Job Retention

PREPARED BY

Shoroog Alsubhi

Problem Description

Problem Description

The job retention prediction project aims to analyze the employees' information in order to build a machine learning models.

Design the Solution

Design the Solution

01

Company employee's information dataset.

02

Analyze this information.

03

Build a machine learning model

Dataset

Dataset

The " HR_comma_sep" dataset is provided by Kaggle platform

The dataset consists of 14999 data points with 10 features.

After conducting an analysis of these features, it was found that eight of them are significant to use on the models.

A dataset sample

satisfaction_level	last_evaluation	number_project	average_montly_hours	time_spend_company	Work_accident	left	promotion_last_5years	Department	salary
0.38	0.53	2	157	3	0	1	0	sales	low
0.80	0.86	5	262	6	0	1	0	sales	medium
0.11	0.88	7	272	4	0	1	0	sales	medium
0.72	0.87	5	223	5	0	1	0	sales	low
0.37	0.52	2	159	3	0	1	0	sales	low

Algorithm

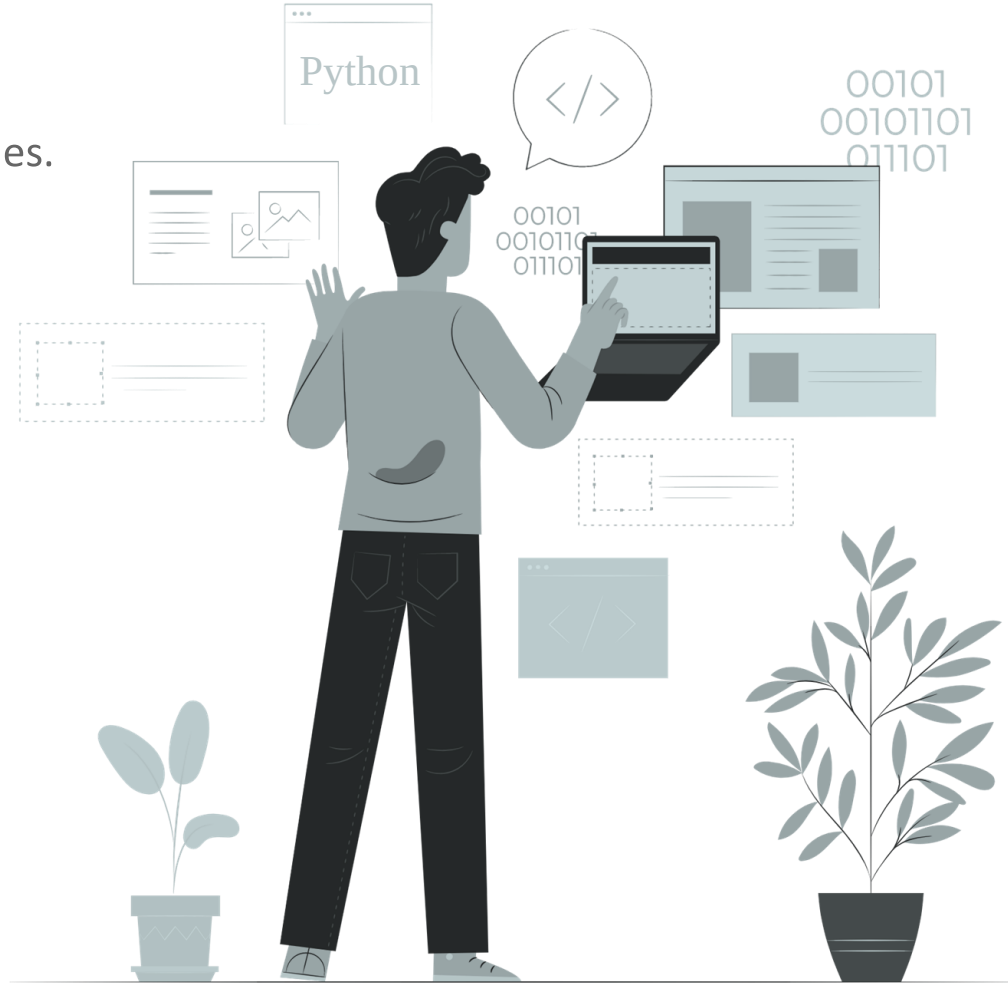
Algorithm

Feature Engineering

1) Eliminate nonimportant features
2) Convert categorical variables into binary dummy variables.

Models

1) Logistic regression.
2) k-nearest neighbors.



Algorithm

Model Evaluation and Selection

Train test split library, the training size is 75% and the test size is 25%.

Train the model with Logistic Regression

	precision	recall	f1-score	support
0	0.79	0.69	0.74	2838
1	0.78	0.85	0.81	3590
accuracy			0.78	6428
macro avg	0.78	0.77	0.78	6428
weighted avg	0.78	0.78	0.78	6428

Train the model with K Neighbors

	precision	recall	f1-score	support
0	0.99	0.96	0.98	2838
1	0.97	1.00	0.98	3590
accuracy			0.98	6428
macro avg	0.98	0.98	0.98	6428
weighted avg	0.98	0.98	0.98	6428

Tools

Tools

Numpy and Pandas for data manipulation.

Scikit-learn for modeling.

Matplotlib for plotting.

THANK YOU FOR LISTENING

Github repo: https://github.com/ShoroogH/JobRetention