Summer Internship Project Work
On

Training and Development

At

STEEL AUTHORITY OF INDIA LIMITED, Ranchi

By

KAUSHIK VATS

BIRLA INSTITUTE OF TECHNOLOGY, MESRA, RANCHI

Under the guidance of
KUMAR SIDDHANTH

TABLE OF CONTENTS

ACKNOWLEDGMENT

We take this opportunity to express our heartfelt gratitude to the Coal & Coke Division of RDCIS, SAIL, to help us in taking up a project on "A Study on Coke Making from Different Blend Compositions". The experimental work carried out in completing the project will surely be useful throughout our careers as an Engineer.

Our sincere thanks to Shri S. K. Kushwaha, Sr. Mgr. (CCT), Shri P. K. Pankaj, DM (CCT), Dr. P.K. Jha, DM (CCT), Dr. B. K. Sahoo, DM (CCL), Shri Bablu Ghosh, AM (CCL) and all the Senior Technicians Mr. A.K. Jaiswal, Mr. A. Kindo & Mr. N. K. Mishra, RDCIS, SAIL, Ranchi. Without their help, encouragement and continuous guidance all throughout, it would not have been possible to complete this project. We would also like to extend our gratitude to Shri S. K. Das, GM (C, E & E), Shri K. K. Manjhi, DGM (CCT), Shri B. Chakraborty, AGM (CCL) and Mrs. Neeta Chakravarty, AGM (CCL) for their encouragement and their consistent support.

We extend our sincere thanks to GM (P&A), Shri Kumar Sidhanth, Jr. Mgr. (HRD) and other members of HRD for their support and guidance throughout the training programme.

Thanking You,

Saurav Kumar Sarkar
Syed Shabbir Ahmad
Shivani Dasgupta
Ruchi Kumari
Anjali Ranjan
Tarunam

Anamica Kirti Barla
Prabha Nisha Toppo
Sameer Kerketta
Divya
Shikha Tirkey
Rohit Mukherjee

Executive Summary

- **Project Title:**

Data-Driven Coke Quality Prediction using Regression Modelling and Machine Learning Techniques

- **Internship Organization:**

Steel Authority of India Limited (SAIL), R&D Centre, Ranchi

- **Introduction & Objective**

Coke quality is a key factor in blast furnace performance in the steelmaking process. Important metrics that affect plant efficiency are M40 (strength), M10 (abrasion), CSR (Coke Strength after Reaction), and CRI (Coke Reactivity Index). This study investigates how these performance measures are impacted by the characteristics of the raw materials, particularly their moisture content and volatile matter.

The main objective was to provide predictive insights that may help with production planning and quality control by utilizing data analytics and regression-based machine learning models. Applying machine learning and statistical reasoning methods to industrial datasets was made possible by this study.

- **Work Completed & Methodologies Used**

✅ Regression Analysis (Excel + Python): Applied Simple and Multivariate Linear Regression models using Excel and Python (sk learn). Split data into training and testing sets to evaluate prediction accuracy.

✅ Statistical Learning Concepts: Covered hypothesis testing, t-scores, p-values, and SSE computations manually and through Python.

✅ Model Assumption Validations: Created and interpreted QQ plots, linearity plots, and residual vs fitted value graphs.

✅ Data Visualization & Handling: Used Excel and Python to clean, explore, and visualize data with scatter plots, heatmaps, etc.

✅ Tools & Platforms: Excel, Python (Pandas, NumPy, Matplotlib, Sklearn), Google Colab, and SAIL Ranchi coke dataset.

- **Outcomes & Learning**

1. Created numerous regression models to accurately forecast the quality of coke.

2. Recognized and represented how moisture and volatile matter affect M40 and CRI.

3.Acquiredpracticalexperiencewithactualindustrydata.

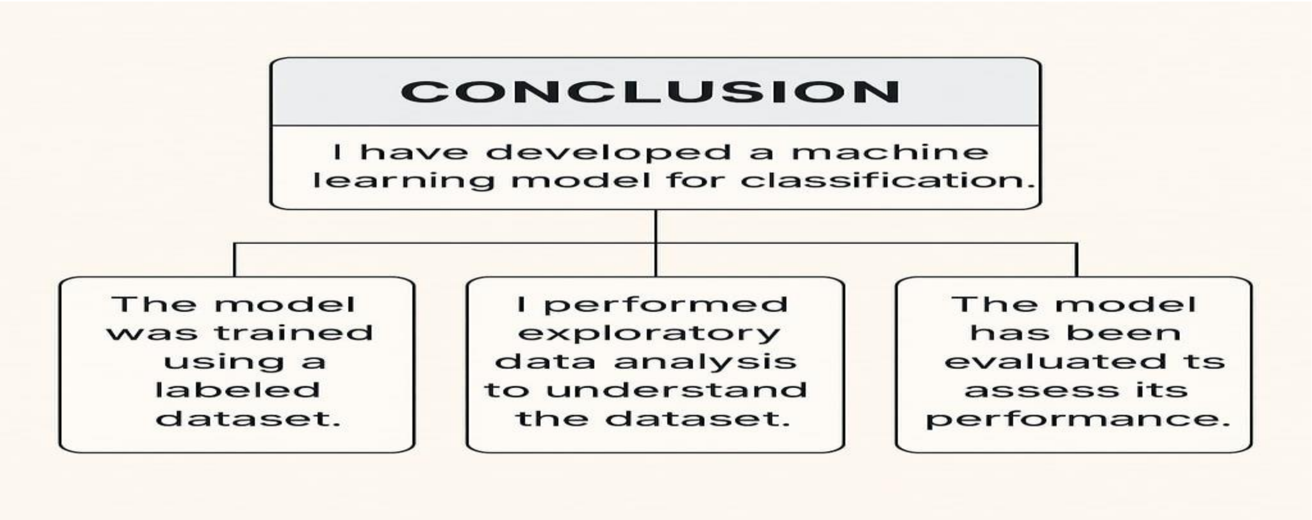
4. Acquired knowledge of how to assess model performance through statistical and visual techniques.

- **Additional Learning & Future Scope**

To enhance the project for ML/Data Science roles, I'm actively exploring:
- Polynomial and Ridge/Lasso Regression
- Feature scaling techniques using Standard Scaler
- Advanced ML models like Decision Trees and Random Forests
- Basic dashboarding with Stream lit
- Hosting project code and visualizations on GitHub

- **In conclusion**

This internship provided a hands-on introduction to data analytics and machine learning. Significant insights were gleaned from SAIL's operational data using statistical modeling, Python scripting, and data visualization. I'm also branching out into feature engineering, deployment, and supervised machine learning methods. This work serves as a solid basis for the expansion of data-driven industrial innovation.



COMPANY PROFILE

3.1 INTRODUCTION

Steel Authority of India Limited (SAIL) is the largest steel-making company in India and one of the seven Maharatna's of the country's Central Public Sector Enterprises. The Government of India owns about 75% of SAIL's equity and retains voting control of the Company. However, SAIL, by virtue of its 'Maharatna' status, enjoys significant operational and financial autonomy.

Steel Authority of India Limited (SAIL) is a company registered under the Indian Companies Act, 1956 and is an enterprise of the Government of India. It has five integrated steel plants at Bhilai (Chhattisgarh), Rourkela (Orissa), Durgapur (West Bengal), Bokaro (Jharkhand) and Burnpur (West Bengal). SAIL has three special and alloy steel plants viz. Alloy Steels Plant at Durgapur (West Bengal), Salem Steel Plant at Salem (Tamil Nadu) and Visvesvaraya Iron & Steel Plant at Bhadravati (Karnataka). In addition, a Ferro Alloy producing plant Maharashtra Elektrosmelt Ltd. at Chandrapur, is a subsidiary of SAIL. SAIL has Research & Development Centre for Iron & Steel (RDCIS), Centre for Engineering & Technology (CET), SAIL Safety Organization (SSO) and Management Training Institute (MTI) all located at Ranchi; Central Coal Supply Organization (CCSO) at Dhanbad; Raw Materials Division (RMD), Environment Management Division (EMD) and Growth Division (GD) at Kolkata. The Central Marketing Organization (CMO), with its headquarters at Kolkata, coordinates the country-wide marketing and distribution network.

The Committee of Public Undertaking of the Fifth Lok Sabha was the first Parliamentary Committee to undertake a significant review of the question of setting up a Holding Company for steel. It was first considered in the Department of Steel in 1971 with the following two objectives:

- Rapid growth of the industrial sector, of the economy, of the state as a leading agent of the growth process; and

- Ability of the Government to divert investment into areas which are strategic from the point of view of future development.

In this context, it was recognized that the Public Sector had to be made more efficient in order that it might be able to contribute far more than it had to the common pool of investible surplus in the economy. Further, such a holding company could perform a number of other important functions like coordination and control of constituent units, planning long term programmes, introduction of necessary technological changes, setting up of an R & D organization and training of managerial personnel for the Public Sector as a whole. Based on the above considerations, the proposal to set up a holding company for Steel and associated input industries was approved by the Government in January 1972. Accordingly, the formation of Steel Authority of India Limited was approved by the Government in December, 1972. The company was incorporated on January 24, 1973 with an authorized capital of Rs.2, 000 crores. In 1978 SAIL was restructured as an operating company.

3.2 PRESENT STATUS OF SAIL

- Steel Authority of India Limited (SAIL) through its five integrated steel plants at Bhilai, Bokaro, Burnpur, Durgapur and Rourkela accounts for major steel production capacity of India.
- Three special steel plants at Bhadravati, Durgapur and Salem produce a wide range of special steels, special alloy steels and stainless steel.
- Today, SAIL is one of the largest corporate entities. Its innate strength lies in its technologists and professionals and a trained manpower of over 1.34 Lakh including subsidiary. It had a sales turnover of over Rs. 45,555 crores during 2007-08.
- It is a fully integrated iron and steel maker, producing both basic and special steels for domestic construction, engineering, power, railway, automotive and defense industries and for sale in export markets.

- Ranked amongst the top ten public sector companies in India in terms of turnover.

- SAIL manufactures and sells a broad range of steel products, including hot and cold rolled sheets and coils, galvanized sheets, electrical sheets, structural, railway products, plates, bars and rods, stainless steel and other alloy steels.
- The company has the distinction of being India‘s largest producer of iron ore and of having the country‘s second largest mines network. This gives SAIL a competitive edge in terms of captive availability of iron ore, limestone, and dolomite which are inputs for steel making.
- SAIL's wide range of long and flat steel products is much in demand in the domestic as well as the international market. This vital responsibility is carried out by SAIL's own Central Marketing Organization (CMO) and the International Trade Division. CMO encompasses a wide network of 34 branch offices and 54 stockyards located in major cities and towns throughout India.
- With technical and managerial expertise and know-how in steel making gained over four decades, SAIL's Consultancy Division (SAILCON) at New Delhi offers services and consultancy to clients world-wide.
- SAIL has a well-equipped Research and Development Centre for Iron and Steel (RDCIS) at Ranchi which helps to produce quality steel and develop new technologies for the steel industry. Besides, SAIL has its own in-house Centre for Engineering and Technology (CET), Management Training Institute (MTI) and Safety Organization at Ranchi. Our captive mines are under the control of the Raw Materials Division in Kolkata. The Environment Management Division and Growth Division of SAIL operate from their headquarters in Kolkata. Almost all our plants and major units are ISO Certified.

3.3 SWOT ANALYSIS OF THE COMPANY

The strengths, weaknesses, opportunities and threats for the Indian steel industry have been tabulated below. The national steel policy lays down the broad roadmap to deal with all of them.

Strengths

1. Availability of iron ore and coal

2. Low labor wage rates

3.Abundance of quality manpower

3. Mature production base

Weaknesses

1. Unscientific mining

2. Low productivity

3. Coking coal import dependence

4. Low R&D investments

5. High cost of debt

6. Inadequate infrastructure

Opportunities

1. Unexplored rural market

2. Growing domestic demand

3. Exports

4. Consolidati

Threats

1. Global steel price fluctuations.
2. Rising competition from private and foreign firms.
3. Strict environmental regulations.

RDCIS

R&D Centre for Iron and Steel

- The Research & Development Centre for Iron & Steel (RDCIS) at Ranchi is the corporate R&D unit of SAIL. Set up in 1972, the Centre has ISO: 9001 certification to its credit. It undertakes R&D projects in diverse realms of Iron & Steel Technology under the categories of Plant Performance Improvement (PPI), Product Development (PD), Scientific Investigation and Development (SID), Basic Research (BR) and Technical Services (TS).
- RDCIS has around 236 dedicated and competent scientists and engineers and its laboratory is equipped with around 300 sophisticated diagnostic research equipment and 5 pilot plant facilities.
- RDCIS provides customers with prompt, innovative and cost-effective R&D solutions; develop and commercialize improved processes and products; continually enhance the capability of its human resources to emerge as a center of excellence. The major efforts are directed towards cost reduction, quality improvement and value-addition to products of SAIL plants and providing application engineering support to SAIL's products at customers' end. RDCIS, along with steel plants, takes initiatives to develop special steel products utilizing the modernized production facilities at steel plants.
- RDCIS also offers technological services to various organizations in the form of Know-how transfer of technologies developed by RDCIS; Consultancy services; specialized testing services; Contract research; Technology Awareness Programmes.

