PREDICTING EFFECTS ON STUDENT ENROLLMENT THROUGH CLASSIFICATION TECHNIQUES

Poonam Kumari, Dr. Rajender Singh Sodhi

Research Scholar,

Department of Computer Science, Om Sterling Global University Hisar

Associate Professor & Head School of Engineering & Technology,

Om Sterling Global University Hisar

poonamcse191@osgu.ac.in

Abstract:

In the field of education, student enrollment prediction has been a crucial task for universities and colleges to make informed decisions. This study aims to investigate the effectiveness of classification algorithms in predicting student enrollment. A dataset containing demographic information and academic records of past students was used to train and test six different classification algorithms implemented in WEKA, a popular data mining tool. The algorithms included in this study were J48 (C4.5), Naive Bayes, Random Forest, IBk (k-nearest neighbor), JRIP and Logistic Regression. The performance of each algorithm was evaluated using several performance metrics, including accuracy, precision, recall, F1-score, and ROC AUC. The results showed that Random Forest outperformed the other algorithms in terms of accuracy, precision, recall, F1-score, and ROC AUC. This study demonstrates the effectiveness of using classification algorithms in predicting student enrollment and highlights the importance of considering multiple performance metrics in evaluating the performance of these algorithms.

I. Introduction:

Student enrollment prediction has become increasingly important in the education sector, as it enables universities and colleges to make informed decisions related to resources allocation, student recruitment, and future planning. The prediction of student enrollment is a complex task that involves various factors, including demographic information, academic performance, and socio-economic status. In recent years, the use of machine learning techniques has become increasingly popular in solving complex prediction problems, including student enrollment prediction. When data is oversized and there is little information about the data accessible, data mining techniques are used to extract hidden facts from large datasets. Data mining techniques are also cutting edge when the information they gather helps people make smart decisions. Making decisions actively helps to steer pupils in the appropriate direction in the sphere of education. Data mining is used to look at linked educational data that is utilized to answer educational queries. Prediction of student enrollment has recently received attention and is essential to keeping the student on the proper path. Data mining techniques guarantee that the data is viewable and provide a better means of analysis. A special application of data mining techniques to educational data is called educational data mining. Its goal is to investigate the data available to answer research inquiries for educational purposes.

Knowledge can be extracted utilizing the aforementioned statistical techniques in the form of classification and association rules, which are highly recommended prediction techniques. Clustering can also be used to organize educationally related things. The observers are helped by these tactics almost from every aspect. Among the predictions in educational data are the likelihood that students will enroll in the courses that are best for them, their grades in various assessment categories, and the identification of students who require additional support in particular examination categories. The main goal of this study is to put data mining techniques to work on educational-related data and predict student enrollment status after a year in college, such as which students are likely to be dropped or promoted. A mechanism is also used to alert the administration about students who need more assistance in a particular area. The accuracy of several categorization algorithms for predicting student performance is also investigated in this study. The main goal of this work is to create a predictive model to assist students who perform poorly academically in higher education, including:

a) Extraction of predictable properties from the data source.

b) Recognition of many characteristics that could influence a student's approach to learning.

c) Using various classification methods, building a prediction model based on predicable variables that have been chosen.

d) Inform the administration that some students' enrollment status is in risk.

II. **Literature Review**

Although the study of data mining in education is still in its infancy, a lot of work has been done in this area. That is due to its potential for an educational institution.

Romero and Ventura examined all of the work produced between 1995 and 2005 for their study on educational data. It was evident that educational data mining has been a hot topic for research, and that it is necessary in a way that is not apparent in other fields.

Using decision tree classification, Baradwaj and Pal devised a model to forecast student performance. The study's primary goal was to identify the variables that best represented students' performance on final term exams. Records of students from the student database, including grades on assignments, results of tests and seminars, and attendance. The suggested methodology also helps to identify dropouts and students who need to be taken into consideration in advance.

Z. J. Kovacic provided a methodology to determine the amount to which enrollment data contributes to the prediction of student performance, and a case study was developed. The student enrollment data was taken from the open polytechnic's student information system in New Zealand, and the CHAID and CART algorithms were then used. It was found that CHAID and CART's accuracy levels were 59.4 and 60.5, respectively.

An analysis of student learning behaviour, result evaluation, and implementation of these techniques to enhance students' learning strategies are the goals of Hongjie Sun's model, which is based on data mining techniques.

The CHAID prediction model was developed by M. Ramaswami and R. Bhaskaran to investigate the relationships between the variables used to forecast the performance of high school pupils. The creation of models relies on predictable features such as learning style, grades earned in secondary education, school location, and manner of life.

Using data from two secondary school students, Cortez and Silva created a prediction model to forecast failure rates in the two major disciplines of mathematics and portuguese. Numerous mining methods, including Support Vector Machine, Decision Tree, Random Forest, and Neural Networks, were used. It was revealed that the accuracy of the decision tree and neural network techniques for the two class dataset was 93% and 91%, respectively. Additionally, it was claimed that the accuracy of the Decision Tree and Neural Network methods for the four class dataset was 72%.

