e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

MACHINE LEARNING-BASED PLACEMENT PREDICTION SYSTEM

Dnyaneshwar Kudande*1, Om Darkunde*2, Omkar Koli*3, Vedika Bhuvad*4, Anuja Pawar*5

*1Professor, Department Of Computer Engineering, SRTTC, Khamshet, Pune, India.

*2,3,4,5Student, Department Of Computer Engineering, SRTTC, Khamshet, Pune, India.

DOI : https://www.doi.org/10.56726/IRJMETS64224

ABSTRACT

The rapid evolution of today's job market challenges educational institutions to prepare students for successful employment. To address this, we introduce a Campus Placement Prediction System that leverages machine learning to forecast placement outcomes, enabling institutions to make data-driven decisions, identify at-risk students, and enhance employability support. By analyzing factors like academic performance, skills, internships, and extracurricular activities, the system provides actionable insights into key employability factors. Using algorithms such as logistic regression, decision trees, and random forests, it predicts outcomes with over 80% accuracy, with random forests performing best. The anonymized dataset, encompassing grades, skills, and experiences, was preprocessed for consistency, and critical predictors like academic scores, internships, and technical skills were identified through feature selection and correlation analysis. This system empowers institutions to optimize curricula, offer tailored support, and align student development with industry demands.

Keywords: Predictive Analytics, Data Preprocessing, Feature Selection, Logistic Regression, Decision Trees, Random Forests, Supervised Learning, Classification Models, Training Data, Model Evaluation.

I. INTRODUCTION

The Campus Placement Prediction project utilizes machine learning to predict a student's chances of getting placed by analyzing factors such as academic performance, skills, certifications, and extracurricular activities. This system provides actionable insights to help students identify areas for improvement and assists placement coordinators in designing targeted training programs. By leveraging advanced algorithms, the project bridges the gap between academia and industry, empowering students to make a smooth transition into professional careers while enhancing the overall effectiveness of campus recruitment processes.

II. LITERATURE REVIEW

1. **Paper Name**: A Comparative Study on Machine Learning Algorithms for Predicting the Placement Information of Under Graduate Students

Author: Tadi Aravind , Bhimavarapu Sasidhar Reddy , Sai Avinash, Jeyakumar G

Abstract: As Machine Learning (ML) algorithms become increasingly popular for solving challenging and interesting real-world prediction problems, there has been a growing interest in the student community to learn the principles of ML and its various algorithms. This includes implementing commonly known machine learning algorithms and testing them by solving simple prediction problems relevant to the student community in the educational system. In this context, this paper aims to solve the student placement prediction problem using various machine learning models: linear regression, k- nearest neighbor regression, decision tree regression, XGBoost regression, gradient boost regression, light GBM regression, and random forest classifier models. The work is carried out in two phases: Phase 1 involves a simple dataset, and Phase 2 involves an extended dataset with additional features about the students. This research presents a comparative performance analysis of these seven models, implemented on both datasets. The performance metrics considered in this study include prediction accuracy and root mean square error

2. **Paper Name**: Markov Prediction Model for Host Load Detection and VM Placement in Live Migration"

Author: Suhib Bani Melhem, Anjali Agarwal, Nishith Goel, and Marzia Zaman

Abstract: The design of good host overload/underload detection and VM placement algorithms play a vital role in assuring the smoothness of VM live migration. The presence of the dynamic environment that leads to a changing load on the VMs motivates us to propose a Markov prediction model to forecast the future load state of the host. We propose a host load detection algorithm to find the future overutilized/underutilized hosts state to avoid immediate VMs migration. Moreover, we propose a VM placement algorithm determine the set of candidates hosts to receive the migrated VMs in a way to reduce their VM migrations in near future. We evaluate our proposed algorithms through CloudSim simulation on different types of PlanetLab real and random workloads. The experimental results show that our proposed algorithms have a significant reduction in terms of SLA violation, number of VM migrations, and other metrics than the other competitive algorithms. give it in normal form

3. **Paper Name**: Placement Prediction using Various Machine Learning Models and their Efficiency Comparison

Author: Irene Treesa Jose ,Daibin Raju ,Jeebu Abraham Aniyankunju, Joel James,Mereen Thomas Vadakkel