Technology Marketing

RDCIS offers technological services in the form of

- Know-how transfer of technologies developed by RDCIS
- Consultancy services / Contract research
- Specialized testing services
- Training

RDCIS Clientele includes:

- Bhabha Atomic Research Centre, Mumbai
- Vikram Sarabhai Space Centre, Thiruvananthapuram
- Nuclear Fuel Complex, Hyderabad
- RDSO,Lucknow
- Southern Railway, Chennai
- CIL, Kolkata
- Indian Oil Corpn.,R&D, Faridabad
- Balmer Lawrie & Co. Ltd., Kolkata
- Hindustan Petroleum Corp., Mumbai
- MECON Ltd., Ranchi
- BHEL-WRI, Tiruchirappalli
- RINL, Visakhapatnam
- Tata Steel Ltd., Jamshedpur
- Uttam Galva Metallics Ltd., Wardha
- Jindal Steel & Power Ltd., Raigarh
- Ispat Industries Ltd., Dolvi
- NTPC-Alstom, New Delhi

Steel Plant	Location	Products
Bhilai Steel Plant	Chattisgarh	Rails (13/26m), Long Rails, (65-260m), Blooms, Billets, Slabs, Channels, Joists, Angles, TMT Rebars, Wire Rods, Crane Rails, Plates, Pig iron & Coal Chemicals
Durgapur Steel Plant	West Bengal	Blooms, Billets, Joists, Narrow Slabs, Channels, Angles, TMT Rebars, Wheels & Axles, Pig iron & Coal Chemicals
Rourkela Steel Plant	Odisha	Plate Mill Plates, HR Plates, HR Coils, Slabs, CR Sheet/ Coil, Galvanised Sheets (plain & Corrugated), ERW Pipes, Spiral Weld pipes, CRNO, Pig iron & Coal Chemicals
Bokaro Steel Plant	Jharkhand	Hr Coils, Slabs, HR Sheets. Plates, CR Coils. Sheets, GP Sheets. coils, GC Sheets, Galvanealed Steel, HRPO, Pig iron & Coal Chemicals
IISCO Steel Plant	West Bengal	Wire rods, Bars & Rebars, Joists, Channels, Angles, Blooms, Billets, Universal & Special section (Z-bar, MS Arch), Pig iron & Coal Chemicals

Special Steel Plants

Steel Plant	Location	Products
Salem Steel Plant	Tamil Nadu	Cold Rolled Stainless Steel, Hot Rolled Carbon & Stainless Steel Products, Micro-Alloyed Carbon Steel
Alloy Steel Plant	West Bengal	Alloy Steel Squares & Rounds, Wear Resistant Plates, Forgings, Carne Wheels, Forged Rolls/ Plaets, Special Quality Slabs & Stainless Setel Slabs (low Ni, 300 & 400 series)
Visveswaraya Iron & Steel Plant	Karnataka	High Quality Rolled & Forged Alloy & Special Steel Products

Ferro Alloy Plant

Steel Plant	Location	Products
Chandrapur Ferro Alloy Plant	Maharashtra	High/ Medium/ Low carbon Ferro-Manganese, Silico-Manganese

Table 1

- **1. Literature Review Section**

For many centuries, coal was burned in small stoves to produce heat in homes and factories. Today, most important use of coal is still use as a fuel. The largest single consumer of coal as a fuel is electrical power industry. For a period of more than 40 yrs. beginning in 1940,

The amount of coal used in U.S. for this purpose doubled in every decade.

Another use of coal is in the manufacture of coke. Coke is nearly pure carbon produce when soft coal is heated in absence of air. In most cases,1 ton of coal will produce 0.7 ton of in this process. Coke is of value in industry because it has a heat value higher than any form of natural coal. It is widely used in steel making and in certain chemical processes.

WHAT IS COAL

Coal is a combustible sedimentary rock composed monthly of carbon and hydrocarbons. It is the most abundant fossil fuel produced in the U.S. but it is a nonrenewable resource. The energy in coke comes from the energy stored by plants that lived 100 of millions of years ago. The heat and pressure from top layers turned the plant remains into coal.

IS ALL COAL THE SAME

Coal can vary in composition even within the same deposit. There are different types of coal, each with differences in energy output as a result of increased pressurization, heat and time.

LIGNITE- Lignite is brownish-black in color and has a carbon content of around 25-35%, a high inherent moisture content sometimes as high as 66%, and an ash content ranging from 6% to 19%.It is used for generating electricity.

Fig: LIGNITE

SUB-BITUMINOUS COAL-This is a dull black coal with a higher heating value than lignite,and is used principally for electricity and space heating.

BITUMINOUS COAL-Bituminous coal is an organic sedimentary rock formed by diagenetic and sub metamorphic compression of peat bog material. Its primary constituents are macerals: vitrinite, and exinite. The carbon content of bituminous coal is around 60-80%; the rest is composed of water, air, hydrogen, and sulphur, which have not been driven off from the macerals. Bank density is approximately 1346 kg/m³ (84 lb/ft³). Bulk density typically runs to 833 kg/m³ (52 lb/ft³). The heat content of bituminous coal ranges from 24 to 35 MJ/kg (21 million to 30 million BTU per short ton) on a moist, mineral-matter-free basis. Formation is usually the result of high pressure being exerted on lignite.

Fig: BITUMINOUS COAL

ANTHRACITE-Anthracite is the most metamorphosed type of coal (but still represents low-grade metamorphism), in which the carbon content is between 92.1% and 98%. It has the highest carbon content, the fewest impurities, and the highest calorific content of all types of coals, which also include bituminous coal and lignite. China accounts for the majority of global production; other producers are Russia, Ukraine, North Korea, Vietnam, the UK, Australia and the US. Total production in 2010 was 670 million tons.

Fig: ANTHRACITE

WHAT IS COAL USED FOR

Coal has many important uses worldwide. The most significant uses are in electricity generation, steel production, cement manufacturing and as liquid fuel. Other important users of coal include alumina refineries, paper manufacturers, and the chemical and pharmaceutical industries. Several chemical products can be produced from the by-products of coal. Refined coal tar is used in the manufacture of chemicals, such as creosote oil, naphthalene, phenol, and benzene.

EARLY USE

The first major boon for coal use occurred in 1830 when the Tom Thumb, the first commercially practical American-built locomotive, was manufactured. The Tom Thumb burned coal, and in rapid fashion, virtually every American locomotive that burned wood was converted to use coal. America's coal industry had begun taking shape. Anthracite (or "hard" coal), clean and smokeless, became the preferred fuel in cities, replacing wood by about 1850. Bituminous (or "soft coal") mining came later. In the mid-century Pittsburgh was the principal market. After 1850 soft coal, which is cheaper but dirtier, came into demand for railway locomotives and stationary steam engines, and was used to make coke for steel after 1870.Outcrop coal was used in Britain during the Bronze Age (2000-3000 years BC), where it has been detected as forming part of the composition of funeral pyres. It was also commonly used in the early period of the Roman occupation: Evidence of trade in coal (dated to about AD 200) has been found at the inland port of Heronbridge, near Chester, and in the Fenlands of East Anglia, where coal from the Midlands was transported via the Car Dyke for use in drying grain. Coal cinders have been found in the hearths of villas and military forts, particularly in Northumberland, wonder of a permanent brazier of coal on the altar of Minerva at Aquae Sulis (modern day Bath) although in fact easily-accessible surface coal from what became the Somerset coalfield was in common use in quite lowly dwellings locally."

COKING AND USE OF COKE

Coke is a solid carbonaceous residue derived from low-ash, low-sulphur bituminous coal from which the volatile constituents are driven off by baking in an oven without oxygen at temp. as high as 1000°C so that the fixed and as a reducing agent in smelting iron ore on a blast furnace .The product is too rich in dissolved carbon and must be treated further to make steel. The coke must be strong enough to resist the weight of over burden in blast furnace which his why coking coal is so important in making steel by the conventional route. However, the alternative route to be direct reduced iron, where any carbonaceous fuel can be used to make sponge or pelletized iron. Coke from coal is grey, hard, and porous and has a heating value of 24.8 million Btu/ton (29.6 MJ/kg). Some coke-making processes produce valuable by-products that include coal tar, ammonia, light oils and "coal gas".

Petroleum coke is the solid residue obtained in oil refining, which resembles coke but contains too many impurities to be useful in metallurgical applications.

According to Wilson & Wells, "Carbonization of coal is destructive heating of the coal in the absence of air with production of a solid, porous, carbonaceous residue, or coke, which remains in the retort, and the evolution of a number of volatile products which escape. The two principle classes of carbonization processes are high temperature coking and low temperature carbonisation. In high temperature carbonisation temperature of coke is above 900^{0}C, whereas in low temperature carbonisation temperature of coke is not more than 700^{0}C.

CHEMICAL ANALYSIS OF COKE

Proximate analysis of the coke samples gives an insight into the contents of moisture (Total & Analytical) volatile matter, and as. Fixed carbon is obtained by subtracting the sum of the percentage of, ultimate analysis made on metallurgical coke with any regularity.

PHYSICAL PROPERTIES OF COKE

Appearance: The coke oven operator constantly observes the physical appearance of the coke he produces, noting such features as the size of the pieces, the fractures, colour, ring, and cellular structure. Representative samples are frequently preserved over

Size: The size analysis is determined by a sieve test. The most common expression for the size of coke destined for blast furnace use is the percentage of total coke retained on a 2 inch screen. The size determination is especially important because it forms an integral part of most of the other tests.

Strength: The strength of coke is a measure of its ability to retain its size during handling. Two tests are used for evaluating this quality; the shatter test and the tumbler test. Shatter test: The shatter test measures the resistance of the coke to impact and is indicative of the degree of fracturing of the coke pieces.

Tumbler test: The tumbler test is designed to measure both the ease of breakage and the abrasion of the coke.

Bulk Density: The bulk density is weight per cubic foot of the coke.

Apparent specific gravity: The apparent specific gravity is the gravity of the irregular shaped pieces of coke with their finely pored structures.

True specific gravity: The true specific gravity is that of the solid coke substance itself and is greater than the apparent specific gravity.

Porosity & Void Volume: The porosity of a coke is the fraction which is represented by the pores in the pieces, and the void volume is the volume of voids in a mass of coke pieces, both of these characteristics are usually expressed as fractions.

Reactivity: Reactivity is a term which denotes the rate at which a coke reacts with one of the oxidizing gases, air, oxygen, steam or carbon dioxide. The reactivity is dependent on a number of factors, the amount of the coke surface exposed to the gas, the character of the surface, and the conditions of exposure. The latter include the kind of gas, its velocity and concentration, and the temperature.

EFFECT OF COKING CONDITION ON COKE

The yield and properties of a coke depend primarily on the coals from which it was produced and on the carbonising conditions. A large amount of data has been accumulated on the influence of these factors on coke quality; it is still insufficient to permit any reliable predictions as to results which can be expected from a change in conditions. The data are of value however, in selecting conditions for trial of a particular coal or coals. Types of coal, pulverization of coal, addition of inert to coal mix have varying effect on coking.

QUALITY OF COKE REQUIRED IN BLAST FURNACE

High temperature coke owes its dominating position as a blast furnace fuel to the unique combination of strength, porosity, density and other properties which it possesses. It seems probable; however, that uniformly in size, structure, and other physical and chemical characteristics is the most desirable characteristics of coke.

Another essential requirement is strength, in order to stand up under furnace conditions. It must be able to bear up under the great weight of the columns of materials above it at high temperatures which prevail in the furnace. The coke should be free from dust and fines. The fines tend to plug the voids in the blast furnace charge and thus restrict the flow of air and gas. A good blast furnace coke contains not over 2 percent volatile matter and 85-90 percent fixed carbon. The sulfur in the blast furnace coke has ranged from 0.6 to 1.5 percent but it is as low as possible, because the coke is the chief source of sulfur in the pig iron. In general, reduction in the ash content of the coke increases the rate of iron production, and several steel companies have shown 3 to 6 percent increases for a 1 percent reduction in ash. However, a coke with uniform ash content near 15 percent is said to produce more pig iron than coke varying ,for instance between 10 and 15 percent. Coke should not be wet; the moisture should be preferably below 5 percent.