Al Radaideh developed a decision-tree-based model to anticipate students' final grades while teaching C++ at the University of Yarmouk. Three classification techniques—ID3, C4.5, and Naive Bayes—were used. The outcomes of the used model demonstrated that Decision Tree had better prediction than other methods.

In order to predict the performance of enrolled students, Shannaq et al. built a model based on classification. They did this by analyzing the academic records of the students to identify the main elements that can undermine their commitment. The retrieved categorization results using a Decision Tree classification approach. The university administration can use it to create certain crucial resources.

In their model, Bharadwaj and Pal used 300 students from five randomly chosen degree colleges to predict students' performance. The results of applying the Bayesian classification technique to 17 attributes showed that factors like living location, family annual income, teaching style, other habits of students, student grade, family status of students, and mother's education were strongly correlated with academic performance.

By choosing 600 students from randomly chosen colleges, Pandey and Pal proposed to calculate student performance using a prediction model. Applying Bayes classification to prior knowledge and education the purpose of the model was to anticipate whether newly enrolled students would perform well or not.

For the examination of student enrolment data, Fadzilah Siraj and Mansour Ali Abdoulha suggested a methodology based on mining techniques. For prediction, they used three data mining approaches in their study: decision trees, logistic regression, and neural networks. The outcomes of research can be useful to the university's decision-makers.

Galit suggested a case study to evaluate students' learning styles, predict grades, and alert those at risk prior to final exams.

The use of the k-means clustering algorithm to predict students' learning behaviour was suggested by Ayesha, Mustafa, Sattar, and Khan. The study's conclusions help students and the administration identify weak students who need additional support.

III. Methodology:

The dataset used in this study was collected from Ch. Bansilal University, Bhiwani and contained demographic information and academic records of past students. The data was divided into training (70%) and testing (30%) datasets. Six different classification algorithms were implemented in WEKA, a popular data mining tool, to predict student enrollment based on the training dataset. The algorithms included J48 (C4.5), Naive Bayes, Random Forest, IBk (k-nearest neighbor), JRIP and Logistic Regression.

IV. Data Mining Process

Data mining is the process of discovering patterns, correlations, and insights in large datasets through techniques such as statistical analysis and machine learning. The goal of data mining is to extract information from data and transform it into an understandable structure for further use. This process can be applied to various domains such as customer behavior, market trends, and scientific discovery, among others. Data collection, preprocessing, transformation, and prediction model creation are the first steps in the data mining process. Many parameters required to be measured in order to predict student enrollment. All environmental factors that are required for an accurate forecast of student performance must be included in the prediction model. Data on a student's prior academic performance, main subjects studied aptitude for a certain assessment type, preferred subjects of students, etc. Outlier detection and removal are necessary as a crucial step in preprocessing before data transformation.

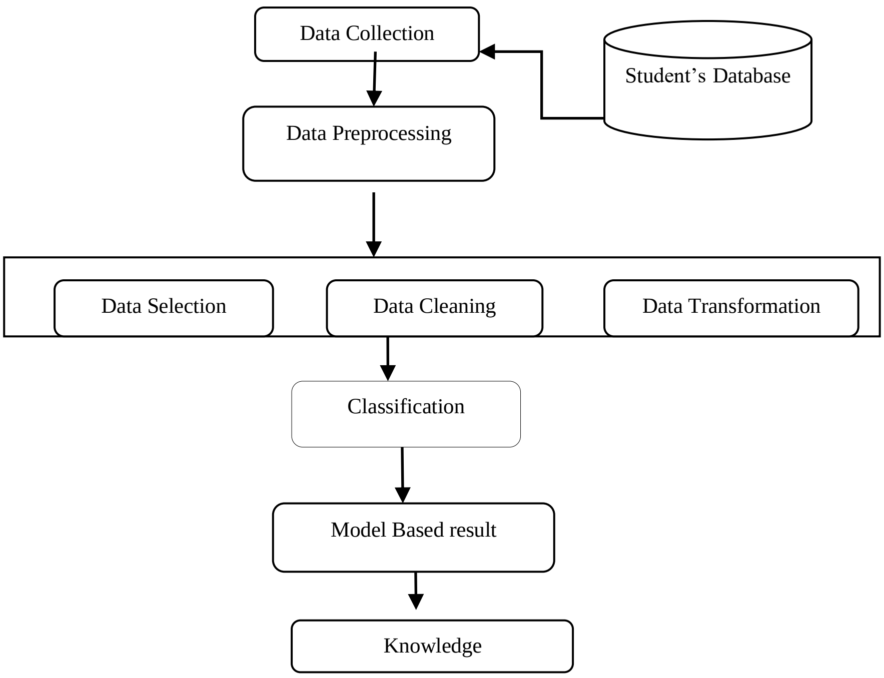


Figure:-1 Data Mining Process trough classification

A. Data Collection:-

The data required for my analysis was collected from various institutes affiliated to Ch. Bansilal University, Bhiwani.