Abstract: To design a placement predictor, we aimed to calculate the likelihood of a student being placed in a company, based on the company's criteria. The predictor takes various parameters to assess the student's skill level, with some data points coming from the university level and others from tests conducted within the placement management system. By combining these inputs, the goal was to accurately predict whether a student will be placed. To train the predictor, we used data from past students. The challenge, however, was to identify the most suitable classification algorithm that could yield maximum accuracy for our dataset. Different algorithms perform differently depending on the nature of the problem and the dataset they handle. Therefore, we selected four algorithms—KNN, SVM, Logistic Regression, and Random Forest—and compared their accuracy levels in the context of our data. The outcome of this comparison would guide us in choosing the best algorithm for implementing the placement predictor in the placement management system. To evaluate the performance of each algorithm, we trained them on the dataset and tested them against separate test data, calculating accuracy based on the True Positive, True Negative, False Positive, and False Negative values. These values allowed us to compute accuracy using the standard accuracy equation..

4. **Paper Name**: PLACEMENT PREDICTION USING MACHINE LEARNING

Author: Dr. Kaveri Kari , Ms. Pranali Shinde , Ms. Nikita Deore, Ms. Shweta Narkhede ,Mr. Piyush Ekade

Abstract: A placement predictor is being designed to calculate the likelihood of a student being placed in a company, based on the company's criteria. The predictor utilizes various parameters to assess the student's skill level. Some of these parameters are derived from university-level data, while others come from tests conducted within the placement management system itself. By combining these data points, the predictor aims to accurately determine whether a student will be placed in a company. To train the predictor, data from past students are used, allowing the system to learn patterns and make more accurate predictions.

5. **Paper Name**: CAMPUS PLACEMENT PREDICTION SYSTEM USING MACHINE LEARNING

Author: Prof. S.S. Kashid, Ashish Badgujar, Vishal Khairnar,Anurag Sagane, Nishant Ahire **Abstract:** A campus placement prediction system is developed to calculate the possibility of a student securing a job in a company through campus placements. The model takes various parameters that provide insights into the student's skill level. Some data are collected from the college level, such as academic performance, CGPA, pointers, and attendance, while other data are obtained from tests conducted within the placement management system. By combining these data points, the model aims to accurately predict whether a student will be placed in a company. Additionally, data from past years' students are used to train the model, utilizing educational data mining to gather real data specific to that college. This helps the machine learning model make more effective predictions tailored to the particular institution.

The purpose of this placement prediction system is to help students improve their academic performance and develop soft skills that will increase their chances of getting placed. The system also aids the faculty by providing valuable insights that can guide training and enhance the placement department's efforts. This analysis will help the institution understand student performance and ensure it meets the needs of recruiters.

For this supervised machine learning approach, logistic regression is particularly effective. Designing the logistic model plays a crucial role in obtaining accurate predictions. This process includes selecting the training data tuples and their known outcomes, often referred to as real data, to train the model.

6. **Paper Name**: Student Placement Prediction Using Machine Learning

Author: P. Archana, Dhathirika Pravallika, Pandila Sindhu Priya3, Sarikonda Sushmitha Sripada,Amitha

Abstract: The placement of scholars is one of the most important activities in academic institutions, as admission and the reputation of institutions often depend on placement rates. The main objective of this paper is to analyze the historical data of previous years' students and predict the placement possibilities of current students, thereby helping to increase the placement percentage of the institution. This system presents a recommendation model that predicts whether a current student will be placed or not. If the student is placed, the system also predicts the company based on the data of previously placed students. For this, two different machine learning classification algorithms, namely the Naive Bayes Classifier and the K-Nearest Neighbors (KNN) algorithm, are used to predict the results. The efficiency of these algorithms is then compared based on the dataset. This model helps the placement cell within the institution to identify potential students, allowing them to focus on and improve both technical and social skills.

7. **Paper Name:** STUDENTS PLACEMENT PREDICTION USING MACHINE LEARNING ALGORITHMS

Author: Dr. Kajal Rai

Abstract: Placements hold immense importance for students and academic institutions. They help students build a strong foundation for their professional careers, while a good placement record gives a competitive edge to a college or university in the education sector. Machine learning is a method of statistical analysis that automates the construction of analytical models. This paper focuses on a machine that predicts whether a student will be placed or not, based on the student's qualifications, historical data, and experience. The predictor uses three machine learning algorithms—Decision Tree, Naïve Bayes, and Random Forest—to forecast the student's placement. An evaluation of these algorithms is then conducted based on the accuracy achieved.

