

DATASET HUMAN CAPITAL

By Hepta Scientist Group



Mentor



Ari Sulistiyo Prabowo



Members of Hepta Scientist Group



Dias Indah
Melisawati

Pendidikan Matematika
Universitas Negeri Semarang



Siti Hafsah

Teknik Informatika
President University



Louis Madaerdo
Sotarjua

Teknik Elektro
Universitas Singaperbangsa
Karawang



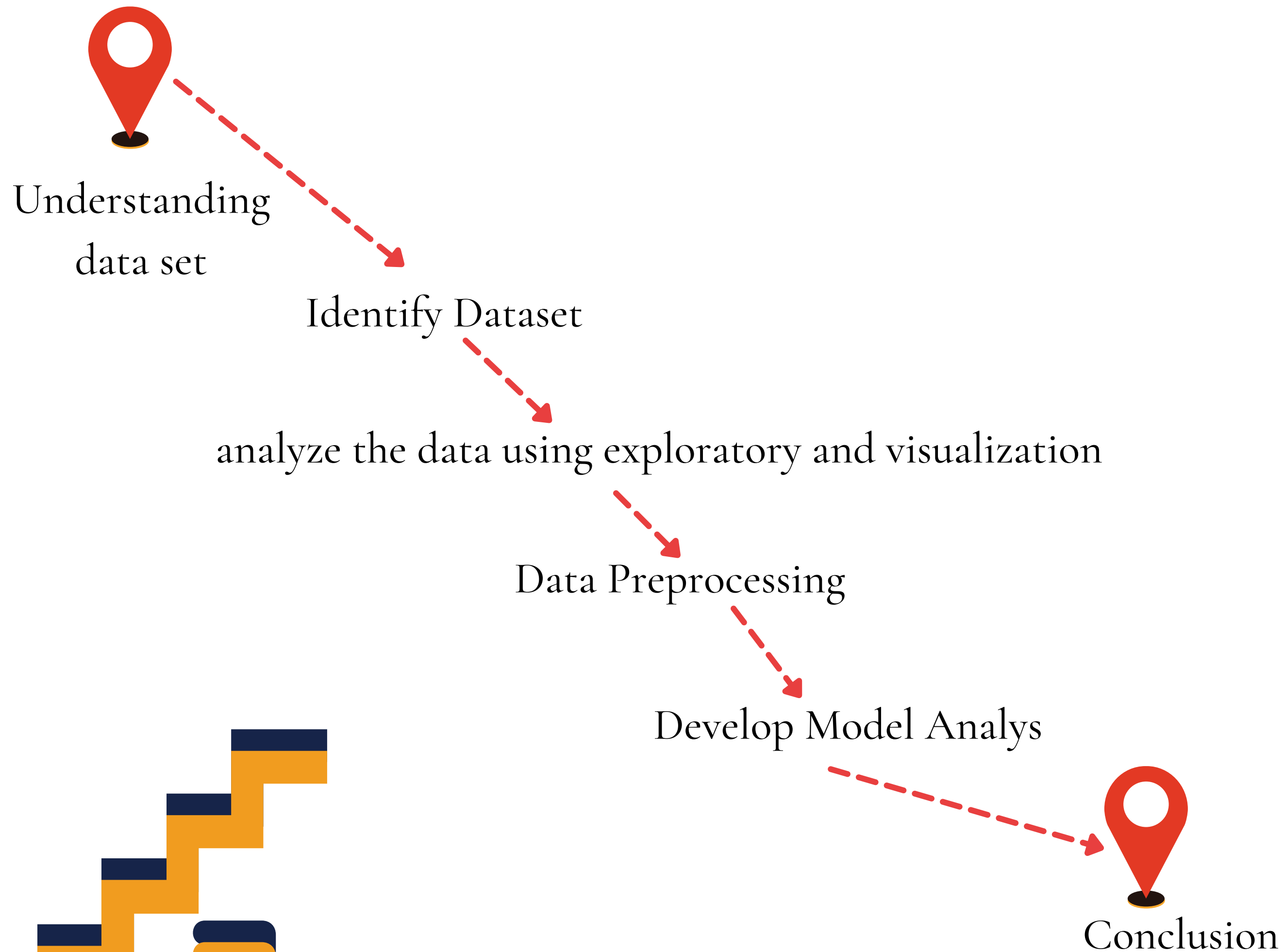
Siti Hamidah

Statistika
Universitas Padjadjaran



Taqiyuddin Yazid
Zaidan

Teknik Elektro
Universitas Pendidikan Indonesia



STEP

BY

STEP



UNDERSTANDING DATASET HUMAN CAPITAL

1



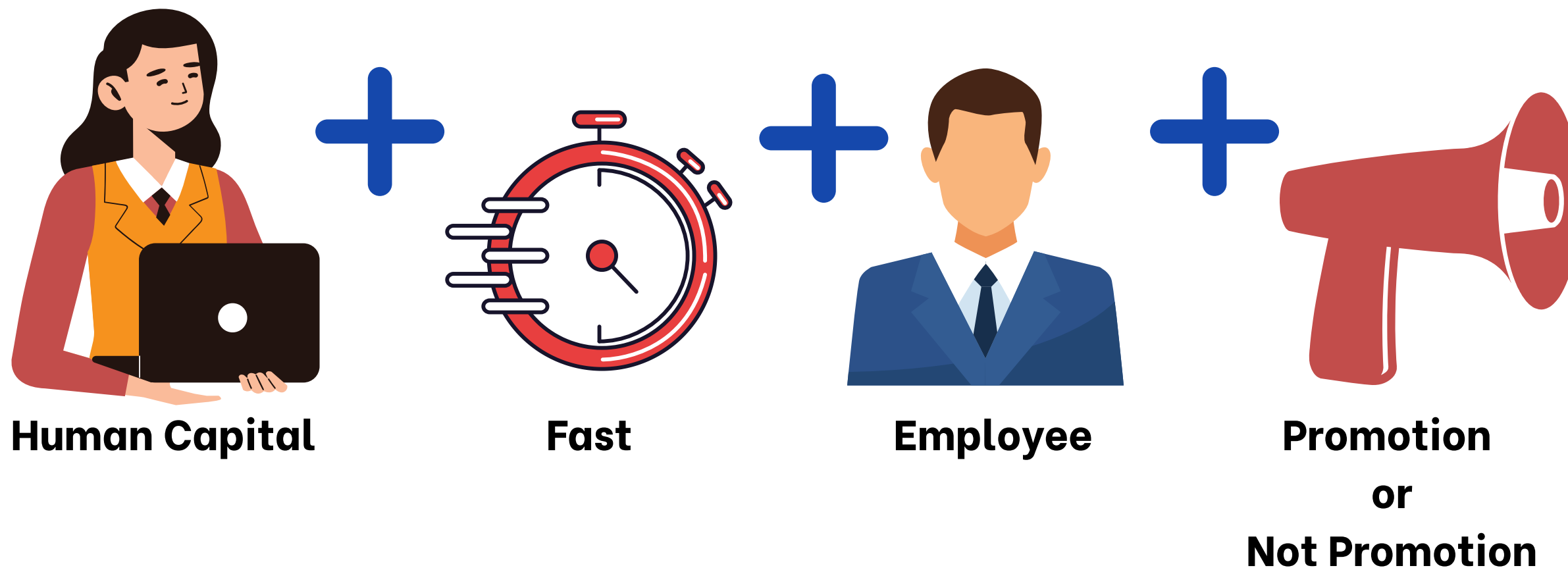
Definition of Human Capital

What is Human Capital?

Human capital is perceived to increase productivity and thus profitability. The more investment a company makes in its employees, the chances of its productivity and success becomes higher.



Problem of Human Capital Dataset



Model Machine Learning

NOTES!!!

How can human capital can classify quickly and precisely which employees belong to the promoted category and which are not promoted.

The variables gender and promoted has abbreviation forms as follows:



M : Male



F : Female



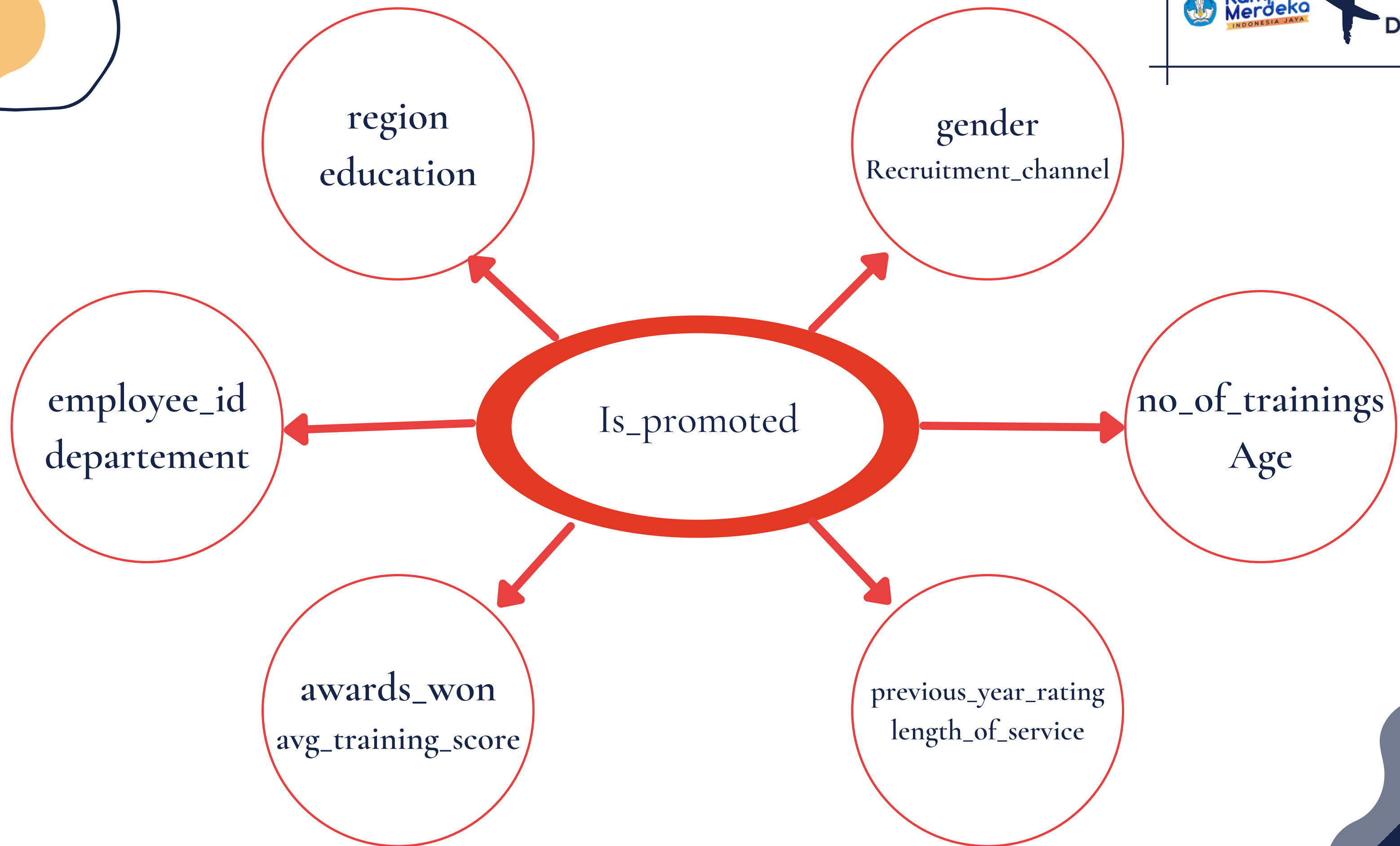
o : not promoted



1 : is promoted

Gender

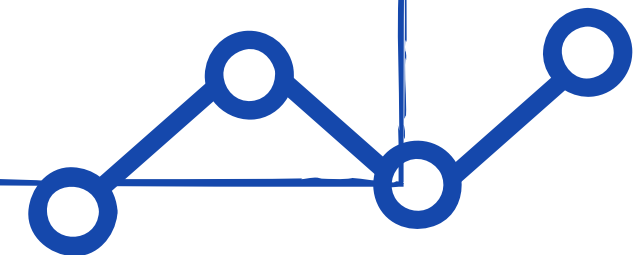
Promoted



IDENTIFY DATASET HUMAN CAPITAL



2

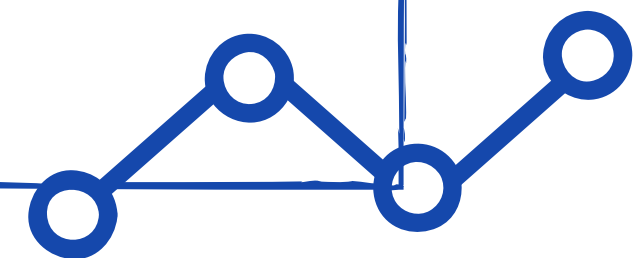


Identify Which Activities Should Be Done

To simplify the work of Human Capital in accelerating the classification of who is promoted and who is not promoted, is to create a machine learning model. There are many algorithms that can be applied in machine learning classification models, including: KNN, Decision Tree, Random Forest, Cross Validation.