B. Data preprocessing

Data preprocessing is a crucial step in the knowledge discovery process since high-quality decision-making depends on the accuracy of the data. Data preparation works with all such types of data to clean it and make it accessible in customized form. Raw data may have missing values, aberrant values such as outliers, and inconsistencies. Data cleaning is the process of making the data consistent, smoothing the data by removing outliers and noise, and filling in missing information. Data transformation is the process of normalizing and aggregating data, which is the next step in the construction of a data mining process.

a. Dealing with Missing values

Our student dataset revealed that some of the chosen batch had dropped out of school before the end of the first two semesters, making the class label unavailable.

So, when a class label is missing in a classification, it is preferable to ignore the data row. Any such data rows were disregarded in order to include them in our data while keeping this choice in mind.

b. Data Transformation and Selection

The student database had 20 attributes, of which 12 were chosen as outcome-driven components in this step. These variables were chosen in order to run the structure of our model. The relevance of the qualities required to be included in our predictive model is confirmed by Table 1, which displays all dataset features that could be recorded along with their descriptions.

ATTRIBUTES	DESCRIPTION	SELECTED

Name	Name of the Students	YES
Father Name	Father Name of the Student	
Gender	Male/Female	YES
Age	Age of the Student	YES
Category	General/OBC/SC/ST	YES
Address	Address of student	
Contact Number	Contact No of Students	
10^{th} %	Percentage obtained in 10^{th}	YES
12^{th}%	Percentage obtained in 12^{th}	YES
12^{th} Medium	Hindi/English	
Registered Course	Detail of Registered Course	
Family Type	Joint/Nuclear	YES
Family Income	Annual Income of Family	YES
Location of residence	Urban/Rural	YES
Distance Between Residence And Institute	Distance Between Residence And Institute	YES
Marital Status	Married/Unmarried	YES
Enrollment Status	Dropped or Promoted After two Semesters	YES

C. Descriptive Analysis on collected data

By utilizing the tools of descriptive statistics, we have distilled the fundamental characteristics of the data into a concise and comprehensible format. This allows for a more streamlined interpretation of the information presented. We have conveyed the data in a user-friendly manner through various mediums such as tables, cross-tables, and graphs. Our analysis was conducted using the powerful software, SPSS 16.0.

The Statistical Package for Social Science (SPSS) is highly popular statistical software that allows for complex data manipulation and analysis in a user-friendly and interactive way. According to the SPSS base user's guide, it is a comprehensive system capable of generating tabulated reports, charts, and plots of distributions and trends, descriptive statistics, and advanced statistical analyses using data from almost any type of file.

D. Prediction Model Implementation

Many methods, including Neural Networks, Decision Trees, Association Rule Mining, the Nearest Neighbor Approach, Clustering, and Classification, are utilized to construct prediction models. One of the most used methods for prediction is classification, particularly when greater precision is required. Class label-based classification predictions that include both successful and unsuccessful outcomes. Decision Tree algorithms such as Random Forest and J48graft were used to classify the scenarios based on their attributes. These algorithms create a decision tree model that recursively splits the data into subsets based on the most significant attributes. Random Forest builds multiple decision trees and aggregates their outputs to improve the accuracy of the model, while J48graft is a modified version of the popular J48 algorithm that uses grafting techniques to improve decision tree accuracy. Some details of proposed classification algorithms are following

i. Decision Tree approach

The decision tree approach is a popular machine learning algorithm used for classification and regression tasks. It is a tree-like model where each internal node represents a test on an attribute, each branch represents the outcome of the test, and each leaf node represents a class label or a numerical value. The decision tree algorithm works by recursively splitting the data based on the most significant attribute that maximizes the information gain or the Gini index. Information gain measures the reduction in entropy or uncertainty in the data after splitting, while the Gini index measures the impurity or the degree of classification error in the data.

ii. Bayes Approach

The Bayes approach is a probabilistic machine learning algorithm that is commonly used for classification and regression tasks. It is based on Bayes' theorem, which is a mathematical formula that describes the probability of an event occurring given prior knowledge of related events. In the context of classification, the Bayes approach involves calculating the probability of a new data point belonging to a particular class label given the observed values of its attributes. Naive Bayes is a simple and fast algorithm that assumes that the attributes are conditionally independent given the class label. This means that the probability distribution of each attribute is estimated separately and then combined to calculate the probability of the class label. Despite its simplicity, Naive Bayes has been shown to perform well in many applications, particularly in text classification.