According to Raymond E. Zimmerman. "Coal is heterogeneous, carbonaceous, fossilized materials of almost infinite variations in consist - from low grade lignite or brown coal to graphite-like meta-anthracite. Among its many properties is the ability of some types of coal to for coke when heated in ovens or retorts. During the application of varying degrees of temperature in an atmosphere which will not support combustion, particles of some types of coal will first coalesce or melt, giving off volatile gases and vapor‟s; will then become fluid; and finally will harden into a porous or cellular substance called coke. Since it is high in fixed carbon, coke is valuable as a fuel and as a reducing agent in many metallurgical and chemical processes. Coking coals, by definition, are coals that when heated to sufficiently high temperatures pass through a transient plastic stage in which they successively soften, swell and re-solidify into a coherent cellular coke."

FACTORS AFFECTING COKE QUALITY

A good quality coke is generally made from carbonisation of good coking coals. Coking coals are defined as those coals that on carbonisation pass through softening, swelling and re-solidification to coke. One important consideration in selecting a coal blend is that it should not exert a high coke oven wall pressure and should contract sufficiently to allow the coke to be pushed from the oven. The properties of coke and coke oven pushing performance are influenced by following coal quality and battery operating variables: rank of coal, petrographic, chemical and rheological characteristics of coal, particle size, moisture content, bulk density, weathering of coal, coking temperature and coking rate, soaking time quenching practice and coke handling. Coke quality variability is low if all these factors are controlled. Coke producers use widely differing coals and employ many procedures to enhance the quality of the coke and to enhance the coke oven productivity and battery life.

High-quality metallurgical coke has been produced from good coking coals. The price rise and reduced availability of prime coking coals on a world wide scale has caused intensive studies of the behavior of coals blends in the so-called co-carbonization process where lower rank coals with some additives were used instead of prime coking coals. Coke is a key substance for the efficient and stable operation of the blast furnace. Hence its cost and quality have been studied for a long time. In a conventional blast furnace operation coke acts as a source of energy and as reducing agent ventilation in the furnace. Coke should, therefore, have sufficient strength to withstand the burden of iron ore sinter and other materials at high temperature apart from its physico-chemical characteristics such as low ash, sulphur and phosphorus contents, and low reactivity towards carbon dioxide etc.

Coke making is a complex process to convert coking coal, through a series of operation into metallurgical coke. The process starts from unloading of the coal at the wagon tripplers and ends at sizing and transportation of coke to blast furnace. The salient aspects of BF coke making are discussed in the following paragraphs.

Types & Sources of Coking Coal:

All coals are not coking coals, i.e. all types of coal can‟t be used for coke making. Coking coals are classified as:

- Prime Coking Coal (PCC)
- Medium Coking Coal (MCC)

above coking coal the following types of coal are also used for coke making in all NINL plants. At present the above Coal not being used at NINL, CO Plant.

- Imported Coking Coal (ICC) – Hard
- Soft Coking Coal (SCC)

Coal is extracted from coal mines & processed in the coal washeries to lower down the ash content to make it fit for coke making.

The different sources of coal are named after the respective washeries and are as follows:

PCC - Bhojudih
- Sudamdih
- Munidih
- Patherdih
- Dugda
- Mahuda

MCC - Kathra
- Swang
- Rajrappa
- Kedla
- Nandan

ICC (Hard) – Australia, New Zealand
USA - Being used at NINL
SCC – Australia – Being used at NINL.

PROPERTIES OF COKING COAL

Inherent Moisture**:** This gives a very good idea about the rank of coal with advancement of rank the inherent moisture comes down.

Volatile Matter (VM): This is the volatile matters present in the coal which goes out as gas during carbonisation.

Percentage of Ash: Lower the ash percentage better is the coal. Indian coal normally contains a high percentage of ash. This is reduced to some extent by suitable

COAL	IM	VM	ASH	FSI	LTGK
PCC	2.5-3.0	21-23	19.0	3.5	E
MCC	2.5-3.0	23-25	18.5	3.0	E
SOFT	2.5-3.0	25-30	8-10	5.0	G
HARD	2.5-3.0	18-20	8-10	5.0	G
NZ	2.5-3.0	25-30	3-4	5.0	G

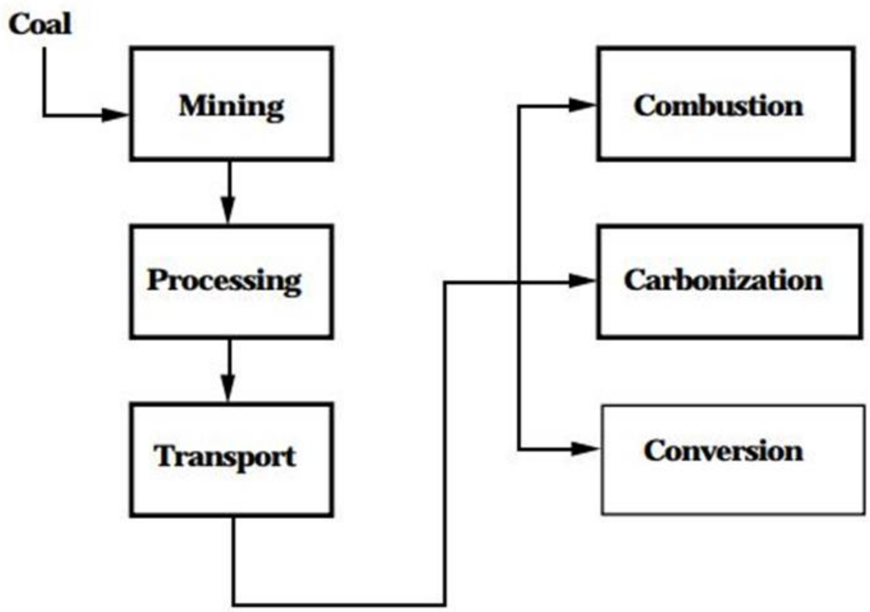
TABLE 1: Properties of various coking coals

LTGK Coke Type : This is another test for agglomerating behaviour of coal.However this is done at a slower rate of heating

COAL HANDLING PLANT

Coke is one of the most important raw materials used to extract iron from the iron ore. The success of Blast Furnace operation depends upon the consistent quality of coke, which is used in Blast Furnace. The quality of coke depends upon the Pre-carbonisation technique, carbonization & Post-carbonisation techniques used in Coke ovens. Pre-carbonisation technique is controlled by Coal handling Plant.

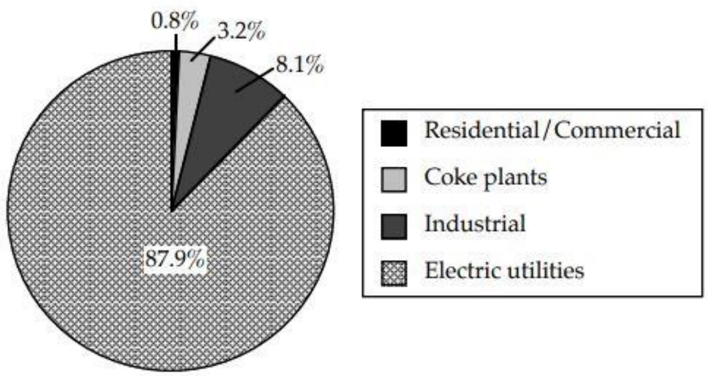
- **THEORY AND RSULT SECTION**



There are three major pathways for coal utilization. These are illustrated in Figure above: -

The principal process by which coal is used is combustion;Combustion involves burning the coal in air to liberate thermal energy (heat). The heat is used as such for comfort or to carry out many industrial processes that require high temperatures; it is also used to generate steam for use in electric power plants. Carbonization is the heating of coal to high temperatures in the absence of air; it is used in manufacturing coke for the metallurgical industry.

Distribution of coal utilization pathways is as follows:-



Coal Carbonization(Coking) and coke

The second largest use of coal is the manufacture of coke for the metallurgical industry. Iron and its alloys (particularly the various kinds of steel) are perhaps the most versatile metallic construction materials known to humankind. Unfortunately, iron occurs in nature not as metallic iron but rather in ores containing iron compounds such as oxides or carbonates. The production of metallic iron requires a chemical reaction to remove the oxygen from the iron compounds and liberate the iron as the metal. The conversion of a metal oxide to metal is an example of a general type of chemical reaction known as reduction (as opposed to the inverse process called oxidation). Substances that we add to the reaction to cause the reduction to take place are termed reducing agents. Among the most powerful types of reducing agents, and among the least expensive, are various forms of carbon or substances rich in carbon.

Coals that pass through a plastic stage on heating are called **caking coals**. Some caking coals re-solidify on heating to form a hard, very strong, carbon-rich porous mass suitable for use as a reducing agent in the metallurgical industry. This material is called **coke**, and the special class of caking coals that yield a satisfactory coke are known as **coking coals**.

To be useful in the metallurgical industry, the coke needs to meet five criteria:-

- To be a useful reducing agent, the coke must have very high carbon content.
- To keep the iron reasonably pure, the coke must have low contents of sulfur and ash.
- To provide ample heat, the coke must have a high content of fixed carbon and have a high calorific value.
- To let air pass through the fuel bed, but yet keep the fuel bed from being collapsed by the weight of iron ore, the coke must be quite porous and very strong.
- Finally, to help keep the cost of the iron low, the coke must be cheap.

PROCESS

Coal carbonization involves heating of coal in the absence of air in coke oven plant. Coke making process is multistep complex process and variety of solid liquids and gaseous products are produced which contain many valuable products. Various products from coal carbonization in addition to coke are coke oven gases. coal tar, light oil, and aqueous solution of ammonia and ammonia salt. Coke oven gases are about 310-340 cum per tone of dry coal which contains gaseous products, coal tar vapours, light oil and water.

COKE OVEN PLANT

Due to the development of iron and steel industry coke oven plant has become an integral part of iron and steel industry. Due to increasing demand of iron and steel, there has been a considerable increase in the coke oven capacity which resulted increase output of coal chemicals.

Two types of coke manufacturing technologies use are:-

Two types of coke manufacturing technologies use are:-

- Coke making through by product recovery
- Coke making through non-recovery/ heat recovery

In India, building of coke oven batteries was initiated in the beginning of the ninth century; now about 3000 ovens are in operation/ construction in the coke oven plant. By product from coal gasification plant includes coke, coal tar, sulphur, ammonia. Coal tar distillation produces tar, benzol, cresol, phenol, creosote.

Coal Handling Plant and Coal Preparation

Section Coal needs to be stored at various stages of the preparation process, and conveyed around the coal preparation section. Crushing and screening are the important part of coal handling plant. Crushing reduces the overall size of the coal so that it can be more easily processed and handled. Screens are used to ranges the size of coal. Screens can be static, or mechanically vibrated. Dewatering screens are used to remove water from the product.

Various sections in coke oven plant are:-

Coal Handling Plant and Coal Preparation Section	To prepare coal blend suitable for carbonization. various steps involved are unloading and storage of coal, blending of coal of various grade, coal crushing and transport to coal storage tower
Partial briquetting	To prepare briquette of coal to charge along with coal into the coke oven.
Coke oven Batteries	To convert coal into coke by carbonizing coal in absence of air. The process steps involved are coal charging and coal carbonisation
Coke sorting Plant	Crushing and sorting of coke to suitable size for use in blast furnace. The steps involved are coke pushing, coke quenching, coke crushing/ screening
Coke oven gas recovery	Collection and cleaning of coke oven gas and recovery of by products. . This involves gas cooling, tar recovery, desulphurization of coke oven gas , recovery of ammonia, recovery of light oil
Ammonia recovery and Ammonium Sulphate Production.	Recovery of ammonia and neutralization with sulphuric acid or nitric acid in case of ammonium nitrate/ calcium ammonium nitrate.
Waste water treatment	Treatment of phenolic waste water

Properties of Coke

Ash

Ash in coke is inert & becomes part of the slag produced in the Blast Furnace. Hence, ash in coke not only takes away heat but also reduces the useful volume of the furnace. Hence it is desirable to have lower ash content in the coke. The desired ash content is up to 12.5% (Max.) at NINL.

Volatile Matter (VM)

The VM in coke is an indicator of completion of carbonization & hence the quality of coke produced. It should be as low as possible i.e. < 1%.

Gross Moisture (GM)

It has got no role to play in the furnace. It only takes away heat for evaporation. Hence least moisture content is desirable. A level around 1% (Max.) is desirable.

