GPA.

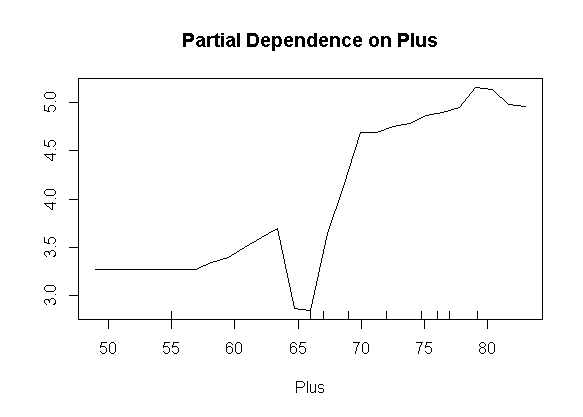
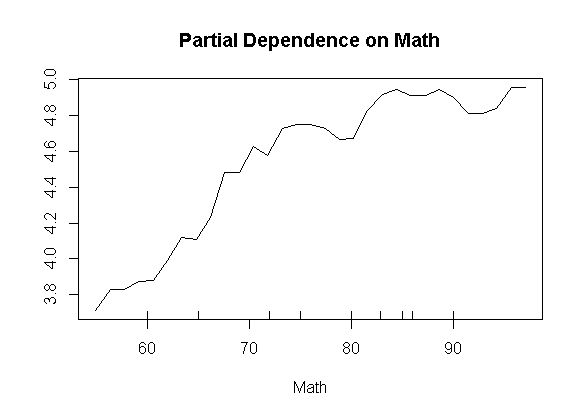
 

Figure 19: Partial Plot of Math’s and Plus2 Grade

From the first partial dependency plot indicates a positive association to predict grade A is when evaluating plus to scored. Similarly, the plus to mathematics score has more potential to predict bachelor grade with A grade. Similarly, the get tree function that describes the root and left and right daughter of each node sides.

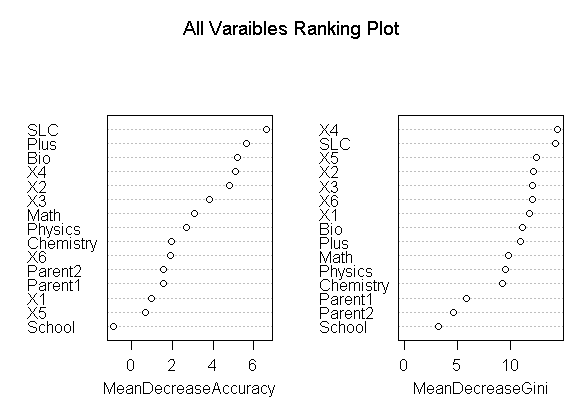
However, the final model after removing the unimportant columns describes using randomForest (GPAAA ~SLC+ School+ Plus+ Physics +Math+ Chemistry+ Bio+Parent1+ Parent2+ X1 +X2+X3+X4+X5+X6, data=train, ntree=500, mtry=4, importance=TRUE, proximity=TRUE) with only previous values. Its confusion matrix of training sets with 98 percent accuracy significant however the test accuracy has not signed with accuracy is lower than percent. Whose mtry become 10 with 80 nodes tree within the top level of 80.

Figure 20: Mean Accuracy and Gini Decrease Plot

The varImpPlot command describes the highest to lowest association form the final model with the percentage of decreasing accuracy when eliminating each predator variable to the model. Thus, when comparing both mean dependency plot indicates accuracy and when eliminating each input variables its model accuracy decrease percentage respectively. Therefore, it is concluded that the internal grade of SLC, plus two percentage, Bio X4, X3, Math has strong non negated components to model formation. Whereas student from private school or government schools has the least role to predict bachelor grade when using random forest prediction. Thus, the school and parent2 education information could be removed for better prediction. Similarly, the partial plot of six internal grade plots describes the similar association to bachelor grades as.

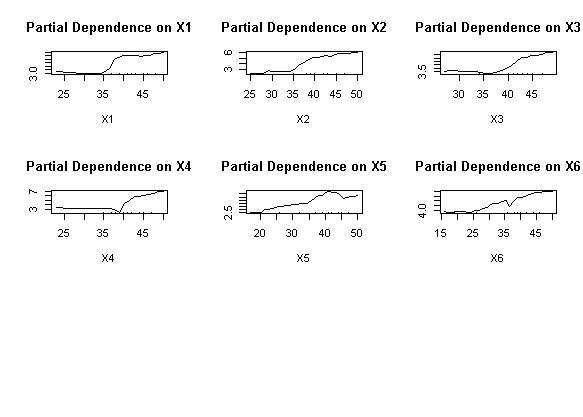


Figure 21: Partial Plot of Internal Marks

When calculating partial plots only with internal marks of student. The first partial plots predict to score A grade in final grade has only positive association when internal marks are greater than 35 towards 50. Similarly, however the column of X5 internal marks could score A grade when internal marks with around 40 but not above 45, when student scored more than 45 had -ve association to scored grade in A in bachelor grade accurately predicted with bachelor grade.

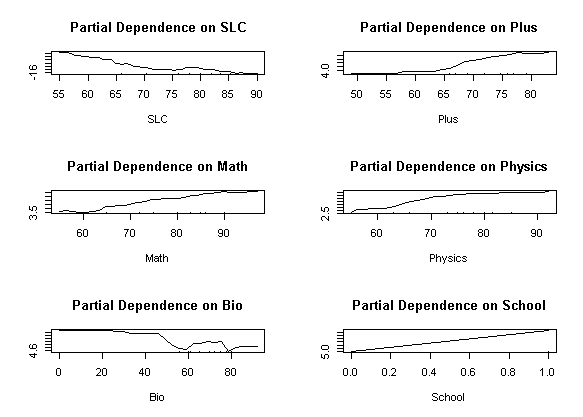


Figure 22 : Partial Plot of Previous Grade

Similarly, when using previous grades of partial plots of SLC to GPA A grade has advised when SLC having 75 and 85 has same potential to predict A grade. The Plus two score has positive increased after 60 in similar incasement to predict grade incasement the bio prediction implies that those who scored 80 in biology have less chance to scored grade A as compared 40 or not having bio as major. The majority those who don’t have bio scored A grade. And the graduation form school has linear association implies unimportant to the model. Thus, the SLC grade to predict bachelor A has -ve association whereas plus two, Math, Physics, has the positive score and the Bio and School of private and government has less prediction to Grade. Similarly, we could easily predict all eighth-grade bachelor students.

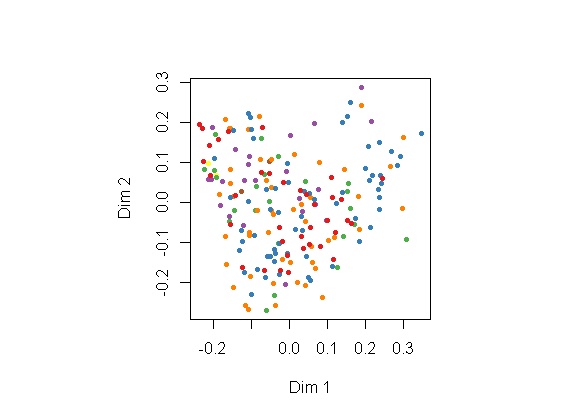


Figure 23: MDS Plot of Grades

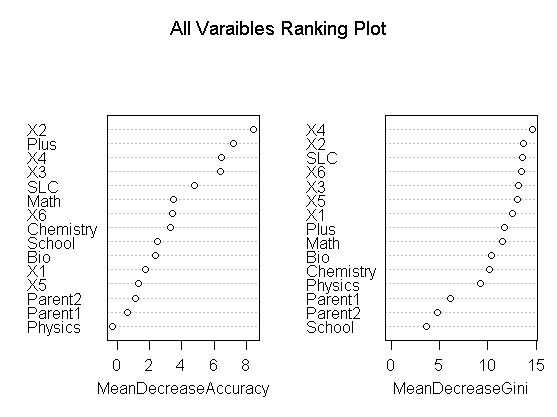
Similarly, the mean density plots the color GPA different grades to display all variables used when predicating before examination independent components. Similarly, the internal marks with score rank ("1" "2" "3" "4" "5" "6" "7") produces similar output using final=randomForest (GPAA ~., data=train) with considering original model and another final=randomForest (GPAA ~SLC+ School+ Plus+ Physics+ Math+ Chemistry+ Bio+ Parent1+ Parent2+ X1+ X2+ X3+ X4+ X5 + X6, data = train, ntree = 500, mtry = 4, importance = TRUE, proximity = TRUE) with correction more than 95 percent accuracy with significant result. However, the second corrected result produces 98 % to 37 percent for final accuracy of both test and training data describe as. 

Figure 24: Mean Accuracy and GINI Decrease Plot

This time model predicts school and physics marks of internal evaluation has least accuracy to contribute GPA grade where as X2 and plus2 grade have highest contribution to formed the model A grade. The prediction of values is as.

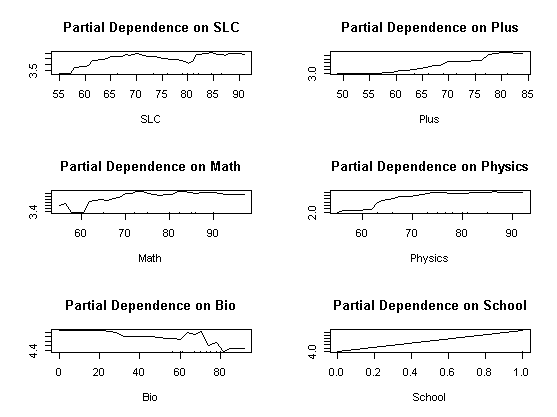
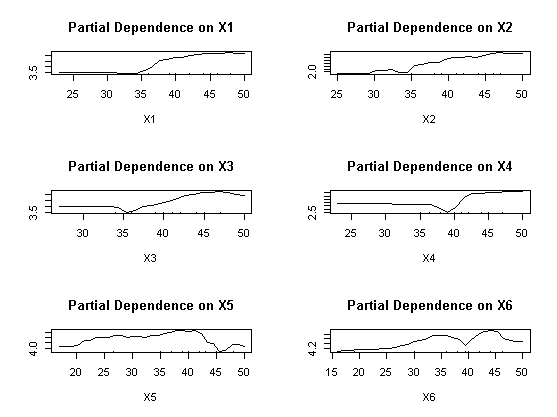
 

Figure 25: Partial Plot of Internal Grades

From the above partial plot predict the second grade A have linear association after 60 % percent of SLC, plus2 Math and Physics of however Bio and internal marks have better A prediction when scored less than 20 in their high school background. Similarly, the mathematics second major subject of internal has a positive association when marks were around 40 to 45 only have positive association. These partial plots concludes that marks scored internally in 40 predict A- grade largely accurate than other marks to other grades.

## **Plus, two Back Propagation Grade Prediction**

The random forest methodology is used for the prediction of student grade here researcher tries to predict old plus2 results using random forest with including bachelor grade using the real concept of backpropagation methodology. The plus2 grade of student having less category as compared bachelor 10-grade point in Nepal. After converting respective grade marks into grade point using grade if (grade >90 A, if(grade>80, if (grade, A- if(grade>70, B+, if(grade> 60, B, if (grade >50, C+ if(grade>40, C, if(grade>25, D else E. Which is based on exactly accordance with higher school centerboard grading criteria in Nepalese system. From the data the different histogram of bachelor and plus2 grading summarized as.



Figure 26: Histogram of Plus to and Bachelor Grade Distribution

According to Nepalese plus2 grading points, letters grading is A, B, B+, C, C+, and D+. although the D+ student could not be getting enrollment in the engineering program. Similarly, the grade points of Bachelor of engineering students have 10 scale grade points like (A, A-, B+, B, B-, C+, C, C-, D+, D- and F grades. However, there were not scoring students in their final bachelor students. The backward propagation of internal marks of bachelor degree were used to predict plus2 was apply for model validation development produced the confusion matrix as.

A A- B B- B+ C C+

A 3 17 3 4 6 0 0

B 20 23 9 9 24 1 2

B+ 24 46 17 14 29 0 1

C 0 1 0 0 1 0 0

C+ 5 6 8 8 7 1 1

This cross-tabulation presents the records of Bachelor final grade plus2 grade of sample data. After sub setting selected columns of these information whose scored grade in the final examination of plus to and their respective final grade in random forest model validation with test and training split of 70 and 30 percent of 290 students. When using final=randomForest (PlustoGrade ~., data = train) model produced to its 79 to 98 percent accuracy. Although this concept is very retrospective thinking

Reference

Prediction A B B+ C C+

A 7 0 1 0 0

B 0 14 5 0 0

B+ 4 6 35 0 2

C 0 0 0 0 0

C+ 1 1 0 0 13

The final tree and mtry plots determine after 200 tree and with 3 mtry the output became constant. Similarly, mtry become 3 for the final model for better increasement of prediction to the model. The histogram was perfectly bell-shaped curved indicates that samples were in uninformedly distributed while model prediction.

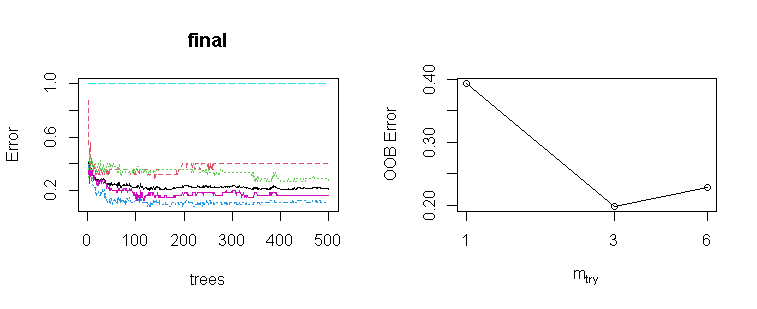


Figure 27: Trees and Mtry Plot

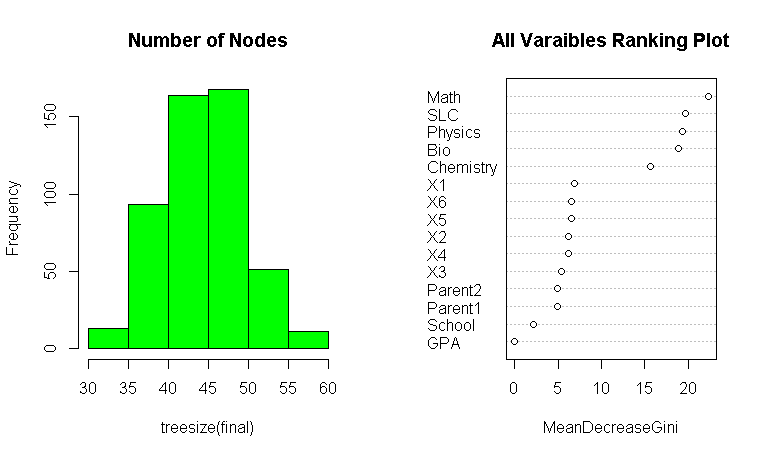


Figure 28: Histogram Nodes and Gini Decrease Plot

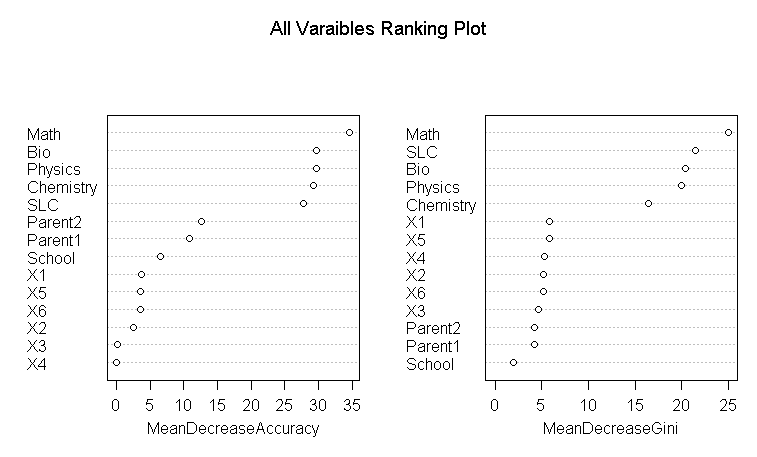
From the mean decrease Gini plot the Math and SLC, Physics have greatest dominant attributes for prediction plus2 grade using backward propagation whereas parent education, school’s information and final GPA were least significantly formed random forest model. Thus, final model final=randomForest (PlustoGrade ~ SLC + School + Physics + Math + Chemistry + Bio + Parent1 + Parent2 + X1+ X2+ X3 +X4 +X5 +X6, data= train, ntree = 500, mtry = 3, importance = TRUE, proximity = TRUE). Which produced the more accurate output 97 and 77 percent accuracy suggests mtry at 3. Therefore, the ultimately the mean decreased plots were plotted as 

Figure 29: Mean Decrease Accuracy and Gini Decrease Plot

The variable significance while using backward propagation mean decreased accuracy plots indicates that Mathematics SLC, and Physics Chemistry had highest percentage contribution as compared parent2 and school information to predict plus2 grade. Similarly, the bachelor’s internal marks would not have least significant than 5 percent for prediction plus to grade.

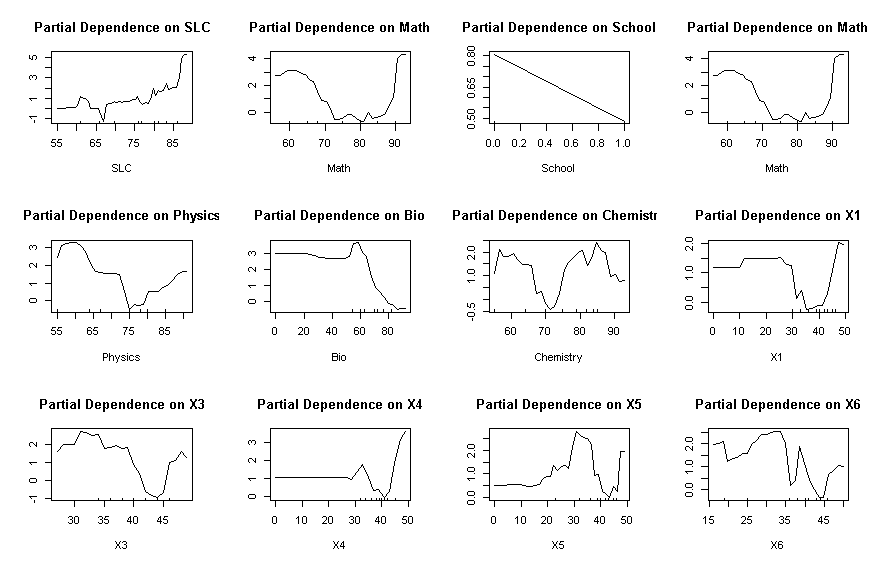


Figure 30: Partial Dependency Plot of Plus2 Grade

Similarly, the partial plot finally implies that the bachelor’s internal grades won’t have a positive association to predict final grade of plus2 however Math, SLC Physic and Bio, and Chemistry score were mixed perdition to predict plus2 A grade.

**Neural Network Result Prediction using Single Layer**

The fully connected neural network conversed at fixed point to predict dependent variable after calculating its weighted sum with bios in each subsequent layer. The result of student grade is either fail or passed a binary numerical variable of all grades before final results. The result of passed student and their high school, intermediate levels their internal and their final grade of each student are the independent variable to predict whether the student has passed grade or not. After loading the library tidyverse packages the subset of data with 21 interdependent variables were passed to the neural network model after normalized independent variables sample 580 students with 401 as training and 179 are testing data sets. The neural network n=neuralnet (pf~ SLC +Plus +Physics+ Math + Chemistry+ Bio+ Parent1 +X1+X2 +X3+X4+ X5+X6+ X111+ X222+ X333+ X444+X555+ X666, data=training, hidden= c (1), err.fct="ce", linear. output= FALSE) produces the plots as.

