

**CA707 Data Mining Lab**  
**Project Report- Nov 2019**

**Roll No: 205118032**

**Name: KUNDAN THAKARE**

**Title of the Project : Income Classification**

**Dataset Description:**

**Name : Adult Data Set**

**URL : <http://archive.ics.uci.edu/ml/datasets/Adult>**

**No of Records: 48842**

**No of Features: 14**

**List all Features( With One Line Description):**

<b>Feature</b>	<b>Description</b>
<b>AGE</b>	Age of the individual (Numerical)
<b>WORKCLASS</b>	Class of work (Categorical)
<b>Fnlwgt</b>	Final weight of how much population it represents (Numerical)
<b>EDUCATION</b>	Educational qualification of an individual (Categorical)
<b>Education_num:</b>	Numerical representation of educational qualification (Numerical)
<b>MARITAL STATUS</b>	Marital status of the individual (Categorical)
<b>OCCUPATION</b>	Occupation of the individual (Categorical)
<b>RELATIONSHIP</b>	Type of relationship (Categorical)
<b>RACE</b>	Race of an individual.(Categorical)
<b>SEX</b>	Gender of an individual (Categorical)

<b>CAPITAL-GAIN</b>	Capital gain is the profit one earns on the sale of an asset like stocks, bonds or real estate (Numerical)
<b>CAPITAL-LOSS</b>	Capital loss is the loss incurred when a capital asset. (Numerical)
<b>HOURS-PER-WEEK</b>	Average number of working hours per week. (Numerical)
<b>NATIVE-COUNTRY</b>	Country of origin (Categorical)

**Type of problem:** Classification

**Algorithm Used:** Naive Bayesian

**Evaluation Metric:** Out of 48842, 37266 predicted correctly and 11576 predicted wrong

**Accuracy Obtained:**

**Using Libraries: 78.5%**

**Your Algorithm:76.3%**

**Date**

**Signature**