

# EMPLOYEE PROMOTION PREDICTION

## CASE STUDY REPORT

**Amrutha R Nair (2217006)**

Department of Computer Science - Rajagiri College of Social Sciences - Kalamassery,  
Kochi, Kerala

[msccs2205@rajagiri.edu](mailto:msccs2205@rajagiri.edu)

### 1.ABSTRACT:

Employee promotion is the process of elevating an employee to a higher position or level within an organization. It is usually done to recognize an employee's exceptional performance, skills, and abilities, and to provide new challenges and opportunities for career growth. The aim of this project is to classify whether an employee of a particular organization get promoted or not. Promotion decisions are usually based on a combination of factors. Based on different criteria, an employee get promoted. When a wrong employee promotion occurs, it can have several negative consequences for both the organization and the individual involved. To mitigate the negative consequences, it is important for organizations to have clear and objective criteria for promotion, and to ensure that the most qualified and deserving candidates are selected. It is fundamental to know whether an employee will promoted or not. The project aims to build multiple classification models using Orange tool and select the model with the best accuracy out of it.

### 2.INTRODUCTION

Employee promotion is an important aspect of any organization's growth and success. Promoting employees from within the organization not only recognizes their hard work but also helps to retain top talent and increases productivity. However, identifying the right candidates for promotion can be a challenging task, especially for large organizations with a diverse workforce. This case study examines a dataset of employee promotion records from a multinational corporation and explores how they can be used to predict employee promotion. By analyzing the dataset and applying various models that can be created by orange tool, we can gain insights into the factors that contribute to employee promotion and develop a predictive model to identify high-performing employees who are likely to be promoted in the future. In today's competitive job market, organizations must strive to retain top talent and provide opportunities for growth and advancement. One way to achieve this is through employee promotion. However, the promotion process can be subjective and prone to bias, which can lead to unequal opportunities for employees. To address this issue, many organizations are turning to data-driven approaches to ensure fairness and objectivity in the promotion process. In this case study, we will explore a dataset on employee promotion and analyze the factors that contribute to a successful promotion outcome. We will also examine how model created by orange tool can be used to predict the likelihood of an employee being promoted, and how this can inform organizational decision-making. Through this case study, we hope to demonstrate the value of data-driven approaches in promoting fairness and equity in the workplace.

### 3.MATERIALS AND METHODS

#### Materials:

- A computer with Orange installed
- Employee promotion dataset (can be downloaded from Kaggle or other sources)

#### Methods:

- Import the dataset into Orange:
- Open Orange and select the "File" option from the top menu
- Select "Open" and browse for the location of the employee promotion dataset
- Select the dataset file and click "Open"

SI NO	ATTRIBUTE	DESCRIPTION	TYPE
1	employee_id	Id of employee	Numeric
2	department	Department of employee	categorical
3	region	Region of employee	categorical
4	education	Education of employee	categorical
5	gender	Employee gender	categorical
6	recruitment_channel	Channel of recruitment for employee	categorical
7	no_of_trainings	Number of trainings attended by employee	Numeric
8	age	Age of employee	Numeric
9	previous_year_rating	Past year rating	Numeric
10	length_of_service	Service year length in years	Numeric

11	awards_won?	Number of awards won	categorical
12	avg_training_score	Average of training score of an employee	Numeric
13	is_promoted	Employee is promoted or not	categorical

#### 4.DATA VISUALIZATION AND INTREPRETATION

Visualization is a powerful tool for understanding and communicating information contained within a dataset. By representing data visually, patterns, trends, and outliers can be easily identified and interpreted, making it easier to draw insights and make data-driven decisions. It enables data analysts to identify outliers and correlations, which would have been challenging to identify otherwise. Additionally, visualization also helps to uncover trends and patterns that are not easily discernible from tabular data. It facilitates communication of insights to stakeholders in a more compelling and easy-to-understand way, which ultimately helps in making data-driven decisions. Therefore, data visualization plays a crucial role in improving the quality of data analysis and is an indispensable part of the analytical toolkit.

