Dear Intern

Interim project report is an inherent component of your internship. We are enclosing a reference table of content for the interim project report.

The key objective of this report is for you to capture how far you have got in completing the internship work against milestones expected to be achieved within a specific duration and seek the mentor’s feedback. Depending on the internship project and your progress (IT/Non-IT, Technical/Business Domain), you may choose to include or exclude or rename sections or leave some sections blank from the table of content mentioned below. You can also add additional sections. You can refer the project presentation to view the milestones related to your internship project. Please populate milestone# (1 / 2 / 3) and the milestone description in the interim project report based on the milestone for which you are submitting the interim project report.

You can refer the project presentation to view the milestones related to your internship project.

|  |  |
| --- | --- |
| Internship Project Title | RIO-210: Salary Prediction Dashboard for HRs |
| Name of the Company | TCS ION |
| Name of the Industry Mentor | Himdweep Walia |
| Name of the Institute | Amity University Online |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Start Date | End Date | | Total Effort (hrs.) | | Project Environment | Tools used |
| 12-04-24 | 24-04-24 | | 70 | | Jupyter Notebook | Python |
| Milestone # | 2 | Milestone: | | Train the dataset and predict the salary of a particular HR based on the training set. | | |

**TABLE OF CONTENT**

* Acknowledgements
* Objective
* Introduction / Description of Internship
* Internship Activities
* Approach / Methodology
* Assumptions
* Exceptions / Exclusions
* Charts, Table, Diagrams
* Algorithms
* Challenges & Opportunities
* Risk Vs Reward
* Reflections on the Internship
* Recommendations
* Outcome / Conclusion
* Enhancement Scope
* Link to code and executable file
* Research questions and responses

**1. Acknowledgements**

We express gratitude to *Himdweep Walia* for their guidance and support throughout this internship 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.

**2. Objective**

The primary objective of this project is to develop a Salary Prediction Dashboard utilizing data analytics and machine learning techniques. I make 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.

**3. 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.

**4. 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.

**5. Data Collection**

In the process of data collection, I begin by leveraging easily accessible sources such as platforms like Kaggle or the UCI Machine Learning Repository. I meticulously select a dataset aligned with my project's objectives and my areas of interest. Upon accessing the dataset, I thoroughly review any accompanying documentation or metadata to gain a comprehensive understanding of its structure and variables. This ensures that I am well-equipped to proceed with subsequent data analysis and interpretation tasks effectively.

**6. Data Interpretation**

In data interpretation, I commence by delving into the dataset's structure, employing fundamental commands or functions within the Python programming language, such as Pandas, to load the dataset and conduct a comprehensive examination of its layout. Subsequently, I proceed to identify the pivotal variables contained within the dataset, meticulously discerning those that are pertinent to my analysis or project objectives. Following this, I embark on visualizing the distribution of the data, crafting straightforward visual representations such as histograms for numerical variables and bar charts for categorical variables. These visualizations serve to illuminate the inherent distributions within the dataset, facilitating a deeper understanding of its underlying characteristics and trends.

**7. Data Visualization**

I initiate by crafting basic plots using prominent Python libraries such as Matplotlib or Seaborn. These plots encompass fundamental visualizations including histograms, bar charts, and scatter plots, providing an initial exploration of the dataset's characteristics. Subsequently, I delve into visualizing relationships between variables, employing techniques such as scatter plots or pair plots to unveil correlations or patterns inherent within the data. This insightful exploration aids in identifying significant associations between key variables, pivotal for subsequent analysis and model development.

Moreover, I emphasize customization of the visualizations tailored to the requirements of the Salary Prediction Dashboard. Through meticulous adjustment of colors, labels, and titles, I strive to enhance the informativeness and visual appeal of the plots. This customization ensures that the visualizations seamlessly integrate with the dashboard interface, effectively conveying insights to end-users clearly and engagingly.

