(ISSN: 2582-5526) www.thevoiceofcreativeresearch.com

The Voice of Creative Research

Vol. 7 & Issue 2 (April 2025)

https://doi.org/10.53032/tvcr/2025.v7n2.25

Research Article

Machine Learning for Placement Prediction: A Study Using Weka, Orange, and Simple ML

Siddhesh Kadam
Assistant Professor,
Dept of Computer Science
Kirti M. Doongursee College of Arts Science and Commerce
(Autonomous), Mumbai
siddhesh.rohidas.kadam@gmail.com

Prabha Siddhesh Kadam
Assistant Professor,
Dept of Computer Science
Kirti M. Doongursee College of Arts Science and Commerce
(Autonomous), Mumbai
prabhapilankar@gmail.com

Abstract

Placement prediction is a crucial application of machine learning in education, helping institutions and students understand employability factors. It enables educational institutions to design effective training programs and assists students in improving their career prospects. This study evaluates the performance of three popular data analysis tools, namely Weka, Orange, and Simple ML, using a publicly available placement dataset. The dataset comprises various features, including academic performance, extracurricular involvement, and technical skills, which play a significant role in determining job placements. The models were assessed based on multiple evaluation metrics, including accuracy, precision, and recall. The findings offer valuable insights into the most effective tool for placement prediction, highlighting strengths, limitations, and potential areas for improvement. This research aims to assist educators, data scientists, and institutions in selecting the most suitable machine learning tool for predictive analytics in placement prediction.

Keywords: Predictive Analytics, Classification Algorithms, Machine Learning, Data Science

1. Introduction

The increasing competition in the job market necessitates the development of efficient, data-driven prediction models to enhance student placement outcomes. Universities and institutions worldwide are leveraging machine learning (ML) tools to analyse historical placement data, identify employability factors, and optimize student training strategies. By utilizing ML techniques, institutions can gain insights into trends, predict hiring probabilities, and personalize career guidance based on data-driven evidence.

This research focuses on the comparative evaluation of three prominent machine learning tools like Weka, Orange, and Simple ML for placement prediction. Each of these tools offers distinct advantages in data preprocessing, model selection, evaluation, and visualization. Weka is well-known for its extensive algorithm library and academic applications, Orange provides a visual and interactive data mining environment, while Simple ML, a Google Sheets add-on, enables seamless integration with spreadsheet-based workflows, making it accessible for non-technical users. The study examines these tools based on parameters like model performance where the model Accuracy, precision, recall, F1-score, and computational efficiency will be checked.

Beyond mere technical comparisons, this study also explores the practical implications of ML-driven placement prediction in academic settings. An efficient placement prediction system enables institutions to proactively identify students at risk of unemployment, suggest skill enhancement programs, and tailor career training modules accordingly. By analysing a publicly available placement dataset, this research aims to determine which tool offers the most effective and scalable solution for placement prediction. The findings provide valuable insights for educational institutions, career counsellors, and researchers, aiding in the adoption of the most suitable ML tool for data-driven career planning and student success enhancement.

2. Literature Review

To predict student placement into departments or work environments based on academic performance and other qualities, several research have investigated different machine learning techniques. For analysis, researchers have used programs like R Studio and WEKA. By providing focused training and support, educational institutions can use the insights gleaned from these predictions to enhance student outcomes (K. S. Rao et al., 2017). WEKA tool is easy to use and produces results in a split second. The Java programming language is used to construct the Weka tool's foundation, which includes a robust library for comprehension and data processing methods. Without much effort, statistical analysis produces and makes available for researchers. To classify students' job profiles as programmers, Kadam P., and Tere G. (2022) employed the Weka application. Kajal Rai (2022) used Random Forest, Naïve Bayes, and decision trees to forecast student placement. The accuracy gained was then used to evaluate these methods. 265 MBA student samples were used in the study. Compared to the two other algorithms that were studied, Random Forest's accuracy with 10-fold-cross validation in the study using the WEKA tool was 86.04%. Machine learning was employed by Bindu and

Dushyanth (2021) to forecast student placement using the WEKA tool. HSC grades, graduation level grades, previous department and history, communication and programming abilities, internships accomplished or not, and details about interest areas are all included in the data collection. To determine the study's characteristics and predict the students' placement, the authors preferred to use entropy and gain. Mavani Umang et al., 2020 In order to accurately anticipate student placements, the study used WEKA, R, Orange, and Python on data set contains the field like SSC, HSC, and engineering marks. The goal was to compare technologies. The authors T.Thilagaraj and Dr.N Sengottaiyan (2017), stated that Python scripting or visual programming are used for data mining in Orange. Both beginners and specialists can use this open-source data visualization and analysis tool. It has machine learning components as well as bioinformatics and text mining add-ons. In 2014, A. Jović, K. Brkić, and N. Bogunović wrote an overview of free software tools for general data mining. However, they noted that Orange's apparent drawback is that there are not many widgets available.