In order to analyze any data and to bring out the important features,

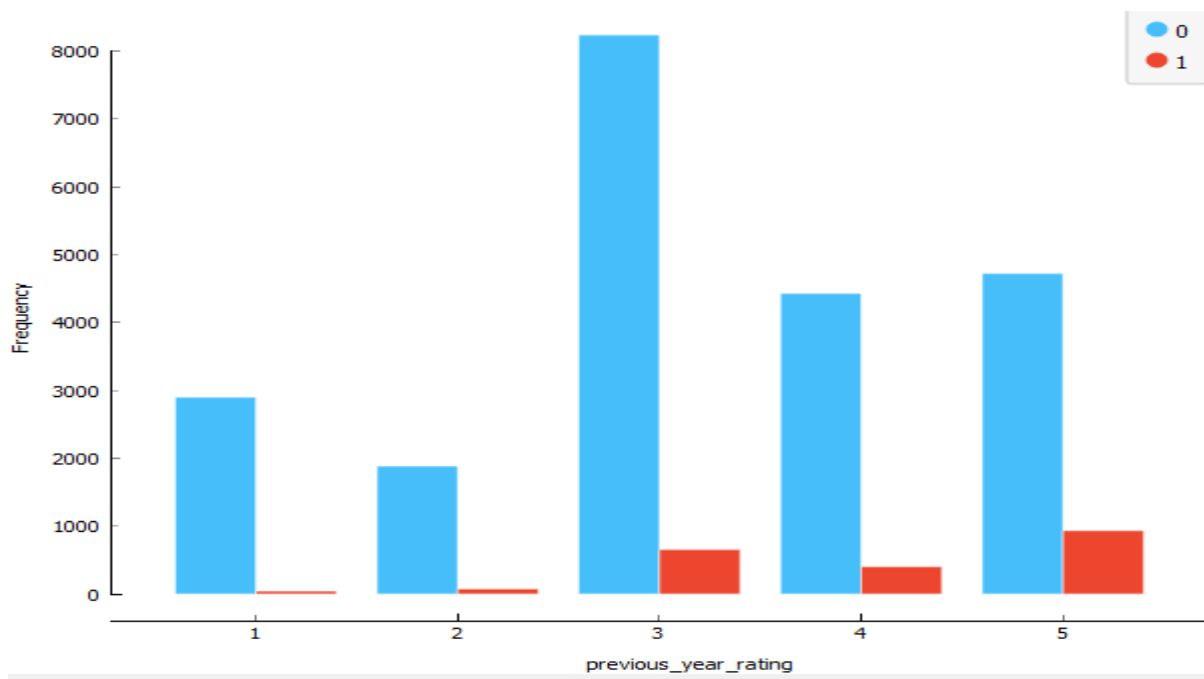


Fig:Previous\_year\_rating v/s is\_promoted

Based on the above graph, we can tell that employees with a previous year rating of 5 are promoted highly. Previous year rating can be an important feature in employee promotion decisions. The rating provides an indication of the employee's performance and achievements in the previous year, which can be a useful reference point for evaluating their potential for

promotion. Generally, organizations look for employees who consistently perform at a high level and demonstrate the potential for growth and development. However, it should be noted that promotion decisions should not be based solely on past performance ratings. Furthermore, it is important to ensure that the rating system is fair and transparent and that employees have a clear understanding of the criteria used for evaluation.

