

CIT – Interview Task

Folder contents

- **Train dataset** - A compressed JSON file that contains Recipe ID, Type of cuisine, and Ingredients.
- **Test dataset** - A compressed JSON file that contains Recipe ID and Ingredients.
- **Sample submission** – A compressed CSV file template for submitting the results.

Data description

Each row in the training dataset is a recipe and it contains list of ingredients used in that recipe and to which cuisine the dish belongs. This is a multi-class classification problem.

Task

- Build a model that will learn the cuisine type based on the training dataset
- Predict the type of cuisine for each recipe ID in the test
- Submit
 - Python or R code
 - Exploratory analysis on the train data
 - Evaluation metric used and why
 - Explain your algorithm
 - Your results in the format as recommended in sample_submission.csv file

Rules:

- A Dropbox folder has been shared and work only in that folder
- Save all the results in the 'Output' folder (within the shared folder)
- Attempt as many of the tasks mentioned above in the given 4 hours
- Do not submit identical results as your fellow candidate
- Can use internet for researching and completing the task
- Should be able to explain your analysis and output