FEM11152 – Data Science for Marketing Analytics

Case 2a

In this **2-page** report, you will create a Logistic Regression model to answer the following research question: will a consumer choose to travel by train or by car to his or her destination? The data set TransportMode.RData consists of 122 consumer choices between travelling by train or by car. A description of the variables can be found below:

- mode: the mode of transportation, either train (0) or car (1);
- ttime_t: the terminal waiting time for the train (minutes);
- invc_t: in-vehicle-cost for traveling by train (dollars);
- invt_t: in-vehicle-time for traveling by train (minutes);
- invc_c: in-vehicle-cost for traveling by car (dollars);
- invt_c: in-vehicle-time for traveling by car (minutes);
- hinc: household income;
- psize: traveling group size.

For the **2-page** report, you should at least do the following:

1. Explain the Logistic Regression model and summarize the problem.
2. Explain your choices in data preparation (if needed).
3. Divide the data into a training and a test set and use the training set to fit a model.
4. Provide an interpretation of your model. Comment on individual coefficients. For example, interpret the signs, magnitude, and significance of the coefficients. Did you expect these results?
5. Use the test set to assess the predictive performance of the model you have created.
6. Comment on your results.

The assignment will be graded according to the rubrics on Canvas