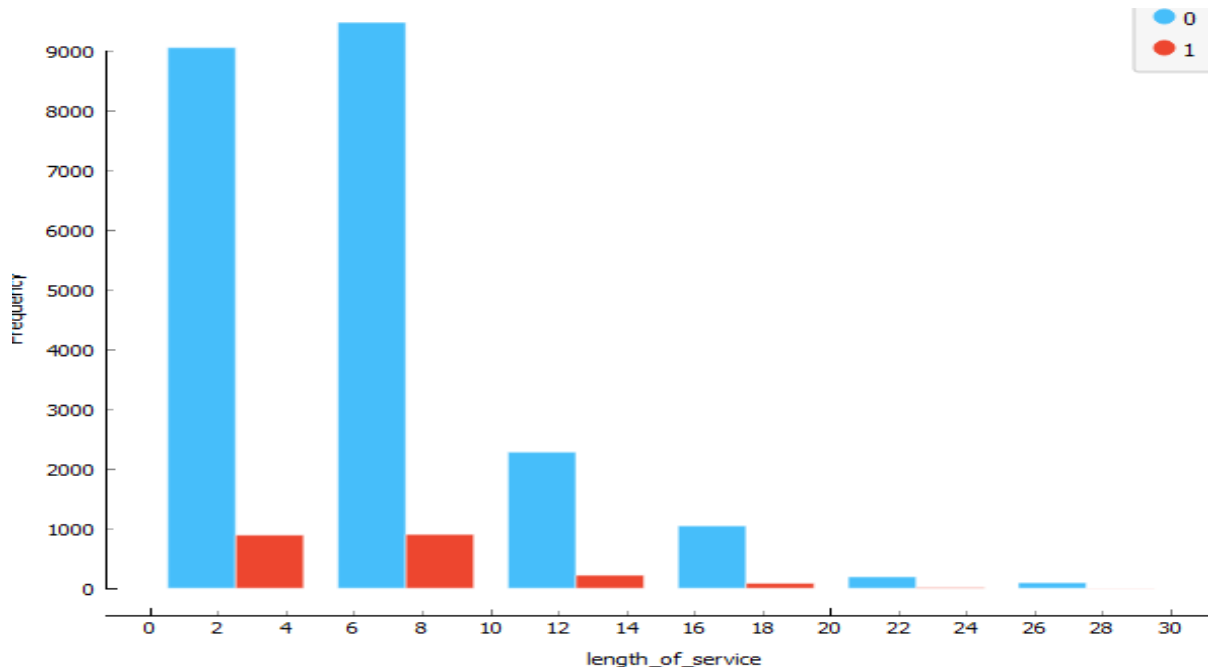


Fig :length\_of\_service v/s is\_promoted

The data presented in this graph makes it very evident that employees who have served for fewer than ten years are more likely to be promoted than those who have served for more than ten years. However, this is not always the case because decisions regarding promotions are often made based on a wide range of considerations, such as job performance, talents, potential for growth and development, and the requirements of the business. Even while employees with fewer than ten years of service may have shown good performance and promise, it is possible that employees with more than ten years of service may have gained additional knowledge and expertise that makes them better fit for certain positions or responsibilities.

In addition, the policies and practices of an organization, such as seniority-based promotion schemes, may give priority to workers who have been with the company for a longer period of time. In the end, choices on promotions should not be based solely on length of service but rather on an evaluation process that is both fair and transparent and takes into consideration a variety of factors.

