MACHINE LEARNING

Final Project

Abinaya Sundari Panneerselvam
ID811224718

Executive Summary:

ABD company has owned several power plants that use fossil fuels to generate electricity. Here I as a Business Analyst, analyze the reference data set “PUDL” - which is public dataset about the US electricity generation, to analyze which fuel is more efficient in producing electricity and maximize the profit as well. Also, we are predicting the scope of fossil fuel in future using Kmean Clustering Algorithm.

Problem Statement:

1)Which fuel that produce high heat content with reasonable cost?

2)Which fuel that produce least sulfur and ash content?

3)What is the future scope of fossil fuel?

4)Which fuel gives more profit?

Technique:

Data cleaning: There was a significant amount of unstructured data was present which was transformed into structured data by imputing NA values and normalization.

The most popular unsupervised machine learning approach for dividing a given data set into a set of k groups is called "K-means clustering", where k denotes the number of groups the analyst has identified using two methods called wss and silhouette. Also, kmean algorithm simple to implement, scales to large dataset and guarantees convergence.

Finding optimal K value:

The value of k obtained by the elbow method is not clear, so I choose silhouette method. The range of the Silhouette value is between +1 and -1. A high value is desirable and indicates that the point is placed in the correct cluster. When a data point receives a score of 1, it means that its cluster is the largest and that it is the furthest apart from all other clusters. The worst number is one. Values close to 0 indicate clusters that overlap. So, the si = 0.78 (i.e.) si > 0, the model is well clustered

Conclusion:

- Cluster 1: Coal,
- Cluster 2: Gas, Oil.

Cluster Analysis:

RStudio: Notebook Output

A tibble: 2 × 6

clr_sil <int>	fuel_received_units <dbl>	fuel_mmbtu_per_unit <dbl>	fuel_cost_per_mmbtu <dbl>	sulfur_content <dbl>	ash_content <dbl>
1	22100.5	22.667	2.733	0.85	8.3
2	21348.0	1.030	7.390	0.00	0.0

2 rows

Finally, from the cluster1 and cluster2 we can address the problems. In cluster 1, fossil fuel "coal" – which receive average quantity of fuel 22100.5 that produce average amount of heat energy of 22.66 with the average cost of 2.73.

Here, in cluster 1, the average content of sulfur produce from heating coal is 0.85 which is moderate but high amount of ash content as 8.3.

In cluster 2, fossil fuel "gas" and "oil" – which produce average amount of heat energy of 1.03 with the average cost of 7.390, but cluster 2, with gas and oil does not emit sulfur and ash while heating it to produce electricity.

Problem 1: Which fuel that produce high heat content with reasonable cost?

Cluster 1, with fossil fuel coal produce high heat content in turn produce electricity with reasonable cost thus maximizing profit.

Problem 2: Which fuel that produce least sulfur and ash content?

Cluster 2, with gas and oil, does not radiate either sulfur or ash while heating the fuel to produce electricity.

Problem 3: What is the future scope of fuel which we can use to produce electricity?

According to cluster 2, In the future gas and oil will play important role in the production of electricity. Although, the coal is the cheapest and efficient fuel, but it will emit more co2 when heating it will cause global warming. So, in future use of coal will get reduced, but use of oil and gas will grow rapidly.

Problem 4: Which fuel gives more profit?

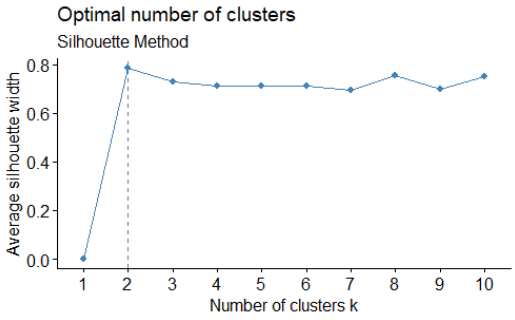
In the term of profit maximization, from cluster 1 we can see, using fossil fuel coal will give more profit because the availability of the coal is high also price required to transport coal is less than that of transporting oil or gas since we need separate pipeline to transport them. Coal is producing more heat at less price which in turn produce more electricity. Even though, the profit is high, burning coal will radiate more sulfur and ash content, but we can recycle the coal ash into products like concrete or wallboard.

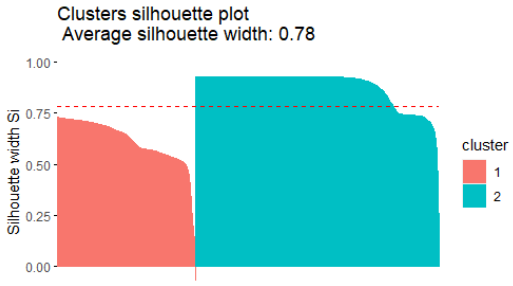
The sulfur emitted by coal also be removed by using method called scrubber. The scrubber works like, after sulfur is burned to produce SO2, the exhaust gas travels through a scrubber where SO2 is reacted with by a spray of limestone (or another chemical agent) and water. The

reaction allows for the removal of SO2 prior to its emission into the atmosphere. The more recent scrubbers used by Duke Energy are typically built to remove 95% or more of the SO2 from exhaust gases. Water vapor can be seen rising from the stack as a white plume.

Appendix:

Silhouette Method to Find Optimal K





Silhouette Average width (si)

Plotting Of Cluster for Fossil Fuels:

Reference:

https://www.duke-energy.com/our-company/environment/air-quality/sulfur-dioxide-scrubbers

https://www.epa.gov/coalash/coal-ash-basics#:~:text=Some%20power%20plants%20may%20dispose,products%20like%20concrete%20or%20wallboard.