

# Task 4: Feature Encoding & Scaling

## AI & ML Internship Report

### 1. Objective

The objective of this task is to understand and apply feature engineering techniques such as categorical feature encoding and numerical feature scaling to prepare data for machine learning models.

### 2. Dataset

The Adult Income Dataset is used in this task. It contains demographic and employment-related information such as age, education, occupation, working hours, and income level.

### 3. Tools & Technologies

- 1 Python
- 2 Pandas
- 3 NumPy
- 4 Scikit-learn
- 5 Jupyter Notebook

### 4. Methodology

First, the dataset was loaded and analyzed to identify categorical and numerical features. Label Encoding was applied to ordered categorical features such as the income column. One-Hot Encoding was applied to non-ordered categorical features like occupation and workclass. Numerical features were scaled using StandardScaler to standardize the data.

### 5. Feature Encoding

Label Encoding converts categorical values into numeric labels where order matters. One-Hot Encoding creates separate binary columns for each category where no order exists.

### 6. Feature Scaling

Feature scaling ensures that numerical features contribute equally to model training. StandardScaler was used to transform features so that they have zero mean and unit variance.

### 7. Importance of Scaling

Scaling improves the performance and convergence speed of machine learning algorithms. Algorithms such as Logistic Regression, KNN, SVM, K-Means, and Neural Networks require scaled data.

### 8. Deliverables

- 1 Preprocessed Dataset
- 2 Jupyter Notebook
- 3 GitHub Repository
- 4 Project Report

## 9. Conclusion

This task provided hands-on experience with feature engineering techniques. The dataset is now fully numerical, scaled, and ready for machine learning model development.