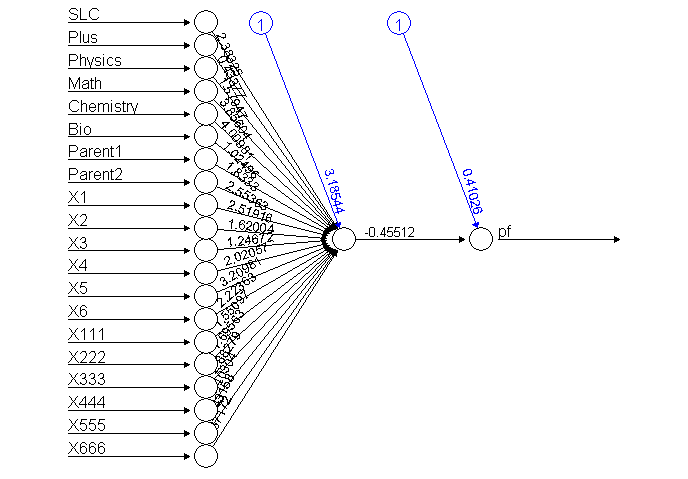
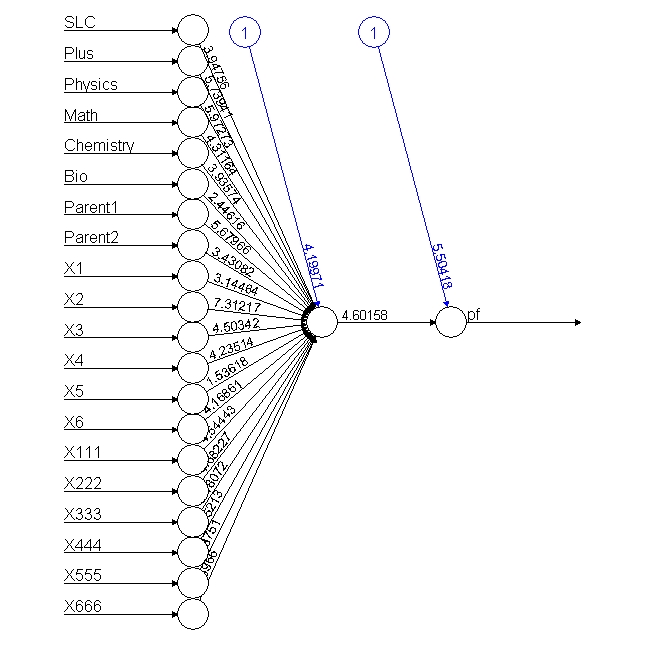
 

Figure 31: Neural Network of 21 Independent Variables

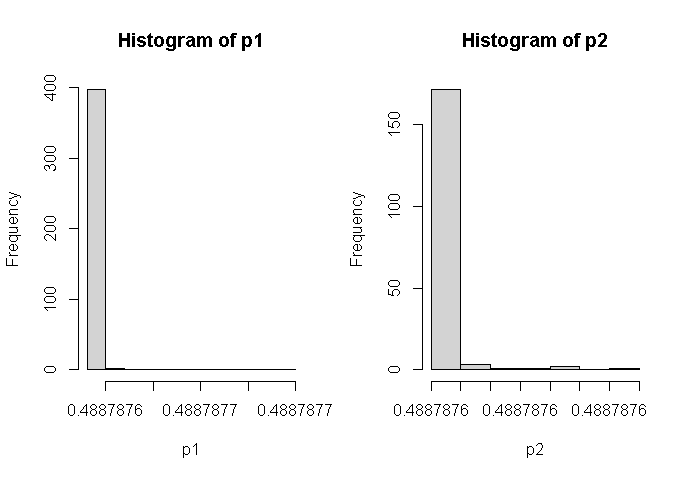
After computing model, the value generated at end node were plotted separately in two histograms with training and test data, the prediction confusion matrix with output data plots the histogram as. 

Figure 32: Histogram of Conversed Output Values

The test and training data output are further categorized with the pass and fail of 401 training data predicts 205 and 196 with using 1-sum(diag(tab1))/sum(tab1) produces the model accuracy is 0.4887781 percent. Similarly, the test 179 data sets calculate with pred2 variables of passed and fail student (both) is 85 to 94 using 1-sum(diag(tab2))/sum(tab2) function. The miss classification errors are better scored 0.525139. Here the result of each student based on 21 independent variables has needed of model improvement using various hyperparameter.

Similarly, the data only using passed student of 290 records apply to predicts the above model produces 100% accuracy on both model with the model of 20 independent variables as n= neuralnet (pf~ SLC + Plus + Physics + Math+ Chemistry+Bio+Parent1+ Parent2+X1 +X2+X3+ X4+X5+ X6+X111 +X222 +X333+ X444+ X555+ X666, data = training, hidden = c (1), err.fct ="ce”, linear. output = FALSE). The confusion matrix prediction with produced using 209 passed student in training data sets with evaluation 1-sum(diag(tab1))/sum(tab1) produces 100 accurate results. Similarly, the accuracy of testing data predicts 81 data of sample 100 percent accurately plots in histogram because all passed student predicts accurately using half model.

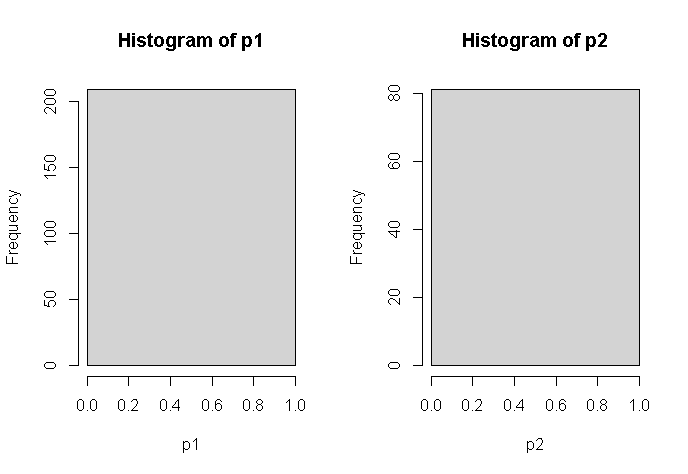


Figure 33: Histogram of Half Model Passed Student

From the above histogram of model produced 100 accurate prediction when using 21 independent sample when using passed student of 290 student records.

## **Neural Network of Pass/Fail Prediction using Single Layer**

As we know neural network model works only in numeric variables, the dependent variable GPA represents with 0 for Fail,1 for A-grade,2 for A- Grade, 3 for B+ grade, and so on having 8 categories respectively is single dependent variable of 20 predictor variables. Based on predictor variable this model used error factor as cross-entropy with the large number of dependent variables which includes high school score marks, plus two marks, school either from private or government, highest academic of their parent1 and parent2 and six internal grades of running score in bachelor degree and their final scored are the independent inputs for prediction to final grade GPA of student.

After loading libraries neuralnet, tfruns, keras and tensorflow, dplyr packages in r console. The read csv command is used to load data in programming workspace. The tydeverse package is used to select dependent and independent variables for the purpose of selecting 20 independent variables from the student database. Here the explanatory result GPA variables having eight final grade score grade of result. After excluding predicator variable all other independent variables were normalized using min max function and whose output is again converted into data frame for analysis purpose to another variable as Sr=as.data. frame (apply (ss [-1, 1:20], 2, function (x) (x - min (x)) / (max (x)- min (x)))). After setting seed value in two samples of training and test are determined using sample with 70 and 30 percent sample partition of data as ind = sample (2, nrow (Sr), replace = T, prob=c (.7, .3)) and training=Sr [ind == 1,1:20] and test= Sr [ind == 2, 1:20]. Here twenty predicator variables in normalized numerical values passes the neural network model is determined with n=neuralnet (pf ~ SLC +Plus +Physics +Math+ Chemistry + Bio +Parent1 +Parent2 +X1+ X2+ X3+ X4+ X5+ X6+ X111+ X222+ X333+ X444+ X555 + X666, data = training, hidden = c (1), err.fct="sse”, linear. output = FALSE) conversed as.

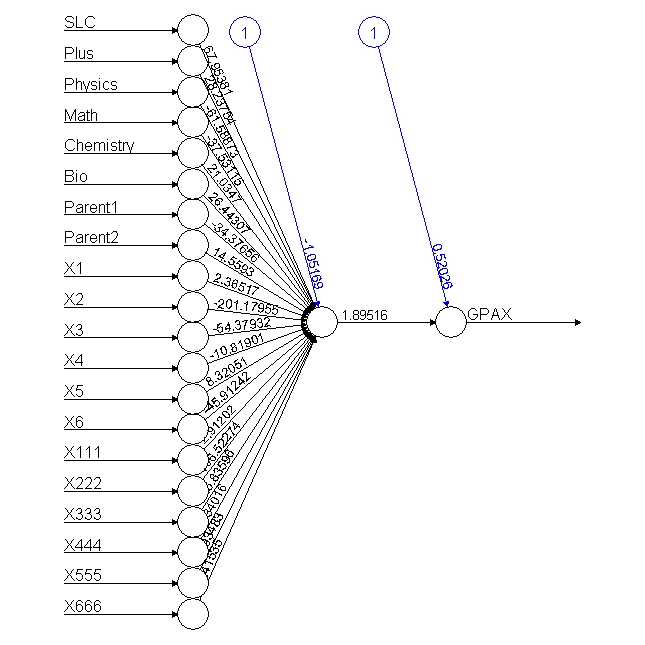
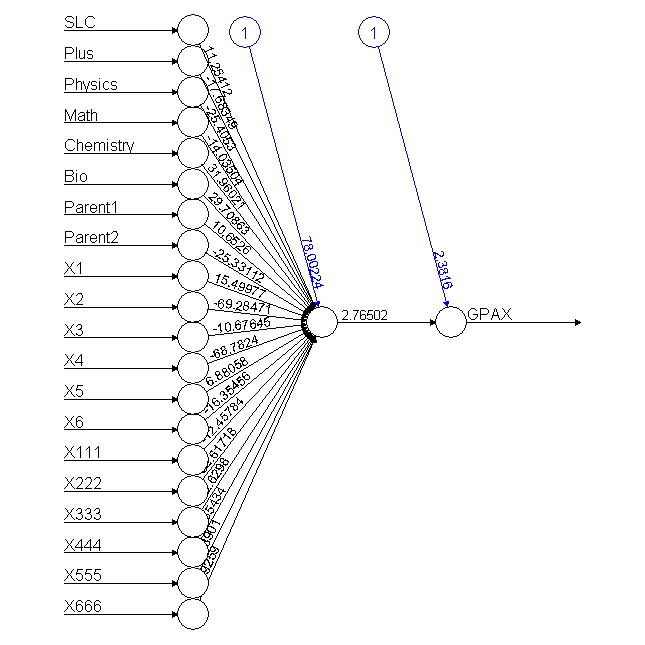
 

Figure 34: Neural Network Singe Layer of All Result

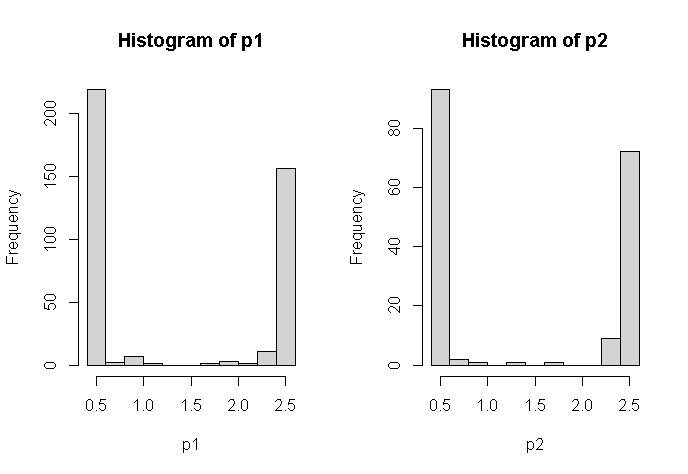
These neural network models are designed using training data with single hidden layer in between. Similarly based on categorical predicator variable error factorial is set out cross- entropy when linear out become false for accurate prediction. The plot n function prints the network diagram. The names’ function displays all the independent variable to formed such model with name. After computing training data with model data using d1=ifelse (p1>=2.2,1, ifelse (p1>=1.83,2, ifelse (p1>=1.5,3, ifelse (p1>=1.2,4, ifelse (p1>=.80,5, ifelse(p1>=.5,6, ifelse (p1>=0.1,7, 0))))))) of all grades for the appropriate grade prediction with consideration with output of following grade points of GPA. The model output of result in histogram is further categorized with respective GPA. 

Figure 35: Histogram of Training and Test Conversed Output

While doing so the independent variable like school information were not included according to from random forest model suggestion for better prediction. Thus, there were 20 predictor variables to predict whether the student has passed or fail grade using a single middle neuron. While conversing to 20th node input vector value and its network weight of every single input using 20th node input weight who’s each weight values were forward using the sigmoid function to another output layer again calculated input to another 21 neurons for prediction final either student result. This model exactly used as passed to the signal impulse from various organs of the human body to the brain. The neuralnet: compute (n, training [, -1]) function computes its output using net. result while both values could be compared with excluding predictor variable. The confusion matrix of data.

pred1 0 1 2 3 4 5 6 7

pred2 0 1 2 3 4 5 6

1 28 6 21 11 8 6 1

3 0 0 0 0 1 0 0

4 0 0 1 0 0 0 0

5 1 0 0 0 0 0 0

6 56 6 10 11 5 7 0

1 35 12 41 34 24 16 4 1

2 0 0 2 2 0 0 0 0

3 0 1 0 0 0 0 0 0

5 5 3 0 0 0 0 0 0

6 165 21 21 7 1 6 0 0

The accuracy of the model produces the error using 1-sum(diag(tab1))/sum(tab1) is 0.91022 percent accuracy of training data sets of 401 data sets and 179 training data sets of same model produces the accuracy after using 1-sum(diag(tab2))/sum(tab2) is 0.8100559 percent accuracy is good model prediction.

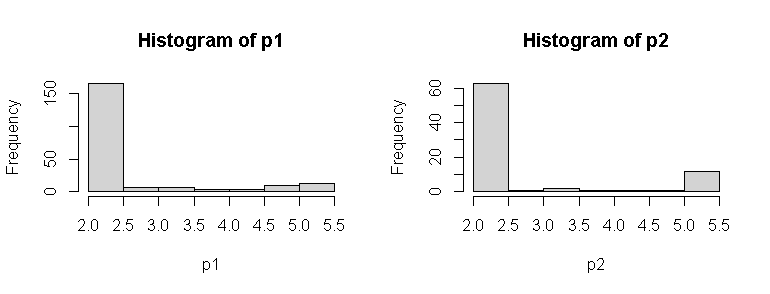
Similarly, when the result with different grading 1" "2" "3" "4" "5" "6" "7" points of the same student with half model conversed for the neural network. After having sub setting using GPA grade is greater than 0 passed students. The ss=data2 %>% select (GPAX, GPA, SLC, Plus, Physics, Math’s, Chemistry, Bio, Parent1, Parent2, X1, X2, X3, X4, X5, X6, X111, X222, X333, X444, X555, X666) %>% filter (GPA>0) and formed 70 30 split ratios of 290 passed student records after normalization of independent variables the neural model is designed with seven categories of grades scored of respective grade equivalence number. After computing, model output with training data of 209 data sets and 81 test datasets produces the output histogram based on the neural network model describes in seven categories of model different seven grade started with highest to lowest grades of students.

Figure 36 : Histogram of Half Neural Network Model

While calculating each grade scale of model output pred2=ifelse (p2> =4.6,1, ifelse (p2> =4.8,2, ifelse (p2>=3.8,3, ifelse(p2>= 3.4,4, ifelse (p2>=2.9,5, ifelse (p2>=2.5,6,7)))))) condition for model output grades produces the confusion matrix of both sets of training and test sets as.

pred1 1 2 3 4 5 6 7

pred2 1 2 3 4 5 6

1 0 2 3 3 5 0

3 0 2 0 0 0 0

4 0 0 0 1 0 0

5 1 0 0 0 0 0

6 0 1 0 0 0 0

7 12 23 11 9 6 2

1 0 0 1 2 12 3 1

3 0 0 0 4 3 0 0

4 0 0 3 2 0 0 0

5 0 0 4 1 0 0 0

6 0 5 1 1 0 0 0

7 36 63 42 16 9 0 0

The model validation and accuracy using 1-sum(diag(tab1))/sum(tab1) is found 0.9808612 percent accuracy of training data sets of 209 records of student where as the test prediction accuracy formed confusion matrix as using cross diagonal sum of its confusion matrix 1-sum (diag (tab2))/ sum (tab2) is 0.9506173 accuracy respectively implies model were good enough.

## **Neural Network Correction with Single Layer with Five Neurons (N=5)**

From the above, fully connected neural network model analyzed 22 independent variables to predict student grade whose accuracy further evaluated with using various model to get the best model for research recommendation. Therefore, more than a single hidden layer in between might test and analyzed for grade prediction of 580 records of Pokhara University Nepal. After loading data, the dependent and independent variables like high school grade, plus2 grade with mathematics, physics bio and chemistry with parent education and whether the student is derived from government schools or private schools and their bachelors’ internal grade for five hidden layers for prediction of student grade are selected using tydeverse package in r console. After selecting 14 independent variables. After using seed value for more reputability, the two samples with 70% for training and 30% for testing purpose 401 and 179 records separated using sample command. As we know the independent variables GPA with includes 0 for fail, 1 for A,2 for A- and so on eight grades ("0" "1" "2" "3" "4" "5" "6" "7") of student scored in result 2019. The model n = neuralnet (GPAA~SLC +Plus+ Physics+ Math+ Chemistry+ Bio+ Parent1+ Parent2+ X1+X2+ X3+ X4 + X5+ X6, data = training, hidden = 5, err.fct = "sse", linear. output = FALSE). The model here used an error factor with the sum of a square and its hidden layer is 5 nodes for predicting student grade result conversed and plotted as.

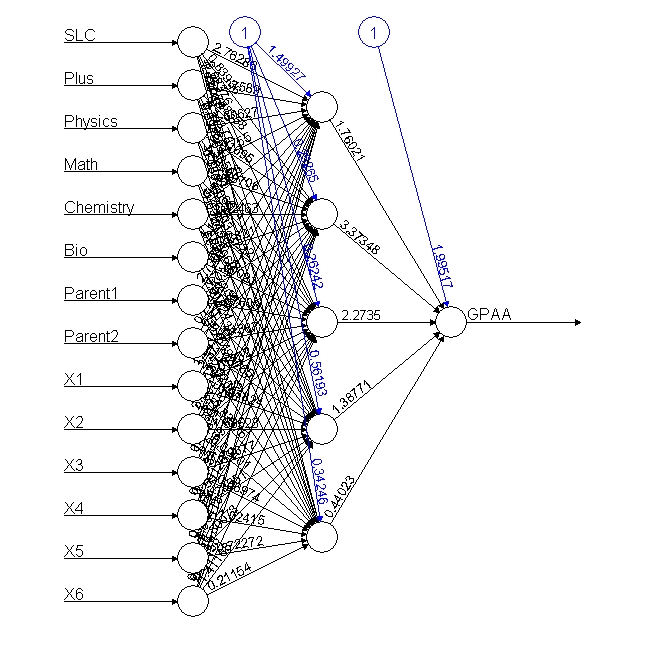
 

Figure 37: Neural Network with Single Layer 5 Neurons

After the fully connected neural network model conversed output is further computed using neuralnet: compute (n, training [, -1]) function and its predicated output printed in r console using output$net. result model function. The prediction value of the output neuron is further categorized on GPA categories of student.

