

Adult Income Prediction

Design a **Machine Learning Model** on the data-set given below:

Download the dataset from this link. [CLICK HERE TO DOWNLOAD DATASET](#)

The detailed description about the dataset can be found here. [Dataset Description](#)

For your reference, kindly have a look at this jupyter notebook. [CLICK HERE](#)

If you are facing any difficulty, follow the steps mentioned below:

Step - 1 - Getting things ready for the project

- Import the required libraries
- Load the data
- Discover and handle missing values (Similar to previous assignment)

Step - 2 - Data Preparation

- Consider 'income' as the target variable (:P)
- Remove unnecessary columns (For eg: 'fnlwgt')
- Standardizing numerical data
- Encoding categorical features
- Train Test Split

Step - 3 - Train the model on Training Data

- Consider 'income' as the target variable (-_-)
- Identify the TASK in hand and choose the appropriate algorithm for the task.
- Since task is Classification, use Logistic Regression

Step - 4 - Predict on Test Data

Step - 5 - Evaluate each Model

- Use evaluation metrics like Accuracy, Confusion Metric, Precision and Recall.
- Print classification report as shown in the figure below

```
In [75]: print(metrics.classification_report(y_test, y_test_pred))
```

Number of features = 94

	precision	recall	f1-score	support
0	0.72	0.60	0.65	3293
1	0.88	0.93	0.90	10192
accuracy			0.85	13485
macro avg	0.80	0.76	0.78	13485
weighted avg	0.84	0.85	0.84	13485

STEP 3: Take a screen shot (Snipping Tool for windows) and post it on the SLACK 'JIET-machine_learning' channel. You can also post it on LinkedIn to showcase your work.

Step - 6 - Use RFE to extract the best features.

- Build Logistic regression using the best features
- Test and Evaluate the new model

Step - 7 - Use PCA to extract the best features and build the model again.

- Since task is Classification, use Logistic Regression, Knn, DT and SVM (Create an individual model for each of the algorithm)
- Test and Evaluate for each Model.
- Print the Classification Report as shown in Step 5.