MACHINE LEARNING FINAL PROJECT REPORT

POWER GENERATION IN US

Icon

Description automatically generated

### Instructor: Dr. Murali Shanker

Presented by

Sai Sree Pulimamidi

Reg No: 811223765

Executive Summary:

This report is the result of a study by the employee of Non – Governmental Organizations (NGO) for the programmatic review of PUDL dataset which have large amount of data regarding Power Generation in US which was published under the review of U.S. Department of Energy’s (DOE) to give their customers affordable, dependable electricity while also having as little of an impact on the environment as possible. By examining and clustering the PUDL dataset using K-means process there are five clusters formed which encompassed multiple findings and provided an output that using natural gas will be the best choice by considering both economic concerns and ecological effects.

Problem Statement:

* The objective of Power plants is to give their customers affordable, dependable electricity while also having as little of an impact on the environment as possible.
* Examining and clustering the PUDL dataset and finding out which cluster have highest Sulphur and ash emission

Techniques:

The PUDL dataset consists of 20 attributes. Firstly, starting with data preparing process removed all the unnecessary columns which consists of dates, code and transportation details as they are not important for finding out the best resource of power generation and dropped all the mandatory columns for future data processing and converted few of the categorical variables into numerical like 'contract\_type\_code' and 'fuel\_group\_code' and standardized the data for getting best results and applied k-Means method for clustering.

Why **K-Means Clustering?**

K-Means clustering is employed because, in contrast to other clustering techniques, it performs well for vast amounts of data and is beneficial for obtaining a specific number "K". In contrast to hierarchical clustering approaches, where decisions cannot be modified once clusters have been formed, decisions can be updated after the construction of clusters in k-means clustering.

Chart, line chart

Description automatically generatedChart, scatter chart

Description automatically generated

k-means clustering – Elbow Method

Graphical user interface, application, Word

Description automatically generated

k-means clustering – Silhouette Method

After using both Elbow and Silhouette k-means algorithms, the appropriate value of k is 5, and five clusters are created for the provided dataset.

Analysis and Discussion:

Bar chart, square

Description automatically generated

**CLUSTER – 0**

Inthis cluster the emission of sulfur and ash content is more compared to other cluster this signifies it consists of mixture of non-renewable resources like coal and natural gas and it is cost efficient, but this cluster will not be the best as it emits large amount of greenhouse gases which effects the environment.

**CLUSTER – 1**

In this cluster the efficiency is more, and the emission of sulfur is less but the emission of ash is more which indicates that it contains maximum amount of coal which is most abundant form of non-renewable resource and less expensive.

**CLUSTER – 2**

In this cluster the efficiency is more but all the harmful chemicals like mercury, ash, chlorine and ash emissions are also more, and it is cost inefficient so choosing this cluster will be not a good choice.

**CLUSTER – 3**

In this cluster the efficiency is more but harmful chemicals like mercury and ash contents are more which in future causes respiratory diseases for living beings and causes acid rains.

**CLUSTER – 4**

In this cluster the efficiency is less compared to above clusters but it releases less amount of sulfur and ash which will be less harmful for the ecology and coming to cost wise also it is comparative less so this will be the best cluster to adapt

Treemap chart

Description automatically generated

Conclusions:

The Evolution of electricity generating techniques are adapting more quickly than ever before because of the development of numerous, economic factors, the application of decentralized energy resources, and the development of digitalization. The best efficient fuel should have properties like availability, renewable, cost efficient and environmentally friendly. By above cluster analysis Natural gas will be the best fossil fuel to adapt as it is having more adaptable properties compared to other sources of power generation.