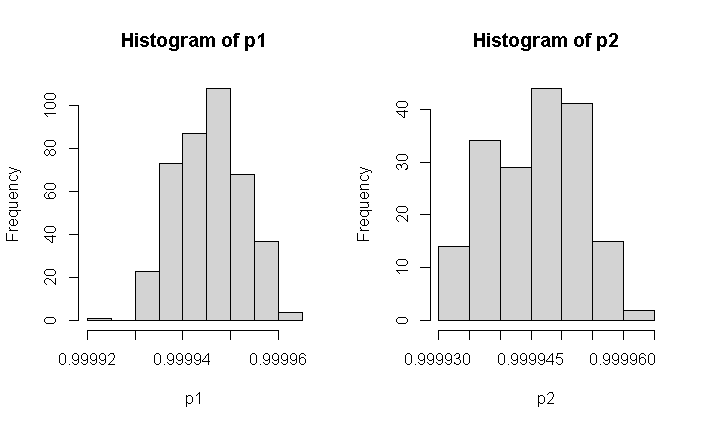


Figure 38: Histogram of Conversed Output of Single Layer 5 Neurons

After selecting training and test output the histogram predict total values are grouped into eight categories with its output weight, using ifelse (p2> =.999955,1, ifelse (p2 >=.99995,2, ifelse (p2>=.999945,3, ifelse (p2>=.99994,4, ifelse (p2>=.999935,5, ifelse (p2>=.99993,6, ifelse (p2> =.999925,7,0))))))) of each training 401 and 179 test samples of 580 records. The training data confusion matrix calculation using 1-sum(diag(tab1))/sum(tab1) produce is0.9127182 accuracy on training data.

pred1 0 1 2 3 4 5 6 7

0 1 0 0 0 0 0 0 0

pred2 0 1 2 3 4 5 6

1 10 1 4 0 2 0 0

2 17 3 10 5 2 3 1

3 19 5 4 7 4 5 0

4 14 2 4 4 3 2 0

5 14 1 10 6 1 2 0

6 11 0 0 0 2 1 0

1 19 2 4 4 9 2 1 0

2 34 6 12 7 5 3 1 0

3 55 10 20 12 5 4 1 1

4 44 7 13 12 3 8 0 0

5 37 9 13 5 3 5 1 0

6 15 3 2 3 0 0 0 0

Similarly, the test data of 179 sample accuracy and confusion matrix using 1- sum (diag (tab2)) /sum (tab2) is 0.8715084 percent implies best model using full model.

Similarly, the model is further analyzed those who had passed results (half model) after sub setting the succussed GPA student 290 passed student 209 and 81 training and testing data using n=neuralnet (GPAA~ SLC+ Plus+ Physics + Math + Chemistry+Bio+Parent1 +Parent2+ X1+X2 +X3 +X4+ X5+ X6, data = training, hidden =5, err.fct ="sse", linear. output = FALSE) plots the output as above second plot. Thus, the categorical variable with "1" "2" "3" "4" "5" "6" "7" seven categories of GPA student scored marks is further computed and predicated with model output as with training and test plots histogram of both training and test as.

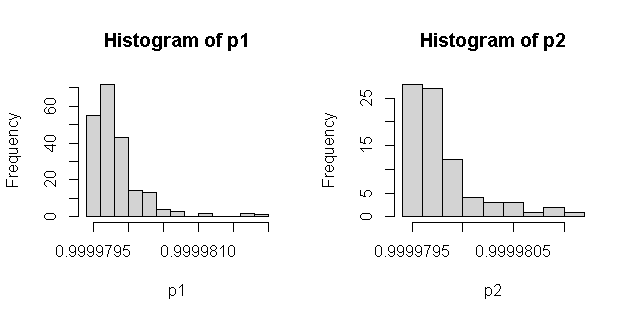


Figure 39: Histogram of Single Layer Conversed Output 5 Neurons

For the purpose of seven categories data GPA, the respective GPA values the histogram final predicated value is further calculated using ifelse(p2>=.9999802,1, ifelse (p2>= .9999801, 2, ifelse (p2> =.9999799,3, ifelse (p2 >= .9999798, 4, ifelse (p2>= .9999797, 5, ifelse (p2 >= .9999796, 6, 7)))))) condition to get proper confusion matrix. The first training confusion matrix accuracy calculated is 0.9282297 using 1-sum (diag(tab1)) /sum (tab1) and the testing confusion validate using similar formula produced is 0.9382716 percent accuracy of model using 5 hidden neurons as.

pred1 1 2 3 4 5 6 7

pred2 1 2 3 4 5 6

1 0 4 1 1 4 0

2 0 1 0 0 0 0

3 0 5 0 2 1 0

4 2 1 1 2 0 1

5 6 2 5 1 2 0

6 1 5 3 1 1 0

7 4 10 4 6 3 1

1 3 0 0 0 2 0 0

2 1 1 0 1 0 0 0

3 0 0 2 1 1 0 0

4 3 11 2 1 1 0 0

5 5 15 15 6 8 2 1

6 13 29 17 8 5 0 0

7 11 12 15 9 7 1 0

## **Neural Network of Two Hidden Layer (N=7, 3)**

For the purpose model improvement evaluation, the more hidden layer can be designed using two hidden layers of 7 and 3 in between single dependent and 13 independent variables to predict final GPA having various "0" "1" "2" "3" "4" "5" "6" "7" eight category GPA grades of 580 student records (full model). After loading subset full data with pass and fail records in final grade 401 training and 179 testing sample of normalized data separation for model testing. After loading neural network package, n=neuralnet (GPAA ~ SLC + Plus+ Physics + Math +Chemistry +Bio+Parent1 +Parent2 +X1+ X2+ X3+ X4+ X5+ X6, data = training, hidden = c (7, 3), err.fct ="sse”, linear. output = FALSE) conversed and plots model as.

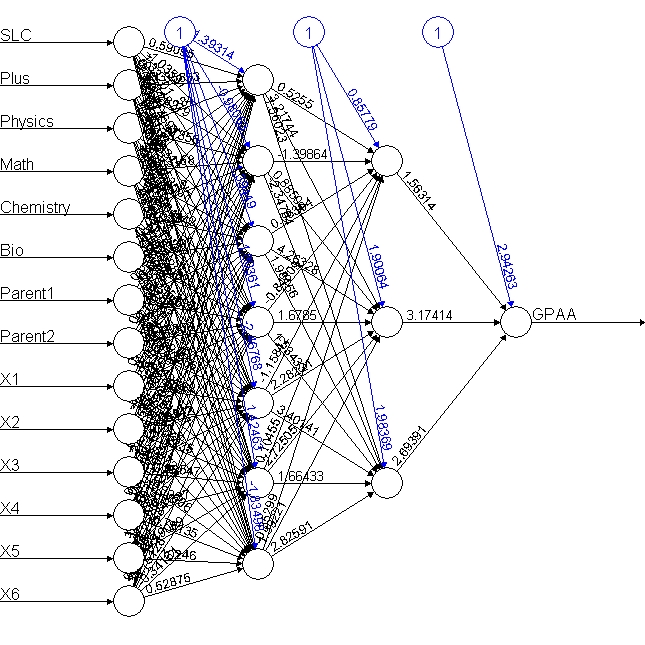
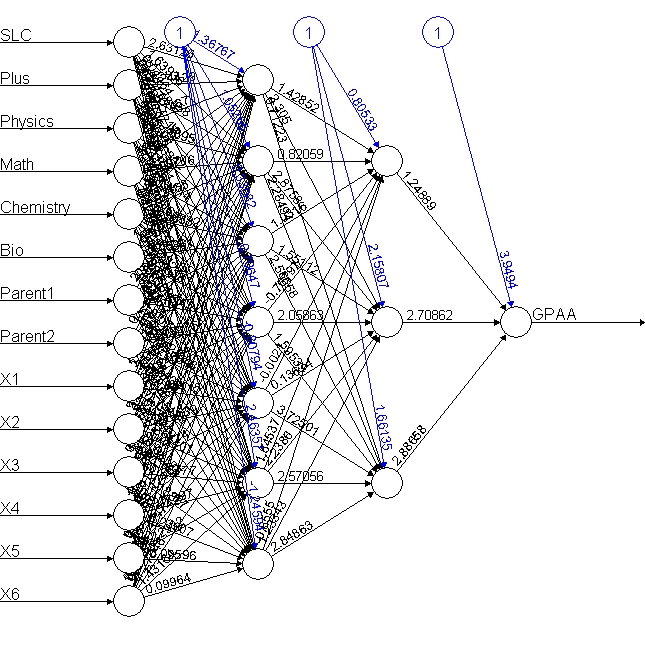
 

Figure 40: Neural Network of Two Hidden Layers ( 7,3)

After computing the neural network model output is further separated training and test data using the condition of ifelse(p1>=.999944,1, ifelse(p1>=.999942 ,2, ifelse (p1>= .999943, 3, ifelse (p1>=.999940,4, ifelse (p1>= .999938,5, ifelse (p1 >=.999936,6, ifelse (p1>= .999936, 6,0))))))) function of output weight display confusion matrix respectively as

pred1 0 1 2 3 4 5 6 7

pred2 0 1 2 3 4 5 6

0 1 0 0 0 2 0 0

1 15 0 3 1 2 1 0

2 17 2 3 6 0 1 0

4 37 8 19 11 5 7 1

5 11 2 5 4 3 1 0

6 4 0 2 0 2 3 0

0 4 1 1 0 0 1 0 0

1 25 6 3 3 1 1 0 0

2 37 3 11 4 1 6 2 0

4 96 15 29 28 11 11 1 1

5 37 8 16 7 9 3 1 0

6 6 4 4 1 3 0 0 0

The accuracy is further calculated using 1-sum (diag (tab1))/ sum (tab1) is 0.8553616 percent of training data and the accuracy using 1-sum(diag(tab2)) /sum(tab2) is 0.8826816 percent accuracy of model seems good.

Figure 41: Histogram Output of Test and Training Data

Similarly, the half model of fully connected neural network with grade of seven grades "1" "2" "3" "4" "5" "6" "7" of those passed student converse formed the model after subletting the 14 independent and GPA predictor variable of two training sample of 70 and 30 percent test sets is evaluated using model and training and test data sets of 209 and 81 data sets of 290 passed student in the final examination in 2019. Plotted the neural network model as second plot above and its histogram after applying seven category of neuron output plotted as.

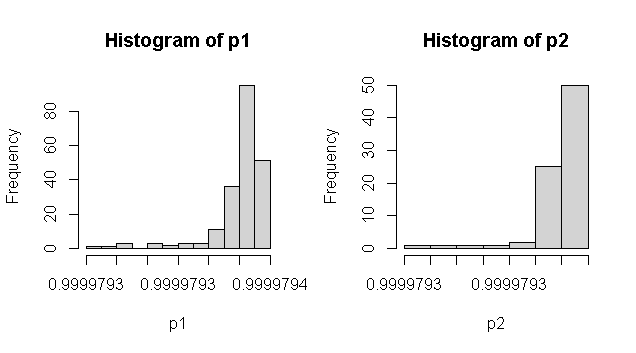


Figure 42: Histogram Output of Training and Test

After plotting the net result in the training histogram and test histogram the seven GPA marks are evaluated its confusion matrix using table command of respective sampled data as.

pred1 1 2 3 4 5 6 7

pred2 1 2 3 4 5 6

1 3 12 3 1 3 1

2 8 12 10 8 6 0

3 0 2 0 1 0 1

4 1 0 1 1 1 0

5 0 1 0 1 0 0

6 1 1 0 1 1 0

1 10 23 17 4 7 1 0

2 24 36 26 18 14 1 1

3 0 6 5 0 0 0 0

5 0 2 0 1 0 0 0

6 1 0 1 2 0 1 0

7 0 1 2 1 1 0 0

Which produce the training accuracy is 0.7511962 percentage and the test model accuracy is 0.8024691 percent respectively. Therefore, half model is far better than full model.

## **Neural Network with Three Hidden Layers (N= 7, 3, 2)**

As neural network model is a large model for conversing the predicated input to the dependent variable with the help of some constant bios and the sum of each input signal and its weight in fully connected neurons in the forward direction. In each node first weight to is calculated before the node and the activation function using a sigmoid function with bios, and 'tanh' are possible for the logistic function and tangent hyperbolic is selected for movement in further pushing. The first neural network model used the formula with dependent and list of independent variables in the model with data on the training sample. The threshold optional parameter a numeric value specifying the threshold for the partial derivatives of the error function as stopping criteria. The step max parameters use to reaching this maximum leads to a stop of the neural network's training process.

All confidence intervals are calculated under the assumption of a local identification of the given neural network. If this assumption is violated, the results will not be reasonable. Therefore, three or more hidden layers are used for testing their prediction. However, when act. fact="tanh" become true the linear.out parameter needs to be TRUE. When the hidden layer becomes 7,3,2 the model won’t have conversed. While selecting the first hidden layers and more the neural network needs to convert the 97 and the error is 566.01552 and time is 0.59 secs otherwise the prediction accuracy of the model could not select. Similarly, the output based on reputation will be selected rep=specified numbers. When hidden layers (7,3,1) and linear out become TRUE the model algorithm did not converge in 5 of 5 repetition(s) within the step max otherwise it will be conversed either (50,14,7,4,1) as many as middle layers. However, the model accuracy and confusion matrix are significant for the model evaluation.

After selecting sub setting data using ss=data2 %>% select (GPAA, SLC, Plus, Physics, Math, Chemistry, Bio, Parent1, Parent2, X1, X2, X3, X4, X5, X6) from large data base then test and training data using ind=sample (2, nrow(ss), replace=T, prob=c (.7,.3)) splits data 401 for training and 179 as test sets for neural network model with three hidden layers. After loading library(neuralnet) packages the model n=neuralnet (GPAA~ SLC+ Plus+ Physics+ Math+ Chemistry+ Bio +Parent1+ X1+X2+X3+X4+X5+X6, data=training, hidden=c (7,3,2), err.fct="sse”, linear. output=FALSE, lifesign='full’, rep=5, algorithm="rprop+”, stepmax = 10000). This model used three hidden layers as (7,3,2) and linear. output equal FALSE otherwise the model wont conversed produced output as hidden: 7, 3, 2 thresh: 0.01 rep: 1/5 steps: 1000 min thresh: 1.15857355426732. the repetition step is 5 indicates that the model hidden: 7, 3, 2 thresh: 0.01 rep: 1/5 steps: 36 error: 577.00933 time: 0.05 secs

hidden: 7, 3, 2 thresh: 0.01 rep: 2/5 steps: 34 error: 577.00915 time: 0.04 secs

hidden: 7, 3, 2 thresh: 0.01 rep: 3/5 steps: 26 error: 577.0094 time: 0.02 secs

hidden: 7, 3, 2 thresh: 0.01 rep: 4/5 steps: 31 error: 577.00874 time: 0.03 secs

hidden: 7, 3, 2 thresh: 0.01 rep: 5/5 steps: 41 error: 577.00957 time: 0.06 secs

The plot of final output will be selected whose output found minimum time as. plot (n, rep=3)

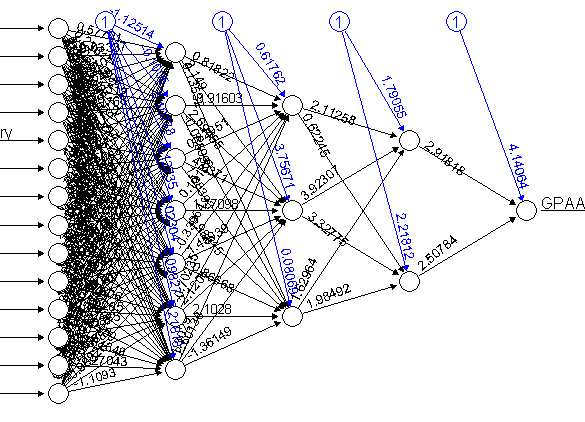
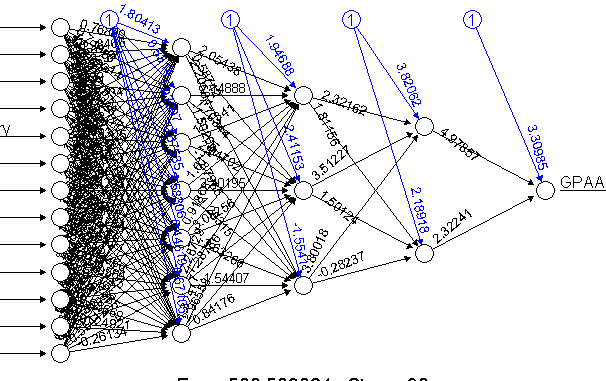
 

Figure 43: Neural Network Model of Three Hidden Layers

After conversed after setting various parameters with 5 repetitions, the best model is selected and computed its output=neuralnet: compute (n, training [, -1], rep = 3) is computed with respect training model for accuracy and confusion matrix prediction of all 580 student records. After computing using output=neuralnet: compute (n, training [, -1], rep=3) function. The categorical GPA values with different levels as "0" "1" "2" "3" "4" "5" "6" "7" could be plotted in histogram for the calculation confusion matrix and model test and training accuracy.

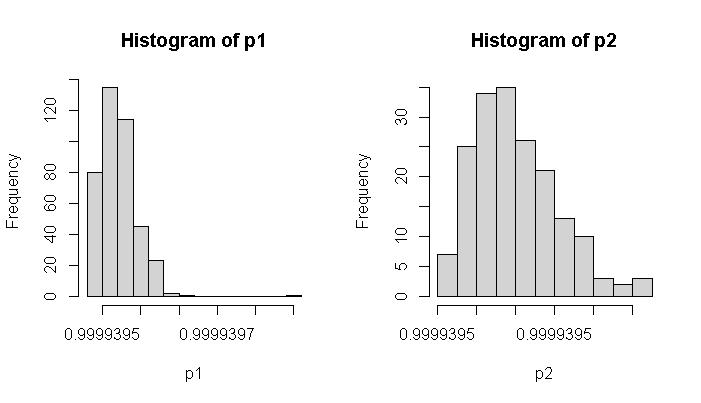


Figure 44: Histogram of Test and Training Output

After evaluating histogram of data, the confusion matrix and model accuracy of training and test sets of 401 and 179 sample datasets using pred1= ifelse (p1> =.999944,1, ifelse (p1>=.999942,2, ifelse (p1 >= .999943,3, ifelse (p1>=.999940,4, ifelse (p1>=.999938,5, ifelse(p1>=.999936,6, ifelse(p1>=.999936,7,0))))))) tab1=table (pred1, training$GPAA) based on network model in different eight GPA grade the confusion and accuracy.

pred2 0 1 2 3 4 5 6

2 85 12 32 22 14 13 1

pred1 0 1 2 3 4 5 6 7

2 205 37 64 43 25 22 4 1

From the above confusion matrix, the test accuracy is calculated using 1-sum(diag(tab1))/ sum(tab1) is 0.4887781 and test accuracy is 0. 5251397.This accuracy is quite low this model therefore this model is rejected to GPA prediction which only matched 52 percent of 580 student records.

Similarly, the half model data is prepared using library(tidyverse) package and select only those columns using ss=data2 %>% select(GPAA,GPA,SLC, Plus, Physics, Math, Chemistry, Bio, Parent1, X1,X2, X3, X4, X5, X6) %>% command then split into two training and test at ratio 70 and 30 (209: 81) then feed these data using model n=neuralnet (GPAA~ SLC+ Plus+ Physics+ Math +Chemistry+ Bio+ Parent1+ X1+ X2+ X3+ X4+ X5+ X6, data=training, hidden=c(7,3,2), err.fct="sse", linear.output=FALSE, lifesign='full', rep=5,

algorithm ="rprop+”, stepmax=10000) produces the output as

hidden: 7, 3, 2 thresh: 0.01 rep: 1/5 steps: 41 error: 500.50669 time: 0.06 secs

hidden: 7, 3, 2 thresh: 0.01 rep: 2/5 steps: 30 error: 500.50908 time: 0.02 secs

hidden: 7, 3, 2 thresh: 0.01 rep: 3/5 steps: 36 error: 500.50753 time: 0.02 secs

hidden: 7, 3, 2 thresh: 0.01 rep: 4/5 steps: 44 error: 500.50897 time: 0.03 secs

hidden: 7, 3, 2 thresh: 0.01 rep: 5/5 steps: 33 error: 500.50792 time: 0.02 secs

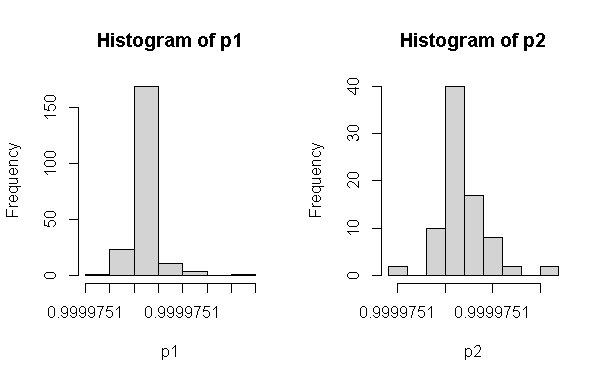
then model print plot (n, rep=2) function as above second model. After computing output=neuralnet: compute (n, training [, -1], rep=2) function the training and test accuracy with confusion matrix plotted two separate histograms as.

