AET 003

Student Placement Prediction and AI-Powered Career Guidance using Machine Learning and NLP

Prithvi Vinayak P
Student,
Dept. of Electronics and Telecomm. Engg.,
Vidyalankar Institute of Technology, Mumbai
prithvi.p@vit.edu.in

Ajinkya Pimpalkar
Student,
Dept. of Electronics and Telecomm. Engg.,
Vidyalankar Institute of Technology, Mumbai
Ajinkya.pimpalkar@vit.edu.in

Viraj Ratadiya
Student,
Dept. of Electronics and Telecomm. Engg.,
Vidyalankar Institute of Technology, Mumbai
Viraj.ratadiya@vit.edu.in

Dr. Saurabh Mehta
Chief Academic Officer & Professor,
Dept. of Electronics and Telecomm. Engg.,
Vidyalankar Institute of Technology, Mumbai
Saurabh.mehta@vit.edu.in

Abstract - **The world of the global job market is changing and is no longer dependent solely on academic performance to predict the employability of a student. Today, students are assessed on a range of academic qualifications, skill sets, certifications, and relevant hands-on experiences. To fill this gap, we propose a two-fold intelligent system based on applying machine learning for placing the cars, and a large language model for the real-time generation of the roadmap from career. Placement Prediction System based on XGBoost, a gradient boosting framework-based parameters, such as CGPA, SSC/HSC performance, certification, internships, and technical skills build using. At the same time, the AI-Powered Career Chatbot is built on top of Meta's LLaMA 3 model and hosted locally with Ollama. Structured four-year road maps are ready to be generated based on the students' interests, background, and current market trends. General and institute-specific (VIT) two modes are offered by the chatbot, and local information on college clubs and events are incorporated in the latter. On top of being predictive, our system goes beyond to articles for students to take action on, bridging the gap between their current qualifications and what they would really like to do for a living. Experimental results and evaluations confirm that this modular platform can be used as an authoritative academic-to-career navigation tool.**

Keywords - Placement Prediction, Machine Learning, XGBoost, Career Guidance, Chatbot, LLaMA, NLP, Educational Data Mining, Ollama

I. Introduction

Transition from academic education to the industry employment is a complicated and sometimes uncertain process for students. Although many students score very high marks in academics, they don't get the suitable jobs because they lack the skill related to respective field, certifications, exposure to industry standards and which are relevant to the job. At present, institutional placement systems enormously depend on academic based performance metrics like CGPA, but do not take into account various other indicators of employability, including internships, project work or communication skills.

As AI is increasingly being used in all sorts of sectors, intelligent systems are becoming tools for making educational decisions. Machine learning can help provide predictions on performance based student data, and Natural Language Processing (NLP) can assist in having live interaction to deliver guidance in a conversational way. However, while there exists prior research that has explored ML models for predicting their academic performance [1][4], few works have presented an end to end integrated solution for improved career readiness where not only are the students to be advised and their performance to be predicted but also the interventions to be prescribed [1].

The first component of our proposed solution is a Placement Prediction System based on XGBoost that predicts the potential placements an individual will secure in the future, and the second is an AI Powered Careeer Chatbot that generates personalized 4-year development plans for each of our students. Together, these systems provide a smart, adaptive, and scalable handing over an employability platform for students to assess and increase their employability.

II. Related Works

Educational Data Mining has been studied for predicting outcomes and getting academic performance. XGBoost [1] by Chen and Guestrin was introduced since then and used widely in academic prediction tasks for its efficiency and performance. Following that, the work of Wang et al. [2] on ensemble-based approaches for predicting student success and that of Kim [3] on predicting dropout likelihood using deep learning are further developed. For example, GritNet predicts outcomes without giving actionable suggestions [3]. Feng et al. [4] utilized feature extraction in EDM as they claimed the relationship between sense of commitment to the class, path to the class, related courses, separate feature sets and academic results. Several studies [5][6] [7] highlighted the effect of certifications and internships in making the gap between industry readiness.