Previous studies have applied machine learning for placement prediction, utilizing classification algorithms like Decision Trees, Random Forest, and Support Vector Machines (SVM). While Weka and Orange have been widely studied, fewer studies focus on Simple ML. This paper aims to fill that gap by evaluating their performance using a common dataset.

3. Objective:

1. To implement machine learning algorithms in Weka, Orange, and Simple ML to identify which tool performs best across various model types.

2. To evaluate and compare the performance of Weka, Orange, and Simple ML in predicting student placements.

4. Methodology:

A publicly available datasets on student placements was obtained from Kaggle. The datasets include attributes such as academic performance, extracurricular activities, certifications, and placement outcomes. The datasets differed in terms of number of records as well as the fields considered for the study.

The details of the three data sets used are given below:

1. File Name:	Placedata v2.0-Synthetic	
# Columns: 12	# Rows :10000	Age, gender, stream, Internship, CGPA, Hostel, History of Backlog, Placed or Not
2. File Name:	Job_placement	
# Columns: 11	# Rows :700	Id, Name, gender, age, degree, stream, college name, placement status, salary, gpa, yrs of experience
3. File Name:	Job_placement_data	
# Columns: 13	# Rows :215	Gender, ssc_percentage, ssc_board, hsc_percentage, hsc_board, hsc_subject, degree_percentage, undergrad_degree, work_experience, emp_test_percentage, specialisatio, mba_percent, status

Table 1: Data Set Description

Data Preprocessing stage performed the operations for handling missing values, feature scaling and encoding categorical variables as well as the splitting data into training and testing sets.

Experimental Setup in the Orange:

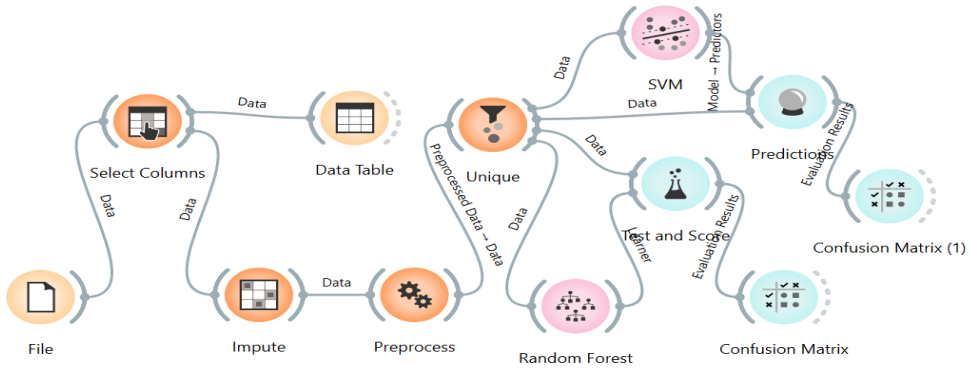
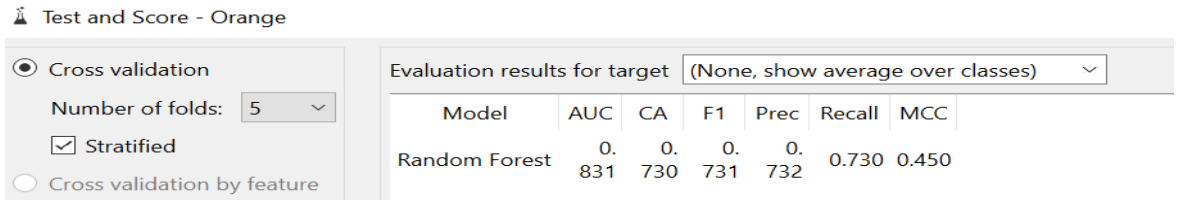