Figure 45: Histogram of Training and Test Data

After selecting same model, the computed output with training and test data describes predicated output. The confusion matrix is achieved using ifelse (p2 > =.9999722,1, ifelse (p2 >= .9999721 ,2, ifelse (p2>=.9999720,3, ifelse (p2> = .9999719,4, ifelse (p1>=.9999718,5, ifelse (p2> =.9999717,6,7)))))) condition of neuron output. The model confusion matrix is as.

pred2 1 2 3 4 5 6 pred1 1 2 3 4 5 6 7

1 13 28 14 13 11 2 1 36 68 51 26 24 3 1

After using 1-sum(diag(tab1))/sum(tab1) model accuracy the training data is 0.9952153 percent and test data accuracy is 0.8395062percentage accurately prediction implies the best model.

Similarly, the head (training [,],88) function compared test with actual and model as.

GPAA GPA SLC Plus Physics Math Chemistry

1 2 3.45 0.6900000 0.8333333 0.81081081 0.5192308 0.78947368

2 3 3.01 0.7916667 0.7222222 0.56756757 0.7500000 0.55263158

3 1 3.82 0.5833333 0.6944444 0.54054054 0.5769231 0.76315789

4 3 3.26 0.8333333 0.7777778 0.67567568 0.8653846 0.39473684

6 5 2.48 0.8888889 0.7222222 0.56756757 0.6346154 0.39473684

7 5 2.53 0.8611111 0.7777778 0.67567568 0.8653846 0.39473684

8 2 3.36 0.5555556 0.5000000 0.75675676 0.3269231 0.02631579

9 2 3.38 0.5833333 0.5000000 0.75675676 0.3269231 0.02631579

Therefore, from the above output on original GPA with respect GPAA variables were largely mismatched. The second record had accurately matched the other records model prediction and test accuracy with original students’ records were mismatched. This model has large misclassification varies from .488 to 82 percent accuracy.

## **Neural Network with Original Datasets (N=7 ,3 ,2)**

The another thought of model validation using neural network with original data sets with three hidden layers comprises (7,3,2) in between before predictor variables of GPA with numerical category of eight different grades of student marks. After loading the data set into r console the preprocessing of data sets with converting as number pattern if the required data sets grouped data2 %>% select (GPAA, SLC, Plus, Physics, Math, Chemistry, Bio, Parent1, X1, X2, X3, X4, X5, X6) with tidyverse package then split into 70 and 30 splits with 401 and 178 random samples. After taking seed values for repeatability the neural network of three hidden layers nn=neuralnet (n=neuralnet (GPAA~ SLC+ Plus+ Physics+ Math+ Chemistry+ Bio+ Parent1 +X1+ X2+X3+ X4+ X5 + X6, data=training, hidden=c (7,3,2), err.fct="sse”, linear. output=FALSE, lifesign ='full’, rep=5, algorithm ="rprop+”, stepmax=10000)

$ SLC: num 79.8 83.5 76 85 86 ...

$ School: int 1 1 0 0 0 1 0 1 1 1 ...

$ Plus: int 78 74 73 76 76 74 76 66 66 69 ...

$ Physics: int 85 76 75 80 80 76 80 83 83 56 ...

$ Math: int 72 84 75 90 90 78 90 62 62 68 ...

$ Chemistry: int 85 76 84 70 70 70 70 56 56 61

$ Bio: int 85 75 70 70 70 80 70 48 48 0 ...

$ Parent1: int 1 4 3 2 2 4 1 1 1 2 ...

$ Parent2: int 3 4 4 4 4 4 4 4 4 3 ...

$ X1: int 44 47 49 40 50 42 50 44 45 46 ...

$ X2: int 39 39 44 41 41 36 41 44 44 45

Sample Original Data

Here researcher used default algorithm resilience propagation with five possible repetitions is 5 reptation, the error factor is set as the sum of a square and step max is 10000 iteration the correction takes place sand default algorithm is rprop+ default activation logistic. This model displays all the information with reached.

hidden: 7, 3, 2 thresh: 0.01 rep: 1/5 steps: 38 error: 576.00928 time: 0.04 secs

hidden: 7, 3, 2 thresh: 0.01 rep: 2/5 steps: 27 error: 576.00805 time: 0.03 secs

hidden: 7, 3, 2 thresh: 0.01 rep: 3/5 steps: 31 error: 576.00722 time: 0.03 secs

hidden: 7, 3, 2 thresh: 0.01 rep: 4/5 steps: 41 error: 576.00792 time: 0.04 secs

hidden: 7, 3, 2 thresh: 0.01 rep: 5/5 steps: 29 error: 576.00938 time: 0.03 secs

After evaluation of network output the plot of neural network is selected using plot (n, rep=2) or plot (n, rep=3) or plot (n, rep=5) plots as.

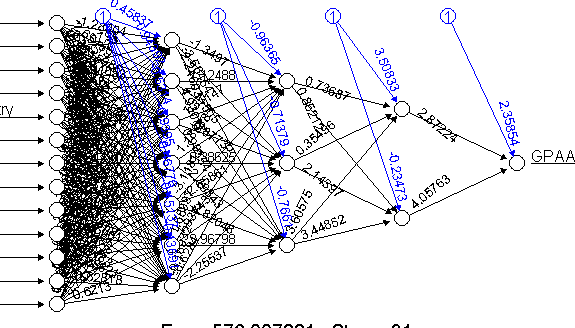
 

Figure 46: Neural Network Model of (7,3,2) with Original Data

After computing the model using output = neuralnet: compute (n, training [, -1], rep=3) command on selected data, the output of the final GPA is further plotted with the histogram of the eighth-grade display using ifelse(p1>=.9999,1, ifelse(p1>=.9998,2, ifelse(p1>=.9997,3, ifelse(p1>=.9996,4, ifelse(p1>=.9995,5, ifelse(p1>=.9994,6, ifelse(p1>=.9993,7,0))))))) command and training histogram and testing histogram.

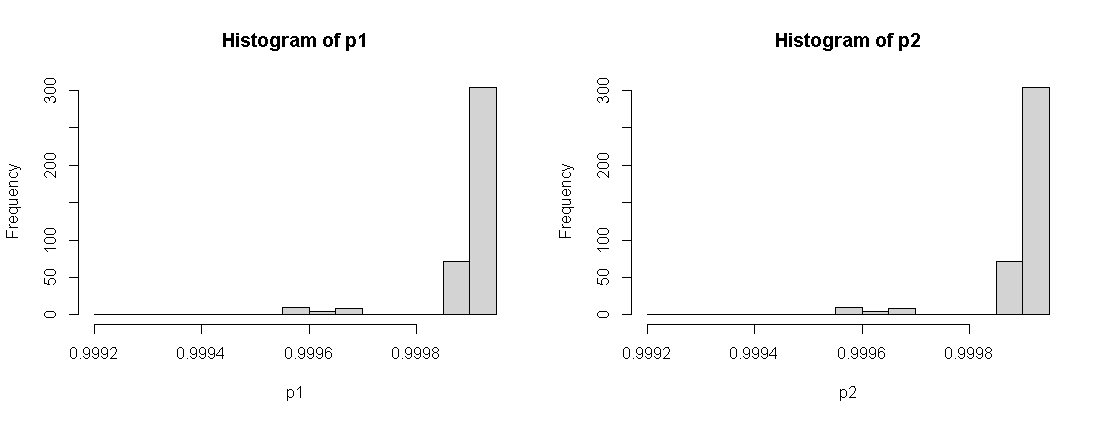


Figure 47: Histogram of Training and Test of Original Data

After histogram the accuracy production using ifelse(p2>=.9999,1, ifelse (p2>=.9998,2,

ifelse (p2>=.9997,3, ifelse(p2>=.9996,4, ifelse (p2>=.9995,5, ifelse(p2>=.9994,6, ifelse (p2>=.9993,7,0))))))) is calculated histogram for eight different gread prediction. The table of matrix of confusion display as.

pred1 0 1 2 3 4 5 6 7

0 2 0 0 0 0 0 0 0

1 144 26 55 36 23 15 2 2

2 40 11 5 8 0 7 1 0

3 0 0 1 0 0 0 0 0

4 12 1 0 0 0 0 0 0

5 7 2 0 1 0 0 0 0

The model accuracy using misclassification error using 1-sum(diag(tab1))/sum(tab1) # accuracy and miss classification is 0.9177057 percentage of input signals. The misclassification is again compared with original values gives the percent of errors with test data using 1-sum(diag(tab2)) /sum (tab2) is 0.8938547 percentage is quite improvement.

Similarly, the half model of the above design after grouping passed student record using tidyverse function. With data splitting 70:30 splits 209 and 81 sampled whose model plot is plotted after plot (n, rep=2) evaluation above diagram. The training neuralnet: compute (n, training [, -1], rep =2) and test neuralnet: compute (t.test [, -1], rep=2) computation using ifelse(p1>=.99,1, ifelse(p1>=.8,2, ifelse(p1>=.7,3, ifelse(p1>=.6,4, ifelse(p1> =.5,5, ifelse (p1>=.4,6,7)))))) of both sets plots the histogram as



Figure 48: Histogram of Three Hidden Layers with Original Datasets

The confusion matrix and test accuracy were calculated respectively using table (pred1, training $GPAA) function whose prediction accuracy is evaluated using 1-sum(diag(tab1)) / sum (tab1) function respectively. The training accuracy confusion matrix as

pred1 1 2 3 4 5 6 7 pred2 1 2 3 4 5 6

1 38 66 52 25 24 2 2 1 14 27 15 12 11 2

Which produces training accuracy is 0.8181818 percentage and test accuracy is 0.8271605 implies that this model has best than three hidden layers with equal both training and test accuracy. The further commands like n i. e model output describes its converse with fixed steps the bias terms make great role to converse average of input. And the plot(n) plots all five possible network plots separately respective weight of each neuron in model formed. The last result indicates each steps error, threshold with respect other variables as. [,1] [,2] [,3] [,4] [,5]

error 502.506696433 502.509034718 502.507491262 502.508920992 502.507875680

reached. Threshold 0.006696345 0.009034558 0.007491151 0.008920836 0.007875559

steps 41.000000000 30.000000000 36.000000000 44.000000000 33.000000000

Intercept.to.1layhid1 0.054168304 1.104125759 -0.943756709 -2.037907228 1.918519241

SLC.to.1layhid1 2.771775079 0.068886056 -2.682168438 0.920408542 0.951375898

Plus.to.1layhid1 -0.349863963 0.284661860 -1.454400609 0.719360861 0.742296172

Physics.to.1layhid1 -1.392210963 -1.969335711 -2.110365322 -1.170741178 0.732048862

Math.to.1layhid1 -0.756881174 -0.196151625 -0.899326340 0.256693833 0.358515170

Chemistry.to.1layhid1 -0.351140566 0.758284045 -2.519944863 -1.822276517 0.322513743

Bio.to.1layhid1 1.578623977 1.122742620 -0.986965968 0.997086131 -1.141343270

Similarly, the summary(covariate) and summary (n$result. matrix) commands the descriptive statists of the model as > summary (n$result. matrix)

V1 V2 V3 V4 V5

Min.: -3.2440 Min.: -2.2058 Min.: -2.6822 Min.: -2.8807 Min.: -2.0316

pred2 0 1 2 3 4 5 6

0 0 0 0 0 1 0 0

1 65 11 25 13 9 8 1

2 15 1 6 5 3 2 0

3 0 0 0 1 0 1 0

4 1 0 1 1 0 1 0

5 4 0 0 2 1 1 0

1st Qu.: -0.4227 1st Qu.: -0.2122 1st Qu.: -0.2330 1st Qu.: -0.3728 1st Qu.: -0.2313

Median: 0.2268 Median: 0.4416 Median: 0.6247 Median: 0.5976 Median: 0.5456

Mean: 4.3319 Mean: 4.4751 Mean: 4.9417 Mean: 4.5588 Mean: 4.7872

3rd Qu.: 1.0754 3rd Qu.: 1.3312 3rd Qu.: 2.3169 3rd Qu.: 1.3883 3rd Qu.: 1.7071

Max. :502.5067 Max. :502.5090 Max. :502.5075 Max. :502.5089 Max. :502.5079

> summary(covariate)

SLC Plus Physics Math Chemistry Bio Parent1

Min.:55.00 Min. :49.00 Min.:55.0 Min.:55.00Min.:55.00Min.:0.00Min.1.000

1st Qu.:75.60 1st Qu.:66.00 1st Qu.:69.0 1st Qu.:69.001st Qu.:70.001st Qu.:60.001st Qu.:2.000

Median :80.0 Median :73.0Median :76.0Median :80.0Median :78.0Median :70.0Median :2.0

Mean:78.37 Mean:70.83 Mean:75.3Mean:77.07Mean:75.95Mean:65.25Mean:2.278

3rd Qu.:84.253rd Qu.:77.003rd Qu.:81.03rd Qu.:85.003rd Qu.:84.003rd Qu.:75.003rd Qu.:3.000

Max.:90.00Max.:84.00Max.:92.0Max.:97.00Max.:93.00Max.:92.00Max:4.000. So that each model output could be analyzed respectively.

## **Back Propagation of One Single Layer**

The reversed propagation of neural network traversed while adjusting the predicated value in reversed order used rep optional parameter uses the number of repetitions for the neural network's training. The learning rate is a numeric value specifying the learning rate used by traditional backpropagation. Used only for traditional backpropagation. The life sign is a string specifying how much the function will print during the calculation of the neural network. 'none', 'minimal' or 'full'. The algorithm to calculate the neural network. The following types are possible: backprop, rprop+, rprop-, sag refers to backpropagation, 'rprop+' and 'rprop-' refer to the resilient backpropagation with and without weight backtracking, while 'sag' and 'slur' induce the usage of the modified globally convergent algorithm (rprop). The error factor is used for the calculation of the error. Alternatively, the strings 'sse' and 'ce' which stand for the sum of squared errors and the cross-entropy can be used. When the error rate is set to act. fact=” tnha” than linear. output becomes true.

The exclude option may exclude the matrix row and column for the model. The likelihood parameter is logical. If the error function is equal to the negative log-likelihood function, the information criteria AIC and BIC will be calculated (Anastasiadis, Stefan, & Frauke, 2005). Furthermore, the usage of confidence. the interval is meaningful. The result a list containing the overall result of the neural network for every repetition (Riedmiller & Braun, 1993) the confidence. interval, a method for objects of class nn, typically produced by a neural net. The confidence intervals of the weights and the network information criteria NIC (Murata, 1994) All confidence intervals are calculated under the assumption of a local identification of the given neural network. If this assumption is violated, the results will not be reasonable. Therefore, three or more hidden layers are used for testing their prediction. However, when act. fact="tanh" become true the linear.out parameter needs to be TRUE. When the hidden layer becomes 7,4,2 the model won’t have conversed. While selecting the first hidden layers and more the neural network needs to convert the 97 and the error is 566.01552 and time is 0.59 secs otherwise the prediction accuracy of the model could not select. Similarly, the output based on reputation will be selected rep=specified numbers. When hidden layers (7,3,1) and linear out become TRUE the model algorithm did not converge in 5 of 5 repetition(s) within the step max otherwise it will be conversed either (50,14,7,4,1) as many as middle layers. However, the model accuracy and confusion matrix are significant for the model evaluation. Here researcher try to explore single neuron network. After loading tidyverse package the selected variables of independent and dependent in another database before model design then converted into numeric pattern. The full model grouped into two training and test sample 401:179).

After taking set. seed (112) the neurol network with back propagation as n=neuralnet (GPAA~ SLC +Plus +Physics+ Math+ Chemistry+ Bio+ Parent1+X1+X2+ X3+ X4+X5 +X6, data = training, hidden=c (1), learning rate = 0.01, err.fct = "sse", linear. output = FALSE, lifesign = 'full’, rep = 5, algorithm = 'backprop’, stepmax =10000). Here all 14 independent columns of student records with data only training sets having single neuron with error factor is sum of square error with repetition using algorithm backpropagation with stepmax=10000 because output variable is categorical so output is true and the default algorithm is rprop+ default activation logistic regression. This model generates the output of five cases then select the best model.

hidden: 1 thresh: 0.01 rep: 1/5 steps: 1000 min thresh: 0.0497338019864384

2000 min thresh: 0.0249254114993073 3000 min thresh: 0.0166314321580569

4000 min thresh: 0.0124793501363943 4994 error: 576.01 time: 1.18 secs

hidden: 1 thresh: 0.01 rep: 2/5 steps: 1000 min thresh: 0.049688784713512

2000 min thresh: 0.0249141041753707 3000 min thresh: 0.0166263979159323

4000 min thresh: 0.0124765157533581 4993 error: 576.01 time: 1.22 secs

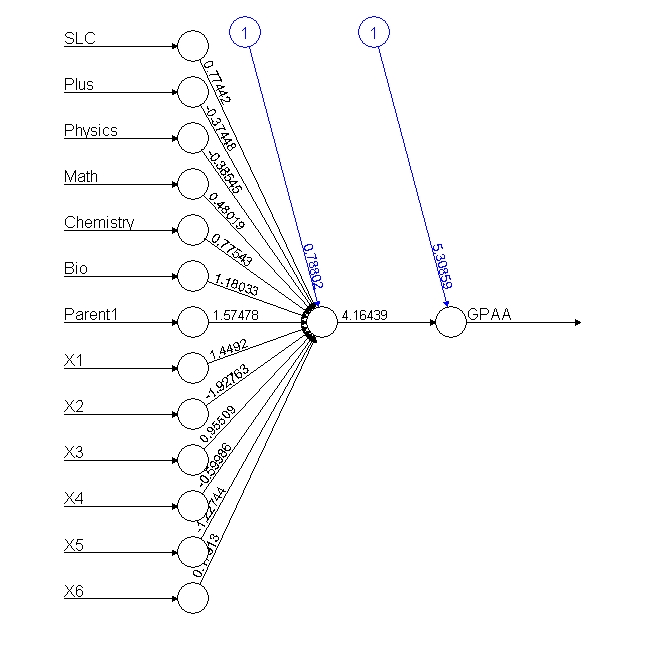
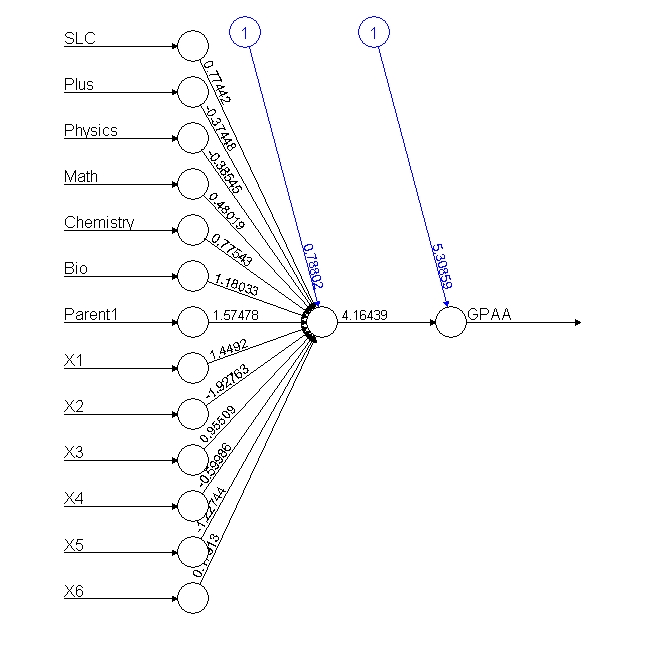
 

Figure 49: Single Layer Backpropagation Neural Network

After converted neural network model whose output further divides using pred1= ifelse (p3 >= .9,1, ifelse (p3>=.8,2, ifelse (p3>=.7,3, ifelse(p3>=.6,4, ifelse(p3> =.5,5, ifelse (p3>=.2,6,0))))))

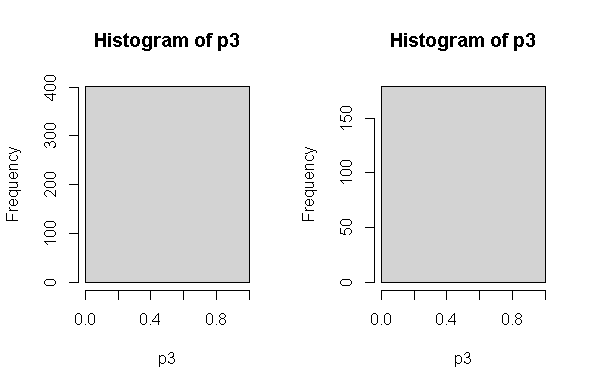


Figure 50: Histogram of Single Hidden Layer Backpropagation

The confusion matrix as.

pred1 0 1 2 3 4 5 6 7 pred1 0 1 2 3 4 5 6

1 205 40 61 45 23 22 3 2 1 85 12 32 22 14 13 1

The test data miss classification and accuracy using 1-sum(diag(tab3)) /sum(tab3) is 0.4887781 and the misclassification Error=mean (training$GPAA! = pred1) the misclassification is again compared with original values gives percent of errors (remember here the neural network model may not produce same output when run another time due to random weight initially taken part) is 0.9002494 is the worst model similarly the test accuracy and misclassification were 0.5251397 and 0.9329609. Therefore, both test and training output predicts this model has 50 percentage accuracy while using single neuron model.

