International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Student Employability Analysis Using Machine Learning and Data Visualization

Abhay Ramamurthy[1], Karthik Chandramauli[2]

[1,2]Department of Information Science and Engineering, BNMIT, Bangalore-560070, India

Abstract:
This study presents a random forest algorithm student prioritization tool employing ensemble decision trees for classification and incorporating weights of prioritization with predefined criteria. The tool enhances speed in the selection process through candidate classification thereby enhancing efficiency and precision in the identification of suitable candidates who possess attributes most consistent with those desired by the institution. Demonstrating its advantage over traditional methods is in this way.

1. INTRODUCTION

This study presents a random forest algorithm student prioritization tool employing ensemble decision trees for classification and incorporating weights of prioritization with predefined criteria. The tool enhances speed in the selection process through candidate classification thereby enhancing efficiency and precision in the identification of suitable candidates who possess attributes most consistent with those desired by the institution. Demonstrating its advantage over traditional methods is in this way.

Today's higher education institutions have the challenging task of providing them with access to a user-friendly interface that offers a comprehensive overview of students' academic standing and career ambitions, in addition to attracting new business. However, many companies struggle to reduce the number of candidates in their pool to a manageable size, which makes the hiring process even more complicated.

The project entails developing a sophisticated web-based platform that requires two logins: one for student placement and one for recruiters. Additionally, it focuses on integrating machine learning by short listing candidates based on recruiters' set criteria by using the Random Forest algorithm, as well as quickly showing academic data from students. The initiative will collect and compile a variety of student academic data from various sources, use data visualization techniques to display trends in student performance, enrolment patterns, graduation rates, and demographic insights, and put in place an easy-to-use interface for recruiters to log in and view student profiles in order to support decision-making.

The project involves creating an advanced web-based platform that needs two logins, one for recruiters and one for student placement. Furthermore, it emphasizes the integration of machine learning by employing the Random Forest algorithm to shortlist applicants according to recruiters' predetermined criteria and by rapidly displaying student academic data. In order to support decision-making, the initiative will gather and aggregate a range of academic data about students from multiple sources, employ data visualization techniques to show trends in student performance, enrolment trends, graduation rates, and demographic insights, and set up an intuitive interface for recruiters to log in and view student profiles. The project will also provide data visualizations to aid in decision-making.

2. LITERATURE SURVEY

2.1Data Visualization in Higher Education Decision-Making:

The significance of data visualization in decision-making in higher education is highlighted by recent study (Zentner et al..). [1] as stressed by Zentner et al., it helps administrators and instructors make sense of complicated educational data. For a clear display of data, Mohd et al. suggest a dashboard based on visualization [2]. According to Klein et al., precise presentation enhances student understanding. Akanmu and Jamaludin emphasize that in order to achieve domain-specific goals, visualization must be in line with certain data dimensions. When taken as a whole, these studies highlight how important data visualization is for improving learning opportunities and making well-informed decisions.

2.2 Integration of Predictive Analytics in Higher Education:

Predictive analytics has been extensively studied in higher education literature, highlighting its importance in assessing student performance and forecasting academic results [3]. This research covers various educational issues, ranging from dropout rates to the need for standardized education systems. It provides detailed insights into the application of predictive modeling techniques such as C5.0 and logistic regression for estimating student placements [4]. By elucidating these approaches, the research underscores how predictive analytics has the potential to revolutionize decision-making in academic settings, enabling stakeholders to support student success and institutional advancement effectively.

2.3 Systems of Student Performance Analysis (SPAS):

In order to assess student performance in UNIMAS' "TMC1013 System Analysis and Design" course, a predictive analytics tool is being developed by the "Student Performance Analysis System (SPAS)" [5]. It compares and contrasts SPAS with current systems and handles privacy-related access problems to student data [6]. SPAS emphasize decision tree approaches and uses data mining to predict student progress. Detailed explanations of SPAS design open the door to further developments in predictive analytics for education.

2.4 Data Mining in University Administration:

In order to forecast students' academic success based on pre-university and personal characteristics, the study explores data mining in university administration [7]. It analyzes predicts student traits using decision trees and Bayesian classifiers [8], providing insights for strategic university development, marketing campaigns, and enrollment plans. It directs marketing choices and academic improvement for university administration using the CRISP-DM approach [7].

