

Project 2 Presentation

- Dataset 1 : Adulte income
- Dataset 2 : Crop recommendation



Dataset 1 : Adulte income

Stakeholders/problem we are solving

- Our stakeholders is U.S census agency , The goal of this project is to accurately predict whether or not an adult makes more than 50 000 US Dollars in an year on the basis of the feautres given.

Data description

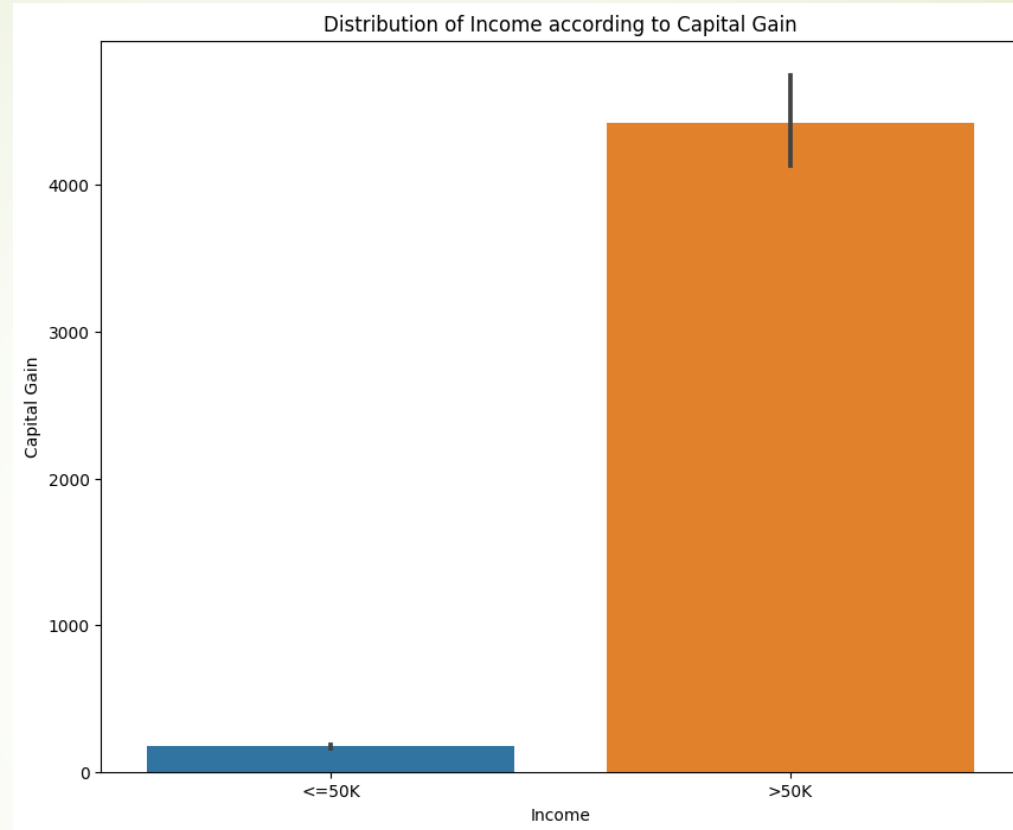
Data contains 48842 rows and contains 15 columns :

Age, Workclass, fnlwgt, education, education-num, marital-status, Occupation, relationship, race, sex, capital-gain, capital-loss, hours-per-week, native-country, salary.

Data contain 6374 duplicated row and no missing value.

Data is imbalanced, and need to be balanced.