Similarly, when using half model with passed student information 290 student the training and test sets of 209 and 81 splits using the same model back propagation as n=neuralnet (GPAA~ SLC + Plus +Physics+ Math+ Chemistry+ Bio+ Parent1+X1+X2+ X3+ X4+X5 +X6, data = training, hidden=c (1), learning rate = 0.01, err.fct = "sse", linear. output = FALSE, lifesign ='full’, rep=5, algorithm='backprop’, stepmax=10000). After conversed the =plots(rep=3) plots.

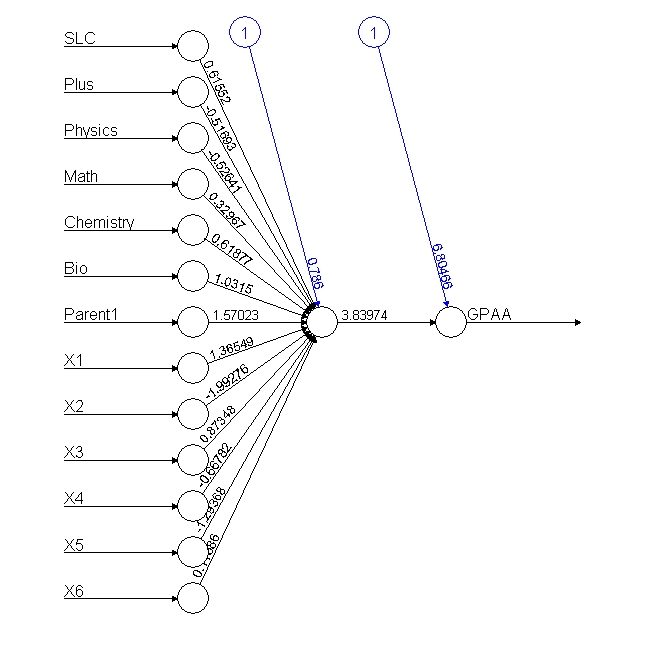
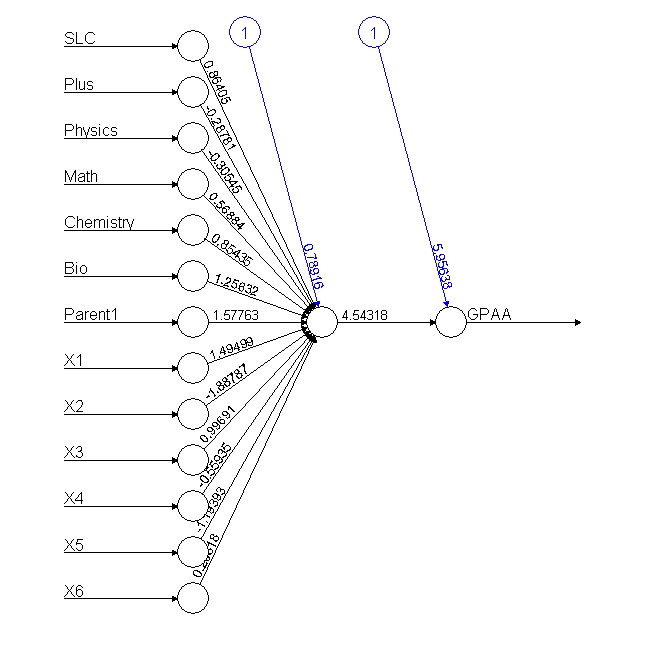
 

Figure 51: Neural Network of Single Layer Backpropagation

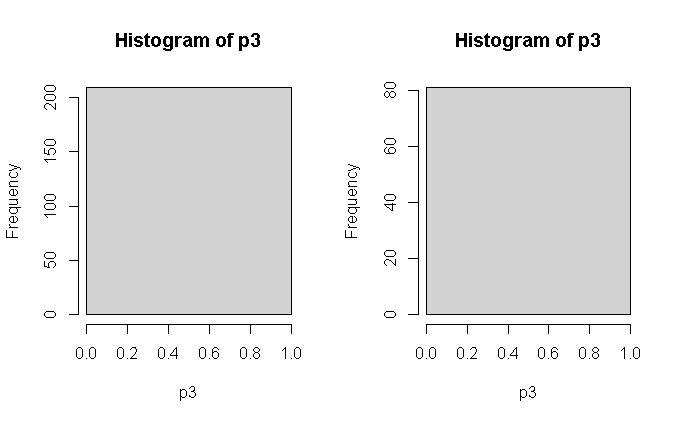
After computing original with output using neuralnet: compute (n, training [, -1]) and test neuralnet: compute (nn, testing [, -1]) using ifelse (p3>=.9,1, ifelse (p3>=.8,2, ifelse(p3>=.7,3, ifelse(p3>=.6,4, ifelse(p3>=.5,5, ifelse (p3> =.2, 6, ifelse (p3>=.1,7,0))))))) produced the histogram of output as 

Figure 52: Histogram of Single Layer Conversed Value

The confusion matrix and test and training accuracy of the model are as

pred1 1 2 3 4 5 6 pred1 1 2 3 4 5 6

1 14 27 15 12 11 2 1 14 27 15 12 11 2

After using training 1-sum(diag(tab3))/ sum(tab3) accuracy is 0.8271605 percentage and test accuracy is 0.8271605 both models exactly same accuracy is best model in half data. Whose mismatched output after cbind (training$GPAA, pred1) function describes each induvial student records.

5 2 1

14 3 1

16 1 1

26 3 1

60 1 1

61 1 1

66 2 1

72 2 1

## **Back Propagation of Five Hidden Layers (N=5)**

After taking set. seed (112) the neurol network with back propagation as n=neuralnet (GPAA~ SLC +Plus +Physics+ Math+ Chemistry+ Bio+ Parent1+X1+X2+ X3+ X4+X5 +X6, data= training, hidden=c (5), learning rate = 0.01, err.fct = "sse", linear. output = FALSE, lifesign='full’, rep=5, algorithm='backprop’, stepmax=10000). Here all 14 independent columns of student records with data only training sets having single neuron with error factor is sum of square error with repetition using algorithm backpropagation with stepmax=10000 because output variable is categorical so output is true backprop activation logistic regression. This model generates the output of five cases then select the best model.

hidden: 5 thresh: 0.01 rep: 1/5 steps: 1000 min thresh: 0.019666708909702

1983 error: 858.01 time: 1.13 secs

hidden: 5 thresh: 0.01 rep: 2/5 steps: 1000 min thresh: 0.0261156887868856

2000 min thresh: 0.013494926120734 2737 error: 858.0083 time: 1.6 secs

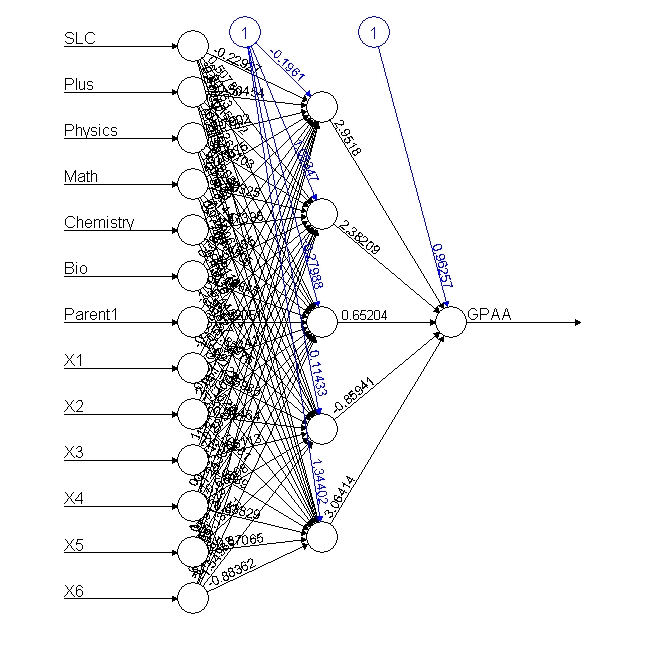
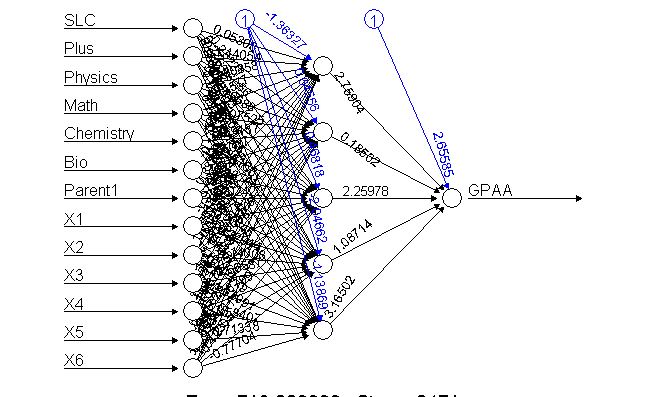
 

Figure 53: Full and Half Model Neural Network of Single Layer 5 Neuron

After neural network conversed the output at last node is the student grading of search grouped up using ifelse(p3>=.9,1, ifelse(p3>=.8,2, ifelse(p3>=.7,3, ifelse(p3>=.6,4, ifelse(p3>=.5,5, ifelse(p3>=.2,6, ifelse(p3>=.1,7,0))))))) condition of respective student grades.

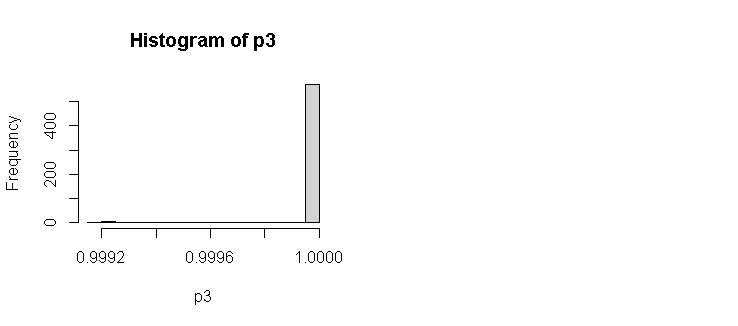
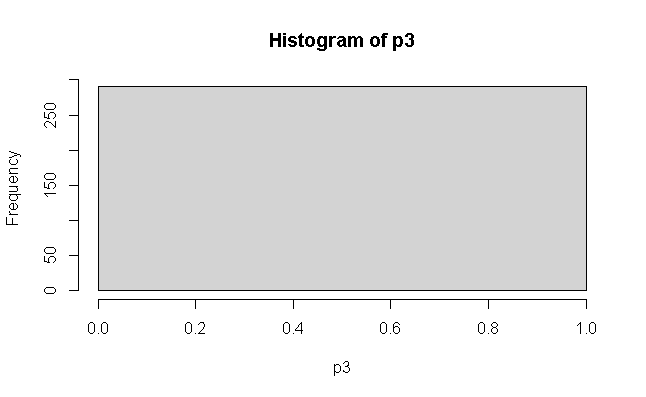
 

Figure 54: Full Model Histograms of 5 Neuron Single Layer

pred1 0 1 2 3 4 5 6 7 pred1 1 2 3 4 5 6 7

1 290 52 93 67 37 35 4 2 1 52 93 67 37 35 4 2

The test data miss classification and accuracy using 1-sum(diag(tab3))/sum(tab3) is 0.5 and the misclassification Error=mean (training$GPAA! = pred1) the misclassification is again compared with original values gives percent of errors (remember here the neural network model may not produce same output when run another time due to random weight initially taken part) is 0.9103448 is the worst model.

Similarly, when using half model with passed student information 290 student the training and test sets of 209 and 81 splits using the same model back propagation as n=neuralnet (GPAA~ SLC + Plus +Physics+ Math+ Chemistry+ Bio+ Parent1+X1+X2+ X3+ X4+X5 +X6, data = training, hidden=c (5), learning rate = 0.01, err.fct = "sse", linear. output = FALSE, lifesign ='full’, rep=5, algorithm='backprop’, stepmax=10000). After conversed the = plots (rep =3) plots. After neuralnet: compute (n, training [, -1]) and test neuralnet: compute (nn, testing [, -1]) using ifelse(p3>=.9,1, ifelse(p3>=.8,2, ifelse(p3>=.7,3, ifelse(p3>=.6,4, ifelse(p3>=.5,5, ifelse (p3 >= .2,6, ifelse (p3>=.1,7,0))))))) produced the histogram of output as.

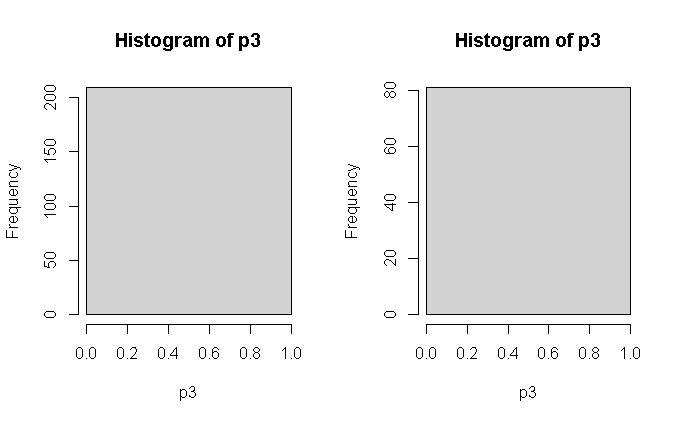


Figure 55: Half Model Histogram of Training and Test data

The confusion matrix and test and training accuracy of the model are as

pred1 1 2 3 4 5 6 pred1 1 2 3 4 5 6

1 14 27 15 12 11 2 1 14 27 15 12 11 2

After using training 1-sum(diag(tab3))/sum(tab3) accuracy is 0.8206897 percentage and test accuracy is 0.8271605 both models exactly same accuracy is best model in half data. Whose mismatched output after cbind (training$GPAA, pred1) function describes each induvial student records.

## **Back Propagation of Two Hidden Layers (N = 7, 3)**

After loading full data base, the sub setting of data with 14 independent and 1 independent variable using tidyverse library with all 580 sample student data. Then training (401) and test 179 data random sample were split 70:30 ratio. After using seed set. seed (222) the neural network with back propagation with two hidden layers of 7,3 as nn=neuralnet (GPAA ~ SLC+ Plus +Physics+ Math+ Chemistry +Bio +Parent1 +X1+ X2+ X3+ X4+ X5 + X6, data = data, hidden = c (7, 3), learning rate = 0.01, err.fct ="sse”, linear. output = FALSE, lifesign='full’, rep=5, algorithm='backprop', stepmax = 10000).

After setting learning rate remover in each iteration is 0.01 and algorithm is backprop satisfied five iteration tests with stepmax is 10000 produced.

hidden: 7, 3 thresh: 0.01 rep: 3/5 steps: 1000 min thresh: 0.0263606432485815

2000 min thresh: 0.0130459729524676 2601 error: 858.01 time: 2.42 secs

hidden: 7, 3 thresh: 0.01 rep: 4/5 steps: 1000 min thresh: 0.0192861629359062

2000 min thresh: 0.010537316400092 2121 error: 858.01 time: 1.83 secs

hidden: 7, 3 thresh: 0.01 rep: 5/5 steps: 1000 min thresh: 0.0341629139786731

2000 min thresh: 0.0169608886403354 3000 min thresh: 0.0112709333735587

3379 error: 858.01 time: 2.98 secs

From the above output the best model plot is chosen to using print (n, rep =3) incase third iteration is less time completion to formed final conversed single output.

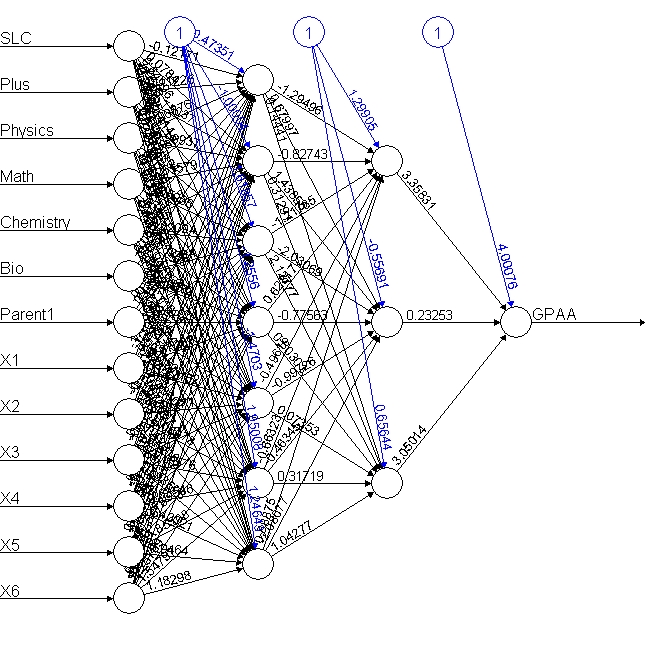
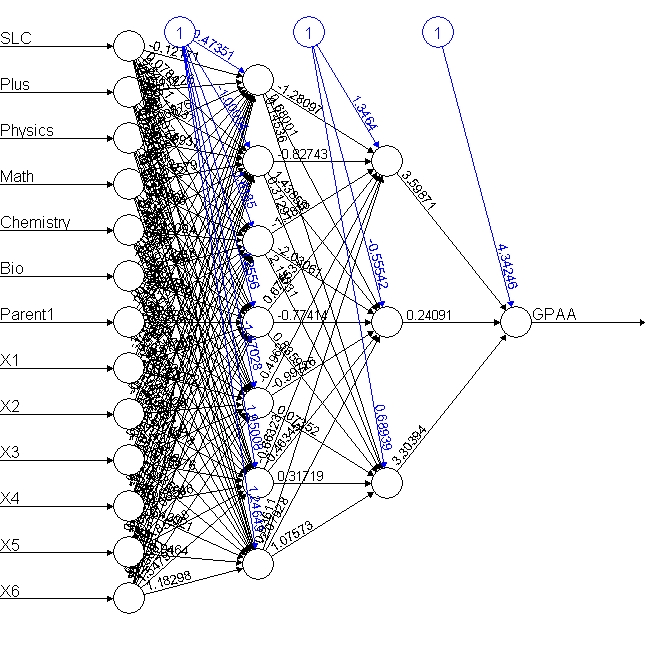
 

Figure 56 : Neural Network with Two Hidden Layer N = 7,3

After model conversed the compute function compute using neuralnet: compute (nn, data [, -1]) function. After using output$net. result function the output plots histogram.

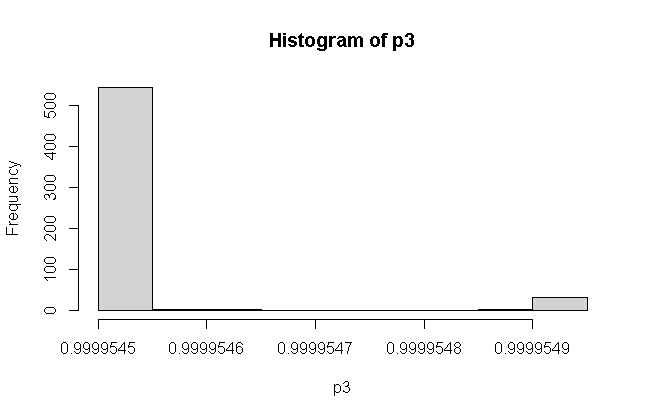


Figure 57: Histogram of Conversed Using Two Layers

After using ifelse(p3>=.99995485,1, ifelse(p3>=.99995480,2, ifelse(p3>=.99995475,3, ifelse(p3>=.9999547,4, ifelse(p3>=.99995465,5, ifelse(p3>=.9999546,6, ifelse (p3 >= 99995455,7,0))))))) to separate different grades of student marks in GPA variable. The confusion matrix is describing training and test accuracy as.

pred1 0 1 2 3 4 5 6 7

pred1 1 2 3 4 5 6 7

1 2 0 2 0 1 0 0

2 1 1 1 0 0 0 0

6 49 92 64 37 34 4 2

0 271 47 90 63 36 33 4 2

1 18 5 3 4 1 2 0 0

6 1 0 0 0 0 0 0 0

The full model accuracy using 1-sum(diag(tab3))/sum(tab3) function is 0.5241379. whereas the half model is 0.7689655 increased largely. the mis classification is again compared with original values gives percent of errors using cbind (data$GPAA, pred1). Therefore, this model is only predicting 52 percent whose model prediction and input were largely unmatched the model predicts failed but student grade is A- so on missed matched as [10,] 2 0 [11,] 1 1 [12,] 2 1 [13,] 5 1 [14,] 5 0 [15,] 3 1 [16,] 2 0 and [31,] 3 1 [32,] 4 6.

