

## Machine Learning Based Scholarship and Credit Pre-Assessment System

Ramakrishna Hegde<sup>1</sup>, Anusha G V<sup>2</sup>, Sindhu Madival<sup>3</sup>, Sowjanya H S<sup>4</sup>, Sushma U<sup>5</sup>, Soumyasri S M<sup>6</sup>

<sup>1,2,3,4,5</sup>Vidyavardhaka college of Engineering, Mysore, India

<sup>6</sup>Sapient College of Commerce and Management, Mysore, India

**ABSTRACT:** The researcher explained the implementation process of finding the scholarship for the students by using machine learning supervised learning algorithm i.e. Naïve Bayes algorithm. Addition to this it includes a small description of naïve bayes classifier which used to be used through the authors. It explains the significance of training facts set and trying out information set in Machine mastering techniques. Machine learning nowadays becomes plenty used technique in the field of IT industry. It is a very effective instrument and technique for many quite a number fields such as education, IT and even in enterprise industry. In this paper, the researcher attempt to find computerized end result reputation of scholarships of college students by way of using naïve bayes classifier algorithm primarily based on the scholar educational performance, conversation skills, greedy power, IHS, income, time management, regularity etc. A scholarship offers a strength and self assurance to a student. It also boosts the performance of students indirectly. Usually scholarships are furnished by governments or authorities organizations. It is very essential for students to recognize their personal potentiality early in their educational profession so that they faster its growth, receiving attention from an employer or corporation helps college students take this step. Students can apply for scholarships primarily based on the eligibility criteria (such as caste category, annual income, etc). The scholarship will be issued based on merit, student performance and career specific. Different schemes of scholarships are provided for the students based on distinct eligibility criteria. By the use of a naïve bayes classifier, the researcher acquired a end result with accuracy of 96.7% and error of 3.3%. The repute of scholarship students was once displayed in the form of yes or no.

**KEYWORDS:** Machine Learning, Scholarship, Naïve Bayes, Bayes Theorem, Supervised Learning Method, Scholarship, Prediction

### 1. INTRODUCTION

This paper explains about the implementation manner of scholarship prediction system via the use of machine learning techniques like naïve bayes classifier. The pros and cons of Machine learning and Data mining methods are discussed. It additionally describes the naïve bayes classifier and working of algorithm.

A Scholarship are financial aid for college students to proceed their studies. Usually scholarships are supplied through governments or authorities organizations. If students are identified for their performances it offers them the confidence to pursue different goals. It is very essential for students to recognize their own potentiality early in their tutorial profession so that they foster its growth, receiving cognizance from an enterprise or company helps college students take this step. Students can follow for scholarships based totally on the eligibility standards (such as caste category, annual income, etc). There are some of the common scholarships will be awarded based on merit, student performance and career specific. Different schemes of scholarships are furnished for the students based totally on different eligibility criteria.

The existing system was very time killing process. The scholarship will be awarded based on taking interviews with applied status and by comparing all students data. Hence it required more time to detect the most eligible student out of huge students. This manner is causing time free and additionally pretty exhausting. On the different hand, it turns into to tough to make goal decisions in some situations. Using artificial intelligence and machine learning in the process of detecting college students who will be given scholarship is essential to consider in each goal and subjective ways to students who apply. This task is designed to supply scholarship to the college students where the eligible college applied students are shortlisted based on the parameters or conditions such as percentage, grade, annual income, aggregate, conversation skills, aptitude etc. by the usage of machine learning techniques.

### 2. LITERATURE SURVEY

As per the paper "Computerized reasoning Based Grant DQG Credit Pre-Evaluation Framework" [4], Grants are given to understudies who are concentrated in college by open foundations and associations by considering various standards. The understudies who have the right to have grant are

identified by inspecting data in the structures that are given to understudy to identify the individual who needs the grant. This cycle make time free and furthermore very debilitating. Then again, it turns out to be too hard to even consider settling on target choices in certain circumstances. Utilizing man-made reasoning and AI during the time spent distinguishing understudies who will be given grant is imperative to assess in both level headed and abstract manners to understudies who apply an exactness of over 80%.

