**Report of Industrial Training**

At

**Indian Oil Corporation Ltd.**

**E1, Udhyog Marg, Block A, Sector 1, Noida, Uttar Pradesh 110096**

***In partial fulfilment of the requirements for the award of the degree of***

**Bachelor of Technology**

**in**

**Electronics and Communication Engineering**

*Submitted by*

**Ritu Srivastava**

0901EC161080

**DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING**

**MADHAV INSTITUTE OF TECHNOLOGY**

(A Constituent College of RGPV)

Gwalior, INDIA

**7,November 2019**

| **CONTENTS** |
| --- |
| Certificate |
| Acknowledgement |
| Introduction |
| Training details |
| Conclusion |

**Acknowledgement:**

I would like to gratefully acknowledge the contribution of all the people who took an active part and provided valuable support to me during the project. To begin with I would like to thank Mr. Abhishek Anurag (AM, L&D) for giving me the wonderful opportunity to earn and bring out this project.

I would like to thank my mentor Ms. SOUMYA SRIVASTAVA and Ms. DEEPTI SAHDEV(ma’am), who has been available at all time to help me whenever I needed and provided valuable suggestions, shared her rich corporate experience and helped me script the exact requisites during my internship.

I would like to enormously thank Abhishek Gupta(Manager, L&D) at pipeline division Head Office, for teaching me invaluable skills and sharing with me their knowledge. They have always had trust in my abilities and have appreciated the same.

**INTRODUCTION:**

In this project in analytics of CP/PSP Report under the INFORMATION SYSTEMS department I have enriched my knowledge in statistic and analytics, During the project I worked in the domain of predictive analysis which includes data fetching, data cleaning, data transform, data visualizing and prediction phase. I have used various languages and software packages during my Internship such as Python, R, Qlik Sense etc.

I worked on the data provided by the pipeline division of the company. In India’s infrastructure, the petroleum pipelines form a crucial part enabling sustained availability of petroleum products in all parts of the country for economic growth. The pipelines transport petroleum products from refineries to demand areas and crude oil from import terminals as well as domestic sources to the inland refineries. India is a vast country, a wide network of pipelines becomes the paramount requirement of transporting petroleum products to interiors from refineries and crude oil to the landlocked refineries.  
Therefore proper management of pipelines is very important. Pipelines starts to corrode with continuous usage and various protective measures such as cathodic protection and coating are taken to prevent or at least slow down the process.

My project was Identification of Pipeline Section for Coat and Wrap Refurbishment and my aim was to try to predict using various sensor values provided to us that when we the chainage might require service in the future.

**TRAINING DETAILS:**

**Indian Oil Corporation Ltd.**

Indian Oil Corporation Limited (IOCL), commonly known as IndianOil is an Indian state government owned oil and gas company headquartered in New Delhi. It is the largest commercial oil company in the country, with a net profit of INR 19,106 crore (USD 2,848 million) for the financial year 2016–17. It is ranked 1st in Fortune India 500 list for year 2016[6] and 137th in Fortune's ‘Global 500’ list of world's largest companies in the year 2018. As of 31 March 2017 IndianOil's employee strength is 33,135, out of which 16,545 are in the officer cadre. It is India's largest downstream oil company, with a work force of more than 33,000 employees, a turnover of Rs. 506,428 crore and a net profit of Rs. 21,346 crore in 2017-18.

Indian Oil accounts for nearly half of India's petroleum products market share, 35% national refining capacity (together with its subsidiary Chennai Petroleum Corporation Ltd., or CPCL), and 71% downstream sector pipelines through capacity. The Indian Oil Group owns and operates 11 of India's 23 refineries with a combined refining capacity of 80.7 MMTPA (million metric tonnes per annum).[22] Indian Oil's cross-country pipeline network, for transportation of crude oil to refineries and finished products to high-demand centers, spans over 13,000 km The company has a throughput capacity of 80.49 MMTPA for crude oil and petroleum products and 9.5 MMSCMD for gas. On 19 November 2017, IOC, in collaboration with Ola, launched India’s first electric charging station at one of its petrol-diesel stations in Nagpur.[23] Indian governments’ National Electric Mobility Mission Plan launched in 2013 aims at gradually ensuring a vehicle population of 6 to 7 million electric and hybrid vehicles in India by 2020.

