Session 6: Building an Income Classification Model

Task: Build a classification model for predicting the income using the Adult Census Income

Dataset.

• Load the dataset

• Check for null values and ? in any columns and handle those values. Check the distribution of target variable income and identify if the dataset is balanced.

• Perform the following Univariate analysis

• Create a barplot for column income

• Create a distribution plot for column age

• Create a barplot for column education

• Create a barplot for Years of Education. Use column education.num

• Create a pie chart for Marital status. Use column marital.status

• Perform the following Bivariate analysis

• Create a countplot of income across columns age, education, marital status, sex

• Draw a heatmap of data correlation and find out the columns to which income is highly

correlated

• Prepare the dataset for modeling

• Label encode all the categorical columns

• Prepare independent variables X and dependent variable Y (Income).

• Perform feature scaling using StandardScaler and fix the imbalance in the dataset

using any one of the techniques like SMOTE or RandomOverSampler

• Perform a train test split in the ratio 80:20 and random\_state 42.

• Perform Data Modeling

• Train Logistic Regression Model, KNN Classifier Model, SVM Classifier, Naive Bayes

Classifier, Decision Tree Classifier and Random Forest Classifier

• Perform model evaluation on Accuracy and F1 score and identify the best model