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-125: Classification Model - Build a Model that Classifies the Side Effects of a Drug |
| Name of the Company | TCS ion |
| Name of the Industry Mentor | Himalaya Aashish |
| Name of the Institute | ICT academy kerala |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Start Date | End Date | | Total Effort (hrs.) | | Project Environment | Tools used |
| 20/02/2020 | 20/03/2020 | | 20 | | Jupyter notebook | Excel, jupyter |
| Milestone # | 1 | Milestone: | | Day5:Student should be able to create, clean and sanitize data and preprocess the data | | |

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

**Acknowledgement**

The internship opportunity I had with TCSion was a great chance for learning and professional development. I express my deepest thanks to Himalaya Ashish, Industry Mentor for taking part in useful decision & giving necessary advices and guidance and arranged all facilities to make life easier.

It is my radiant sentiment to place on record my best regards, deepest sense of gratitude to all faculty member of ICT Academy of Kerala for their careful and precious guidance which were extremely valuable for my study both theoretically and practically.

I perceive as this opportunity as a big milestone in my career development. I will strive to use gained skills and knowledge in the best possible way, and I will continue to work on their improvement, in order to attain desired career objectives. Hope to continue cooperation with all of you in the future.

**Objective**

In the Pharma industry, the most common and high-priority question to be answered is “whether a particular drug has side effects over various types of people?”

Now a days massive data generated from the search engines has widened the perspective of the market research and analysis in the drug data. With the help of other parameters we will predict whether a drug is safe or not. Its success is based on the available data on <https://archive.ics.uci.edu/ml/datasets/Drug+Review+Dataset+%28Druglib.com%29> . The main given objective of this project to build a classification model that classifies the side effects of a particular drug by age, gender and race. The model need to have good amount of accuracy and have to meet the industry standards.

**Introduction / Description of Internship**

The project guidelines clearly mentioned that we are expected to create a model that classifies the trial data of a drug based on their age, gender and race. We also entrusted to create a dataset of 4, 00,000 patients containing the following details for each patient based on various attributes according to the data. At the end of the project we should be able to create a dataset, clean the dataset, sanitize it and preprocess the data to perform data partitioning and handle missing values. Create training and testing sets. Build a classifier and fit the data to the model

**Internship Activities**

The activity mainly concentrates on how we make up to the objective of the internship. The given resources were very useful to kick start our internship and the day wise plan helps us to calculate the overall time and amount of work to be done each day and what extra we can do about it. We can explore different aspects of this data which vary from EDA to the final prediction model for the 30days

**Approach / Methodology**

The Approach / Methodology used here will be the Linear Strategy which consist in sequential phases with no feedback loops. The project solution is not released until the final phase is reached. This strategy is characterized by clearly defined goal solution and requirements, zero or few change request of the scope, routine and repetitive process inside the project, use of pre-established formulas and templates. The pre-defined steps includes data cleaning, EDA, PCA, data preprocessing , feature processing, splitting to test and train set, applying machine learning algorithms, comparison of machine learning algorithms. Opting the best prediction model

**Assumptions**

By various Exploratory data analysis we can come an assumption that the drug are rated good for the body by chemist, it have a slight side effect of the dataset mainly for depression from the

The condition attribute mainly concentrate on insomnia, depression and other mental problem related to brain issues