Software Architecture Design for Classroom Occupancy Prediction

by

Abraham Montilla
Nikolay Bandura
Kristen McIntyre
Mengdi Yue
Svetlana Zolotareva

Submitted in partial fulfillment of the requirements for Data Science Certificate Program

Georgetown University School of Continuing Studies
2017

The purpose of the project is to create a room occupancy multi-class classifier using the following sensors: humidity, temperature, light , CO_2, bluetooth signals, camera, motion, door and noise. Some of these sensors collect data on regular basis with one minute intervals, the others trigger asynchronously, when captured by a sensor. To label the data, the number of people going in and out the room will be monitored by manual/automatic logging during class. For the exploration phase, the project will include external data such as weather information and readings from another similar project in an attempt to have more features and observations for the machine learning phase. Depending on correlations the team will decide if this data is worthwhile.

All the input data will be stored in MySQL Raw WORM Database. By wrangling such information we will generate a new database in MySQL representing the normalized wrangled data. This process includes standardization of timeframes across all data, as well as proper aggregation of information.

Using python libraries, such as pandas and matplotlib, we will perform exploratory data analysis and finalize our prediction model concept. Upon the successful completion of the concept, we are considering TensorFlow and/or Scikit-learn as our machine learning tools in the project. After this phase we will be ready to summarize our results and support or contradict original hypothesis. Also, our intended software to make such reporting includes Tableau, Matplotlib and Jupyter Notebook and PDF reporting. Please refer to figure 1 for the full architecture proposed for the project.
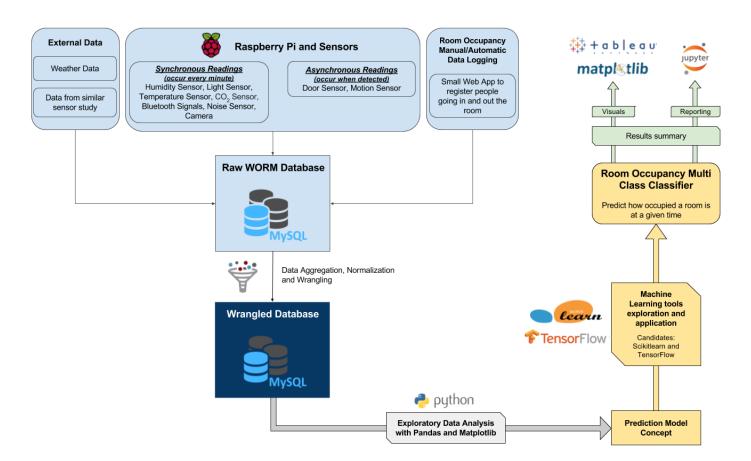
Figure 1. Architecture Diagram