Figure 1, Model designed for Binary Placement Prediction



Test and Score - Orange

Cross validation

Number of folds: 5

Stratified

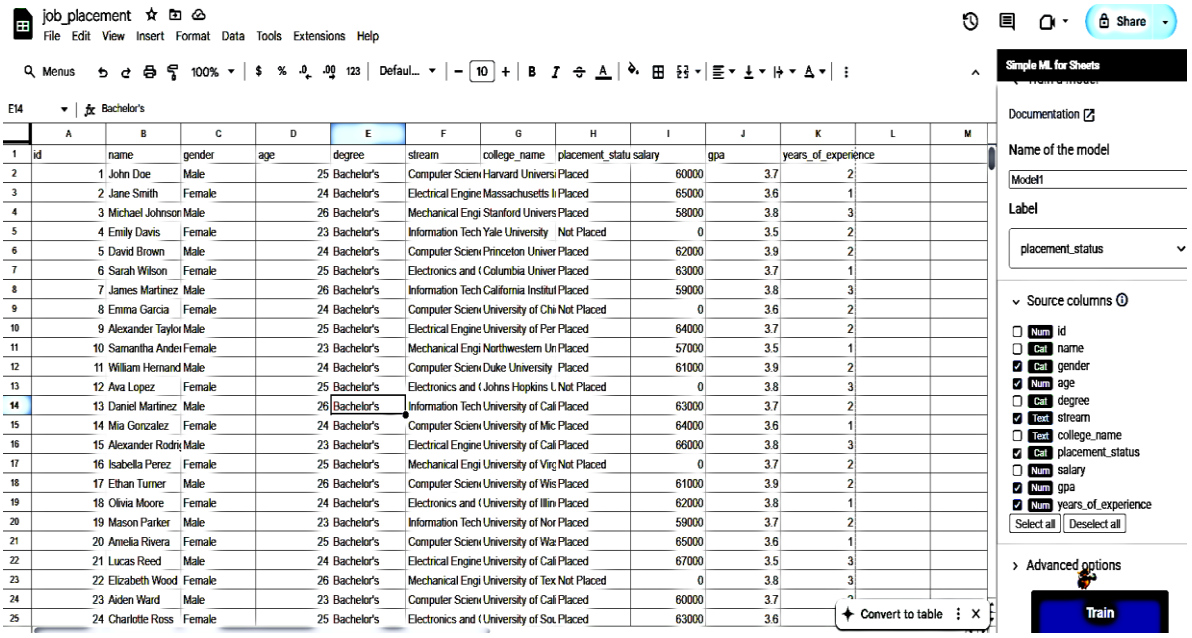
Cross validation by feature

Evaluation results for target (None, show average over classes)

Model	AUC	CA	F1	Prec	Recall	MCC
Random Forest	0.831	0.730	0.731	0.732	0.730	0.450

Figure 2, Model Evaluation Random Forest

Use of Simple ML for Placement Prediction:



job_placement

File Edit View Insert Format Data Tools Extensions Help

Share

Simple ML for Sheets

Documentation

Name of the model

Model1

Label

placement_status

Source columns

Num id
Cat name
Cat gender
Num age
Cat degree
Text stream
Text college_name
Cat placement_status
Num salary
Num gpa
Num years_of_experience

