JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Classification of Regional Coals using Unsupervised Clustering Method

[1]Shaikh Mohammed Raihan, [2]Ghugare Suhas B

[1]Postgraduate student at AISSMS College of Engineering, Pune, [2]Assistant Professor at AISSMS College of Engineering, Pune

[1]Department of Chemical Engineering,

[1]AISSMS College of Engineering, Pune, India

Abstract: In the current study, the coal's classification utilized in thermal power plants is reported. The coal samples were collected from India, Pakistan, Turkey, Indonesia and New Zealand. These coals were categorized using the conventional unsupervised clustering approach K-Means Clustering and Auto-Model, which selects the best clustering model iteratively. To carry out the aforementioned classification, coal can be divided in to water content, ashes, volatile matter, combustible carbon, and gross calorific value (GCV). Based on the lower Davies Bouldin index, the Auto-model based classifier performed better then k-means classifier. As stated by the study's classification findings, Power coals from the aforementioned specific regions may be categorized most efficiently using conventional and artificial intelligence methods into seven groups.

Keywords: *K-Means, Auto-Model, Clustering, Coal data.*

I. INTRODUCTION

Coal is a flammable black or brownish-black sedimentary rock that occurs in the form of coal seam. Coal is primarily composed of carbon, with trace amounts of other elements such as hydrogen, Sulphur, oxygen, and nitrogen. Heat and pressure of deep burial over millions of years convert dead plant matter into peat, which further converted into coal. Massive coal deposits were formed in former wetlands known as coal forests, which covered much of the Earth's tropical land areas during the late Carboniferous (Pennsylvanian) and Permian periods. Since many notable coal deposits are older, they originate during the Mesozoic and Cenozoic eras. [15]

Coal is primarily used as a fuel. Coal has been known and used for thousands of years, but its application was limited until the Industrial Revolution. Coal consumption increased with the invention of the steam engine. Coal provided about a quarter of the world's primary energy and more than a third of its electricity in 2020. [15] Coal is used in the production of iron and steel, as well as in other industrial processes.

Coal is the most important source of energy for electric power generation in the World, accounting for more than 39.1% of annual coal production. This is due to its abundant availability. It is estimated that over 2,400 coal fired power plant exist around the globe with combined capacity of over 2000 gigawatts. Steel, fertilizer, chemical, paper, and cement are the other major coal-consuming industries, in addition to thermal power plants. Carbon (38-60), volatile matter (1-36), Moisture (3-43), silicon oxide (45-63), aluminium oxide (15-36), iron oxides (2-20), calcium oxide (trace-12), magnesium oxide (trace), ash (3-60), Sulphur (0.3-8.3), and phosphorus (0.5) are the ranges (%) of various constituents of coal. Coal resources can be found in 29 major coalfields across World. According to estimates, World's total proven coal reserves of anthracite, bituminous, sub-bituminous, and lignite coals are 1,054,782 million tons, with anthracite and bituminous coals accounting for nearly 69.7%. Of the total coal output in 2021, 87.3% was estimated to be thermal coal, and 12.7% was coking coal. [17]

1.1 Coal classification

Coal classification is critical in technological and economic applications. There are several coal classification systems in use today, and new schemes are constantly being introduced. Coal classification serves three major goals: selection of coal for a specific industrial application, determination of coal grade or price for commercial purposes, and quantity/constituents/property-based categorization for resource assessment. Coal is commonly classified based on its rank and type. The rank of a coal describes the degree of metamorphism that it goes through during coalification as it matures from peat to anthracite. The rank has a significant impact on the physical and chemical properties of coal. Anthracite is at the top of the ranking scale, with higher carbon and energy contents and a lower moisture content. Low ranking coals, like lignite, are browner and softer friable materials with a

dull, earthy appearance. These contain a high amount of oxygen (up to 30%), a low amount of carbon (60-75% on a dry basis), and a high amount of moisture (30-70%). Generally thermal power plants, low rank coals are typically used. The energy content of these coals are low due to their high moisture and low carbon percentages. Another method of classifying coals is based on the organic debris, known as "Macerals," from which the coal was formed. [17]

1.2 Types

Under the right circumstances, dead biotic material is subjected to pressure by geological processes throughout time, increasing its metamorphic grade or rank into:

1. Lignite, or brown coal, is the lowest rank of coal, the most hazardous to health, and is almost entirely used as a fuel for electric power generation.
2. Jet is a compact form of lignite that is sometimes polished and has been used as an ornamental stone since the Upper Paleolithic period.
3. Peat, a precursor of coal
4. Sub-bituminous coal, whose properties are intermediate between lignite and bituminous coal, is primarily used as a fuel for steam-electric power generation Heavy Oil Recovery and Upgrading.
5. Bituminous coal is a dense sedimentary rock that is usually black but can also be dark brown and has well-defined bands of bright and dull material. It is primarily used as a fuel in the generation of steam-electric power and in the production of coke. Known as steam coal in the United Kingdom, it was historically used to generate steam in steam locomotives and ships.
6. Anthracite coal, the highest rank of coal, is a harder, glossy black coal that is primarily used for heating residential and commercial spaces.
7. Graphite is difficult to ignite and is not commonly used as a fuel; it is most commonly used in pencils or as a lubricant powder.
8. Cannel coal (also known as "candle coal") is a fine-grained, high-rank coal with a high hydrogen content that is primarily composed of liptinite. [25]

1.3 Composition

The composition of coal is reported either as a proximate analysis (moisture, volatile matter, fixed carbon, and ash) or as an ultimate analysis (moisture, volatile matter, fixed carbon, and ash) (ash, carbon, hydrogen, nitrogen, oxygen, and sulfur). The term "volatile matter" does not exist in and of itself (except for some adsorbed methane), but rather refers to the volatile compounds produced and released by heating coal. On a dry, ash-free basis, a typical bituminous coal may have an ultimate analysis of 84.4% carbon, 5.4% hydrogen, 6.7% oxygen, 1.7% nitrogen, and 1.8% Sulphur.

