DATASET HUMAN CAPITAL

By Hepta Scientist Group

Hepta Scientist Group

Ari Sulistiyo Prabowo
As Mentor

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 data set

Identify Dataset

analyze the data using exploratory and visualization

Data Preprocessing

Develop Model Analys

Conclusion

UNDERSTANDING DATASET HUMAN CAPITAL

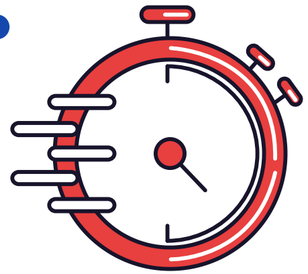
1

Definition of 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

Human Capital



Fast

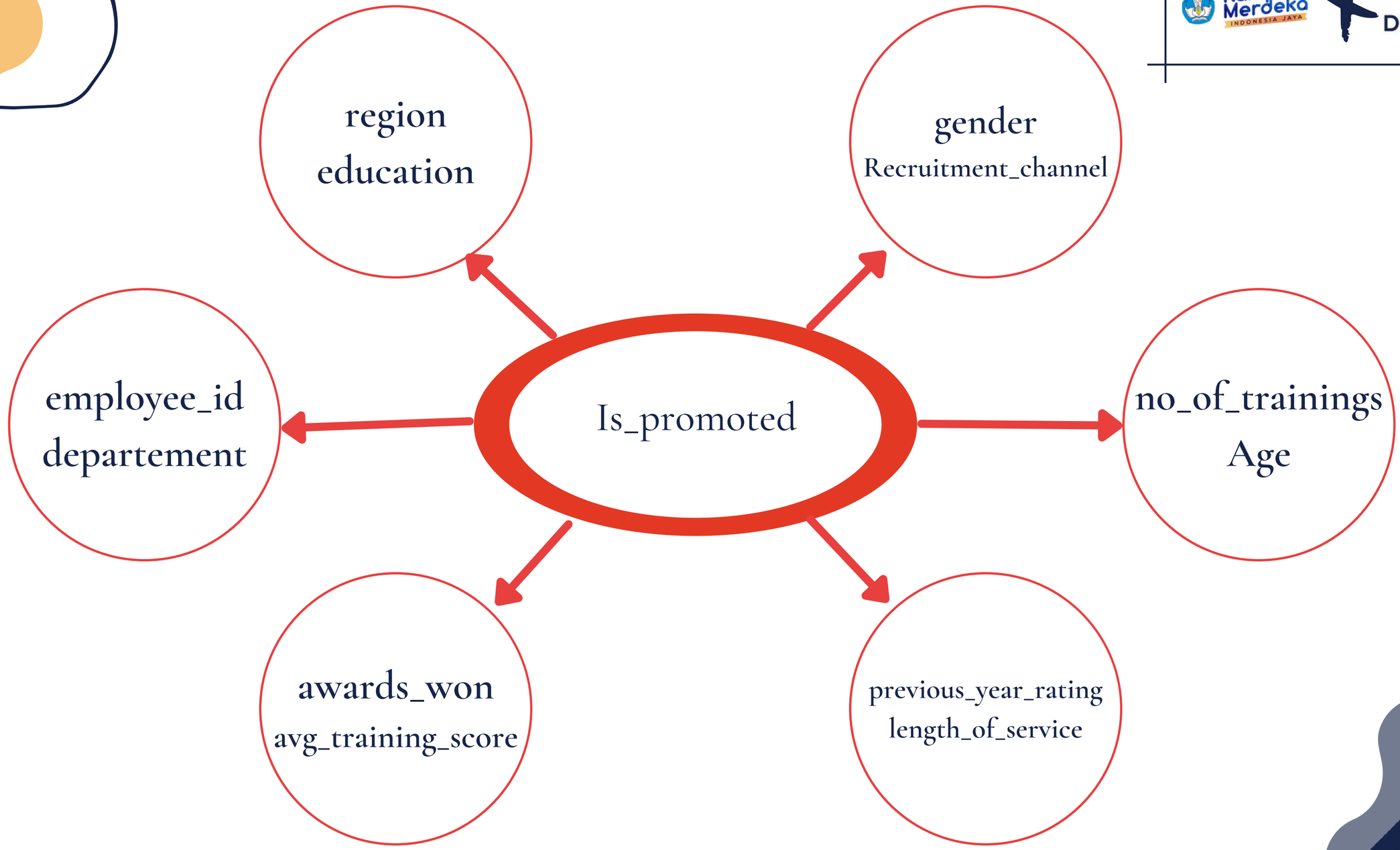
Employee

Promotion
or
Not Promotion

= 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.


