**Industrial Copper Modeling**

**Introduction:**

The Industrial Copper Modeling project focuses on predicting the selling price and status (won or lost) in the industrial copper market using machine learning regression and classification algorithms. By exploring the dataset, performing data cleaning and preprocessing, and applying various machine learning techniques, we aim to develop models that can accurately predict the selling price and status in the copper market.

**Requirements:**

To run this project, the following libraries are needed:

- NumPy: A library for numerical computations in Python.

- Pandas: A library for data manipulation and analysis.

- Scikit-learn: A machine learning library that provides various regression and classification algorithms.

- Matplotlib: A plotting library for creating visualizations.

- Seaborn: A data visualization library built on top of Matplotlib.

Make sure these libraries are installed in your Python environment before running the project

**Dataset:**

The dataset used for this analysis contains information about industrial copper transactions, including variables such as selling price, quantities, and status (won or lost). It provides a comprehensive view of the copper market and factors that influence the outcomes of transactions.

**Project Learnings**

The main learnings from this project are as follows:

1. **Exploring Skewness and Outliers**: Analyse the distribution of variables in the dataset and identify skewness and outliers. This step helps in understanding the data quality and potential issues that may affect the model performance.

2. **Data Transformation and Cleaning**: Transform the data into a suitable format for analysis and perform necessary cleaning steps. This includes handling missing values, encoding categorical variables, and scaling numerical features.

3. **Machine Learning Regression Algorithms**: Apply various machine learning regression algorithms to predict the selling price of industrial copper. Compare the performance of algorithms such as linear regression, decision trees, random forests, or gradient boosting.

Regression model details:

The copper industry deals with less complex data related to sales and pricing. However, this data may suffer from issues such as skewness and noisy data, which can affect the accuracy of manual predictions. Dealing with these challenges manually can be time-consuming and may not result in optimal pricing decisions. A machine learning regression model can address these issues by utilizing advanced techniques such as data normalization, outlier detection and handling, handling data in wrong format, identifying the distribution of features, and leveraging tree-based models, specifically the decision tree algorithm.

4. **Machine Learning Classification Algorithms**: Apply different machine learning classification algorithms to predict the status (won or lost) of copper transactions. Explore algorithms such as logistic regression, support vector machines, or random forests to classify the outcomes.

Classification model details:

Another area where the copper industry faces challenges is in capturing the leads. A lead classification model is a system for evaluating and classifying leads based on how likely they are to become a customer. You can use the STATUS variable with WON being considered as Success and LOST being considered as Failure and remove data points other than WON, LOST STATUS values.

5. **Evaluation and Model Selection**: Evaluate the performance of regression and classification models using appropriate metrics such as mean squared error (MSE), accuracy, precision, and recall. Select the best-performing models based on these metrics.

**Methodology:**

1. **Data Loading**: Load the industrial copper dataset into the code using pandas library. Perform initial data exploration to understand the structure and content of the dataset.

2. **Data Cleaning and Preprocessing**: Handle missing values, remove outliers if necessary, and perform necessary data transformations such as encoding categorical variables. This step ensures the data is in a suitable format for analysis.

3. **Exploratory Data Analysis (EDA)**: Use pandas, matplotlib, and seaborn libraries to explore the dataset. Analyze different variables, their distributions, and relationships. Generate visualizations such as histograms, scatter plots, or box plots to gain insights into the data.

4. **Machine Learning Regression**: Apply various machine learning regression algorithms to predict the selling price of industrial copper. Split the dataset into training and testing sets, train the models, and evaluate their performance using metrics such as mean squared error (MSE).

5. **Machine Learning Classification**: Apply different machine learning classification algorithms to predict the status (won or lost) of copper transactions. Split the dataset into training and testing sets, train the models, and evaluate their performance using metrics such as accuracy, precision, and recall.

6. **Documentation**: Prepare a comprehensive documentation summarizing the steps involved in the analysis, including the preprocessing techniques, machine learning algorithms used, and their performance. Include visualizations and interpretations to effectively communicate the results.

**Conclusion:**

The Industrial Copper Modelling project aims to predict the selling price and status in the industrial copper market using machine learning techniques

Video Link:

Githublink: https://github.com/kanis11/IndustrialCopperModeling

Video link: https://www.linkedin.com/feed/update/urn:li:activity:7236606735736586240/