Following are the oxides present in the ASH

Table 1: oxide content in ash

Content	Percentage
SiO_2	20–40
Al_2O_3	10–35
Fe_2O_3	5–35
CaO	1–20
MgO	0.3–4
TiO_2	0.5–2.5
Na_2O& K_2O	1–4
SO_3	0.1–12

1.4 Emission intensity

Emission intensity is the amount of greenhouse gas emitted per unit of electricity generated by a generator over its lifetime. Coal power plants have a high emission intensity, emitting around 1000 g of CO_2eq per kWh generated, whereas natural gas has a medium emission intensity, emitting around 500 g CO_2eq per kWh. Coal emission intensity varies by type and generator technology, and in some countries exceeds 1200 g per kWh.

II. METHODOLOGY

2.1. Coal analysis

Coal analysis techniques are specific analytical methods for determining the physical and chemical properties of coals. These methods are primarily used to assess the suitability of coal for coking, power generation, or iron ore smelting in the steelmaking process.

2.1.1 Chemical properties of coal

Coal is classified into four types or ranks: lignite (brown coal), bituminous coal (black coal), anthracite, and graphite. Each type of coal has a unique set of physical properties that are primarily influenced by moisture, volatile content (in terms of aliphatic or aromatic hydrocarbons), and carbon content.

2.1.1.1 Moisture

Because all coal is mined wet, moisture is an important property of coal. The water absorbed in the ground can be called as adventitious moisture, and it evaporates quickly. The moisture contained within the coal itself is referred to as inherent moisture and is quantified. Within the coal, moisture can take four different forms:

Surface moisture: Water held on the surface of coal particles or macerals is referred to as surface moisture.

Hygroscopic moisture: The capillary action present inside the coal's microfractures which holds water, this water is known as Hygroscopic moisture

Decomposition moisture: water contained within the decomposed organic compounds of coal.

Mineral moisture: It is water that is part of the crystal structure of hydrous silicates like clays.

Total moisture is determined by calculating the mass difference between an untreated sample and the sample after it has been analyzed.

This can be done by any of the following methods:

- Heating the coal with toluene.
- Drying in a nitrogen atmosphere in a minimum free-space oven at 150 °C (302 °F).
- Drying in hot air which is at 100 to 105 °C (212 to 221 °F) and determining the relative mass loss.

From the above methods, first & second can be used for low-rank coals, but third method can be used only for high-rank coals because free air drying low-rank coals may promote oxidation. Inherent moisture is examined in the same way, though it can be done in a vacuum.

2.1.1.2 Volatile matter

Volatile matter in coal refers to the components of coal, except for moisture, which are liberated at high temperature in the absence of air. This is usually a mixture of short- and long-chain hydrocarbons, aromatic hydrocarbons and some sulfur. Volatile matter also evaluates the adsorption application of an activated carbon. The volatile matter of coal is determined under rigidly controlled standards. In Australian and British laboratories this involves heating the coal sample to 900 ± 5 °C (1650 ±10 °F) for 7 min. Also, as the rank of coal increases the volatile matter decreases (AMK).

2.1.1.3 ASH

Ash content of coal is the non-combustible residue left after coal is burnt. It represents the bulk mineral matter after carbon, oxygen, sulfur and water (including from clays) has been driven off during combustion. Analysis is fairly straightforward, with the coal thoroughly burnt and the ash material expressed as a percentage of the original weight. It can also give an indication about the quality of coal. Ash content may be determined as air dried basis and on oven dried basis. The main difference between the two is that the latter is determined after expelling the moisture content in the sample of coal.

2.1.1.4 Fixed carbon

The fixed carbon content of the coal is the carbon found in the material which is left after volatile materials are driven off. This differs from the ultimate carbon content of the coal because some carbon is lost in hydrocarbons with the volatiles. Fixed carbon is used as an estimate of the amount of coke that will be yielded from a sample of coal. Fixed carbon is determined by removing the mass of volatiles determined by the volatility test, above, from the original mass of the coal sample.

2.1.2 Special combustion tests

2.1.2.1Specific energy

Aside from physical or chemical analyses to determine the handling and pollutant profile of a coal, the energy output of a coal is determined using a bomb calorimeter which measures the specific energy output of a coal during complete combustion. This is required particularly for coals used in steam generation.

2.1.2.2Ash fusion test

The behavior of the coal's ash residue at high temperature is a critical factor in selecting coals for steam power generation. Most furnaces are designed to remove ash as a powdery residue. Coal which has ash that fuses into a hard glassy slag known as clinker is usually unsatisfactory in furnaces as it requires cleaning. However, furnaces can be designed to handle the clinker, generally by removing it as a molten liquid.

Ash fusion temperatures are determined by viewing a molded specimen of the coal ash through an observation window in a high-temperature furnace. The ash, in the form of a cone, pyramid or cube, is heated steadily past 1000 °C to as high a temperature as possible, preferably 1,600 °C (2,910 °F). The following temperatures are recorded;

- Deformation temperature: This is reached when the corners of the mold first become rounded.
- Softening (sphere) temperature: This is reached when the top of the mold takes on a spherical shape.
- Hemisphere temperature: This is reached when the entire mold takes on a hemisphere shape.
- Flow (fluid) temperature: This is reached when the molten ash collapses to a flattened button on the furnace floor.

2.1.2.3 Crucible swelling index (free swelling index)

The free swelling index test is the most basic method for determining whether a coal is suitable for coke production. A small sample of coal is heated in a standardized crucible to around 800 degrees Celsius (1500 degrees Fahrenheit).

A small coke button remains in the crucible after heating for a specified time, or until all volatiles are driven off. The Free Swelling Index is determined by comparing the cross-sectional profile of this coke button to a set of standardized profiles [9].

2.1.3 Coal classification by rank

Several international standards classify coals by their rank, where increasing rank corresponds to coal with a higher carbon content. The rank of coal is correlated with its geologic history, as described in Hilt's law.

