

EMERGING TRENDS IN COMPUTATION & ARTIFICIAL INTELLIGENCE

Editors

Dr. K. Santhosh Kumar

Dr. H. Sivalingan

Mrs. L. Sankara Maheswari



Emerging Trends in Computation & Artificial Intelligence

First Edition

Editor

Dr. K. Santhosh Kumar

Head & Assistant Professor,
PG & Research Department of Computer Science,
Providence College for Women (Autonomous),
Coonoor, The Nilgiris, Tamil Nadu, India.

Co-Editors

Dr. H. Sivalingan

Head & Assistant Professor,
Department of Data Science,
Providence College for Women (Autonomous),
Coonoor, The Nilgiris, Tamil Nadu, India.

Mrs. L. Sankara Maheswari

Head & Associate Professor,
Department of Information Technology,
Sri.G.V.G Visalakshi College for Women (Autonomous),
Udumalpet, TamilNadu, India - 642 128.

Published by

CiiT Publications

#156, 3rd Floor, Kalidas Road, Ramnagar,
Coimbatore – 641009, Tamil Nadu, India.

Phone: 0422 - 4377821

www.ciitresearch.org

All Rights Reserved.

Original English Language Edition 2025 © Copyright by CiiT Publications, a unit of **Coimbatore Institute of Information Technology**.

This book may not be duplicated in anyway without the express written consent of the publisher, except in the form of brief excerpts or quotations for the purpose of review. The information contained herein is for the personal use of the reader and may not be incorporated in any commercial programs, other books, database, or any kind of software without written consent of the publisher. Making copies of this book or any portion thereof for any purpose other than your own is a violation of copyright laws.

This edition has been published by CiiT Publications, **Coimbatore**.

Limits of Liability/Disclaimer of Warranty: The author and publisher have used their effort in preparing this “Emerging Trends in Computation & Artificial Intelligence” book and author makes no representation or warranties with respect to accuracy or completeness of the contents of this book, and specifically disclaims any implied warranties of merchantability or fitness for any particular purpose. There are no warranties which extend beyond the descriptions contained in this paragraph. No warranty may be created or extended by sales representatives or written sales materials. Neither CiiT nor author shall be liable for any loss of profit or any other commercial damage, including but limited to special, incidental, consequential, or other damages.

Trademarks: All brand names and product names used in this book are trademarks, registered trademarks, or trade names of their respective holders.

ISBN 978-93-6126-962-2

This book is printed in 70 gsm papers.

Printed in India by CiiT Publications.

MRP Rs. 700/-

CiiT Publications

#156, 3rd Floor, Kalidas Road, Ramnagar,

Coimbatore – 641009, Tamil Nadu, India.

Phone: 0422 - 4377821

www.ciiiresearch.org

11	AI-DRIVEN ECONOMIC GROWTH AND WORKFORCE TRANSFORMATION IN INDIA Dr. S. Sathiyapriya	63
12	REVIEW: APPLICATIONS OF BLOCK CHAIN TECHNOLOGIES WITH ASSISTIVE AI FOR SUSTAINABLE AGRICULTURE Mrs. P. Gangalakshmi	67
13	SOFT COMPUTING INNOVATIONS AND BEST PRACTICES FOR APPLICATIONS IN ADVANCED COMPUTING Dr. Hema Deenadayalan	79
14	NAVIGATING THE EVOLVING CYBER SECURITY LANDSCAPE Dr. R. Jayaprakash	90
15	LEVERAGING DECISION STUMP CLASSIFICATION FOR DATA-DRIVEN STUDENT PLACEMENT OUTCOME PREDICTIONS Dr. B. Kalaiselvi	95
16	EMBODIED NEUROMORPHIC INTELLIGENCE S. Lavanya	102
17	ADAPTIVE BILATERAL REGION GROWING CORRELATION BASED FEATURE SELECTION WITH INCEPTION V3 MAMMOGRAM CLASSIFICATION Mrs. T. Leena Prema Kumari, Dr. K. Perumal	107
18	EMERGING TRENDS IN COMPUTATION AND ARTIFICIAL INTELLIGENCE FOR CLOUD COMPUTING OPTIMIZATION Maria Sofia R. B, Dr. R. Parameswari	117
19	DEEP LEARNING AND GENETIC ALGORITHMS FOR AGRICULTURAL ASSOCIATION RULE OPTIMIZATION Mrs. N. Amirtha Gowri	127
20	INTELLIGENT SYSTEMS FOR REAL-TIME DECISION-MAKING IN ENGINEERING Mrs. R. Nirmala	131
21	A DETAIL STUDY ON INTRUSION DETECTION SYSTEM Dr. P. Sudha	137
22	THE MATHEMATICS BEHIND MACHINE LEARNING: FROM THEORY TO IMPLEMENTATION Priyadharsini	143
23	A REVIEW: NASCENT APPLICATIONS OF GENERATIVE AI IN HEALTHCARE Dr. V. Punithavathi	146

