**In Class Assignment1 (200 points)**

**CS430-01**

**Machine Learning on Cloud**

**Fall 2022**

**Exploratory Data Analysis, train the data using regression techniques (100 points)**

**Goal:** The goal of this assignment is to use Pandas/Matplotlib/Seaborn to explore the dataset, use Sciket-learn libraries do the data preprocessing, split the data and train models with Regression Techniques.

**The dataset you will have to use :**

[**https://www.kaggle.com/datasets/rhuebner/human-resources-data-set**](https://www.kaggle.com/datasets/rhuebner/human-resources-data-set)

**The target variable is salary calculation. Use Randomized search Cross Validation technique to detect the best model for salary prediction out of:**

* 1. **Linear Regression**
  2. **SGD Regression**
  3. **Ridge Regression**
  4. **Support vector Regression**
  5. **Decision Tree Regression**

Ensure your notebook is organized and has proper **Markdown comments on data preprocessing, model performed,** etc. You can assume that after someone see the raw notebook, so it should be clear.

1. **Assignment Submission:** Upload a link to your GitHub repository for the project in the area provided in Moodle by the deadline specified.
2. Prepare a PowerPoint on your findings and present it on 10/12/2022 end of class.