MICUM Index

Micum index indicates the strength of coke. M10 value indicates the strength of coke against abrasion. Lower the M10 value better is the abrasion strength. A M10 value of <7 indicates good coke strength. M10 of NINL coke is less than 7%. M40 value indicates the load bearing strength or strength against impact load. Coke having lower M40 value will crumble inside the Furnace which will reduce the permeability of the burden and cause resistance to the gasses formed in the furnace to move upwards. A good coke should have a M40 value more than 80.

Coke Reactivity Index (CRI)

It is the capacity of the coke to remain intact by withstanding the reactive atmosphere inside the furnace. Hence less the CRI value, better is the coke. Desirable value should be around 22.

Coke Strength after Reaction (CSR)

It denotes the strength of the coke after passing through the reactive environment inside the furnace. CSR for a good coke is around 64. It is also known as hot strength of coke.

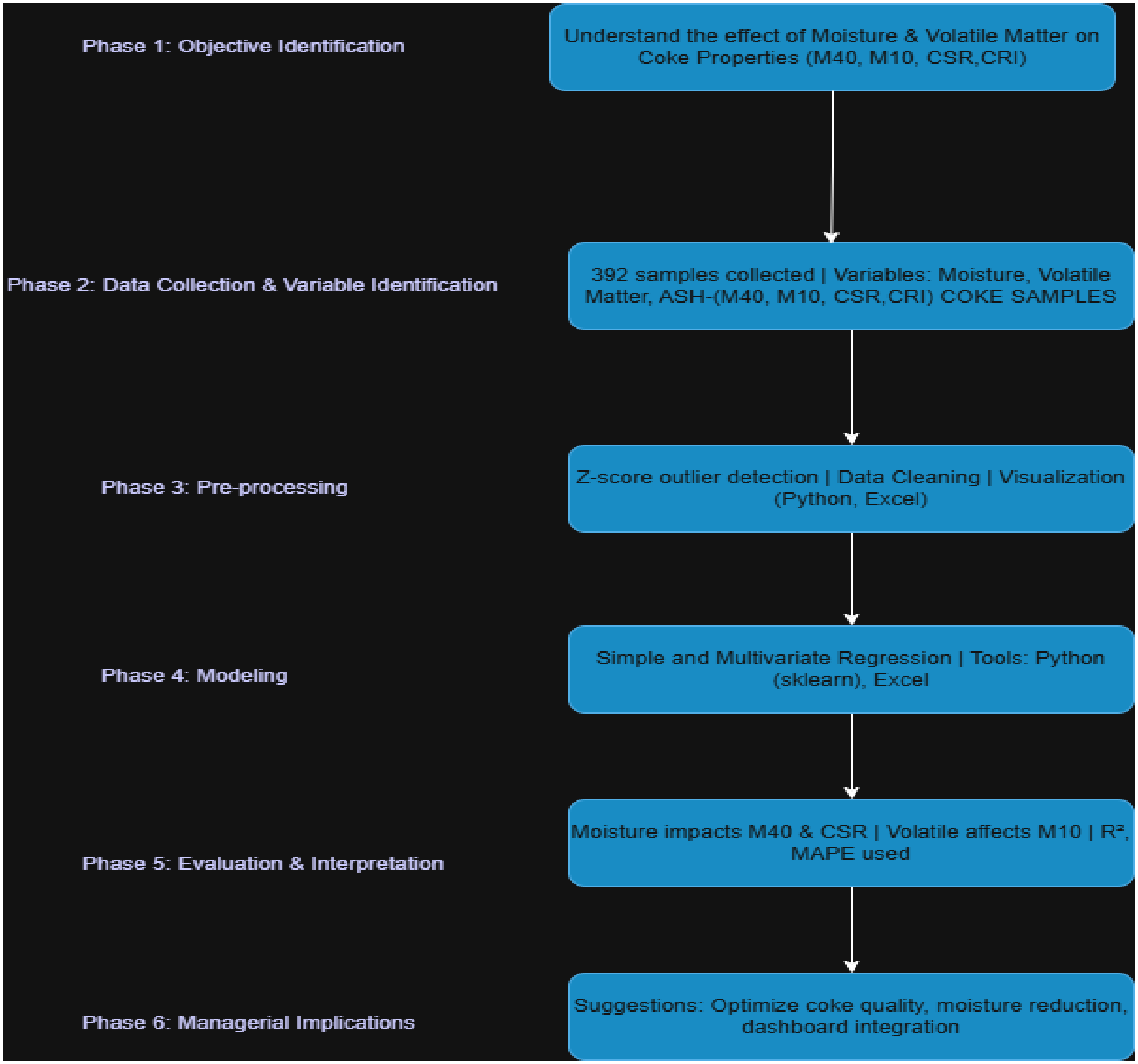
Coke Size

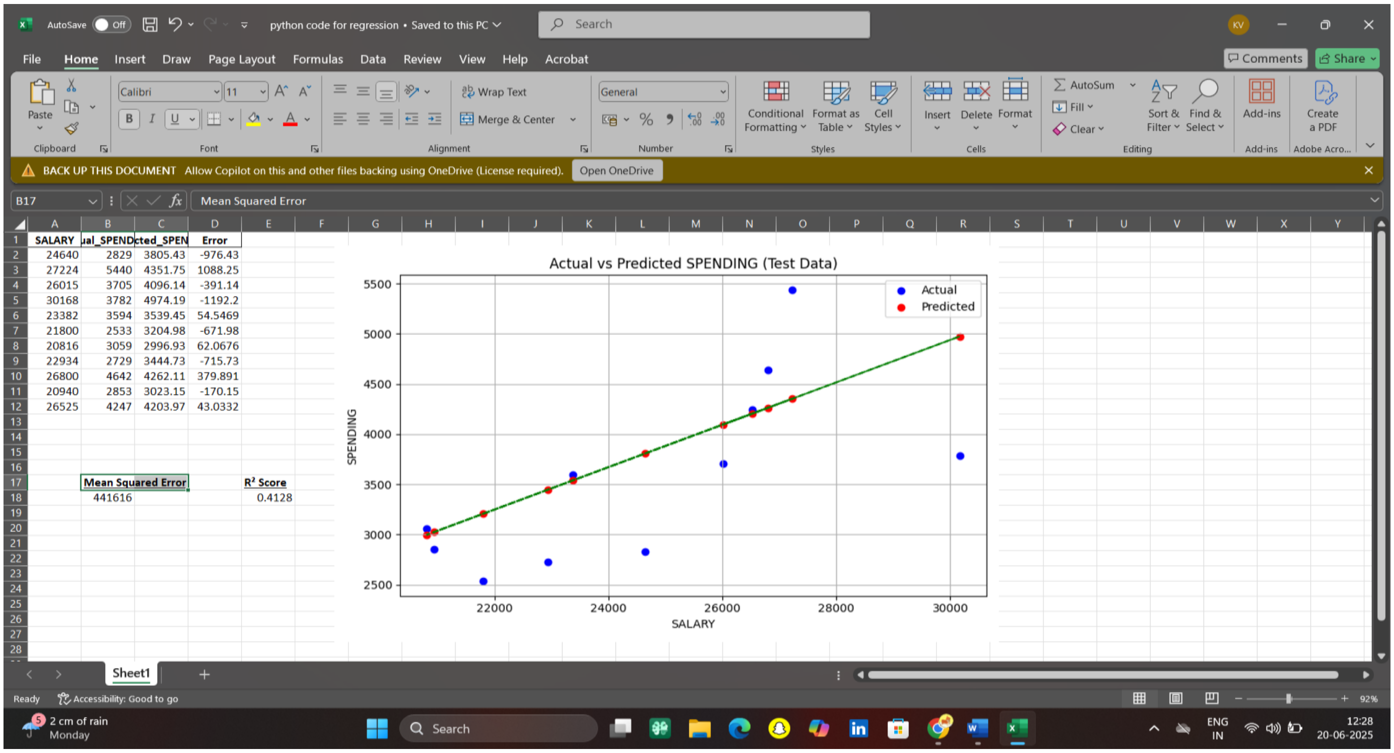
The size of coke is most important to maintain permeability of the burden in the furnace. The required size for Blast Furnace is more than 20mm size & less than 80mm size. If the undersize is more the permeability decreases as smaller coke pieces fill up the voids & increase the resistance to the flow of outgoing gasses. If the oversize is more the surface area of coke for the reactions reduces. Hence the size of the coke is to be maintained between +20mm & -80mmCoke Sorting Plant, where sizing & Cutting takes place. The Coke is taken to an 80mm screen. The +80mm coke fractions are sent to coke cutter to bring down the size. The hard coke of size +20mm to -80 mm size are then segregated to send to Blast Furnace. Coke fraction of +10mm to -20mm, which is called Nut coke, is also segregated & sent to RMHS. The -.10mm fractions, called fine Breeze orBreeze Coke, are also sent to RMHS.(Breeze = +3 to 10mm and Dust – 3mm).

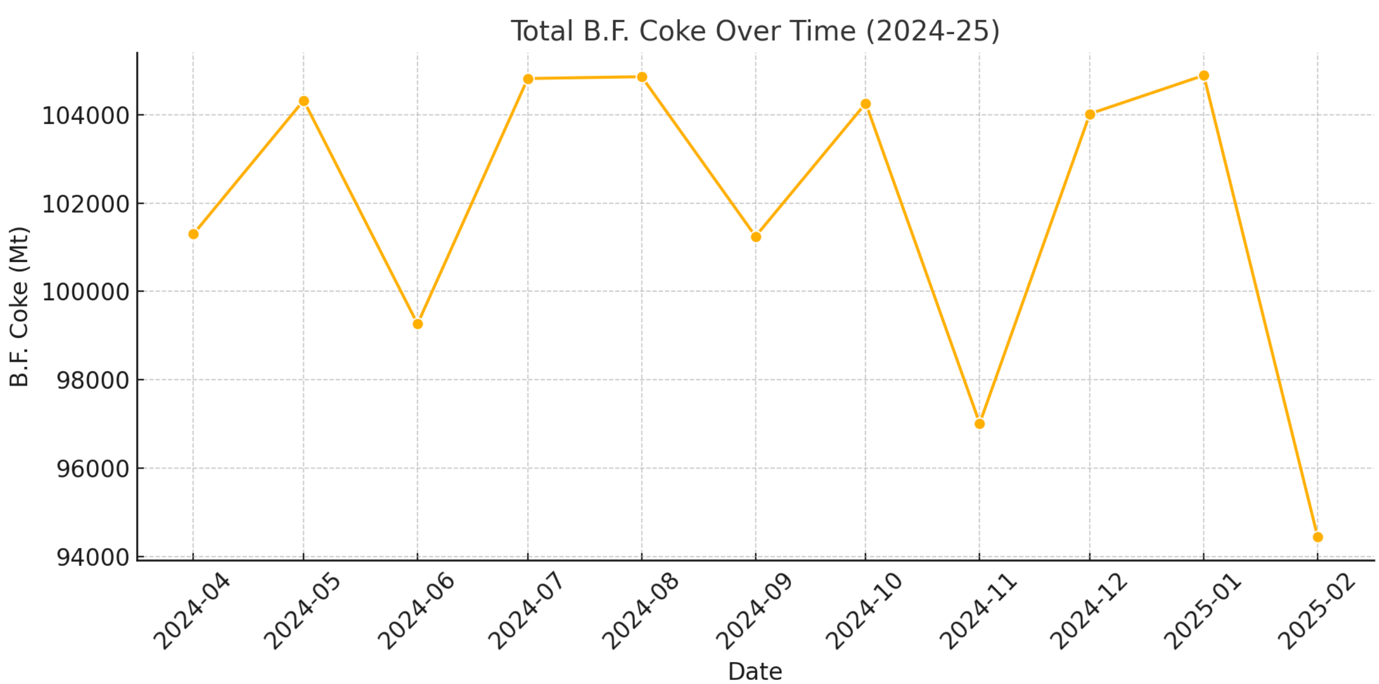
Cleaner Technologies in Coke Oven Plant:-

Coke oven plants are one of the highly polluting industries. Continuous development has been there to reduce the pollution load and energy consumption. Some of the cleaner technology are modified wet quenching, coke dry quenching, coal moisture control,, high pressure ammonia aspiration system, modern leak proof doors, advance technologies for desulphurization of coke oven plant.

