**Class #23: Feature Engineering with Scikit-learn**

**1. Understanding feature scaling.**

**2. Hands-on with feature scaling techniques.**

**3. Introduction to Principal Component Analysis (PCA).**

**4. Working with PCA in real-world examples.**

**5. Practical exercises with label encoding.**

**6. Hands-on with ordinal encoding.**

**7. Working with one-hot encoding through practical examples.**

**8. Removing outliers in real-world datasets. Assignment #23:**

Sclearn scalling করে ।

**1. Understanding feature scaling.**

কেন করবো ? কারণ, যদি সবগুলা ডেটার মধ্যে ১টা ডেটার মধ্যে বেশি পার্থক্য হয়ে যায়, তখন scaling করতে হয় ।

Min max scaller

|  |
| --- |
| import pandas as pd  df = pd.read\_excel("car\_sales\_data.xlsx")  from sklearn.preprocessing import MinMaxScaler  scaler = MinMaxScaler()  # Apply scaling to the numerical columns (assuming all columns are numerical)  df\_scaled = scaler.fit\_transform(df)  # Show the scaled dataframe  print(df\_scaled) |

|  |
| --- |
| import pandas as pd df = pd.read\_excel("car\_sales\_data.xlsx") from sklearn.preprocessing import StandardScaler object1 = StandardScaler() a = object1.fit\_transform(df) print(a) |

**3. Introduction to Principal Component Analysis (PCA).**

ছবি হচ্ছে Array. অপ্রয়োজনীয় অংশ দূর করে দেয় । PCA করলে scaling করতে হবে এবং accuracy কমবে ।

Remove ourleyer

Heatmap

|  |
| --- |
| import matplotlib.pyplot as plt import pandas as pd import seaborn as sns import pandas as pd df = sns.load\_dataset("iris") corr= df.corr(numeric\_only=True) sns.heatmap(corr, annot=True, cmap='coolwarm') plt.show() |

Co-relation

Distribution

Classitaring

Data missing