Introduction

- There is a need to improve power management in terms of cost, reliability, and efficient service.
- Large amount of power related data gets accumulated every year, we can mine these data to get useful result.
- Analyzing relationship and patterns among all the events will help in better preparation for future crisis.

Objective

- Determining factors behind outage and analyzing relationship and patterns between various outage data.
- Building learning model with help of data mining techniques, which will be helpful in preventing future crisis.
- Automating power crisis alert system: it would be cost efficient, and helpful in better resource management.

Approach

- Preparing dataset in different phases of data gathering, data cleaning and merging.
- Two Approaches have been taken: Supervised Machine Learning and Unsupervised Machine Learning.
- Recursing through steps to improve the results; In terms of calculation and data manipulation.

Results

- Higher silhouette coefficient signifies better cluster.
- Below result shows most significant number of cluster is six.

System Overview

- Data collected from www.oe.netl.doe.gov
- Data Cleaning includes merging heterogeneous data set with five years of data. Removing outliers and missing values.
- Dataset contains mixed type of numeric, ordinal and nominal values.
- Significant Data Attributes: AreaAffected, NERCRegion, EventType, DateEventBegan, TimeEventBegan, DateofRestoration, TimeofRestoration
- Dataset Consist of around eleven attribute and twelve hundred training instances.
- For Supervised learning: classifier have been created with six classes. Each class signifies a level of emergency depending on the issue resolution time.
- For Unsupervised Learning: AreaAffected have been replaced with respective latitude and longitude values.
- Date and time have been converted into respective Julian and Numeric format.

Classification Results: Model is trained using J-48

- Below result shows Level A being most severe event but, its occurrence is significantly higher and comparable to other events.

Architecture

- PAM (Partitioning around medoids) is used from family of K-medoids, as it can operate on mixed data types.
- For PAM algorithm Fourteen distinct EventType values have been converted into binary format.
- In each iteration of K clustering. Silhouette coefficient has been used to determine cluster quality.
- Apart from natural disaster, various other events have been considered to better understand power outage scenario throughout the year, and with respect to affected areas.

Conclusions And Future Work

- Apart from natural disaster, there are other important factors to be considered like: vandalism, load shading.
- Power outage has some major severity issues, which are as frequent as other less severe issues.
- Clustering results can be improved significantly by building neural network with respect to the dataset.
- Types of attributes in the dataset are good but result can be improved with increasing the size of dataset.