**Income Prediction**

**Abstract:**

An individual’s annual income results from various factors. Intuitively, it is influenced by the individual’s education level, age, gender, occupation, and etc.

**Problem Statement:**

Explore the possibility in predicting income level based on the individual’s personal information.

**Dataset Information:**

The dataset contains 16 columns target field is Income where the income is divide into two classes. There are a total of 14 attributes which are the demographics and other features to describe a person

**Variable Description:**

|  |  |
| --- | --- |
| **Column** | **Description** |
| work-class | Private, Self-emp-not-inc, Self-emp-inc, Federal-gov, Local-gov, State-gov, Without-pay, Never-worked. |
| fnlwgt | The weights on the CPS files are controlled to independent estimates of the civilian noninstitutional population of the US. These are prepared monthly for us by Population Division here at the Census Bureau. |
| education | Bachelors, Some-college, 11th, HS-grad, Prof-school, Assocacdm, Assoc-voc, 9th, 7th-8th, 12th, Masters, 1st-4th, 10th, Doctorate, 5th-6th, Preschool. |
| marital-status | Married-civ-spouse, Divorced, Never-married, Separated, Widowed, Married-spouse-absent, Married-AF-spouse. |
| occupation | Tech-support, Craft-repair, Other-service, Sales, Execmanagerial, Prof-specialty, Handlers-cleaners, Machine-opinspct, Adm-clerical, Farming-fishing, Transport-moving, Privhouse-serv, Protective-serv, Armed-Forces. |

|  |  |
| --- | --- |
| relationship | Wife, Own-child, Husband, Not-in-family, Other-relative, Unmarried. |
| race | White, Asian-Pac-Islander, Amer-Indian-Eskimo, Other, Black. |
| sex | Female, Male. |
| native-country | Individual’s native country |
| age | Age of an individual |
| capital-gain | Capital gain |
| capital-loss | continuous |
| hours-per-week | Individual’s working hour per week |

**Scope:**

* Explore the data and understand the factors that responsible for a person’s income
* Display important features that affect person’s income
* Find the optimum number of nearest neighbours
* Compare the knn classification model after EDA with the base knn model

**Learning Outcome:**

The students will get a better understanding of how certain parameters play an important role when it comes to an individual’s income and how knn uses the idea of nearest neighbours for classification.