region
education
gender
Recruitment_channel
employee_id
departement
Is_promoted
no_of_trainings
Age
awards_won
avg_training_score
previous_year_rating
length_of_service

IDENTIFY DATASET HUMAN CAPITAL

2

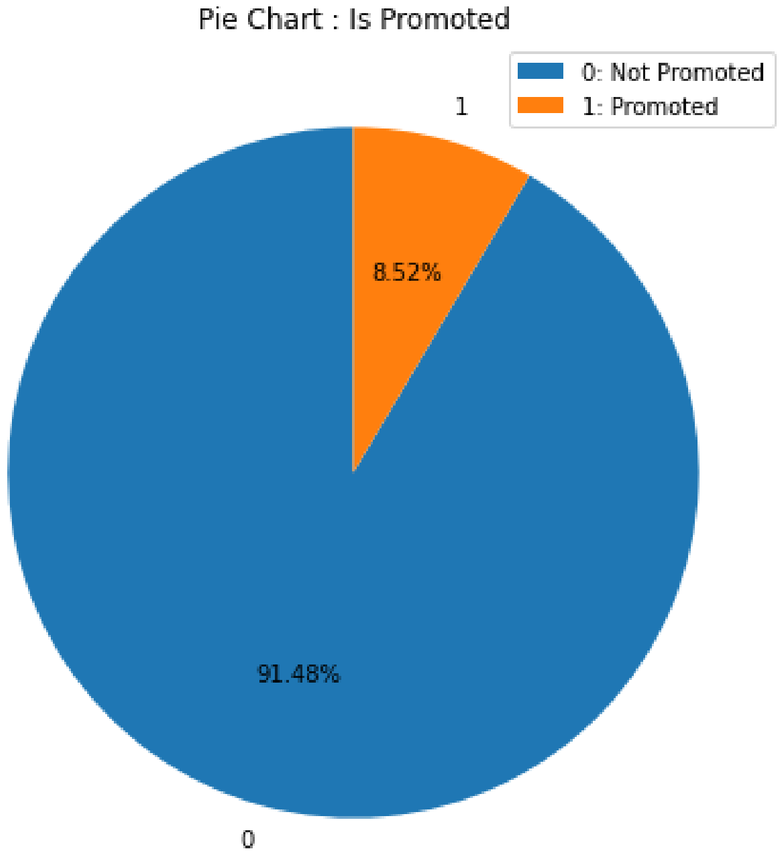
Identify Which Activities Should Be Done

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 are classification models, including: KNN, Decision Tree, and Random Forest.