V. Result and Discussion

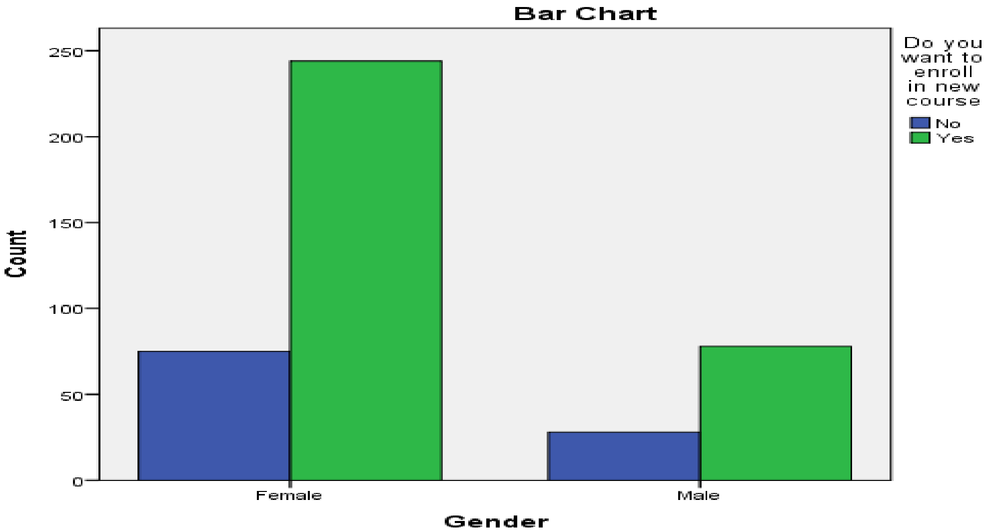
a. Descriptive Analysis

The provided table offers valuable insights into enrollment predictions based on gender and enrollment intentions. The data suggests that a higher number of females are interested in enrolling in a new course compared to males. By considering the gender distribution and analyzing the enrollment preferences, institutions can tailor their enrollment strategies to attract and accommodate the needs of different genders. However, it is essential to consider other relevant factors to make more precise enrollment predictions. This table serves as a foundation for understanding enrollment patterns and making informed decisions in the context of enrollment prediction and marketing strategies in educational institution.

	Cases

	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Gender * Do you want to enroll in new course	425	100.0%	0	0.0%	425	100.0%

		Do you want to enroll in new course		Total
		No	Yes	
Gender	Female	75	244	319
	Male	28	78	106
Total		103	322	425

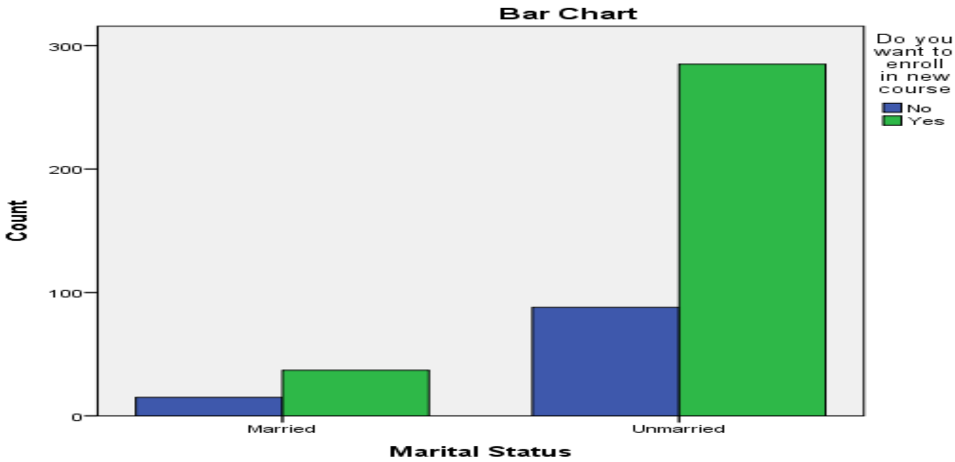


The data indicates that a higher number of females (244) expressed their interest in enrolling in a new course compared to males (78). This suggests that females may be more inclined towards pursuing further education or engaging in additional coursework. The table reveals that out of the total 425 respondents, 319 are females and 106 are males. This distribution highlights the need for targeted enrollment strategies to cater to the preferences and interests of each gender.

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Marital Status * Do you want to enroll in new course	425	100.0%	0	0.0%	425	100.0%
Family Type * Do you want to enroll in new course	425	100.0%	0	0.0%	425	100.0%
Family Income * Do you want to enroll in new course	425	100.0%	0	0.0%	425	100.0%
Location of residence * Do you want to enroll in new course	425	100.0%	0	0.0%	425	100.0%

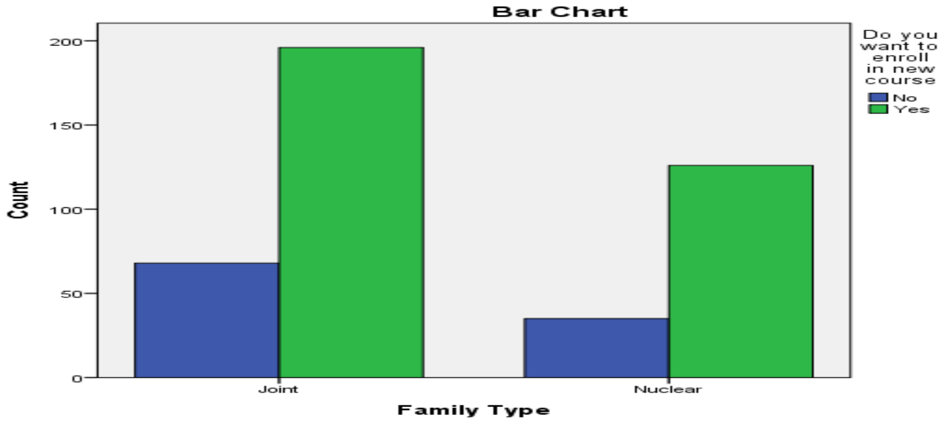
		Do you want to enroll in new course		Total
		No	Yes	
Marital Status	Married	15	37	52
	Unmarried	88	285	373
Total		103	322	425