**PROTECTION OF PIPELINES :**

**Cathodic Protection :**

Cathodic protection (CP) is a method to control corrosion by using a direct electrical current which neutralizes external corrosion typically associated with metal pipe. It is generally used when a pipeline is buried underground or in water. When executed on a new pipeline, cathodic protection can prevent corrosion from the start. On an older pipeline, cathodic protection can impede existing corrosion of the line.

**Coatings and Linings :**

Coatings and linings applied to pipelines whether above or below ground and often used in combination with cathodic protection. The Trenton Corporation produces efficient, long-lasting anticorrosion materials, including Wax-Tape® wraps, primers, outer wraps and hot-applied coatings. Fiber-reinforced polymers may also strengthen and repair pipes.

**Corrosion Inhibitors :**

Corrosion inhibitors are compounds added to the upstream pipeline can inhibit the corrosion of carbon and low-alloy steel pipe.

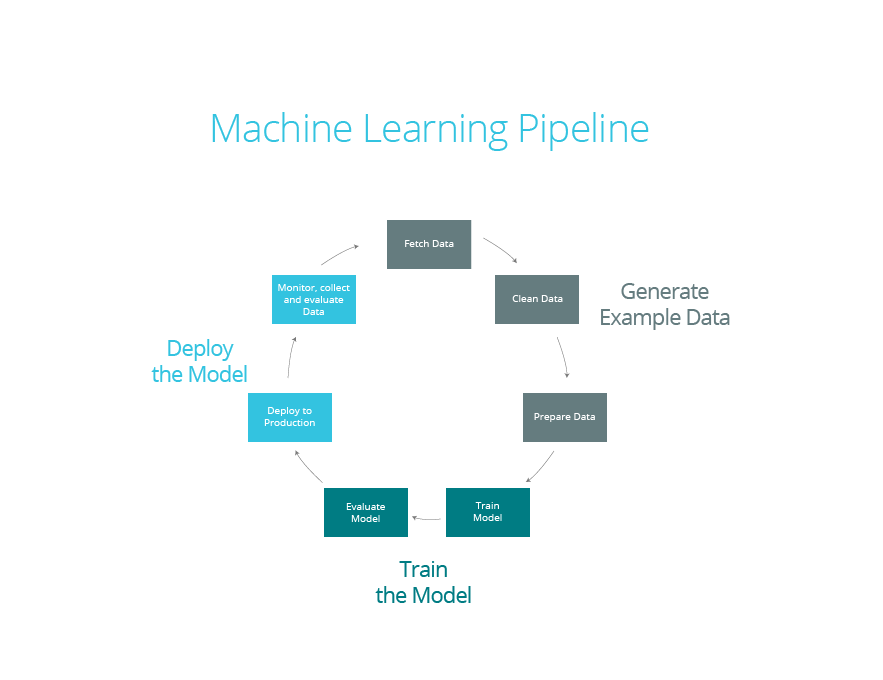
**Pipeline Material :**

Pipeline material used will also significantly influence corrosion. Using materials like plastic, stainless steel or special alloys enhance the lifetime of the pipeline, while steel or steel reinforced concrete is subject to corrosion.

Linc Energy Systems is a distributor of anti-corrosion materials and provides cathodic protection services. If you need assistance with your projects, submit a "request a quote" now.

**PREDICITVE ANALYSIS:**

Predictive analytics is an area of statistics that deals with extracting information from data and using it to predict trends and behavior patterns. The enhancement of predictive web analytics calculates statistical probabilities of future events online. Predictive analytics statistical techniques include data modeling, machine learning, AI, deep learning algorithms and data mining.Often the unknown event of interest is in the future, but predictive analytics can be applied to any type of unknown whether it be in the past, present or future. For example, identifying suspects after a crime has been committed, or credit card fraud as it occurs. The core of predictive analytics relies on capturing relationships between explanatory variables and the predicted variables from past occurrences, and exploiting them to predict the unknown outcome. It is important to note, however, that the accuracy and usability of results will depend greatly on the level of data analysis and the quality of assumptions.



1. **Define project** : Define the project outcomes, deliverable, scope of the effort, business objectives, identify the data sets that are going to be used.
2. **Data collection** : Data mining for predictive analytics prepares data from multiple sources for analysis. This provides a complete view of customer interactions.
3. **Data analysis** : Data Analysis is the process of inspecting, cleaning and modelling data with the objective of discovering useful information, arriving at conclusion
4. **Statistics** : Statistical Analysis enables to validate the assumptions, hypothesis and test them using standard statistical models.
5. **Modelling** : Predictive modelling provides the ability to automatically create accurate predictive models about future. There are also options to choose the best solution with multi-modal evaluation.
6. **Deployment** : Predictive model deployment provides the option to deploy the analytical results into everyday decision making process to get results, reports and output by automating the decisions based on the modelling.