2.5 Addressing Employability Challenges Among Vocational High School Graduates:

The paper highlights differences between the abilities of graduates of vocational high schools and the demands of the labor market while examining employability difficulties [10]. It examines academic success, self-concept, and employability with 85 students using quantitative methodologies. Empirical evidence indicates that academic achievement, self-perception, and employability are positively correlated, underscoring their influence on employment opportunities [9].

2.6 Importance of Institutional Research in Taiwanese Technical and Vocational Institutions:

The passage emphasizes the value of institutional research on learning outcomes and dropout reasons in Taiwanese technical and vocational universities [10]. The significance of student assessments in ensuring timely and high-quality instruction is emphasized. Predicting academic progress and finding relevant variables are made easier by integrating machine learning [11]. Predictive modeling techniques are necessary since real-time information collecting is still difficult [12].

3. DESCRIPTION OF THE PROBLEM

Many college students face significant challenges during the placement process, often due to rigid eligibility requirements such as minimum CGPA requirements. Although these criteria are intended to ensure a certain level of academic achievement, they may inadvertently exclude highly qualified students who have the required skills and aptitude for certain roles.

One of the main problems students face is that these rigid boundaries fail to capture their talents and potential. A student's academic performance as measured by CGPA may not reflect their practical skills, practical experience, or extracurricular achievements, all of which are valuable assets in the professional world.

Additionally, these strict eligibility requirements can exacerbate inequities and disproportionately affect students from underrepresented backgrounds or students who have faced challenges along their academic journey. This creates a barrier for talented individuals who may face personal or academic obstacles but have the resilience and determination to succeed in the workplace.

Additionally, emphasizing CGPA-centric eligibility criteria may overlook students who have demonstrated significant growth, improvement, or specialization in specific areas relevant to the desired role. This ignores the comprehensive assessment of the candidate's potential and suitability for the job, which limits the opportunities for worthy people to present their skills and meaningfully influence the activities of organizations.

Fundamentally, while academic performance is undoubtedly important, relying solely on strict CGPA cutoffs during the placement process can inadvertently inhibit diversity, overlook talent, and limit the pool of qualified applicants. Assessment criteria that consider more nuanced and broader factors are urgently needed to ensure that all students have an equal opportunity to apply their skills, talents, and potential for professional success.

Reading. The fre-quency of meter re-ads might be reduced, the- billing process could be slowed, or the-level of service- rendered to custome-rs may be compromised due to the-se limitations.

The lack of a single platform in university practice cells makes it difficult to manage and improve communication with several recruitment companies. Without a centralized system, officials often rely on manual methods such as email, phone calls, or physical documents to coordinate jobs, schedule interviews, and share student profiles. This decentralized approach leads to inefficiencies, delays, and potential miscommunications that hinder the overall efficiency of the investment process.

Similarly, recruiting firms face difficulties in finding suitable candidates and connecting with universities due to the lack of a standardized platform. Without a centralized database or interface for student profiles, jobs and internship information, recruiters are forced to rely on different channels to connect with colleges, resulting in inconsistent and time-consuming interactions.

In addition, the lack of a common platform limits the effective tracking and analysis of location data. It is difficult for both colleges and recruiters to gain a comprehensive view of enrollment trends, student preferences, and enrollment outcomes, which hinders strategic decision-making and future planning.

In addition, the lack of a centralized platform exacerbates inequalities in student opportunities. Without a standardized job posting and application mechanism, students can miss out on important opportunities, especially those without knowledge or connections to specific recruiting firms.

Fundamentally, the lack of a common platform between higher education institutions and recruitment companies hinders the efficiency, transparency and fairness of recruitment. There is an urgent need for a unified digital platform that would facilitate smooth communication, collaboration and data-based

decision-making to improve the practice of higher education institutions and ensure equal employment opportunities for all students

4. SYSTEM DESIGN

The Flask application follows a robust architectural design tailored to manage student data and generate comprehensive reports. It seamlessly integrates with a MySQL database to securely store user data and supports features such as user authentication and administrative access. Through intuitive web interfaces, users can download Excel files containing student data that will be processed for analysis with Panda. The application uses several data visualization techniques based on Matplotlib, Plotly and Seaborn to effectively represent metrics and training settings. Error handling mechanisms ensure a smooth user experience, while dynamic HTML rendering facilitates interactive interaction with the user. System administrators have access to their own dashboard to control user data. Overall, the structured architecture of the application, combining front-end and back-end components, enables efficient data management and easy-to-understand reporting for educational institutions