As indicated by paper "The Correlation of Straight Relapse Strategy and K-Closest Neighbours in Grant Beneficiary" [6], The grants grant for understudies are frequently abstract, not straightforward, un-quantifiable, and less exact on track. One of the PC advancements used to deal with enormous information, for example, grant beneficiary issues is information mining. Different strategies for information mining can be utilized to anticipate the attainability of information, for example, K-Closest Neighbours (KNN) and Direct Relapse [4]. This investigation thinks about the two techniques in tackling the grant beneficiary issue. The credits utilized are Semester period, Grade Point Normal (GPA), Proclamation Letter of Dynamic Understudy, Letters of Help, Understudy Personality Card, Character Card, Family Card, Study Result Card, Explanation Letter, Ledger, and Articulation Letter of Passed Organization. The factors utilized in the correlation interaction incorporate Exactness, Accuracy, Review, Order Mistake, Supreme Blunder, and Root Mean Square Mistake (RMSE). Information from 8212 grant beneficiaries are tried through re-enactment testing of preparing information and testing information 90:10, 70:30, 50:50, 30:70, and 10:90. In this, Fast digger is utilized as an apparatus to see the consequences of examination from the two techniques. Subsequently, the two strategies for information recreation 90:10 and 70:30 give 100% of exactness, accuracy, and review. The outcome has shown that KNN strategy is more powerful and proficient instead of Direct Relapse technique. This gives new information commitment. Ideally, the determination interaction of grant beneficiaries can be carried out much better, straightforward, as of now not abstract, and right on the objective.

Likewise, as indicated by kola Soyibo, (1998) [5] portrays that all understudy execution is generally significant on SEB. It discovers the understudies with contrasts in secondary school understudy execution when they getting bunch.

In 2015 Ahmad [7] likewise utilized same AI strategies in his paper called "The expectation of understudy's scholastic execution utilizing grouping information mining procedures". Creator gathered information of four years, 2012-2015. The gathered information go through some pre-handling steps to eliminate loud information. At long last the analyst got a right calculation to discover the understudy's scholastic execution and the tips to work on the presentation of every understudy.

As indicated by creator Bunker, Kamal (2012) [3] clarifies that data mining grouping helps in improving and analyzed the information of understudies. The information mining strategy like k-implies, choice tree, Bayesian classifier calculations are ordinarily utilized for the expectation of scholarly execution of college understudies. They got an outcome with best exactness.

As per Thalia An Agnos, Eva Shirring (2018) [1] portrays grant is vital for each understudy to help monetarily in their investigations and it likewise propel the under studies individual, wellbeing or monetary issues may prompt lose their grant openings the educators and they speaks with industry experts prompts consequence in the program. The analyst recommended that given framework offer kind help and projects to spur and furthermore help to privileged understudies to complete their science certificates with specific abilities, information and demeanour which arise understudies as a pioneer. This paper depicts the boost effect of grants and beneficiary's authority abilities and mentalities, their future ways and their college experience yet the proposed framework sets aside more effort for information handling and productivity is low.

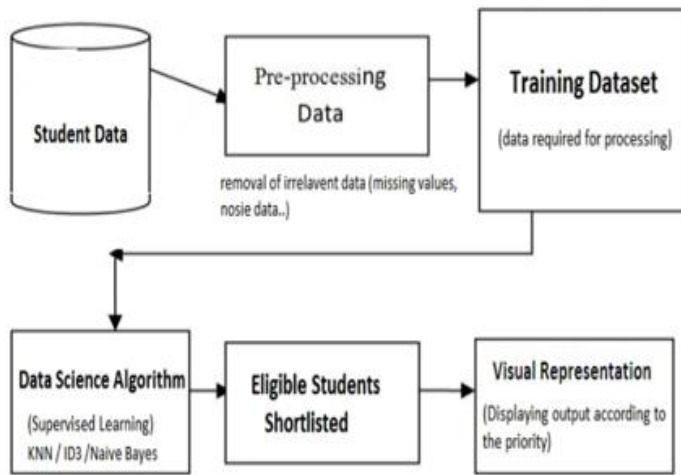
### 3. METHODOLOGY

#### A. Data Gathering

Data Gathering is a process of collecting all previous scholarships data. In research time the researchers conducted an interview to collect information about the process of scholarship. they collected data in the form of excel sheet format. The researcher used rapid miner tool for pre-processing the collected data.