ANALYZE THE DATA USING EXPLORATORY AND VISUALIZATION

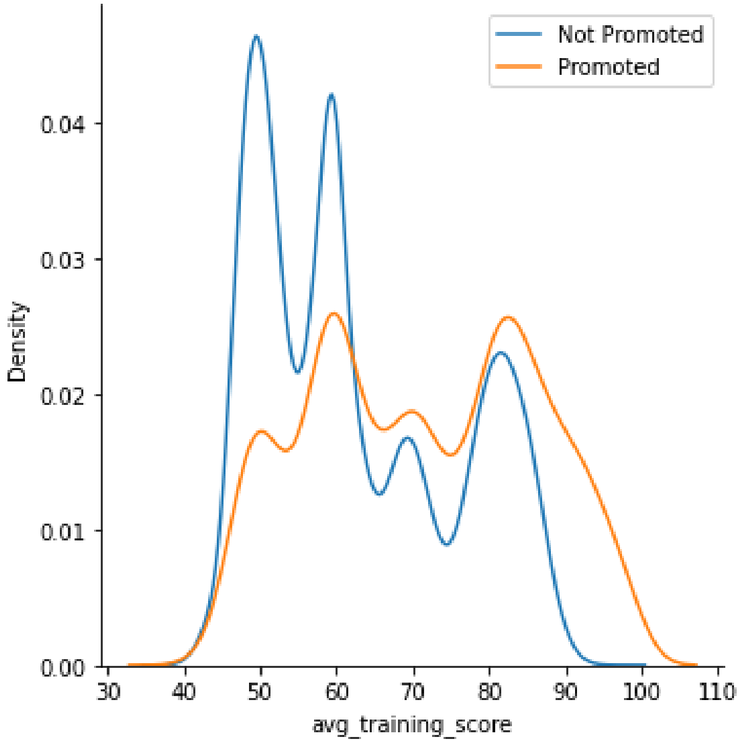
3

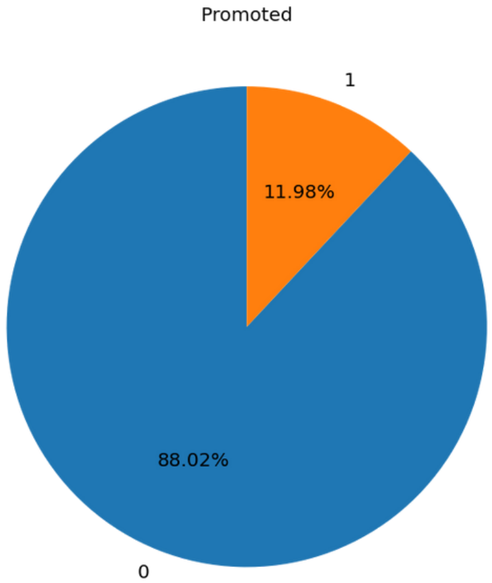
Composition of Dependent Variables

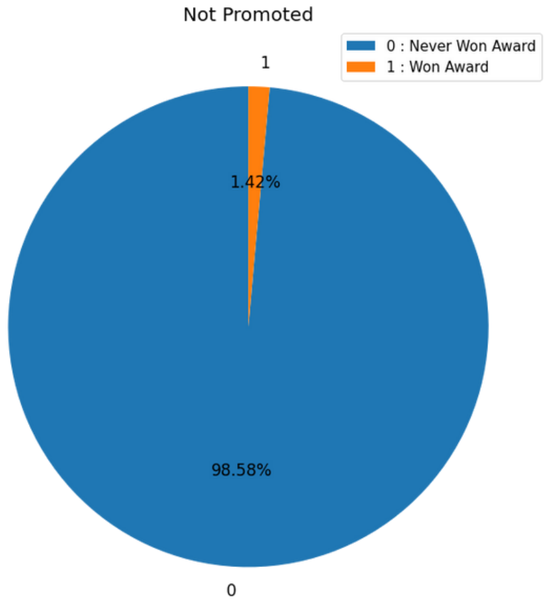


List of Variables that have the different pattern either promoted or not promoted