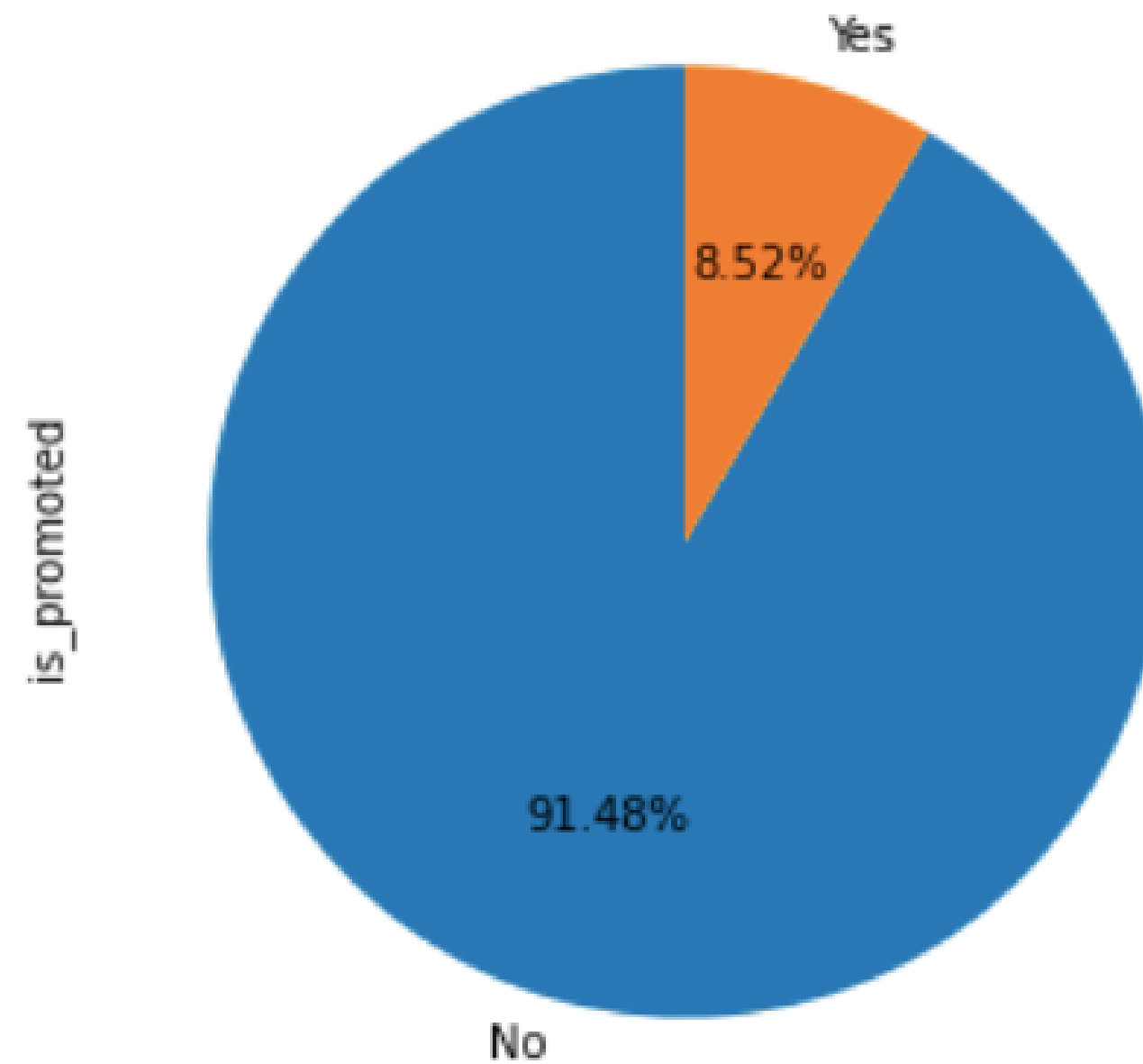
ANALYZE THE DATA USING EXPLORATORY AND VISUALIZATION

