UUM540E – Engineering Project in Aerospace: Artificial Intelligence

with Applications to Aerospace Systems

Spring 2018-2019

Project #1

Due Date: 28 March 2019, 23:00

(Please do not forget to hand in all associated code with your HW)

Project – 1: Supervised Learning Algorithms

Problem Definition: *In This Project, you are going to investigate the performance of some supervised learning algorithms on real-word IoT data. You aim is to detect whether any human is in the room. Using the provided data, you are going to train supervised machine learning algorithms and test their performance on detection and classification.*

Project Steps:

1. **Read the paper:**
   a. ***"Accurate occupancy detection of an office room from light, temperature, humidity and CO2 measurements using statistical learning models"*** by Luis M. Candanedo, Véronique Feldheim. (http://www.sciencedirect.com/science/article/pii/S0378778815304357)
   b. You can also read more papers about this topic. If you do this, please attach those papers, add references to your report and explain the importance and contribution to your work.
2. **Investigate the data:**
   a. Provide important visualizations for data and explain the correlations, problems, probabilistic distributions …etc.
   b. Clean the data if necessary
   c. Complete the missing parts of the data if necessary
   d. Select your features and explain why? Provide detailed explanation about your feature selection process.
   e. You have to add at least ***one new feature*** to pass this project. Add at least one new feature and explain its effect on your project. (Adding more features to your model will be graded as more points for this project!). Provide detailed explanation about your feature engineering process.

3. **Build Train, Test and Validation Data Sets:**
   a. According to your choice, build your train-test or train-test-validation sets. (It is up to you to combine and rebuilt data sets or use them as is.)
   b. If you use some nonlinear transformation on your data sets provide this information and don't forget to apply the same operation on your test and validation data sets.

c. Three data sets are provided to you to train and test the classification models. For all the data sets, the temperature, humidity, the derived humidity ratio, light, CO2, occupancy status (0 for non-occupied, 1 for occupied) and time stamp are defined.

   i. **Training**, 8143 of 7 variables. ***Measurements taken mostly with the door closed during occupied status***

   ii. **Testing 1**, 2665 of 7 variables. ***Measurements taken mostly with the door closed during occupied status***

   iii. **Testing 2**, 9752 of 7 variables. ***Measurements taken mostly with the door open during occupied status***

4. **Build, Train and Test Your Statistical Machine Learning Algorithms:**
   a. Build your Machine Learning Models for Classification task given below:
      i. *Logistic Regression*
      ii. *Naïve Bayes*
      iii. *K-Nearest Neighbors*
      iv. *Decision Tree*
      v. *Random Forest*
      vi. *Gradient Boosting Machines*
      vii. *Kernelized Support Vector*
      viii. *Other classification algorithms you choose (Optional, graded more points!)*
   b. For every algorithm you use, tune its hyper parameters and note its accuracy, precision, recall values and built confusion matrix. Give all the results in a table format.
   c. For some algorithms, plot the feature importance results.
   d. Explain the results and investigate underfitting or overfitting and how you resolve these problems.
   e. Try to use different feature combinations with your model and explain the effects.
   f. **Your performance will be evaluated and graded for *Test Set 2*.**

5. **Report:**
   a. After completed your study, prepare a report which includes all your steps, outcomes and results in a great detail.
   b. Your code must be submitted as a python notebook.
   c. Your report must be in pdf format.

Notes:

- ***Any form of plagiarism is strictly prohibited. Violation will be subjected to 'no credit' for the course.***

Good Luck!

Dr. M. Umut DEMİREZEN

Prof. Gokhan INALHAN