**Scrap Probability Prediction**

**Problem Statement:**

Develop a machine learning model to predict the probability of a product being scrapped. Thereby taking preventive actions by identifying the products with high Scrap Probability and identifying the likelihood of products being scrapped for planning better productive schedules, allocate resources more efficiently and improve quality control.

**Description:**

This Model focuses on creating a predictive model to estimate the likelihood of products being scrapped based on historical warehouse data. In warehouse management, scrapping products due to defects, damage, or expiration is a major concern. High scrap rates can lead to increased costs, wasted resources, and inefficiencies in inventory management. By predicting which products are likely to be scrapped, warehouses can proactively address these issues, reducing waste and enhancing operational efficiency. The ultimate goal is to minimize scrap rates and enhance the overall effectiveness of warehouse operations. Identifying and mitigating risks associated with high scrap probability products, ensuring that potential issues are addressed before they escalate.

**Business Value:**

Developing a machine learning model to predict the probability of products being scrapped in warehouse management offers significant business value across multiple dimensions.

* **Waste Minimization**: By predicting and preventing high-probability scrap incidents, the warehouse can significantly reduce waste, leading to lower disposal and handling costs.
* **Financial Performance**: Optimizes costs and resource utilization, improving financial performance.
* **Improved Disposal Planning:** Predictive scrap probability allows warehouses to plan disposal activities more effectively by anticipating which products are likely to be scrapped in the future.

**Dataset Research:**

The dataset is created by assuming the appropriate features involved in this process. Dummy dataset is created to train the model.

We believe that CRISP-ML Q (Cross Industry Standard Process for Machine Learning with Quality Assurance) will ease obtaining efficient results as it follows step-by-step process. Starting by understanding the business and data, preprocessing the data then model building it and finally evaluate the model to make sure it’s performing well.

**`CRISP-ML(Q)`** process model describes six phases:

1. Business and Data Understanding
2. Data Preparation
3. Model Building
4. Model Evaluation
5. Deployment
6. Monitoring and Maintenance

**Business and Data Understanding**

**Objective(s):** Minimize the Number of Scrapped products.

**Constraint(s):** Maximize the Business revenue by reducing the Scrap losses.

**Success Criteria:**

**Business Success Criteria**: Reduce total Scrap products by 25% within the first year of implementation.

**Machine Learning Success Criteria**: Achieve good performance metrics with R-Squared value greater than 0.85

**Economic Success Criteria**: Boost overall sales revenue by reducing the likelihood of products to be Scrapped

**Data Collection/Description**  
**Data:** This is the dummy data created for the testing purpose for the model building with some relating columns.

**Data Dictionary:**

* Dataset contains 8 columns/features
* Dataset contains 100 records

**Description:**

* **Item\_ID** - Unique identifier for each item
* **Item\_Name** - Item Name or product Name
* **Item\_Category** - Category name into which category the following product/item falls.
* **Scrap\_Description** - Reason for the Scrapping of the product/item
* **Quantity** - Number of Quantity Scrapped
* **Disposal\_Method** - Type of Disposing the Scrapped products
* **Cost** - Cost of the product/item for the given quantity
* **Scrap\_Probability** - Probability of an item more likely to be Scrapped

**Exploratory Data Analysis**

As part or EDA, we used D-Tale and SweetViz, which are an Auto EDA libraries which gives the complete insights on data. We can able to see Descriptive Statistics and basic analytics of the data.

**Data Preprocessing/Data Wrangling/Data Cleaning**

Data Preprocessing involves removing of unwanted columns, checking for missing values, handling duplicates, Dummy variables creation for categorical values and splitting the dataset.

**Model Building/Modelling**

**Linear Regression:** Linear Regression is the practical and reliable choice to determine the relation between the dependent and independent variables. It aims to predict the value of the dependent variable based on the values of the independent variable(s). It forms the basis for more complex regression techniques and is widely applicable across various domains.

**Training:** Data training is achieved by splitting the data into train and test dataset. Scrap \_Probability is the Dependent variable, which needs to predict.

**Metrics:**

* **Mean absolute error**: 0.0224
* **Mean squared error**: 0.0007
* **R-Squared error**: 0.93

**Conclusion:**

The development of a machine learning model to predict the probability of products being scrapped in warehouse management presents a significant opportunity for enhancing operational efficiency, reducing costs, and improving overall business performance. By leveraging techniques such as linear regression, warehouses can gain valuable insights into the factors contributing to product scrapping and take proactive measures to mitigate risks and optimize resource allocation.