



PROJECT REPORT

Fundamentals of Machine Learning



DECEMBER 12, 2022

NIHARIKA DOBANABOINA
Graduate student

Table of Contents:

S.no	Contents	Page no.
1.	Executive Summary	2
2.	Introduction	3
3.	Problem Statement	3
4.	Analysis and Discussions	3-4
5.	Conclusion	4

Executive Summary:

The objective of this project is to analyze various factors associated with the power generation in the U.S.A and suggest the type of fuel that can be excluded from the power generation process as the U.S. government is planning to preserve fossil fuels. As a result, analysis has been done based on the amount that is spent on each fuel to understand the type of fuel on which the government is spending the least. This type of fuel is assumed to be used the least and the government is willing to save such fuel for coming decades.

Using a machine learning algorithm, data has been grouped into 3 groups, each of them having one type of fuel. Further analysis has been done based on the average fuel cost, number of units of fuel received at power plants and their chemical composition. As per the results, Gas is the fuel type on which the government is spending the most and Oil is the one on which the U.S government is spending the least. Despite having lesser per unit cost for MMBtu, Coal is not widely used, and this could be because of the presence of impurities like ash, mercury, and sulfur in it.

Introduction:

The data has information related to the monthly fuel Contracts, Purchases and Costs. It has more than 20 variables ranging from the mine_ids from which fuels are to be supplied to different power plants by various suppliers to the type of transportation used to supply these fuels. For effective analysis purpose, data has been cleaned and certain variables with missed, and redundant data were excluded. From 608,565 rows, 2 percent of data has been sampled with a set. Seed (2467) which helps in better interpretation of the data. Furthermore, *fuel type code pudl* variable has been converted into a numerical variable by creating 3 dummy variables for 3 types of fuel.

Problem Statement:

The U.S Power generation Unit has hired a data analyst to study historical data of monthly_fuel_contract, Purchases, and Costs information to identify the type of fossil fuel on which they are spending the least money on. They are willing to cut down expenses on the type of fossil fuel on which they are spending the **least** money and exclude it from power generation. The reason stated behind this new approach is to preserve fossil fuels for future use which are not being used much.

Questions to focus on:

On which fuel are they spending the most and the least money?

Are there any other factors to consider while excluding a particular type of fuel from power generation?

As a data analyst, do you suggest any other alternative recommendations?

Analysis and Discussion:

After thoroughly analyzing the data using a machine learning algorithm, below are my findings:

- The average amount spent on Gas, Coal, and Oil are as follows:
 - Gas – 963,957.95 dollars
 - Oil – 53,227 dollars
 - Coal- 82,689.86 dollars

**Amount spent is calculated by multiplying the average number of fuel units received and the average per unit cost.*

- The U.S.A power generation is spending the maximum amount on **Gas** as it is the fuel that is received in the greatest number of units with lesser cost per unit.
- **Oil** is the fuel type on which they are spending the least. However, Oil is not just used in power generation but also to propel vehicles and in petrochemical industry to make products such as plastics, Solvents, and hundreds of other intermediate and end-user goods as per the information published by U.S. Energy Information Administration.

Source: [Link](#)

Therefore, it is not recommended to exclude Oil from power generation as this oil can be used for multiple reasons.

- In comparison to Oil, the amount spent on **Coal** is higher and as per the chemical composition of coal, it contains percentages of ash, mercury, and Sulfur impurities. In power plants, when power is generated by this type of fuel with impurities, additional expenses are to be covered as the greater the degree of these impurities, the higher would be the associated cost. The government should also consider these additional costs.
- When talking about preserving fossil fuels for future generations, as per the article posted by *International Energy Agency*, all unabated coal generation ends by 2040. To get on track with the Net Zero by 2050 Scenario, an annual average reduction of emissions from coal-fired power plants of around 8% is needed through 2030.
Source: [Link](#)
Hence, there is a need to preserve Coal for future generations.
- I would like to recommend the Government to deploy more efficient technologies like *Carbon Capture Utilisation and Storage* to generate power to preserve fossil fuels.

Conclusion:

The type of fuel to be excluded from the power generation is **Coal** as there is an immediate need to cut down the usage of Coal and preserve it for the future generations. Moreover, the amount spent on Coal is way too higher than on Gas and Oil. Hence, instead of excluding the fuel on which they are spending the least, I would suggest considering the fuel on which they are spending the most, especially when it is not even the best type of fuel.

Therefore, instead of spending huge amount on Coal, the government can redirect the same amount in Oil as it exists in its purest form which can also be used for multiple purposes.