#### B. Data Pre-Processing

Researcher selected some of the variables from the data which was collected. Some of the attributes had missing values, null values. To get the accurate result, all those data had to be modify or clean. All collected data undergo pre-processing step which also called as data cleansing. This step helps to fill the missing values in the dataset, identifying and removing the un-correct data and normalization of data. The researcher add an extra field in the dataset i.e. status of scholarship of a students. The dataset has 992 rows and 9 columns and 8928 instances. The pre-processed data step helps to get normalized data.



**Figure No 1.** Data flow diagram of scholarship prediction system

The researcher spitted the dataset into two parts i.e. training dataset and testing dataset. The training dataset helps to predicting the scholarship of the students. The testing dataset helps to check the accuracy of the output result. Using a naive bayes classifier researcher were able to find out the most eligible students for getting scholarship in bunk of applied students. These all eligible students were visually displayed in the screen.

**Table 1.** List of attributes used in scholarship prediction system

Attributes	Description
Regno	Register number of a students.
Income	Annual income of the students.
Regular	Students regularity and attendance in class.
Communication skills	Students conveying capability and sharing of ideas effectively.
Time management	Organizing and planning the time of students.
Grasping ability	Understanding status of the students.
Ex-Activities	The state of activeness of the students.
IHS	Interest in higher studies.
status	Yes or No status of scholarship.

In this system, we have integrated the data from the kaagle website to train the model to predict scholarship based on the dataset provided by any organization. The data related to the scholarship prediction is taken from kaagle website which contains 9 different attributes and such attributes reg no, income, Time management, Ex-activities, Grasping power, HIS, Regular, communication skills, Aggregate. And these attributes of scholarship prediction dataset will be in

string or integer format based on attributes and before using initially taken scholarship prediction dataset, the dataset need to be processed. The entire dataset is converted to the required format and if the dataset contains irrelevant data or any missing data or any duplicate data then all these data will be rectified.

In this study the researcher used machine learning technique for prediction of scholarship of the students. The algorithm used by the researcher was Naive Bayes classifier to get an accurate result. The naive bayes classifier belongs to the family of probabilistic classifiers. Data will be trained effectively in supervised learning method. The speciality of the naive bayes classifier it can predict the result or output by using small kind of training data.

It used the bayes theorem, the conditional probability. It can be represented as

$$\text{Posterior} = \frac{\text{prior} * \text{likelihood}}{\text{Evidence}}$$

The above conditional probability equation also represented as

$$P(C_r | x_1, \dots, x_n) \propto p(C_r | x_1, \dots, x_n) p(C_r) p(x_1 | C_r) p(x_2, \dots, x_n)$$

$$p(C_r) \prod_{i=1}^n p(x_i | C_r)$$

where  $\alpha$  represents the probability.

This means the conditional distribution over class C is:

$$P(C_r | x_1, \dots, x_n) = \frac{1}{Z} p(C_r) \prod_{i=1}^n p(x_i | C_r)$$

**Algorithm**

**Step 1:** Dataset has to be scanned. (storage servers)

**Step 2:** Finding the probability of each attribute i.e. n, n\_c, m, p.

**Step 3:** Next apply the below formulae to find the probability value of an attribute.

$$P(\text{attribute value}(a_i)/\text{subject value}(v_j)) = (n_c + mp)/(n+m)$$

Where:

- n = the no. of training data for which v = v<sub>j</sub>
- n\_c = number of data for which v = v<sub>j</sub> and a = a<sub>i</sub>
- p = a Prior estimate for P(a<sub>i</sub>|v<sub>j</sub>)
- m = size of testing data

