MILLENNIAL ENERGY CHALLENGE 2021 RULEBOOK

Perform data wrangling and develop a machine learning model to predict the time taken to travel across two points in a city.

Case study:

As the new intern of a logistic company in London, your supervisor has assigned you to a new data science project with the goal of reducing waste (i.e., time and fuel expenses) by optimizing the route of transportation for the delivery of goods. Thus, the first step in achieving the goal is to develop a model in predicting the travel time (which corresponds to value of money) across two points of location in the targeted city (London).

Timeline:

Registration – Upon publicity of event's promotional poster, until the mid of competition duration.

Competition starts – 3rd Oct 2021

Competition duration – 1 month

Evaluation of top 5 submissions – 2-3 Nov 2021

Winner announcement – 3 Nov 2021

Competition Guidelines:

- Dataset provided:

1) london_lsoa.json: The boundaries of London in geospatial (GeoJSON) format, including Zone IDs used in the other export options.
2) training_WeeklyAggregate.xlsx (This training set is a small subset of a larger dataset of approximately 2.7 million rows): Contains the arithmetic mean for aggregated travel times over the first quarter of 2020 between every zone pair in London.
    a. sourceid – Source location ID as per london_lsoa.json
    b. dstid – Destination location ID as per london_lsoa.json
    c. dow – Days of week, where 1: Monday, 2: Tuesday and so on.
    d. mean_travel_time (label): The average travel time as per the shortest distance travelled by car from its source location to destination on the particular day of week.
3) Testing dataset, similar to training dataset but without labels.
4) Sample_submission.csv, the example of a submission file.

- The language to be used should be Python, and IDE should be Jupyter Notebook (for presentation purpose).

- Participants are allowed multiple submission throughout the competition duration, with the limit of 1 submission per day.

- The file to be submitted should be a '.csv' file with exactly 4 columns with header row (namely 'sourceid', 'dstid' and 'dow' from testing dataset, and 'predicted_mean_travel_time' from your model prediction) for all rows of the testing dataset. The sequence of the rows should be exactly the same as the testing dataset.

- The sample of the submission file is as sample_submission.csv.

- Your model will be evaluated using the MSE (mean square error) metrics from sklearn.metrics.mean_squared_error(), between the predicted values of your model for the testing dataset and the true value of the testing dataset from the organizer after your submission.

General Information:

- The maximum number of submissions is 1 per day. Resubmission is only allowed the following day.
- Data from any other source is not permitted. Please only use data from the files provided by the organizer.
- All group should consist of 3-5 persons.
- The organizer will contact you regarding the python script evaluation process via email within 24 hours after the end of the competition if your latest submission has the 5 lowest error score after being tested by the organizer.
- All scripts should have their respective output readily displayed (in Jupyter Notebook) during the evaluation process, where the final predicted result should be the same as the submitted file.

- Features that are used to train the model after feature engineering should be logical and relevant
- The contacted group will be disqualified from the competition if they are not able to elaborate and reproduce the result based on their data preprocessing and model training script, and if the submitted file is not from their own original work. In the event of such situation, the next group with the lower error score will be contacted for script evaluation process via email within 24 hours from the evaluation of the former group. E.g. If the fifth-ranked group could not elaborate on their scripts, they will be disqualified. The sixth-ranked group will be contacted for script evaluation within 24 hours after the previous fifth-ranked group presenter.
- In the event of a tied score, the earlier submission will be considered the winner.
- After the script evaluation process, winners will be announced on 20 Oct 2021.

Acknowledgement:

1) OPIS DATA © OPIS, LLC. All Rights reserved.
2) Google search results may be reordered or augmented by Uber.
3) INEGI

   Red Nacional de Caminos 2016

   Información Vectorial de Localidades Amanzanadas y Números Exteriores 2016

   Marco Geoestadístico Nacional 2016
4) Data © OpenStreetMap contributors, http://www.openstreetmap.org/copyright
5) (TIGER/Line® files) U.S. Census Bureau, http://www.census.gov/
6) Data Source ©TomTom: For United Kingdom (excluding Northern Ireland): Contains Ordnance Survey data © Crown copyright and database right 2015; Code-Point® Open data: Contains Royal Mail data © Royal Mail copyright and database right 2015; Contains National Statistics data © Crown copyright and database right 2015. All rights reserved. This material is proprietary and the subject of copyright protection and other intellectual property rights owned or licensed to TomTom.
7) For London: © Greater London Authority. Greater London Authority does not warrant quality or accuracy of the data.

© EuroGeographics for certain administrative boundaries

RULES AND REGULATIONS

Participation

- Academic institution with any programs may participate in the competition.
- A team of three to five students is expected to analyze and work on the competition.
- Undergraduate (UG), graduates with less than 1 year of experience in industry and postgraduate (PG) students may take part in the challenge. A maximum of 2 PhD students per team is allowed.
- Students are only allowed to sign up in a team of the same university and more than one team from a university is welcomed to participate.
- Each participating team will need to pay for fee with the price of RM200.

All payment is non-refundable

Academic Advisors (optional)

- The academic advisor is responsible for the maintenance of the rules and procedures of the competition.
- Only one academic advisor is allowed for each team.
- The role of the academic advisor is to advise only.
- Every participating student team needs to be confirmed and approved by the academic advisor.

Withdrawal of team

- If a university decides to withdraw from the competition, no penalty will be considered in case the problem statement was not yet sent to the university. If the team withdraws from the competition after the problem statement was made available, IHR, CSI and SEACaRL may decide on a penalty depending on the reasons for withdrawal.

Plagiarism

- The inability of participants to explain their work will be considered as plagiarism.
- Therefore, every student should be aware of what plagiarism is and how to avoid it.
- IHR, CSI and SEACaRL takes plagiarism very seriously. Therefore, if the participants are unable to explain their work, it will be rejected on the base of plagiarism.

Case study

- Without the confirmation of IHR, CSI, SEACaRL and case study provider, the work of the students, as well as the problem statement itself, must remain within the walls of the represented university (host).
- No part of the case study may be reproduced or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without prior written permission of the provider. Violation of this rule will result in a university declared ineligible for the competition for two years.
- The assignment, or parts of it, cannot be presented to any person/company, apart from the Faculty Advisor (if any) only as a rehearsal, and the competition jury in the final stage presentation.

Consultancy

- All students must present only their own work or the work of the team members in their presentations. Universities that include work of others (fellow students, professors or other industry professionals) shall be disqualified from the competition.
- Only the team of three to five students should do the actual work. If fellow students, consultants or the faculty advisor would perform the work of the student, the team will be disqualified.

Prizes

- 5 finalist teams will be chosen and will receive an invitation to present a 15-minute presentation on their work.
- All teams will receive a certificate of participation.