

## **Northeastern University**

### **7390:**

### **Reading Summary**

In Energy and Building, the authors believe GBM is one of the best among four predictive models: (a) multiple linear regression, (b) support vector machine with radial kernel, (c) random forest and (d) gradient boosting machines (GBM). In order to find the best model, there are several steps.

First, specify data sources: Since the domestic appliances used by the occupants would leave traceable signals to the environment like temperature, humidity, vibrations, light and noise. The data can get from the sensor in the houses. Data like Appliances energy consumption, Light energy consumption, T1 Temperature in kitchen and so on (list in Table 2)

Second, Exploratory Data Analysis. Create graphs to understand the dataset. Some graph has been drawn for analysis. Fig.1 Shows different proportion of energy consumption for different devices (different data). Fig.7 shows the different days of the appliance consumption. Some new variable has been created to analyze the data. Like number of seconds from midnight for each Day (NSM), the week status (weekend and workday) and the day of the week.

Third, Feature engineering. Find out the important features.

Fourth, Prediction algorithms. Four prediction algorithms have been applied to the research. They are (a) multiple linear regression, (b) support vector machine with radial kernel, (c) random forest and (d) gradient boosting machines (GBM).

Fifth, Feature Selection. The conclusion for this part is all the predictors can be considered relevant to minimize the RMSE, they all will be used to test four regression models.

Sixth, Model Validation and Selection. (d) gradient boosting machines (GBM) is believed as the best prediction algorithm.

The conclusion after this whole process is

- a) NSM was ranked as the most important to predict the appliances' consumption.
- b) The GBM ranked the pressure as the most important weather variable followed by the outdoor temperature, dew point temperature and so on.
- c) Data from a wireless sensor network has been proven to increase the prediction accuracy.