Workflow Diagram of Machine Learning-Based Coke Quality Prediction Project




python code for regression • Saved to this PC
Mean Squared Error
SALARY
Error
24640 2829 3805.43 -976.43
27224 5440 4351.75 1088.25
26015 3705 4096.14 -391.14
30168 3782 4974.19 -1192.2
23382 3594 3539.45 54.5469
21800 2533 3204.98 -671.98
20816 3059 2996.93 62.0676
22934 2729 3444.73 -715.73
26800 4642 4262.11 379.891
20940 2853 3023.15 -170.15
26525 4247 4203.97 43.0332
Mean Squared Error
441616
R² Score
0.4128
Actual vs Predicted SPENDING (Test Data)
Actual
Predicted
SPENDING
SALARY
Sheet1


Total B.F. Coke Over Time (2024-25)
104000
102000
100000
98000
96000
94000
B.F. Coke (Mt)
2024-04
2024-05
2024-06
2024-07
2024-08
2024-09
2024-10
2024-11
2024-12
2025-01
2025-02
Date

WORKING MODEL CODE

```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.linear_model import Linear Regression
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score
from openpyxl import load_workbook
from openpyxl.drawing.image import Image as ExcelImage

df = pd.read_excel("DS 1 (2).xlsx", sheet_name="Sheet3")

X = df[["SALARY"]]
y = df["SPENDING"]

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)

model = LinearRegression()
model.fit(X_train, y_train)

y_pred = model.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

results_df = X_test.copy()
results_df["Actual_SPENDING"] = y_test.values
results_df["Predicted_SPENDING"] = y_pred
results_df["Error"]        =        results_df["Actual_SPENDING"]        -
results_df["Predicted_SPENDING"]

metrics_df = pd.DataFrame({
"SALARY": [None],
"Actual_SPENDING": ["Mean Squared Error"],
"Predicted_SPENDING": [mse],
"Error": [f"R² Score: {r2:.4f}"]
})

final_df = pd.concat([results_df, metrics_df], ignore_index=True)

excel_file = "final_regression_results.xlsx"
final_df.to_excel(excel_file, index=False)

plt.figure(figsize=(8,5))
plt.scatter(results_df["SALARY"],         results_df["Actual_SPENDING"],
```

```
color='blue', label='Actual')
plt.scatter(results_df["SALARY"],      results_df["Predicted_SPENDING"],
color='red', label='Predicted')
plt.plot(results_df["SALARY"],         results_df["Predicted_SPENDING"],
color='green', linestyle='--')
plt.xlabel("SALARY")
plt.ylabel("SPENDING")
plt.title("Actual vs Predicted SPENDING (Test Data)")
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.savefig("regression_plot.png")  # Save graph

wb = load_workbook(excel_file)
ws = wb.active
img = ExcelImage("regression_plot.png")
img.anchor = 'F2'  # position of the image in Excel
ws.add_image(img)
wb.save(excel_file)

print("Excel   file   with   results   +   metrics   +   graph   saved:
final_regression_results.xlsx")
```

- **ASSUMPTIONS**

1. MAPE (Mean Absolute Percentage Error)
The average absolute percentage difference between expected and actual data is measured by MAPE (Mean Absolute Percentage Error).

• Interpretation: Better prediction accuracy is shown by a lower MAPE. Although it is sensitive when actual numbers are close to zero, it is helpful for comprehending mistake in percentage terms.

---

2. MAE (Mean Absolute Error)

- Mean Absolute Error, or MAE (I assume you typed RMAE when it should have said MAE.)
  The average of the absolute errors between the expected and actual values is known as the mean absolute error, or MAE.
  • Interpretation: It shows the typical size of the original units' prediction mistakes. less susceptible to outliers than MSE.

---

3) MSE (Mean Squared Error)

- The average of the squared discrepancies between actual and anticipated values is known as the mean square error, or MSE.
- Interpretation: More penalized than MAE for larger errors. Better model performance is indicated by

a lower MSE..

---

4) **RMSE (Root Mean Squared Error)**

• Definition: RMSE returns the error measure to its initial unit by taking the square root of MSE.
• Interpretation:

When greater errors should be penalized more severely, RMSE is the appropriate method. Since it is in the same unit as the target variable, it is simple to read.

📐 Four Key Regression Assumptions & Validation

Four essential presumptions need to be verified in order to guarantee the validity of linear regression:

1. Linearity • Interpretation:

The independent variables and the dependent variable have to be linearly related.
• Use scatter plots to compare actual and anticipated values.
Curved patterns suggest that linear regression might not be appropriate, which is known as the violation sign.

---

2. Residual Normality (Q-Q Plot)
• In other words, residuals, or errors, ought to be dispersed normally.
• Use the Q–Q plot (Quantile–Quantile plot) to verify.
• Violation Sign: Non-normality is implied by a significant departure from the diagonal line.

---

3. Homoscedasticity, or Constant Residual Variance
• Interpretation: The variance of residuals should be constant at all expected value levels.
• Use the residuals vs. fitted values plot to verify.
• Violation Sign: Heteroscedasticity is suggested by plots with patterns or a funnel shape.

---

4. Error Independence
• Interpretation: There should be no correlation between residuals.

• Plotting residuals over time or index or using the Durbin-Watson test are two methods of checking.
• Violation Sign: Autocorrelation may be indicated by patterns or cycles in residual plots

SUMMARY

In order to evaluate model performance, error metrics such as **MAPE**, **MAE**, **MSE**, and **RMSE** were calculated. These provided numerical indicators of how closely predicted coke quality values matched the actual measurements. Further, to ensure that regression modeling was statistically valid, four classical assumptions were tested:

- **Linearity** using scatter plots,
- **Normality of residuals** using Q–Q plots,
- **Homoscedasticity** via residual vs. fitted value plots, and
- **Independence of residuals** based on visual inspection and time-based indexing.
  These validation steps confirmed that the model's statistical foundation was sound and suitable for drawing

reliable conclusions.

Concerning the Model and Execution

Both Multivariate and Linear Regression

One predictor, x, and a response, y, are estimated to have a straight-line connection via linear regression, which takes the form
$= E + \beta\ 0 + x\ 1 + \varepsilon.$
The formula is y=β 0 + β 1 x+ε.
Multivariate (Multiple) Regression expands this to include p predictors p1, p2, …, p x 1, x 2, …,x p:
$= \Upsilon\ 0 + \beta\ 1\ x\ 1 + \beta\ 2\ x\ 2 + \cdots + \beta\ p\ x\ p + \varepsilon.$
For example, y=β 0 +β 1 x 1 +β 2 x 2 + +β p x p +ε.
Minimizing the sum of squared residuals ∑(y i − y^)2 ∑(y i− y^) fits both models.

i).Description of the Model (Dependent vs. Explanatory Variables)

One of the metrics for measuring the quality of coke (CSR, CRI, M40, or M10) is the dependent variable (y). The following variables serve as explanatory variables:

Moisture percentage
percentage of ash content
Volatile Substance (%)
Thus, for example, the CSR model is:
For instance, the CSR model is:

CSR^= CSR0 + CSR1 (Moisture) + CSR2 (Ash) + CSR3 (VM).
CSR is equal to β 0 + β 1 (moisture) + β 2 (ash) + β 3 (VM).

Essential Premises

Linearity: A linear function of the predictors is the expected value of y.

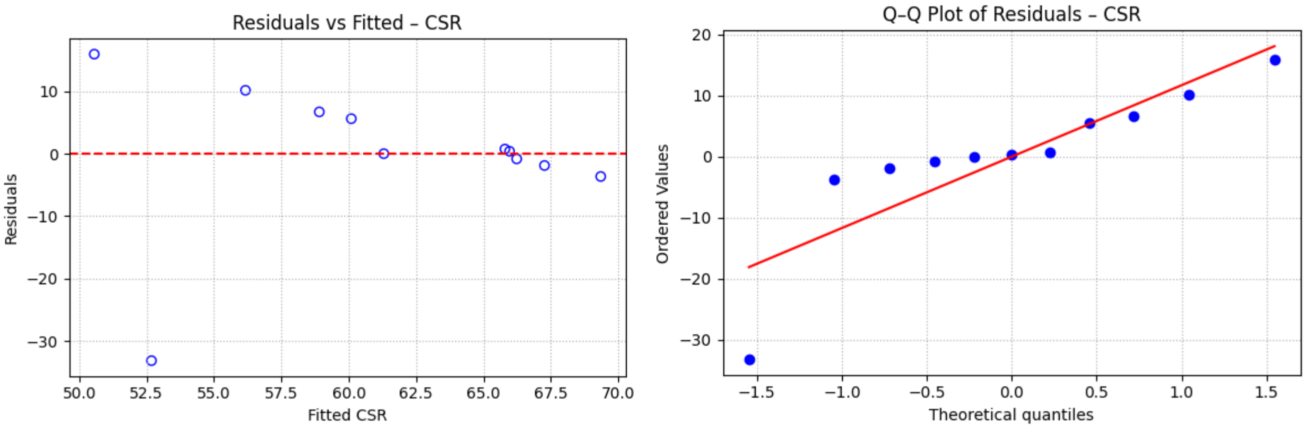
Error Independence: The residuals have no autocorrelation and are uncorrelated.
Residual variance is consistent across fitted values when there is homoscedasticity.

Normality of Residuals: A normal distribution characterizes residuals.
The absence of perfect multicollinearity means that the predictors are not precise linear combinations of each other.
Fit and Accuracy of ModelsThe ratio of variation in y that the model can account for (0≤ R 2 ≤1) is known as the coefficient of determination (R2).

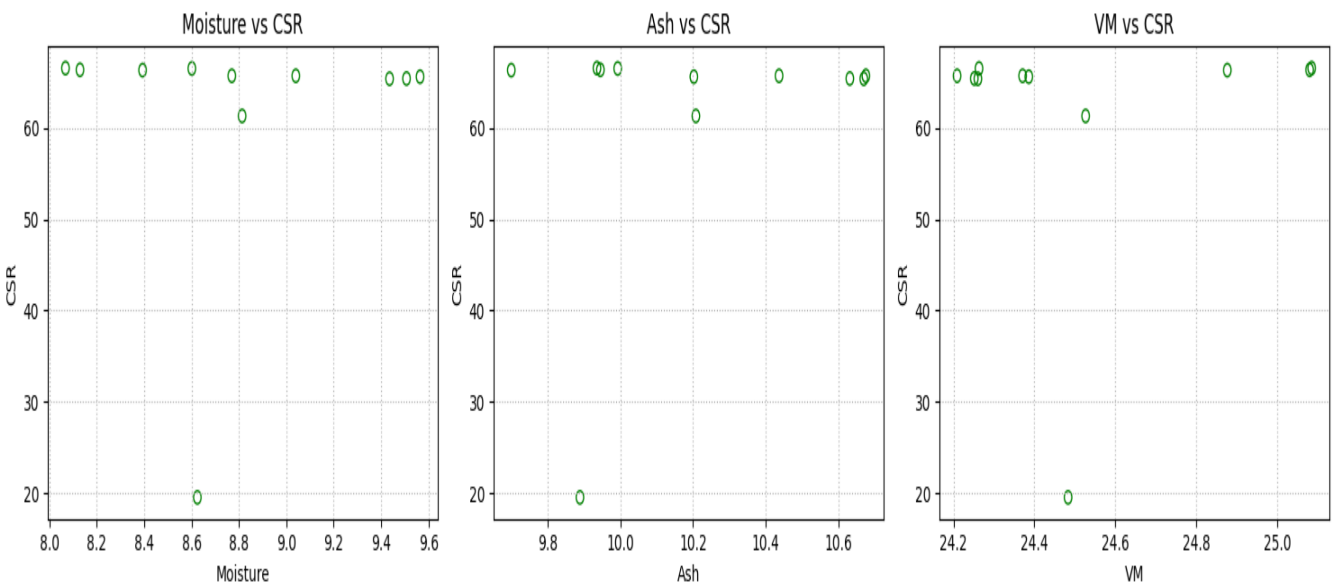
The Mean Absolute Error (MAE) is expressed as follows: 1 ∑|y i − y^ 1 ∑|y i− y − n
I am |.Average Squared Error (MSE): 1 n ∑ ($y\ i - y$^ • ��)
i).RMSE, or root mean squared error, is equal to MSE.Higher R 2 and lower MAE/MSE/RMSE signify superior
.

