

Paper ID: MISS21\_97

# A Study on Data Cleaning of Hydrocarbon Resource Under Deep Sea Water Using Imputation Technique Based Approach of Data Science.

Dipjyoti Deb

Department of Electronics & Communication Engineering, Techno College of Engineering Agartala Maheshkhola, West Tripura, India

Corresponding author

## Abstract

**Background:** A Hydrocarbon which is an organic compound is a combination of hydrogen and oxygen and also in all type crude oil, petroleum and natural gases, hydrocarbon found naturally. Hydrogen and carbon are present in different trees and plants in the surface. Unfortunately, the surface resource of hydrocarbon gets reducing everyday due to the rapid increase of industrialization [1]. Deep oceans are also an alternative resource of hydrocarbon. Millions of years ago microorganisms deposited and buried in the sediments over a period, and now by achieving sufficient pressure and temperature releasing the hydrocarbons. This hydrocarbon in the form of gas, leakages and spreads in the ocean and causes hazards for the sea animals [2][3]. This under sea fossil fuel can be the giant resource of future energy and can become a factor of development of blue economy.

**Objectives:** This paper mainly focuses in the data cleaning or data wrangling processes, consists of removing unwanted observations, missing data handling, structural error solving, outlier's management which is one of the most necessary phase of data science life cycle upon which the remaining phases namely exploratory data analysis (EDA), data modelling, model evaluation and model deployment are dependent.

**Methodology:** Same method of data cleansing may not be suitable for every domain. In this paper an imputation technique-based approach is applied to clean the dataset and to achieve a structured data.

**Results and discussions:** We have taken Polycyclic Aromatic Hydrocarbon Samples dataset (2011 to 2021) from California Department of Fish and Wildlife. A fuzzy matched and clustering based approach is given in [4]. This paper will provide a fast and more suitable approach to find the result compared with the above-mentioned method.

**Conclusion and future work:** Data Cleaning is essential to make analysis accurate and for error-free machine learning models otherwise whole efforts will be ruined. Data cleansing techniques of single-source and multi-source-based data quality problem is described in [5] may be further applied to get more structured and unique data.

## References

- [1] John Mawdsley, P. Goel. "Renewable vs Hydrocarbon-The energy Reality". Institutional Research AltraCrop Inc.
- [2] OLEG G. MIRONOV." Hydrocarbon pollution of the sea and its influence on marine organisms".Institute of Biology of the South Seas, A.S. Ukr. SSR Sevastopol, USSR. Helgoländer wiss. Meeresunters. 17, 335-339 (1968).
- [3] G. D. Floodgate ."Some environmental aspects of marine hydrocarbon bacteriology". School of Ocean Sciences. Marine Science Laboratories, University of Wales at Bangor, Gwynedd LL59 5EY, United Kingdom.Vol.9:3-11,1995.Published April 28.
- [4] Vaishali Chandrakant Wangikar, Ratnadeep R. Deshmukh." Data cleaning: Current approaches and issues", (<https://www.researchgate.net/publication/278301609>).
- [5] Erharo Rahm, Hong Hai Do. "Data cleaning: Problems and Current Approaches", <http://dbs.uni-leipzig.de>.

