Use Of Big Data To Predict Competent Students' Higher Studies And Bridge Them With Foundations

M. Dharanika [1], M. Jayasri [2], S. Kaminidevi [3], Mrs. V. Revathy [4]

UG Student, Department of Computer Science and Engineering, Arasu Engineering College, Kumbakonam,

Tamil Nadu, India[1, 2, 3]

Assistant Professor, Department of Computer Science and Engineering, Arasu Engineering College, Kumbakonam,

Tamil Nadu, India[4]

Abstract: Our project's major goal is to raise the quality of education in our country by identifying talented students and providing them with financial aid for further study. In this study, data mining and machine learning are combined with big data to forecast which kids will be capable of pursuing higher education. The anticipated student list will then be connected to charitable foundations and private sponsors. The ability of the learner is predicted using a machine learning system. Big data is utilized to increase forecast accuracy. Data mining is used to examine the necessary data for prediction and to identify relationships in the student data set. By applying the logistic regression approach to forecast the values of pupils, the accuracy is finally assessed.

Keywords: Machine learning, logistic regression, big data

I.   INTRODUCTION

In most cases, mining refers to the removal of precious materials from the ground. Data mining is the process of extracting significant information from a vast volume of data in computer science. Data mining is sometimes referred to as knowledge extraction or knowledge discovery [1]. The technique and approach of data mining are highlighted by Abhijit et al. Data mining is now utilized practically anywhere that a significant volume of data is kept. Most often, data mining information is utilized to forecast future trends and enable commercial decision-making. Data mining is frequently utilized in applications such as financial analysis, biological analysis, scientific analysis, intrusion detection, fraud detection and research analysis.

II.   RELATED WORKS

The effective method of predicting student ability and motivation is based on logistic regression and Sci-kit methods. To collect students and support content, use pandas framework, a software library in python that can manage and analyze data. This data is sent to pre-process where the data can be cleaned and logistic regression performed. The algorithm processed results into the next stage, called training. In this training level sci-kit machine learning library training files, training files can be sent to tensor flow, free machine learning open source library dedicated to training shows data and finally working data is linked and data is uploaded to cloud.

Classification [3] is one of the most important data mining techniques used to classify predefined datasets used in healthcare decision making, to make better diagnosis and treatment for patients. Thyroid disease has also been diagnosed in patients worldwide and has become a major endocrine health concern and concern. The k-Nearest Neighbor, Support Vector Machine, Decision Tree and Naive Bayes are used to classify the data. Compared to other classifiers, Decision Tree achieved the highest accuracy with 98.88%.

Classification is a classic problem in machine learning. In machine learning, algorithms that can perform classification tasks are often called classifiers. Unequal data [9] usually refers to the classification problem in problems where the class distribution is unequal. Unstable information such as spam calls, credit card fraud and natural disaster calls are part of our daily lives. In general, there are two ways to solve the dataset imbalance problem. One focuses on the classifier, the other on the data. Running the Logistic Regression Algorithm to create a more robust logistic regression classifier for classifying fewer classes, the analysis was applied with a Python program.

Increasing reliance on social networking sites [7] has led to the generation of big data, also known as big data. It's usually caused by three issues, including packaging, advertising, and many other things .Data provides only a few techniques for identifying good information from big data sources such as patterns and rules. They attempt to provide an overview of all the foundations of social

analysis by creating paths, mining areas, problems and different challenges. Social media from various file formats including text, video, image, audio, PDF and PowerPoint. Therefore, it is difficult to detect fake news. Facebook is used more and more by many people. These networks allow users to post content about them and connect with their friends. There are social media chart methods community research, recommendation methods, opinion mining and many methods for opinion mining.

III. METHODOLOGY

Data Collection: Collect data from various sources such as academic records, socio-economic profiles, geographic location, and other relevant information about students in rural areas.

Data Pre-processing: Clean and pre-process the collected data to remove any missing or incorrect values, outliers, duplicates, and other inconsistencies in the data.

Data Integration: Combine and integrate the pre-processed data from various sources into a single dataset.

Data Analysis: Analyze the integrated dataset using various big data analytics techniques such as data mining, machine learning, and statistical analysis to identify the students who have the potential for success in higher studies.

Prediction Model Development: Develop a prediction model based on the results of the data analysis that can accurately predict competent students who are likely to excel in higher education.

Partner with Foundations: Partner with charitable foundations, non-profit organizations, and individual sponsors who are willing to provide financial support and other forms of assistance to eligible students in need.

Connect Eligible Students: Connect the eligible students identified by the prediction model with the partnered foundations based on their individual needs and requirements.

