|  |  |
| --- | --- |
| Internship Project Title | TCS iON RIO-125: HR Salary Dashboard - Train the Dataset and Predict Salary |
| Name of the Company | TCS iON |
| Name of the Industry Mentor | Debashis Roy |
| Name of the Institute | B. K. Birla College of Arts, Science & Commerce (Autonomous), Kalyan |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Start Date | End Date | | Total Effort (hrs.) | | Project Environment | Tools used |
| 01-10-2023 | 10-10-2023 | | 40 hrs | | Google Colab | Python3 |
| Milestone # | 2 | Milestone: | | Train the dataset and predict the salary of particular HR based on the dataset. | | |

**TABLE OF CONTENT**

* Acknowledgements
* Objective
* Introduction
* Internship Activities
* Approach / Methodology
* Outcome
* Link to code and executable file

**ACKNOWLEDGEMENTS**

I am conveying my sincere gratitude towards my industry mentor, Debashis Roy, for helping me throughout this project till now and providing me this wonderful platform to complete this project. I am also thankful for answering my queries at every phase of the project. I also want to thank all my friends who helped me with valuable suggestions during this project.

**OBJECTIVE**

The objective of this project is to build a salary prediction dashboard for human resource management. This dashboard will use machine learning algorithms to predict the salary of job candidates based on factors like their experience, age, and qualifications. This information can help HR managers to make better decisions when hiring candidates for job positions.

**INTRODUCTION**

The human resources department often has a lot of job applicants to process and must choose the best candidates for each job. Candidates often consider salary when deciding whether to accept a job offer, so it's important for HR to offer competitive salaries.

In this project, we will use a dataset that includes information on over 32,000 job candidates, such as their experience and salary. This dataset is good for our analysis because it has a wide range of job profiles and salaries. We will use this data to build a salary prediction dashboard to help HR managers make better decisions about salaries for job candidates.

In the dataset I have chosen, the target salary only has two categories: less than or equal to 50K and greater than 50K. Therefore, the model needs to predict which of these two classes the salary belongs to. This means that our model will be a binary classification model. There are several methods we can use for binary classification, such as SVM, logistic regression, random forest, etc. I have trained and tested my data using Logistic Regression and KNN, compared them to select the best model.

**INTERNSHIP ACTIVITIES**

* + Watched the welcome kit videos.
  + Done preparations for RIO – pre-assessment.
  + Attended the RIO – pre-assessment test.
  + Went through the day-wise plan.
  + Read the project reference material.
  + Read the industry project material.
  + Watched webinar 1.
  + Watched webinar 2.
  + Gone through all posts in the digital discussion room.
  + Watched lectures and other videos to gain a better understanding of the topic.
  + Created a GitHub account to store and share my project files.
  + Found a suitable data set for the project.
  + Wrote activity reports to document my progress.
  + Verified that the data set had enough data for the project.
  + Read articles and learned how to clean and sanitize the data.
  + Applied data cleaning and sanitization techniques to the data set.
  + Conducted exploratory data analysis to identify patterns and trends in the data.
  + I watched videos about how to train a model.
  + I used logistic regression and trained it on my data.
  + I also used a KNN (k-nearest neighbor) classifier and trained it.

**APPROACH / METHODOLOGY**

To complete the second milestone of my internship project, I took the following approach:

* I began by understanding the concepts and requirements of the project through reading articles and watching videos.
* I used Google Colab for programming because it allows me to quickly write and execute code.
* I created a GitHub account to publish my code and share it with others.

Overall, this approach allowed me to gain the knowledge and skills necessary to work on the project and make progress towards the first milestone.

**OUTCOME**

Over the course of 15 days, I have gained a lot of knowledge about a variety of topics. I have learned and understood many concepts and ideas related to my internship project, and I feel that I have made good progress in my learning.

**LINK TO CODE AND EXECUTABLE FILE**

* Link to the code:

https://colab.research.google.com/drive/1cJzfULcTTTWFLM3aQjOrZfOUu17MUGsQ#scrollTo=oTy\_U34CLkXU