**PROJECT - Identification of Pipeline Section for Coat and Wrap Refurbishment:**

**AIM:**

Predictive Analysis of Coating on Pipelines Sections based on PSP data, Coating History, and DCVG Reports. It will help locate coating faults, third party threats and highlight deficiencies in the Cathodic Protection (CP) strategies thus enabling efficient Pipeline operation.

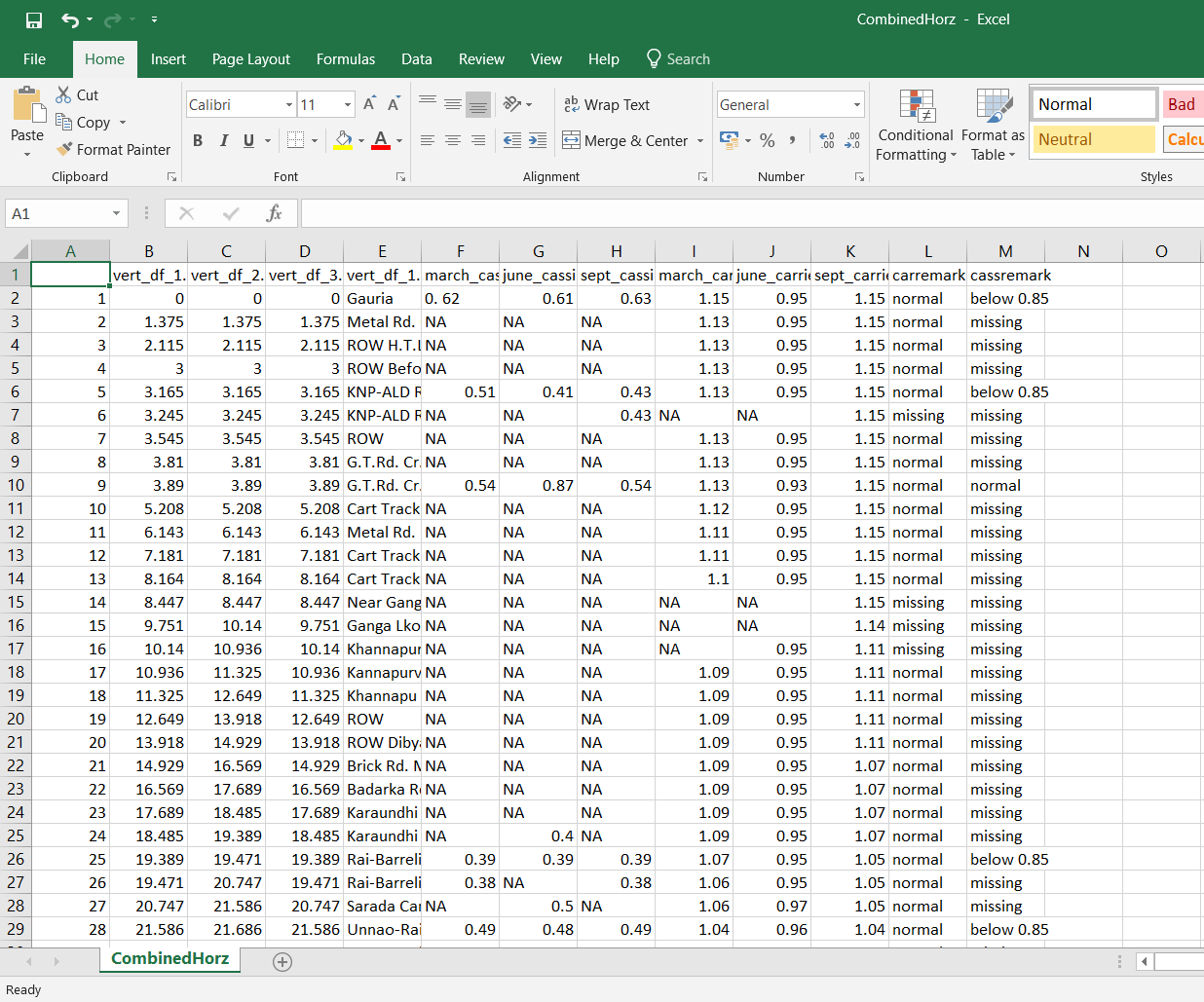
**Parameters Available:**

* Carrier OFF Potential
* Power Availability
* Current Density
* CIPL Reports (Close Interval Potential Logging)
* Circuit Resistance
* Casing ON Potential

**Data Cleaning:**

The data provided to us was messy. There was no particular order or format of recording the data and it varied from chainage to chainage. So, first and the longest part of our analysis was data cleaning.

