Table of Contents

i. Title Page  
ii. Problem Statement   
iii. Business context, stakeholders and value

* Overall understanding of the business domain
* Explanation of the business context
* Formulation of the business question
* Understanding and engagement of stakeholders
* Estimation of the business value

iv. Data description, sources, quality

* Translation of the business question into a data question
* Defining what data is needed to answer the business question
* Understand how to source the data
* Understanding of how the data was generated
* Understanding of the quality of data and its limitations
* Understanding of how the data can be sourced in the future

v. Data exploration, analysis and visualisation

* Data exploration showing the key entities and their business significance
* Using effective visualisation to communicate key aspects of the data

….

vii. Summary

Title Page

Towards clean and Green E-mail Communication for Managers, Scientists, Engineers, and Doctors by optimized Email Visualization using smart AI Algorithms.

ii. Problem Statement

Creating clean and green email communication for professionals like managers, scientists, engineers, and doctors is important. Optimizing email visualization using smart AI algorithms can significantly enhance email communication's efficiency and environmental impact.

Motivation for the Project: Clean and Green Singapore. [https://www.cgs.gov.sg/](https://colab.research.google.com/corgiredirector?site=https%3A%2F%2Fwww.cgs.gov.sg%2F). The requirements of this project were provided by David Woon, from Continental Automotive Singapore.

**Proposed solution:**

1. **Data Collection and Labeling:**
   * A significantly large and diverse dataset of emails from enron (Reference 1) is considered for training AI models. The dataset was modified based on the person, who suggested the project. The modification is provided in preparatory.ipynb
   * First, a dataset ([smartcontinental.csv](https://github.com/venuannamdas/E_clean_AI/blob/master/smartcontinental.csv)) of 9571 observations was obtained from an enormously large dataset. These emails are labelled according to specified categories (time\_waste, reply, required\_response).
   * Next, a dataset ([email\_details.csv](https://github.com/venuannamdas/E_clean_AI/blob/master/email_details.csv)) of 4521 observations was obtained to understand the body of the email and its contents are studied and summarized for insights.
2. **Priority Ranking:**
   * Develop a priority ranking algorithm that takes into account the labels assigned to each email (time\_waste, reply, required\_response) and any additional factors like sender importance or email urgency.
3. **Recommendation System:**
   * Implement a recommendation system that suggests actions based on the priority ranking. For example, if an email is classified as "required\_response," the system could recommend replying promptly.
4. **Natural Language Understanding (NLU):**
   * To understand the content of emails, NLU techniques, such as sentiment analysis, entity recognition, and topic modelling is considered. This can provide insights into the email's content.

By considering these points, we can ensure that the email response recommendation and classification system is robust, customizable, and capable of meeting the specific needs of different organizations and users.

Top of Form

Reference 1: [The Enron Email Dataset | Kaggle](https://www.kaggle.com/datasets/wcukierski/enron-email-dataset)

iii. Business context, stakeholders and value

* Overall understanding of the business domain
* Explanation of the business context
* Formulation of the business question
* Understanding and engagement of stakeholders
* Estimation of the business value