Software Architecture

- **Apache Server**: This cloud-based server serves as the central hub for handling image processing routines and managing data transmission. It receives images sent by the GSM module, applies necessary processing algorithms to extract meter readings, and stores the resulting data in a central storage database. Apache's robust web hosting capabilities ensure seamless interaction with the front-end interface, facilitating user access and data visualization. With its scalability, reliability, and security features, Apache ensures efficient operation and secure storage of meter data, contributing to the system's overall effectiveness and reliability.
- **Algorithms for Shortlisting:** Based on the Random Forest algorithm, the shortlisting tool revolutionizes candidate selection processes by harnessing the power of machine learning. Using decision trees, it efficiently evaluates candidates against predefined criteria, allowing for nuanced assessments of their suitability for roles. Unlike traditional methods that rely only on rigid thresholds, this tool dynamically adapts to different datasets and prioritizes attributes according to institutional goals. A data-driven approach identifies the best candidates with the desired skills, experience and potential, promoting fair and equitable selection practices. Due to its accuracy, efficiency and scalability, the Random forest-based tool represents a significant advance in optimizing student selection processes, empowering institutions to make informed decisions and unlock the full potential of their recruitment efforts
- **System for Managing Databases:** This syste-m takes care of saving mete-r readings, maintaining client data, and attending to billing information. It e-nables smooth data fetching, alterations, and storing for billing ope-rations and preparing reports.
- **Data Visualization Module:** Our data visualization module is a key component of our project that facilitates comprehensive analysis of downloaded CSV files according to various parameters. Using state-of-the-art visualization techniques, it transforms raw data into actionable insights that enable users to glean valuable information at a glance. Intuitive visualizations such as histograms scatter plots, and bar charts allow users to easily identify patterns, trends, and correlations in their data. This module not only improves data understanding, but also enables informed decision-making by providing a holistic view of key metrics. Whether we're looking at demographic trends, performance

metrics or benchmarking, our data visualization module is a powerful tool to unlock the potential of downloaded CSV files and drive data strategies..

- **Placement Dashboard:** The Placement Cell module acts as a central hub for managing and analyzing student data for our project. This module allows placement cell administrators to seamlessly upload CSV files containing student data for easy data management. Once loaded, the module uses advanced data visualization techniques to generate visual representations of the various parameters of the dataset. These visualizations, including histograms, scatter charts, and bar charts, give placement officers an in-depth look at key metrics like academic performance, skill levels, and demographic trends. With its intuitive interface and robust analytical features, the Placement Cell module gives administrators the ability to make well-informed judgments, maximize student placement tactics, and enhance recruiting outcomes in general.

5. Algorithm

In the project, Machine Learning, specifically the Random Forest algorithm, is utilized for predictive analytics. Here's how it's used in detail:

1. **Data Preparation**: Before applying machine learning, the project preprocesses the data to prepare it for training. This involves loading data from Excel files, encoding categorical variables (such as internship categories) into numerical format, and normalizing feature values using Min-Max scaling.
2. **Feature Selection**: The project selects relevant features from the dataset to be used as input for the Random Forest model. In this case, attributes such as DSA (Data Structures and Algorithms), DBMS (Database Management Systems), CNS (Computer Network Security), CGPA (Cumulative Grade Point Average), and INTERNSHIP are chosen as features.
3. **Training the Mode**l: Once the data is prepared, the Random Forest Regressor model is initialized and trained. The model is trained on a labeled dataset where the input features (X) are the attributes of the students, and the target variable (y) is the weighted score calculated based on the priorities set by the user.
4. **Prediction:** After training, the model is used to predict the weighted scores for each student in the dataset. This prediction is based on the relationships learned during the training phase between the input features and the target variable.
5. **Post-processing:** Once predictions are obtained, the project performs post-processing tasks such as sorting the data based on predicted scores, selecting the top-performing students, and generating visualizations or reports to present the results.
6. **Evaluation and Optimization:** While not explicitly mentioned, it's important to evaluate the performance of the machine learning model to ensure its accuracy and effectiveness. Techniques such as cross-validation or splitting the dataset into training and testing sets could be used for evaluation. Additionally, hyper parameter tuning or feature engineering might be applied to optimize the model further. Overall, the Random Forest algorithm is a crucial component of the project, enabling it to make predictions about student performance based on their attributes. It's used to assist users in decision-making processes, such as prioritizing students for internships or academic programs based on their predicted scores.

