Machine Learning with Accelerometer and GPS Data
Deliverable #6

## Data Files

- **exploratory/los_angeles_10/los_angeles_10.csv**
  - **Date**: the timestamp of a data point
  - **Speed**: traveling speed of the vehicle
  - **X**: forward acceleration (front and back) of the vehicle
  - **Y**: horizontal acceleration (left and right) of the vehicle
  - **Z**: vertical acceleration (up and down) with natural gravity (-1G)
- **exploratory/los_angeles_10/los_angeles_10_labeled.csv**
  - **Date**: the timestamp of a data point
  - **Speed**: traveling speed of the vehicle
  - **speedbump**: whether this data point is a speed bump
  - **forw_accel**: forward acceleration (front and back) of the vehicle
  - **hori_accel**: horizontal acceleration (left and right) of the vehicle
  - **vert_accel_G**: vertical acceleration (up and down) of the vehicle with natural gravity (-1G)
  - **vert_accel**: vertical acceleration (up and down) of the vehicle without natural gravity
  - **vert_accel_ratio_speed**: the ratio between vertical acceleration (without natural gravity) and traveling speed
  - **sq_vert_accel_ratio_speed**: the ratio between squared vertical acceleration (without natural gravity) and traveling speed
  - **vert_jolt**: vertical jolt of the vehicle (incremental change of vertical acceleration)
  - **vert_jolt_ratio_speed**: the ratio between vertical jolt and traveling speed
  - **sq_vert_jolt_ratio_speed**: the ratio between squared vertical jolt and traveling speed
  - **vert_jolt_mean**: 5-sliding-window mean of vertical jolt
  - **vert_jolt_sd**: 5-sliding-window standard deviation of vertical jolt
  - **vert_jolt_min**: 5-sliding-window minimum of vertical jolt
  - **vert_jolt_max**: 5-sliding-window maximum of vertical jolt
  - **vert_jolt_range**: 5-sliding-window range of vertical jolt
- **exploratory/los_angeles_11/los_angeles_11.csv**
  - **Date**: the timestamp of a data point
  - **Speed**: traveling speed of the vehicle
  - **X**: forward acceleration (front and back) of the vehicle
  - **Y**: horizontal acceleration (left and right) of the vehicle
  - **Z**: vertical acceleration (up and down) with natural gravity (-1G)
- **exploratory/los_angeles_11/los_angeles_11_labeled.csv**
  - **Date**: the timestamp of a data point

    - **Speed**: traveling speed of the vehicle
    - **speedbump**: whether this data point is a speed bump
    - **forw_accel**: forward acceleration (front and back) of the vehicle
    - **hori_accel**: horizontal acceleration (left and right) of the vehicle
    - **vert_accel_G**: vertical acceleration (up and down) of the vehicle with natural gravity (-1G)
    - **vert_accel**: vertical acceleration (up and down) of the vehicle without natural gravity
    - **vert_accel_ratio_speed**: the ratio between vertical acceleration (without natural gravity) and traveling speed
    - **sq_vert_accel_ratio_speed**: the ratio between squared vertical acceleration (without natural gravity) and traveling speed
    - **vert_jolt**: vertical jolt of the vehicle (incremental change of vertical acceleration)
    - **vert_jolt_ratio_speed**: the ratio between vertical jolt and traveling speed
    - **sq_vert_jolt_ratio_speed**: the ratio between squared vertical jolt and traveling speed
    - **vert_jolt_mean**: 5-sliding-window mean of vertical jolt
    - **vert_jolt_sd**: 5-sliding-window standard deviation of vertical jolt
    - **vert_jolt_min**: 5-sliding-window minimum of vertical jolt
    - **vert_jolt_max**: 5-sliding-window maximum of vertical jolt
    - **vert_jolt_range**: 5-sliding-window range of vertical jolt
- **exploratory/los_angeles_12/los_angeles_12.csv**
    - **Date**: the timestamp of a data point
    - **Speed**: traveling speed of the vehicle
    - **X**: forward acceleration (front and back) of the vehicle
    - **Y**: horizontal acceleration (left and right) of the vehicle
    - **Z**: vertical acceleration (up and down) with natural gravity (-1G)
- **exploratory/los_angeles_12/los_angeles_12_labeled.csv**
    - **Date**: the timestamp of a data point
    - **Speed**: traveling speed of the vehicle
    - **speedbump**: whether this data point is a speed bump
    - **forw_accel**: forward acceleration (front and back) of the vehicle
    - **hori_accel**: horizontal acceleration (left and right) of the vehicle
    - **vert_accel_G**: vertical acceleration (up and down) of the vehicle with natural gravity (-1G)
    - **vert_accel**: vertical acceleration (up and down) of the vehicle without natural gravity

