

Final Project Timeline

The final project includes three components: presentation and final project report. The percentages of each part to your final project score is given below:

1. Presentation on April 28 and 30 during our regular meeting time. Each group will give a 15-20 minutes' presentation and leave 5 minutes for questions and comments. This part contributes 20 percent to your final project score;
2. Final project report is due May 3, 2019. The final project report contributes 30 percent to your final project score. This will be evaluated based on your writing, interpretation, and completeness.
3. The models and methods you tried in your final project contributes to 20 percent to your final project. This will be evaluated based on the appropriateness of your models and methods, and the number of ways attempted.
4. Model prediction accuracy contributes 30 percent to your final project score. All the teams will work on the same training and test data sets. You will build the models based on the training data set and predict the results using test data set. The team with the highest prediction accuracy will receive the full 30 percent. The team with the second highest prediction accuracy will receive 27 percent. The third highest will receive 24 percent. More specifically, the team with the i -th highest prediction accuracy will receive $30 - (i - 1) * 3$ percent ($i = 1, 2, 3, 4, 5, 6, \dots$).

Dataset: Transportation Mode Detection with Unconstrained Smartphones Sensors

In this final project, we will work on a transportation mode detection data set that was downloaded from the following link:

<https://www.kaggle.com/fschwartz/tmd-dataset-5-seconds-sliding-window>

Identify user's transportation modes through observations of the user, or observation of the environment, is a growing topic of research, with many applications in the field of Internet of Things (IoT).

Transportation mode detection can provide context information useful to offer appropriate services based on user's needs and possibilities of interaction.

Sensors included in the first set (parameter 1) are accelerometer, sound, and gyroscope. These three sensors have the highest values of accuracy taken individually.

First dataset Dfirst is formed by twelve features, four for each sensor. Second dataset Dsecond is formed by eight sensor and thirty-two features. Third dataset Dthird formed by all nine relevant sensors and thirty-six features, differ from previous Dsecond only for speed derived features.

All the data sets contain 5894 rows and each row is considered as a sample. In your project, please use the first 4000 rows as training data set and use the remaining 1894 rows as testing data set.