DECEMBER 12, 2022

MACHINE LEARNING

FINAL PROJECT

AKANKSHA NADUKULA
811237609
anadukul@kent.edu

Executive Summary:

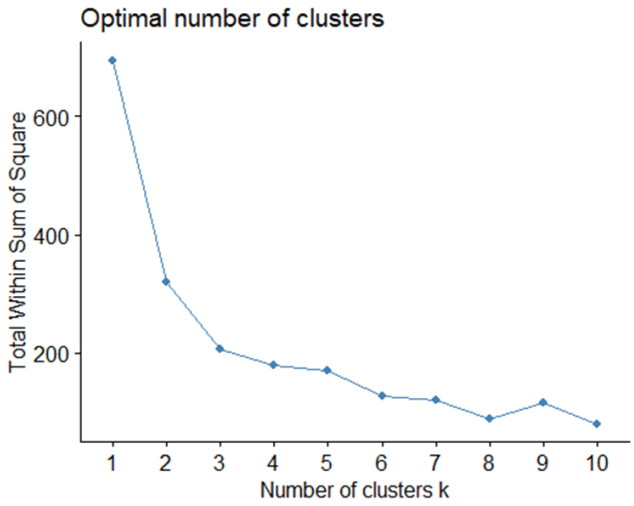
In this project, the data is taken from USA power generation dataset for analyzing power consumption by the people based on the wastage emission and cost per unit. For this purpose a 2% sample is extracted and then all the unwanted variables are removed. From that 75% of sample is taken as train and remaining as test set. Later clustering analysis is performed based on k means technique and obtained 2 clusters. The clusters determined the usage summary and based on that summary several conclusions has been made.

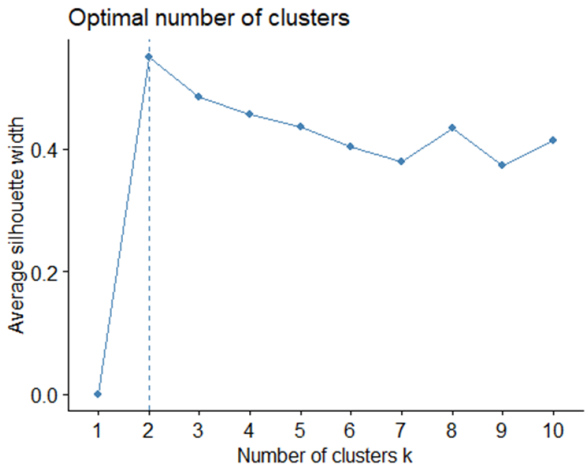
Problem statement:

Analyzing the fuel consumption by people based on wastage emissions like Sulphur, Ash and also cost per unit and provide a detailed report on the preferences of people power consumption.

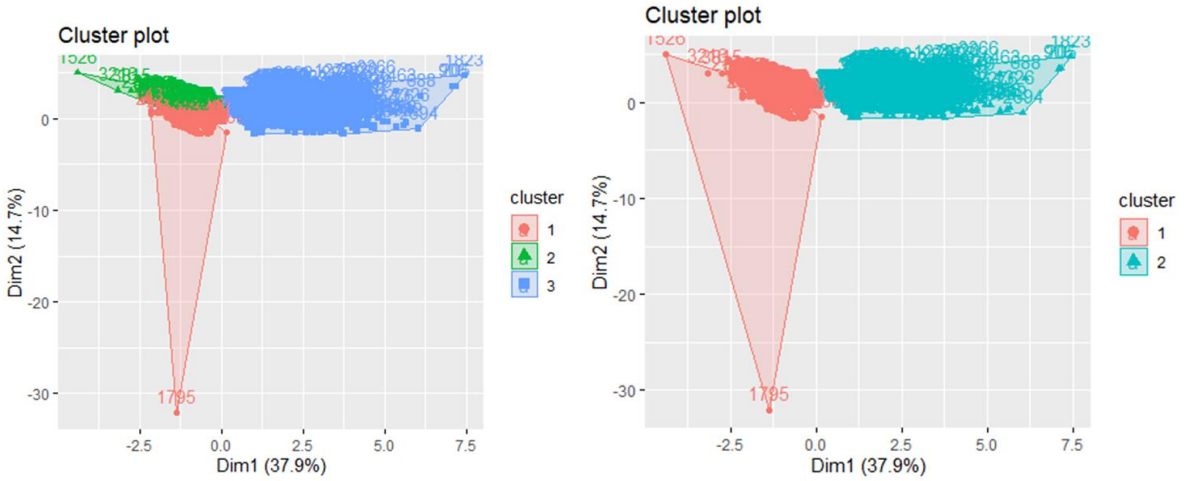
Techniques used:

A data sample from USA power generation dataset is taken and after handling all NA value exceptions. Clustering analysis is performed for that data after partitioning into train and validation sets. A method called K-Means divides data into clusters to examine commonalities. Since the data is huge, hierarchal clustering takes large amount to cluster hence we are choosing K means for clustering analysis. Using the K means performed graphs with two methods WSS and Silhouette results are 3, 2 centers respectively which can be seen from the plots.





Below is the plots for k means clusters with k=2 and k=3



We can observe that 2 clusters have no overlaps on them when compared to 3 cluster division.

Hence we selected with 2 clusters and further continued are analysis based on the 2 clusters that are obtained.

Conclusion:

From the above clusters, we can observe the average values for each parameter used

cluster <int>	Avg_receivedunits <dbl>	content_of_sulphur <dbl>	avg_ash <dbl>	avg_fuel_cost <dbl>
1	296268.67	0.002984683	0.000000	1.606212
2	56122.84	1.315212766	7.968337	20.927188

Cluster 1 has 296268.67 average units consumed when there is nearly 0 sulphur and ash contents whereas Cluster 2 has 56122.84 average units consumed when there is nearly 1.3% of sulphur and 7.9% ash content.

The clusters show that individuals tend to use less of those fuels and switch to more ecologically friendly fuels whenever there are high sulfur and ash emission levels. Additionally, we can see that costs rise as waste emissions increase. These are well described by cluster 1, and by doing so, we may conclude that people are becoming more environmentally aware and efficient energy users.

Appendix:

Code:

https://github.com/Nadukula/machine-learning.git

Data:

https://data.catalyst.coop/pudl/fuel_receipts_costs_eia92