ML Problem Statement

Mentorness Internship Program

Project Name: Salary Predictions of Data Professions

Welcome to the Machine Learning Internship, focused on predicting the salaries of data professionals. In this project, you will dive into the world of regression tasks and gain hands-on experience in data analysis, feature engineering, and machine learning model development. The goal is to predict the salaries of data professionals based on a rich dataset.

Problem Statement:

Salaries in the field of data professions vary widely based on factors such as experience, job role, and performance. Accurately predicting salaries for data professionals is essential for both job seekers and employers.

Your Mission:

Your mission in this internship is to build a predictive model for salaries of data professionals. You will follow these key steps:

1. Exploratory Data Analysis (EDA): Dive into the dataset, conduct comprehensive EDA, and unveil valuable insights about data professionals' salaries. EDA will involve data visualization, summary statistics, and identifying patterns in the data.

2. Feature Engineering: Create new features or transform existing ones that can provide additional insights or improve model performance. Feature engineering might involve deriving features related to experience, job role, and performance.

3. Data Preprocessing: Prepare the data for model training. This includes handling missing values, encoding categorical variables, and scaling or normalizing features as needed.

4. Machine Learning Model Development: Train various machine learning regression models to predict salaries. Experiment with different algorithms such as linear regression, decision trees, random forests, and gradient boosting to find the best-performing model.

5. Model Evaluation: Assess the performance of your models using appropriate evaluation metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R2) score. Identify the model that provides the most accurate salary predictions.

6. ML Pipelines and Model Deployment: Create ML Pipelines to streamline the end-to-end machine learning process, from data preprocessing to model training. Additionally, deploy a model that can generate predictions for unseen data. Consider using frameworks like Flask or FastAPI for deployment.

7. Recommendations: Based on your findings and insights from the model, provide actionable recommendations. These recommendations can be related to factors influencing salaries and potential strategies for improving earnings in data professions.

ML Problem Statement

Mentorness Internship Program

Dataset Overview:

The dataset contains the following columns:

- `FIRST NAME`: First name
- `LAST NAME`: Last name
- `SEX`: Gender
- `DOJ`: Date of joining the company
- `CURRENT DATE`: Current date of data
- `DESIGNATION`: Job role/designation
- `AGE`: Age
- `SALARY`: Target variable, the salary of the data professional
- `UNIT`: Business unit or department
- `LEAVES USED`: Number of leaves used
- `LEAVES REMAINING`: Number of leaves remaining
- `RATINGS`: Ratings or performance ratings
- `PAST EXP`: Past work experience

Conclusion:

This internship offers an exciting opportunity to apply machine learning techniques to predict salaries in the data profession domain. By the end of the internship, you will have gained valuable skills in data analysis, preprocessing, feature engineering, regression model development, ML Pipelines, model deployment, and recommendation generation. Your work will contribute to providing valuable insights for job seekers and employers in the field of data professions. Are you ready to embark on this exciting journey of predicting salaries for data professionals and making an impact in the job market? Let's get started!