**Step 4:** Product the probabilities by p

**Step 5:** classify the values of attributes to another set of classes by comparing.

# “Machine Learning Based Scholarship and Credit Pre-Assessment System”

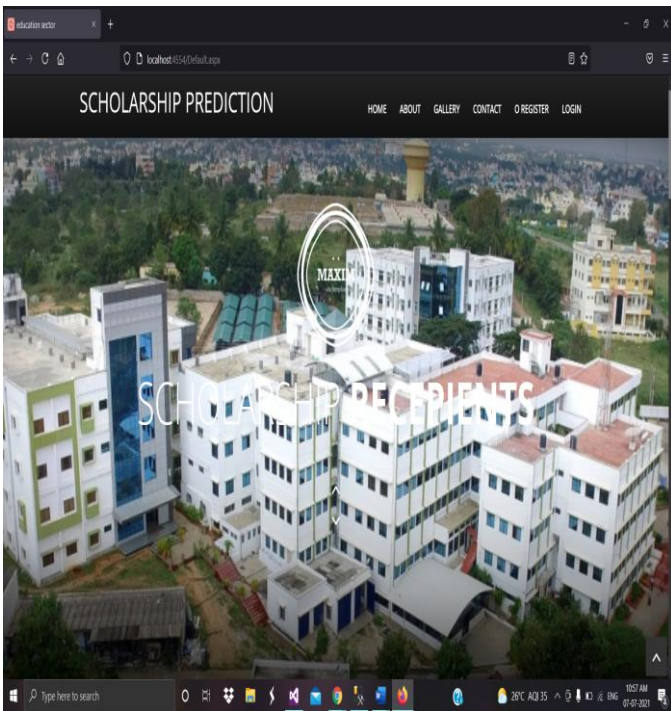


Figure No 2. Home page

In the figure 2 the first window of the scholarship prediction system it shows so many options in the top i.e. home, about us, gallery, contact, register and login. In this system we have 3 types of users namely admin, student and organization. By entering a correct id and password they were able to login.

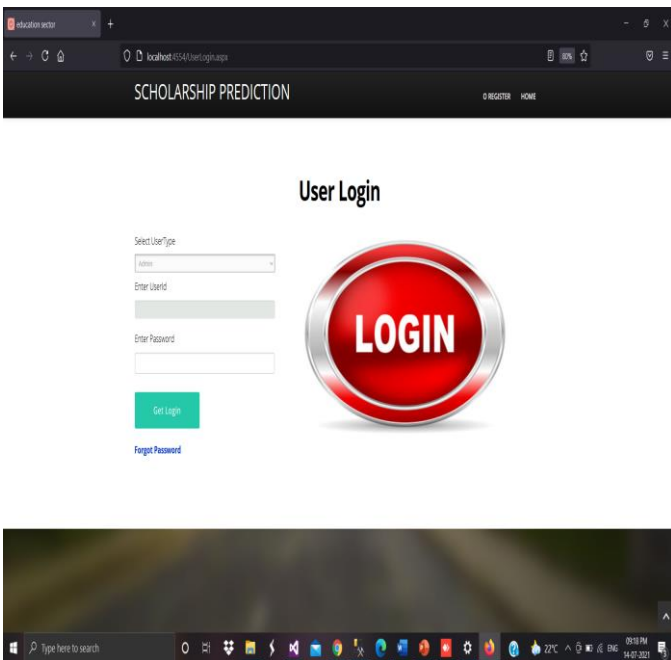


Figure No 3. User login page

The figure 4 shows adding of student details by the admin to the system. In this page it is required to enter the name, mobile number, e-mail id, course and semester of the student. After entering the details admin has to press add student button to save the details in the database.