In spite of their success, most models to date are only for predictive analysis. The main appeal of our system is that we

provide a two layer solution which encompasses predictive analytics with XGBoost, as well as proactive career roadmap generation using NLP and LLMs. In addition to predicting employability of a worker, this hybrid model also curates custom skill development roadmaps for the employee.

III. Objectives

The goal for this project is to design a viable system that will help determine a student's placement readiness and offer real time and personalized career guidance. The project bridges the gap between the academic performance and employability through combining machine learning with conversational AI. The main objectives of our project are as follows:

1. Placement Prediction Using XGBoost
2. Top Company and Role Recommendations
3. AI-Based Career Chatbot
4. Career Roadmap Generation
5. Analysis of Placement Trends

Together, these objectives help to ensure that the project delivers a tool that is useful and interactive to use in helping students to improve their placement readiness and to make informed decisions regarding career options.

IV. Fundamentals of Technologies

A. Machine Learning & XGBoost

Machine learning is a single aspect of Artificial Intelligence that these algorithms can learn from the data to predict something, or decide. In this project, the supervised machine learning is used and is trained towards the labelled data to generalize to predict the outcome of an unseen data.

XGBoost (Extreme Gradient Boosting) is a distributed, scalable, optimized parallelized gradient boosting library for R, Python, Weka, Java, Julia, Octave, Julia, F# and more. It is useful for structured data problems, which is why it became one of the most popular tools in the world of machine learning competitions, as it has high performance. Second, XGBoost does so in an iterative fashion, adding decision trees that rectify the particular errors of the previously added trees and minimizes the loss via gradient descent. Moreover, some regularization techniques that help prevent overfitting are applied.

We use the XGBoost on the student attributes including CGPA, internships, certifications and technical skills to find the placements probability in the Placement Prediction System. We exploit the ensemble of decision trees, train the model on historical placement data and model makes accurate predictions.

Fig 1: XGBoost Model Architecture

B. Language Models: LLaMA, BERT, BART

As stressful as it may be, it is simply the emergent power of large language models (LLMs) which, trained on humongous text corpora, are able to understand and generate our human language. Many natural language processing (NLP) task rely on them.

1. LLaMA (Large Language Model Meta AI)

Meta AI's LLA family includes LLMs with 7B to 65B parameters. LLaMA is trained on various internet text and excels at producing coherent and contextually relevant responses. During this project you are using LLaMA to power the Career Chatbot, which can offer real time career advice and is able to generate a personalized roadmap for students.

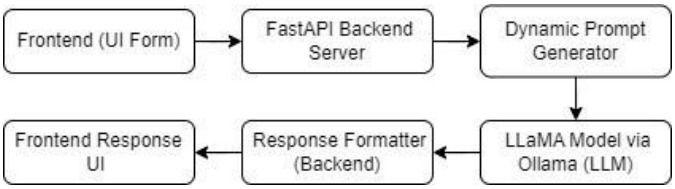

Fig 2: LLaMA via Ollama Architecture Flowchart

2. BERT (Bidirectional Encoder Representations from Transformers)

Google released the BERT model in 2018, which is a transformer-based model. It can understand word contexts taken into account not only the left context, but also the right context at the same time [10]. In the Career Chatbot, BERT is used for entity recognition, searching for aspects such as "AI/ML" or "Web Development", to better respond to user's questions.

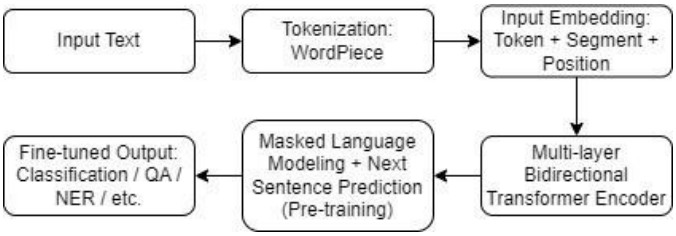

Fig 3: BERT Workflow

3. BART (Bidirectional and Auto-Regressive Transformers)

BART is a sequence to sequence task denoising autoencoder which combines bidirectional encoding as provided by BERT with autoregressive decoding that is provided by GPT. Corrupting text and training to reconstruct it, it is versatile on both generation and comprehension tasks [11]. As part of this project, BART is used to perform intent classification, such as "How can I prepare for a career in AI/ML?", meaning the chatbot knows what questions the users are querying the chatbot with. to generate appropriate responses.