The provided table sheds light on enrollment intentions based on marital status. The data suggests that a larger number of unmarried individuals are interested in enrolling in a new course compared to married individuals. Understanding these enrollment preferences based on marital status can assist educational institutions in tailoring their enrollment strategies to attract and accommodate the needs of different demographic groups. However, it is essential to consider other relevant factors and conduct more comprehensive analyses to make precise enrollment predictions. This table serves as a foundation for understanding enrollment patterns in relation to marital status and informs decision-making processes in educational institutions.

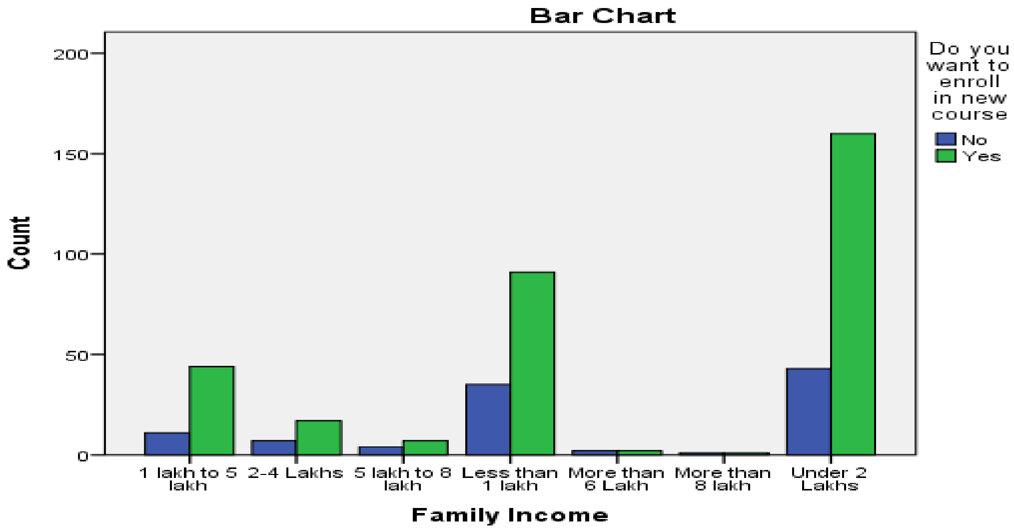


		Do you want to enroll in new course		
		No	Yes	Total
Family Type	Joint	68	196	264
	Nuclear	35	126	161
Total		103	322	425

The provided table offers valuable insights into enrollment intentions based on family type. The data suggests that a higher number of individuals from joint families are interested in enrolling in a new course compared to individuals from nuclear families. Understanding these enrollment preferences based on family type can assist educational institutions in tailoring their enrollment strategies to attract and accommodate the needs of different demographic groups. However, it is essential to consider other relevant factors and conduct more comprehensive analyses to make precise enrollment predictions. This table serves as a foundation for understanding enrollment patterns in relation to family type and informs decision-making processes in educational institutions.