We wrote scripts in R for different format of input data. We had to take special precautions for missing values. In the end it was decided that some of the missing values were replaced with median and other with NULL.



**Data Mapping:**

In computing and data management, data mapping is the process of creating data element mappings between two distinct data models. Data mapping is used as the first step for a wide variety of data integration tasks, including:

• Data transformation or data mediation between a data source and a destination

• Identification of data relationships as part of data lineage analysis

• Discovery of hidden sensitive data such as the last four digits of a social security number hidden in another user id as part of a data masking or de-identification project

• Consolidation of multiple databases into a single database and identifying redundant columns of data for consolidation or elimination

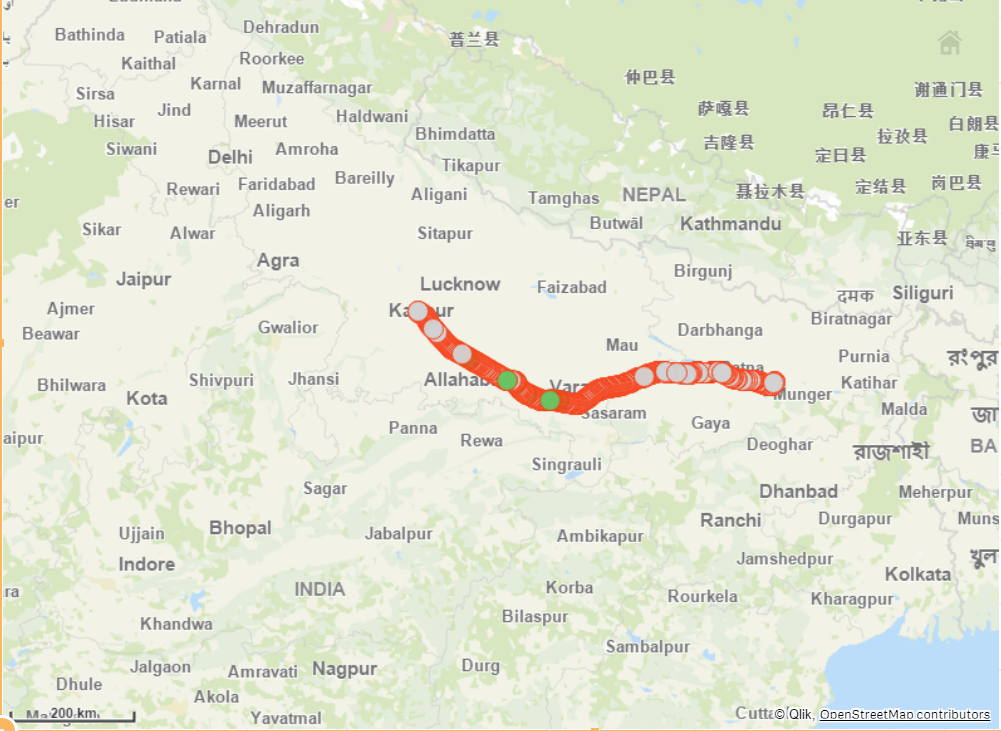
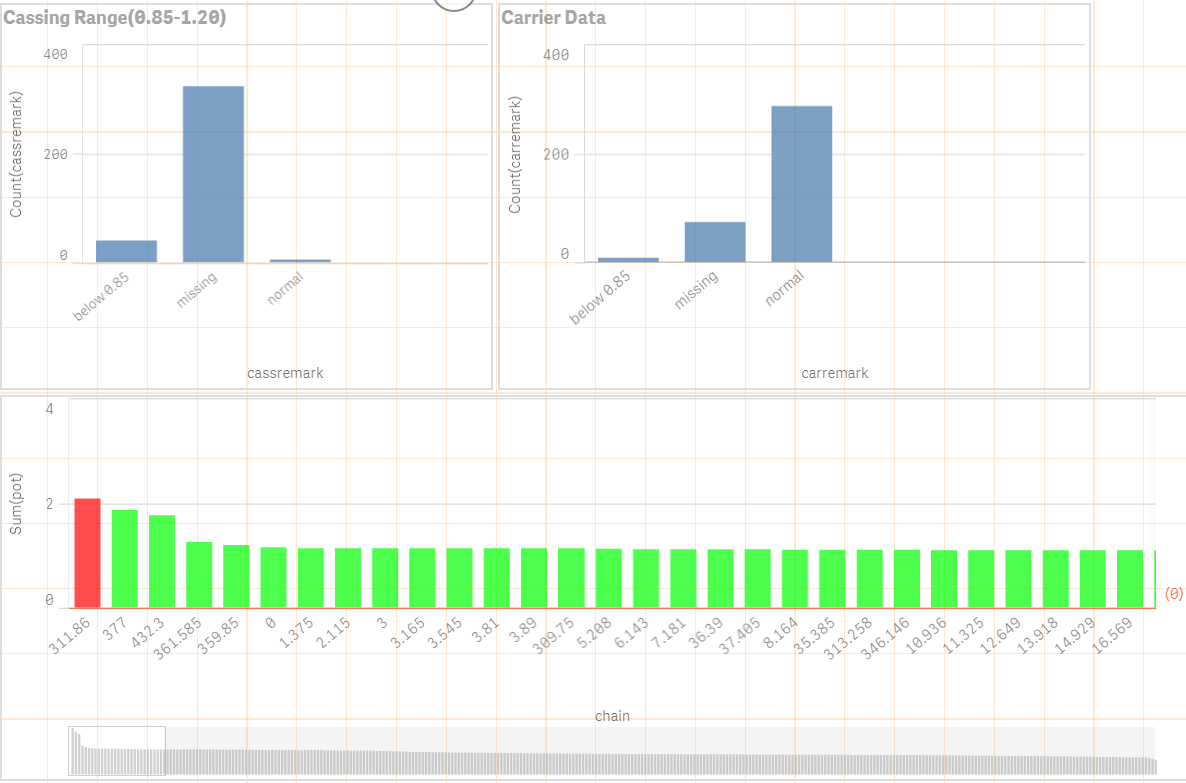
We worked on the four region of pipeline that is south region pipeline, eastern region pipeline, western region pipeline and northern region pipeline where we do the data cleaning also putting all required data together and eliminating all unnecessary data from the parent data sheet.

