

Detailed Project Report

Adult Census Income Prediction

Ritesh Naik

Darshan Naik

Objective

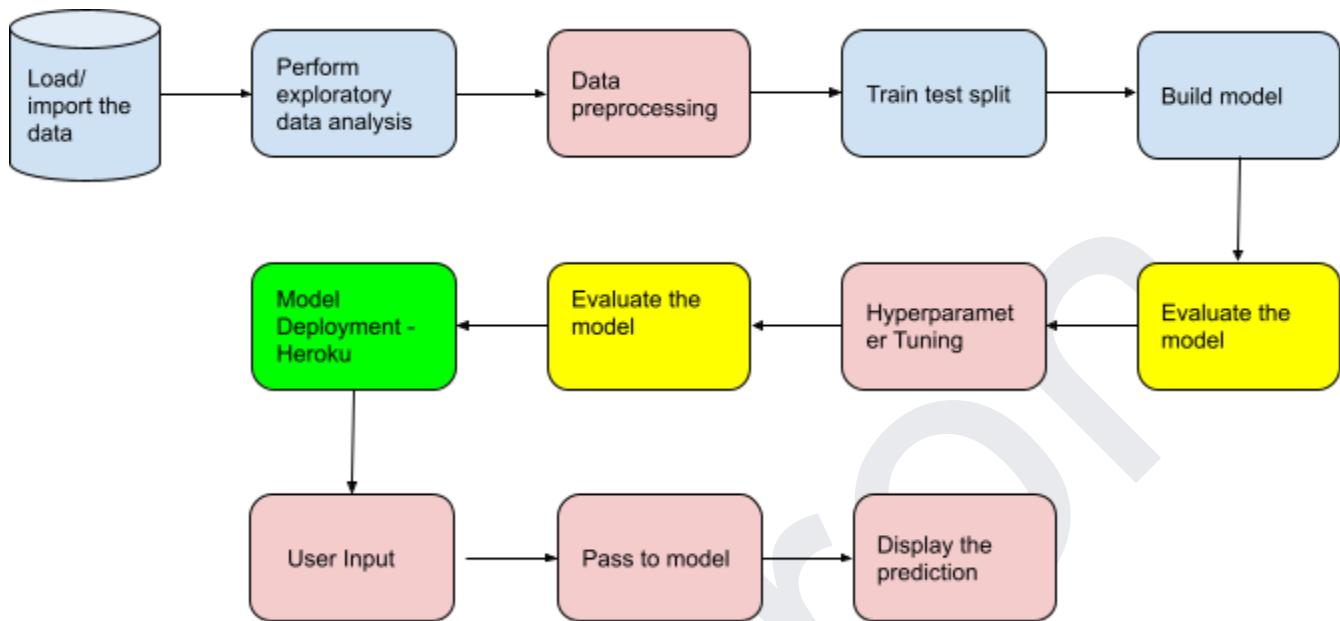
The major goal of this project is to develop an online application called Adult Census Income Prediction that uses classification-based Supervised Machine Learning algorithms to categorize people's incomes as either higher than or equal to \$50,000. Our website makes it possible for users to determine whether or not a person makes more than \$50,000 annually. This application's goal is to identify trends in income data, analyze those trends, and then forecast future outcomes based on ten complex patterns.

With the help of machine learning, we can find the pattern & ultimately find the answer whether a person makes \$ 50K annually or not.

Advantages

- These statistics can be used to understand income trends, more difficult educational requirements, salary in relation to age, etc.
- The main goal of the machine learning-based online application is to identify interesting patterns in the user's income data and then make predictions based on them.
- A person's income can now be maximized and government efficiency improved by finding strategies to predict income. Get a sense of the person's income, whether it is more than \$50,000 or less than \$50,000.
- App can be used to understand the gender gap in income.

Architecture



Collect & Import the data

- ❖ **Collect the data** : Dataset is provided by iNeuron company. Lets download & save it locally.
- ❖ **Import the dataset** : Lets import the data using Pandas library in python
- ❖ **Exploratory Data Analysis** : Lets load & analyze the data.
 - Try to understand the missing values, special characters, total number of rows vs columns present in the dataset.
 - Understand the relationship among the features.

Data Transformation

❖ Data Transformation :

- Remove special characters - Data consist of ("?") , lets replace them with NaN
- Remove unwanted columns - Dropping 'fnlwgt' & 'education-num' columns as they are not necessary

❖ Data Preprocessing:

- Categorical Variable : Let's use OneHotEncoding to deal with categorical values
- Replace “NaN” with mode

Model Selection & Tuning Hyperparameters

- ## ❖ Model Selection Hyperparameter Tuning :
- Out of the following models built, "Decision Tree", "Random Forest Classifier", "XGBoost", "Logistic Regression" , we have selected "XG Boost" as our final model , as this model gave highest accuracy & F1 Score.

Hosting to cloud (Deployment)

❖ Deployment

- We have used Html & CSS for front end development.
- The backend is Flask, however the HTML markup formats that are returned to the user via an HTTP request that are created using Jinja2, a templating language.
- After testing successfully in local environment, we have used Heroku cloud platform to host the application
- We have hosted the app in Heroku,which is a platform as a service (PaaS) that allows users to create, launch, and manage all of their apps in the cloud.

Prediction

- ❖ Prediction

- Here is our app <https://ineuron-salary-pred.herokuapp.com>