Name: Vaibhav Khamesra

Email: 21ucs224@lnmiit.ac.in , vaibhavkhamesra643@gmail.com

Phone No: +917023659757

Institute: The LNM Institute of Information Technology

Branch of Study: B.Tech in Computer Science and Engineering (2021 – 2025)

Greetings,

This is the Documentation part of the solution for the Problem Statement of Adobe Papyrus Nebulae Hackathon 2023.

I would like to begin by expressing my Gratitude to everyone involved in the management side for the creative problem statement and the exposure I received from it also kudos to the API team for creating the pdf extraction API really a great achievement.

Language Used: Python

The whole solution is divided into 5 Code which are:

1. API_code.py : This program calls the API request through the *get_data_from_api(file_index)* function This function uses the boiler plate code with some minimal changes to receive the data in *.json and .csv* formats for the pdf with a given file index.

2. zip_extracter.py: This program extracts the received output from the above code which is stored in a zip file using the function *extract(file_index)*

3. data_extraction_code.py: This program extracts the data from the *.json and .csv* files through the function extract(file_index). The function handles all the formats in which data is present in the *.csv and .json* file and appends it to the resultant *.csv* file as required by the problem statement.

4. directory_handler.py: This programs creates the necessary directories for the main_program.py and other programs to function

5. main_program.py: This program uses the above four programs and modules and automates the task for any given number of pdf files