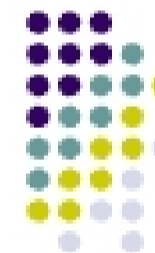


Data Mining



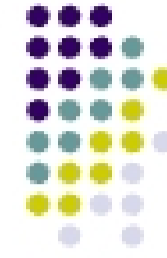
Introduction



- University of Iowa Power Plant is conducting study on their Boiler 11
- Wants to use this data to predict efficiency and classification accuracy



Problem Definition



- Data has been collected measuring 14 different variables within the boiler
- The original data set was mined
 - the rules determined from this data set only consider the static data
- The project will try to determine how the rules will change classification accuracy (CA) if you consider the trends of the data, instead of focusing on what the value was at a particular time



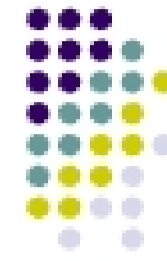
Project Goals



- The goal of the project is to determine if boiler efficiency can be predicted more accurately if the dynamic nature of the process is considered
- Create new features based on trends and current values
- Find which discretization level impacts the classification accuracy best
- Learned how to use data mining software



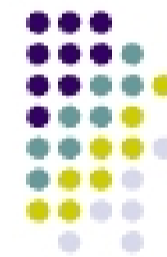
Data Collection



- Iowa City Power Plant raw data
 - 14 different variables
 - I.E. Boiler Master, Air Fuel Ratio, ect
 - About 10,000 different observations for each variable
- Data collected was collected at one minute intervals on randomly selected days
- Trends were created by looking at 20 minute intervals



Slope Calculation



- Calculate slope of raw data over time
 - Understood how the data was changing from point to point
 - Used 20 minute intervals
- Standardize Slope of data
 - Eased comparison
 - Improved transformation process

