

#Feature Encoding & Scaling

##Project Title

Feature Encoding & Scaling on Adult Income Dataset

##Project Description

This project focuses on preprocessing the Adult Income dataset by applying feature encoding and scaling techniques. The goal is to convert categorical data into numerical format and scale numerical features so that the dataset becomes suitable for machine learning algorithms.

Feature engineering is an important step in the ML pipeline, and this task helps in understanding how raw data is transformed into model-ready data.

##Objectives

- Identify categorical and numerical features
- Apply Label Encoding where order exists
- Apply One-Hot Encoding where no order exists
- Scale numerical features using StandardScaler
- Compare dataset before and after scaling
- Understand the impact of scaling on ML algorithms
- Save the preprocessed dataset

##Dataset Used

Adult Income Dataset

Target:

- Income

Types of Features:

- Categorical: Workclass, Education, Marital Status, Occupation, etc.
- Numerical: Age, Hours-per-week, Capital-gain, Capital-loss, etc.

##Tools & Libraries Used

- Python
- Pandas
- NumPy
- Scikit-learn
- Jupyter Notebook

##Preprocessing Steps

1. Loaded the dataset and explored its structure
2. Identified categorical and numerical columns
3. Applied Label Encoding to ordinal features
4. Applied One-Hot Encoding to nominal features
5. Scaled numerical features using StandardScaler

6. Compared data ranges before and after scaling

7. Saved the final processed dataset

##Key Learnings

- Encoding converts text data into numerical format
- One-Hot Encoding avoids false ordering
- Label Encoding is useful when order exists
- Scaling ensures all features contribute equally
- Scaling improves performance of distance-based algorithms

##Files in this Repository

- adult.csv – Original dataset
- processed_adult.csv – Preprocessed dataset
- Task4_Feature_Encoding_Scaling.ipynb – Notebook with all steps
- README.md – Project documentation

##How to Run

1. Clone this repository
2. Open the Jupyter Notebook
3. Run all cells step by step
4. View the processed dataset and outputs

##Conclusion

This task helped in understanding the importance of feature engineering in machine learning. Encoding and scaling are essential steps that improve model performance and make the data suitable for training ML models.