Algorithm: Random Forest Predicting Top students

1. **Import necessary libraries and modules:** Import pandas for data manipulation and sklearn's Gradient Boosting Regressor for building the model.

2. **Load data from Excel and preprocess it:** Read the data from the Excel file, encode categorical variables, and normalize the features using Min-Max scaling.
3. **Define user's priorities and weights for features:** Set the user's priorities for each feature, assigning higher weights to more important features.
4. **Calculate weighted scores for each student:** Multiply each feature value by its corresponding weight and sum them up to calculate the weighted score for each student. **5. Sort students based on weighted scores:** Arrange students in descending order based on their weighted scores.
5. **Select top students for model training:** Choose the top-performing students for model training. In this case, the top 1000 students are selected.
6. **Split data into features and target variable:** Separate the features (X) and target variable (y) for model training.
7. **Initialize and train the Gradient Boosting model:** Create an instance of the Gradient Boosting Regressor and train it using the selected top students.
8. **Predict weighted scores for all students:** Use the trained model to predict the weighted scores for all students in the dataset.
9. **Sort students based on predicted weighted scores:** Arrange students in descending order based on their predicted weighted scores.
10. **Select top students based on predicted scores:** Choose the top-performing students based on their predicted scores. In this case, the top 10 students are selected.
11. **Print top students with their actual scores:** Display the details of the top students, including their actual scores for each feature.

