Potential Project Topics

Yun Choi

Ames, Iowa Housing Prices

- Problem Statement: Determine the sale price of a house given the home features provided in the Ames Housing dataset. (Regression problem)

- Data: 79 explanatory variables and sale price of 2,930 residential properties in Ames, Iowa from 2006-2010. (Source: Kaggle)

- Hypothesis: Home features (including number of rooms, lot size, year built, etc.) will help us determine the sale price of a given house.

HR Analytics

- Problem Statement: Determine which employees will leave the company using employee data from Kaggle which include salary level, satisfaction level, last evaluation, average monthly hours, etc. (Classification problem)
  - Different types of employees leaving (clustering problem)

- Data: 9 explanatory variables and the outcome variable (whether the employee has left) for 15,000 current and former employees. (Source: Kaggle)

- Hypothesis: Employees who work long hours, receive poor evaluation and low salary have a higher probability of leaving the company than those who don’t.

Prudential Life Insurance

- Problem Statement: Predict the risk rating of each life insurance applicant using applicant's age, physical attributes, employment, medical history of the applicant and family, etc. (Classification problem)

  - Determine which attributes are important (dimensionality reduction problem)

- Data: 100+ explanatory variables and risk rating (1-8; ordinal) of ~60,000 life insurance applicants. (Source: Kaggle)

- Hypothesis: Applicant's attributes will help determine their risk rating.