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."

EDUCATION

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

EXPERIENCE(3.5+ years)

April 2021 - Present

Data Scientist • JSW Group

- 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. **Classification** of **pellet cars** as defective or normal using **Computer Vision.**

- Gathered a dataset of images from camera placed strategically along the pellet cars.
- Preprocessed the images by **normalizing pixel values** to ensure uniformity and enhance model performance.
- **Data Augmentation** technique like rotation, flipping and contrast has been used to increase model robustness.
- Designed the architecture of **Convolutional Neural Network(CNN)** model to serve the purpose of classification.
- Deployed the model to run in real time during plant operation and trigger an alert when defective car is identified.

2. **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.
- **Linear regression** model developed and all the 5 important **assumptions of linear regression** have been tested. Other models developed including **Lasso, Random Forest, XGBoost**, etc.
- **Randomized SearchCV** has been used for the purpose of **hyperparameter tuning**.
- Finally, these features are controlled accordingly, resulting in high TI due to **waste reduction by 6%**

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

- This project aimed to deliver the right quality of pellet to customer.
- 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.
- After that, **Box- Cox Transformation** is applied to make the distribution of the variable into a more normal distribution.
- **Classifier algorithms** such as, **Random Forest, logistic regression, SVM**, etc. has been applied**.**
- Created **Classification report** to get the **Accuracy, Precision, Recall** and **F1 Score.**
- To improve the performance furthermore, **Target Class distribution** balanced using **SMOTE- TOMEK.** And then, above steps are repeated accordingly.

4. **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.
- 3 Clustering algorithms including **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 curv**e has been employed which plots '**Within Cluster Sum of Squared Distances**' for every value of k.
- **Silhouette score** has been considered to measure the accuracy of models.

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.

SKILLS

- Programming: Python, SQL
- Framework: Sci-kit Learn, Tensorflow, Keras, OpenCV, Numpy, Pandas, Matplotlib, Seaborn
- ML/DL algorithms: Linear/Logistic Regression, SVM, Naive-Bayes, Bagging, Boosting, K-Means, PCA, Neural Networks
- Data Science Skills : Statistics, EDA, Feature Engineering, Hyperparameter Tuning, Model Development
- Tools: GIT, power BI, Advanced Excel