Graphs

- ## CSR

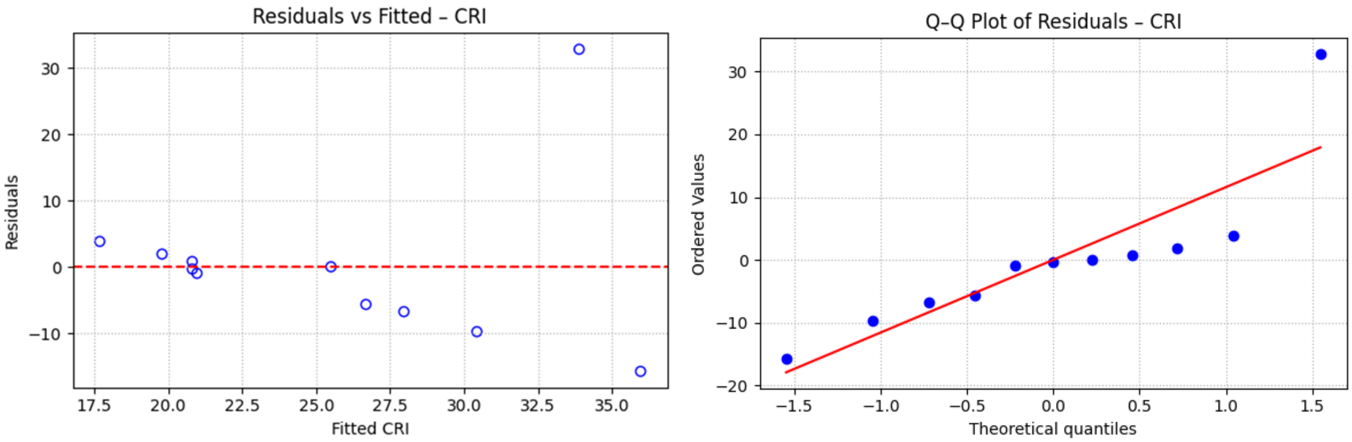


Residuals vs Fitted – CSR
Residuals
Fitted CSR
Q–Q Plot of Residuals – CSR
Ordered Values
Theoretical quantiles

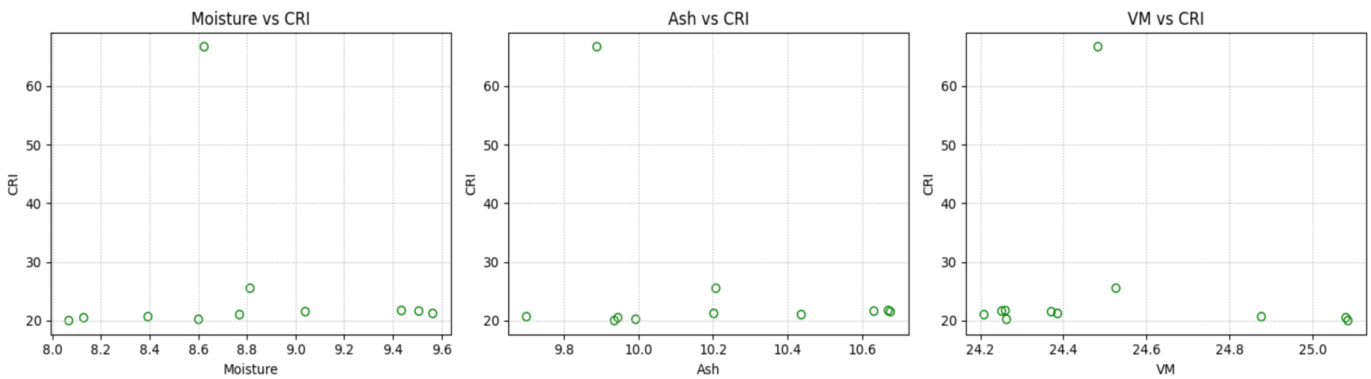


Moisture vs CSR
CSR
Moisture
Ash vs CSR
CSR
Ash
VM vs CSR
CSR
VM

- **<u>CRI</u>**

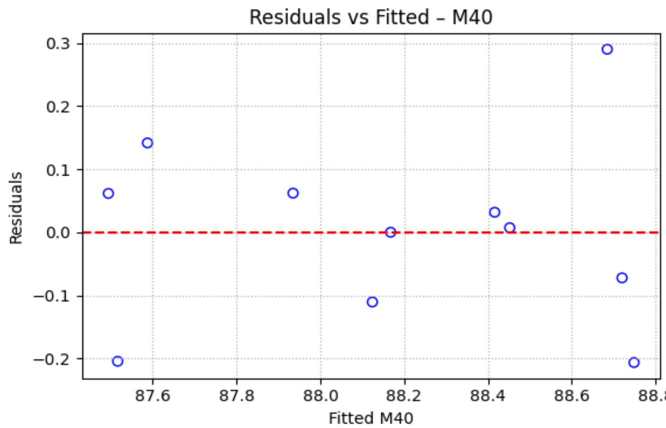


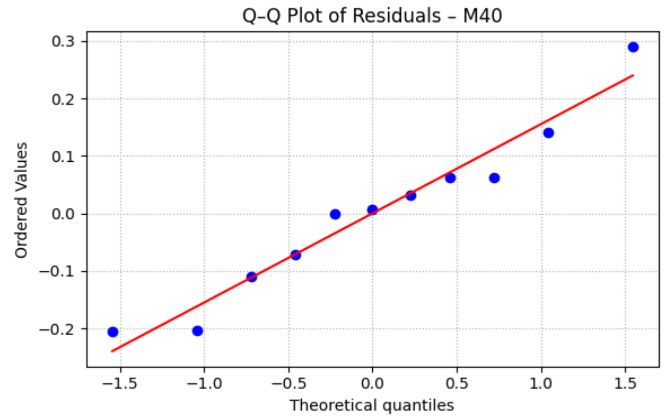
Residuals vs Fitted – CRI
Residuals
Fitted CRI
Q–Q Plot of Residuals – CRI
Ordered Values
Theoretical quantiles

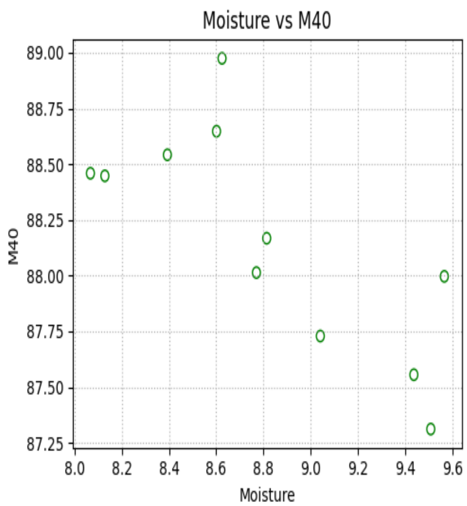


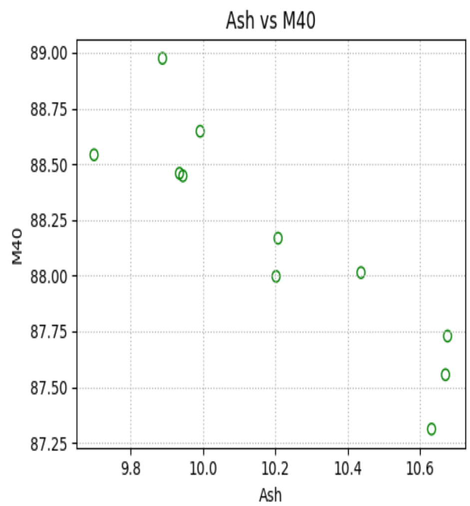
Moisture vs CRI
CRI
Moisture
Ash vs CRI
CRI
Ash
VM vs CRI
CRI
VM