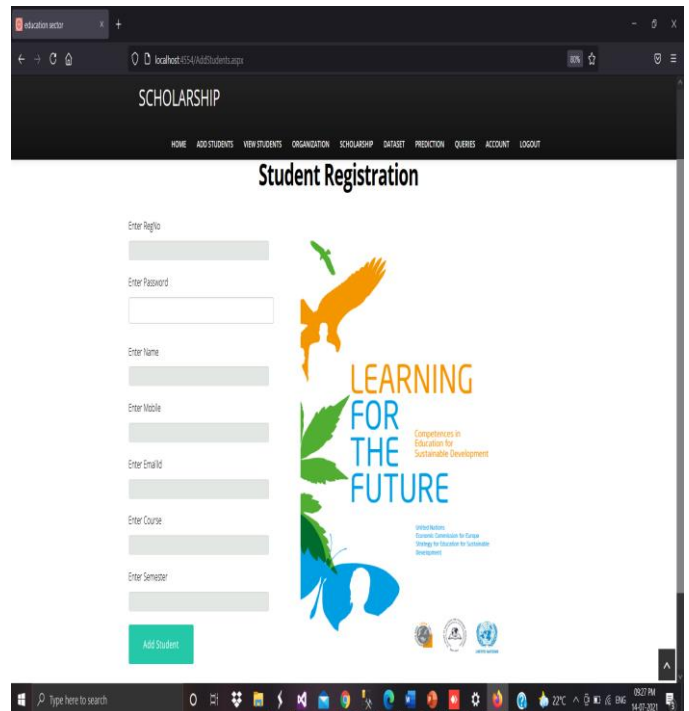


Figure No 4. Student registration by Admin

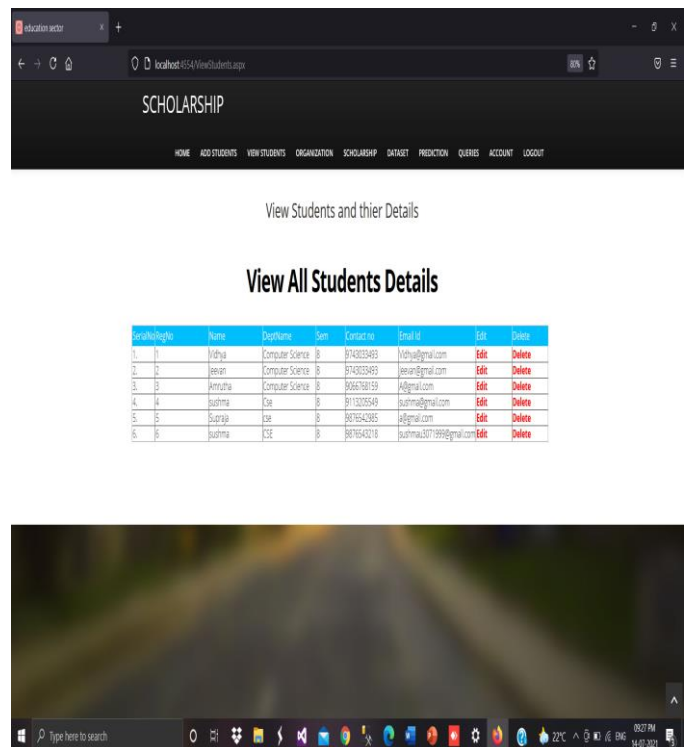


Figure No 5. Registered Students details page

The figure 5 represents all the registered students details by admin. It had the extra feature of edit and delete the record of particular student.



# “Machine Learning Based Scholarship and Credit Pre-Assessment System”

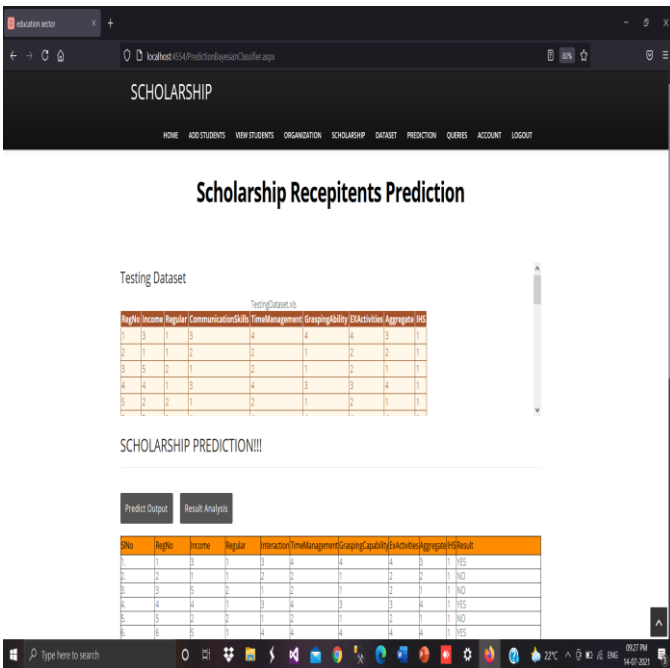


Figure No 6. Scholarship prediction page

In figure 6 shows the tracking of scholarship result by the students. The student can view the status of result whether he/she get a scholarship or not. The student performance and the scholarship status automatically get highlighted when the student press the prediction button.

