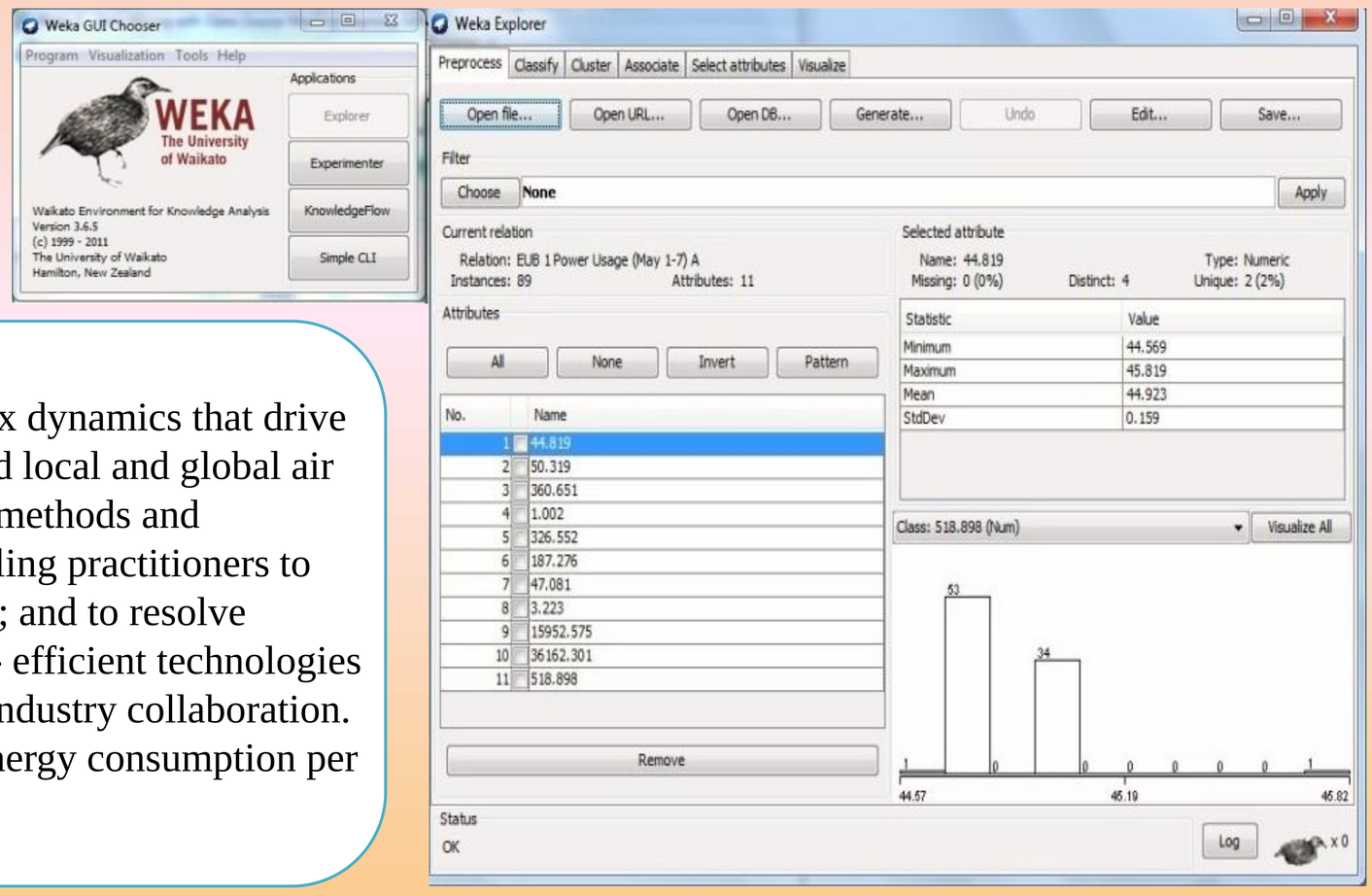
SAN DIEGO SUPERCOMPUTER CENTER

Introduction

Data mining, at its core, is the transformation of large amounts of data into meaningful patterns and rules. It can be further broken down into two types: directed and undirected. In directed data mining, you are trying to predict a particular data point — the sales price of a house given information about other houses for sale in the neighborhood, for example.



Abstract

UCSD seeks to advance the understanding of the complex dynamics that drive community-scale, end-use energy demand, and associated local and global air emissions; to apply this knowledge to generate planning methods and community design models and municipal processes enabling practitioners to build energy-efficient, low- carbon development projects; and to resolve market barriers and risks impeding integration of energy- efficient technologies into development projects through energy-development industry collaboration. This can be aided by monitoring and analyzing data of energy consumption per entity.