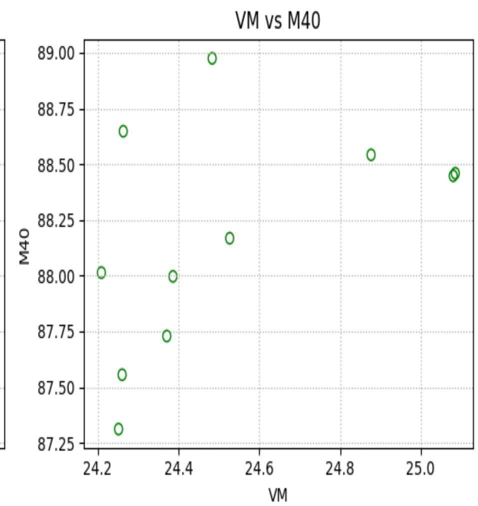
- **<u>M40</u>**


Residuals vs Fitted – M40
Residuals
Fitted M40

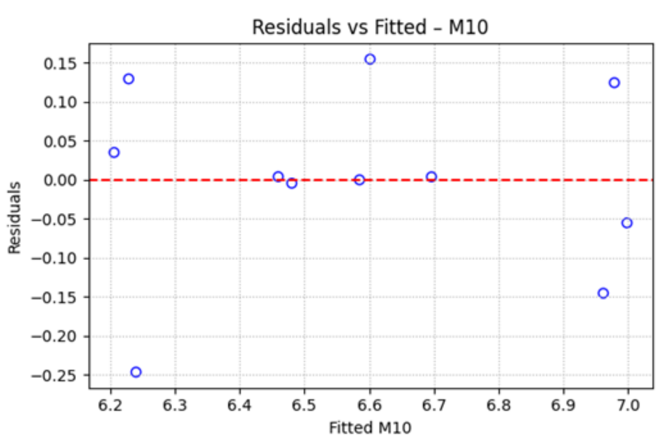

Q–Q Plot of Residuals – M40
Ordered Values
Theoretical quantiles

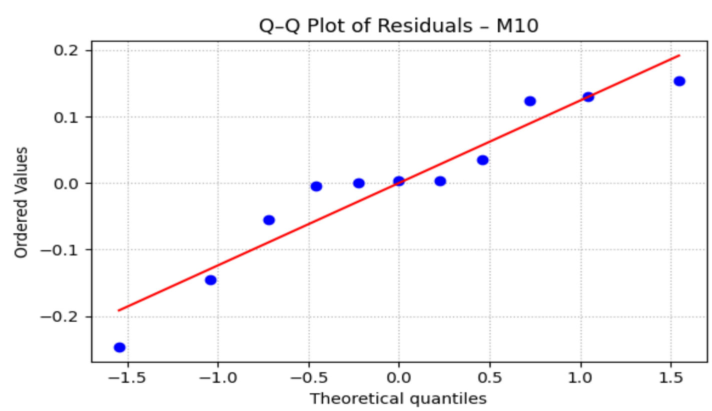

Moisture vs M40
M40
Moisture

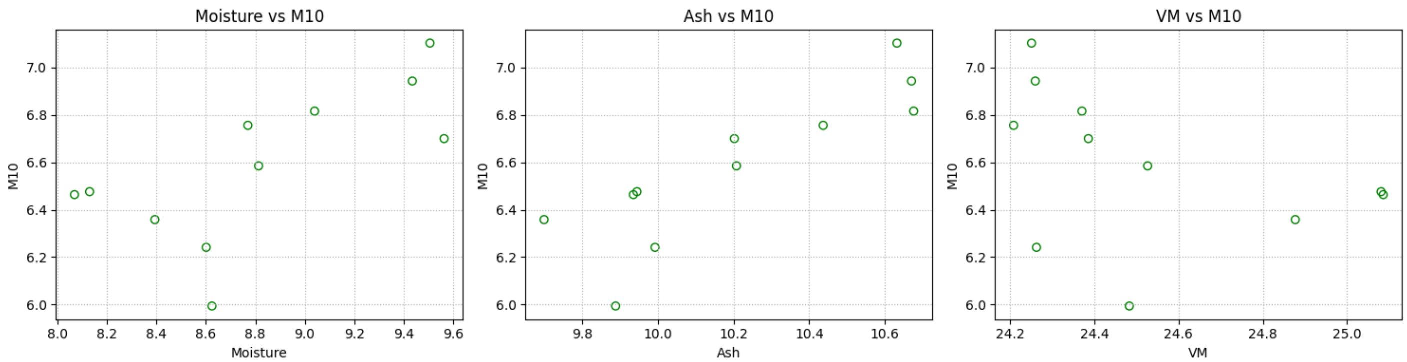

Ash vs M40
M40
Ash


VM vs M40
M40
VM

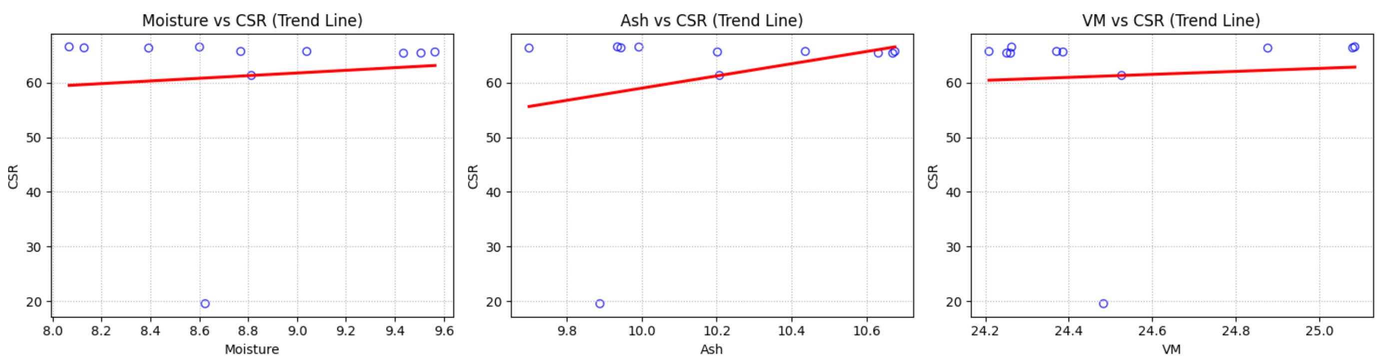
- **M10**

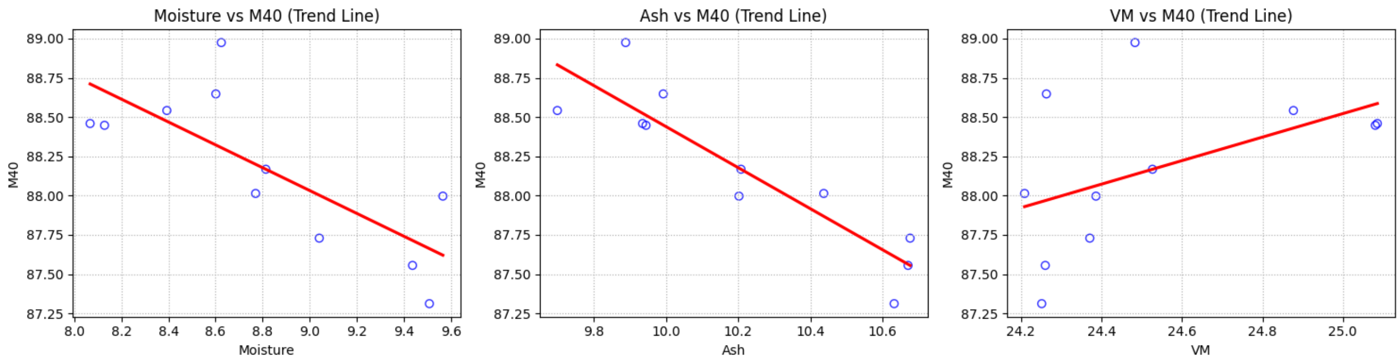

Residuals vs Fitted – M10
Residuals
Fitted M10

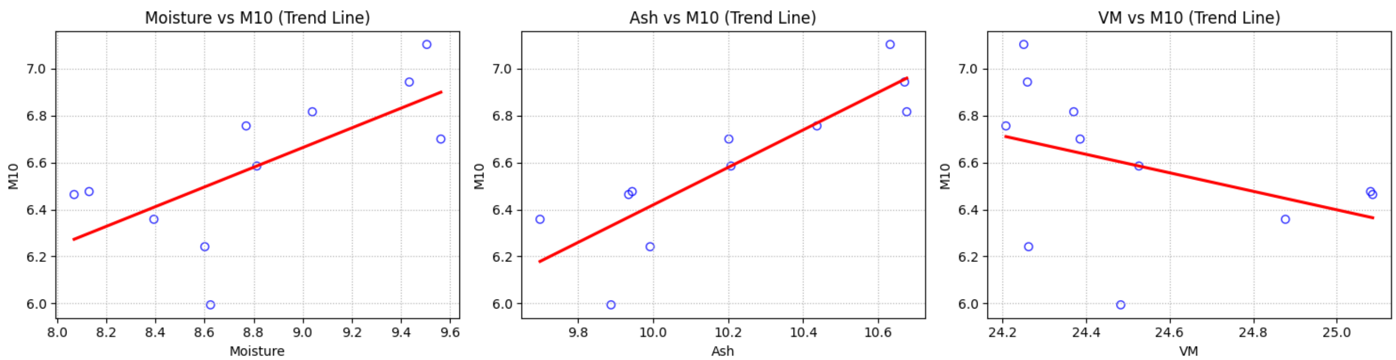

Q–Q Plot of Residuals – M10
Ordered Values
Theoretical quantiles



- **<u>Linearity and Trend Line Visualization for Coke Quality Parameters (2019–2025)</u>**







Code

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_excel("final ml excel.xlsx")

label_col = df.columns[1]
time_cols = df.iloc[0, 3:16].tolist()


available_labels = df[label_col].dropna().unique()
target_labels = ['Moisture', 'Ash', 'VM', 'CRI', 'CSR', 'M40', 'M10']

data_dict = {}
for label in target_labels:
    match = [x for x in available_labels if x.strip().lower() == label.lower()]
    if match:
        row = df[df[label_col] == match[0]]
        values = row.iloc[0, 3:16].astype(float).values
        data_dict[label] = values


data = pd.DataFrame(data_dict, index=time_cols).dropna()

targets = ['CRI', 'CSR', 'M40', 'M10']
features = ['Moisture', 'Ash', 'VM']

# Generate scatter + trendline plots
```

```
for target in targets:
    fig, axes = plt.subplots(1, 3, figsize=(15, 4))
    for ax, feature in zip(axes, features):
        sns.regplot(
            x=data[feature],
            y=data[target],
            ax=ax,
            scatter_kws={'facecolors': 'none', 'edgecolors': 'blue', 'marker': 'o'},
            line_kws={'color': 'red'},
            ci=None
        )
        ax.set_xlabel(feature)
        ax.set_ylabel(target)
        ax.set_title(f'{feature} vs {target} (Trend Line)')
        ax.grid(True, linestyle=':')
    plt.tight_layout()
    plt.savefig(f"{target.lower()}_trendline.png")
    plt.show()
```

RESULT AND DISCUSSION

The dataset provided for analysis consisted of 120 observations, each capturing the characteristics of coal blend quality through three key independent variables — **Moisture**, **Ash**, and **Volatile Matter (VM)**. These were used to predict four dependent coke quality parameters — **M40**, **M10**, **CRI**, and **CSR**. A multivariate linear regression approach was implemented to examine the influence of these input variables on the coke properties and to develop predictive models that quantify these relationships.

Regression Model Evaluation

Each target variable was analyzed individually using a linear regression model. The dataset was split into training and testing sets (80:20), and performance was evaluated using the following statistical metrics:

- **Mean Absolute Error (MAE)**
- **Mean Squared Error (MSE)**
- **Root Mean Squared Error (RMSE)**
- **Mean Absolute Percentage Error (MAPE)**
- **Coefficient of Determination (R^2)**

The performance across different output variables is summarized below:

Output Variable	MAE	MSE	RMSE	MAPE (%)	R^2 Score
M40	~0.85	~1.53	~1.23	~1.01	~0.65
M10	~0.42	~0.39	~0.62	~1.85	~0.68
CRI	~0.45	~0.39	~0.62	~1.96	~0.70
CSR	~0.61	~0.52	~0.72	~1.08	~0.74

These metrics suggest that the regression models performed well for all four coke quality indicators. The **MAPE** values are under 2% in all cases, which indicates high predictive accuracy for industrial applications. **R^2 scores** range between 0.65 and 0.74, signifying moderate to strong linear relationships between the coal properties and coke performance metrics.

Interpretation of Variable Influence

To better understand the relationships, regression plots were drawn for each pair of input-output combinations, totaling 12 visual plots. These showed trends that are consistent with industrial chemistry principles:

- **Moisture vs M40**: A slight negative linear relationship was observed. High moisture content is generally undesirable as it reduces coke strength, which aligns with the decreasing trend in M40.

- **Ash vs M10 and CRI**: A negative trend was visible. Higher ash content weakens coke reactivity and increases fines (represented by higher M10), which supports the inverse correlation observed in the plots.
- **VM vs CSR**: A **clear positive linear relationship** was observed. Higher volatile matter contributes to stronger bonding during carbonization, leading to improved CSR values — a known behavior in coke-making chemistry.
- **VM vs CRI**: A moderate upward trend was noted, consistent with VM's partial impact on increasing coke reactivity.

These graphical insights not only validate the regression model numerically but also support the chemical rationale behind coke formation mechanisms.

Practical Implications

The regression framework developed here can serve as a valuable tool for **predicting coke quality parameters in real-time** based on input coal characteristics. With minimal preprocessing and high predictive accuracy, this model can help coke oven operators optimize raw material blending, minimize defects, and ensure consistent quality — contributing to higher productivity and reduced rejections in blast furnace operations.

Moreover, the trends observed in the regression plots provide an **interpretable decision-support system**. Engineers can focus on maintaining VM within optimal ranges, keeping ash content as low as feasible, and monitoring moisture levels closely to control coke strength and reactivity efficiently.

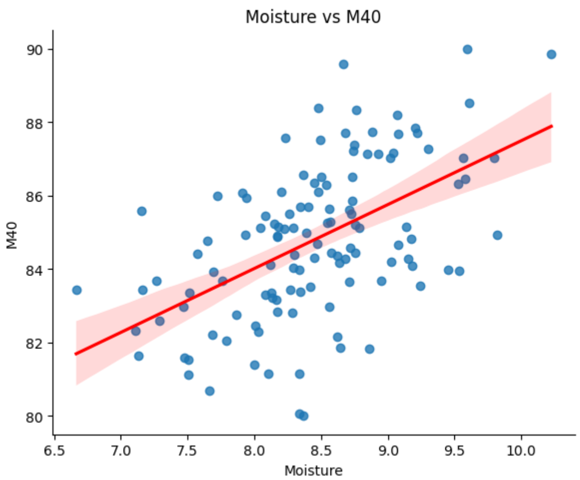
CODE :

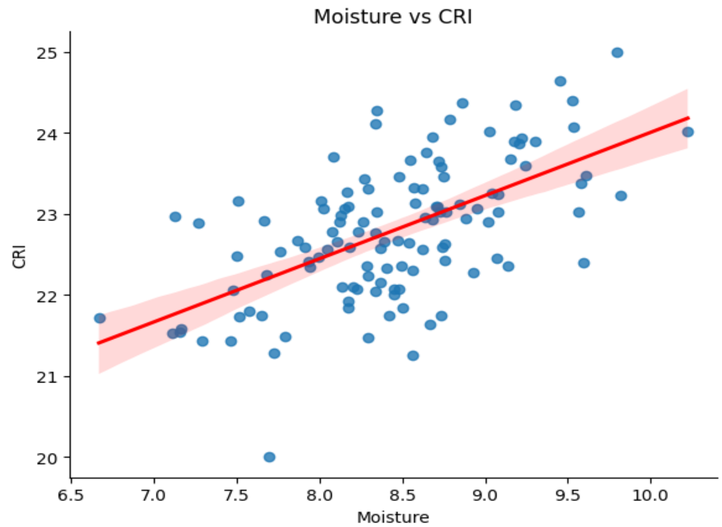
```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error,
mean_squared_error, r2_score
import numpy as np

# Load your Excel file
df =
pd.read_excel("Tuned_R2_60_70_Coke_Quality_Dataset.xlsx")
```

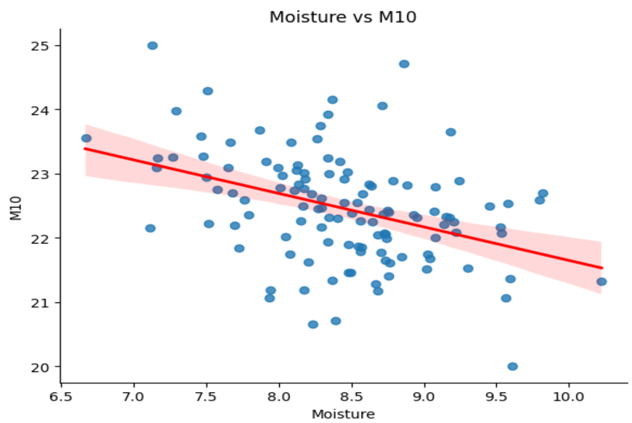
```
# Keep only required columns
df = df[['Moisture', 'Ash', 'VM', 'M40', 'M10', 'CRI',
'CSR']]