In the ASTM system, any coal with more than 69% fixed carbon is classified by its content of carbon and volatiles. Coal with less than 69% fixed carbon is classified by its heating value. Volatiles and carbon are on a dry mineral free base; heating value is based on the moisture content as mined, but without any free water.

The ISO also has a coal ranking system, though its subdivisions do not align with the ASTM standard.

Table 2 constituent of coal [12]

Class	**Group**	**Fixed Carbon % Dry, mineral free**	**Volatile Matter % Dry, mineral free**	**Heating Value MJ/kg Moist, mineral free**
Anthracite	Meta Anthracite	>98	<2	
	Anthracite	92–98	2–8	
	Semi Anthracite	86–92	8–14	
Bituminous	Low Volatile	78–86	14–22	
	Medium Volatile	69–78	22–31	
	High Volatile A	<69	>31	>32.6
	High Volatile B			30.2–32.6
	High Volatile C			26.7–30.2
Sub bituminous	Sub bituminous A			24.4–26.7
	Sub bituminous B			22.1–24.4
	Sub bituminous C			19.3–22.1
Lignite	Lignite A			14.7–19.3
	Lignite B			<14.7

2.2. Cluster analysis

Cluster analysis, also known as clustering, is the task of grouping a set of objects so that objects in the same group (called a cluster) are more similar (in some ways) to those in other groups (clusters). It is a primary task of exploratory data analysis and a common statistical data analysis technique used in a variety of fields such as pattern recognition, image analysis, information retrieval, bioinformatics, data compression, computer graphics, and machine learning. One of the reasons for the proliferation of clustering algorithms is that the concept of a "cluster" cannot be precisely defined. A common denominator exists as a collection of data objects. However, different researchers use different cluster models, and different algorithms can be given for each of these cluster models. The properties of a cluster as discovered by different algorithms vary significantly. Understanding these "cluster models" is critical to comprehend the differences among the various algorithms.

Typical cluster models include the following:

- Connectivity models: for example, hierarchical clustering builds models based on distance connectivity.
- Centroid models: for example, the k-means algorithm represents each cluster by a single mean vector.
- Distribution models: clusters are modeled using statistical distributions, such as multivariate normal distributions used by the expectation-maximization algorithm.
- Density models: for example, DBSCAN and OPTICS define clusters as connected dense regions in the data space.

- Subspace models: in biclustering (also known as co-clustering or two-mode-clustering), clusters are modeled with both cluster members and relevant attributes.
- Group models: some algorithms do not provide a refined model for their results and just provide the grouping information.
- Graph-based models: A clique, defined as a subset of nodes in a graph where every two nodes in the subset are connected by an edge, is a prototypical form of cluster. As in the HCS clustering algorithm, relaxations of the complete connectivity requirement (a fraction of the edges can be missing) are known as quasi-cliques.
- Signed graph models: Every path in a signed graph has a sign from the product of the signs on the edges. Under the assumptions of balance theory, edges may change signs and result in a bifurcated graph. The weaker "cluster ability axiom" (no cycle has exactly one negative edge) yields results with more than two clusters or subgraphs with only positive edges.
- Neural models: the most well-known unsupervised neural network is the self-organizing map and these models can usually be characterized as similar to one or more of the above models, including subspace models when neural networks implement a form of Principal Component Analysis or Independent Component Analysis.

2.2.1 Centroid-based clustering

Each cluster in centroid-based clustering is represented by a central vector that is not necessarily a member of the data set. When the number of clusters is fixed to k, k-means clustering has a formal definition as an optimization problem: find the k cluster centers and assign the objects to the cluster center with the shortest squared distance.
Because the optimization problem is known to be NP-hard, the common approach is to look only for approximate solutions. Lloyd's algorithm Least squares quantization in PCM, also known as the "k-means algorithm," is a well-known approximate method (although another algorithm introduced this name). However, it only finds a local optimum and is frequently repeated with different random initializations. k-means variants frequently include optimizations such as selecting the best of multiple runs, but also restricting the centroids to members of the data set (k-medoids), selecting medians (k-medians clustering), selecting the initial centers less randomly (k-means++), or allowing a fuzzy cluster assignment (fuzzy c-means). K-means has several intriguing theoretical properties. First, it divides the data space into what is known as a Voronoi diagram. Second, it is similar in concept to nearest neighbor classification and, as such, is widely used in machine learning. Third, it can be viewed as a variant of model-based clustering, and Lloyd's algorithm as a variant of the Expectation-maximization algorithm for this model, which is discussed further below.

2.3. K-means clustering

K-means clustering is a popular unsupervised machine learning algorithm used for partitioning a dataset into K distinct clusters. The goal of K-means clustering is to group similar data points together and separate dissimilar data points, based on their feature similarity.

Following steps are used in the K-means clustering algorithm:

1) Initialization: Specify the number of clusters, K, which is to identify in a given dataset. Randomly initialize K cluster centroids.

2) Assignment: Assign each data point to the nearest centroid. This is done by calculating the distance between each data point and each centroid and assigning the data point to the cluster with the closest centroid.

3) Update: After assigning all data points to clusters, update the centroids of each cluster by calculating the mean of all the data points assigned to that cluster. This step moves the centroid closer to the center of its assigned data points.

4) Iteration: Repeat the assignment and update step till convergence is achieved. Convergence occurs when the centroids no longer move significantly or when a specified number of iterations is reached.

5) Result: Once the algorithm converges, K numbers of clusters receives, each represented by its centroid. Then these centroids are used to classify new data points based on their proximity to the centroids.

K-means clustering works best when the clusters are well-separated and have a roughly spherical shape. It is sensitive to the initial random centroid selection and can converge to different results with different initializations. To mitigate this issue, the algorithm is often run multiple times with different initializations, and the best clustering solution is chosen based on a defined evaluation criterion, such as minimizing the within-cluster sum of squares.

K-means clustering has various applications, including customer segmentation, image compression, anomaly detection, and document clustering, among others.