    - **vert_accel_ratio_speed**: the ratio between vertical acceleration (without natural gravity) and traveling speed
    - **sq_vert_accel_ratio_speed**: the ratio between squared vertical acceleration (without natural gravity) and traveling speed
    - **vert_jolt**: vertical jolt of the vehicle (incremental change of vertical acceleration)
    - **vert_jolt_ratio_speed**: the ratio between vertical jolt and traveling speed
    - **sq_vert_jolt_ratio_speed**: the ratio between squared vertical jolt and traveling speed
    - **vert_jolt_mean**: 5-sliding-window mean of vertical jolt
    - **vert_jolt_sd**: 5-sliding-window standard deviation of vertical jolt
    - **vert_jolt_min**: 5-sliding-window minimum of vertical jolt
    - **vert_jolt_max**: 5-sliding-window maximum of vertical jolt
    - **vert_jolt_range**: 5-sliding-window range of vertical jolt
- **exploratory/los_angeles_13/los_angeles_13.csv**
    - **Date**: the timestamp of a data point
    - **Speed**: traveling speed of the vehicle
    - **X**: forward acceleration (front and back) of the vehicle
    - **Y**: horizontal acceleration (left and right) of the vehicle
    - **Z**: vertical acceleration (up and down) with natural gravity (-1G)
- **exploratory/los_angeles_13/los_angeles_13_labeled.csv**
    - **Date**: the timestamp of a data point
    - **Speed**: traveling speed of the vehicle
    - **speedbump**: whether this data point is a speed bump
    - **forw_accel**: forward acceleration (front and back) of the vehicle
    - **hori_accel**: horizontal acceleration (left and right) of the vehicle
    - **vert_accel_G**: vertical acceleration (up and down) of the vehicle with natural gravity (-1G)
    - **vert_accel**: vertical acceleration (up and down) of the vehicle without natural gravity
    - **vert_accel_ratio_speed**: the ratio between vertical acceleration (without natural gravity) and traveling speed
    - **sq_vert_accel_ratio_speed**: the ratio between squared vertical acceleration (without natural gravity) and traveling speed
    - **vert_jolt**: vertical jolt of the vehicle (incremental change of vertical acceleration)
    - **vert_jolt_ratio_speed**: the ratio between vertical jolt and traveling speed
    - **sq_vert_jolt_ratio_speed**: the ratio between squared vertical jolt and traveling speed

    - **vert_jolt_mean**: 5-sliding-window mean of vertical jolt
    - **vert_jolt_sd**: 5-sliding-window standard deviation of vertical jolt
    - **vert_jolt_min**: 5-sliding-window minimum of vertical jolt
    - **vert_jolt_max**: 5-sliding-window maximum of vertical jolt
    - **vert_jolt_range**: 5-sliding-window range of vertical jolt
- **exploratory/los_angeles_14/los_angeles_14.csv**
    - **Date**: the timestamp of a data point
    - **Speed**: traveling speed of the vehicle
    - **X**: forward acceleration (front and back) of the vehicle
    - **Y**: horizontal acceleration (left and right) of the vehicle
    - **Z**: vertical acceleration (up and down) with natural gravity (-1G)
- **exploratory/los_angeles_14/los_angeles_14_labeled.csv**
    - **Date**: the timestamp of a data point
    - **Speed**: traveling speed of the vehicle
    - **speedbump**: whether this data point is a speed bump
    - **forw_accel**: forward acceleration (front and back) of the vehicle
    - **hori_accel**: horizontal acceleration (left and right) of the vehicle
    - **vert_accel_G**: vertical acceleration (up and down) of the vehicle with natural gravity (-1G)
    - **vert_accel**: vertical acceleration (up and down) of the vehicle without natural gravity
    - **vert_accel_ratio_speed**: the ratio between vertical acceleration (without natural gravity) and traveling speed
    - **sq_vert_accel_ratio_speed**: the ratio between squared vertical acceleration (without natural gravity) and traveling speed
    - **vert_jolt**: vertical jolt of the vehicle (incremental change of vertical acceleration)
    - **vert_jolt_ratio_speed**: the ratio between vertical jolt and traveling speed
    - **sq_vert_jolt_ratio_speed**: the ratio between squared vertical jolt and traveling speed
    - **vert_jolt_mean**: 5-sliding-window mean of vertical jolt
    - **vert_jolt_sd**: 5-sliding-window standard deviation of vertical jolt
    - **vert_jolt_min**: 5-sliding-window minimum of vertical jolt
    - **vert_jolt_max**: 5-sliding-window maximum of vertical jolt
    - **vert_jolt_range**: 5-sliding-window range of vertical jolt
- **exploratory/los_angeles_video/los_angeles_video.csv**
    - **Date**: the timestamp of a data point
    - **Speed**: traveling speed of the vehicle
    - **X**: forward acceleration (front and back) of the vehicle