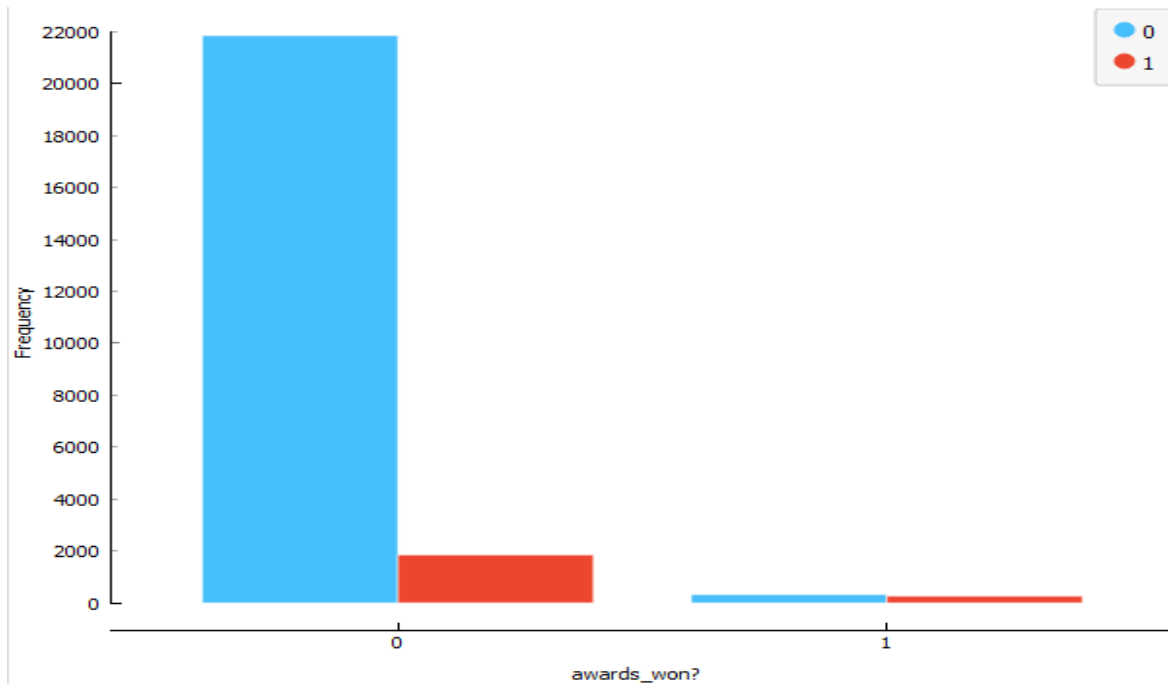


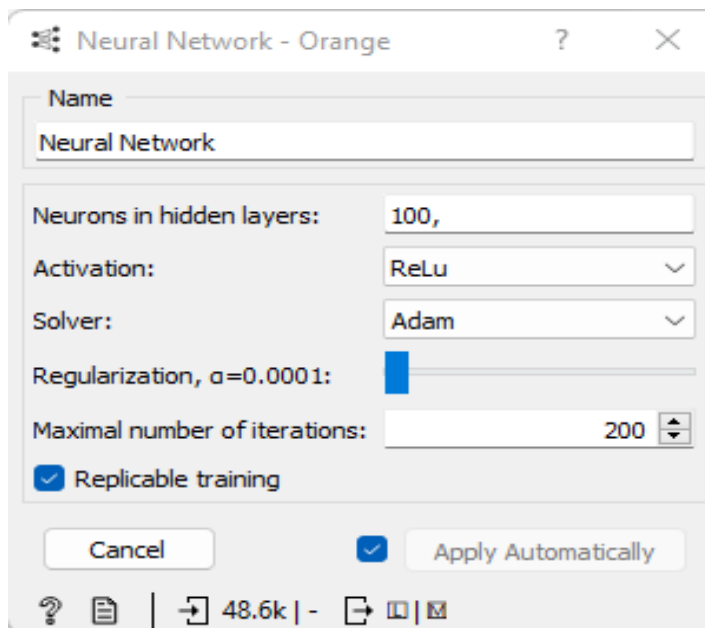
Fig:awards\_won? v/s is\_promoted

From this graph, it is clear that about 45% of employees who have won at least 1 award get promoted. It suggests that winning an award may be an important factor in an employee's promotion. However, it is important to note that there may be other factors at play as well, such as job performance, skills, potential for growth and development, and organizational needs. Additionally, it is possible that the relationship between winning an award and getting promoted is not causal, but rather that employees who are high performers and have demonstrated potential for growth and development are more likely to both win awards and get promoted. Ultimately, promotion decisions should be based on a fair and transparent evaluation process that considers a range of factors, rather than on any single factor alone.

## 5.PREPROCESSING AND FEATURE SELECTION

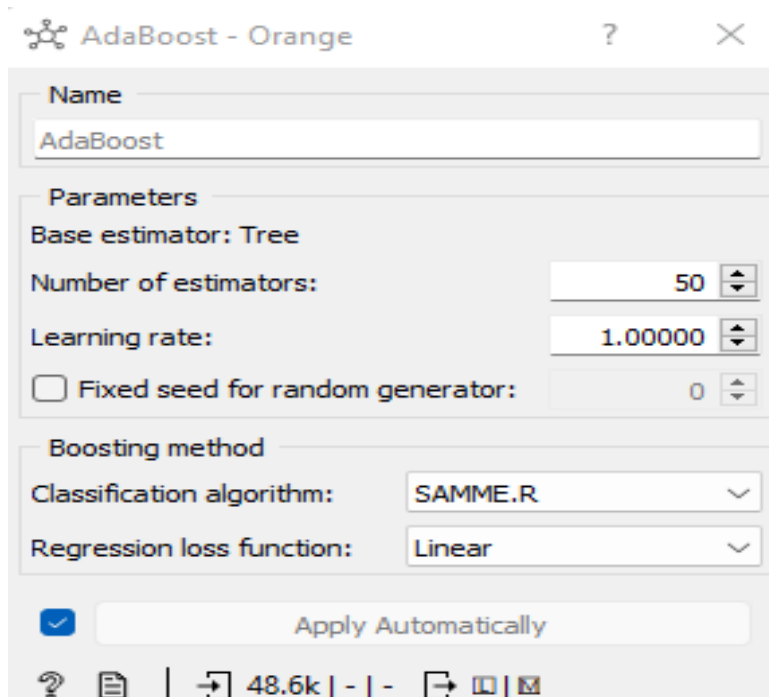
- 1) Load the dataset: Open the Orange tool and load the employee promotion dataset by selecting "File" -> "Open" and selecting the dataset.
- 2)set the target variable in this dataset is\_promoted is the target variable.
- 3) Check for missing values: Use the "Data Table" widget to visualize the data and identify if there are any missing values in the dataset. If there are any missing values, handle them by selecting "Impute" widget and remove the missing values.
- 4) Perform feature selection: Use the "Select Columns" widget to select the relevant columns for analysis and remove any irrelevant or redundant columns. In this dataset there are three columns were removed. The removed features are employee\_id,region ,recruitment\_channel.
- 5) Check for outliers: Use the "Box Plot" widget to identify any outliers in the dataset. If there are any outliers, handle them by removing outliers.