2.4. Auto Model clustering

The Auto Model Clustering feature is an automated approach to clustering analysis. It is designed to simplify the process of finding the optimal clustering solution by automatically testing and evaluating different clustering algorithms and configurations.

Steps involved in the auto model clustering: -

1) Data Preparation: First, the dataset is to be loaded and any necessary data pre-processing, such as handling missing values, scaling, or transforming variables should be performed.

2) Auto Model Clustering Operator: In Rapid Miner, the Auto Model Clustering operator can be found, which is a high-level operator that automates the clustering process. it can be accessed through the operator search bar or by navigating through the operator toolbox.

3) Configuration: When the Auto Model Clustering operator is applied, certain settings need to be specified, such as the number of clusters (K), the evaluation measure to optimize (e.g., silhouette coefficient), and the maximum number of models to consider.

4) Algorithm Selection: Rapid Miner will automatically evaluate different clustering algorithms, such as K-means, DBSCAN, or hierarchical clustering, along with various parameter configurations for each algorithm. It will train and evaluate different models using a cross-validation or holdout approach.

5) Model Evaluation: The Auto Model Clustering feature compares the performance of different clustering models using the specified evaluation measure. It selects the best model based on the chosen criterion, such as maximizing the silhouette coefficient or minimizing the sum of squared distances.

6) Results: Once the Auto Model Clustering process is complete, Rapid Miner provides the best clustering solution based on the evaluation measure. the results can be examined, including cluster assignments, cluster centroids, and any other relevant information.

The advantage of using Auto Model Clustering in Rapid Miner is that it automates the selection and evaluation of clustering algorithms and their configurations. This can save time and effort, especially when you are unsure about the optimal clustering approach for your dataset. However, it is still important to interpret and validate the results obtained from the automated process to ensure they align with your domain knowledge and objectives.

III. MODELING

For modeling of both K-mean and Auto model, the software RapidMiner is used.

3.1 For K-mean clustering

3.1.1 Steps involved to perform clustering

i. Import data in local repository of which the clustering can be performed.
ii. Add an operator named Set role, to set the role for given input data e.g Label, id, weight, regular etc.
iii. Add an operator named K-mean, which is inside Segmentation in Modeling
iv. Connect the example set output of Set role operator to example set input of K-means operator
v. Connect both the cluster set output and cluster model output to result as shown in the figure.
vi. Select the K-mean operator and check the parameter box, select add cluster attribute and then put the value of k (number of cluster).
vii. Run the process.
viii. Analyze the statistical data from results.

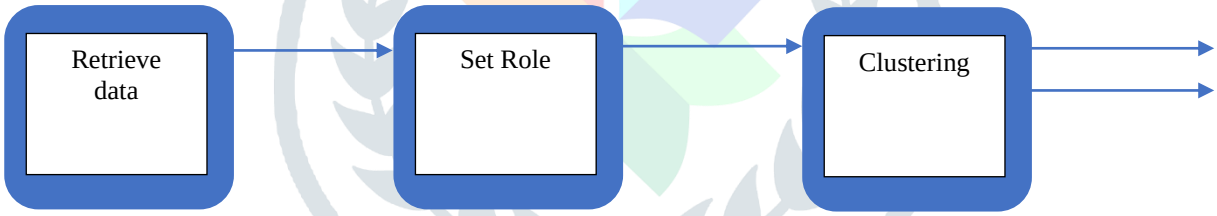

Fig. 1: K-mean clustering process

3.1.2 Collection of necessary data

As we connected both clusters set output and cluster model output, results contain both example set and cluster model. So first check whether the number of cluster obtained are correct or not by selecting cluster model, then collect the table chart of all the necessary data like Fixed carbon, Volatile matter, Ash, Gross calorific value/High heat value etc. for the particular cluster by selecting example set in result.

3.1.3 Davies Bouldin Index

A method used to determine a dataset's clustering quality is the Davies-Bouldin Index (DBI). It calculates the average resemblance between every cluster and the cluster that is closest to it, relative to the average dissimilarity within each cluster. A lower DBI value indicates better clustering quality.

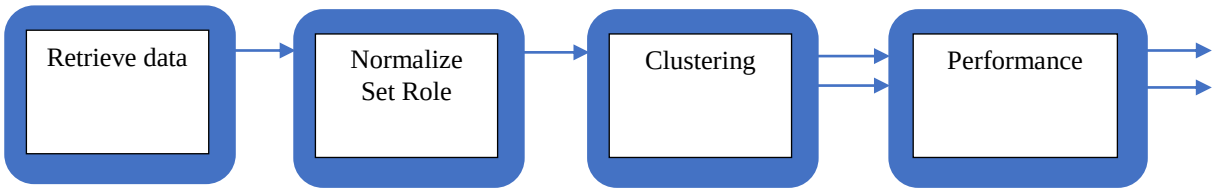

Fig. 2: Db index performance evaluation process

3.2 For Auto-Model

i. Select the Auto Model mode in views box.
ii. In load data, select the file which contain the data of coal sample and press next.
iii. Put the value in number of clusters below the K-Mean Clustering and disable X-Mean option.
iv. Run the process.
v. Analyze the statistical data from results.

IV. RESULT AND DISCUSSION

Scatter plots of the coal dataset are plotted to observe the groups formed among the variables involved in the dataset which are useful for clustering. Fig. 3 to 10 indicate the scatter plots.

