



INTERSHIP PROGRAM 2023

PROJECT REPORT

MACHINE LEARNING

PlacementPal : Graduation and Placement Predictor

Created By:	Dhananjay Gaur	Approved By:	<Domain Lead Name>
Created On:	13-09-2023	Approved On:	DD-MMM-YYYY

Page left blank intentionally

INDEX

1	PROJECT DETAILS	1
2	SUMMARY	1
3	INTRODUCTION	1
3.1	Background	1
3.2	Stakeholders	1
3.3	Objectives	1
4	METHODOLOGY	2
4.1	Considerations & Assumption	2
4.2	Approach	2
4.3	Activities	3
5	TARGETTED V/S ACHIEVED OUTPUT	3
6	CONCLUSION	3
7	APPENDICES	4
7.1	Appendix A – Decision Tree Classifier	4
7.2	Appendix B – Linear Regression	6

1 PROJECT DETAILS

Project Name	PlacementPal : Graduation and Placement Predictor		
Project Sponsor	Tushar Topale		
Project Manager	Harshada Topale		
Start Date	01-08-2023	Completion Date	14-09-2023

2 SUMMARY

The project's methodology included stringent data pre-treatment, including the removal of irrelevant data, to address the issue at hand. A number of conventional classification models were evaluated when two distinct subproblems, Placement Status Prediction and Year of Graduation Prediction, were looked at. These models included Logistic Regression, Naive Bayes, Decision Tree, Random Forest, and Artificial Neural Networks. The placement prediction task was handled expertly by the Decision Tree Algorithm, which provided a variety of solutions with astounding accuracy. Finally, by consistently producing reliable results across the sample, linear regression proved to be a workable alternative for predicting the year of graduation. The project was completed by using these carefully selected models on the test dataset, appending the results to the original dataset, and saving them in a new Excel file. This resulted in significant insights and predictive projections for students' academic trajectories and career options.

3 INTRODUCTION

3.1 Background

Lack of clear academic planning and career guidance is the primary cause of the issue, which leaves students uncertain of their graduation date and without a job offer. Students may find it difficult to finish their coursework on time if they do not receive timely information on their graduation status. Additionally, people could not be effectively prepared for the job market without a prior understanding of the placement-specific skills, which could make it difficult for them to get work and impede their chances of having successful careers in general.

3.2 Stakeholders

The stakeholders of the project include:

- Process Holder: Harshada Topale
- End Customers: Cloud Counselage Pvt. Ltd.
- Project Team: Dhananjay Gaur

3.3 Objectives

Make a prediction model that can accurately anticipate a student's graduation year and placement status.

4 METHODOLOGY

4.1 Considerations & Assumption

The Consideration and Assumptions of the project are:

- The model assumes that the available data is comprehensive enough to capture all relevant factors influencing graduation and job outcomes.
- The model assumes that future events or changes in policies that could impact graduation rates and job prospects are not considered.
- For model training, sufficient and relevant historical data will be accessible.
- The data provided would be suitable for the use of Machine Learning models.

4.2 Approach

To address the provided challenge, the data was cleaned and superfluous data that was not required for model training was removed. I then separated the problem into two subproblems:

1. Prediction of placement status:

Because placement prediction is a classification issue, I initially tested a few typical classification Machine Learning models to see which performed the best. The models used were as follows:

1. Logistic Regression Algorithm
2. The Naive Bayes Algorithm
3. The Decision Tree Algorithm
4. Artificial Neural Networks with the Random Forest Algorithm

It was discovered that the Decision Tree Algorithm worked best on the supplied dataset, yielding different solutions with good accuracy.

2. Predicted graduation year:

Because predicting the year of graduation is a regression issue, I utilized Linear Regression on the same dataset that was used to train the model for placement prediction. Because linear regression performed well on the dataset, it was chosen to be used on the final test dataset.

Once the models for both subproblems were chosen, the test dataset was loaded and these models were applied to it to create the final output.