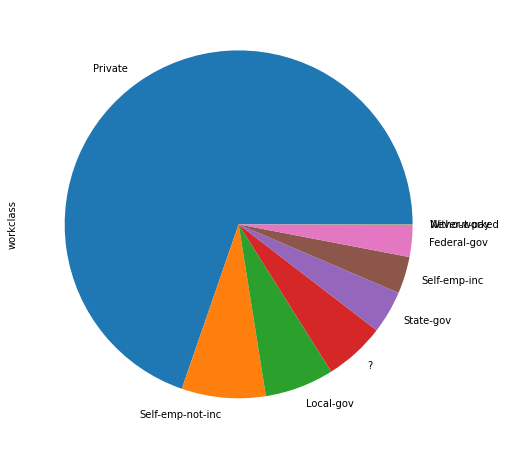
**8. Approach/Methodology**

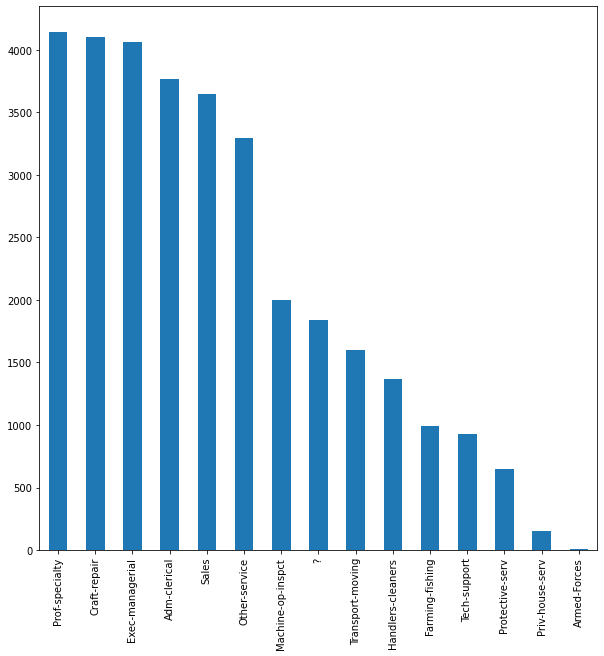
To complete the second milestone of my internship project, I took the following approach:

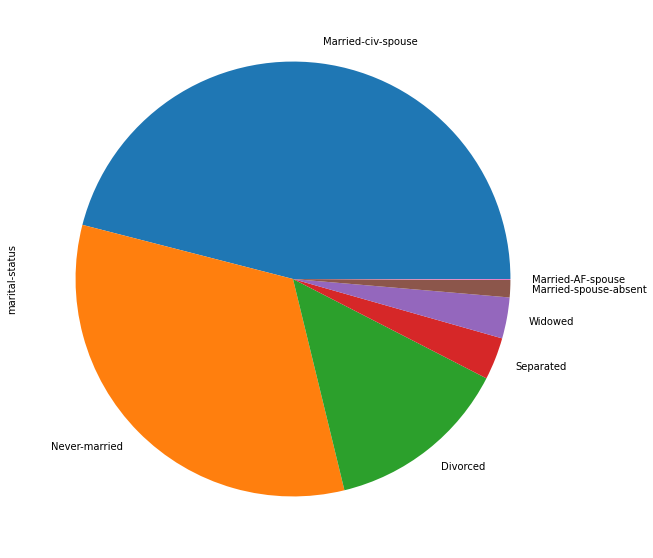
* I began by understanding the concepts and requirements of the project through reading articles and watching videos.
* I used Jupyter Notebook for programming because it allows me to quickly write and execute code.
* I created a GitHub account to publish my code and share it with others.

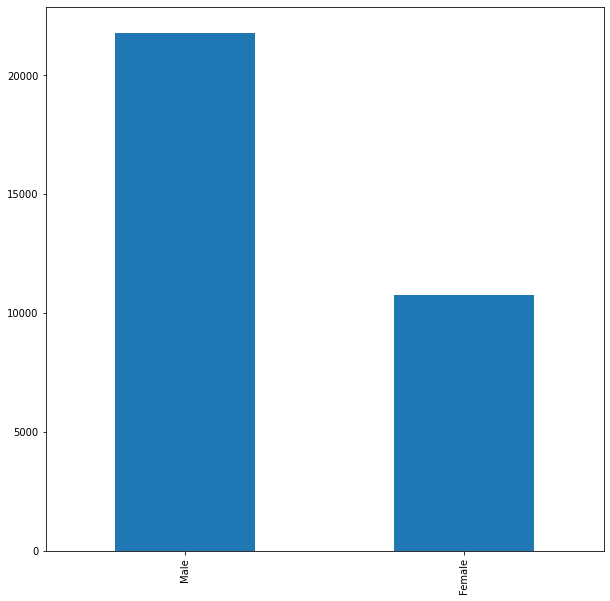
Overall, this approach allowed me to gain the knowledge and skills necessary to work on the project and make progress towards the first milestone.

Charts, Table, Diagrams









**9. Recommendations**

To refine our project, we should narrow down our focus to specific industries or job positions, tailoring our salary predictions accordingly. Additionally, we need to delve into advanced data handling techniques, such as addressing missing values and outliers more effectively. Experimenting with different algorithms and fine-tuning their parameters could also enhance the accuracy of our model. These strategies will contribute to the development of a more precise and actionable Salary Prediction Dashboard.

**10. Outcome / Conclusion**

Completing the second milestone of this internship project, I've progressed to training the dataset and predicting the salary of a specific HR professional based on the training set. 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.

**11. ENHANCEMENT SCOPE**

This industry project encompasses a wide scope, focusing on predicting individual salaries by analyzing resumes or CVs. Leveraging natural language processing techniques will be pivotal in crafting this application, aiding in the extraction of pertinent information from textual data.

**LINK TO CODE AND EXECUTABLE FILE**

* Link to the code:

<https://colab.research.google.com/drive/1cEw3FYyiBpg1qQMJ0vPB72uc_BQgBpPQ?usp=sharing>

* Executable file:

<https://github.com/SahilGitMaster/Salary-Prediction-Dashboard-for-HRs>