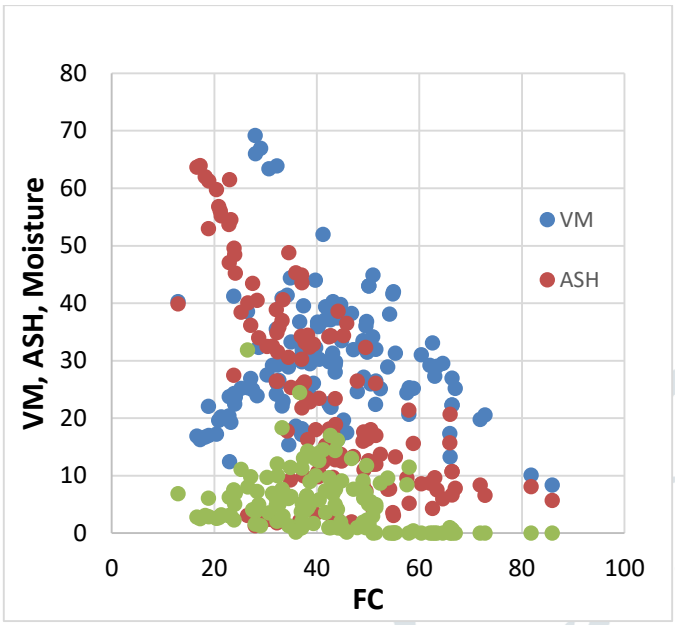

Fig. 3: Scatter plot of FC vs VM, ASH, Moisture

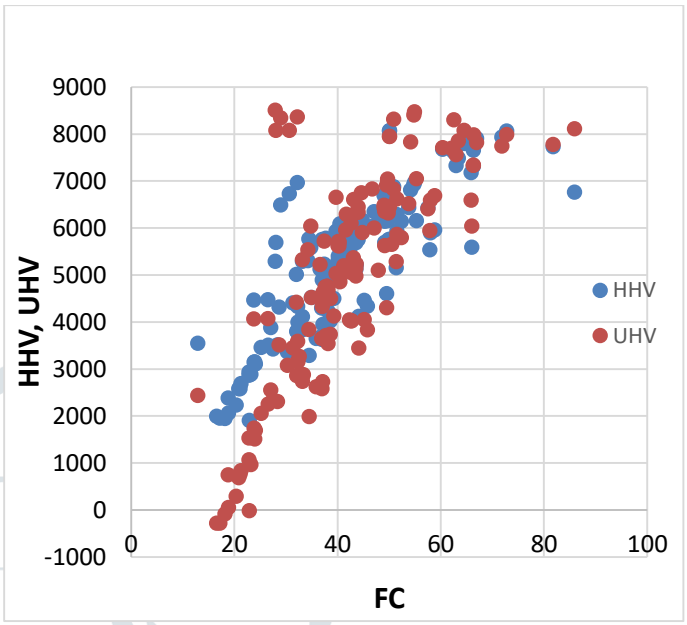

Fig.4: Scatter plot of FC vs HHV, UHV

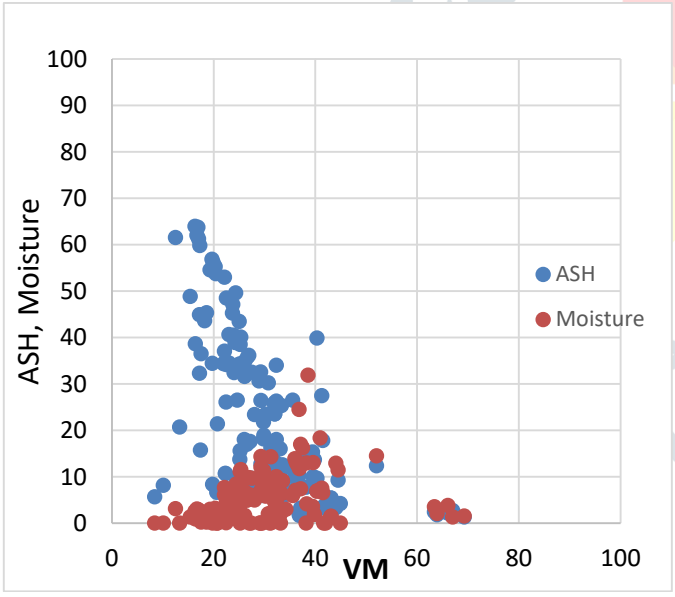

Fig. 5: Scatter plot of VM vs ASH, Moisture

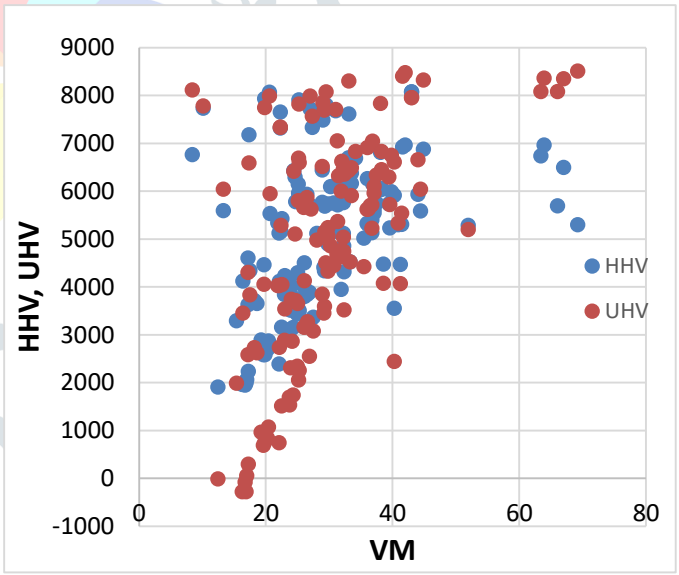

Fig. 6: Scatter plot of VM vs HHV, UHV

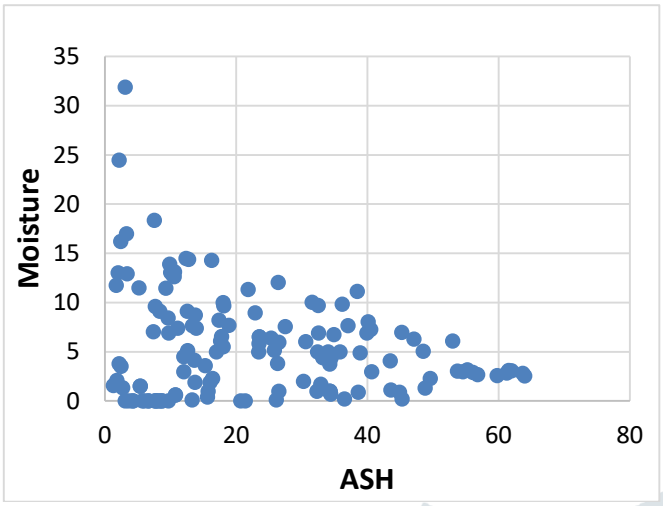

Fig. 7: Scatter plot of ASH vs Moisture

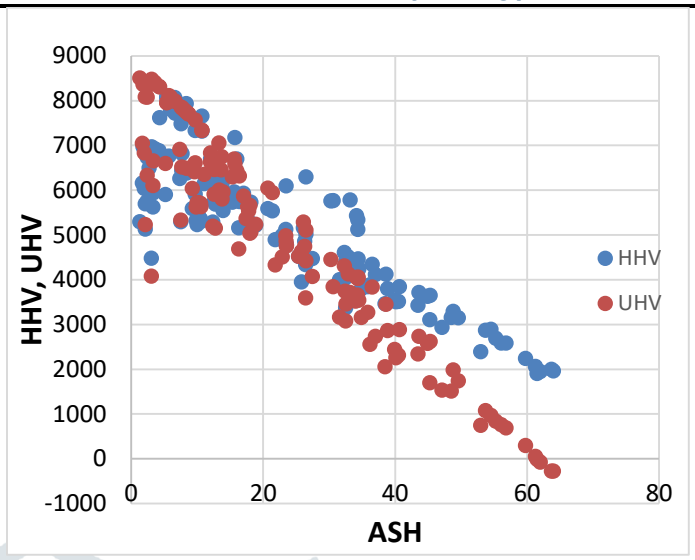

Fig. 8: Scatter plot of VM vs HHV, UHV

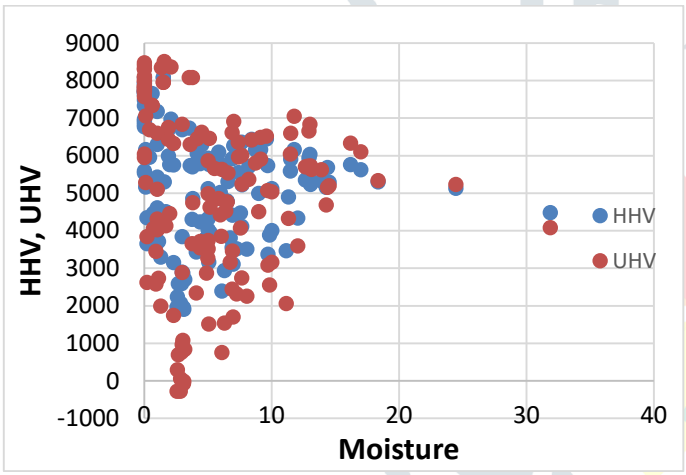

Fig. 9: Scatter plot of Moisture vs HHV, UHV

