



MACHINE LEARNING PROJECT

Badiganti, Karthik
kbadigan@kent.edu

EXECUTIVE SUMMARY

ABC Inc. is a big Multi-National Company decided to come into power generation industry. They wanted to know about the trends and insights about the fuel which is used by their competitors and the cost factors and recommendation of type of fuel composition needed for them to start the company. For this reason, a sample USA Power Generation dataset is taken to do the analysis. As part of analysis, sulphur and ash emissions are considered along with the type of fuel used and the cost per unit too. With the help of this data, clustering analysis has been performed and concluded that power generated using certain fuel type and certain amounts of sulphur and ash content, which will be briefly discussed below have higher chance of success in the industry.

Problem Statement

ABC Inc. is a big Multi-National Company wanted to get into power generation industry. They wanted to know about the market trends and insights about the companies which are already present in the industry and provide recommendations about the type of fuel used and what type of fuel composition is required to achieve success in the industry.

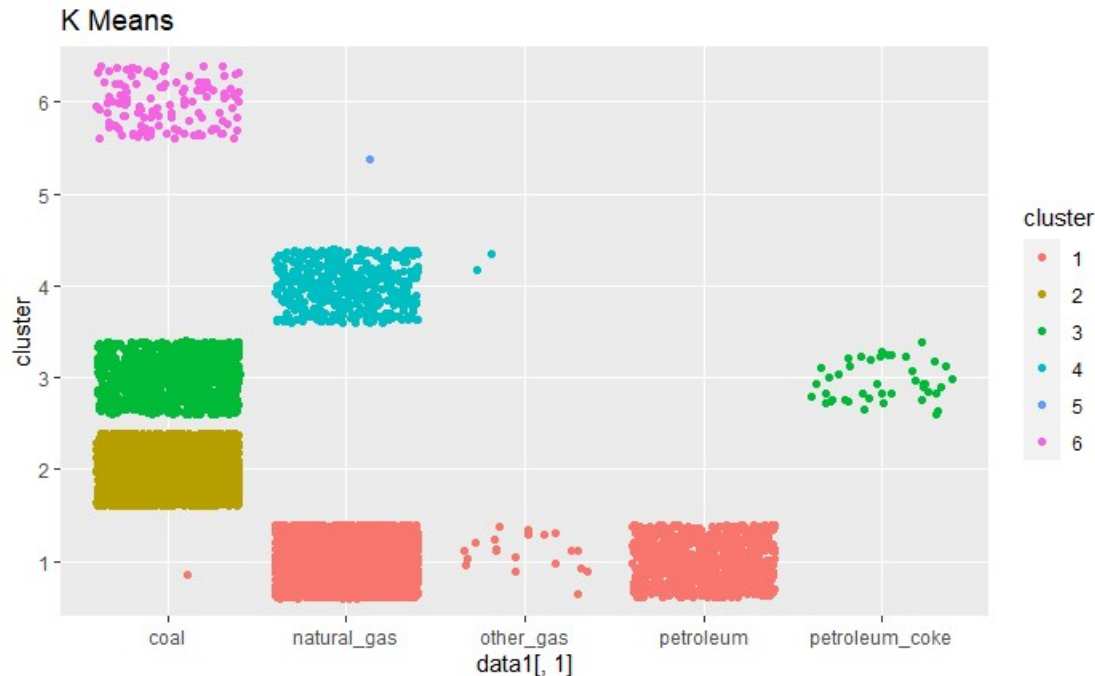
Technique Used

A sample of data is taken from USA power Generation dataset and sulphur content, ash content, fuel group, Suppliers, received fuel units and cost per unit. Initially it can be observed that natural gas has more than 54% while coal-based fuel has 34% market share. The average cost per unit for natural gas is 7.7, though the market share is more with natural gas-based power, choosing coal would be an ideal solution but the total amount of waste emission from coal is high.

fuel_group_code	avg_sulphur	avg_ash	avg_units	avg_mmbt	avg_cost	supplier_count
other_gas	0	0	659113.2	0.869136	4.538227	22
petroleum_coke	5.456428571	0.466429	20965.55	28.16493	2.235476	42
petroleum	0.185771567	0	5103.603	5.827943	14.89242	823
coal	1.347053653	9.930048	48735.37	21.20211	3.080754	3299
natural_gas	0	0	399887.3	1.029325	7.74306	4944

To reduce waste emission, suppliers who produce energy using coal with less emissions needs to be considered with high impact in the industry and less cost. For this purpose, clustering analysis is performed.

K-Means technique is used for clustering as it processes large data more efficiently and quickly when compared to other techniques. Clustering analysis is performed and there are 6 clusters formed in which coal is clustered further into 3 clusters ignoring cluster with one outlier point. It can be observed from the above data that cluster 2 has sulphur percentage of less than 1 percent and relative cost per unit is also less. Also, there are 3 more clusters classified along with the fuel groups which can be seen below.



Conclusion

From the above clustering, clusters 2,3,6 are the clusters which are clustered within coal-based power generation suppliers. Cluster 6 has high wastage emissions when compared to cluster 2 and cluster 2 has very less average cost per unit and maximum suppliers.

cluster	avg_sulphur	avg_ash	avg_units	avg_mmbt	avg_cost	supplier_count
1	0.4	0	258	12.51	2.65	1
6	1.307301587	40.76429	21225.59	13.13602	11.47267	126
3	2.888667306	10.5349	36806.12	23.66547	2.780894	1043
2	0.594612494	7.813542	56230.4	20.47676	2.731202	2129

From the above clustering data, the cluster 2 would be the optimal solution for the business. The recommendation would be as follows,

The company should follow a business strategy of suppliers who are currently producing power based on **coal with average Sulphur emission of 0.59% and ash emission of 7.8%.** This way the cost is optimal along with the competition being very less when compared to natural gas-based power suppliers.