- Term Project – Machine Learning Model to predict Adult income

- Team Members –
    1. Nikhil Thakur
    2. Meet Naik
    3. Krishma
    4. Avanee Shah

Title: Predicting Income Levels using Machine Learning on the Adult Data Set

Introduction:

The Adult Data Set is a widely used data set for classification tasks in machine learning. It contains demographic and financial information of individuals from various backgrounds, which includes both continuous and categorical variables. The objective of this project is to build a model that can accurately predict the income level of an individual based on the features provided in the Adult Data Set. This project will use machine learning algorithms to explore the dataset, preprocess the data, train a model, and evaluate the model's performance.

Research Questions:

1. What is the accuracy of the machine learning model in predicting income levels based on the Adult Data Set?
2. Which machine learning algorithm(s) perform best on the Adult Data Set in predicting income levels?
3. What are the key factors that contribute to high income levels according to the model?

Methods:

- The proposed study will use the Adult Data Set to train and test several machine learning algorithms for predicting income levels. The following steps will be performed:

- Data Preprocessing: The data set will be preprocessed to remove any missing values, duplicate data, and irrelevant columns. Categorical variables will be transformed into numerical values using one-hot encoding. Feature scaling will be applied to normalize continuous variables.

- Model Selection: Several machine learning algorithms will be tested to identify the best algorithm for the task. The algorithms to be considered include decision trees, random forests, support vector machines, and logistic regression.

- Model Training: The selected algorithm will be trained on the preprocessed data set using k-fold cross-validation to optimize the model's performance.

- Model Evaluation: The model's performance will be evaluated using various metrics such as accuracy, precision, recall, and F1-score.

- Feature Importance: The trained model will be used to identify the key features that contribute to high income levels.

Expected Outcomes:

The expected outcomes of this project are:

- A machine learning model that accurately predicts income levels based on the Adult Data Set.
- Identification of the best machine learning algorithm for the task.
- Identification of the key features that contribute to high income levels.

Implications:
The proposed study will have several implications for future research and practical applications. This study will contribute to the development of machine learning algorithms that can predict income levels accurately. The results of this study may also be useful for policymakers and economists to identify the factors that contribute to high income levels. Finally, the findings of this study may help individuals make informed decisions about their financial future.

Conclusion:
This project aims to develop a machine learning model that can accurately predict income levels based on the Adult Data Set. The proposed study will use a variety of machine learning algorithms, preprocess the data, train and test the model, and evaluate the model's performance. The expected outcomes of this study are a model that accurately predicts income levels, identification of the best machine learning algorithm, and identification of the key features that contribute to high income levels.