3



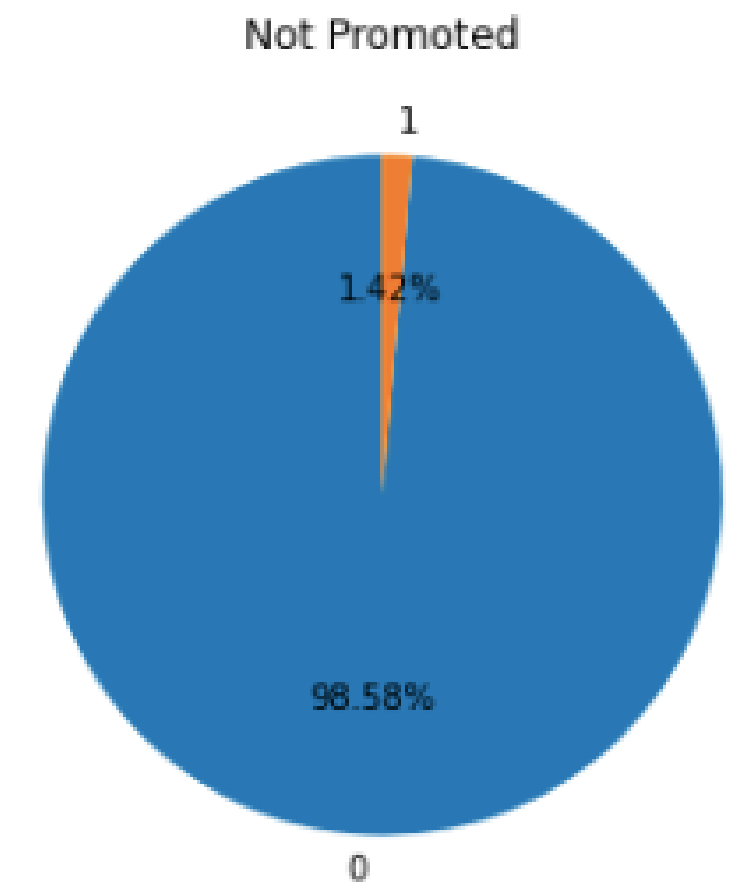
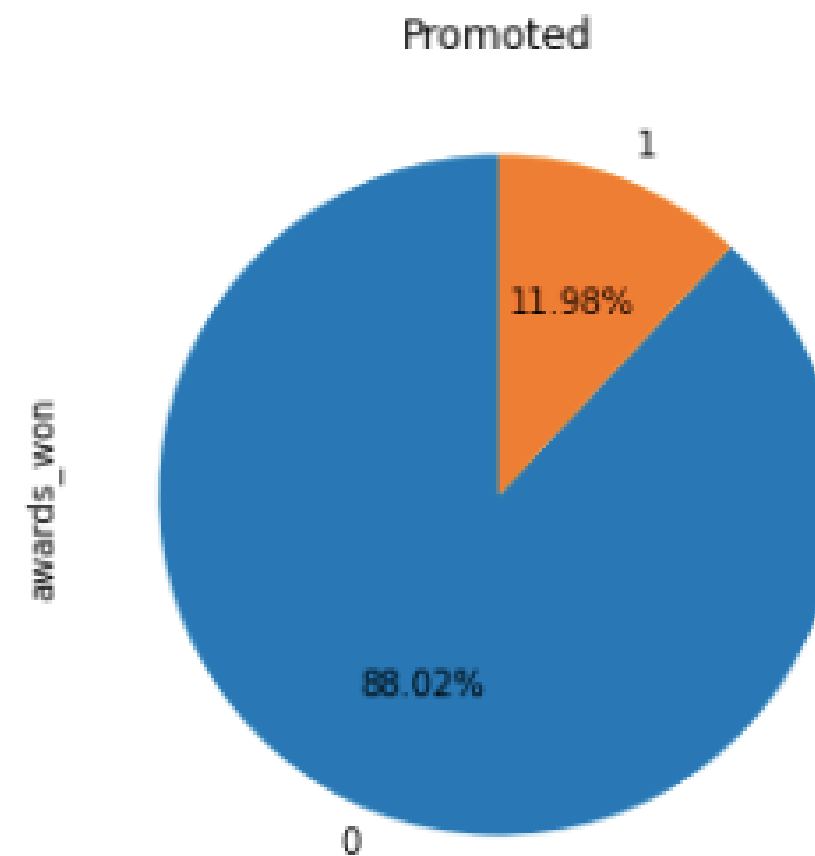