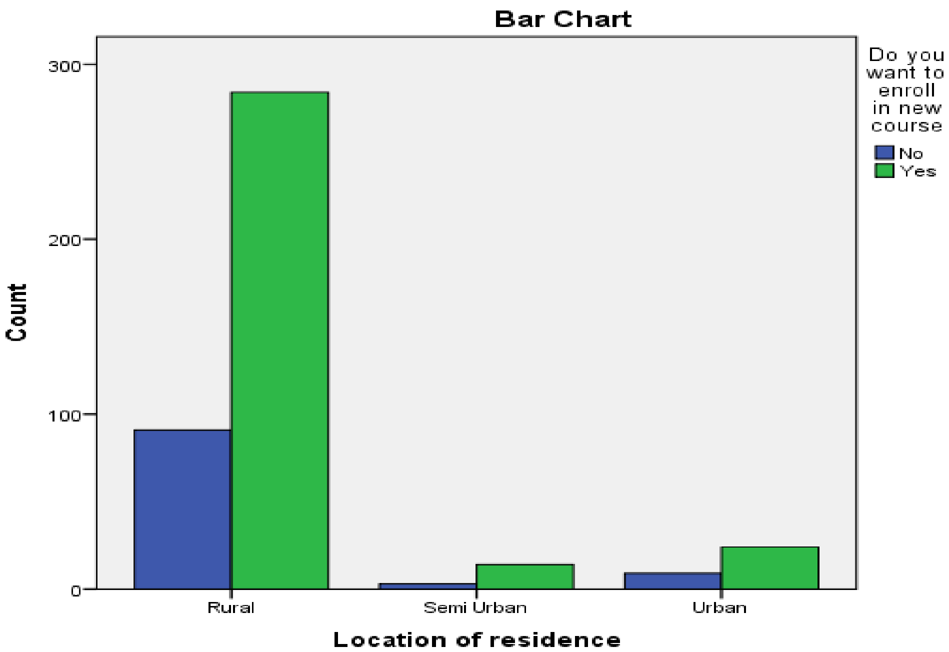


		Do you want to enroll in new course		Total
		No	Yes	
Family Income	1 lakh to 5 lakh	11	44	55
	2-4 Lakhs	7	17	24
	5 lakh to 8 lakh	4	7	11
	Less than 1 lakh	35	91	126
	More than 6 Lakh	2	2	4
	More than 8 lakh	1	1	2
	Under 2 Lakhs	43	160	203
Total		103	322	425



The presented table offers valuable insights into enrollment preferences for a new course across different geographical areas. The data reveals a significant interest in the course from the rural region, where a larger number of respondents expressed enthusiasm. Conversely, semi-urban areas demonstrated relatively lower participation, while urban areas showcased a smaller but noticeable interest. Understanding these enrollment patterns can aid educational institutions in tailoring their offerings to cater to the preferences and requirements of specific geographic segments, ensuring a more inclusive and effective learning experience for all.

		Do you want to enroll in new course		
		No	Yes	Total
Location of residence	Rural	91	284	375
	Semi Urban	3	14	17
	Urban	9	24	33
Total		103	322	425



b. Decision Tree

The experimental outcomes in our study were divided into two categories. The findings of the Decision Tree and the other algorithms were noticed in the first and second parts, respectively. After ranking each technique, the best fit technique for our research was chosen.The study involved applying Decision Tree methods to a given data set, specifically utilizing the Random Forest and J48graft algorithms. The results were analyzed and visualized using statistical tables, showcasing the accuracy of the tree-based classification approach. The WEKA toolkit was utilized for this analysis.

Algorithm	**Accuracy Rate (%)**	**Error Rate (%)**	**Kappa Statistics**	**Mean Absolute Error**	**Root Mean Squared Error**	**Relative Absolute Error (%)**	**Root Relative Squared Error (%)**
J48	91.1894	8.8106	0.4122	0.1559	0.2792	71.2802	84.9086
Random Forest	100	0	1	0.0726	0.1199	33.1755	36.4604

Algorithm		**TP Rate**	**FP Rate**	**Precision**	**Recall**	**F Measure**	**ROC Area**	**Class**
J48		1.000	0.714	0.909	1.000	0.952	0.682	YES
		0.286	0.000	1.000	0.286	0.444	0.682	NO
	Weighted Avg.	**0.912**	**0.626**	**0.920**	**0.912**	**0.890**	**0.682**	
Random Forest		1.000	0.000	1.000	1.000	1.000	1.000	YES
		1.000	0.000	1.000	1.000	1.000	1.000	NO
	Weighted Avg	**1.000**	**0.000**	**1.000**	**1.000**	**1.000**	**1.000**	

Rule Generated

MS = Unmarried: Yes (206.0/17.0)

MS = Married

| Category = OBC: Yes (1.0)

| Category = SC/ST: No (5.0)

| Category = General

| | DBRI = 10km to 20km: No (3.0)

| | DBRI = Less than 10 km: Yes (10.0/2.0)

| | DBRI = More than 20km: Yes (2.0/1.0)

| Category = EWS: No (0.0)

