18.1

Describe analytics models and data that could be used to make good recommendations to the power company.

Summary:
I would consider a logistic regression model to attain the probability of a customer not paying , setting the threshold at 50% . Among those who identify as not paying ,I next run a KNN model to determine those who cannot(don't have the abilty to) pay and who just doesn't. Then, I will impose ARIMA on those households predicting their costs of power for the next time period. Finally, I use optimization to determine the most cost efficient way to determine who to cut off.

Logistic Regression:

Given: a) Binary variable of whether payment haven't been paid for over 3 months
b) Credit Score
c)Payment default history (to any company)
d)past payment history

Use: Logistic Regression with threshold of 50%

To: Identify households into two groups "Not going to pay" and "Paid/Will pay eventually"

Support Vector Model

Given: a) income
b) household size
c)employment status
d)whether or not receiving subsidies

Use: SVM

To: Classify Among those identified as "Not going to pay" into "" Unable to pay" and "Can pay and not paying.

ARIMA

Given: past usage data

Use: ARIMA

To: Estimate cost of power for next time period

Optimization

Given: a) current work force of technicians

b) time needed to shut down power

c) gas price/ (per kilometer)

d) location data

e) estimated cost of electricity

f) marginal cost of hiring new technician

Use: Optimization over a network model

To : determine best who to shut down and shortest routes