|  |  |
| --- | --- |
| 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 | ICT Academy of Kerala |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Start Date | End Date | | Total Effort (hrs.) | | Project Environment | Tools used |
| 17-04-2023 | 16-07-2023 | | 21.5 | | Google Colab | Python3 |
| Milestone # | 1 | Milestone: | | Create dataset, clean dataset, and sanitize dataset | | |

**TABLE OF CONTENT**

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

**ACKNOWLEDGEMENTS**

I would like to express my heartfelt appreciation to my industry mentor, Debashis Roy, for his invaluable guidance and support throughout this project. His assistance and the opportunity he provided me with to work on this wonderful platform have been instrumental in helping me successfully complete this endeavor. I am truly grateful for his willingness to address my queries at every stage of the project.

I would also like to extend my gratitude to all my friends who offered their valuable suggestions during this project. Their input has greatly enriched my work, and I am thankful for their unwavering support.

Once again, I extend my sincerest thanks to Debashis Roy and all my friends for their contributions, encouragement, and belief in me.

**OBJECTIVE**

The aim of this project is to develop a salary prediction dashboard specifically designed for human resource management. This dashboard will leverage the power of machine learning algorithms to forecast job candidates' salaries, taking into account various factors such as their experience, age, and qualifications. By providing this valuable information, the dashboard aims to assist HR managers in making more informed decisions during the candidate selection process and improve the overall effectiveness of hiring for job positions.

**INTRODUCTION**

The human resources department faces the challenging task of evaluating numerous job applicants and selecting the most suitable candidates for each position. Salary considerations often play a significant role in a candidate's decision to accept a job offer, highlighting the importance of offering competitive compensation packages.

In this project, we will leverage a comprehensive dataset comprising information on more than 32,000 job candidates, including their experience levels and salaries. This dataset proves to be highly valuable for our analysis as it encompasses a wide range of job profiles and associated salaries. Our objective is to utilize this dataset to construct a salary prediction dashboard, empowering HR managers to make well-informed decisions regarding appropriate salary offers for job candidates.

During the initial phase of the project, I have successfully collected the dataset and proceeded to diligently clean and refine it. As a result, the dataset now stands primed for the subsequent training phase, where it will be utilized for constructing a robust salary prediction 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.

**APPROACH / METHODOLOGY**

To accomplish the initial milestone of my internship project, I adopted the following methodology:

1. Familiarization: I commenced by immersing myself in the project's concepts and requirements through extensive reading of articles and watching informative videos. This preliminary step enabled me to grasp the fundamental aspects and gain a comprehensive understanding of the project's scope.

2. Programming Environment: I opted to utilize Google Colab as my programming environment. This choice stemmed from its advantages, including the ability to swiftly write and execute code. Leveraging the platform's features allowed me to streamline my workflow and expedite development tasks.

3. Version Control: Recognizing the significance of collaboration and code management, I established a GitHub account. This facilitated seamless code sharing and provided a platform for version control, enabling efficient collaboration with peers and potential reviewers.

By implementing this methodology, I was able to acquire the requisite knowledge and skills essential for project engagement, paving the way for substantial progress towards achieving the initial milestone.

**OUTCOME**

Upon the successful completion of the first milestone in this internship project, I have gained valuable insights into the process of cleaning and preparing datasets for model construction. As part of this endeavor, I employed several techniques to enhance the cleanliness and accuracy of the dataset.

To begin with, I eliminated redundant columns such as "capital gain" and "capital loss" during the model building process. Additionally, I identified the "education-num" column as unnecessary and subsequently removed it from the dataset. These actions ensured that the dataset was streamlined, omitting irrelevant information and promoting data accuracy.

Furthermore, I conducted exploratory data analysis (EDA) to identify and subsequently remove records containing superfluous values. By doing so, I was able to improve the overall quality and precision of the data, eliminating any potential outliers or inconsistencies that may have hindered the model's performance.

Overall, these outcomes have equipped me with the necessary skills and knowledge to effectively clean and prepare datasets, setting a strong foundation for subsequent stages of the project.

**LINK TO CODE AND EXECUTABLE FILE**

* Link to the code:

<https://colab.research.google.com/drive/1JO8Y2ufpab-GM5o3JOr_HWgypNjaEC5_?usp=drive_link>

* Executable file:

<https://github.com/Zandubai/TCS_iON-RIO-125.git>