

Problem Statement

In the vehicle Make-Model tagging universe, the names of vehicles with Make, Variants, Seating Capacity, Petrol-Diesel etc form a very large list.

Insurance companies offer Insurance packages for vehicles based on these lists (which are Excel spreadsheets). Since each Insurance company forms their own lists and naming conventions, it is difficult to search for the right record in the Master List (which is an Excel sheet).

Direct Lookup is usually time consuming and returns multiple result sets. The solution requires the closest-record from the Master List to be selected as Output.

Hackathon Approach

Students would get various Excel sheets with heterogeneous naming of cars and models. Given any row in any of the Excel sheet, they have to use Machine Learning algorithm or something similar to it and select the right record from Master database.

Inputs

A master vehicle data set would be given which would have the vehicle details like vehicle name, seating capacity, fuel capacity, and cubic capacity as input and would have a unique code against the same.

Now each team would be given some similar dataset with the same input params, but the vehicle naming convention would be different, or let's say we won't have any data in these datasets matching any other inputs.

Now the teams need to develop an ML algorithm, such that given vehicle details with these fixed input params, the algorithm should return the code as an output from the master vehicle dataset.

Master Dataset link

Training Sets link

Testing Sets -> would be given during submission round (more details below).

Submission Round

1. Before judging round, the teams would be given a test sheet at 5pm, and they're expected to submit the response in below format by 6pm.
2. https://docs.google.com/spreadsheets/d/1EgcLYrLZgr8wqwU2L7D_qaQahXqFadL5rPt6_C3zFCY/edit?usp=sharing (to be used as an example). The original submission link will be shared at 5 pm.

Description

local_code -> The unique code provided in the test sheet.

mapped_code -> The code mapped from the 1st master sheet.

3. Post submission, we would check the probability of matches and qualified teams would be moved to judging round.