    - **Y**: horizontal acceleration (left and right) of the vehicle
    - **Z**: vertical acceleration (up and down) with natural gravity (-1G)
- **exploratory/los_angeles_video/los_angeles_video_labeled.csv**
    - **Date**: the timestamp of a data point
    - **Speed**: traveling speed of the vehicle
    - **speedbump**: whether this data point is a speed bump
    - **forw_accel**: forward acceleration (front and back) of the vehicle
    - **hori_accel**: horizontal acceleration (left and right) of the vehicle
    - **vert_accel_G**: vertical acceleration (up and down) of the vehicle with natural gravity (-1G)
    - **vert_accel**: vertical acceleration (up and down) of the vehicle without natural gravity
    - **vert_accel_ratio_speed**: the ratio between vertical acceleration (without natural gravity) and traveling speed
    - **sq_vert_accel_ratio_speed**: the ratio between squared vertical acceleration (without natural gravity) and traveling speed
    - **vert_jolt**: vertical jolt of the vehicle (incremental change of vertical acceleration)
    - **vert_jolt_ratio_speed**: the ratio between vertical jolt and traveling speed
    - **sq_vert_jolt_ratio_speed**: the ratio between squared vertical jolt and traveling speed
    - **vert_jolt_mean**: 5-sliding-window mean of vertical jolt
    - **vert_jolt_sd**: 5-sliding-window standard deviation of vertical jolt
    - **vert_jolt_min**: 5-sliding-window minimum of vertical jolt
    - **vert_jolt_max**: 5-sliding-window maximum of vertical jolt
    - **vert_jolt_range**: 5-sliding-window range of vertical jolt

## R Scripts:

- **exploratory/los_angeles_10/los_angeles_10.Rmd**
    - Input data file: **los_angeles_10.csv**
    - Data Cleaning & Exploratory Data Analysis
        - Epoch 1: Display of Vertical Acceleration with Natural Gravity
        - Epoch 2: Display of Vertical Acceleration without Natural Gravity
        - Epoch 3: Display of Ratio between Vertical Acceleration (w/o Natural G) and Speed
        - Epoch 4: Display of Ratio between Vertical Acceleration (w/o Natural G) and Speed

        - Epoch 5: Display of Vertical Jolt (Incremental Change of Vertical Acceleration)
        - Epoch 6: Display of Ratio between Vertical Jolt and Speed
        - Epoch 7: Display of Ratio between Squared Vertical Jolt and Speed
        - Epoch 8: Display of Sliding-Window Statistics of Vertical Jolt
        - Epoch 9: Comparative Display of Two Most Promising Factors
    - Output data file: **los_angeles_10_labeled.csv**
- **exploratory/los_angeles_11/los_angeles_11.Rmd**
    - Input data file: **los_angeles_11.csv**
    - Data Cleaning & Exploratory Data Analysis
        - Epoch 1: Display of Vertical Acceleration with Natural Gravity
        - Epoch 2: Display of Vertical Acceleration without Natural Gravity
        - Epoch 3: Display of Ratio between Vertical Acceleration (w/o Natural G) and Speed
        - Epoch 4: Display of Ratio between Vertical Acceleration (w/o Natural G) and Speed
        - Epoch 5: Display of Vertical Jolt (Incremental Change of Vertical Acceleration)
        - Epoch 6: Display of Ratio between Vertical Jolt and Speed
        - Epoch 7: Display of Ratio between Squared Vertical Jolt and Speed
        - Epoch 8: Display of Sliding-Window Statistics of Vertical Jolt
        - Epoch 9: Comparative Display of Two Most Promising Factors
    - Output data file: **los_angeles_11_labeled.csv**
- **exploratory/los_angeles_12/los_angeles_12.Rmd**
    - Input data file: **los_angeles_12.csv**
    - Data Cleaning & Exploratory Data Analysis
        - Epoch 1: Display of Vertical Acceleration with Natural Gravity
        - Epoch 2: Display of Vertical Acceleration without Natural Gravity
        - Epoch 3: Display of Ratio between Vertical Acceleration (w/o Natural G) and Speed
        - Epoch 4: Display of Ratio between Vertical Acceleration (w/o Natural G) and Speed
        - Epoch 5: Display of Vertical Jolt (Incremental Change of Vertical Acceleration)
        - Epoch 6: Display of Ratio between Vertical Jolt and Speed
        - Epoch 7: Display of Ratio between Squared Vertical Jolt and Speed
        - Epoch 8: Display of Sliding-Window Statistics of Vertical Jolt
        - Epoch 9: Comparative Display of Two Most Promising Factors
    - Output data file: **los_angeles_12_labeled.csv**