III. MOTIVATION

The motivation for developing a Campus Placement Prediction System arises from challenges and opportunities within education and employment. With the job market becoming increasingly competitive, graduates face immense pressure to secure placements, highlighting the need for institutions to adapt by offering timely support through career counseling, resume-building workshops, and mock interviews. Building strong connections with industry professionals and providing skill development aligned with market demands can significantly enhance employability. Additionally, leveraging data analytics enables institutions to make informed decisions by analyzing student performance, alumni career trends, and skill gaps. This data-driven approach allows career services to tailor programs, allocate resources effectively, and refine strategies to address the evolving needs of students and employers, ultimately improving placement outcomes.

IV. OBJECTIVE

- **Accurate Prediction**: Develop a machine learning model to predict a student's likelihood of securing placement based on academic performance, skills, and extracurricular activities.
- **Personalized Insights**: Provide actionable insights to students, helping them identify strengths and areas for improvement to enhance their employability.
- **Data-Driven Decisions**: Empower institutions to utilize predictive analytics for better decision- making in designing training programs and allocating resources effectively.
- **Enhanced Recruitment Process**: Assist placement coordinators in streamlining recruitment efforts by matching students with suitable job roles based on their profiles.
- **Improved Industry-Academia Collaboration**: Facilitate better engagement between institutions and industries to align student preparation with real-world job market demands.
- **Scalability and Adaptability**: Build a system that can adapt to different educational institutions and industries, ensuring flexibility and scalability.

V. PROBLEM STATEMENT

The rising competition in the job market and the increasing number of graduates seeking employment have created challenges for educational institutions in effectively preparing students for placements. Traditional

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

methods of assessing student readiness and matching them with job opportunities often fail to utilize data effectively, leading to inefficient resource allocation, overlooked skill gaps, and reduced employability. This project aims to develop a machine learning-based Campus Placement Prediction System that analyzes student profiles, including academic performance, skills, and extracurricular activities, to accurately predict placement outcomes. The system seeks to provide actionable insights for students, enable data-driven decision-making for institutions, and streamline the placement process by aligning education with industry demands.

VI. METHODOLOGY

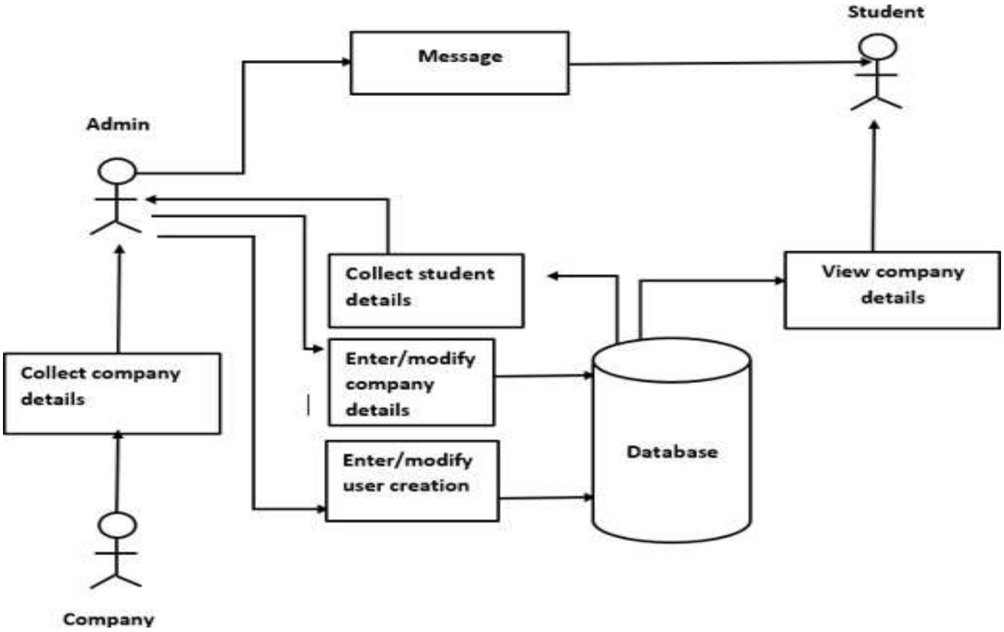

Fig 1: System Architecture

Hardware Requirement

- System: Intel I3 8th Gen 2.9 Ghz
- Hard Disk: 256 GB (Min)
- Monitor: 15 VGA Color
- Mouse: Logitech.
- Ram: 4 GB (Min)