## **Back Propagation of Three Hidden Layers (N= 7,3,2)**

Back prorogation is the concept of reverse traversing for the purpose of balancing each neuron weight between input and output weight so that the model converse quickly requires the learning rate as hyperparameter to dropout some learning rate value quickly in each step repetition. After using the same data sets. nn=neuralnet (GPAA~ SLC+ Plus +Physics+ Math+ Chemistry +Bio+ Parent1+ X1+X2+ X3+X4+ X5+ X6, data = ss, hidden = c (7,3,2), learning rate=0.01, algorithm = 'backprop' err.fct ="sse", linear. output = FALSE). Back prorogation conversed very faster than forward traversing and produced as the similar output of each neuron calculation because each iteration the learning rate (0.01) in each parameter reduced with drop out and their respective weight calculation adjusted when reverse traversing.

The nn model describes its error rate threshold and steps as SLC.to.1layhid1 -1.650261e+00

Plus.to.1layhid1 -3.660134e-01 Physics.to.1layhid1 -3.162966e-01 Math.to.1layhid1 -198817e-01 Bio.to.1layhid1 -6.228716e-01 Parent1.to.1layhid1 -3.340415e-01 variables with start weights [1,] 0.6687143 0.27589340 -0.26063939 -0.83758243 -0.1984163 0.70073352 -0.6061511 [2,] -1.6501009 0.50627262 -0.41441971 and plots the conversed neural network as.

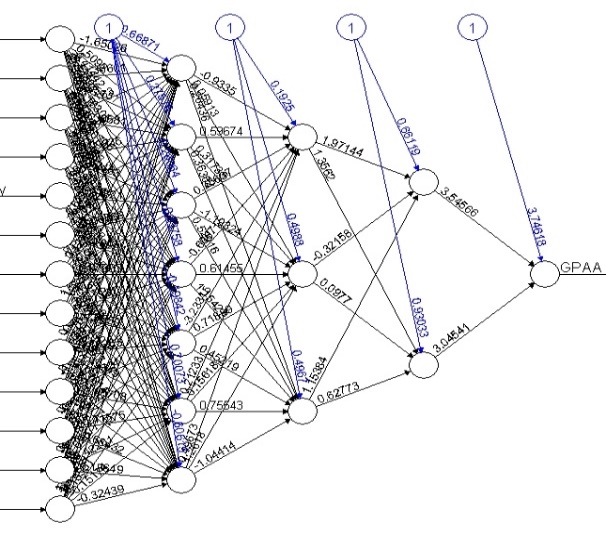
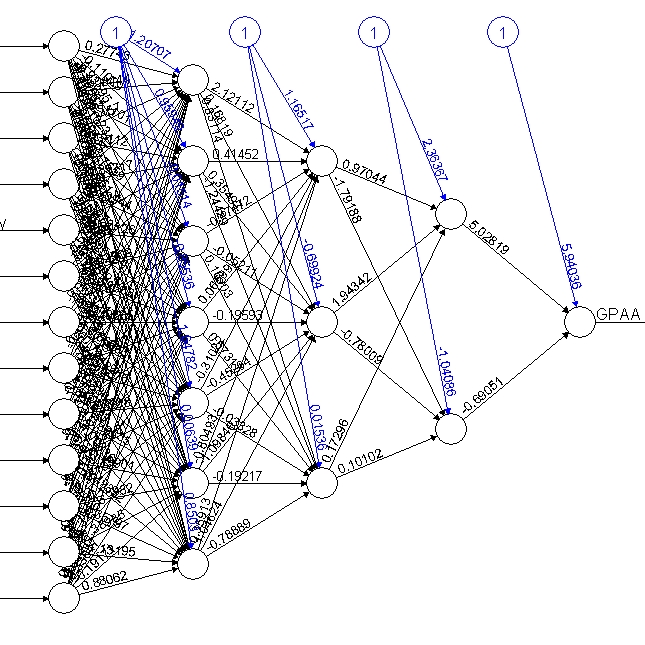
. 

Figure 58: Full and Half Neural Network Model using N= 7,3,2

The final plot conversed as above however this neural network conversed all others model with increasing hidden numbers but our primary concern is to converse in a single prediction. The model $net. result display the final weight generated at last GPA node further computed using compute with original full data for the histogram plot.

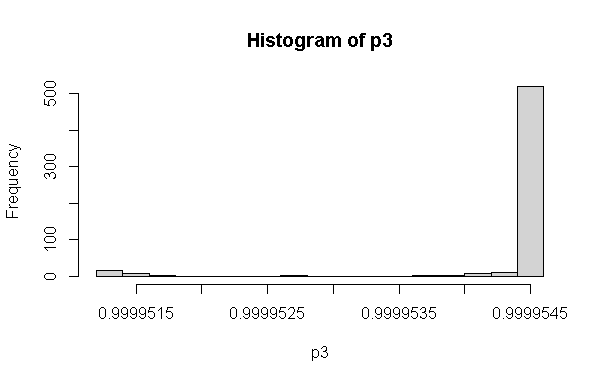
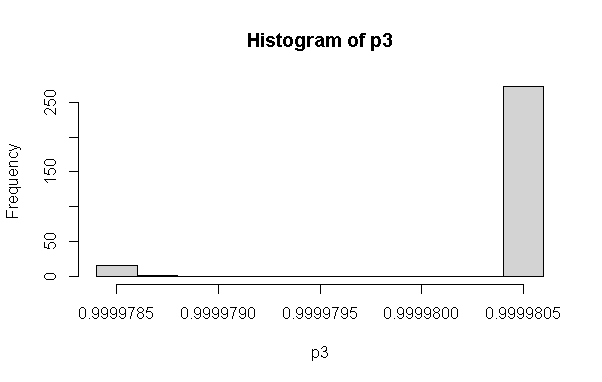
 

Figure 59: Full and Half Model Histogram of Neural Network

The model output from the neuron network is further categorized on histogram data sets

pred1=ifelse(p3>=.9999544,1, ifelse(p3>=.9999540,2, ifelse (p3 >=.9999535,3, ifelse(p3> =.9999530,4, ifelse(p3>=.9999527,5, ifelse(p3>=.9999525,6, ifelse(p3> 9999520,7,0))))))) concept of all records and confusion matrix is generated using tab3=table (pred1, ss$GPAA).

pred1 0 1 2 3 4 5 6 7

0 20 3 4 2 1 0 0 0

1 253 45 88 61 34 34 4 2

2 12 3 0 2 1 1 0 0

3 2 1 0 1 1 0 0 0

4 1 0 0 1 0 0 0 0

5 2 0 1 0 0 0 0 0

pred1 1 2 3 4 5 6 7

1 46 90 62 36 33 4 2

7 6 3 5 1 2 0 0

When using 1-sum(diag(tab3))/sum(tab3) produce full model is 0.8862069 percent total accuracy whose mis classification using mean (data$GPAA! = pred1) is 0.8862069. similarly, the half model produce 0.8310345 percentage and misclassification is 0.8413793. Then after using cbind (data$GPAA, pred1) to compare individual miss matched data as [478,] 0 1 [479,] 5 1 [480,] 3 1 [481,] 3 0 [482,] 2 1 [483,] 4 1 [286,] 3 1 [287,] 1 1[288,] 3 1 [289,] 2 1 [290,] 3 1 the mismatched model prediction vs actual grades of sample students.

Here researcher tries to predicts with model output of other students’ grades based on their pass and fail grade of eight grade variable of new predication based on model. The new. output=neuralnet: compute (predication based = matrix (c (1,1,1,1,1,1,1,1,1,1,1,1,1,

1,1,1,1,1,1,1,0,0,0,0,0,0,

1,1,1,1,1,1,1,1,1,1,1,1,0,

1,1,1,0,0,0,0,0,0,0,0,0,0),

byrow = TRUE, ncol = 13))

As from the model takes 13 different input signals formed the output after converse in single point model the new. output$net. result interpretation results with percentage of above input values [,1] [1,] 0.9999804 [2,] 0.9999800 [3,] 0.9999804 [4,] 0.9999799 suggest that first three individual input is higher than the last input lowest scored as compared with fourth candidate whose 13 input values are in the range of (0 to 1) grades of GPA values.

Similarly new. output=neuralnet: compute (nn, covariate = matrix (c (1,1,1,1,1,1,1,1,1,1,1,1,7,

1,1,1,7,1,1,1,1,1,1,1,6,1,

1,1,1,1,1,1,1,1,1,1,5,1,1,

1,1,1,1,1,1,1,1,1,1,1,1,1),

byrow = TRUE, ncol = 13))

Which produced new. output$net. result indicates the result with percentage of above input values are [,1] [1,] 0.9999805 [2,] 0.9999771 [3,] 0.9999804 [4,] 0.9999804 respectively.

When comparing using new. output=neuralnet: compute (nn, covariate = matrix (c (

80,70,55,70,70,68,1,40,45,45,45,45,40,

60,1,60,60,1,60,1,30,30,0,30,1,30,

88,80,0,60,60,60,1,35,35,35,35,35,30,

1,1,1,1,1,1,1,1,1,1,1,1,1),

byrow = TRUE, ncol = 13))

The new. Output$net. result# interpretation result with percentage of above input values are [,1] [1,] 0.9999805 [2,] 0.9999771 [3,] 0.9999804 [4,] 0.9999804 receptively parentage.

## **Neural Network with Four Hidden Layers (N= 25, 12, 7, 3)**

After loading data sets of original data base, the data management before processing is designed to converted into numeric data then model of those student who had passed in grade using four hidden layer neural model (25,12,7,3) as nn=neuralnet (GPAA ~SLC+ Plus+ Physics + Math+ Chemistry+ Bio+Parent1+ X1+ X2 +X3+ X4+X5+X6, data = data, hidden =c (25,12,7,3), learning rate=0.01, algorithm='backprop', err.fct="sse", linear. output =FALSE). The model produces all the weights and calculation of each weight of intermediator layers neurons as 3layhid6.to.4layhid3 6.500746e-01 Intercept.to.GPAA 4.283506e+00

4layhid1.to.GPAA 8.516410e-01 4layhid2.to.GPAA 5.310168e+00

4layhid3.to.GPAA 5.215656e-01 and the plot of neural network

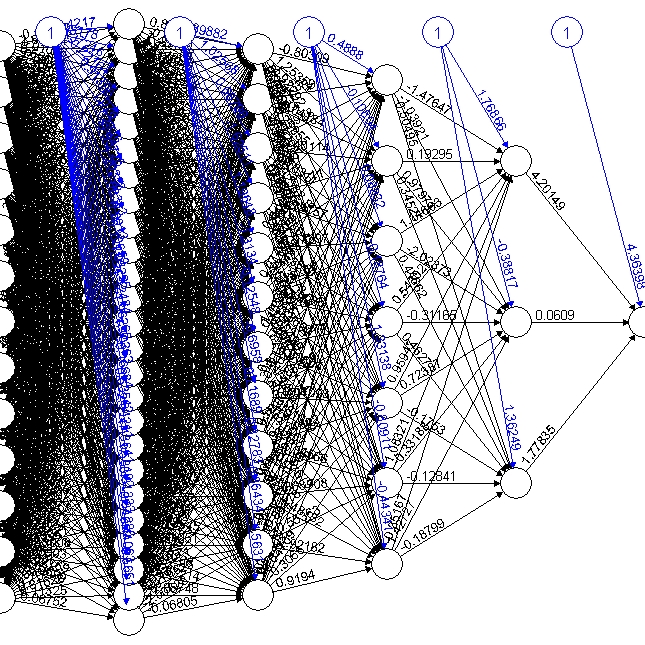
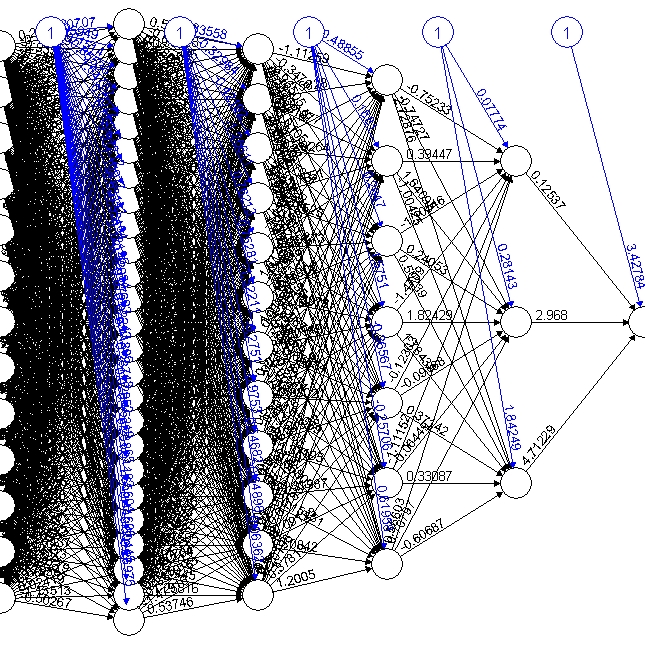
 

Figure 60: Full and Half Model Neural Network of (N= 25,12,7,3)

This model converses quickly with the output at last node [576,] 0.9999556 [577,] 0.9999553

[578,] 0.9999565 [579,] 0.9999560, [580,] 0.9999548 indicates the various grade of student. After output=neuralnet: compute (nn, data [, -1]) and p3=output$net.result the histogram

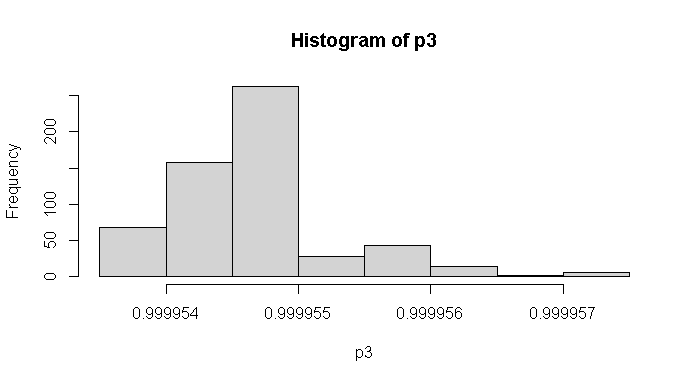
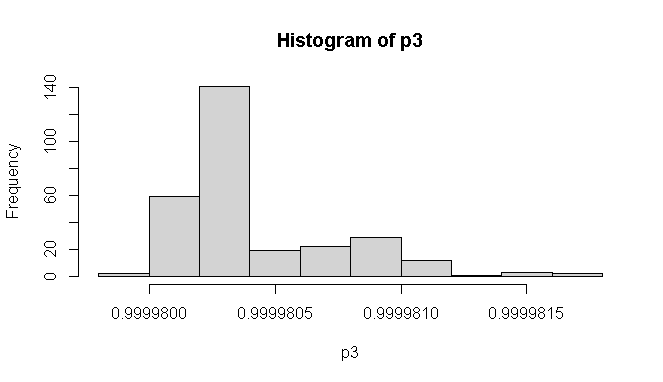
 

Figure 61: Histogram Output Neuron of Four Layer Network

After calculating the predication using pred1= ifelse (p3>=.9999565,1, ifelse (p3>=.99995627,2, ifelse (p3>=.99995585,3, ifelse (p3>=.99995542,4, ifelse (p3>= .9999550,5, ifelse (p3 >= .9999546, 6, ifelse (p3>=.9999542,7,0))))))). The prediction is computed with confusion matrix tab3=table (pred1, ss$GPAA)

pred1 0 1 2 3 4 5 6 7

0 55 6 17 8 3 5 0 0

1 5 0 0 1 0 0 0 0

2 1 0 0 2 0 0 0 0

3 11 4 1 1 0 2 0 0

4 19 2 8 4 0 3 0 1

5 19 2 2 1 2 0 0 0

6 35 3 10 6 3 5 2 0

7 145 35 55 44 29 20 2 1

pred1 1 2 3 4 5 6 7

1 0 1 2 0 2 0 0

2 1 0 0 0 0 0 0

3 4 2 5 0 1 0 0

4 6 19 9 4 7 0 1

5 8 2 2 3 0 0 0

6 25 54 30 17 20 3 1

7 8 15 19 13 5 1 0

The total accuracy using full model 1-sum(diag(tab3))/ sum(tab3) is calculated [1] 0.8982759 percent and half model is 0.9586207 percentage however there were mismatched records [213,] 5 6 [214,] 2 7 [215,] 4 6 [216,] 2 6 [217,] 4 7 [218,] 5 6 [86,] 4 4 [87,] 4 4 [88,] 4 7 [89,] 5 6. Therefore, while calculating neuron in between input to output neuron, it is always recommended to use the square root of neuron needs to be selected in between neural network model. So that more accurately produced. Likewise, if we eliminate the variables like school, parent2 and X5 in the model there is more result inaccuracy will be achieved.

## **Hypothesis Testing**

Hypothesis testing is an essential procedure in statistics. A hypothesis test evaluates two mutually exclusive statements about a population to determine which statement is best supported by the sample data (Anusiewicz & Patrician, 2021). This finding is statistically significant based on a hypothesis test. The variables used in association with other variables are concerning their significance or not. The null hypothesis and alternative hypothesis are statements regarding the differences or effects that occur in the population. A test of significance is a formal procedure for comparing observed data with a claim of the truth of which is being assessed.

After loading the dataset, the final GPA with respective marks and GPA with eight-level categorical numerical values which had passed grade were analyzed. The t-test with average grade between internal with final grade further tested. The t.test (ss$Int, ss$FinalGPA, alternative ='two. sided’, mu=0) with two-sided tested. Which produces the output Welch two-sample t-test data as ss$Int and ss$FinalGPA t = 9.1437, df = 484.33, p-value < 2.2e-16. And the alternative hypothesis is a true difference in means is not equal to 0. 95 percent confidence interval is 5.479536 8.479085 and sample estimates is mean of x to mean of y 83.25517 of internal marks mean and 76.27586 is final marks mean implies internal marks is 7 marks higher than final marks scored of SOE student. The p-value is significant than the reject null hypothesis is that internal and final marks are large differences between when two side tests. After making factorial variables of the different program of 13 currently running Pokhara University the one-way ANOVA test in various groups data the ANOVA is analysis in group data.

Using ano=aov(sss$FinalGPA~sss$Prog) and summary(ano) produces the result as.

Df Sum Sq Mean Sq F value Pr(>F)

sss$Prog 12 1547 128.92 2.944 0.000715 \*\*\*

Residuals 277 12131 43.79

The final GPA marks of13 programs of the school of engineering apply in one-way ANOVA test indicates there is significance when comparing its p values conclude that there is no random chance to become such association between internal and final grade therefore further analysis is taken as Tukey test. Tukey multiple comparisons test of each means within 13 programs 95% family-wise confidence level produces the output with lower and upper range.

diff lwr upr p adj

Civil III-Civil I 0.04166667 -6.654460 6.7377932 1.0000000

Civil V-Civil I 3.32427536 -2.244791 8.8933420 0.7354310

Civil VII-Civil I -1.08333333 -6.793809 4.6271425 0.9999891

Civil VIII-Civil I 0.74945887 -4.420983 5.9199004 0.9999995

Computer-Civil I -4.98774510 -11.998742 2.0232518 0.4630066

Elex III-Civil I -2.22756410 -9.843819 5.3886912 0.9989535

Software-ELEX VIII -4.30555556 -14.058050 5.4469386 0.9603802

Rural-Elex 4.16666667 -6.891622 15.2249556 0.9890248

Software-Elex -1.38888889 -11.141383 8.3636053 0.9999996

Software-Rural -5.55555556 -17.212016 6.1009045 0.9309362

Therefore, final p adjusted values indicate each program wise association of p values have not found significant indicates these variables among various groups have association but not significant with comparing p value with <0.05 confidence.