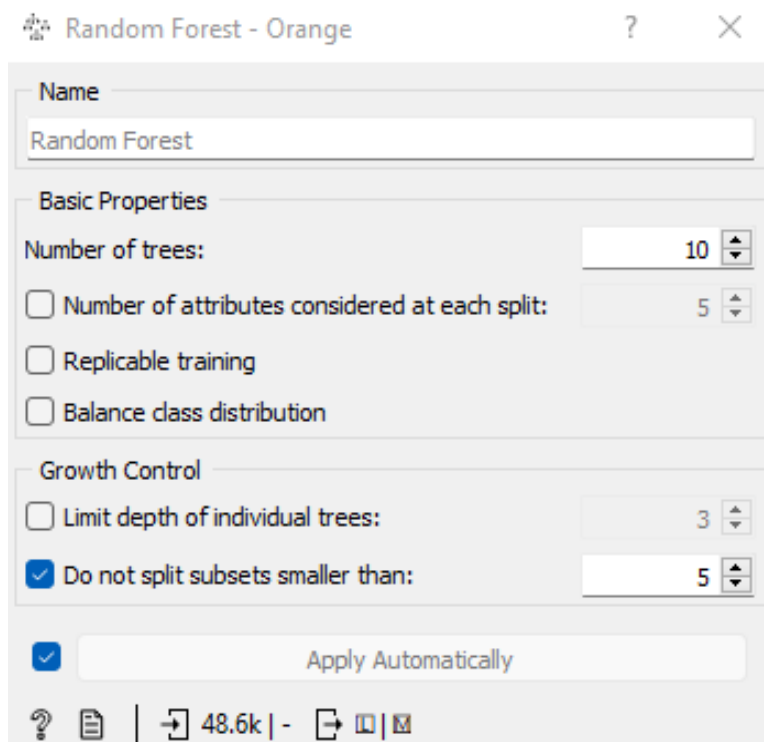
## 2)Adaboost

An Adaboost model for this employee promotion dataset may be generated quickly and easily in Orange by following a few straightforward steps. To begin, the dataset needs to be loaded, and then it needs to be preprocessed so that it can be used in the modelling. After that, one of the available versions of the Adaboost model may be selected, and its settings can be customized by changing the amount of base classifiers, the learning rate, and the stopping criterion. Following this, the model may be "trained" on the training data and "evaluated" using a variety of metrics to see how well it performs on the validation data. In conclusion, the model may, if necessary, undergo additional tuning in order to improve the accuracy of its predictions. In general, Orange offers a straightforward interface for the creation and evaluation of Adaboost models relevant to the employee promotion dataset.