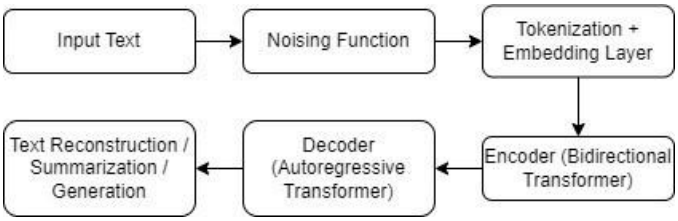

Fig 3: BART Workflow

These dual system technologies provide the backbone to the accuracy of student predictions and personalization of guidance.

V. Methodology

There are several key stages that the development of the dual system solution entailed: data collection, data preprocessing, training the Placement Prediction System model, development of the AI Powered Career Chatbot, and system integration.

A. Data Collection

Data was collected from the student record of Vidyalankar Institute of Technology, Mumbai. The data set consisted of academic metric CGPA, Secondary School Certificate (SSC), and Higher Secondary Certificate (HSC), as well as non-academic factor like internships, certifications, technical skills. Manually creating these models from scratch would not be possible without these comprehensive data.

B. Data Processing

To be suitable for machine learning algorithms, the raw data was processed very thoroughly. Some of those involved in handling missing values through imputation, some normalizing numerical features to ensure consistent scales, some encoding categorical variables by means such as one hot encoding. To increase the power of predicting, feature engineering was done to create new features: number of internships, different kinds of technical skills, etc.

C. Placement Prediction System

Using the XGBoost algorithm as its powerful classification algorithm, the Placement Prediction System was developed and evaluated. Features of the model included CGPA, SSC/HSC score, certificates, experience through internships, and technical skills, and it was trained on the pre-processed dataset. The binary target variable indicated whether a student was placed.

Model performance was evaluated by cross validating the models and calculating the accuracy, precision, recall and F1 score. Using the trained model, the placement probabilities of new students were predicted and top companies and job roles were recommended for them based on their profiles.

D. AI Powered Career Chatbot

To create a real time interaction, I developed the Career Chatbot using LLaMA language model and deployed it too using Ollama. For entity recognition of user input, BERT was used whereas BART for intent classification (identification of query type). Using pre-defined templates and data for the user, the chatbot generates personalized four year roadmaps for subjects such as AI and ML, Cybersecurity and Web Development.

Two modes of operation are available for the chatbot: general mode which provides general career advice and the other is the institution specific mode (VIT mode) wherein the advice is provided based on the institute's curriculum and resources.

E. System Integration

A web based platform is created between the backend via FastAPI and the front end using HTML/CSS/JavaScript with inclusion of the Placement Prediction System and Career Chatbot. This has a user friendly interface which lets the students input the data, get the predictions of his place and have a conversation with the chatbot related to his career. It is scalable and the architecture will allow for future feature integration and expansion of the system to other institutions.

VI. Results & Discussion

The dual system solution also provides actionable insights to students as to where they are likely to be placed, and their related career paths.

A. Placement Prediction System

The Placement Prediction System outputs a list of top companies with corresponding placement probabilities. Additionally, the system displays expected salary ranges and preferred locations as they represent informed decisions.

XG Boost is leveraged by the Usage of Placement Prediction System to make accurate predictions using many attributes of student but no longer leaving CGPA as sole determinant for academic performance. It is a holistic approach of students with different strengths like technical skills or internship experience.