## **Multi-nominal Logistic Regression**

The multi-nominal logistic regression is applying when a predicative variable is in numeric with dependent more than one categorical target variables has not the order of grade and parent education status may play significant role (Davidson, 2014). For example, the color, red, blue, yellow is the category of three income high, middle and low in order but not specific order. The multinomial or multi probity model was suitable but the assumption behind is the error terms were being equal distribution in all terms of values are being tested value. The multinomial logistic regression is used to predict categorical placement in or the probability of category membership on a dependent variable based on multiple independent variables. The independent variables can be either dichotomous (i.e., binary) or continuous format.

After loading data probity <- read.csv ("D:/dataanalysisfinal/SOE/rcode/Dataanalysiscsv.csv", stringsAsFactors = TRUE) command loads database into r console. The str command display structure of the database. After using tidyverse package for sub setting of data those whose data type in numeric selects using select command. After converting data into a data frame, the predictor variable GPA is converted into a numeric ordinal factor of 0,1,2,3,4,5,6 and 7 (eight levels grades). After using seed value, the sample of 70, 30 percent split of training and test (410,170) respectively. To run multinominal regression the predictor variable needs, relevel based with fail 0 levels. After loading nnet package the multinomial regression with training data set is executes as mymodel = multinom (GPAA ~ SLC + Physics + School + Plus + Physics + Math + Chemistry + Bio + Parent1 + Parent2 + X1 + X2 + X3 + X4 + X5 + X6, data = training). The model weights are 136 (112 variables) with initial value 852.571032, iter 10 value 635.228852, iter 20 value 623.903040, iter 30 value 617.383953 iter 40 value 601.397270, iter 50 value 574.168467, iter 60 value 560.407016, iter 70 value 554.530189, iter 80 value 543.392866, iter 90 value 533.969108, iter 100 value 520.413656. and the final value 520.413656, stopped after 100 iterations. This indicates initially the model has 852 accuracy when it reached 100 iteration it converses and minimized 520 accuracy in model development.

The summary model describes the coefficients and standard errors as.

(Intercept) SLC Physics School Plus Math Chemistry Bio Parent1

1 -10.9127982 0.029127015 -0.07860349 -0.9305316 0.01720708 0.03500580 0.044396725 -0.004091035 0.71126846 2 -10.7562741 0.021874206 0.04449264 -0.1189089 0.03321832 -0.04184734 -0.007991546 0.006490877 0.20853753 3 -9.4884325 0.006798686 0.01526304 0.4489333 -0.01672852 -0.01175935 0.017425611 -0.006293013 0.03648324

After comparing the coefficient of each variable with 0.05 threshold, the model input for the prediction to student GPA is rejected to these variables are more than the p-value. The explanation is when log odds of category base zero when other components rise either + or - negative by the number of respective intercepts. The summary is calculated as z=summary (mymodel) $coefficients / summary (mymodel) $ standard. errors the z value of each variable to its coefficient and standard errors. The probability p=(1-pnorm(abs(z),0,1)) \*2 0 means and 1 standard deviation 2 for two tail tests. Which produces each probability of with reference 0 fail value regression. If all values of variables are greater than 0.5 will be drop to create final model. The final model weights are 104 (84 variable) conversed at 524 after 100 iterations. The summary command describes coefficient and standard errors. The prediction final model to training test data of 410 records with residual deviance is 1049.079 and AIC is 1217.079. The prediction of p=predict (mymodel, training) with training data produce the confusion matrix is 0.4560976 accuracy and test accuracy are 0.5823529 is low enough.

p1 0 1 2 3 4 5 6 7

0 65 15 24 22 12 9 2 1

1 1 0 0 0 0 0 0 0

2 4 3 6 3 0 0 0 0

3 0 0 1 0 0 0 0 0

4 0 0 0 0 0 0 0 0

5 0 0 0 0 0 0 0 0

6 1 0 0 0 0 0 0 0

7 1 0 0 0 0 0 0 0

p 0 1 2 3 4 5 6 7

0 212 25 56 39 24 26 0 0

1 0 4 2 2 0 0 0 0

2 3 5 4 1 1 0 0 0

3 0 0 0 0 0 0 0 0

4 2 0 0 0 0 0 0 0

5 1 0 0 0 0 0 0 0

6 0 0 0 0 0 0 2 0

7 0 0 0 0 0 0 0 1

However, the percentage is calculated of each grade is calculated of its diagonal are (n/sum(n)

zero fail is 0.53, 1 A+ is 0.08, 2 A- is 0.15,3 B+ is 0.06,4 B is0.06 and B- is 0 respectively.

And training and test accuracy of each category GPA as.

p 0 1 2 3 4 5 6 7

p1 0 1 2 3 4 5 6 7

0 0.90 0.21 0.33 0.31 0.17 0.12 0.03 0.01

1 0.06 0.00 0.00 0.00 0.00 0.00 0.00 0.00

2 0.13 0.10 0.19 0.10 0.00 0.00 0.00 0.00

3 0.00 0.00 0.04 0.00 0.00 0.00 0.00 0.00

4 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00

5 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00

6 0.50 0.00 0.00 0.00 0.00 0.00 0.00 0.00

7 1.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00

0 0.97 0.11 0.26 0.18 0.11 0.12 0.00 0.00

1 0.00 0.12 0.06 0.06 0.00 0.00 0.00 0.00

2 0.05 0.08 0.06 0.02 0.02 0.00 0.00 0.00

3 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00

4 0.08 0.00 0.00 0.00 0.00 0.00 0.00 0.00

5 0.04 0.00 0.00 0.00 0.00 0.00 0.00 0.00

6 0.00 0.00 0.00 0.00 0.00 0.00 1.00 0.00

7 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1.00

Therefore, from the above output, both training and test accuracy of fail student using multinomial is 90 to 97 percent accuracy. The A grade student is 11 to 21 percent, A- is 26 to 33 percent, and so on. Thus, a multinomial test of regression is highly applicable in eight grade prediction of student grade prediction.

## **Correlation and Linear Regression**

On consideration student GPA prediction of bachelor of engineering-grade from the high school grade, plus to grade and their major science scored like (Physics, Math’s, Chemistry, Bio) and their semester internal grade performance scored. The student performance of the university student test using Pearson correlation shows that there is a positive correlation between entry test and Cumulative Grade Point Average (CGPA) between HSSC and CGPA. The correlation coefficients (r) between the entry test with CGPA, HSSC and CGPA are very close indicating both entry test and HSSC are equally important in predicting the CGP. The correlation coefficients are used to measure how strong a relationship is between two variables (Iannaccone, Conte, Cortis, & Fusco, 2021). There are several types of correlation coefficients, but the most popular is Pearson. The Pearson’s correlation is a correlation coefficient commonly used in linear regression. Pearson's correlation is utilized when you have two quantitative variables and if there is a linear relationship between those variables. Our research hypothesis stating that one score affects the other in a certain way. The correlation is affected by the size and the sign of the linear regression is the next step up after correlation. Linear regression models are used to predict the relationship between two variables or factors. The factor that is being predicted is called the dependent variable. The linear regression use lm () assumes that the error terms associated between the observation vector and predictor variables are independent and identically distributed with mean zero and constant variance. Whereas general linear regression uses glm () to generalize the linear model into what is known as the exponential family which allows for the observations to be taken on different random variables like binomial and multinomial etc. When using glm () specify the family that the observations fall under, and in the linear assumption that the random variable is normal (family=Gaussian).

After loading datasets, the average internal and final marks of GPA scored student background information were converted into their respective numbers of only successful of students of 290 sample datasets. The scatter plots are generally viewing before analyzing the data using plot (ss$FinalGPA, ss$Int/2, main="Scatterplot”, ylab = "GPA”, xlab = "Internal", las=1). The correlation between internal and final grade is calculated using cor (ss$Int, ss$FinalGPA, method ="Pearson") function intercept is 0.1183738 which indicates these two variables have a positive but not strong correlation. Similarly, the cor.test (ss$Int, ss$FinalGPA, method = "Pearson”, alt= "greater”, conf. level=.99) while calculating correlation use Pearson with confidence level also produce the similar output. The cor.test function display the output Pearson’s product on ss$Int and ss$FinalGPA variables is 2.0231, df is 288 and p-value is 0.02199 indicates these two variables are significant with 99 percent confidence interval lies in between -0.01838648 to 1.00000000 is 0.1183738 percent. The linear regression between them is calculated using model=lm(ss$FinalGPA~ss$Int) describes its descriptive statists and confident indicates both model and the internal variable is significant whose variability in model is 14 percent.

            Estimate Std. Error t value Pr(>|t|)

(Intercept) 70.12823 3.06518 22.879 <2e-16 \*\*\*

ss$Int 0.07384 0.03650 2.023 0.044 \* ---

Residual standard error is 6.843 on 288 degrees of freedom, the multiple r-squared is0.01401, adjusted r-squared is 0.01059, the f-statistic is 4.093 on 1 and 288 and model p-value is 0.04399 significant. After loading the corrplot library the correlation of all independent variables describes the internal marks of 6 subjects has the highest correlation with varying 71 % to 50 % association. The subject Chemistry, Bio, School from and Parent educations have least and -ve association among them which plots describes the overall correlation of research data. The final student GPA mean is 76.25 whereas the internal mean is 80.9 indicates the internal marks of the student are overall higher than the final student grade.

.

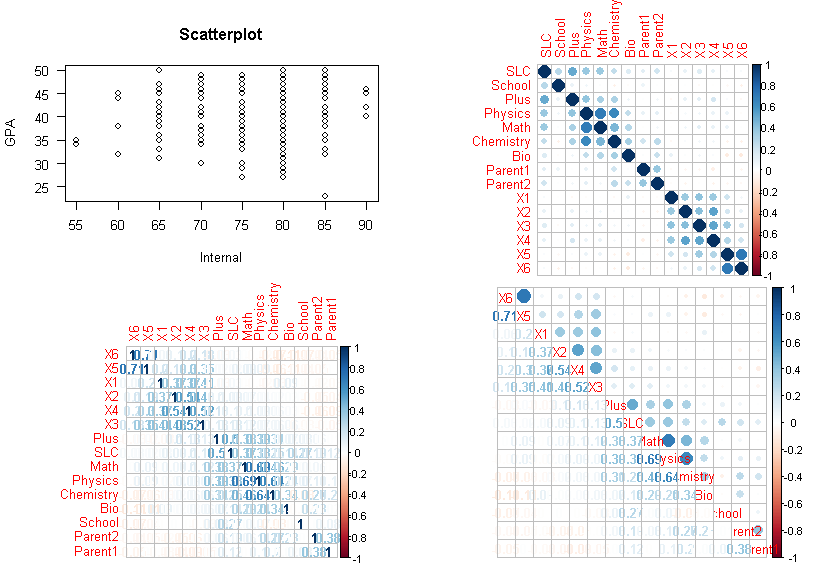


Figure 62: Correlation Plots

Thus, to the correlation between internal and final marks the internal marks need to be reduced up to 7 marks deflection of all internal subjects. After applying internal and final GPA marks in linear model produce’s the model plots indicate similar results.

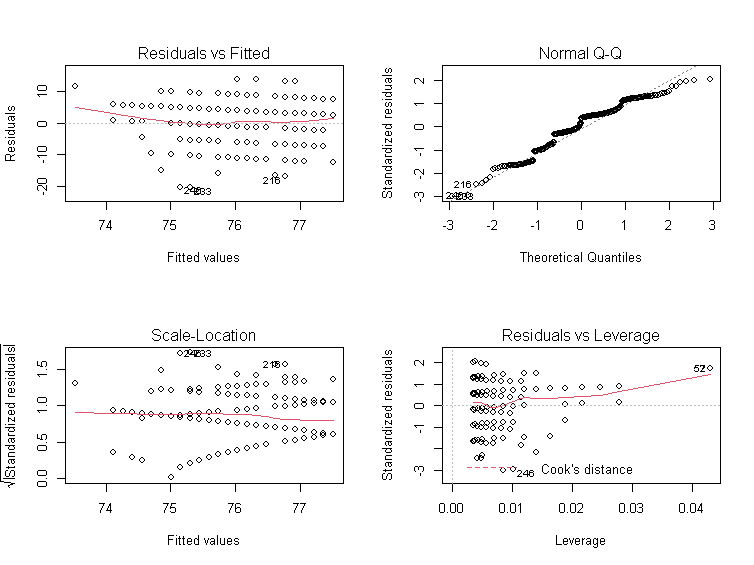


Figure 63: Residual and QQ Plots

After loading the catools package for splitting data into train and test set the sample. split command split data into 80% to 20 percent of 580 student records then the subset (463:117) for model evaluation. After omitting NA values from the database, the predictor variable pf stored 1 for passed student and 0 for fail students is the dependent variable to the model. The general regression model with two binary predictor variables with other 14 independent variables apply general linear regression model with binomial family. The model=glm (pf~SLC+School+Plus +Physics+Math+Chemistry+Bio+Parent1+Parent2+X1+X2+X3+X4+X5+X6, data=train, family="binomial") is used for analysis its coefficient and intercept produce the output using summary model. The coefficients and descriptive statists describe as.

Estimate Std. Error z value Pr(>|z|)

(Intercept) -8.8560840 1.5933912 -5.558 2.73e-08 \*\*\*

SLC 0.0336644 0.0157239 2.141 0.03228 \*

School -0.1063724 0.2175241 -0.489 0.62483

Plus 0.0063139 0.0143357 0.440 0.65962

Physics 0.0136562 0.0176905 0.772 0.44014

Math -0.0089775 0.0128445 -0.699 0.48459

Chemistry 0.0005617 0.0148844 0.038 0.96990

Bio 0.0013824 0.0059483 0.232 0.81623

Parent1 0.1446625 0.1052733 1.374 0.16939

Parent2 -0.0619289 0.1256502 -0.493 0.62211

X1 0.0638051 0.0195620 3.262 0.00111 \*\*

X2 0.0496246 0.0216459 2.293 0.02187 \*

X3 -0.0409040 0.0273335 -1.496 0.13453

X4 0.0562667 0.0254340 2.212 0.02695 \*

X5 0.0184650 0.0188549 0.979 0.32742

X6 -0.0165861 0.0161270 -1.028 0.30373

From the above output, the -8.856 is intercepted (B0) with each variable with a respective coefficient to predict student result. The -ve intercept variables whose intercept indicates the -ve association therefore the final model becomes less model=glm (pf~SLC+Plus+Physics +Parent1+X1+X2+X4, data=train, family="binomial") become the model with the training set with eliminating non-significant values. From the above p-value, the internal marks of bachelor’s grade have the highest significance to variables like SLC, plus marks, physics and parent education have a strong association for the prediction to result in binary classification. After calculating prediction train and test with response the confusion matrix cmatrix=table (Actual=train$pf, predicted=res>0.5) is compared with greater than 0.5 percent threshold accuracy. The accuracy is calculated using accuracy= (cmatrix [[1,1]] +cmatrix [ [2,2]])/sum(cmatrix) in training test is 0.6349892 percent and test set is 0.5897436 respectively.

Predicted

Predicted

Actual FALSE TRUE

0 30 25

1 23 39

Actual FALSE TRUE

0 146 89

1 80 148

The first prediction is 463 student records and second 117 of test data sets. Similar prediction is found in the output plots using model.

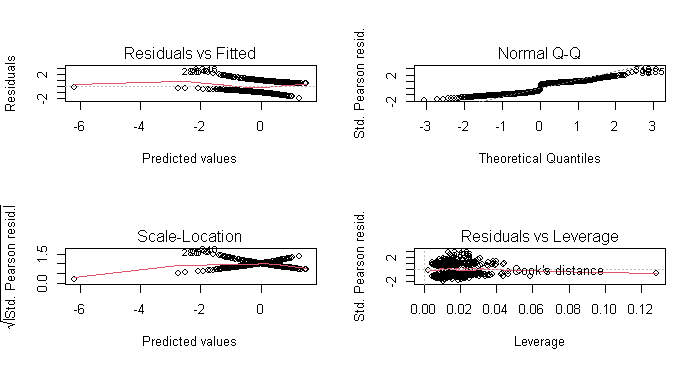


Figure 64: Residual and QQ Plot

Similarly, the program-wise data of each program regression could analyze after converting program data into factor of each program record. The table records of each program are displayed using a table(waterproof) command describes the total sample of program wise records. Civil I Civil III Civil V Civil VII Civil VIII Comp Elex III Elex V Elex VII ELEX VIII ElexI Rural Software as 48 40 92 80 154 34 26 12 16 24 24 12 18. The model dot matrix function creating tables of dummy variables for use in statistical. Note that there was no header row in the data file, so the variable name had to be added model. Matrix(~parasite-1) then header of converted dummy variables of each program is rewrite names(k)[names(k) == "data$ProgCivil I"] <-"CivilI" like this of all 13 programs of Pokhara University engineering courses. The xtab detail information between GPA with program will be display as.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| GPAOR | Civil I | Civil III | Civil V | Civil  VII | Civil  VIII | Comp | Elex  III | Elex  V | Elex  VII | Elex  VIII | Elex I | Rural | Soft |
| (F)0 | 24 | 20 | 46 | 40 | 77 | 17 | 13 | 6 | 8 | 12 | 12 | 6 | 9 |
| (A)1 | 3 | 2 | 18 | 4 | 16 | 1 | 2 | 2 | 1 | 1 | 1 | 0 | 1 |
| (A-)2 | 9 | 9 | 13 | 16 | 24 | 3 | 3 | 1 | 3 | 3 | 3 | 4 | 2 |
| (B+)3 | 8 | 4 | 10 | 9 | 19 | 3 | 1 | 2 | 2 | 6 | 3 | 0 | 0 |
| (B)47 | 0 | 3 | 3 | 3 | 13 | 5 | 5 | 0 | 0 | 0 | 1 | 2 | 2 |
| (B-)5 | 4 | 2 | 2 | 6 | 4 | 4 | 2 | 1 | 2 | 2 | 3 | 0 | 3 |
| (C+6 | 0 | 0 | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| (C)7 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |

Table 5: Summary Table by Program

After converting dummy variables of each program data sample is split into test and training sets of 468 and 112 testing sets of 580 observations. Then the liner model with ordinal GPA with 13 programs. mod=lm (OrdGPA~CivilI+CivilIII+CivilV+CivilVII+CivilVIII+Computer+ElexIII +ElexV +ElexVII +ElexVIII +ElexI +Rural +Software, data, data=train) is used and display using summary statical as.

Estimate Std. Error t value Pr(>|t|)

(Intercept) 3.68750 1.07109 3.443 0.000628 \*\*\*

CivilI 0.72276 1.27197 0.568 0.570161

CivilI 0.04583 1.32631 0.035 0.972448

Civil 1.07566 1.17846 0.913 0.361839

CivilI 0.18015 1.19045 0.151 0.879784

CivilI 0.47083 1.14026 0.413 0.679859

Computer -0.28750 1.32631 -0.217 0.828485

ElexI 0.31250 1.40769 0.222 0.824416

Elex -0.18750 1.72708 -0.109 0.913595

ElexI 0.24583 1.53979 0.160 0.873223

ElexI 0.55060 1.42173 0.387 0.698734

ElexI 0.31250 1.45373 0.215 0.829890

Rural 1.31250 1.72708 0.760 0.447670

Software 1.31250 1.72708 0.760 0.447670

Residual standard error is 4.284 on 463 degrees of freedom. The multiple r-squared is 0.008771 and adjusted r-squared is -0.01692. The f-statistic: 0.3414 on 12 and 463 DF, p-value is 0.9811. The descriptive statists with min, max with coefficient explain with the explain model in average the intercept of each program has change increase or decrease. And the program like Computer, ElexV and Software have -ve association with eliminating these columns for final model. The final p values implies that there is no significance program wise regression model. However, training model accuracy is 0.8802521 and test accuracy is 0.9615385 percent respectively.