The findings were appended to the original dataset as the final output, and the final output was saved in a new Excel file.

4.3 Activities

The activities performed during the project include:

- **Requirement Gathering-**
 - Requirement Elicitation
 - SRS Documentation
- **Project Planning-**
 - Creating a Work Breakdown Structure
 - Upkeep of the RAID Log
 - Keeping a Lessons Learned Log
- **Project Execution**
- **Weekly Meetings**
- **Project Monitoring**
- **Project Closure-**
 - Project Report Generation
 - Submission of the project

5 TARGETTED V/S ACHIEVED OUTPUT

Target Output:

Predict the year of graduation and student placement with high accuracy.

Finished Product:

The accuracy for predicting the year of graduation and student placement was 71%.

6 CONCLUSION

Benefits for Stakeholders:

1. Students

Graduation on Time: Students may better plan their academic route by receiving specific forecasts of the year they will graduate, ensuring they complete their studies on time and meet graduation standards.

Improved Career Preparation: By learning about the placement-specific skills required, students may focus on developing these abilities early on and become more competitive in the labour market.

Improved Employability: By employing exact placement projections, students may correct their defects and capitalize on their strengths, increasing their chances of obtaining a job before graduation.

2. Educational Institutions

Enhanced Student Services: Depending on the students' expected graduation year and career prospects, the institution may provide more personalized academic counselling, increasing student satisfaction ratings.

Improved Resource Allocation: Because accurate predictions can anticipate student demand for courses and support services, they can help institutions allocate resources more effectively.

3. Employers:

Quality Recruitment: Employers may organize their recruitment efforts more efficiently by identifying potential candidates from upcoming graduating batches and ensuring a steady supply of skilled experts.

Reduced Time-to-Hire: If employers are aware of the availability of graduating students ahead of time, they may speed up their recruitment operations and reduce the time it takes to hire new staff.

Future Scope:

1. **Career Path Suggestions:** The project may evolve to give personalised career path suggestions based on students' choices, talents, and market trends, directing them to industries with the best opportunities.
2. **Industry Trend examination:** The project's examination of historical placement data might show market patterns and help schools change their courses more effectively to suit labor market demands.

7 APPENDICES

7.1 Appendix A – Decision Tree Classifier

1. Problem Formulation

The primary goal of our project's use of the Decision Tree Classifier was to forecast where students will be assigned. You must examine a dataset of student characteristics, academic success, extracurricular activities, and placement results for this categorization assignment.

2. Methodology

- **Data Preprocessing:** We handled missing values, coded categorical variables, and separated the dataset into training and testing sets to prepare the data.
- **Feature Selection:** To train the Decision Tree Classifier, relevant features were chosen based on qualities that are likely to impact placement outcomes. These included metrics for academic progress, course advancement, and extracurricular activity.
- **Model Training:** The Decision Tree Classifier was trained on the training dataset using the Python scikit-learn module. We tried a number of hyperparameters to ensure the optimum model performance.
- **Model Evaluation:** We evaluated the model's performance on the testing dataset using a variety of metrics, including accuracy, precision, recall, and F1-score. We also visualized the decision tree that was produced for easy comprehension.

3. Results

The Decision Tree Classifier produced good results, with an accuracy of about 0.71.

7.2 Appendix B – Linear Regression

1. Problem Formulation

To solve the problem of predicting graduation year, we employed Linear Regression. The goal was to create a prediction model that, given relevant academic data, forecasts the year a student is likely to graduate.

2. Methodology

- **Data Pre-processing:** Similar to the classification task, we used data preprocessing to manage missing values and prepare the dataset for training and testing.
- **Feature Selection:** We chose relevant academic variables such as the current academic year, course progress, and past academic success.
- **Model Training:** The linear regression model was trained using the scikit-learn package. We utilized a feature-engineered dataset to predict the graduation year.
- **Model Evaluation:** On the testing dataset, model assessment was done using metrics like Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).

3. Results

The Linear Regression Model produced good findings.