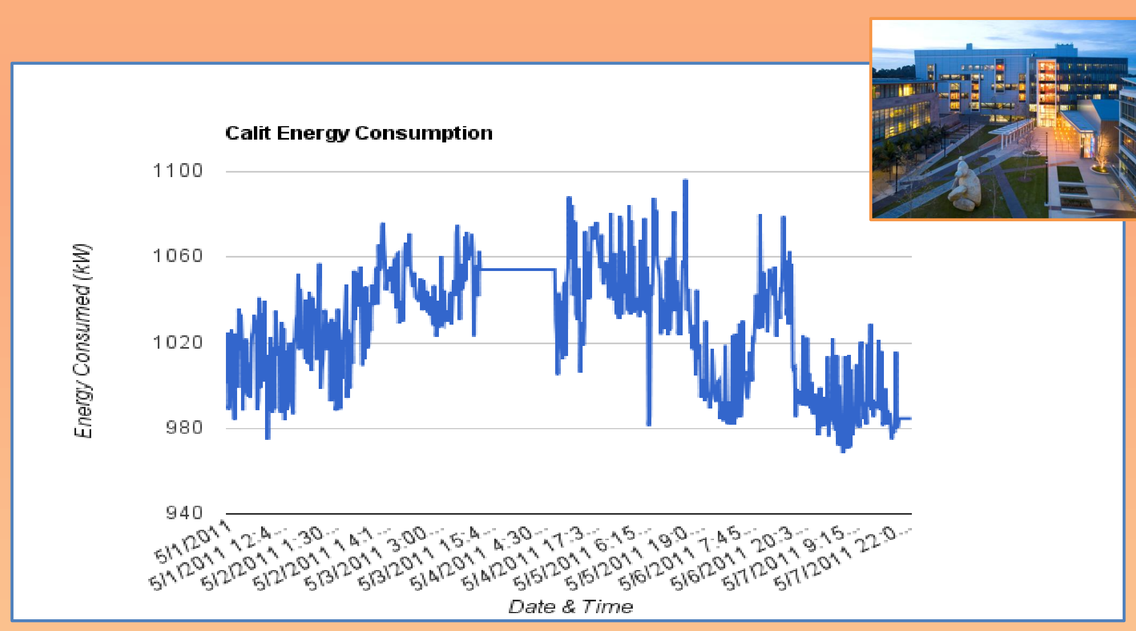
Renewable Energy / Microgrid System Analysis

Poster Created by: Goni Dubnov

Sponsored by the San Diego Supercomputer Center at UC San Diego

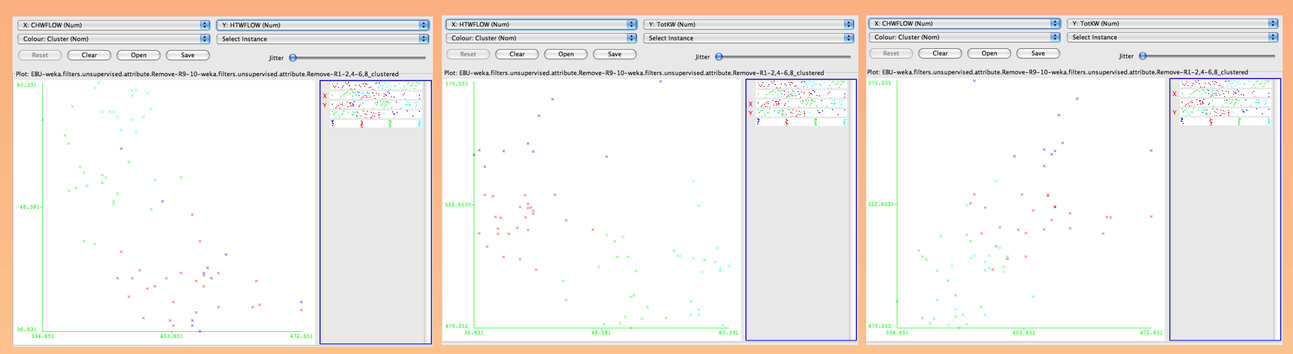
Story

Having a sufficient energy supply is vital for the community. The utility companies are responsible for maintaining and improving their service continuously. To facilitate better planning, utility companies maintain databases that capture the energy consumption and usage patterns of major appliances. These databases are then used to identify the trend and the usage of domestic energy. Currently, the utility companies use various statistical forecasting and regression approaches to estimate their future energy demand. However, with the recent rapid development of data mining, alternative estimation approaches such as decision trees and neural networks have become more popular and easier to operate. These decision trees and neural network analysis were run on the data below.



Methods

In this project we used a data mining program called WEKA in order to form clusters, showing relations or grouping of different attributes and visualize them in statistical models.

