# Problem : To analyze various aspects of an individual and predict what class of income he belongs to (>50k or <=50k) by using census data.

# We have to build a model for classifying whether the income of a particular individual is more or less than 50K annually and determine the accuracy.

Solution :

Since the data predicts two values (>50k or <=50K), this clearly is a classification problem, and we will train the classification models to predict the desired outputs.

1. Age — The age of an individual, this ranges from 17 to 90.

2. Workclass — The class of work to which an individual belongs.

3. Fnlwgt — The weight assigned to the combination of features (an estimate of how many people belong to this set of combination)

4. Education — Highest level of education

From the dataset:

1. Ratio of people earning more than 50K is higher in case Workclass is ‘Self-emp-inc’.

2. People with education level as ‘Masters/Doctorate/Prof-school’ have higher ratios of >50K earning, than <=50K. Bachelors degree also has around 10:7 ratio of <=50K : >50K.

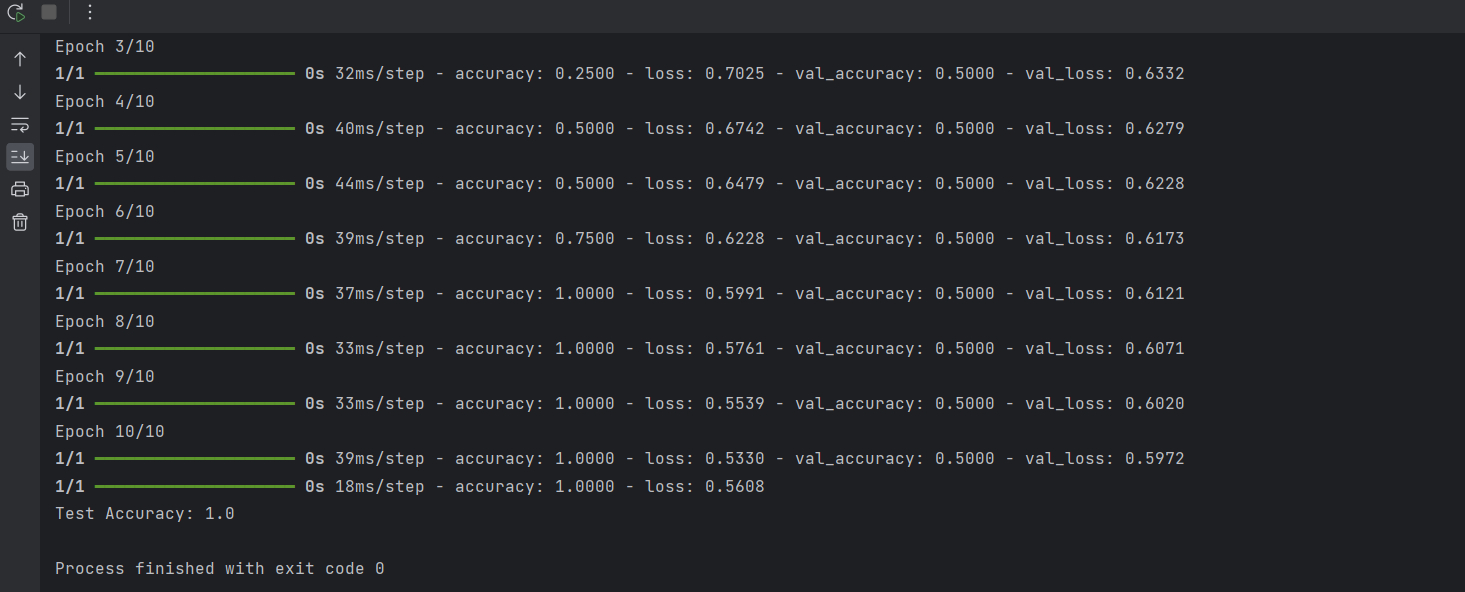
3. From the scatterplot between age, hours\_per\_week and income, we observe that a person needs to be >30 to be earning more than 50K, else needs to work at least 60 hours\_per\_week to earn >50K.

CODE:

<https://github.com/Abinayasen/DEEP-LEARNING>





OUTPUT:  


Important:

Functions:-

Tensorflow- TensorFlow is used to build and train deep learning models as it facilitates the creation of computational graphs

Pandas- It has functions for analyzing, cleaning, exploring, and manipulating data.

Sklearn- Machine Learning in Python · Simple and efficient tools for predictive data analysis

BODY:-

we collect the data as datasets and mark it as dependent and independent variables

df is a variable that uses panda to convert data to frame

le is used to encode labels to the education, income being the dependent variable for the question , it is mapped or split as binary accordingly, where

<=50k = 0

>50k = 1

Let us consider X to represent the independent data frame , i.e age, hours-per,week, education

And y to represent the income

The fit\_transform() method of the StandardScaler, first computes the mean and standard deviation of each feature in X to learn the scaling parameters. By standardizing the features, we ensure that each feature contributes equally to the learning process and prevents features with larger scales from dominating the learning process. X is now transformed to a new array of numerical values.

Next process is the test- train process , here we have assigned test\_size of 0.2 that is 20% of data is collected for testing and the other 80% for training .

By splitting the data into training and testing sets, we ensure that the model is trained on one portion of the data and evaluated on another portion that it hasn't seen during training. This allows us to assess the model's performance on unseen data and detect issues like overfitting.

TENSORFLOW:

Here the process involves

1)building the model

2)compiling the model

3)training the model and

4)evaluating the model

BUILDING:

We use Keras , here the layers of data sets is linealrized to a single layer,

The code consists of two RELU layers and their activation

And converting it to sigmoid function

COMPILATION:

Here the model is compiled , the error is calculated , the solving method for error chosen for this problem is ‘binary\_crossentropy’.

TRAINING:

The model is trained to do 5 iterations back and through the neural network process,

Batch size of 4 defines the number of samples per gradient

Validation split of Specifies the fraction of the training data to be used as validation data. In this case, 20% of the training data will be used for validation during training, as mentioned above.

EVALUATING:

The model is then tested for test accuracy , this model has an accuracy of 0.1