Select all Deselect all

Advanced options

Train

Convert to table

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	id	name	gender	age	degree	stream	college_name	placement_statu	salary	gpa	years_of_experience		
2	1	John Doe	Male	25	Bachelor's	Computer Scien	Harvard Univers	Placed	60000	3.7	2		
3	2	Jane Smith	Female	24	Bachelor's	Electrical Engine	Massachusetts I	Placed	65000	3.6	1		
4	3	Michael Johnson	Male	26	Bachelor's	Mechanical Engi	Stanford Univers	Placed	58000	3.8	3		
5	4	Emily Davis	Female	23	Bachelor's	Information Tech	Yale University	Not Placed	0	3.5	2		
6	5	David Brown	Male	24	Bachelor's	Computer Scien	Princeton Univer	Placed	62000	3.9	2		
7	6	Sarah Wilson	Female	25	Bachelor's	Electronics and (	Columbia Univer	Placed	63000	3.7	1		
8	7	James Martinez	Male	26	Bachelor's	Information Tech	California Institut	Placed	59000	3.8	3		
9	8	Emma Garcia	Female	24	Bachelor's	Computer Scien	University of Chi	Not Placed	0	3.6	2		
10	9	Alexander Taylor	Male	25	Bachelor's	Electrical Engine	University of Per	Placed	64000	3.7	2		
11	10	Samantha Ander	Female	23	Bachelor's	Mechanical Engi	Northwestern Un	Placed	57000	3.5	1		
12	11	William Hernand	Male	24	Bachelor's	Computer Scien	Duke University	Placed	61000	3.9	2		
13	12	Ava Lopez	Female	25	Bachelor's	Electronics and (	Johns Hopkins U	Not Placed	0	3.8	3		
14	13	Daniel Martinez	Male	26	Bachelor's	Information Tech	University of Cal	Placed	63000	3.7	2		
15	14	Mia Gonzalez	Female	24	Bachelor's	Computer Scien	University of Mic	Placed	64000	3.6	1		
16	15	Alexander Rodri	Male	23	Bachelor's	Electrical Engine	University of Cal	Placed	66000	3.8	3		
17	16	Isabella Perez	Female	25	Bachelor's	Mechanical Engi	University of Virg	Not Placed	0	3.7	2		
18	17	Ethan Turner	Male	26	Bachelor's	Computer Scien	University of Wis	Placed	61000	3.9	2		
19	18	Olivia Moore	Female	24	Bachelor's	Electronics and (	University of Illin	Placed	62000	3.8	1		
20	19	Mason Parker	Male	23	Bachelor's	Information Tech	University of Nor	Placed	59000	3.7	2		
21	20	Amelia Rivera	Female	25	Bachelor's	Computer Scien	University of Wa	Placed	65000	3.6	1		
22	21	Lucas Reed	Male	24	Bachelor's	Electrical Engine	University of Cal	Placed	67000	3.5	3		
23	22	Elizabeth Wood	Female	26	Bachelor's	Mechanical Engi	University of Tex	Not Placed	0	3.8	3		
24	23	Aiden Ward	Male	23	Bachelor's	Computer Scien	University of Cal	Placed	60000	3.7			
25	24	Charlotte Ross	Female	25	Bachelor's	Electronics and (	University of Sou	Placed	63000	3.6			

Figure 3, Training Phase of the Model1 (Random Forest)

Model development using WEKA:

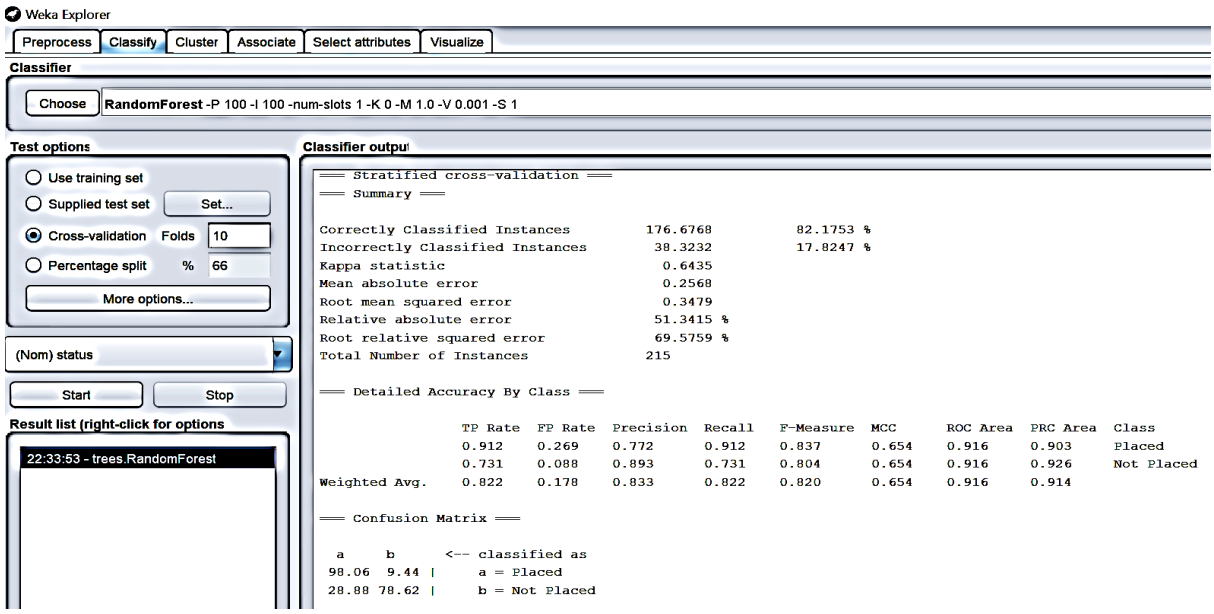


Figure 4, Binary Classification using Random Forest

The datasets were analysed using classification algorithms in Weka, Orange, and Simple ML. Decision Tree (DT), Random Forest (RF), and Naïve Bayes were used to train the models. Accuracy, precision, recall, and F1-score were among the evaluation metrics.