Software Requirement

- Operating System: Windows 8 onwards.
- Python Version: Python 3.10.8 & Above.
- Coding Language: Python
- Back-end(DB) : Django
- Web Server: Apache Tomcat/XAMPP.
- IDE: Eclipse Oxygen

VII. PROJECT SCHEDULE

Project task set:

Major Tasks in the Project stages are:

Task 1: Correctness

Task 2: Availability

Task 3: Integrity

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

Task network:

Figure 2: Task Network

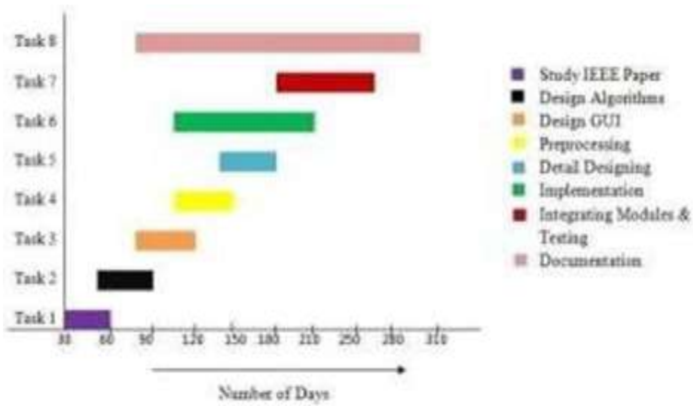

Figure 3: Gantt chart of project

VIII. DATA FLOW DIAGRAMS

- The DFD is also called a bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system.
- The data flow diagram (DFD) is one of the most important modeling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.
- DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output.
- DFD is also known as bubble chart. A DFD may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail.

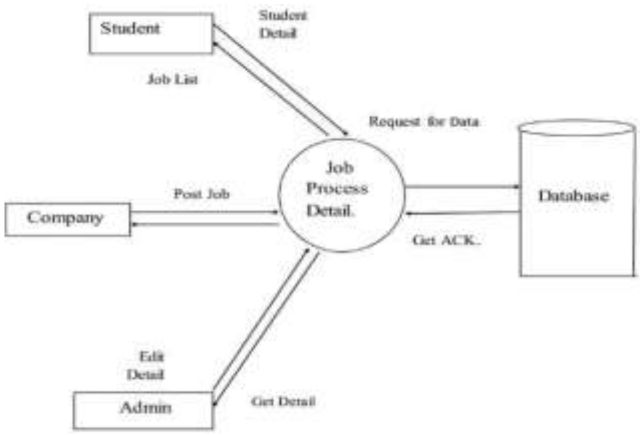

Figure 4:

Level 0 Data Flow Diagram

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

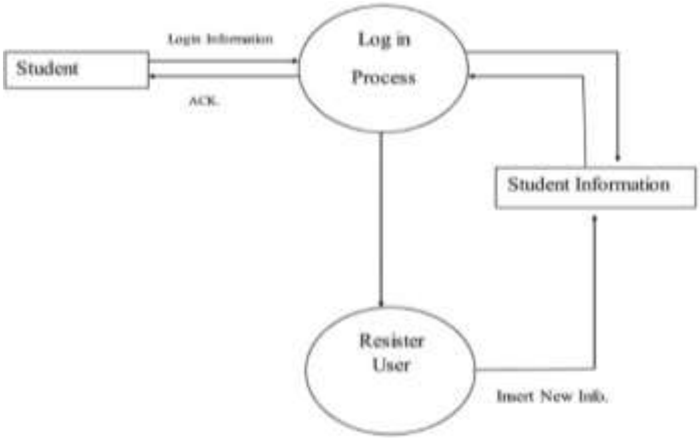

Figure 5:

Level 1 Data Flow Diagram

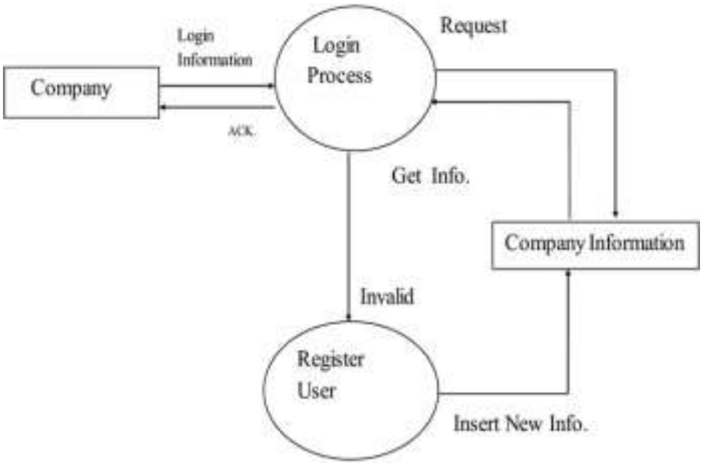

Figure 6:

Level 2 Data Flow Diagram

IX. OVERVIEW OF PROJECT MODULES

1. User Registration and Profile Creation

Students register on the platform by providing details such as academic records, skill certifications, extracurricular activities, and other relevant information. Secure authentication methods ensure that only authorized users can access the system.

2. Data Collection and Preprocessing

The system collects data from registered students, and preprocesses it by handling missing values, normalizing data, and encoding categorical variables, ensuring the dataset is ready for analysis and prediction.

3. Feature Engineering

Relevant features, such as GPA, technical skills, internships, and extracurricular activities, are extracted and engineered to optimize the performance of the machine learning model.

4. Machine Learning Model Training

The processed data is used to train machine learning models, such as logistic regression, decision trees, or ensemble methods, to predict the placement outcomes of students.

5. Placement Prediction

Once trained, the model predicts whether a student is likely to secure placement and provides insights into the factors influencing the prediction.

6. Performance Evaluation and Optimization

The system evaluates model performance using metrics like accuracy, precision, recall, and F1- score.

Continuous optimization is performed to enhance prediction reliability.

7. Dashboard for Insights

Students and placement coordinators gain access to an intuitive dashboard that displays predictions, skill gap analysis, and actionable recommendations for improving employability.

8. Feedback and Recommendations

Based on predictions and analysis, the system generates personalized feedback for students, suggesting skill development programs, certifications, or training to enhance placement prospects.

9. Scalability and Adaptability

The system is designed to handle large datasets and adapt to different institutions by incorporating institution-specific data and requirements, ensuring scalability and usability across diverse environments.

X. PERFORMANCE EVALUATION

In the Campus Placement Prediction System, the time taken to predict placement outcomes depends on the complexity of the machine learning model and the number of features being processed. As the number of student records increases, the system's processing time also rises due to the increased computation required for feature extraction, model inference, and result generation. For individual predictions, the execution times for core functions are as follows: 12.5 seconds for data preprocessing,

18.3 seconds for feature engineering, 5.8 seconds for model inference, and 8.2 seconds for generating personalized recommendations. However, when multiple users simultaneously access the system, processing times may increase due to shared computational resources.

Throughput: Throughput in this system refers to the number of predictions generated per second, influenced by the model's computational efficiency and server performance. Using tools like Apache JMeter, simulated user loads ranging from 50 to 500 were analyzed. The results indicate that as user requests increase, the system's throughput initially improves due to parallel processing but eventually plateaus as the computational resources reach their limits. This behavior highlights the need for scalable architecture to handle high user loads efficiently.

XI. ALGORITHM DETAILS

11.1 Logistic Regression Algorithm:

Input: Training dataset with features (X) and labels (Y).

Output: Predicted probabilities or binary classification (placed/not placed).

1. Initialize weights and bias values.
2. Compute the linear combination of features: Z=XW+bZ = XW + bZ=XW+b.
3. Apply the sigmoid activation function: σ(Z)=11+e−Z\sigma(Z) = \frac{1}{1 + e^{- Z}}σ(Z)=1+e−Z1 to map values to probabilities.
4. Determine the class labels based on a threshold (e.g., 0.5):

- If σ(Z)≥0.5\sigma(Z) \geq 0.5σ(Z)≥0.5, predict placement.
- Otherwise, predict no placement.

