Microsoft Azure Virtual Hackathon
Phase 2

Problem Definition:

Given the following attributes of a trip represented by a tuple:

- latitude_origin
- longitude_origin
- latitude_destination
- longitude_destination
- hour_of_day
- day_of_week

Build a ML model / algorithm to predict Expected Time of Arrival (ETA) in seconds.
You are required to implement your system and deploy your model as an Azure service. We will call your model for ETA prediction via an endpoint.

Submission:

1. Endpoint of deployed model
    - The inputs to the endpoint should include:

Inputs	Data Type	Examples	Notes
latitude_origin	float	-6.141255	Origin
longitude_origin	float	106.692710	Origin
latitude_destination	float	-6.141150	Destination
longitude_destination	float	106.693154	Destination
timestamp	bigint	1590487113	Departure time of a trip in UTC format.

hour_of_day	int	9	Converted from timestamp. Hour from 0 to 23 in UTC format.
day_of_week	int	1	Converted from timestamp. Monday is 0 and Sunday is 6 In UTC format.

- Return/output from the endpoint includes:

Output	Data Type	Example	Notes
eta	int	360	in seconds

- Timestamp is not required as an input to your model. But may be useful if you need to associate with external data sources.
- Please provide instructions and examples of how to use your endpoint in your PPT slides.

2. PPT slides of implementation
   - Please indicate your endpoint on the first page of your PPT.
   - Please document core implementations, including data sources, data preprocessing, feature transformation, modeling, experiments, model deployment, Azure architecture to support model serving etc..

3. Code repository or code snippets
   - Please provide codes of core components via a repository link, or code snippets, for verification purpose.

Evaluations:

Model will be tested on a hidden dataset, which will NOT be released to you:

1. Hidden datasets
   You can choose one of the following categories to test your model prediction accuracy. Please indicate your preference on the first page of your PPT.

	Dataset	Data collection period	Sample size
1	Singapore - Car	May 2019	10000

2	Jakarta - Car	May 2019	10000
3	Jakarta - Motorcycle	May 2019	10000

2. Metrics
   RMSE is used as a metric to evaluate the error of prediction against actual travel time. A model with the lowest RMSE value is considered as the best-performing one. Here, we have given a RMSE baseline of each category for your reference. It will be great if your model can exceed the baseline.

	Dataset	**RMSE baseline**
1	Singapore - Car	240
2	Jakarta - Car	320
3	Jakarta - Motorcycle	400

3. Implementations
   Besides prediction accuracy, we will also evaluate your actual implementations including modeling, experiments, deployment, scalability, Azure architecture.

FAQ

1. **What if my implementation in Phase 2 is different from what was proposed in Phase 1?**
   If some of the components in your proposal are hard to implement, you can look for alternative solutions. But please try to keep your implementation the same as what you had planned originally.

2. **Will there be a sub-competition according to the category of the hidden dataset?**
   No. There is no sub-competition in terms of the hidden dataset you choose.

3. **Is RMSE considered as the only evaluation criteria for Phase 2?**
   No. RMSE is only to evaluate your prediction accuracy. Modeling, experiments, deployment, scalability, Azure architecture will be evaluated from your code and PPT.

4. **Can I use ETA from Azure Maps or other ETA service providers?**
   No. You cannot use ETA values returned from other ETA services directly or as an input to your model. But you are allowed to get routes from those services and convert them as features to your model. If you do, please describe clearly in your PPT how you have used these routes in your solution.