# **Chapter 5: Conclusion and Recommendation**

## **Performance Measures**

As the final student grade GPA is the collective efforts of their previous and ongoing efforts of each semester may predict accurately using the neural network which receives the input weight of each matrix elements of variables to another complete next neuron is calculated though output of previous nodes weight sum with bias terms using sigmoid function which pushes for producing signal weight to converse GPA grades of the student while doing so the researcher first collects all the background information of bachelor of engineering students database of 580. The full model and half model of neural network were designed using various parameters to predict the best model. The full model includes all 580 student records where half model includes 290 student those having pass result in fall 2019 examination. The neural network passes with single hidden layer first passes 22-input signal to formed either GPA whose confusion matrix, accuracy of both raining and test were predicated from the output neuron. Based on the dependent variables neural network output is again calculated its value into the number of subdivision grades. Due to the presence of some unimportant variable with similar pattern of marks to the model the prediction accuracy using forward and back propagation were being summarized in the following table.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **S. N** | **Predicator Variable** | **No of Input** | **No Hidden** | **Training Accuracy** | **Test Accuracy** |
| **1** | **P/F (0/1)** | **22 Full** | **1** | **91** | **81** |
|  | **P/F ()/1)** | **22 Half** | **1** | **98** | **95** |
| **2** | **GPA(Deci.)** | **22 Full** | **1** | **38** | **41** |
|  | **GPA(Deci.)** | **22 Half** | **1** | **82** | **83** |
| **3** | **GPAX(Cat.)** | **22 Full** | **1** | **98** | **95** |
|  | **GPAX(Cat.)** | **22 Half** | **1** | **98** | **95** |
| **4** | **GPA** | **13 Full** | **1** | **83** | **84** |
|  | **GPA** | **13 Half** | **1** | **99** | **93** |
| **5** | **GPA** | **13 Full** | **5** | **91** | **87** |
|  | **GPA** | **13 Half** | **5** | **92** | **93** |
| **6** | **GPA** | **13 Full** | **7,3** | **85** | **88** |
|  | **GPA** | **13 Half** | **7,3** | **75** | **80** |
| **7** | **GPA** | **13 Full** | **7,3,2** | **91** | **89** |
|  | **GPA** | **13 Half** | **7,3,2** | **81** | **82** |
|  | **Back Propagation** | | | | |
| **1** | **GPA** | **13 Full** | **1** | **48** | **52** |
|  | **GPA** | **13 Half** | **1** | **82** | **82** |
| **2** | **GPA** | **13 Full** | **5** | **50** | **91** |
|  | **GPA** | **13 Half** | **5** | **82** | **82** |
| **3** | **GPA** | **13 Full** | **7,3** | **52** | **52** |
|  | **GPA** | **13 Half** | **7,3** | **76** | **76** |
| **4** | **GPA** | **13 Full** | **7,3,2** | **88** | **80** |
|  | **GPA** | **13 Half** | **7,3,2** | **83** | **84** |
| **5** | **GPA** | **13 Full** | **25,12,7,3** | **82** | **82** |
|  | **GPA** | **13 Half** | **25,12,7,3** | **73** | **73** |

Table 6: Neural Network Model Accuracy

This research completes with the design 24 different neural network models whose accuracy on training and test set samples were presented in the table explains the best model. However, based on data input and their weight neural conversed to predicts student grade can be more increased if we reject these variables according to Boruta and Random Forest methodology suggestion. The three-layer backpropagations of original data of student scored prediction is more than 90 percent is a good result. When doing forward neural network with more variables as independent the GPAX with categorical numeric model with single neuron is best recommended where as 13 input and single with more neuron (n=5) is the best for both training and test accuracy.

## **Conclusion and Recommendation**

After applying various data preprocessing and machine learning algorithms of 580 student marks of Pokhara University School of Engineering student, we have achieved quite a remarkable prediction performance (varying between 48-99 % accuracy) of using 24 different neural networks model. The following useful conclusions have been drawn by this study:

Grade prediction model with 22 associative independent associative variables predicts 99% accuracy when using a single hidden layer of student can be accurately predicted even with the result of the final exam different grades. Similarly, when using 13 associative attributes input neurons with 7,3,2 three hidden layers predicts (92 to 98 percent accuracy of both single or full student grade prediction. Similarly, when using 7,3,2 with back propagation neural network produce 90 percent accuracy. This implies that the performance of students in various previous records before the final exam is highly correlated to their performance in the final exam except for second parent education. Another interesting point is that the F grade is the easiest to predict followed by the C grade, while B+ is the most difficult grade to predict. Furthermore, the number of students whose grades slip because of their performance on the final exam is more than those who improve it because of the final exam. The random forest methodology is used for the test of both classification and regression for data management whose output on test and training data sets is achieved 99 percent accuracy validate to our model. Likewise, the Boruta algorithm is used to eliminate many associative variables to predicts strong associative variables suggest 17 very significant variables out of 22 variables to the model.

This prediction will help the management to plan the courses that they are to offer in the upcoming semester more effectively thus making scoring better grades among faculty members less uncertain. This would help students in deciding whether to drop a course or not and lastly, counselors can timely warn students if they are failing. In the future, we would like to extend our model to recommend the courses that would maximize his GPA taking into consideration the type of the course, teacher of the course, and his past performance in such courses.

The prediction of student performance is getting difficult day by day. In this research, we have developed an ordinal regression based on 13 programs which will help students in knowing final grade in a particular subject predicts the accuracy (48 to 53%) accuracy. after testing the multicollinearity test on student data, the multinomial test predicts 90 to 97 % accuracy of student grade prediction. Similarly, the hypothesis testing on internal and final grade marks has a significant association the single ANOVA and Tukey test also express the similar result when program-wise result with final GPA of the student. To accomplish this research, internal exam marks out of 50 are taken into consideration. Then the marks are converted into 100 (percentage) to have a uniform benchmark in data management and preparation.

**Limitation and Delimitation of the Study**

Inclusion and exclusion criteria: Respondents of the selected batch student which were currently being studying were included.

**Exclusion Criteria**

Denials of the sampled participation student were excluded. We note that the reported findings of this study have been based on the dataset of the performance of the undergraduate students from Pokhara university School of Engineering. The dataset used in the study is limited with GPAs available for students in the particular courses based on the result of fall 2019.

**Ethical Consideration**

This study is going to explore the grade prediction and correlation of internal and final marks of student students. This observational and questionnaire survey study does not require any health-related issues of the respondent, therefore, the researcher himself will take first research consent from the School of Engineering, Pokhara University will be taken then multiple, choice questionnaire, short question answer and open questionnaire will be designed to measure the subject wise measurement of sample data.

## **References**

Adepoju, G., Ogunjuyigbe, S., & Alawode, K. (2007). Application of Neural Network to Load Forecasting in Nigerian Electrical Power System”. The Pacific Journal of Science and Technology.8(1):68-72.

Agoritsa, P., & George, K. (2019). Grade Prediction with Course and Student Speciic Model.

Agoritsa, P., & Karypis, G. (2016). Grade prediction with models specific to students and courses. *Received: 20 July 2016 / Accepted: 6 September 2016 / Published online: 22 September*.

Anastasiadis, A., Stefan, F., & Frauke, G. (2005). New globally convergent training scheme based on the resilient propagation algorithm. Neurocomputing 64, pages 253-270.

Anusiewicz, C., & Patrician, P. (2021). Measuring workplace bullying in a US nursing population with the Short Negative Acts Questionnaire. . *Research in Nursing & Health, 44(2), 319-328.*

Bhardwaj, B. (2011). “Data Mining􀯗: A prediction for performance improvement using classification. *Int. J. Comput. Sci. Inf. Secur., vol.*

Brecko, B. (2004). How family background influences student achievement. In Proceedings of the IRC-2004.

Brijesh Kumar, B., & Saurabh, P. (2012). Data Mining: A prediction for performance improvement using classification. *(IJCSIS) International Journal of Computer Science and Information Security, Vol. 9, No. 4, April 2011, pp 136-140*.

Brijesh Kumar, B., & Saurabha, P. (2011). E preceding of 4th Global summit. *Mining education Data Analyze student performance*.

Brinton, G., & Chiang, M. (2016). *Mining MOOC clickstreams: Video-watching behavior vs. in-video quiz performance.* Signal Process., vol. 64, no. 14, pp. 3677–3692, Jul.

Cohen, W. (1995). Effective rule induction. *Twelves international conference on machine larning*.

Daguang, W., & Xie, Z. (2017). Enhancing Teaching and Learning through. *Enhancing Teaching and Learning through Internal Quality Assurance Xiamen University, China*.

David, G., Rijt, G. v., Filip, D., & Janine, v. d. (2008). Students’ assessment preferences, perceptions. *High Educ (2008) 56:645–658*.

Davidson, J. H. (2014). Predicting persistence for first-time undergraduate adult students at four-year institutions using first-term academic behaviors and outcomes. *The Journal of Continuing Higher Education, 62(2), 78-89.*

El, M., & Yasser, M. (2008). Machine learning approaches for epitope. *Iowa State University*.

Emaan Abdul, M. (2016). GRADE PREDICTION USING SUPERVISED MACHINE LEARNING TECHNIQUES.

Filip, D. a. (2008). Students’ assessment preferences, perceptions. *High Educ (2008) 56:645–658 Open access*.

Fisher, J. (2014). Applied Learning in Online Education:A Comparative Study Employing. *Utah Valley University*.

Hamsa, H., Indiradevi, H., & Kizhakkethottam, J. (2016). Student Academic Performance Prediction Model Using Decision Tree and Fuzzy Genetic. *Procedia Technol., vol. 25, pp. 326–332,*.

Havan, A., & Harshil, M. (2015). Student Performance Prediction using Machine. *International Journal of Engineering Research & Technology (IJERT)*.

Hijazi, S., & Naqvi, R. (2006). Factors Affecting Student’s Performance: A Case of Private Colleges. . *Bangladesh e-Journal of Sociology. Vol. 3, No. 1.*

Iannaccone, A., Conte, D., Cortis, C., & Fusco, A. (2021). Usefulness of Linear Mixed-Effects Models to Assess the Relationship between Objective and Subjective Internal Load in Team Sports. . *International journal of environmental research and public healt*.

John, A., & Manabete, S. (2015). Procedural INflurance on Internal and External Assessment score of Undergraduage. *Journal of Education and Practices*.

Junemann, M. L. (2007). Neural networks to predict schooling failure/success. *Lecture Notes in Computer Science, Nature Inspired Problem-Solving Methods in Knowledge Engineering. Berlin: Sp*.

Kalejaye, B., Folorunso, O., & Usman, L. (2015). PREDICTING STUDENTS’ GRADE SCORES USING TRAINING FUNCTIONS OF ARTIFICIAL NEURAL NETWORK. *Journal of Natural Sciences, Engineering and Technology, Volume 14, 2015 (formerly ASSET Journal Series B)*.

Karamouzis, S. (2000). *Using Analogies to predict Student performance.* In Proc. IASTED’s Artificial Intelligence and soft comp. Banf, Canada, P.355-360.

Lima, K., Mohanan, D. T., & Harichandran, S. (2017). Association of class attendance and academic performance of MBBS. *National Journal of Physiology, Pharmacy and Pharmacology*.

Lorenzen, S., Pham, N., & Alstrup, S. (2017). On predicting student performance using low-rank matrix factorization.

Lykourentzou, G., Mpardis, G., & Nikolopoulos, L. (2009). Early and Dynamic Student Achievement Prediction in E-Learning Courses Using Neural Networks.

Lykourentzou, I., Giannoukos, I., Mpardis, G., Nikolopoulos, V., & Loumos, V. (2009). Early and Dynamic Student Achievement Prediction in E-Learning Courses Using Neural Networks. Journal of the American Society for Information Science and Technology, 60(2):372-380.

Maghari, A. (2017). Students Performance Prediction Using KNN and Naïve Bayesian.

Meier, Y., Xu, A., & Van, M. (2016). Predicting grades. IEEE Transactions.

Miranda, L., & Martin, A. (2013). Analysis of student performnce using genetic Algorithum,. *Computer Ccience Journal*.

Mubarak, A., & Umar, L. (2019). *Student Academic Performance Prediction using Artificial Neural Networks: A Case Study.* China: International Journal of Computer Applications.

Muhammad, A., & Aijaz Ahmed, G. (2007). CORRELATION BETWEEN MARKS OBTAINED BY STUDENTS. *Turkish Online Journal of Distance Education-TOJDE April 2007 ISSN 1302–6488, Volume: 8 Number: 2 Article: 10*.

Murata, A. (1994). Network information criterion - determining the number of hidden units for an artificial neural network model. IEEE Transactions on Neural Networks 5 (6), pages 865-871.

Naqui, H. S. (2006). Bangledesh E Journals . *Factors affecting student performance*.

Nikolopoulos, V. (2009). Dropout prediction in e-learning courses through the combination of machine learning techniques.

Nnodium, J. (1992). Multiple choice testingin anatomy. *PubMED*, J Med Educ. 1992;26:301–09.

Osmanbegović, E., & Suljić, M. (2012). *Data Mining Approach for Predicting students performances.* J. Econ. Bus., vol. X, no. 1, pp. 3–12,.

Pandey, U., & Pal, S. (2012). A prediction of performer using classification. *International jpurnal of computer science*.

Patel, C. (2017). STUDENT FINAL GRADE PREDICTION. *Indian Journal of Computer Science and Engineering (IJCSE)*.

Ramaswami, M., & Rasthnasbapathy, M. (2010). A CHAID Based Performance Prediction Model in Educatioal Data mining. *International Journal of computer Science*.

Ranjita, S., Sushobhan, P., Ananya, M., Nina, D., & Patralekha, R. (2014). A Study on the Performance of Medical Students in Internal Assessment and its Correlates to Final Examinations of 2nd MBBS Pharmacology Curriculum in a Medical College of Eastern India. *The new examination system reforms- A must. Studies High Educ. 2002;27:221–31.*

Riedmiller, M., & Braun, H. (1993). A direct adaptive method for faster backpropagation learning: The RPROP algorithm. Proceedings of the IEEE International Conference on Neural Networks (ICNN), pages 586-591. San Francisco.

Romero, C., & Ventura, S. (2010). *Educational Data Mininig: A reviwe of the state of the art.* IEEE Trans. Syst. Man. Cybern., vol. 40, no. X, pp. 601–619.

Rushika, G., Juilee, K., Pooja, M., Sachee, N., & Priya, R. (2018). Prediction of Crop Yield using Machine Learning. *International Research Journal of Engineering and Technology (IRJET)*.

Saleem, S., Kathiry, N., Al-Osimi, S., & Badr, G. (2015). Mining Educational Data to Predict Students’ Academic Performance’. *, Mach. Learn. Data Min. Pattern Recognition, Mldm 2015, vol. 9166, no. 6, pp. 403–414,*.

Shahiri, A., Husain, W., & Rashid, N. (2015). A Review on Predicting Student’s Performance Using Data Mining Techniques. *Procedia Comput. Sci., vol. 72, pp. 414–422,*.

Shoon Lei, W., Zaw Zaw, H., Faridah, Y., & Ibrahim, N. (2014). CANCER RECURRENCE PREDICTION USING. *International Journal of Computational Science and Information Technology (IJCSITY) Vol.2, No.2, May 2014*.

Sidath, A. (2018). *Machine Learning Classifiers.* https://towardsdatascience.com/machine-learning-classifiers-a5cc4e1b0623.

Silva, F. (2018). *Importance performance analysis as tool in evaluating higher edcation*. Retrieved from https://www.researchgate.net/publication/277035463.

Sorour, S., Mine, K., & Hirokawa, S. (2014). *Predicting students’ grades based on free style comments data by artificial neural network. In 2014 IEEE Frontiers in Education Conference (FIE) Proceedings, page 19,.*

Tair, M. M., & Alaa, M. E.-H. (2012). Mining Educational Data to Improve Students’ Performance: A Case Study. . *International Journal of Information and Communication Technology Research, Vol.2 No. 2*.

Verdes, P., Granitto, M., Navone, D., & Ceccatto, H. (2010). FROST PREDICTION WITH MACHINE LEARNING TECHNIQUES. *Instituto de Física Rosario (CONICET-UNR) Argentina*.

Wongkhamdi, T., & Seresangtakul, P. (2010). A Comparison of Classical Discriminant Analysis and Artificial Neural Networks in Predicting Student Graduation Outcomes. *Proc. Second Int. Conf. Knowl. Smart Technol., pp. 29–34,*.

Woodford, C. (2017). *Neural networks*. Retrieved from https://www.explainthatstuff.com/introduction-to-neural-networks.html

Yahya, F., Mohamed, A.-A., & Abdelfatah, A. (2011). Procedures for Evaluating the Teaching Process in computer Discpline.

Yang, T., & Brinton, C. (2017). *Behavior-Based Grade Prediction for MOOCs Via Time Series Neural Networks.* IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, VOL. 11, NO. 5, AUGUST.

Yupei, Z., Yue, Y., Huan, D., Jiaqi, C., & Xuequn, S. (2020). *Graphs Regularized Robust Matrix Factorization and Its Application on Student Grade Prediction.* School of Computer Science, Northwestern Polytechnical University, Xi’an 710129, Shaanxi, China;.

Zacharis, N. (2016). Predicting Student Academic Performance in Blended Learning Using Artificial Neural Networks. *Int. J. Artif. Intell. Appl., vol. 7, no. 5, pp. 17–29,*.

## **Annex1: Research Questionnaire**

Respondent Name: Symbol No:

Marks of Student (in Percent)

SLC Total Percent Select High School Education Govt. Private:

+2 Total Percent Physics: Math: Chemistry: Bio:

Please select your parent highest education label (Please Tick)

a) SLC b) +2 c) Bachelor d) Master e) Ph. D

Please select your other family member’s highest education label (if only have)

a) SLC b) +2 c) Bachelor d) Master e) Ph. D

THANK YOU

-----------------------------------------------------------------------------------------

Research Questionnaire

Respondent Name: Symbol No:

Marks of Student (in Percent)

SLC Total Percent Select High School Education Govt. Private:

+2 Total Percent Physics: Math: Chemistry: Bio:

Please select your parent highest education label (Please Tick)

a) SLC b) +2 c) Bachelor d) Master e) Ph. D

Please select your other family member’s highest education label (if only have)

a) SLC b) +2 c) Bachelor d) Master e) Ph. D

THANK YOU

## **Annex 3: Summary Statists of Sample Data**

GPA SLC School Plus Physics

Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000

1st Qu.:0.0000 1st Qu.:0.4708 1st Qu.:0.0000 1st Qu.:0.4722 1st Qu.:0.2973

Median :0.2487 Median :0.6771 Median :0.0000 Median :0.6667 Median :0.5405

Mean :0.4073 Mean :0.6088 Mean :0.4138 Mean :0.6031 Mean :0.5165

3rd Qu.:0.8244 3rd Qu.:0.7694 3rd Qu.:1.0000 3rd Qu.:0.7778 3rd Qu.:0.6757

Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000

Math Chemistry Bio Parent1 Parent2

Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000

1st Qu.:0.4423 1st Qu.:0.3882 1st Qu.:0.6522 1st Qu.:0.0000 1st Qu.:0.6667

Median :0.5962 Median :0.5263 Median :0.7609 Median :0.6667 Median :0.6667

Mean :0.6002 Mean :0.5190 Mean :0.7064 Mean :0.6201 Mean :0.7753

3rd Qu.:0.7692 3rd Qu.:0.7171 3rd Qu.:0.8152 3rd Qu.:1.0000 3rd Qu.:1.0000

Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000

X1 X2 X3 X4 X5

Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000

1st Qu.:0.7350 1st Qu.:0.7600 1st Qu.:0.4583 1st Qu.:0.7200 1st Qu.:0.7200

Median :0.8400 Median :0.8200 Median :0.6250 Median :0.8000 Median :0.8000

Mean :0.8091 Mean :0.8068 Mean :0.6010 Mean :0.7937 Mean :0.7647

3rd Qu.:0.9000 3rd Qu.:0.9000 3rd Qu.:0.7500 3rd Qu.:0.8800 3rd Qu.:0.8800

Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000

X6 X111 X222 X333 X444

Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000

1st Qu.:0.6400 1st Qu.:0.0000 1st Qu.:0.6111 1st Qu.:0.0000 1st Qu.:0.0000

Median :0.8000 Median :0.7778 Median :0.7222 Median :0.7222 Median :0.7778

Mean :0.7339 Mean :0.6091 Mean :0.6275 Mean :0.5995 Mean :0.5656

3rd Qu.:0.8800 3rd Qu.:0.8889 3rd Qu.:0.8889 3rd Qu.:0.8889 3rd Qu.:0.8889

Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000

## **Annex: Photographs**