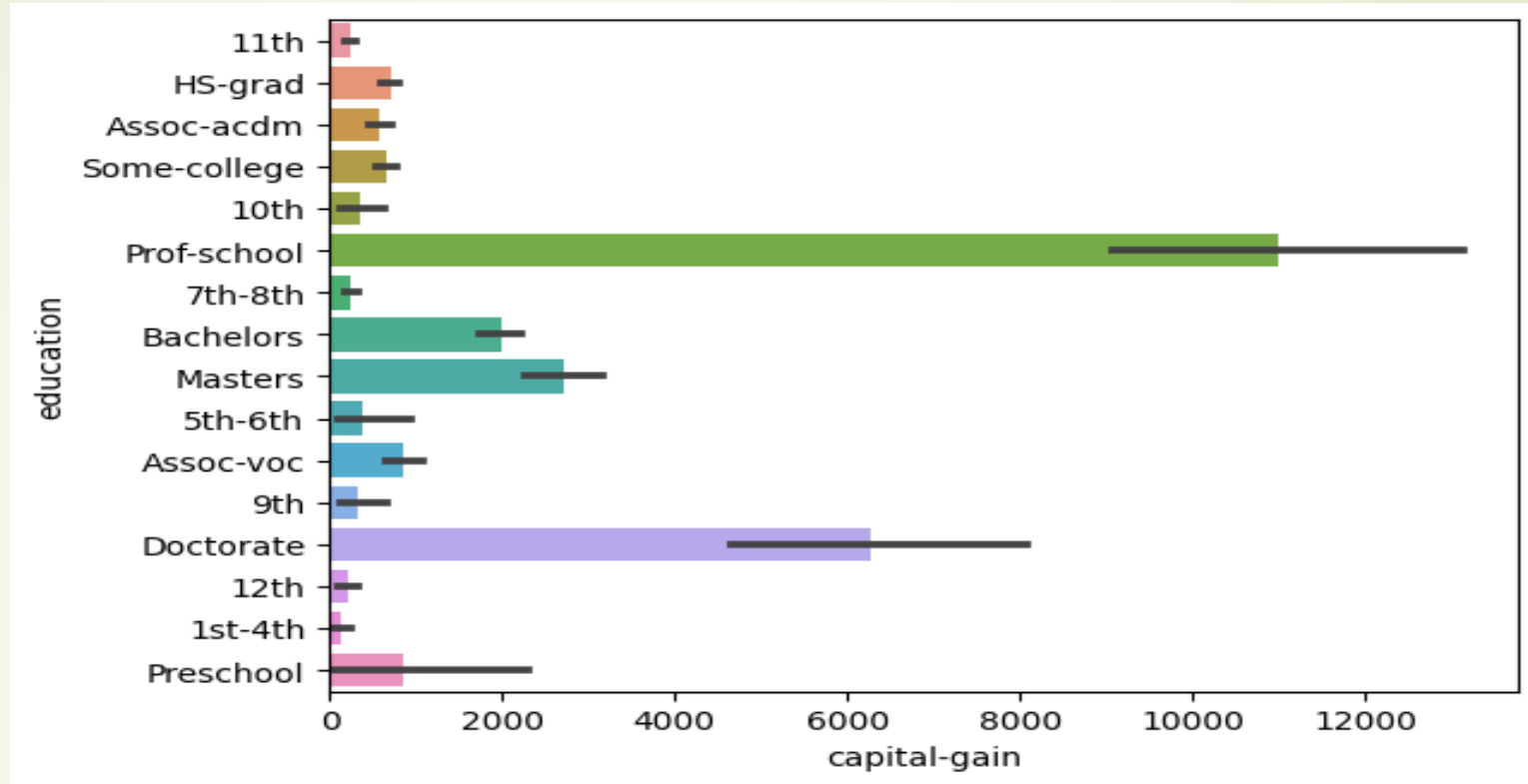
Visual 1



The barplot above represents the distribution of income according to capital gain, this barplot shows us a significant difference between the two classes in terms of the value of capital gain ($>50K$ and $\leq 50K$)

The higher the capital gain, the higher the income too. So we can conclude that the capital gain is a very important factor in determining income

Visual 2



The bar plot above represents the distribution of capital gain according to occupation. This barplot shows us that certain occupations have a high capital gain such as prof-speciality, exec-managerial and sales occupation while the other occupations have corresponding low capital gain.

Since we have already determined previously that capital gain is an important factor in determining income. We reveal the importance of occupation for determining income

Strengths and limitations of your model

High Accuracy (0.84):

This means that it correctly classifies instances with a high degree of accuracy(84%), indicating that it can reliably distinguish between adult who make more than 50K dollars and other who make less .