## 4. RESULT AND DISCUSSIONS

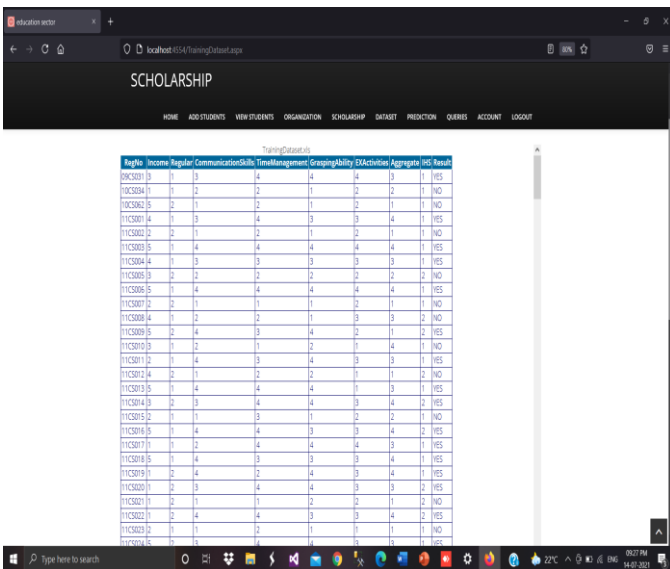


Figure No 7. Scholarship status of each students

In the figure 7 shows prediction page where it contains prediction output and the accuracy of the result. The status of the prediction output will be displayed in the form of yes or no. If the status was "Yes" it means the student was eligible for the scholarship. If the status was "No" it means student was not eligible for getting scholarship from the institution or organization.

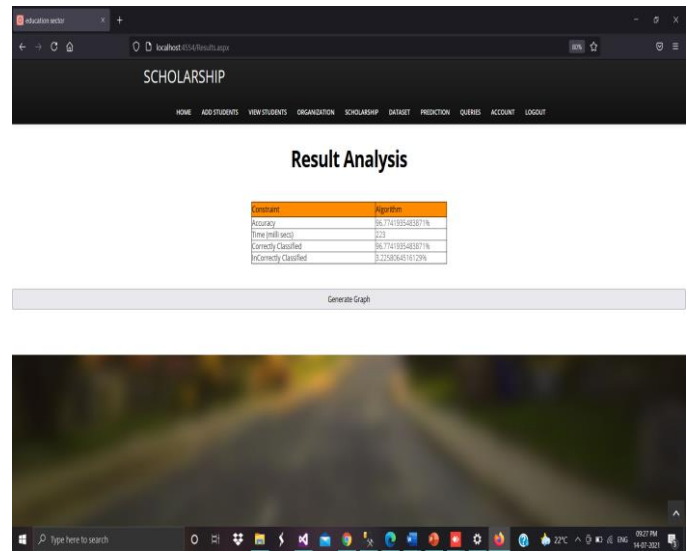
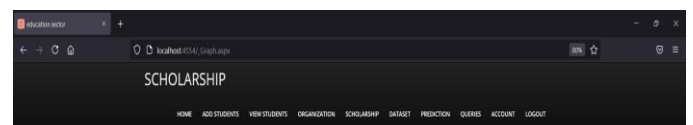


Figure No 8. Result Analysis

In the figure 8 show the output analysis part where the accuracy of this system was 6.7 and the error of 3.3. Due to high accuracy it concluded that the naïve bayes classifier give a best result for the prediction of scholarship.



Graph Representation (Algorithm Vs Accuracy)!!!