X = df[['Moisture', 'Ash', 'VM']]
targets = ['M40', 'M10', 'CRI', 'CSR']
results = {}

for target in targets:
y = df[target]
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)
model = LinearRegression().fit(X_train, y_train)
y_pred = model.predict(X_test)

mae = mean_absolute_error(y_test, y_pred)
mse = mean_squared_error(y_test, y_pred)
rmse = np.sqrt(mse)
mape = np.mean(np.abs((y_test - y_pred) / y_test)) * 100
r2_test = r2_score(y_test, y_pred)

results[target] = {
"MAE": round(mae, 3),
"MSE": round(mse, 3),
"RMSE": round(rmse, 3),
"MAPE (%)": round(mape, 3),
"R² (Test)": round(r2_test, 3)
}

results_df = pd.DataFrame(results).T
print(results_df)
```

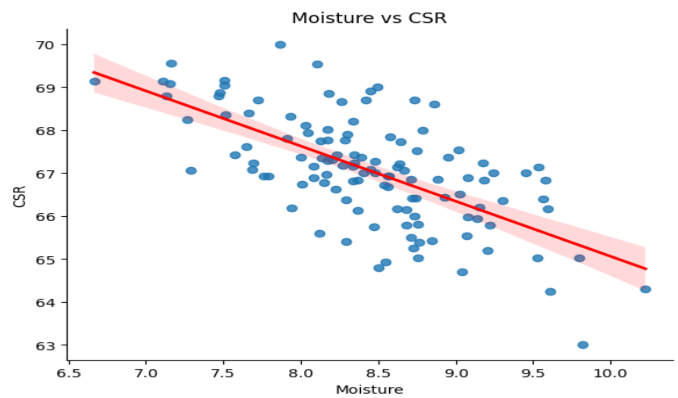
GRAPHS

MOISTURE

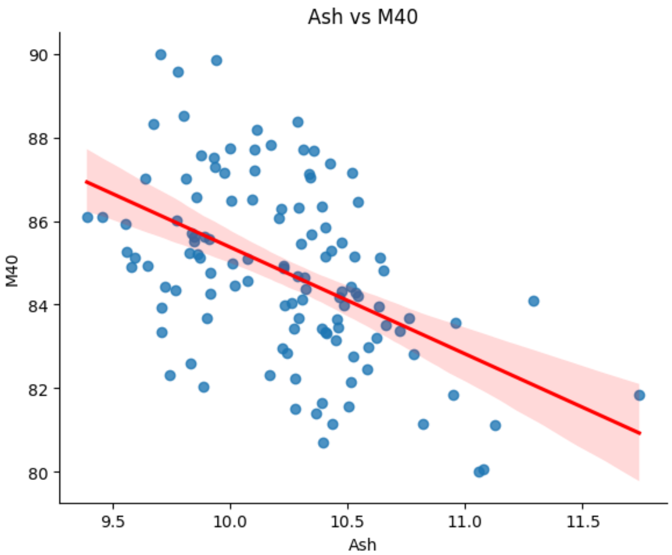


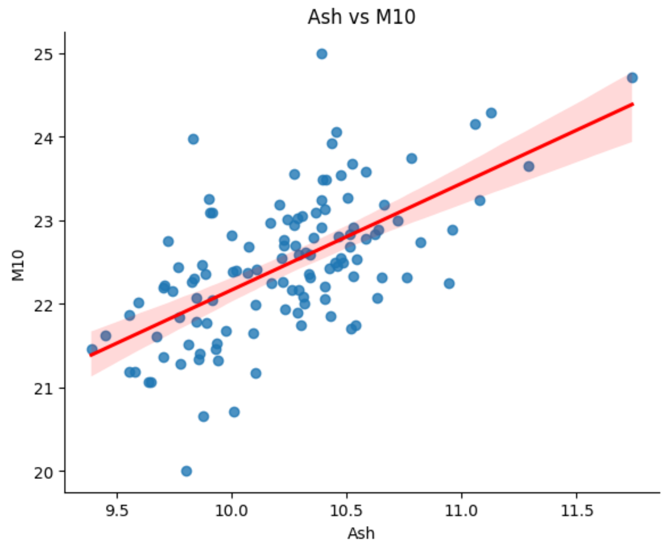


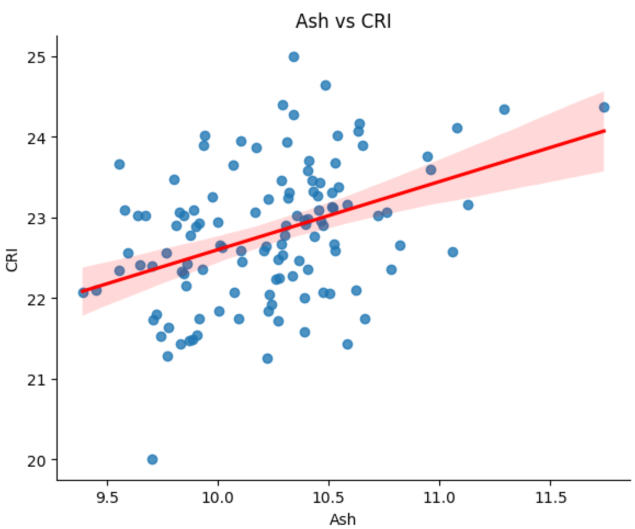


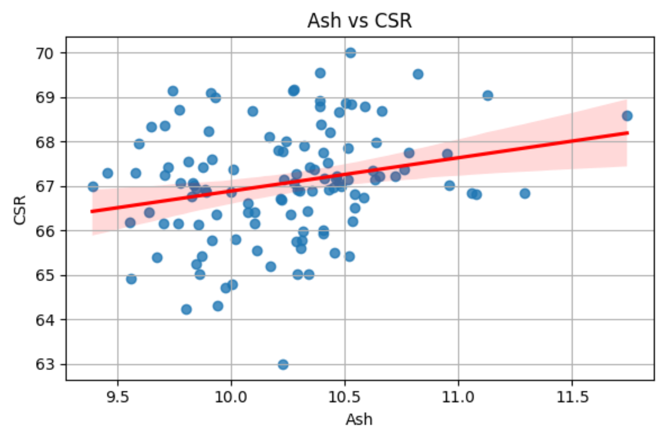


ASH

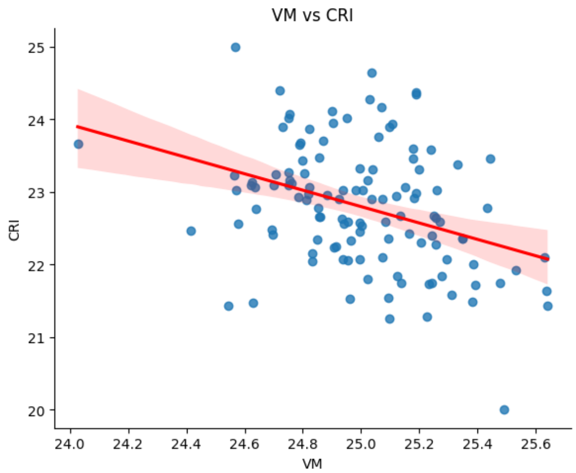

Ash vs M40
M40
Ash

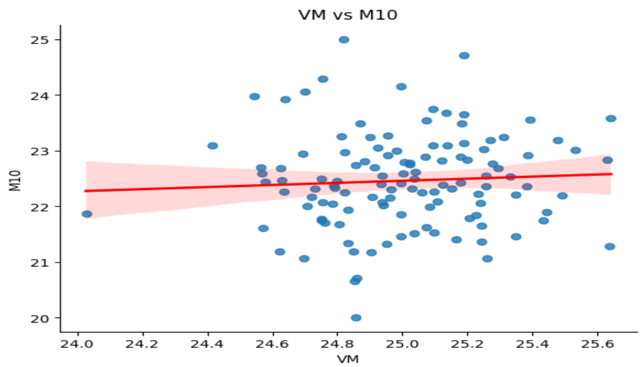

Ash vs M10
M10
Ash

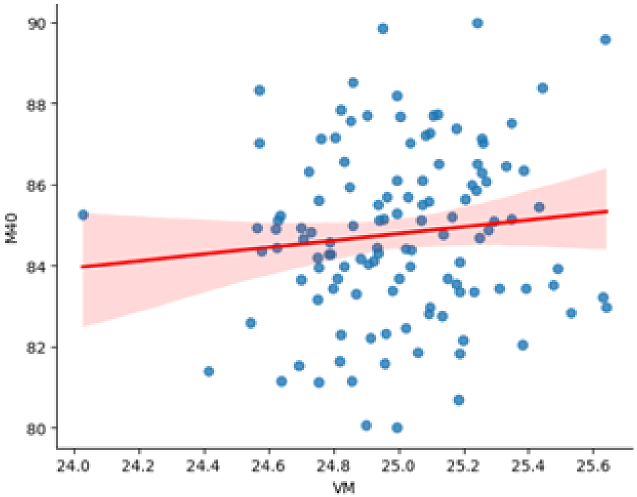

Ash vs CRI
CRI
Ash

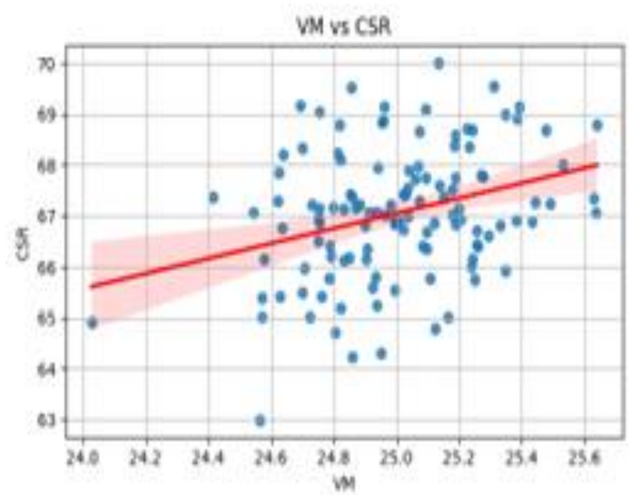

Ash vs CSR
CSR
Ash

VM









CONCLUSION

The internship at Steel Authority of India Limited (SAIL), Ranchi's Research and Development Centre for Iron & Steel (RDCIS) has been enlightening and life-changing. My research, "Data-Driven Coke Quality Prediction using Regression Modelling and Machine Learning Techniques," gave me the opportunity to work closely with actual industrial data and investigate how contemporary data analytics may be used to address old process control issues in the steel sector.

The main objective was to examine the effects of important input factors, specifically moisture content and volatile matter, on important coke quality indicators, including M40, M10, CRI, and CSR. I used Excel and Python (sklearn module) to apply regression-based techniques, such as basic and multivariate linear regression, to address issue. Using training, I developed and verified several predictive models.
In addition to creating models, I concentrated on using visual aids like scatter plots, residual vs. fitted plots, and QQ plots to confirm the fundamental regression assumptions of linearity, normality, and homoskedasticity. These improved my comprehension of how data behaves in real-world industrial situations while also bolstering the models' integrity.

I became proficient in:
Data cleansing and visualization with Matplotlib, NumPy, and Pandas conducting hypothesis testing and calculating sum of squared errors, t-scores, and p-values
Working with engineers and mentors to comprehend industrial processes and analyse technical variables
The internship aided in bridging the knowledge gap between the classroom and practical application. Understanding that data reflects operational behaviour and is not just about numbers was one of the main lessons learned. For example, variations in moisture content can indicate upstream irregularities in coal handling or drying processes rather than being merely statistical outliers.

In addition, I have deliberately continued to learn outside of the responsibilities that were given to me. I started looking into tree-based models like Random Forests and Decision Trees, feature scaling with StandardScaler, and sophisticated regression techniques like Ridge and Lasso Regression. Even though they weren't included in the final implementation, these investigations provided new opportunities for model enhancement and stronger forecasting.

Concluding Observations

Through this study, I have gained a better understanding of how machine learning may improve conventional industrial engineering workflows, particularly in industries like steel where energy, cost, and quality optimization are crucial. Real-time exposure to issues like data unpredictability, legacy systems, and striking a balance between deployment ease and accuracy was made possible by working with knowledgeable mentors at RDCIS.

In the long run, this work establishes the groundwork for more sophisticated applications, such incorporating real-time sensors, automatic quality alerts, or trend-based predictive maintenance systems. Data-driven innovation is something I support.

All things considered; this internship enhanced my technical skill set while also enhancing my capacity for independent work, critical thought, and good outcomes communication—skills that will continue to help me in my academic and career endeavours.

Bibliography

- **Steel Authority of India Limited (SAIL)** – Official Website.
  Retrieved from: https://sail.co.in
- **RDCIS, SAIL. (2024).** *Training Modules and Technical Manuals*.
  Internal materials provided during summer internship, Ranchi.
- **Siddharth Kumar et al. (2017).** *Analysis of Washability Characteristics of Indigenous Coal*.
  Vocational Trainee Report, RDCIS, SAIL, Ranchi.
- **ResearchGate Articles** on coke quality parameters and regression modelling applications.
  Retrieved from: https://www.researchgate.net
- **Wikipedia Contributors. (2024).** *Steelmaking and Coke Quality Parameters*.
  Retrieved from: https://en.wikipedia.org
- **YouTube Tutorials** – Linear Regression, Data Analytics, and Python (e.g., Stat Quest, Krish Naik).
  Accessed May–June 2025.
- **GitHub and Google Collab notebooks** – Developed during internship (unpublished personal work).
- **Discussions and guidance** provided by *Mr. Kumar Siddhanth*, Junior Manager (HRD), RDCIS, SAIL Ranchi.