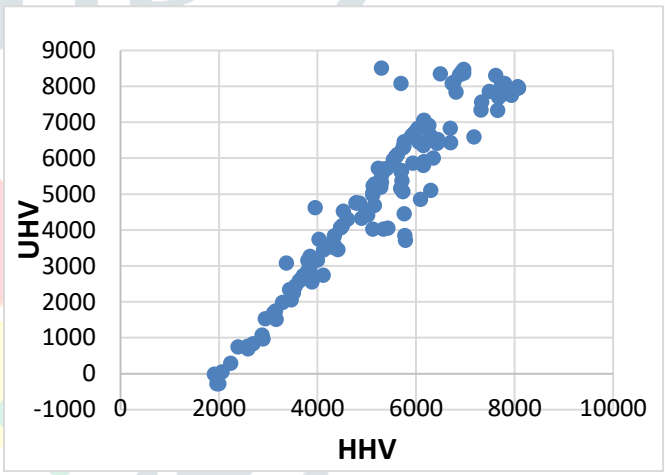

Fig. 10: Scatter plot of HHV vs UHV

The simulation tokes places in two steps. The first one is through K-mean Clustering technique, the second one is Auto Model Clustering. the DB index magnitude is lowest for cluster number (K) equal to seven, thus indicating the presence of seven clusters in the five-dimensional coal dataset. While using K-mean method Davies bouldin index is used to estimate the number of clusters. The lowest value of Db index is calculated by iteration method to determine the number of clusters.