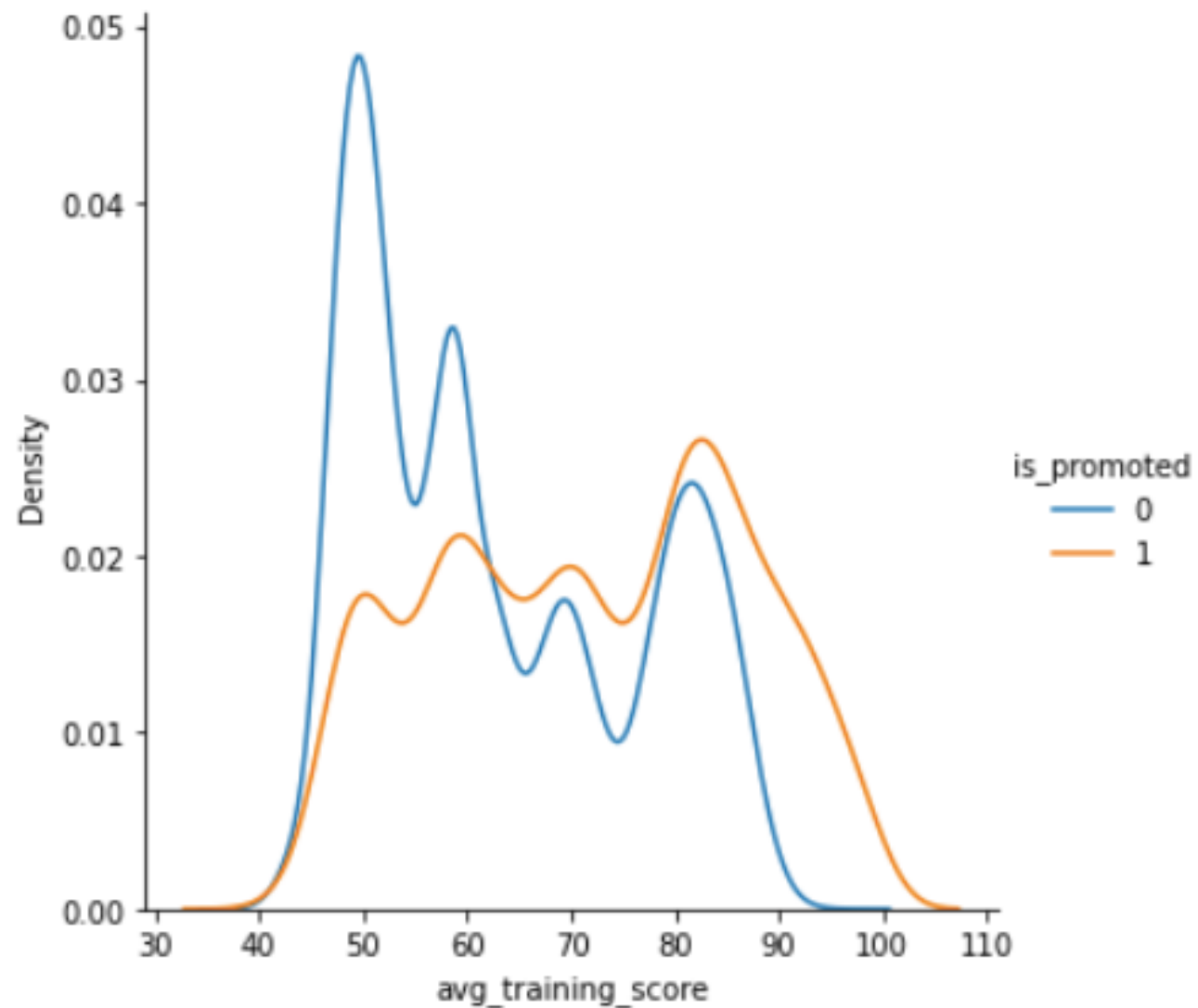
Composition of Dependent Variables