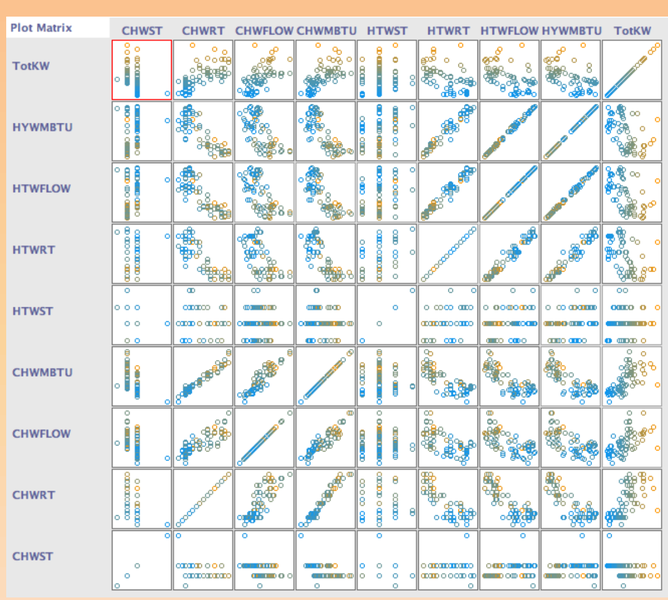


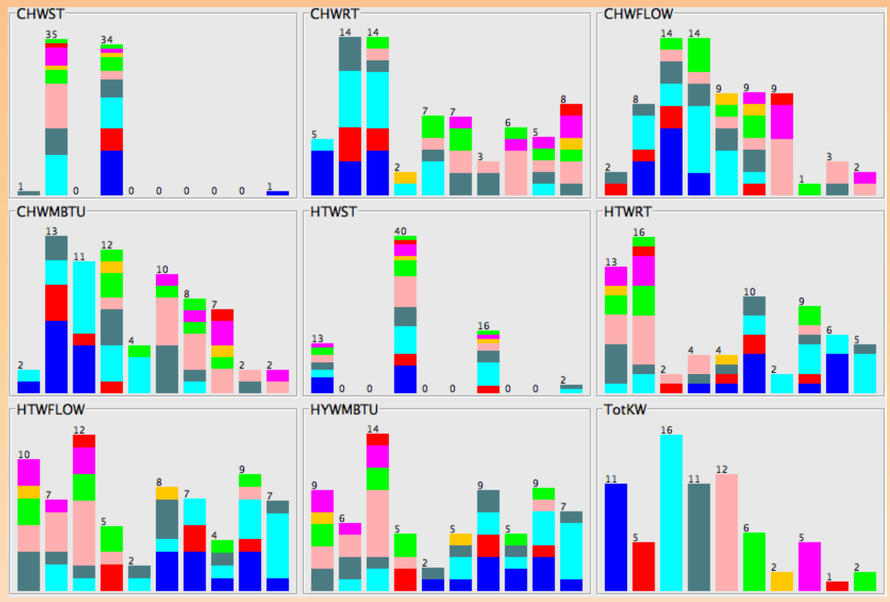
```
J48 pruned tree
------------------

CHWRT <= 50.819: '(-inf-505.186333]' (27.0/1.0)
CHWRT > 50.819
|   CHWFLOW <= 408.651
|   |   CHWRT <= 52.569
|   |   |   CHWFLOW <= 380.651: '(505.186333-540.120667]' (8.0/2.0)
|   |   |   CHWFLOW > 380.651: '(-inf-505.186333]' (11.0/4.0)
|   |   CHWRT > 52.569: '(505.186333-540.120667]' (5.0/1.0)
|   CHWFLOW > 408.651
|   |   HTWRT <= 185.276
|   |   |   HYWMBTU <= 2.91
|   |   |   |   HYWMBTU <= 2.676: '(540.120667-inf)' (2.0)
|   |   |   |   HYWMBTU > 2.676: '(505.186333-540.120667]' (10.0/1.0)
|   |   |   HYWMBTU > 2.91: '(540.120667-inf)' (4.0)
|   |   HTWRT > 185.276: '(505.186333-540.120667]' (4.0)

=== Summary ===

Correctly Classified Instances          45               63.3803 %
Incorrectly Classified Instances        26               36.6197 %
```





Summary

We applied directed and undirected data mining to data from the UCSD Smart Grid. Clustering and decision trees were explored for the purpose of statistical forecasting of energy consumption on UCSD Calit2 and EBU buildings. The research project included collection, preprocessing and applying WEKA data mining program to this data. After seeing the clustering, we wanted to see if the total energy consumption (TotKW) can be predicted from all the measured attributes. In order to do so we filtered the data into different number of nominal values and ran the J48 prediction tree. We found that 5 levels gives a good prediction that uses a lot of the measured attributes.