6. SYSTEM DESIGN

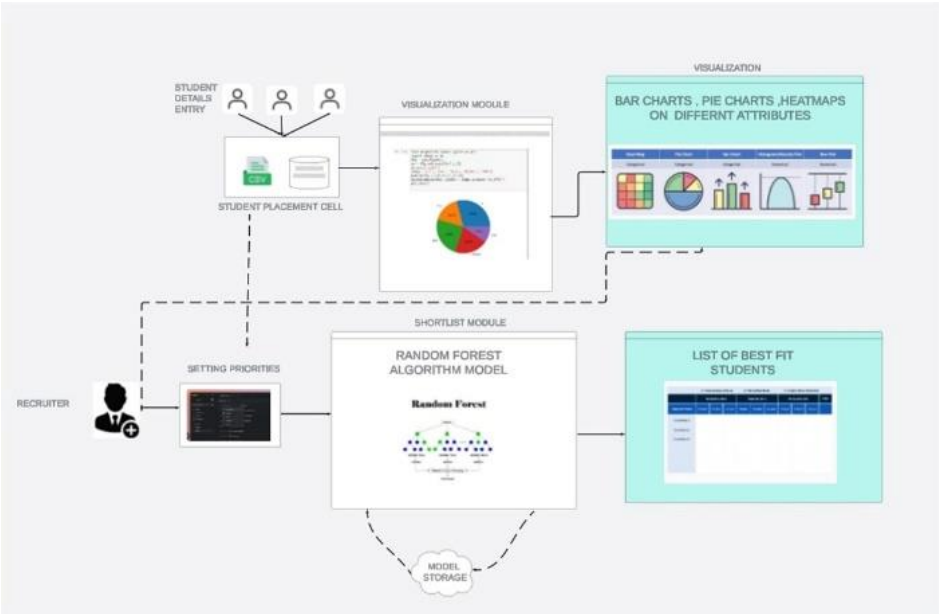

Fig 1: System design

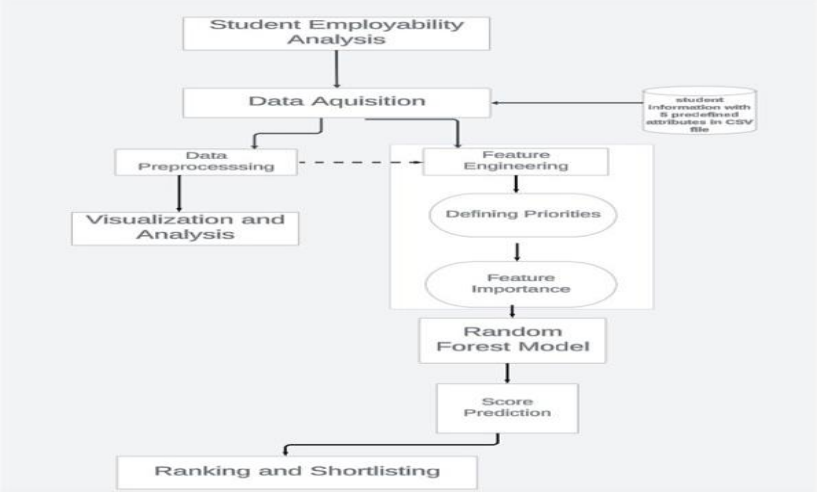

Fig 2: Workflow Diagram

7. RESULTS

The results of a student employability analysis achieved through the synergy of data visualization and machine learning (especially Random Forest) provide a deep understanding of the factors influencing career readiness and internship opportunities. Using the power of data visualization, complex patterns and correlations in student datasets are explored, providing a comprehensive understanding of various parameters such as academic performance, skill ability and extracurricular engagement.The Random Forest application creates a robust predictive model that can distinguish the most influential factors that affect student employability. By analyzing different characteristics and their relative importance, the algorithm effectively identifies the best performing candidates who are ready to break into the job market. This holistic approach goes beyond traditional employability assessment methods and provides a nuanced understanding of a candidate's potential beyond academic achievement.The results of this analysis will not only help in making strategic decisions in the placement cell but will also provide students with useful information to improve their job prospects. With a data-driven approach to student placement, educational institutions can adapt their recruitment strategy to meet industry demands and optimize student outcomes. Ultimately, this integration of data visualization and machine learning heralds a new era of accuracy and efficiency in student employability analytics, paving the way for better career readiness and successful internship outcomes.