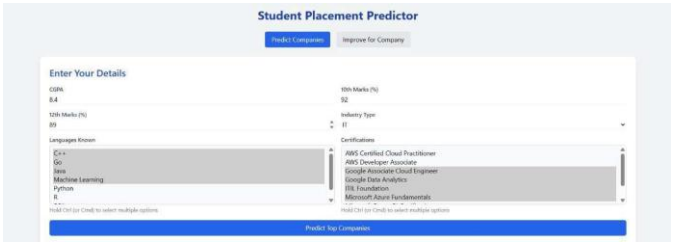

Fig 4: Input for Predicting Top Placement Companies

Fig 5: Output for Predicting Top Placement Companies

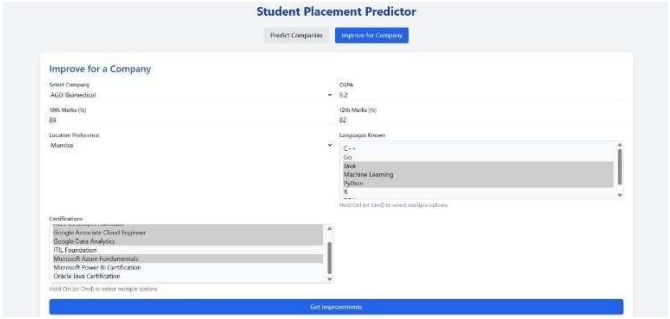

Fig 6: Input for Predicting Placement in a Specific Company

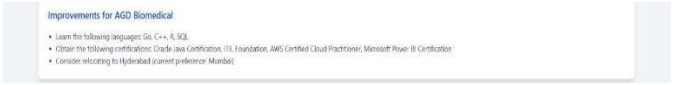

Fig 7: Output for Predicting Placement in a Specific Company

B. AI-Powered Career Chatbot

Career Chatbot enables Natural Language interaction profile students have access to the conversation with the career related questions such as career paths, skill development and job market trends. For example, an AI/ML student would get a step by step road map of the course work to study, skills to acquire and milestone when done with the course in 4 years. Real–time responses from the chatbot improve students' notion of career options.