Pie Chart : Is Promoted

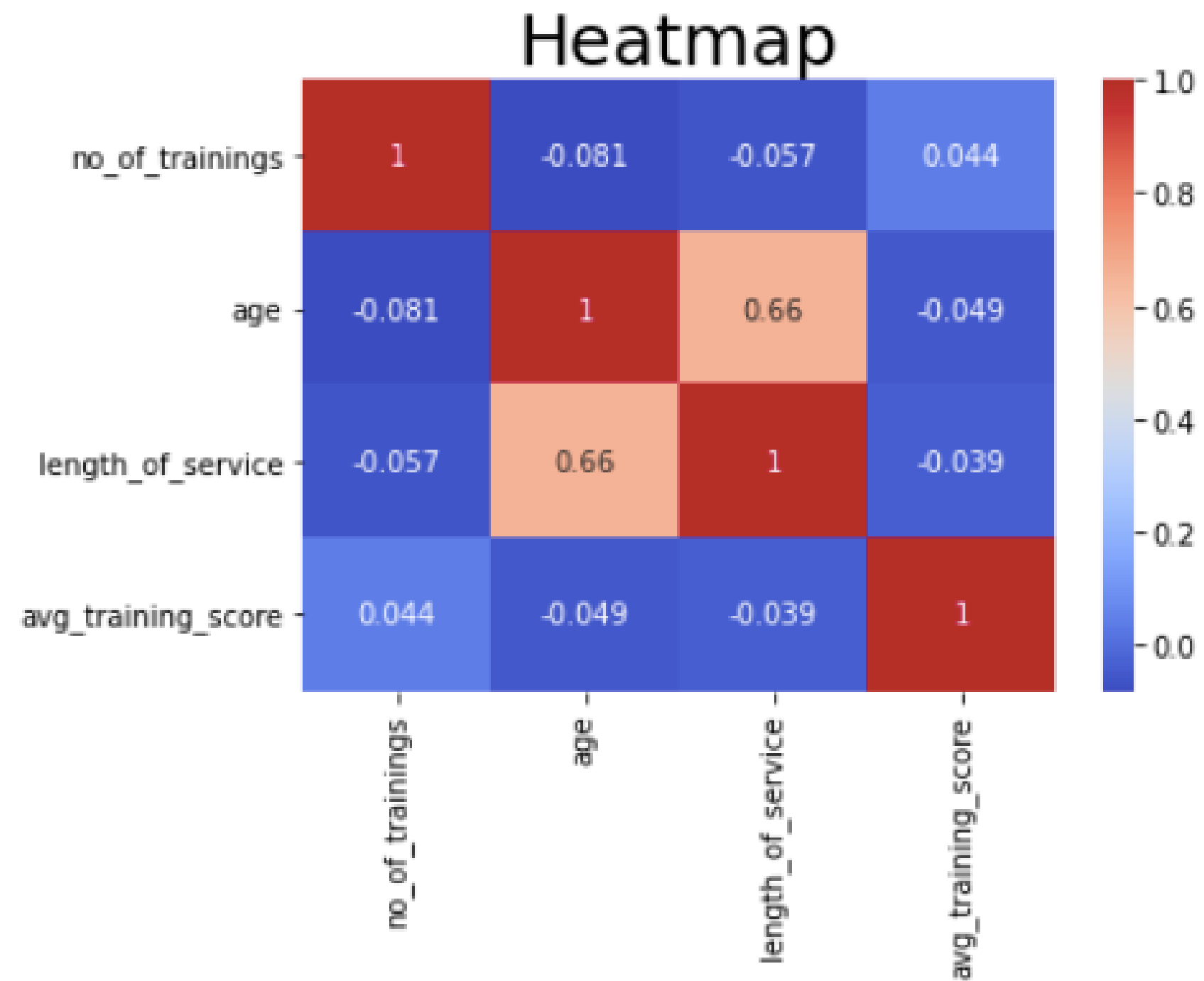


List of variables that have the same pattern either promoted or not promoted

- Average Training Score
- Awards Won
- Region
- Previous Year Rating
- Department