### 3) Random Forest

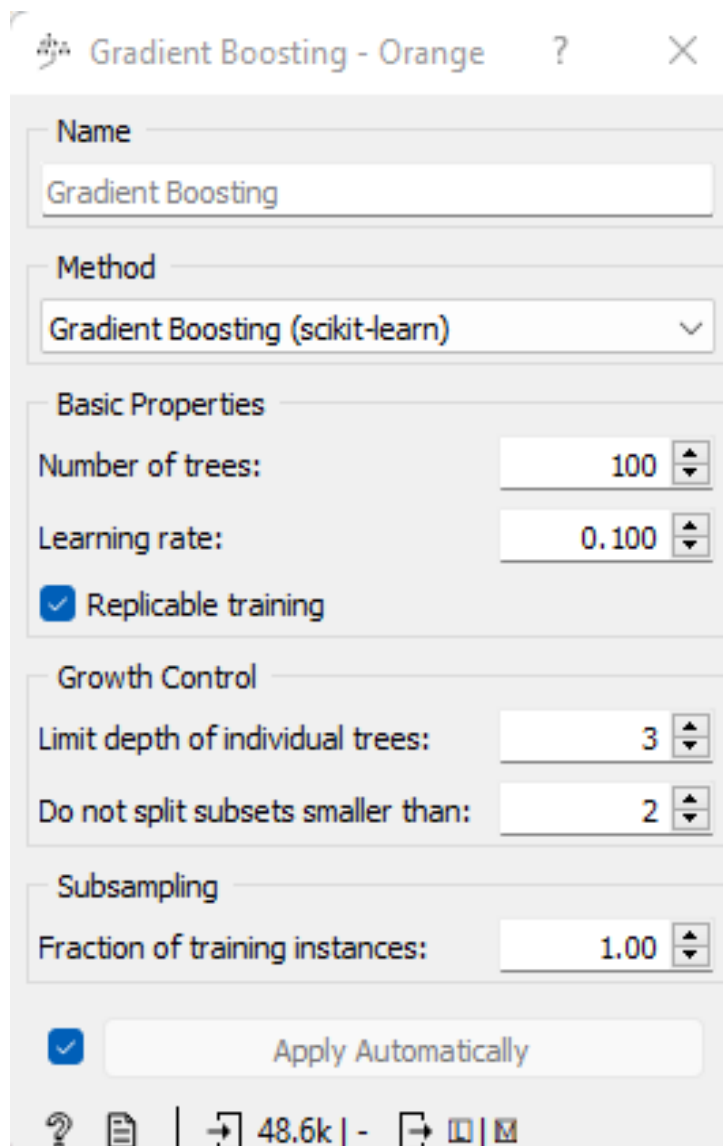
For this employee promotion dataset in Orange, a random forest model can be made by taking a few easy steps. First, the dataset needs to be loaded and preprocessed so that it is ready to be used for modelling. After selecting the random forest model from the list, set the number of trees, maximum depth, and splitting criteria. Training and testing the model with training and validation data follows. Its performance can be judged using different metrics. Finally, the model can be adjusted for accuracy. Orange makes creating and testing random forest models for employee promotion dataset easier.



