Cost Estimation Problem

Introduction

Schenker acts mostly as an intermediary between customers who want to ship some goods and carriers who transport the goods.

In land transport, dispatchers receive orders of shipments and find suitable carriers. This is a process that relies almost entirely on human decisions and interactions. The dispatcher has some tools to access a list of available carriers, and contacts them until he finds someone with whom he agrees on a price. When the price is agreed, the carrier takes the responsibility of delivering the given shipment from the origin to the destination.

It is easy to imagine that deciding the price for a shipment is not an easy task, given the number of human factors that enter into play. For this reason, it is desirable to build a tool able to predict a price that the dispatchers can use as a guideline.

Your task

You are asked to build an accurate model able to predict the costs of a shipment. For this purpose, we are providing you some historical data, which we have divided into two disjoint data sets: `train_data.csv` and `test_data.csv`. In the latter one we have omitted the cost field, which is the target of the predictions. A description of the data columns can be found in the following page.

You are asked to

- train a model using the train data contained in `train_data.csv` and compute cost predictions for the cases contained in `test_data.csv`. Please provide us a file named `cost_predictions.csv` containing your predictions.
  Note 1: we provide you a file named `example_predicted_costs.csv` that contains an example of how the answer file should look like. Please conform to its format.
  Note 2: we will evaluate the quality of your predictions against the true values, using the R^2 (coefficient of determination) metric.
- provide us the code you produced for this task. Keep in mind that we will look at it, so it should be readable. Readability is an important factor that we will consider during the evaluation.

Provide us all the requested documents as attachments to the email:
david.zibriczky@dbschenker.com

Data Description

We have gathered some shipment records, encoding the following information:

- `origin_latitude`: latitude of the shipment's origin
- `origin_longitude`: longitude of the shipment's origin
- `destination_latitude`: latitude of the shipment's destination
- `destination_longitude`: longitude of the shipment's destination
- `weight`: weight of the shipped goods
- `loading_meters`: how many meters the shipment takes up on the truck's longest side
- `is_adr`: is it a dangerous good? (Boolean field)
- `shipping_date`: date when the shipment took place
- `cost`: how much did Schenker pay the carrier?

The fields `weight`, `loading_meters`, and `cost` have been normalized. All float values have been rounded to the third digit, except for the geo-coordinates, which have been rounded to the second digit.