Minor Project

Project Name:

Machine Learning May Minor Project

Project Description:

Problem statement: Create a Regression model to predict the Salary of data scientists.

Context: With the growing demand of data in the industry more and more employees join the organization into the domain of data science. Based on the feature details, the requirement is to predict their salaries.

Dataset:

https://drive.google.com/file/d/1w6anaOWIGbIFXhp2GV7nM3vH_Ov5V_wf/view?usp=sharing

Details of features:

Data Science Job Salaries Dataset contains 11 columns, each are:

1. work_year: The year the salary was paid.
2. experience_level: The experience level in the job during the year
3. employment_type: The type of employment for the role
4. job_title: The role worked in during the year.
5. salary: The total gross salary amount paid.
6. salary_currency: The currency of the salary paid as an ISO 4217 currency code.
7. salaryinusd : The salary in USD
8. employee_residence: Employee's primary country of residence in during the work year as an ISO 3166 country code.
9. remote_ratio: The overall amount of work done remotely
10. company_location : The country of the employer's main office or contracting branch
11. company_size : The median number of people that worked for the company during the year

Steps to consider:

1. Remove handle null values (if any).
2. Apply the necessary data preprocessing like
    a. Handling null values
    b. Handling duplicate records
    c. Label Encoding or Dummy variable creation for encoding categorical columns
3. Select x (independent variable) and y (dependent variable) - Exited(1 or 0)
4. Split data into training and test data.
5. Apply the ANN Classification model on the training dataset and generate the predicted value for the test dataset
6. Predict the exited status for test data
7. Compute Classification metrics – confusion Matrix, classification Report.
8. Report the ANN model accuracy.