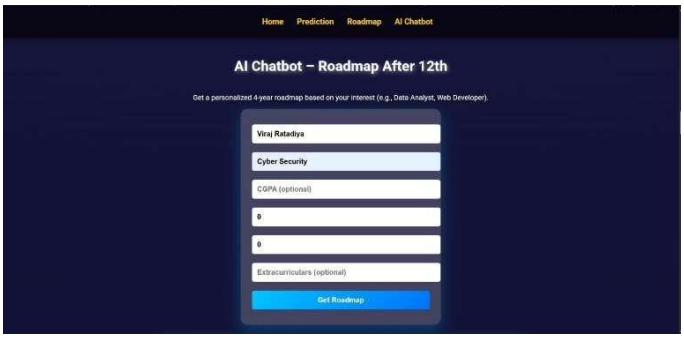

Fig 8: Input for AI Career Chatbot

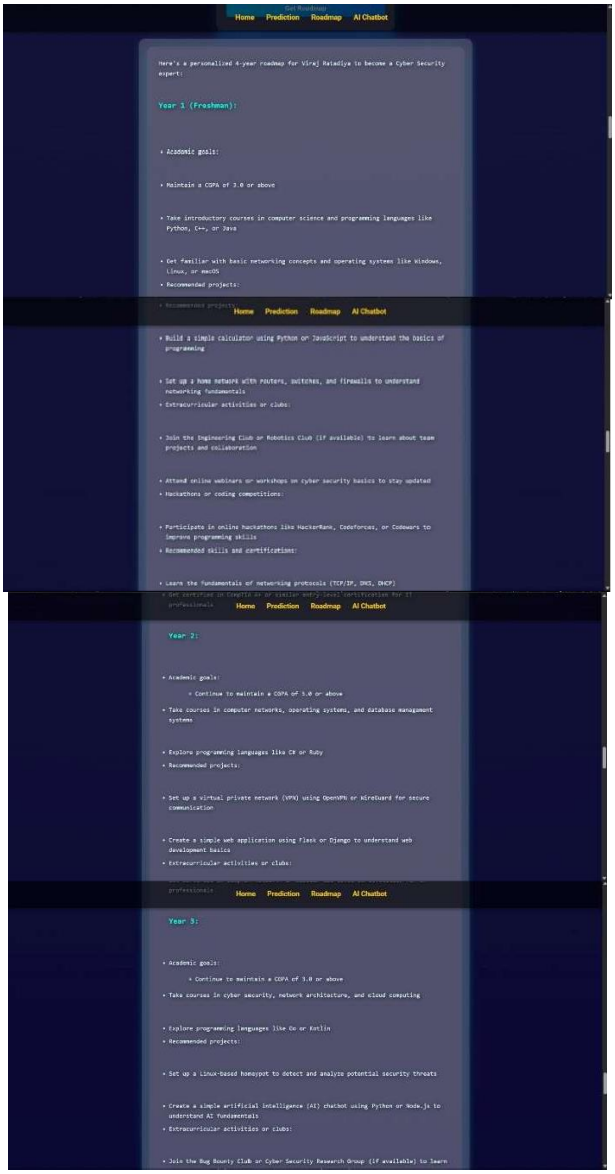

Fig 9: Output for AI Career Chatbot

Ready to give personalized roadmaps, and placing skilled workers accurately, the system proved it bridges the gap between academic performance and employability.

However, limitations exist. Accuracy of the Placement Prediction System depends on the quality and completeness of the training data. To improve predictive power, one could expand the dataset to include differing student profiles and institutions. The Career Chatbot needs working updates because it is updated on the current job market trends.

The next thing that could be added is real time job postings, explore deep learning model to improve accuracy, and making mobile apps that would allow for better accessibility. This study shows how machine learning and AI have the power to influence the supporting aspects of career guidance and placement at higher education.

VII. Conclusion

Based on this, this paper introduced the gap between the academic performance and the employability in engineering education and proposed a dual system solution of a Placement Prediction System and an AI Powered Career Chatbot. The Placement Prediction System is designed to accurately predict the placement probabilities as well as recommend companies and job roles, whereas the Career Chatbot is a real time, personalized guidance and roadmaps service.

The system is developed by data collection, preprocessing, model training and integration into a web platform to improve students' career planning and employability. Potentially future improvements include extending the dataset, preserving real time job trend, and exploring advanced AI model. In this sense, this research represents an important opening step for using technology to help students pass from education to employment.

VIII. References

[1] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. KDD 2016. https://doi.org/10.1145/2939672.2939785

[2] Wang, Y., Ding, A., Guan, K., Wu, S., & Du, Y. (2021). Graph-based Ensemble Machine Learning for Student Performance Prediction. arXiv preprint arXiv:2112.07893. https://arxiv.org/abs/2112.07893

[3] Kim, B. H. (2019). Deep Learning to Predict Student Outcomes. arXiv preprint arXiv:1905.02530. https://arxiv.org/abs/1905.02530

[4] Feng, G., Fan, M., & Chen, Y. (2022). Analysis and Prediction of Students' Academic Performance Based on Educational Data Mining. IEEE Access, 10, 19558–19572. https://doi.org/10.1109/ACCESS.2022.3151652

[5] Agrawal, H., & Mavani, H. (2015). Student Performance Prediction using Machine Learning. IJERT, 4(3). https://www.ijert.org/student-performance-prediction-using-machine-learning

[6] Yağci, M. (2022). Educational Data Mining: Prediction of Students' Academic Performance Using Machine Learning Algorithms. Smart Learning Environments, 9(11). https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z

[7] Chen, Z., Cen, G., Wei, Y., & Li, Z. (2023). Student Performance Prediction Approach Based on Educational Data Mining. IEEE Access, 11, 131260–131274. https://doi.org/10.1109/ACCESS.2023.3335985

[8] Sayana, T. S. (2015). Prediction of Students Academic Performance using Data Mining: Analysis. IJERT, 3(30). https://www.ijert.org/prediction-of-students-academic-performance-using-data-mining-analysis

[9] Bhegade, K. B., & Shinde, S. V. (2016). Student Performance Prediction System with Educational Data Mining. IJCA, 146(5). https://doi.org/10.5120/ijca2016910704