CHAPTER -15
LEVERAGING DECISION STUMP CLASSIFICATION FOR DATA-DRIVEN STUDENT
PLACEMENT OUTCOME PREDICTIONS

Dr. B. Kalaiselvi

Assistant Professor, Department of Information Technology,
Nallamuthu Gounder Mahalingam College, Pollachi.

ABSTRACT

Student placement prediction plays a crucial role in academic institutions to enhance career opportunities and improve employment rates. Providing better placement for the students is the crucial goal of the educational institutions. Assessment of the student technical and numerical skills is very important for the institutions in order to make them prepared in the lagging venture. Classification and prediction algorithm comes into the rescue [3]. Traditional predictive models often require complex computations, making them less efficient for real-time decision-making. In this study, we propose a **Decision Stump Algorithm** for student placement prediction, leveraging a single-level decision tree to classify students based on academic performance, extracurricular activities, and relevant skills. The Decision Stump algorithm offers a lightweight yet effective solution by focusing on a single attribute split criterion, reducing computational complexity while maintaining high prediction accuracy. This model is trained on historical student placement data, and its performance is evaluated using key metrics such as accuracy, precision, and recall. Experimental results indicate that the proposed model of decision stump based placement prediction provides competitive results with minimal processing time, making it a feasible option for institutions looking for a simple yet efficient placement prediction model. Future work aims to enhance the model by integrating ensemble learning techniques to improve robustness and generalization.

KEYWORDS

Placement, Classification, Prediction, Predictive accuracy, Decision Stump.

1. INTRODUCTION

Student placement prediction has become an essential aspect of academic institutions, helping students and educators make informed career-related decisions. Predicting whether a student will secure a job offer based on academic performance, technical skills, and extracurricular activities can significantly improve training programs, resource allocation, and overall placement success rates. Various machine learning techniques have been employed for this purpose, ranging from simple statistical models to complex deep learning algorithms [4]. However, many advanced models require high computational power and large datasets, making them difficult to implement in real-time decision-making scenarios. Tree based machine learning algorithms offer best solutions [5]. A **Decision Stump** is a simplified version of a decision tree that makes decisions based on a single attribute, offering a balance between interpretability and computational efficiency. This approach is particularly useful when quick and straightforward predictions are required. By analyzing historical placement data, our model determines the most influential factor affecting student placements and classifies students into placed or non-placed categories based on a single decision boundary. The key advantages of using a Decision Stump include reduced processing time, ease of implementation, and transparency in decision-making. Despite its simplicity, it can serve as a baseline model for placement prediction and can be further enhanced by integrating ensemble methods such as **Boosting** to improve accuracy and generalization.

2. RELATED WORK

Aman[1] authors developed an LMT (Logistic Model Tree) prediction model using real student data from

the University of Peshawar, considering academic, demographic, and socioeconomic factors. They compared its performance with J48 and Random Forest models. The results showed that the LMT model achieved 83.1% accuracy, making it an effective method for predicting student outcomes.

B. Kalaiselvi [2], the authors developed a hybrid model to analyze student placement data using the AdaBoost classifier along with Decision Stump, NB Tree, and Random Forest classifiers. They found that combining AdaBoost with Random Forest improved accuracy to 87.09%, making it more effective than Decision Stump and NB Tree. In comparison, Random Forest alone achieved only 79.85% accuracy, showing that AdaBoost helps enhance its performance.