5. Optimize weights using gradient descent to minimize the logistic loss function:

- L=−1n∑i=1n[yilog⁡(y^i)+(1−yi)log⁡(1−y^i)]L = -\frac{1}{n} \sum_{i=1}^{n} [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]L=−n1∑i=1n[yilog(y^i)+(1−yi)log(1−y^i)].

6. Iterate until the model converges or a specified number of iterations is reached.

11.2 Random Forest Algorithm:

1. Input: Training dataset with features (X) and labels (Y). Output: Predicted class (placed/not placed).
2. Randomly sample subsets of the training data with replacement (bootstrap sampling).
3. For each subset, construct a decision tree:
4. Randomly select a subset of features for splitting at each node.

5. Split nodes based on criteria like Gini Impurity or Information Gain.
6. Continue until the tree reaches a maximum depth or cannot be split further.
7. Repeat this process to create multiple decision trees.
8. For prediction:
9. Each tree independently predicts the class for the given input.
10. Aggregate the predictions (majority vote for classification).
11. Return the final prediction based on the aggregated results from all trees.

Both algorithms were chosen for their complementary strengths: logistic regression provides interpretability and simplicity, while random forests offer high accuracy and robustness for complex datasets with nonlinear relationships. Together, they enhance the system's predictive performance.

XII. CONCLUSION

The proposed Campus Placement Prediction System provides an effective and data-driven approach to predict student placement outcomes. By leveraging machine learning models such as Logistic Regression and Random Forest, the system ensures accurate predictions while offering actionable insights for students and institutions. This framework empowers institutions to allocate resources efficiently and tailor training programs to improve employability. The implemented algorithms are computationally efficient and capable of handling diverse datasets, ensuring scalability and reliability. This system bridges the gap between education and industry by aligning student preparation with job market demands, ultimately enhancing the overall placement process.

XIII. REFERENCES

[1] Ishizue, R., Sakamoto, K., Washizaki, H., & Fukazawa, Y. (2018). "Student placement and skill ranking predictors for programming classes using class attitude, psychological scales, and code metrics," Research and Practice in Technology Enhanced Learning, 3(7).

[2] Sreenath, K., & Jeyakumar, G. (2016). "Evolutionary Algorithm Based Rule(s) Generation for Personalized Courseware Construction in Educational Data Mining," Proceedings of the IEEE International Conference on Computational Intelligence and Computing Research, 609-615.

[3] Sucharitha, V., Reshma Reddy, R., & Jeyakumar, G. (2014). "Learning Interest Curve generation using PECS Agent based Model," Proceedings of International Conference on Communication and Computing, 152-160.

[4] Jeyakumar, G., Raaghul, K., Pavithra, N., Kavya, D., & Aswathi, B. (2016). "A Prototype for Student Learning Style Modelling Using Felder-Silverman Learning Style Model," Proceedings of International Conference on Smart Structures & Systems, 178-182.

[5] Roshini, K., Bavya, B., & Jeyakumar, G. (2015). "RoBaJe – A Simulated Computational Model for Human Memory to Illustrate Encoding and Decoding of Information," International Journal of Applied Engineering Research, 9(24), 26957-26970.

[6] Deivanayagam, K., Gayathiri, D., Manikandan, A., Raghul Karthik K R., & Jeyakumar, G. (2014). "Learning to Identify Bad Coding Practice," International Journal of Applied Engineering Research, 9(20), 6747-6755.

[7] Pal, A. K., & Pal, S. (2013). "Classification Model of Prediction for Placement of Student," International Journal of Modern Education and Computer Science, 11, 49-57.

[8] Giri, A., Vignesh, M. V., Bagavath, V., Pruthvi, B., & Dubey, N. (2016). "A Placement Prediction System using K-nearest neighbors' classifier," Proceedings of the Second International Conference on Cognitive Computing and Information Processing.

[9] Thangavel, S. K., Bharathi, D. P., & Shankar, A. (2017). "Student placement analyzer: A recommendation system using machine learning," Proceedings of International Conference on Advanced Computing and Communication Systems.

[10] Prasad, A., Harinath, S., Suma, H. S., Suraksha, A., & Mathew, T. (2019). "Student Placement Prediction Using Machine Learning," International Research Journal of Engineering and Technology (IRJET), 6(4).