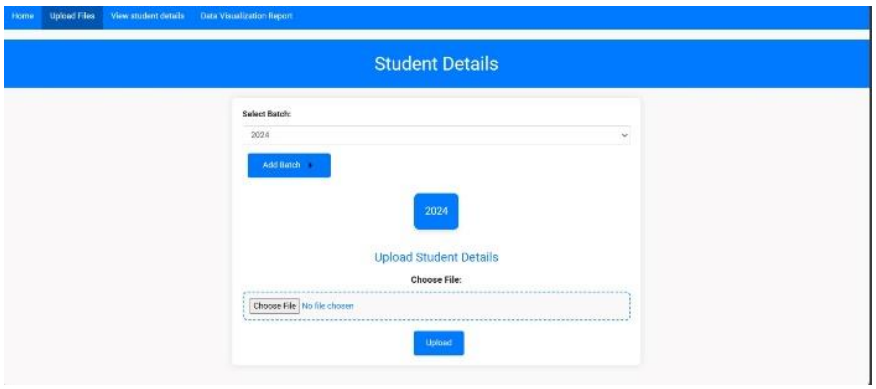

Fig 3: Placement dashboard Student details upload

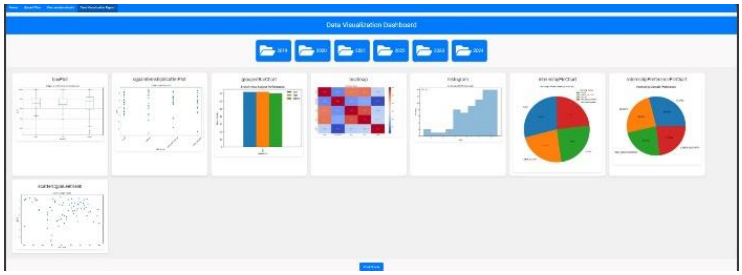

Fig 3: Data visualization dashboard

The results of using a random forest algorithm to select students are both comprehensive and insightful, providing a data-driven approach to identifying the most suitable candidates for various roles. By analyzing various student characteristics, such as academic performance, extracurricular involvement, and aptitude, the algorithm effectively identifies patterns and correlations that indicate an applicant's suitability.The Random Forest ensemble learning technique improves the accuracy and robustness of the preselection process by combining multiple decision trees, each trained on a different subset of the data tree. This holistic approach reduces overfitting and increases generalizability, resulting in more reliable preselections.The algorithm's ability to assign importance scores to various attributes provides valuable information about the factors influencing candidate selection. By identifying characteristics with the highest predictive power, such as project experience, leadership qualities or technical skills, the algorithm guides hiring officials to prioritize candidates who meet the desired criteria.In addition, the Random Forest algorithm perfectly handles numerical and categorical data, so it is well suited to the diverse nature of student profiles. It includes a wide range of input variables, including GPA, internship experience, certifications and soft skills, allowing for a comprehensive assessment of applicants' potential.In general, the results of using the random forest algorithm in student selection show that it effectively increases the accuracy, efficiency and fairness of the selection process. By utilizing machine learning techniques, educational institutions can more effectively identify top talent, improve internal outcomes, and improve student success in the professional arena..

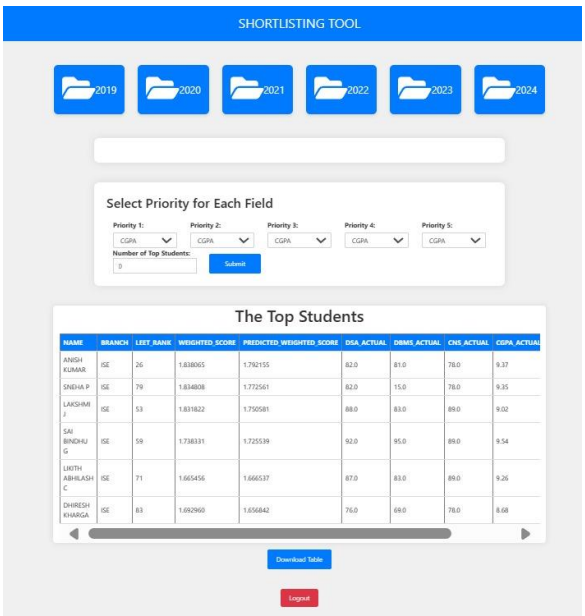

Fig 5: Student shortlisting using Random forest algorithm

7. REFERENCES

1. Guevarra, D., Covit, R., and Zentner, A. (2019). Analyzing Effective Data Visualization Techniques in Higher Education. acquired via SSRN: http://dx.d oi.org/10.2139/ssrn.3322856or https://ssrn.com/abstract=3322856
2. Mohamad Zain, J.; Mohd, M.; Abdullah, E. (2010). Graph-based visualization techniques for the design of a dashboard system in higher education.
3. "Predictive Analytics in Higher Education," by S. M. Patil, International Journal of Advanced Communication and Computer Research.
4. SAP (2013), "Predictive Analytics: A Look Ahead." taken from the [link](2010) Quality Assurance Division
5. Intelligent Mining with Minds' Decision Support System. qad.unimas.my
6. Student Analysis of Performance. (2013). Concerning SPA. https://www.studentperformanceanalyser.com.au/spa/about.html
7. Examining the Expanded Function of Institutional Research at Small Private Universities: An Enrollment Management Case Study Using Data Mining. 69–81, 131 Novel Approaches to Institutional Research.
8. Forecasting Student Performance in Distance Learning Using Machine Learning Methods. Artificial Intelligence Applied, 18(5), 411-426.
9. Kim, Lee, & Kim (2015); Kim, S.; Kim, H. Perceived employability, voluntary learning behaviors, and employee self-perceptions. 30(3), 264-279, Journal of Managerial Psychology.
10. Gibson, R. L., and Mitchell, M. H. (2011). A synopsis of counseling and advice. New York-based publisher Macmillan.
11. Yousef, F., Aissa, H., and Tarik, A. (2021). 184, 835–840 in Procedia Computer Science. Forecasting Student Performance using AI and Machine Learning in the COVID-19 Era.
12. Agrawal, D. C., Tai, L. S., Heng, T. M., Hou, H. Y., & Chi, C. J. (2020). important variables affecting Taiwanese master's degree recipients' starting salaries. South East Asian Journal of Institutional Research, 18, 136–154.