B. Kalaiselvi [3] the study used a machine learning method called J48, which builds decision trees, to predict how likely students are to get placed in a job after graduating. They tested this method on a dataset of students who had already graduated and found that it was able to make accurate predictions 87% of the time. This means that the model correctly predicted whether or not students would be placed in jobs in most of the cases it was tested on.

B. Kalaiselvi [4], the authors used the J48 classifier to categorize student academic data and predict their performance during the COVID-19 pandemic. The model worked with real-time student data to predict end-of-semester test results, achieving an impressive 96.42% accuracy.

IT Jose [5] this paper, the authors explored predicting student placement using different machine learning models, including Support Vector Machine (SVM), Logistic Regression (LR), K-Nearest Neighbors (KNN), and Random Forest. They compared the accuracy and performance of each model to see which one worked best. The models used various factors, such as students' scores in verbal skills, technical programming, reasoning, numeric aptitude, and academic CGPA, as well as data on any backlogs and certifications.

L H Son [6], the authors developed a **machine learning model** to predict student performance using data from **VNU University of Science** and three educational datasets from **KDD**. They used **MANFIS with RS**, which achieved higher accuracy compared to earlier **fuzzy and tree-based models**, as shown by experimental results.

Shreyas [7], the authors created a **machine learning model** to predict whether current students will be placed, using data from previous students. They used **Naive Bayes** and **KNN** models, with the training data coming from students who had already graduated, including their placement status.

Sultana [8] the authors used a type of deep learning model called Convolutional Neural Network (CNN) to predict how well students would perform based on past data. The CNN model was able to make predictions with 97.5% accuracy, which was better than other models they tested. This means the CNN model was very good at predicting student performance.

3. PROPOSED METHODOLOGY

The steps involved in this prediction process includes data pre-processing, feature selection, proposed methodology, implementation, model training and model evaluation which is represented in fig.1.

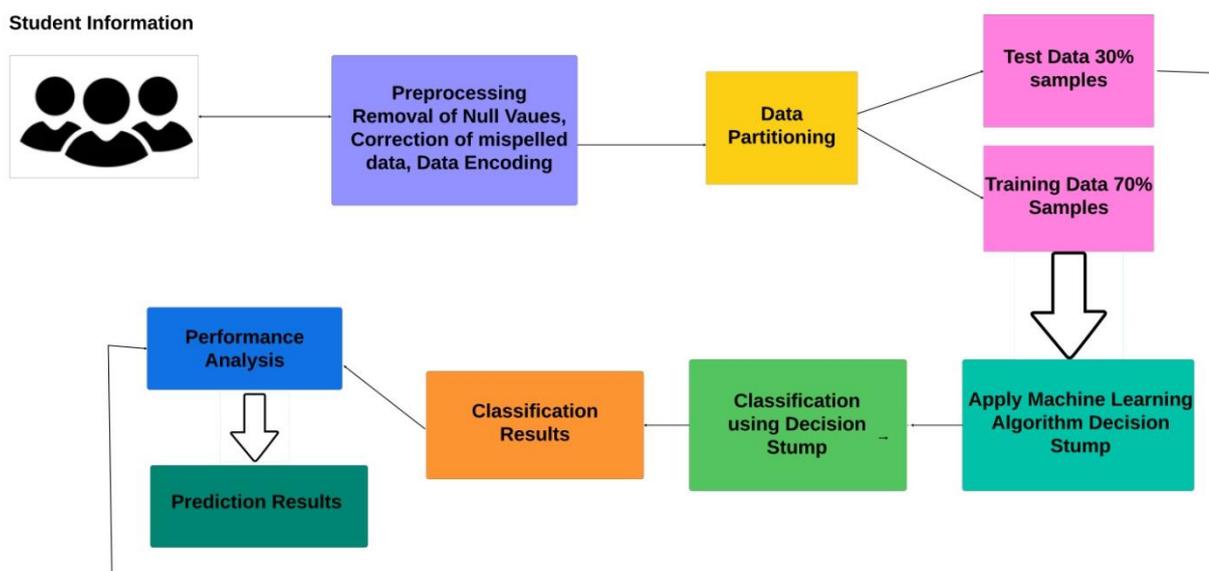


Fig. 1 Proposed Framework

3.1 Dataset

The dataset consists of placement records of students belonging to Nallamuthu Gounder Mahalingam College, Pollachi. Table 1 show the details of attributes such as academic percentage, certifications obtained, course of study, placement training class attended, HSC marks, medium of study, attendance percentage, etc. [7].