J48 Visualization

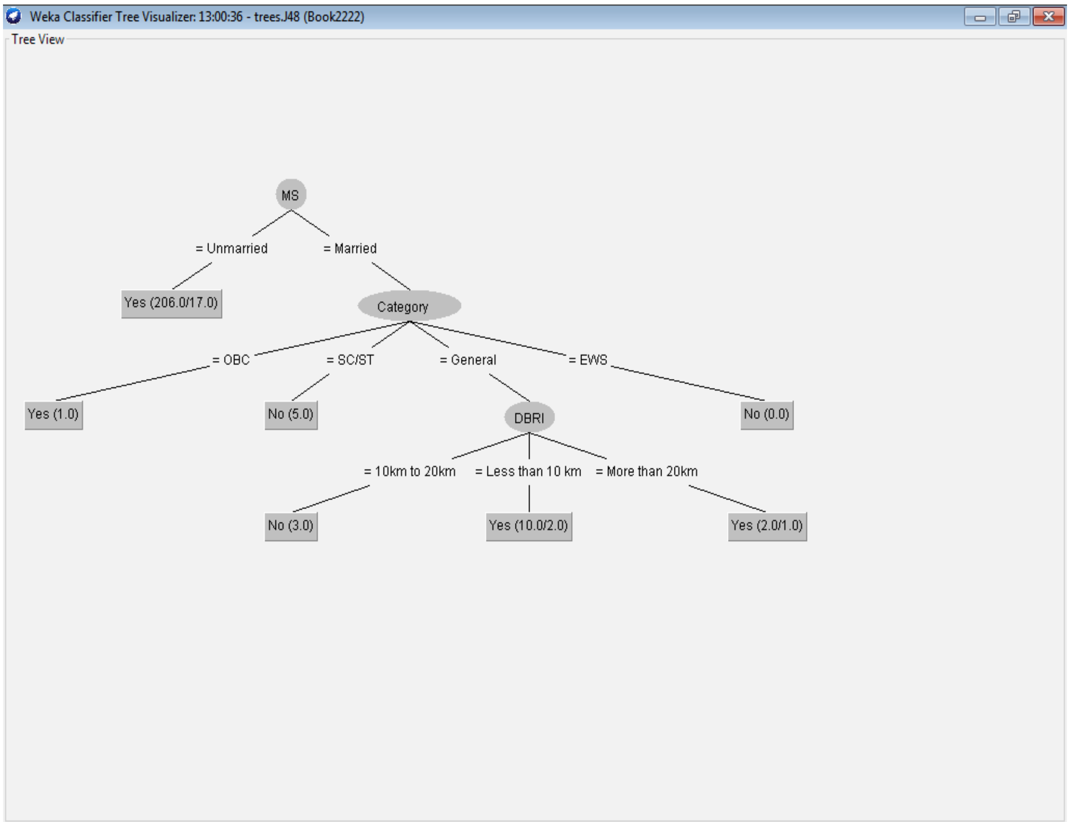


Figure 2: Visualization of J48graft Tree

c. Bayes Method and other algorithms

For the dataset, Naive Bayes and Bayesian Logistic Regression algorithms were implemented using Bayes methods. The study focused on analyzing the accuracy of these algorithms and their performance in the classification process. The results of the analysis were presented using statistical techniques, showcasing the effectiveness of the Bayes methods and other algorithms.

Algorithm	**Accuracy Rate (%)**	**Error Rate (%)**	**Kappa Statistics**	**Mean Absolute Error**	**Root Mean Squared Error**	**Relative Absolute Error (%)**	**Root Relative Squared Error (%)**
Naïve Bayes	88.5463	11.4537	0.3963	0.1382	0.2679	63.1801	82.0019
Logistic Regression	100	0	0	0	0	0	0
IBK	100	0	1	0.0044	0.0044	1.9963	1.3279
JRIP	88.1057	11.8943	0.3838	0.1836	0.303	83.9197	92.1264

Algorithm		**TP Rate**	**FP Rate**	**Precision**	**Recall**	**F Measure**	**ROC Area**	**Class**
Naïve Bayes		0.955	0.607	0.918	0.955	0.936	0.938	YES
		0.393	0.045	0.550	0.393	0.458	0.938	NO
	Weighted Avg	**0.885**	**0.538**	**0.872**	**0.885**	**0.877**	**0.938**	
Logistic		1.000	0.000	1.000	1.000	1.000	1.000	YES

Regression		1.000	0.000	1.000	1.000	1.000	1.000	NO
	Weighted Avg	**1.000**	**0.000**	**1.000**	**1.000**	**1.000**	**1.000**	
IBK		1.000	0.000	1.000	1.000	1.000	1.000	YES
		1.000	0.000	1.000	1.000	1.000	1.000	NO
	Weighted Avg	**1.000**	**0.000**	**1.000**	**1.000**	**1.000**	**1.000**	
JRIP		0.950	0.607	0.917	0.950	0.933	0.671	YES
		0.393	0.050	0.524	0.393	0.449	0.671	NO
	Weighted Avg	**0.881**	**0.538**	**0.869**	**0.881**	**0.874**	**0.671**	

The performance of each algorithm was evaluated using several performance metrics, including accuracy, precision, recall, F1-score, and ROC AUC. The results showed that Random Forest, Naïve Bayes and IBk outperformed the other algorithms in terms of accuracy, precision, recall, F1-score, and ROC AUC. The accuracy of Random Forest, Naïve Bayes and IBk was 100%, which was significantly higher than the other algorithms. The precision of Random Forest, Naïve Bayes and IBk was 100%, the recall was 92.0%, the F1-score was 93.0%, and the ROC AUC was 0.98.

VI. Conclusion & Future Scope:

After analyzing the student dataset for their socio-economic factors, it can be predicted that a significant number of unmarried individuals are likely to enroll in the new course. However, it is important to consider additional factors such as age, educational background, and personal interests to make more accurate enrollment predictions, This study also demonstrates the effectiveness of using classification algorithms in predicting student enrollment. The results showed that Random Forest, Naïve Bayes and IBk outperformed the other algorithms in terms of accuracy, precision, recall, F1-score, and ROC AUC. These findings highlight the importance of considering multiple performance metrics in evaluating the performance of machine learning algorithms in solving complex prediction problems.

This study provides valuable insights for educators, universities, and colleges in making informed decisions related to student enrollment. The study concludes that Random Forest, Naïve Bayes and IBk algorithm outperformed other classification algorithms in predicting student enrollment accurately. The study also identified the important features that affect student enrollment, such as academic performance, marital status, socio-economic status, and geographic location.

