

Problem Statement - Feature Engineering and cross validation

Case Study: Job Change Prediction

Context:

An Ed-Tech company wants to hire data scientists among people who have successfully passed some courses and then signed up for training. The company wants to know which of the people are really looking for a job change and will prefer working with them, after completion of training because it helps to reduce the cost and time for categorization of candidates.

Information related to demographics, education, experience is in hands from candidates signup and enrollment.

Problem:

The data-set aims to answer the following key questions:

What are the factors affecting a person looking for a job change?

Does imbalance in the data affect model predictions?

What are the key business recommendations based on analysis and model?

Attribute Information:

The data contains characteristics of the people

enrollee_id: Unique ID for the candidate

city: City code

city_development_index: Development index of the city (scaled)

gender: Gender of the candidate

relevent_experience: Relevant experience of the candidate in years

enrolled_university: Type of University course enrolled if any

education_level: Education level of candidate

major_discipline: Education major discipline of the candidate

experience: Candidate total experience in years

company_size: No of employees in current employer's company

company_type: Type of current employer

lastnewjob: Difference in years between previous job and current job

training_hours: training hours completed

target: 0 – Not looking for a job change, 1 – Looking for a job change

Learning Outcomes:

Exploratory Data Analysis

K-fold Cross-validation

Handling imbalanced data

Regularization to reduce overfitting

Steps and Tasks:

Import Libraries and Load Dataset

Overview of data

Data Visualization

Data preparation

Choose model

Over and under-sample train set to balance the classes

Perform regularization, if needed

Conclusion