Table 1: Attributes of the Dataset

VARIABLES	DETAIL	POSSIBLE VALUES
Sl.No	Serial Number	Numerical Sequence
Register No	Register Number	Alphanumeric Sequence
StName	Student Name	Name of the Student
Course	Course Name	UG Course Name
Gender	Gender	{Male, Female}
SSLC Mark	Percentage of Marks in SSLC	{35% to 100%}

HSCMark	Percentage of Marks in HSC	{35% to 100%}
UGMarks	Percentage of UG	{35% to 100%}
Placed	Placed in Campus Interview	{Placed,Not Placed}

3.2 Pre-processing

This phase is a crucial part of preparing data for machine learning models. Here's a breakdown of the steps involved:

1. Removing Null or Missing Values: If some data points are missing in a column, they can negatively impact model performance. In this step, columns that have too many missing values might be removed, or the missing values could be replaced with a placeholder, such as the mean, median, or mode of the column, depending on the type of data.
2. Correcting Misspelled Data: Inconsistent or misspelled data can lead to incorrect interpretations. For instance, if students' names, subjects, or other important terms are misspelled in different rows, this can cause issues in analysis. This step involves fixing these errors to ensure consistency.

3. Data Encoding: Machine learning models require numerical input, so categorical data (like gender, city, or department) needs to be converted into numbers. This can be done through various encoding techniques such as:

- Label Encoding: Assigning a unique number to each category.
- One-Hot Encoding: Creating binary columns for each category, where each column represents the presence (1) or absence (0) of a particular category.

3.3 Feature Selection

Analysis of dataset in order to identify the relevant features for placement prediction has to be done. CFS Subset Evaluator is used for attribute selection [6]. The **CFS (Correlation-based Feature Selection)** Subset Evaluator in Weka is used for selecting the most relevant attributes (or features) from a dataset [8]. It evaluates subsets of attributes based on their predictive ability by taking into account both the individual predictive power of each feature and the degree of redundancy among the selected features.

Working of CFS Subset Evaluator

- Evaluating Individual Features: Each feature is considered based on its correlation with the class (target variable). Features that are strongly correlated with the class are prioritized.
- Redundancy Check: Features are also evaluated based on how much redundancy they have with other selected features. Features that are highly correlated with one another (and therefore provide similar information) are considered redundant and are less likely to be selected together.
- Subset Evaluation: The algorithm searches for a subset of features that maximizes the correlation with the class while minimizing the inter-correlation among the features in the subset.

Classification and Train using Decision Stump

Decision stump is a simple machine learning model

that is a single-level decision tree, typically used in classification problems [2]. It selects one feature and a threshold to make a decision. In the context of **student placement prediction**, we can use a Decision Stump to classify students as **placed** or **not placed** based on features like Gender, Area of Living, Community, First Graduate, Higher Secondary Percentage, SSLC percentage, Medium of study, type of school, placement training attended and certifications if any acquired [1]. Information Gain is calculated for each attribute and highest IG Value attribute is assigned as a root node. Based on the feature it fixes the threshold value for the selected feature by applying different threshold and calculates the impurity, then divides the data into two subsets based on the feature value is below and above the threshold. Then it assigns the class label to each data point based on the subset they belong which leads to prediction of class label.

Testing and Performance Measures

The model was trained using 73 instances as training data out of 110 and balance 37 instances as test data. It predicts the placement possibilities. The performance measures such as classification accuracy, precision values, True Positive, False positive values which is present in the confusion matrix are analyzed.

4. IMPLEMENTATION AND RESULTS

The decision stump algorithm finds the best split by analyzing the student data, and fixed the threshold value. Based on the threshold value the model predicts the student's placement prediction. The following process are performed in the implementation phase using Decision Stump classifier in Weka;

- Students data classification is done based on class attribute place (Placed , Not placed),
- Split the students data set into training(70% of instance) and test data set(30% of instances),
- Train the model and test the placement prediction using training and test data sets.