5. Results and Discussion:

A comparative analysis of the tools yielded the following key findings:

Data Set:	**Placedata v2.0-Synthetic**			**Job_placement**			**Job_placement_data**		
Algorithms:	**DT**	**RF**	**Naïve Bayes**	**DT**	**RF**	**Naïve Bayes**	**DT**	**RF**	**Naïve Bayes**
Weka									
accuracy	77.28	78.3	79.54	79.23	81.33	77.19	74	82.18	80.5
precision	77.3	78.4	79.6	79.2	79.4	73.9	74	83.3	80.5
recall	77.2	78.3	79.5	79.2	81.3	77.2	74	82.2	80.5
F1-score	77.2	78.3	79.5	79.2	79.9	75	74	82	80.5
Orange									
accuracy	73.3	78.2	79.6	82.7	82.6	77.9	76.7	82.8	80.5
precision	73.1	78.1	79.8	81.4	80.8	77.4	76.7	82.4	80.6
recall	73.3	78.2	79.6	82.7	82.6	77.9	76.7	82.8	80.5
F1-score	73.2	78.1	79.6	81.9	81.4	77.6	76.7	82.4	80.5

Table 2, Result Analysis of the Models Build using the Tools

The comparative analysis between Weka and Orange highlights slight variations in model performance. Although Naïve Bayes achieved the highest accuracy in Weka, Random Forest proved to be the best-performing model in Orange. Decision Tree exhibited the lowest

accuracy, suggesting that it may not be the most suitable model for job placement prediction in this dataset. While both tools produced comparable results, Orange reported slightly higher accuracy for Random Forest, suggesting potential differences in internal algorithmic optimizations.

The study evaluated the performance of Decision Tree and Random Forest algorithms using the Simple ML environment on three datasets: Placedata v2.0, Job_placement, and Job_placement_data. The Decision Tree algorithm achieved accuracy rates of 82%, 89%, and 87.44% for the respective datasets. In comparison, the Random Forest algorithm outperformed Decision Tree, achieving accuracy rates of 91.96%, 89.92%, and 96.74%. These results indicate that the Random Forest algorithm provides better predictive performance across all datasets. The models were developed using the default settings in Simple ML, ensuring a standardized evaluation framework.

The above outcomes obtained from the multiple models build using the three tools indicates;

Weka: Performed well with Naïve Bayes, offering high accuracy but requiring careful parameter tuning.

Orange: Provided an intuitive interface, suitable for beginners but had limited advanced customization.

Simple ML: Easy to integrate with Google Sheets, provided an accessible and user-friendly platform for model development, making it suitable for quick analysis and non-programmers, but with limited flexibility for advanced model customization.

6. Conclusion and Future Work:

The results indicate that Random Forest and Naïve Bayes are the most suitable models for job placement prediction, offering high accuracy and balanced performance across multiple evaluation metrics. The selection of the best model may depend on the specific application needs, Naïve Bayes for probabilistic classification and Random Forest for robustness against overfitting. Future research could explore hybrid models or ensemble techniques to further enhance predictive accuracy.

7. References

1. K., Sreenivasa, Rao., N., Swapna., Pankaj, Kumar. (2017). Educational data mining for student placement prediction using machine learning algorithms. International journal of engineering and technology, doi: 10.14419/IJET.V7I1.2.8988
2. Kadam, P., & Tere, G. (2022). Placement Prediction for Undergraduate Computer Science Student Using Ensemble Learning. International Journal of Research and Analytical Reviews IJRAR, 5(3), 25–29.
3. Hima Bindu, J., & Dushyanth, B. (2021). Student placement prediction using machine learning. International Journal of Creative Research Thoughts (IJCRT), 9(6), IJCRT2106788. Retrieved from www.ijcrt.org

4. Rai, K. (June 2022). Students' placement prediction using machine learning algorithms. South Asia Journal of Multidisciplinary Studies SAJMS, 8(5), 54-60.
5. Thilagaraj, T., & Sengottaiyan, N. (2017). A review of educational data mining in higher education system. Proceedings of the Second International Conference on Research in Intelligent and Computing in Engineering, 349–358. DOI: 10.15439/2017R87
6. Jović, A., Brkić, K., & Bogunović, N. (2014). An overview of free software tools for general data mining. Proceedings of MIPRO 2014, 26-30 May 2014, Opatija, Croatia, 1112–1117.