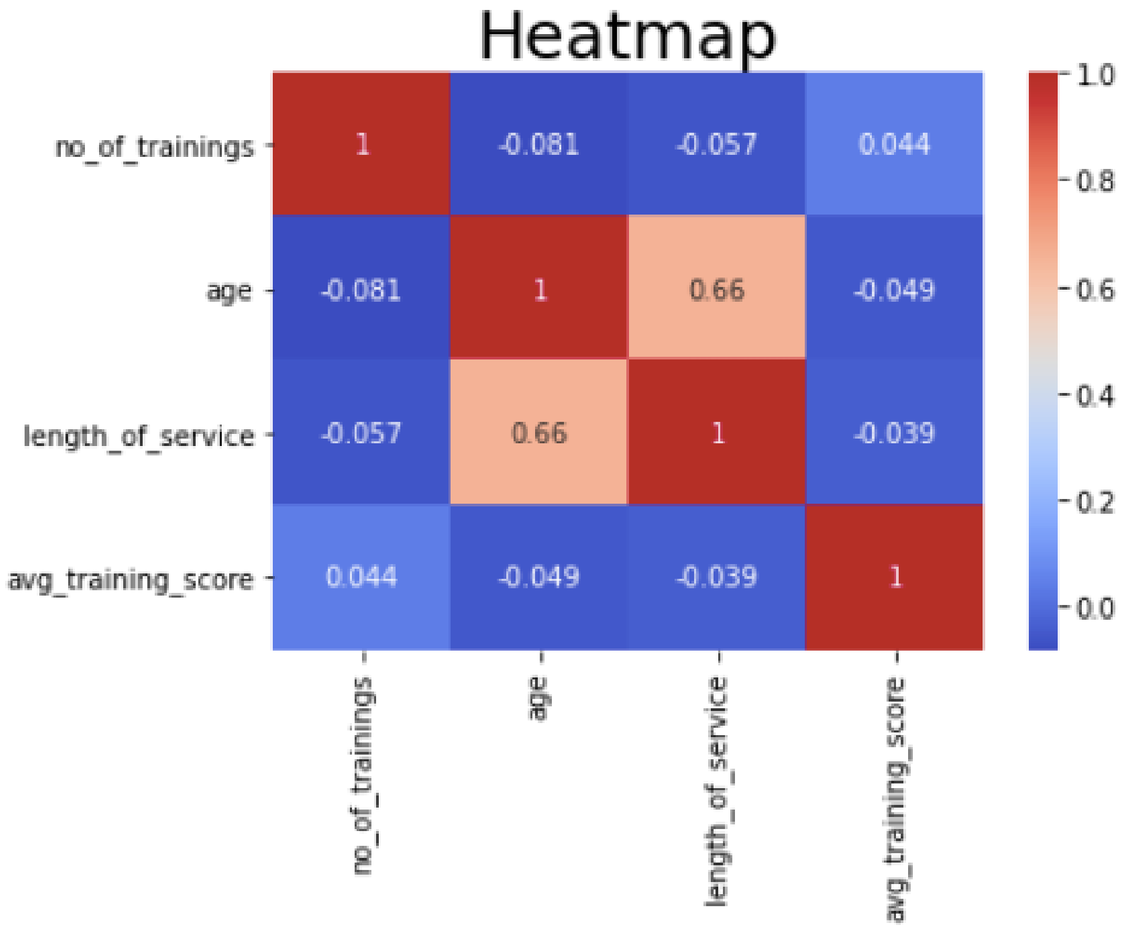
- Average Training Score
- Awards Won
- Region
- Previous Year Rating
- Department