### 4) Gradient boosting

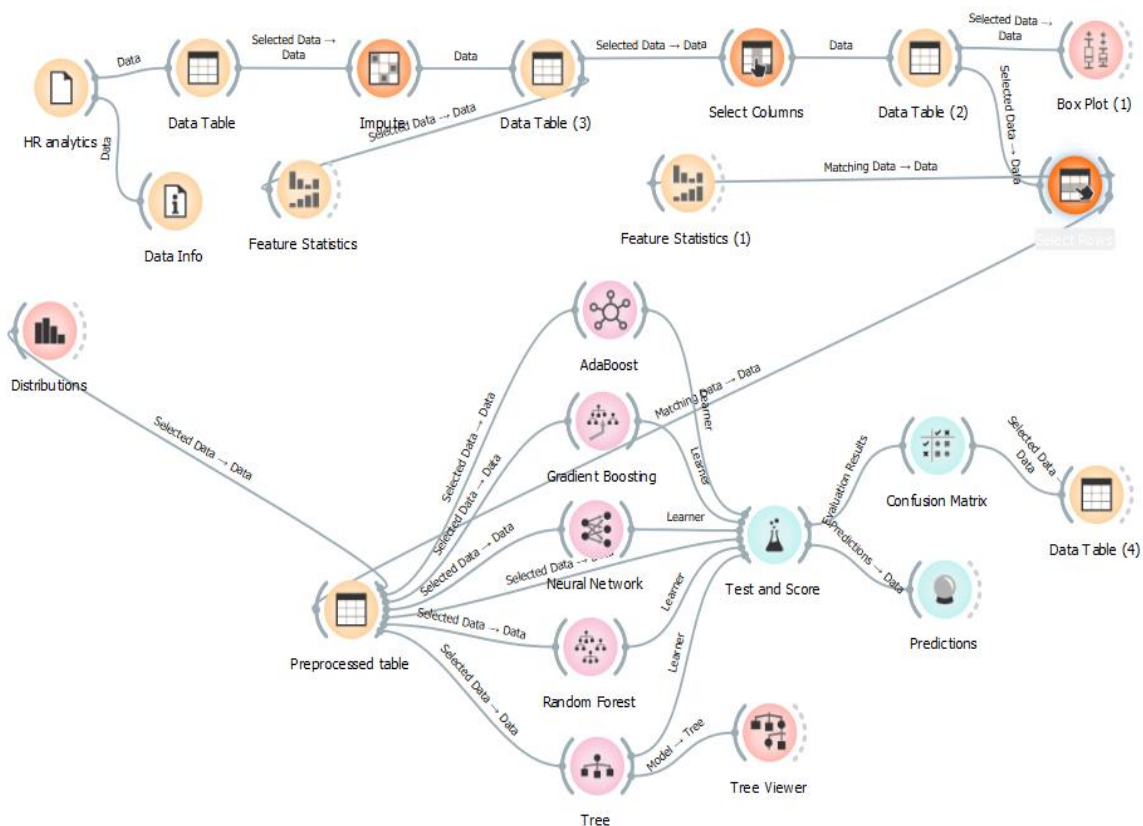
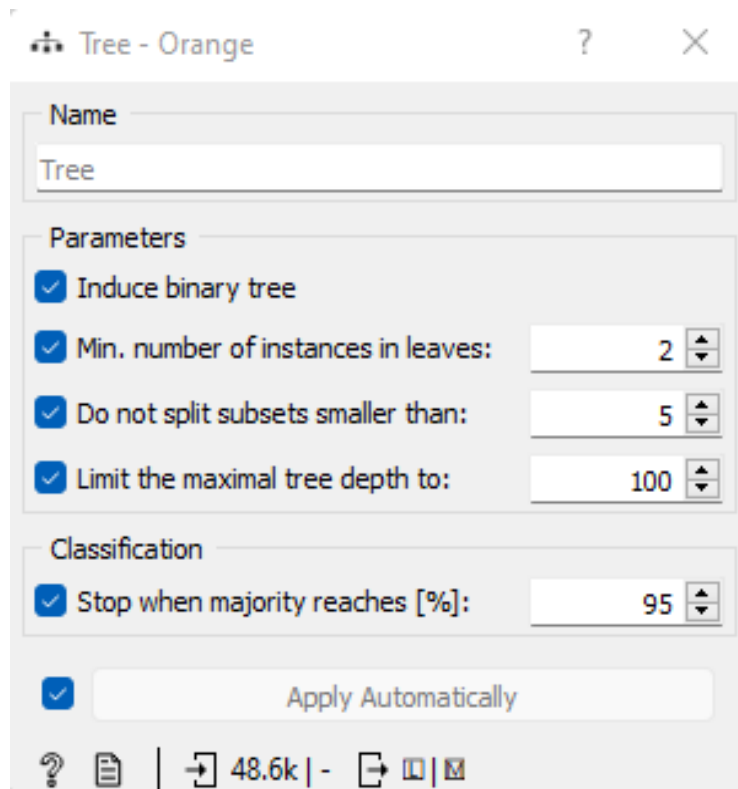
Gradient boosting can assess the employee promotion dataset. Orange's Gradient Boosting widget may train a model by adding decision trees to decrease the previous tree's error. To optimise model performance, change the learning rate and tree count. Cross-validation and Orange widgets like Model Accuracy may evaluate and illustrate the model. Gradient boosting can help analyse the employee promotion dataset and discover contributing elements.