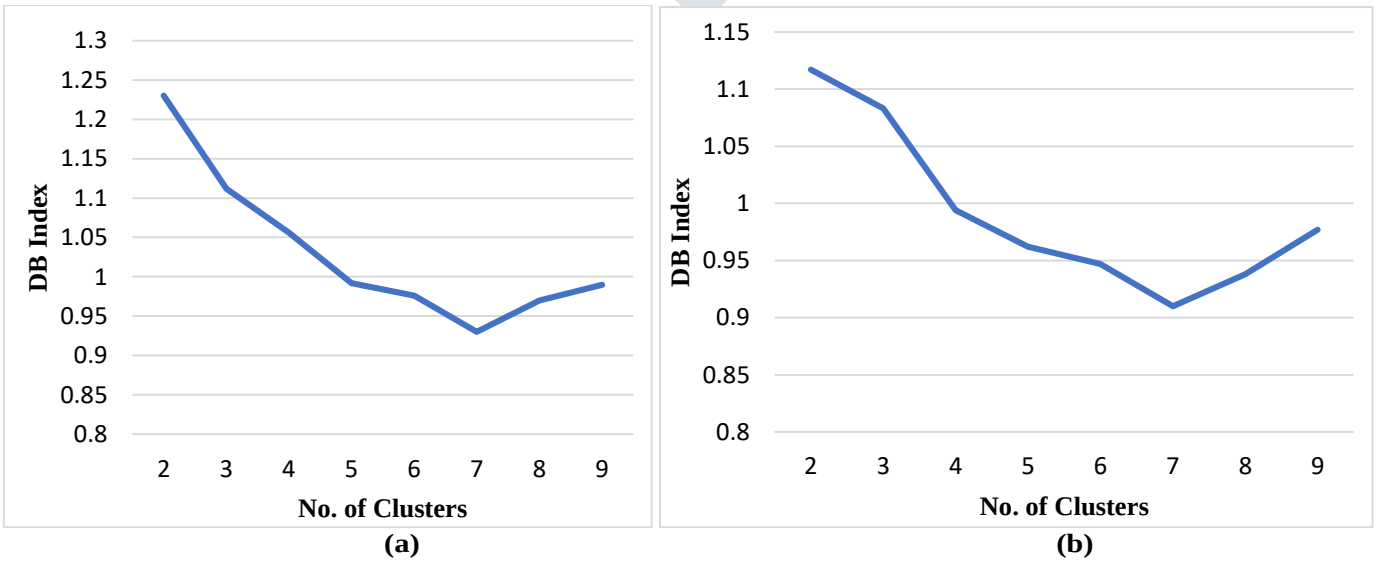

Fig. 11: No. of classes vs Davies Bouldin for (a) K-mean and (b) Auto-Model

The above graphs states that at DB Index have lowest value when it has 7 cluster in both the methods. So, the data is classified into 7 Clusters for both the methods.

Table 3: Class wise data range of proximate analysis with average value

Sr no.	Attribute	Methods	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
1	FC	K-Mean	12.9-37.1	28.4-66	50.1-80.76	23.77-49.5	28.9-85.9	16.58-22.94	27.93-51.4
		Avg	27.25	46.01	64.56	35.23	50.8	19.64	39.04
		Auto model	12.9-39.73	33.41-51.4	57.9-85.9	26.49-49.69	27.93-32.19	16.58-34.5	34.23-58
		Avg	31.86	41.31	66.82	39.71	29.56	22.18	48.61
2	Volatile Matter	K-Mean	15.4-40.3	13.3-66.3	10.11-43.3	16.4-41.22	8.4-66.99	12.4-22.1	21.89-69.21
		Avg	23.15	32.94	27.51	26.34	36.96	17.85	34.4
		Auto model	22.13-41.22	16.4-30.75	8.4-33.13	29.31-51.97	63.39-69.21	12.46-24.98	24.43-44.9
		Avg	29.75	22.15	22.86	38.34	65.91	19.6	33.98
3	Ash	K-Mean	32.5-54.5	1.7-33.19	4.32-10.73	3.7-40.65	1.8-26.5	52.99-63.96	1.3-34.3
		Avg	44.76	14.1	7.64	32.06	8.88	59.33	18.42
		Auto model	18-40.5	23.41-45.3	4.32-21.4	1.7-16.27	1.3-2.69	43.47-63.96	3.1-23.47
		Avg	30.82	34.63	10.17	7.49	2.07	54.68	12.21
4	Moisture	K-Mean	0.2-11.13	0-17	0-1.5	0.2-31.88	0-9.59	2.56-6.09	0.1-24.46
		Avg	4.82	6.93	0.28	6.34	3.34	3.18	8.92
		Auto model	3.8-12.5	0.1-5	0-1	11.47-31.88	1.33-3.78	1.3-6.97	0-11.5
		Avg	7.57	1.91	0.15	15.63	2.46	3.54	5.2
5	GCV	K-Mean	2870.4-3714.1	5535.2-6163.1	7315.2-8075.1	3800.69-4605.1	6260.7-7177.4	1905.1-2693.2	4782.1-5426.9
		Avg	3326.54	5843.1	7737.7	4197.94	6641.7	2235.13	5162.14
		Auto model	3366.5-5767.72	3628.1-6295.6	5535.2-8067	4476.1-6155.8	5295.4-6965.5	1905.1-3429.2	5216.9-8075.1
		Avg	4282.27	4671.11	7274.3	5454.18	6236.43	2621.2	6232.84

Table 4: Grading for Non-coking coal in UHV[17]

Grade	Useful heat value (UHV) (kcal/kg)	Ash (%) + moisture (%) at 60% RH and 40 C	Gross calorific value (GCV) (kcal/kg) range
A	>6200	19.5	>6454
B	5600–6200	19.6–23.8	6049–6454
C	5600–6200	23.9–28.6	5597–6049
D	4200–5600	28.7–34.0	5089–5597
E	3360–4200	34.1–40.0	4324–5089
F	2400–3360	40.1–47.0	3865–4324
G	1300–2400	47.1–55.0	3113–3865

Comparing the table 3 and 4, it found that

- Coal samples belongs to 3rd and 5th Classes are A rank coal due to High GCV.
- Some of the coals samples from 2nd class are B rank. While remaining are C rank.
- Coal samples from 1st, 4th and 7th are belongs to G, C, F and D rank.
- While 6th class coals have lower rank than G.

The class having these coal samples can be determined using the ranges of the five coal properties shown in Table 3. The elements of coal and their magnitudes determine its applicability for a particular industrial application. As a result, coals from classes i, iv, and vi with low GCV and high ash content can be used in cement and brick businesses that require moderate heating. Coals from classes III and V have high GCV and are thus ideal for power production via combustion and gasification routes, while their high ash content is a disadvantage. When blended with biomass, coals from classes v and VII make excellent candidates for co-gasification in a variety of sectors.

V. CONCLUSIONS

This work results in the classification of coals used in thermal power plants using the K-means clustering and the Auto Model K-mean approach, which is based on artificial intelligence. The five characteristics of coal moisture, ash, volatile matter, carbon content, and gross calorific value were used to execute the aforementioned classification. According to the classification outcomes, coals with various geographic origins mainly Turkey, Pakistan and New Zealand can be divided into seven classes. Also it is observes that both the method used for clustering have 70.8% similarity. The classes are also compared with a commonly utilized UHV-based grading system and it was observed that the class analyzed by both K-mean and Auto model exhibit similarity with some grade. Based on the lower Davies Bouldin index, the Auto-model based classifier performed better then k-means classifier. The classification of the power coals as well as the class-wise ranges of the five coal qualities offered in this study can be effectively used to choose coals that are priced appropriately for particular applications. Additionally, the classification methodology shown here can be applied to other fuels like Mineral oil.

