

AVANISH RAJ

Phone: +918789646689 Email: rajavanish68@gmail.com

[linkedin](#)

[github](#)

[kaggle](#)

"Data Scientist seeking opportunities in the field of Data Science to drive strategic business decisions through impactful insights and data-driven decision making."

SKILLS

- Programming: Python, SQL
- Framework: Numpy, Pandas, Sci-kit Learn, Tensorflow
- ML algorithms: Linear/Logistic Regression, SVM, Naive-Bayes, Bagging, Boosting, K-Means, PCA
- Data Science Skills : Statistics, Data Cleaning, EDA, Feature Selection, Hyperparameter Tuning, Model Development
- Data Visualization: Matplotlib, Seaborn, power BI

EXPERIENCE(3.2 years)

April 2021 - Present

Data Scientist • **JSW Steel Limited**

- Analysis of various operating parameters to have a better control over the pelletizing process.
- Along with this, some quality improvement projects have been undertaken in which **machine learning model** built.

• Projects-

1. Multivariate Regression of TI(Tumbler Index) for cost optimization.

- Increase in TI(above 94%) of pellet results in lesser waste generation and further cost reduction.
- Conducted rigorous experimentation to determine the impact of 35 variables like particle size fraction, mill feed rate, cyclone parameters, mixed gas calorific value on the value of TI, resulting in actionable insights.
- Leveraged the power of **matplotlib** and **seaborn** libraries to perform **Exploratory Data Analysis**, visualizing data through box plots, distribution plots, and count plots, enabling data-driven insights.
- Performed **Hypothesis Testing** to determine if null hypothesis is rejected or not based on **p-value**.
- Linear regression model developed and all the 5 important **assumptions of linear regression** have been tested.
- Best fit model predicted out of 4 models (**Lasso**, **Linear Regression**, **Decision Tree** and **XGBoost**).
- To find the best fit model, **GridSearchCV** has been used which also served the purpose of **hyperparameter tuning**.
- Finally, these features are controlled accordingly, resulting in high TI due to **waste reduction by 6%** which sums up **cost saving of 400 cr/year**.

2. Multiclass Classification of CCS(Cold crushing strength) as **Low(<225)**, **Medium(225-250)** and **High(>250)**.

- Here, CCS is a **label data** which depends on 21 **features** such as pellet mean size, FeO content, MgO content, Blaine no., BTT, blended LOI, magnetic property, etc.
- Performed **EDA** using **Histogram** and **Box plot** and further, **Distribution Analysis** has been carried out using **Stats** package under **Scipy**. **PP plot**, **QQ plot** and **Density plot** is plotted.
- After that, **Box- Cox Transformation** is applied to make the distribution of the variable into a more Gaussian(normal) distribution.
- To check for **multicollinearity**, Column having variance inflation factor(**VIF**) score (>10) removed.
- **Classifier algorithms** such as **XGBoost**, **Random Forest**, **logistic regression**, etc. has been applied.
- **Classification report** is created to get the **Accuracy**, **Precision**, **Recall** and **F1 Score**.
- To improve the performance furthermore, **Target Class distribution** is balanced using **SMOTE- TOMEK** which is a combination of **oversampling** and **undersampling** techniques. And then, above steps are repeated accordingly.

3. Clustering of fritted pellet formation.

- This is the case of **unsupervised learning** since dataset contain only features but not label.
- Clusters are identified using 9 features like firing zone temperature, DR Main temperature, mixed gas pressure, carbon addition, Bentonite addition, Cooling Zone temp, machine speed, bed height, etc.
- 2 Clustering algorithms, **K-Means** and **DBSCAN**, have been considered to identify number of clusters.
- To find optimal number of **K** in K-means, along with the business inputs, statistical methods like **elbow curve** has been deployed which plots '**Within Cluster Sum of Squared Distances**' for every value of k.
- **Silhouette score** has been considered to measure the accuracy of models.
- Obtained 5 clusters labelled as 'very low', 'low', 'medium', 'high' and 'severe' and all the five stages of fritted pellet formation have been analysed further in detail.

August 2020- April 2021

Graduate Engineer Trainee • **Shri Balaji Industrial Products Ltd.**

- Contributed as a design engineer to develop intricate castings using **Magma** simulation software.
- Acquired the skill of people management.

EDUCATION

- PG Diploma in Data Science, National Institute of Advanced Manufacturing Technology, Ranchi 2021
- B.Tech, Metallurgy and Materials Engineering, NIAMT Ranchi 2020