## 5)Tree

Tree models can be used by Orange to conduct an analysis of this employee promotion dataset, which contains both numerical and categorical data. The Classification Tree widget has the ability to construct a decision tree model that, given an employee's age, education level, and performance evaluations, may determine whether or not the employee will be promoted. Using the widget, select the variable to divide on as the target, the maximum tree depth, and the splitting criterion. The Tree Viewer widget illustrates the model's interpretation of the tree it was given. In addition to this, the model can be evaluated using cross-validation and orange widgets such as the confusion matrix and the ROC curve. The elements that go into employee promotions can be better understood and predicted with tree models.



*Fig :Model building for Employee promotion dataset*

## 7.PERFORMANCEN AND EVALUATION OF MODELS

The dataset ‘HR analytics: Employee promotion dataset’ consist of 54808 instances and 13 features. Out of 13 features in train data, there are 6 numeric variables and 7 are categorical variables. The selected target variable is ‘is\_promoted’. Models applied in the datasets are Tree, Random Forest, Neural network, Gradient Boosting and Adaboost. The performance value obtained for each model is given below

Model	AUC	CA	F1	Precision	Recall
Tree	0.684	0.932	0.922	0.923	0.932
Random Forest	0.751	0.935	0.922	0.927	0.935
Neural Network	0.807	0.942	0.929	0.941	0.942
Gradient Boosting	0.810	0.941	0.927	0.941	0.941
AdaBoost	0.729	0.915	0.908	0.903	0.915

Among the applied models, Neural Network has highest accuracy which is 94.2%. The confusion matrix of Neural network is given below

		Predicted		$\Sigma$
		0	1	
Actual	0	44258	119	44377
	1	2721	1508	4229
$\Sigma$		46979	1627	48606

All the features except employee\_id,region ,recruitment\_channel are necessary for the Model building for the dataset.

## 8.RESULT AND DISCUSSION

The employee promotion dataset can provide valuable insights into which factors contribute to employee promotion and help organizations make informed decisions about employee development and career advancement. Using Orange Tool, you can build predictive models that can help identify which employees are more likely to be promoted based on their characteristics and job-related factors. By understanding the factors that contribute to employee promotion,

organizations can take steps to ensure a fair and equitable promotion process and improve employee retention. Here among the applied models, Neural Network has highest accuracy which is 94.2%.

## 9.CONCLUSION

In conclusion, constructing an employee promotion prediction model is difficult and involves careful consideration of many parameters. First, comprehend the dataset, problem statement, and variables that may affect the outcome. Model building encompasses data cleaning and preprocessing, feature engineering, model selection, and hyperparameter tuning. Choose a model that captures data patterns and relationships without overfitting. Trees, random forests, Neural network and gradient boosting can predict employee promotions. Interpretability, accuracy, and scalability determine the model. The model's accuracy, precision, recall, and F1 score should be assessed. Cross-validation helps the model generalize to fresh data. Lastly, install a model in production and monitor its performance. This comprises integrating the model into existing systems, verifying its accuracy and reliability, and updating as data and requirements change

There are several things that can be learned from the employee promotion dataset. First, people are more likely to be promoted if they have higher performance ratings, more experience, and better education. Second, the rate of promotion is different for men and women, with men getting promoted more often than women. Third, the promotion process may be biased based on race or department, since some groups have lower promotion rates than others. Lastly, the data suggests that some factors, like age and length of service, may not have a big effect on promotion rates. Overall, the employee promotion dataset gives useful information about the factors that affect employee promotion and shows where the promotion process could be improved. Among the applied models, Neural Network has highest accuracy which is 94.2%.

## 10.REFERENCES

- 1)<https://www.kaggle.com/datasets/arashnic/hr-ana?select=test.csv>
- 2)<https://blog.vantagecircle.com/employee-promotion/#:~:text=Vantage%20Rewards%20today-.What%20is%20Employee%20Promotion%3F.and%20loyalty%20towards%20an%20organization.>
- 3) [https://en.wikipedia.org/wiki/Orange\\_\(software\)](https://en.wikipedia.org/wiki/Orange_(software))
- 4)<https://www.linkedin.com/pulse/promotion-dilemma-case-study-dinesh-g-lakhara>
- 5)<https://www.geeksforgeeks.org/data-preprocessing-in-data-mining/>
- 6)[https://www.popcouncil.org/uploads/pdfs/events/2010MENAWkshop\\_10.pdf](https://www.popcouncil.org/uploads/pdfs/events/2010MENAWkshop_10.pdf)