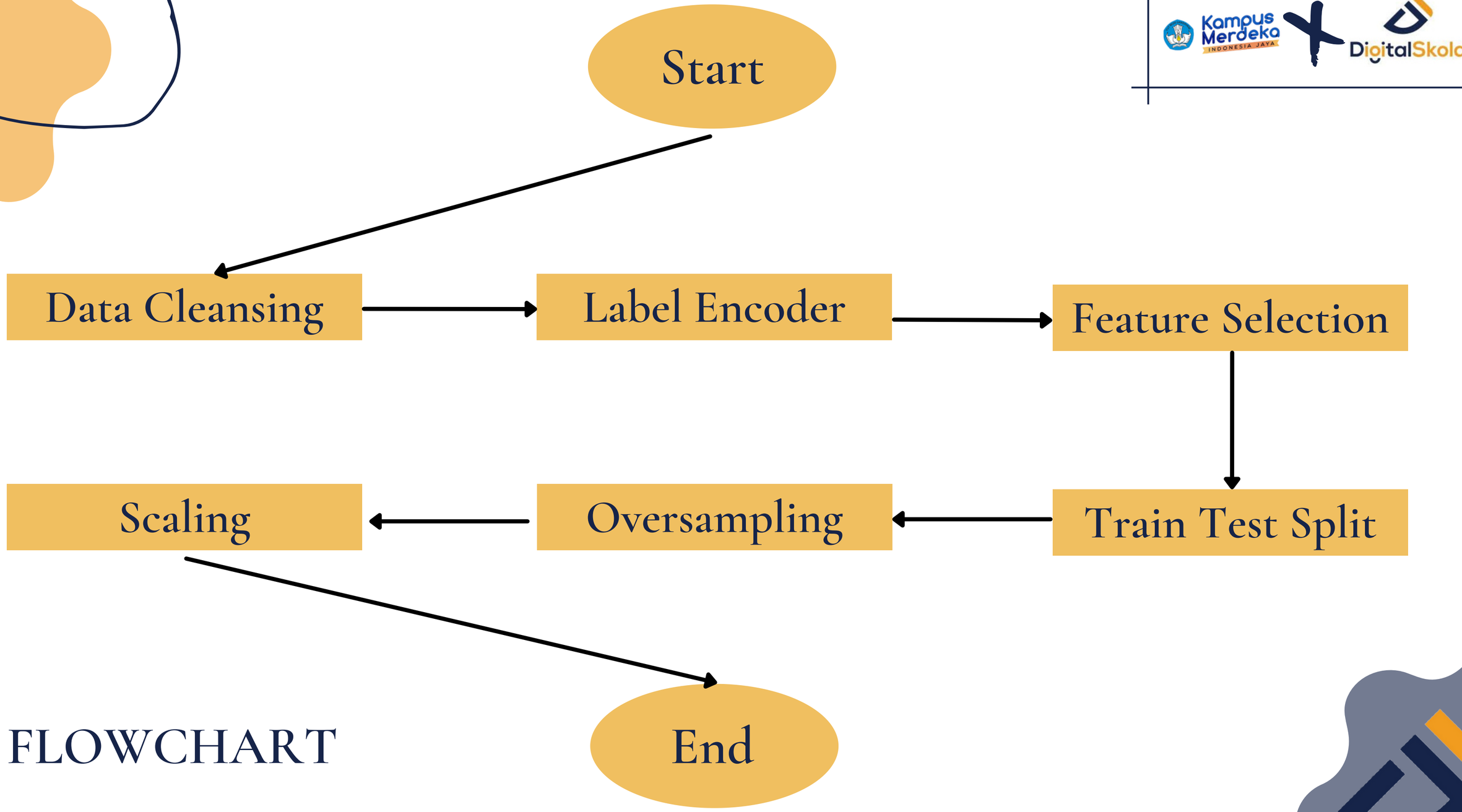


Heatmap Correlation



DATA PREPROCESSING

4


Start
Data Cleansing
Label Encoder
Feature Selection
Train Test Split
Oversampling
Scaling
End

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

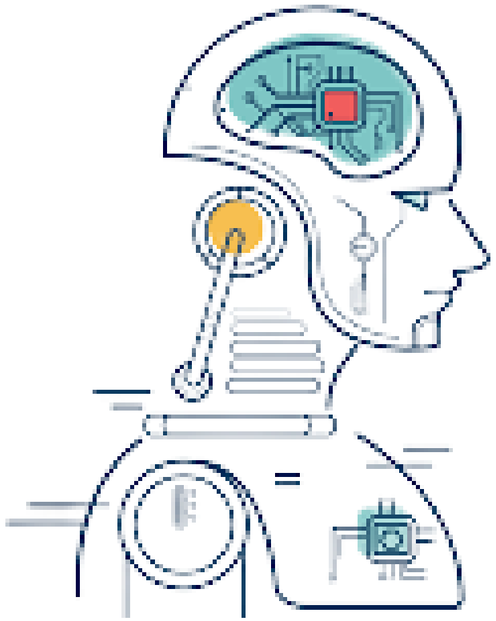
Picture of Testing Performance

CONCLUSION

6

Conclusion

The recommended machine learning model to use is Random Forest



List order of variables that determine employee promotion

1. Previous Year Rating

2. Average Training Score

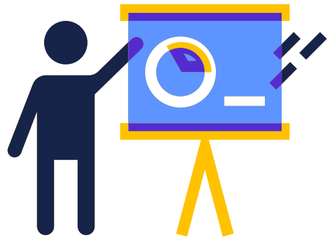
3. Region

4. Age

8. Awards Won

7. Education

6. Length of Service



5. Number of Trainings


THANK
YOU


Q
A