Monitor and Evaluate: Monitor and evaluate the performance of the prediction model and the effectiveness of the partnered foundations in bridging the education gap and empowering the next generation of rural leaders. This methodology aims to use big data analytics to identify competent students in rural areas who have the potential for success in higher education, and bridge them with the best higher education opportunities available by collaborating with charitable foundations and individual sponsors. The methodology focuses on using various analytical techniques and prediction models to ensure accurate predictions, connecting eligible students with relevant foundations, and monitoring the overall performance of the program.
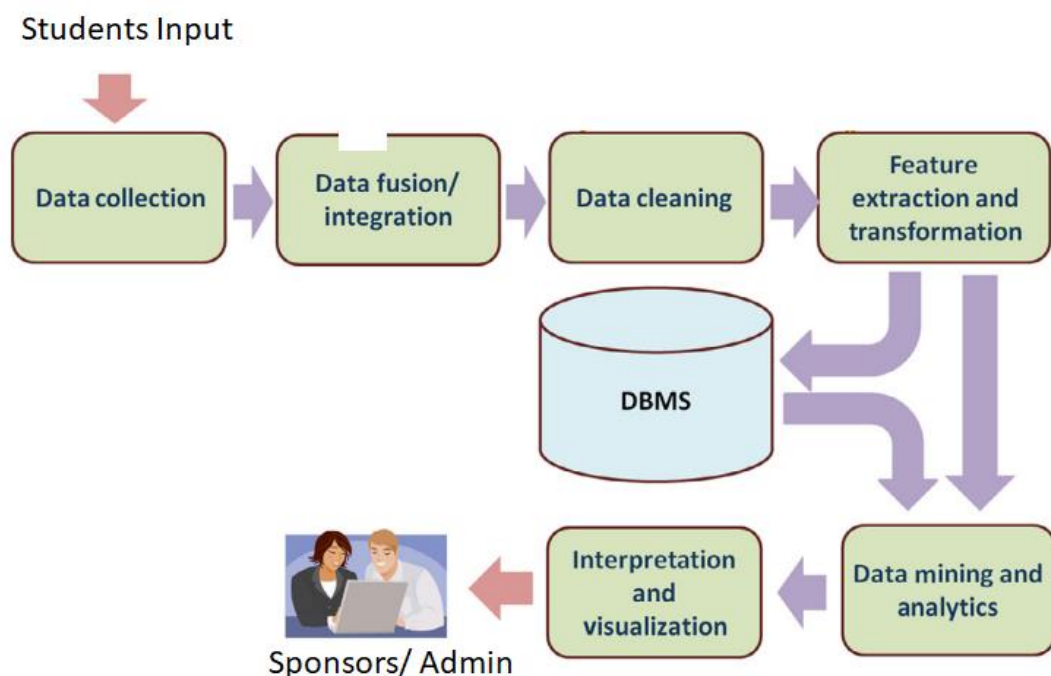
IV.   EXPERIMENTAL RESULTS

We extract relevant data from students who have completed subjects. We did the data frame creation and maintenance of the data, i.e. removing unnecessary features from the data and exporting the necessary data for the project. Checks the conditions regarding whether the student can attend medicine or engineering or other courses. Identify potential students and provide them with confidence and foundation. This system helps identify students who are unable to complete advanced courses and provides them with financial aid. Using this program we are improving the state of education. Predicts more accurately due to the use of machine learning algorithms.

V.　　CONCLUSION

This article describes students who previously estimated should support their higher education and sent their details to the appropriate sponsor. The main purpose of our work is for everyone to receive education and we have achieved our goal. When this process is implemented by the government, the government uses the service to determine how many doctors, engineers, and other professionals exceed in a year. Therefore, the unemployment rate may be very low. Many jobs such as database administrator and cloud management will change as a result of this service.

REFERENCES

[1] Abhijit A. Sawant and P.M. Chawan, "Study of Data Mining Technique for Financial Data Analysis", International Journal of Engineering Science and Innovative Technology (IJESIT).

[2] "Learning Analytics: From Big Data to Meaningful Data Journal of Learning Analytics", Agathe Merceron Beuth, 2015.

[3] Parkavi.A, BibiAminaBegum, "Prediction of thyroid Disease Using Data Mining Techniques," 5th International Conference on Advanced Computing & Communication Systems (ICACCS), 2019.

[4] Lakshmi B.N. and Raghunandhan G.H. "A Conceptual Overview of Data Mining" Proceedings of the National Conference on Innovations in Emerging Technology, 2011.

[5] Machine Learning Applied to Cervical Cancer Data, I.J. Mathematical Sciences and Computing, 2019, pp. 53-64 (Dhwaani Parikh and Vineet Menon).

[6] "Learning Analytics and Educational Data Mining" by George Siemens, Ryan S., and J.D. Baker Jyoti, "More Issues in Mining Techniques in Social Media", IJSRCSEIT, 2017.

[8] Keisuke Abe, "Data Mining and Machine Learning Applications for Educational Big Data in the University", 2019.

[9] "Improving Prediction Accuracy for Logistic Regression On Imbalanced Datasets", Lixictao, Pabhairbattachraya, and Yingquai, 2019.

[10] Shlok Gilda evaluated "Machine Learning Algorithms for Fake News Detection" in 2017 at the Pune Institute of Computer Technology in Pune, India.