So, once cleaned we merged data of various regions using mapping on Qlik Sense.

**Data Visualization:**

Data Visualization is a very important part of Data analytics pipeline. Not only data visualization helps in better understanding the data and detecting outliers, in today’s world it has also become a important tool for data scientists in order to explain their thought processes and suggestions to their superiors and clients.

We used R for data visualization when we had raw data and then we used QLIK SENSE once the data was cleaned and merged.



**Machine Learning:**

Machine learning, a branch of artificial intelligence, was originally employed to develop techniques to enable computers to learn. Today, since it includes a number of advanced statistical methods for regression and classification, it finds application in a wide variety of fields including medical diagnostics, credit card fraud detection, face and speech recognition and analysis of the stock market. In certain applications it is sufficient to directly predict the dependent variable without focusing on the underlying relationships between variables. In other cases, the underlying relationships can be very complex and the mathematical form of the dependencies unknown. For such cases, machine learning techniques emulate human cognition and learn from training examples to predict future events.

Our first choice was to use Neural Nets as they give best results with least amount of hyperparameter tuning and input data augmentations. But, after several failed attempts we decided to look for other alternatives. The main reason for this was lack of sufficient training data.

After some research we came upon Support Vector Machines. In Brief, A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane. In other words, given labelled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples. In two-dimensional space this hyperplane is a line dividing a plane in two parts where in each class lay in either side.

Using SVM we got accuracy around 60% in most cases. After further hyperparameter tuning we got accuracy up to 73%.

**CONCLUSION:**

The 73% accuracy we were able to get on training set was sub-par and it was even worse on test data set. We were not able to improve it any further as our training period was over.

The main issue we encountered was lack of any standard method or format for recording and storing the data. This incre3ased the time it took for data cleaning and processing by many folds as we had to write scripts for every possible format and this made impossible to automate the entire analytics pipeline. Much of the data was in pdf portion which we were unable to process.

In the end it was decided that standard format will be decided and a common database will be created where all the offices will upload data in standard format. This will help gather more and accurate data, which in turn will improve accuracy of ML algorithms.

**References:**

[1.] https://www.qlik.com/us/products/qlik-sense

[2.] https://www.r-project.org/

[3.] https://products.office.com/en-in/excel

[4.] https://stackoverflow.com

**Plagiarism Check :**