Visualized representation of the dataset is shown in fig. 4. The collected data is passed into decision stump classifier with 10 x fold validation the results are obtained. The model correctly classifies 100

instances out of 110 and obtained 90.90% of accuracy with 0.979 as true positive rate which is represented in fig. 2. The margin curve for the classification is represented in fig. 3.

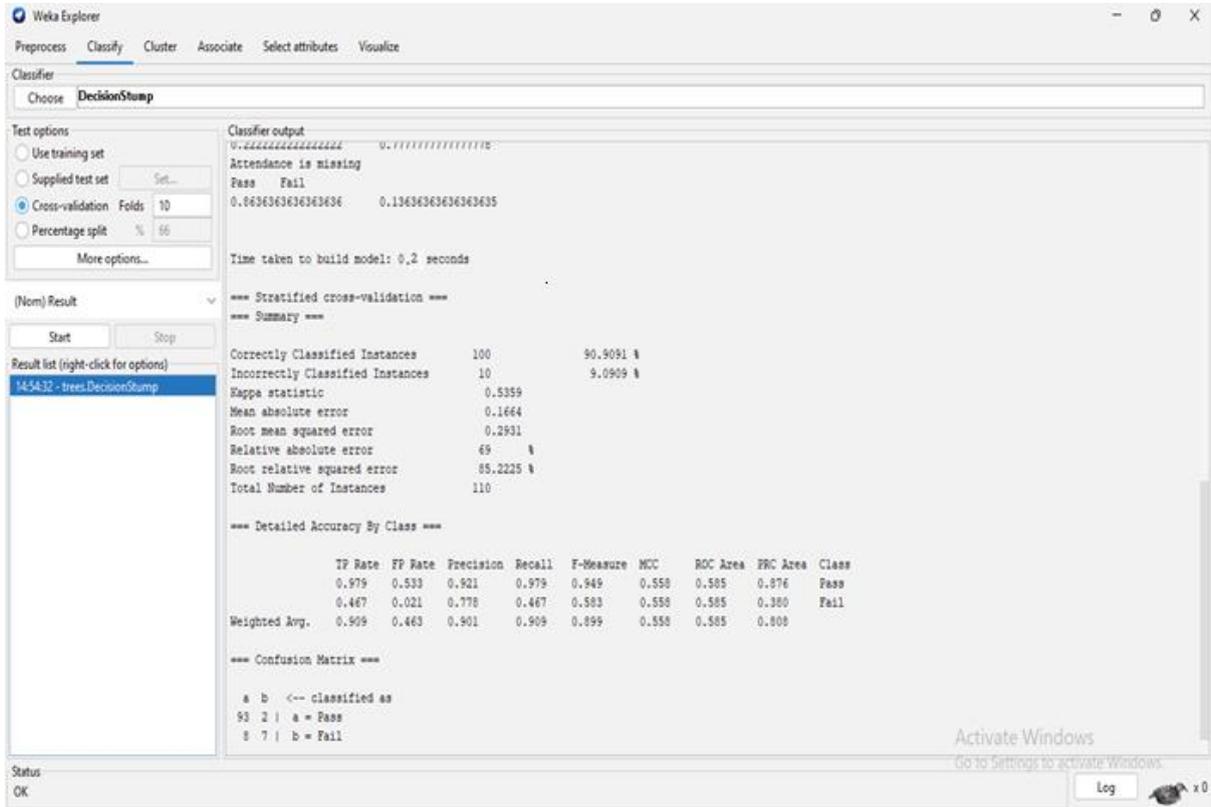


Fig. 2 Classification Results

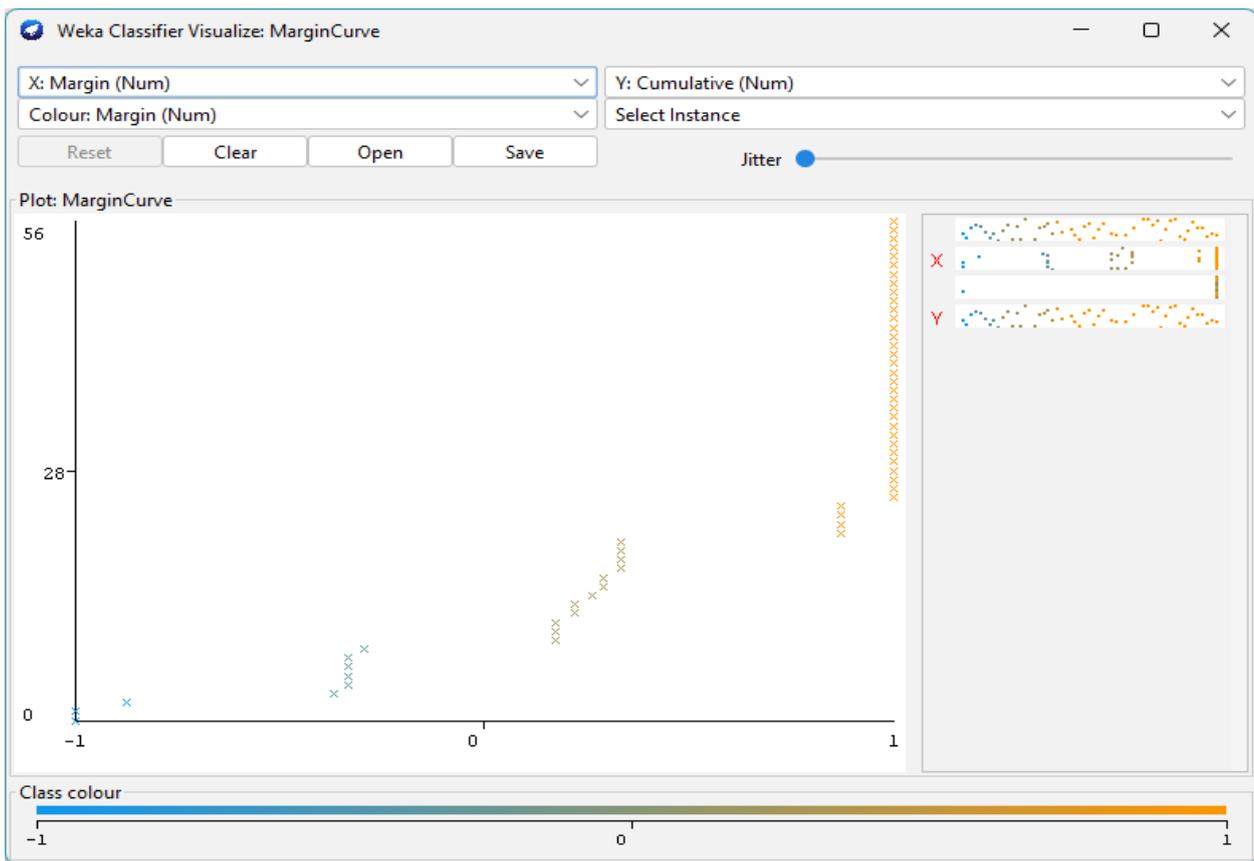


Fig. 3 Margin curve for Dataset

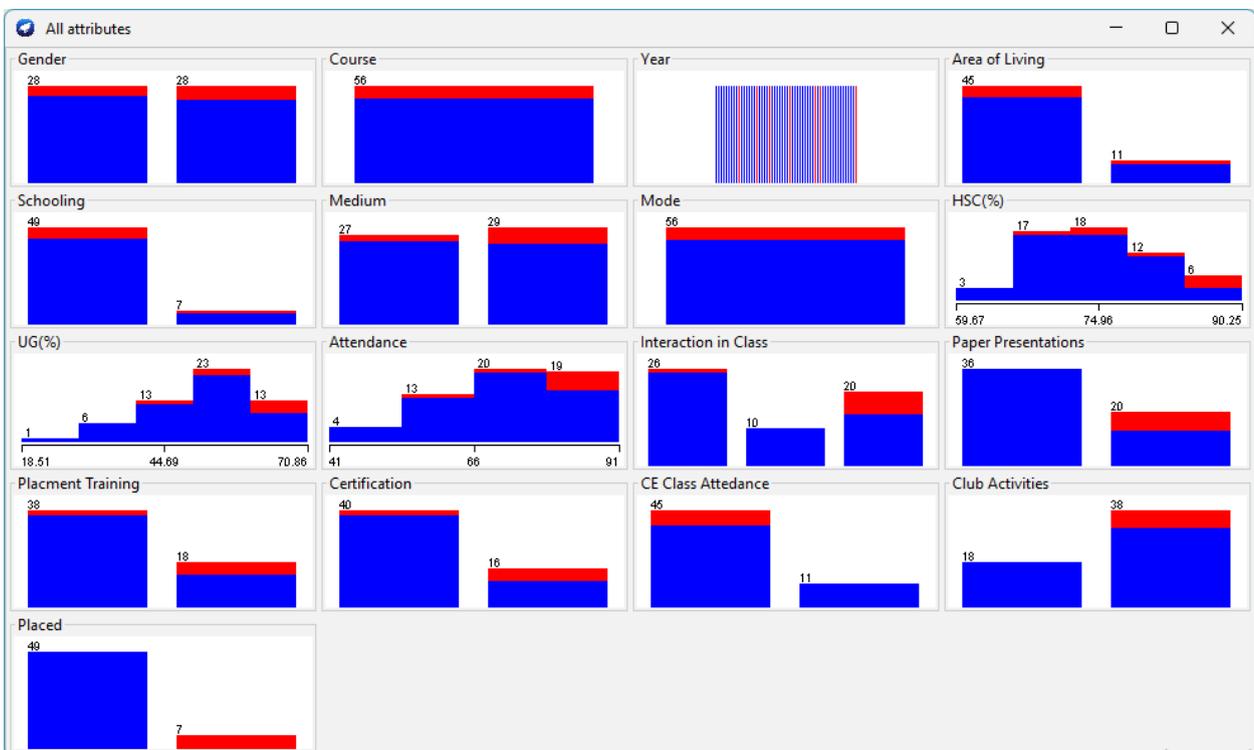


Fig. 4 Visualized Representation of the Dataset

5. DISCUSSION

Out of 110 data the system predicted 100 instances correctly in the classification process carried out by decision stump classifier based on the placement status. Based on thresh hold value the algorithm constructs the decision tree. The tree predicts the pre-processed student data based on the constructed model. It produces 90.90% of accuracy with 0.539 as kappa statistics value 0.1664 as mean absolute error 0.2931 as root mean squared error and its build time is 0.2 seconds. The model can be enhanced to improve its accuracy and minimize the errors by means of ensemble techniques.