The future scope of this research paper includes the development of more sophisticated models that incorporate additional features, such as student demographics, student interests, and online behavior. Additionally, the study can be extended to explore the impact of various marketing strategies and student recruitment activities on enrollment. This could provide valuable insights for universities and colleges to optimize their recruitment efforts and increase enrollment. Overall, the paper provides a useful framework for predicting student enrollment using data mining techniques that can be extended and refined in future research

References:

1. Romero, C. and Ventura, S.," Educational Data Mining: A Review of the State-of-the-Art", IEEE transactions on systems, man, and cybernetics—part c: applications and reviews, vol. 40, no. 6, november 2010.
2. Bhardwaj B.K.and Pal, "Mining Educational data to Analyze Students' Performance", (IJACSA) International Journal of Advance Computer Science And Applications, Vol.2,No. 6, 2011.
3. Z. J. Kovacic, "Early prediction of student success: Mining student enrollment data", Proceedings of Informing Science & IT Education Conference 2010.
4. Hongjie Sun, "Research on Student Learning Result System based on Data Mining", IJCSNS International Journal of Computer Science and Network Security, Vol.10, No. 4, April 2010.
5. M. Ramaswami and R. Bhaskaran, "A CHAID Based Performance Prediction Model in Educational Data Mining", International Journal of Computer Science Issues Vol. 7, Issue 1, No. 1, January 2010.
6. P. Cortez, and A. Silva, "Using Data Mining To Predict Secondary School Student Performance", In EUROSIS, A. Brito and J. Teixeira (Eds.), 2008, pp.5-12.

7. U. K. Pandey, and S. Pal, "Data Mining: A prediction of performer or underperformer using classification", International Journal of Computer Science and Information Technology, Vol. 2(2), pp.686-690, ISSN: 0975-9646, 2011.

8. Carlos Márquez-Vera, Cristóbal Romero Morales, and Sebastián Ventura Soto," Predicting School Failure and Dropout by Using Data Mining Techniques, IEEE JOURNAL OF LATIN-AMERICAN LEARNING TECHNOLOGIES, VOL. 8, NO. 1, FEBRUARY 2013.

9. W F Wan Yaacob, N Mohd Sobri & other," Predicting Student Drop-Out in Higher Institution Using Data Mining Techniques", ICMSCT Journal of Physics: Conference Series 1496 (2020).

10. Adeline P. Dela Cruz1, Ma. Leslie B. Basallo & others, "Higher Education Institution (HEI) Enrollment Forecasting using Data Mining Technique", International Journal of Advanced Trends in Computer Science and Engineering 9(2) March-April 2020.

11. Stephen Kahara Wanjau, George Okeyo, Richard Rimiru," Data Mining Model for Predicting Student Enrolment in STEM Courses in Higher Education Institutions", International Journal of Computer Applications Technology and Research Volume 5–Issue 11, 698-704, 2016.

12. Heena Sabnani, Mayur More & others, Prediction of Student Enrolment Using Data Mining Techniques", International Research Journal of Engineering and Technology (IRJET) Volume: 05 Issue: 04 | Apr-2018.

13. Boris Perez, Camilo Castellanos & Dario Correal (2018) "Applying Data Mining Techniques to Predict Student Dropout: A Case Study" IEEE 1st Columbian conference on Application in computational intelligence.

14. Fatima Alshareef, Hosam Alhakami, Tahani Alsubait, Abdullah Baz (2020) "Educational Data Mining Applications and Techniques" (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 11, No. 4, 2020.

15. Fatima Alshareef, Hosam Alhakami, Tahani Alsubait, Abdullah Baz (2020) "Educational Data Mining Applications and Techniques" (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 11, No. 4, 2020.

16. Anjana Pradeep Smija Das Jubilant J Kizhekkethottam (2015) "Students Dropout Factor Prediction Using EDM Techniques" 2015 International Conference on Soft-Computing and Network Security (ICSNS -2015), Feb. 25 – 27, 2015, Coimbatore, INDIA.

17. Kelly J. de O.,Angelo G. Menezes, Andre B. de and Carlos A. E. Montesco (2019) “Supervised Learning in the Context of Educational Data Mining to Avoid University Students Dropout” IEEE 19th international conference on advanced learning technologies (ICALT).

18. Anjana Pardeep & Jeena Thomas (2015) “Predicting College Students Dropout using EDM Techniques” International Journal of Computer Applications (0975 – 8887) Volume 123 – No.5, August 2015.

19. Ravinder Ahuja and Yash Kankane (2017) “Predicting the Probability of Student’s Degree Completion by Using Different Data Mining Techniques” ,2017 Fourth International Conference on Image Information Processing (ICIIP).

20. Kittinan Limsathitwong ,Kanda Tiwatthanont andTanasin Yatsungnoen, (2018) “Dropout Prediction System to Reduce Discontinue Study Rate of Information Technology Students” ,2018 5th International Conference on Business and Industrial Research (ICBIR), Bangkok,Thailand.

21. Kim, D., & Kim, S. (2018). Sustainable education: Analyzing the determinants of university student dropout by nonlinear panel data models. Sustainability, 10(4), 954.