REFERENCES

1) Awangga, R.M., Pane, S.F., Tunnisa, K. and Suwardi, I.S., 2018. K means clustering and meanshift analysis for grouping the data of coal term in puslitbang tekmira. TELKOMNIKA (Telecommunication Computing Electronics and Control), 16(3), pp.1351-1357.
2) B. Basil Beamish (1994) Proximate analysis of New Zealand and Australian coals by thermogravimetry, New Zealand Journal of Geology and Geophysics, 37:4, 387-392.
3) Behera, Bhaskar & Sahu, Himanshu. (2017). EXPERIMENTAL EVALUATION OF SUSCEPTIBILITY TO SPONTANEOUS HEATING OF SOME INDIAN COALS. Minetech. 37. 41-50.
4) Coates, Adam; Ng, Andrew Y. (2012). "Learning feature representations with k-means" (PDF). In Montavon, G.; Orr, G. B.; Müller, K.-R. (eds.). Neural Networks: Tricks of the Trade. Springer.
5) Csurka, Gabriella; Dance, Christopher C.; Fan, Lixin; Willamowski, Jutta; Bray, Cédric (2004). Visual categorization with bags of keypoints (PDF). ECCV Workshop on Statistical Learning in Computer Vision.
6) Dai, S., Yan, X., Ward, C.R., Hower, J.C., Zhao, L., Wang, X., Zhao, L., Ren, D. and Finkelman, R.B., 2018. Valuable elements in Chinese coals: A review. International Geology Review, 60(5-6), pp.590-620.
7) Dave Osborne (ed), The Coal Handbook: Towards Cleaner Production: Volume 1: Coal Production Elsevier, 2013 ISBN 085709730X, table 2.5-page 47 Coal Analytical Methods Blackwell Scientific Press, 1984.

8) Estivill-Castro, Vladimir (20 June 2002). "Why so many clustering algorithms – A Position Paper". ACM SIGKDD Explorations Newsletter. 4 (1): 65–75. doi:10.1145/568574.568575. S2CID 7329935.
9) Hower, J.C., Bagherieh, A.H., Dindarloo, S.R., Trimble, A.S. and Chelgani, S.C., 2021. Soft modelling of the Hardgrove grindability index of bituminous coals: An overview. International Journal of Coal Geology, 247, p.103846.
10) Hutagalung, J., Ginantra, N.L.W.S.R., Bhawika, G.W., Parwita, W.G.S., Wanto, A. and Panjaitan, P.D., 2021, February. Covid-19 cases and deaths in southeast Asia clustering using k-means algorithm. In Journal of Physics: Conference Series (Vol. 1783, No. 1, p. 012027). IOP Publishing.
11) Kohonen, Teuvo; Honkela, Timo (2011). "Kohonen network". Scholarpedia. 2 (1): 1568. Bibcode:2007SchpJ...2.1568K. doi:10.4249/scholarpedia.1568.
12) Larry T., 2002, Coal Geology, Wiley Publisher.
13) Lovd, S. (1982). "Least squares quantization in PCM". IEEE Transactions on Information Theory. 28 (2): 129–137. doi:10.1109/TIT.1982.1056489.
14) Mehmet Bilen 2019 IOP Conf. Ser.: Earth Environ. Sci. 362 012092
15) Niel de Clerk, (2011) 96012031, Department of Chemical Engineering University of Pretoria, Flash Dryer Unit Optimization through Advanced Process Control
16) Organo-Refining to Produce Near Zero Ash Coals: Determination of Elemental Concentration in Clean Coals Heena Dhawan, Sreedevi Upadhyayula, and D. K. SharmaEnergy & Fuels 2018 32 (6), 6535-6544 DOI: 10.1021/acs.energyfuels.8b00549.
17) Pandit, Y.P., Badhe, Y.P., Sharma, B.K., Tambe, S.S. and Kulkarni, B.D., 2011. Classification of Indian power coals using K-means clustering and Self Organizing Map neural network. Fuel, 90(1), pp.339-347.
18) Pratama, R. and Widodo, S., 2023. Study of increasing the calorific value of coal using waste cooking oil. In IOP Conference Series: Earth and Environmental Science (Vol. 1134, No. 1, p. 012041). IOP Publishing.
19) Seervi, K., 2015. Prediction of calorific value of Indian coals by artificial neural network (Doctoral dissertation).
20) Ullah SM, Zahid U and Masood T. Ultimate and proximate analysis of coal samples from different regions in pakistan for their future utilization (2019) J of Heterocyclics 1: 39-41.
21) Vesanto, J. and Alhoniemi, E., 2000. Clustering of the self-organizing map. IEEE Transactions on neural networks, 11(3), pp.586-600.
22) Vora, P. and Oza, B., 2013. A survey on k-mean clustering and particle swarm optimization. International Journal of Science and Modern Engineering, 1(3), pp.24-26.
23) Vrieze, O.J. (1995). "Kohonen Network" (PDF). Artificial Neural Networks. Springer. Lecture Notes in Computer Science. Vol. 931. University of Limburg, Maastricht. pp. 83–100. doi:10.1007/BFb0027024. ISBN 978-3-540-59488-8. Retrieved 1 July 2020.
24) Wulandari, Winny & Sasongko, Dwiwahju & Rizkiana, Jenny & Anggayana, Komang & Agus, Haris & Widayat, Aghietyas & Choirun, Az & Zahra, Muhammad & Azhari,. (2019). Non-isothermal kinetics analysis of various Indonesian coal thermal characteristics for underground coal gasification purposes. AIP Conference Proceedings. 2085. 20022.
25) Yoro, Kelvin O., and Michael O. Daramola. "CO_2 emission sources, greenhouse gases, and the global warming effect." In Advances in carbon capture, pp. 3-28. Woodhead Publishing, 2020.