6. CONCLUSION

This model achieves 90.90% of classification accuracy under the decision stump algorithm for predicting the placement status of students which will be useful for the academic institutions to improve the skill set for the students to achieve better placements. The tree based algorithm suits best for student placement prediction. Since this machine learning algorithm looks more simple and interpretable. It takes lesser training time. More number of decision trees can be bagged as an ensemble method to contribute themselves into a better performance.

REFERENCES

[1] Aman, Fazal and Ali., 2019. "A Predictive Model For Predicting Students Academic Performance", 10th Intl. Conf. on Infor., Intelligence on Info., Intelligence, Sys. andAppli. (IISA), 2019.

[2] B.Kalaiselvi and Dr.S.Geetha, "Ensemble Machine Learning AdaBoost with NBTree Model for Placement Data Analysis", International Conference on Intelligent Technologies (CONIT), IEEE-2022.

[4] B.Kalaiselvi and Dr.S.Geetha, "Academic performance prediction for first year students in covid pandemic period using data mining technique", Journal of the Maharaja Sayajirao University of Baroda, Volume-54, No.2 (XVI) 2020.

[6] B.Kalaiselvi and Dr.S.Geetha, "Analysis of placement performance prediction on students data using machine learning algorithm", Journal of the Maharaja Sayajirao University of Baroda, Volume-56, No.1 (IX) 2022

[8] I.T Jose, D Raju, , JA Aniyar, J James, and MT Vadakkal, "Placement Prediction using Various Machine Learning Models and their Efficiency Comparison", IJISRT – May 2020.

[3] LH Son and H Fujita, "Neural-fuzzy with representative sets for prediction of student performance", Applied Intelligence, Springer, 2018.

[5] Shreyas H, Aksha P, and Suraksha. "Student Placement Prediction using ML", Intl. Research Journal of Engineering and Technology (IRJIET) Volume: 06 Issue: 04 April 2019

[7] Sultana, Jabeen, MU Rani, and Farquad. "Student's performance prediction using deep learning and data mining methods." Int. Journal. Recent Technol. Eng 8.1S4 (2019): 1018- 1021.