- **exploratory/los_angeles_13/los_angeles_13.Rmd**
  - Input data file: **los_angeles_13.csv**
  - Data Cleaning & Exploratory Data Analysis
    - Epoch 1: Display of Vertical Acceleration with Natural Gravity
    - Epoch 2: Display of Vertical Acceleration without Natural Gravity
    - Epoch 3: Display of Ratio between Vertical Acceleration (w/o Natural G) and Speed
    - Epoch 4: Display of Ratio between Vertical Acceleration (w/o Natural G) and Speed
    - Epoch 5: Display of Vertical Jolt (Incremental Change of Vertical Acceleration)
    - Epoch 6: Display of Ratio between Vertical Jolt and Speed
    - Epoch 7: Display of Ratio between Squared Vertical Jolt and Speed
    - Epoch 8: Display of Sliding-Window Statistics of Vertical Jolt
    - Epoch 9: Comparative Display of Two Most Promising Factors
  - Output data file: **los_angeles_13_labeled.csv**
- **exploratory/los_angeles_14/los_angeles_14.Rmd**
  - Input data file: **los_angeles_14.csv**
  - Data Cleaning & Exploratory Data Analysis
    - Epoch 1: Display of Vertical Acceleration with Natural Gravity
    - Epoch 2: Display of Vertical Acceleration without Natural Gravity
    - Epoch 3: Display of Ratio between Vertical Acceleration (w/o Natural G) and Speed
    - Epoch 4: Display of Ratio between Vertical Acceleration (w/o Natural G) and Speed
    - Epoch 5: Display of Vertical Jolt (Incremental Change of Vertical Acceleration)
    - Epoch 6: Display of Ratio between Vertical Jolt and Speed
    - Epoch 7: Display of Ratio between Squared Vertical Jolt and Speed
    - Epoch 8: Display of Sliding-Window Statistics of Vertical Jolt
    - Epoch 9: Comparative Display of Two Most Promising Factors
  - Output data file: **los_angeles_14_labeled.csv**
- **exploratory/los_angeles_video/los_angeles_video.Rmd**
  - Input data file: **los_angeles_video.csv**
  - Data Cleaning & Exploratory Data Analysis
    - Epoch 1: Display of Vertical Acceleration with Natural Gravity
    - Epoch 2: Display of Vertical Acceleration without Natural Gravity
    - Epoch 3: Display of Ratio between Vertical Acceleration (w/o Natural G) and Speed

        - Epoch 4: Display of Ratio between Vertical Acceleration (w/o Natural G) and Speed
        - Epoch 5: Display of Vertical Jolt (Incremental Change of Vertical Acceleration)
        - Epoch 6: Display of Ratio between Vertical Jolt and Speed
        - Epoch 7: Display of Ratio between Squared Vertical Jolt and Speed
        - Epoch 8: Display of Sliding-Window Statistics of Vertical Jolt
        - Epoch 9: Comparative Display of Two Most Promising Factors
    - Output data file: **los_angeles_video_labeled.csv**

## Python Scripts:

- **sklearn_Models/sklearn_CVGrid.py**
    - Parameter estimation using grid search with cross validation
    - sklearn.model_selection.GridSearchCV Model 1
        - Features: speed, X-accel, Y-accel, Z-accel, Z-jolt
        - Labels: speedbump (1 = yes, 0 = no)
        - Average F1 score: 0.96.3
        - StdDev F1 score: 0.052
        - Learning rate = 0.01
        - Minimum sample leafs = 1
        - N estimators = 150
        - Minimum sample split = 6
        - Max features = 4
- **sklearn_Kenya/sklearn_DTClassifier.py**
    - Locate speed bumps using sklearn.tree.DecisionTreeClassifier
    - Testing scores (F1 score):
        - Features: speed, X-accel, Y-accel, Z-accel, Z-jolt (From the data point and the three previous data points)
        - Labels: speedbump (1 = yes, 0 = no)
        - Average F1 score: 1.0
        - StdDev F1 score: 0.0
        - Median F1 score: 1.0
        - IQR F1 score: 0.0
        - Skewness F1 score: 0.0
        - Zero F1 score: 0