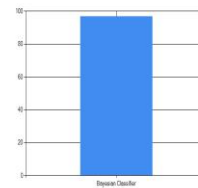


Figure No 9. Graphical representation accuracy page

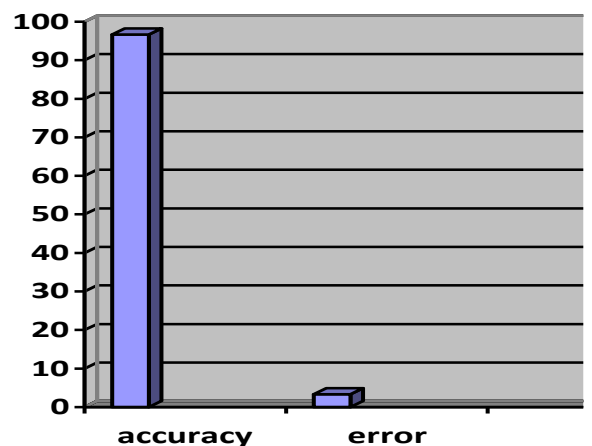
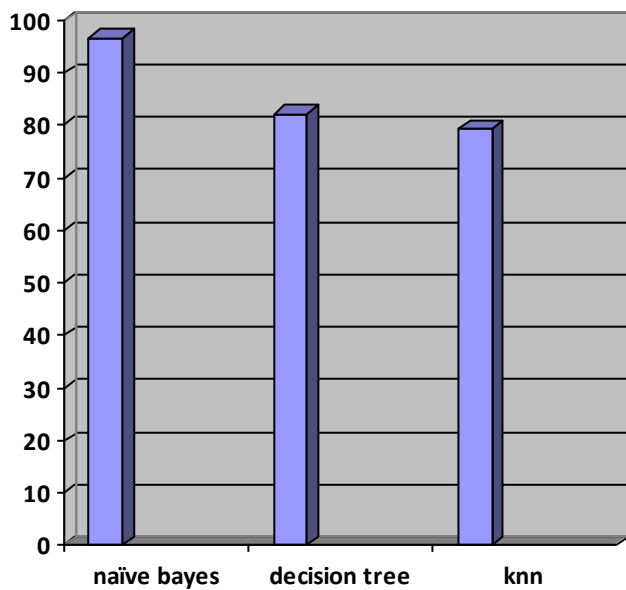


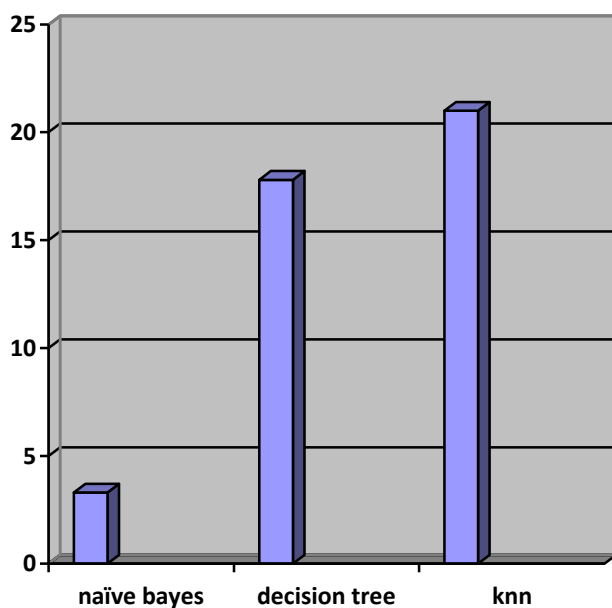
Figure No 10. Accuracy and error of output

This figure 10 shows accuracy and error between Naïve Bayes. Naïve Bayes algorithm has 3% error. IT concludes that naïve bayes classifier gives most accurate result with high performance.



**Figure No 11.** Comparison of accuracy of naïve bayes algorithm, decision tree algorithm and knn algorithm

The figure 11 explains the comparison of accuracy of the system of naïve bayes, decision tree and knn algorithms used to prediction of scholarship of students. The naïve bayes algorithm have high accuracy of 96.7 compared with other algorithms.



**Figure No 12.** Comparison of error of naïve bayes, decision tree and knn algorithm

The figure 12 explains the comparison of error between naïve bayes, decision tree and knn algorithm used to prediction of scholarship of students. The naïve bayes algorithm gives less error compared with other algorithms.

