|  |  |
| --- | --- |
| Internship Project Title | TCS ion RIO-125: Automate Detection and Recognition of Grammatical Errors |
| Name of the Company | TCS iON |
| Name of the Industry Mentor | Snehal Tandel |
| Name of the Institute | University College of Engineering Arni, Thatchur |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Start Date | End Date | | Total Effort (hrs.) | | Project Environment | Tools used |
| 19.04.2023 | 19.07.2021 | | 25 | | Jupyter Notebook | Python3 |
| Milestone # | 2 | Milestone: | | Making the data ready to create a model. | | |

**TABLE OF CONTENT**

|  |  |
| --- | --- |
| S. No | TOPICS |
| 1. | Acknowledgement |
| 2. | Objective |
| 3. | Introduction |
| 4. | Internship Activities |
| 5. | Approach/ Methodology |
| 6. | Outcome |
| 7. | Link to the code and Executable file |
| 8. | Conclusion |

1. **ACKNOWLEDGEMENTS**

I am conveying my sincere gratitude towards my Industry Mentor, Snehal Tandel and Academic Mentor, Mrs. Shoba for helping me throughout this project till now and providing me this wonderful platform to complete this project. I am 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.

1. **OBJECTIVE**

The objective of this Internship is to create a model that should detect and recognize the Grammatical errors of the given sentences in the English language with the help of the Machine Learning concepts.

1. **INTRODUCTION**

In TCS Ion Rio-125 internship, I have selected the Machine Learning Project on Automate Detection and Recognition of Grammatical Errors. In this project, I have to create a model that should detects the grammatical errors in the given sentences which is in English language. The project includes collection of dataset, data pre-processing, analyses of data, splitting of train and train data and creation of model.

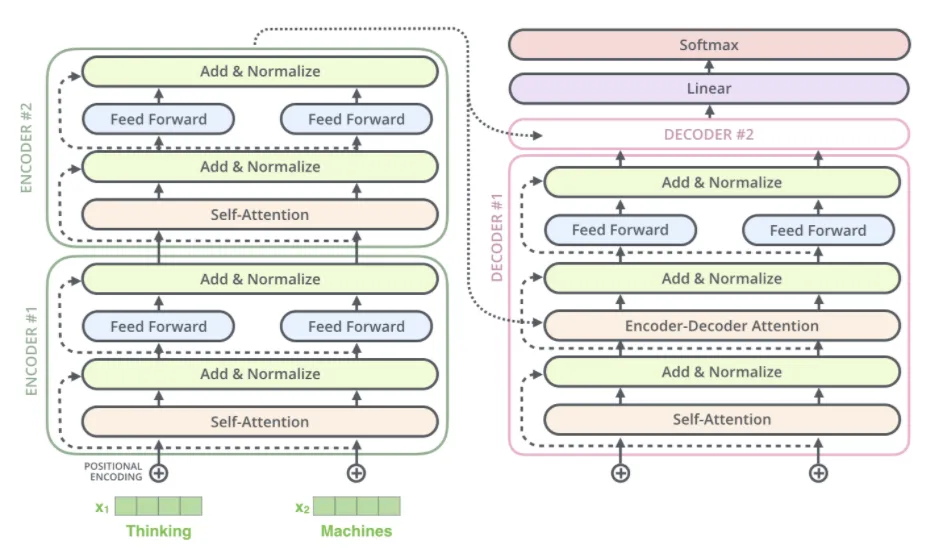
1. **INTERNSHIP ACTIVITIES**
   * Watched all the welcome kit videos.
   * Attended the RIO – pre-assessment test.
   * Went through the day-wise plan.
   * Read and learnt some concepts mentioned in the project reference material.
   * Read the industry project material.
   * Watched webinar 1.
   * Watched webinar 2.
   * Daily checking for announcement in Digital Discussion Room.
   * Created a GitHub Repository for Project.
   * Searched and found out a proper dataset for this project.
   * Wrote activity reports.
   * Checked and clarified the data set whether it has enough data for the project.
   * Read articles and find out how to clean and sanitize the data.
2. **APPROACH / METHODOLOGY**

The approach of completing the 2nd milestone is preparing the data that is, the dataset which I searched and downloaded and with this dataset, I tried many methods for the model creation that is using different types of libraries in different ways. Finally I decided the a way to build a model with this dataset.

1. **ALGORITHM**

In the process of Grammatical Error Detection model, I have used the T5-base. From this T5-base ( **T**ext–**t**o-Text **T**ext **T**ransfer **T**ransformer), we can do various types of model creation and text generation. It is a system that generates the text-to-text as the outcome. With this, I have build the grammatical Error detection model.

Many tasks are cast into this framework: machine translation, classification task, regression task ( for example, predict how similar two sentences are, the similarity score is in range 1 to 5), other sequence to sequence tasks like document summarization (for example, summarizing articles from CNN daily mail corpus).



1. **CHALLENGES**

One of the main challenges in the process of building model is finding a dataset that is suitable for the grammatical error detection. The dataset used in this model is ‘jfleg’. This dataset is widely used in many Grammatical error detection models. Another challenge involves in the understanding of the dataset and the different errors and its corrections in the sentences of the dataset.

1. **REFLECTIONS ON THE INTERNSHIP**

By this Internship program, I have learnt many concepts in the machine learning and deep learning. I have gained some experience on pre-processing the data, building models and learning many more things through this project. I got interested to work on the Machine learning projects and exited to

1. **OUTCOME**

After the 1st milestone of this internship project, I have learned about Natural Language Processing and understood how it works. And I have installed some libraries and imported it in the program for dataset pre-processing and for creating detection model and learnt the uses of the libraries that are used in this project.

After the 2nd milestone I gained the knowledge that how a machine learning model should be created and trained. Building of Machine learning models may involves variety of modules, each of them is different way to build a model in different perspective or the same with different ways.

1. **LINK TO CODE AND EXECUTABLE FILE**

* Link to the code:

1. **CONCLUSION**

The project is further developed for the model creation and detection and recognition of grammatical errors. The dataset has to be analyzed and EDA is processed with the dataset. After that, the process of splitting data and finding patterns in the data is to be completed.