

You've recently joined DataCorp, a leading company in Big Data and Data Science, as a data scientist. Your primary responsibility is to develop a predictive model that helps the company identify candidates who are likely to stay with the company after completing their training program. This model will assist DataCorp in optimising their training programs, reducing costs, and improving candidate categorization.

**Objective:** Your task is to build a machine learning model that predicts the probability of a candidate seeking a new job or staying with DataCorp after completing the training. The model should leverage demographic, educational, and experiential attributes of the candidates to make these predictions. This will help DataCorp focus resources on candidates more likely to stay, enhancing the quality and efficiency of their training programs.

**Description:** The dataset provided contains anonymized information about candidates' demographics, education, experience, and other relevant features. It includes features such as city development index, gender, relevant experience, education level, major discipline, total experience, company size, company type, last new job, and training hours. The target variable indicates whether a candidate is looking for a job change (1) or not (0).

## Features

- enrollee\_id : Unique ID for candidate.
- city: City code.
- city\_development\_index : Development index of the city (scaled).
- gender: Gender of candidate
- relevant\_experience: Relevant experience of candidate
- enrolled\_university: Type of University course enrolled if any
- education\_level: Education level of candidate
- major\_discipline :Education major discipline of candidate
- experience: Candidate total experience in years
- company\_size: No of employees in current employer's company
- company\_type : Type of current employer
- last\_new\_job: Difference in years between previous job and current job
- training\_hours: training hours completed
- target: 0 – Not looking for job change, 1 – Looking for a job change

## Deliverables:

1. **Python Code:**
  - Implement the predictive model(s) with clear and comprehensive documentation for reproducibility.

**2. Detailed Report:**

- Outline the methodology, key findings, model performance metrics, and insights derived from the analysis.
- Highlight the implications of the model's predictions for DataCorp's training program planning and candidate categorization.

**3. Visualisations:**

- Provide relevant visualisations that aid in understanding the data and the model's behaviour.

**4. Recommendations:**

- Suggest actionable steps DataCorp can take to enhance their training programs and candidate retention based on the model's insights.