**5. CONCLUSION**

Scholarships are given to students to reach their goals by considering different criterion’s. The existing system was very time killing process. The scholarship will be awarded based on taking interviews with applied status and by comparing all students data. Hence it required more time to detect the most eligible student out of huge students. This causes time loose and also quite exhausting. Also it feels difficult to take certain decisions in some situations. Proposed system helps in scholarship recipients using data science technique. To predict the scholarship, the data science method Naïve Bayes supervised learning algorithm used. This system gives the results with higher accuracy. Proposed system for scholarship prediction using machine learning or AI techniques.

This system is designed for an educational sector which helps to predicting the scholarship based on the scholarship parameters. It identifies the students who are eligible to get scholarship by using this system in an efficient and faster manner.

**6. FUTURE WORK**

We have made the maximum utilization of our potential in developing this project. But gaining knowledge is a continuous process. Therefore in this section we present some of the ideas which can be used to enhance the functionalities of our project to widen its applications. In the proposed system few constraints are used such as grade, annual income, percentage category, communication skills, aptitude etc... In addition we can use more parameter for the prediction.

**Message or E-mail Module**

In this system, admin has a access to assigns Id and password for students those who applied for scholarship and is feels manually. So we can add message or E-mail feature as a future enhancement where students receives an message or E-mail about their Id and password.

**REFERENCES**

1. Thalia Anagnos, Eva Schiorring, “Helping Students Get More than Teir Money from an Engineering Scholarship Program” IEEE, 2018.
2. Angela R. Bielefeldt, “Characteristics of Engineering Faculty Engagedin the Scholarship of Teaching and Learning” IEEE, 2015.
3. K. Bunkar, "Data Mining: Prediction for Performance Improvement of Graduate Students using," IEEE, 2012.

4. Sercan, Saatci, Hande cansiz, gulsha aslan,erkan ozhan “Artificial intelligence-based Scholarship DQG Credit pre-assessment system” IEEE, 2017.
5. G. Upton, "The Effects of Merit-Based Scholarships on Educational Outcomes," *Journal of Labor Research*, vol. 37, no. 2, pp. 235-261.
6. Okfalisa, Ratik a fitriani, Yelli Vitriana, “ The Comparison of Linear Regression Method and k-Nearest Neighbors in Scholarship Recipient” IEEE, 2018.
7. N. T. Nghe, "A Comparative Analysis of Techniques for Predicting Academic Performance," IEEE, 2007.
8. Jonalyn Joy B .Labayne, Lester L. Mercado, Jheanel Espiritu Estrada, “Model Development Of Students Scholarship Status at First at Asia Institute of technology And Humanities(FAITH)” IEEE,2018.
9. Suma, V., and Shavige Malleshwara Hills. “Data Mining based Prediction of Demand in Indian Market for Refurbished Electronics.” *Journal of Soft Computing Paradigm (JSCP)* 2, no. 02 (2020): 101-110
10. Kumar, T. Senthil. “Data Mining Based Marketing Decision Support System Using Hybrid Machine Learning Algorithm.” *Journal of Artificial Intelligence* 2, no. 03 (2020): 185-193.
11. Chakrabarty, Navoneel. "A Regression Approach to Distribution and Trend Analysis of Quarterly Foreign Tourist Arrivals in India." *Journal of Soft Computing Paradigm (JSCP)* 2, no. 01 (2020): 57-82.
12. Mitra, Ayushi. "Sentiment Analysis Using Machine Learning Approaches (Lexicon based on movie review dataset)." *Journal of Ubiquitous Computing and Communication Technologies (UCCT)* 2, no. 03 (2020): 145-152.
13. Hur, Minhoe, Pilsung Kang, and Sungzoon Cho. "Box-office forecasting based on sentiments of movie reviews and Independent subspace method." *Information Sciences* 372 (2016): 608-624
14. Vohra, S. M., and J. B. Teraiya. "A comparative study of sentiment analysis techniques." *Journal JIKRCE* 2, no. 2 (2013): 313-317.
15. Parvathy G, Bindhu JS (2016) A probabilistic generative model for mining cybercriminal network from online social media.