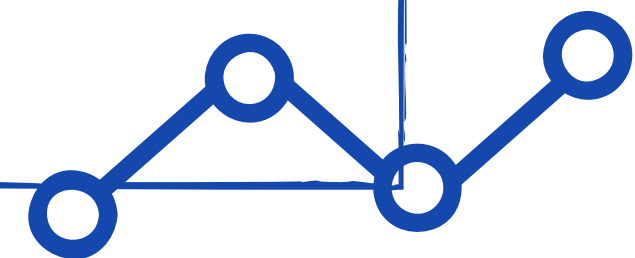


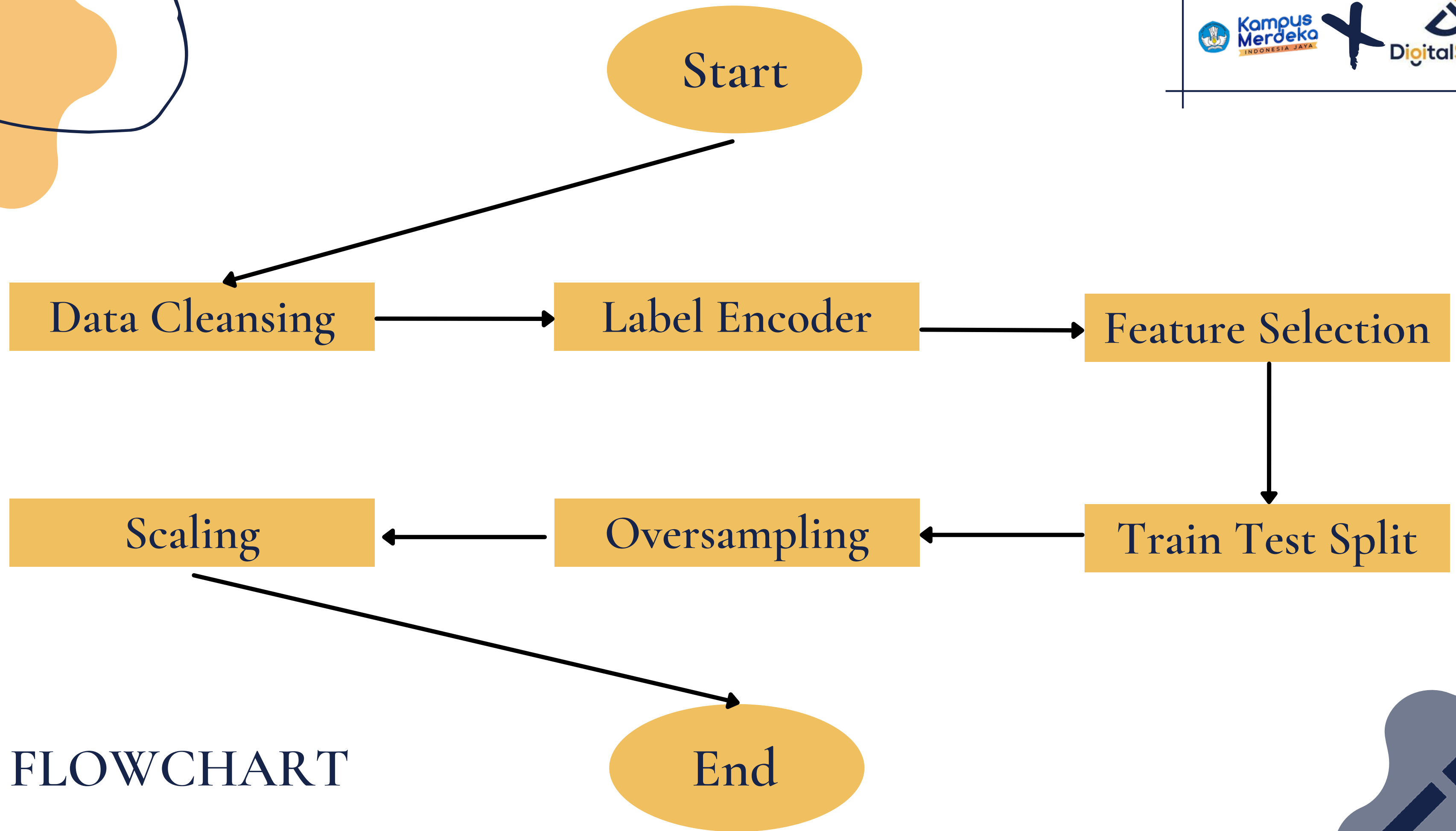
Heatmap Correlation



DATA PREPROCESSING

4





FLOWCHART



DEVELOP MODEL

5

Cross Validation

	KNN	Decision Tree	Random Forest
Accuracy	0.995	0.917	0.928
Precision	0.996	0.926	0.942
Recall	0.994	0.906	0.912
F1 score	0.995	0.916	0.927
Specificity	0.996	0.928	0.944

Picture of Training Performance

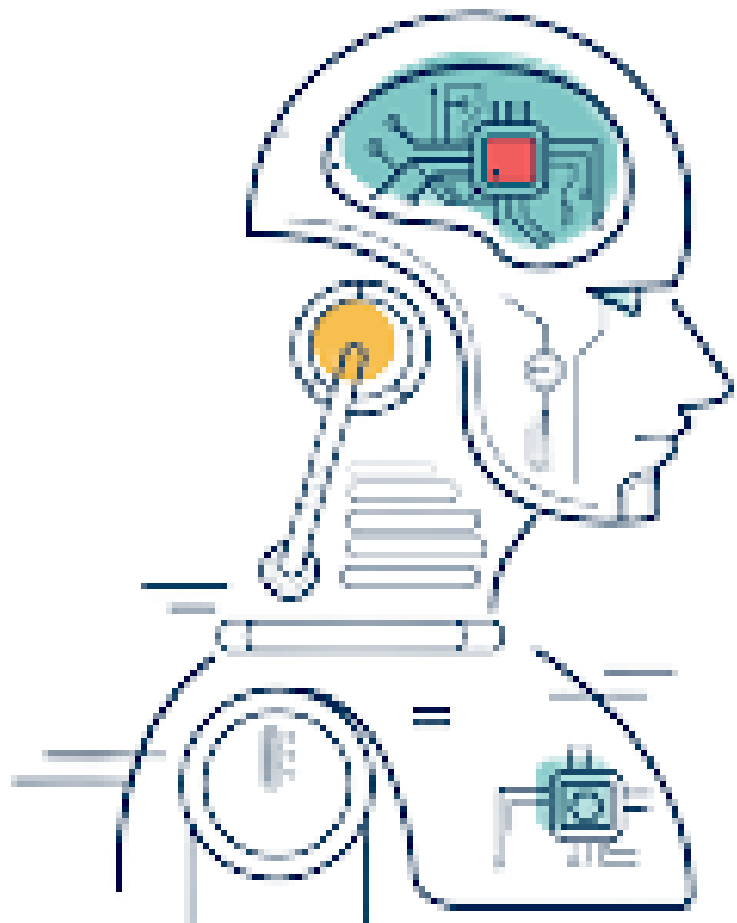
	KNN	Decision Tree	Random Forest
Accuracy	0.891	0.875	0.891
Precision	0.909	0.881	0.909
Recall	0.868	0.868	0.868
F1 score	0.888	0.874	0.888
Specificity	0.914	0.883	0.914

Picture of Testing Performance

CONCLUSION

Conclusion

The recommended machine learning model to use is
Random Forest



Order of variables that determine employee promotion



1. Previous Year
Rating



2. Average
Training Score



3. Region



4. Age



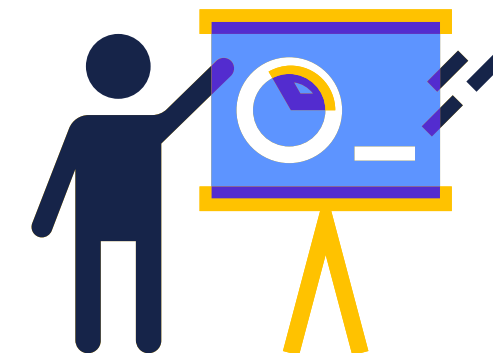
8. Awards Won



7. Education



6. Length of Service



5. Number of Trainings

THANK
YOU

QA