High Precision (0.87):

This means that when the model predicts an adult make more than 50K dollars it is correct 86% of the time.

High Recall (0.93):

With a recall of 93%, the model is capable of identifying a substantial portion of adults who make more than 50K dollars correctly. This is essential for ensuring that adults who make more than 50K dollars are not missed, as their strengths can be leveraged for other adult to improve their incomes.

High F1-Score (0.90):

The F1-Score of 0.90 is a robust measure of the model's overall effectiveness. It balances precision and recall, indicating that Model can simultaneously provide accurate predictions and capture a significant proportion of adults who make more than 50 K dollars .

Recommendation

- To improve Adults income ,they should :
 - Have a High education level (prof-school and Doctorate)
 - Choose a high paid occupation such as prof-speciality ,exec-managerial and sales occupation,
- The State must also take care of certain occupations with low capital gain (Private house service) and support adults who have such occupation.



➡ **Dataset 2 : Crop recommendation**

Stakeholders/problem we are solving

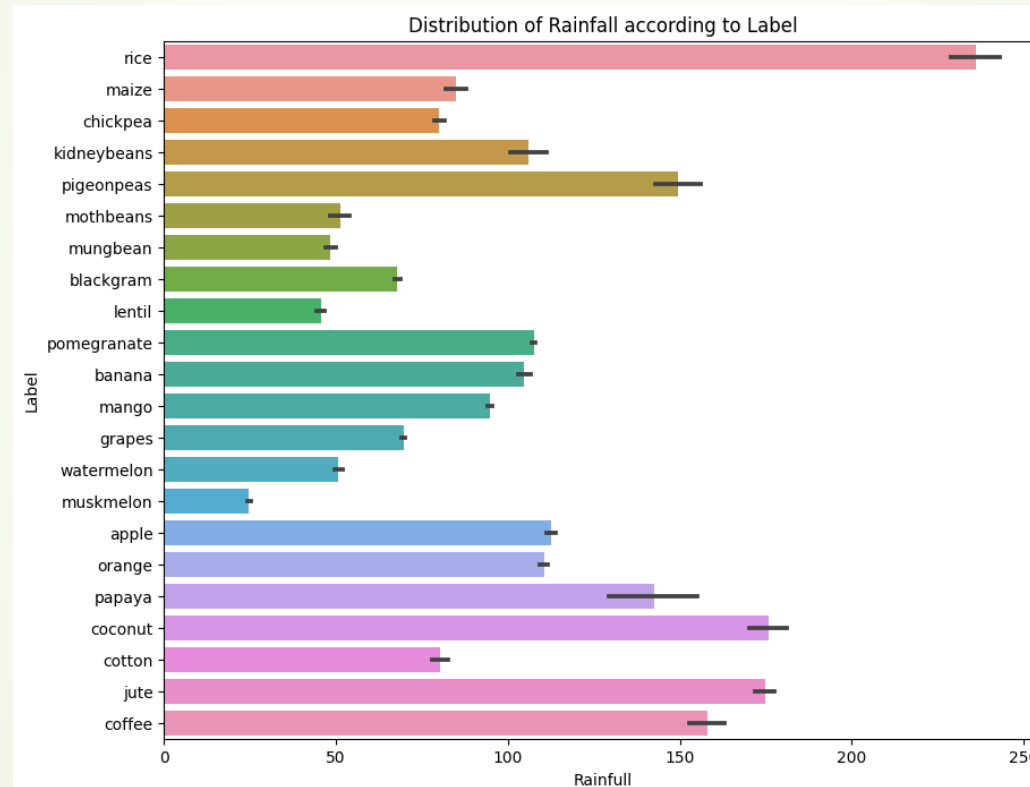
- Our stakeholders are farmer and farm manager , The goal of this project is to build a predictive model to recommend the most suitable crops to grow in a particular farm based on various parameters.

Data description

This dataset contains information on the levels of nitrogen, phosphorus, and potassium in soil, as well as temperature, humidity, pH, and rainfall, and their impact on the growth of crops. Data contain 8 columns and 2200 rows ,it is pretty clear without duplicated row or missing values and pretty balanced.

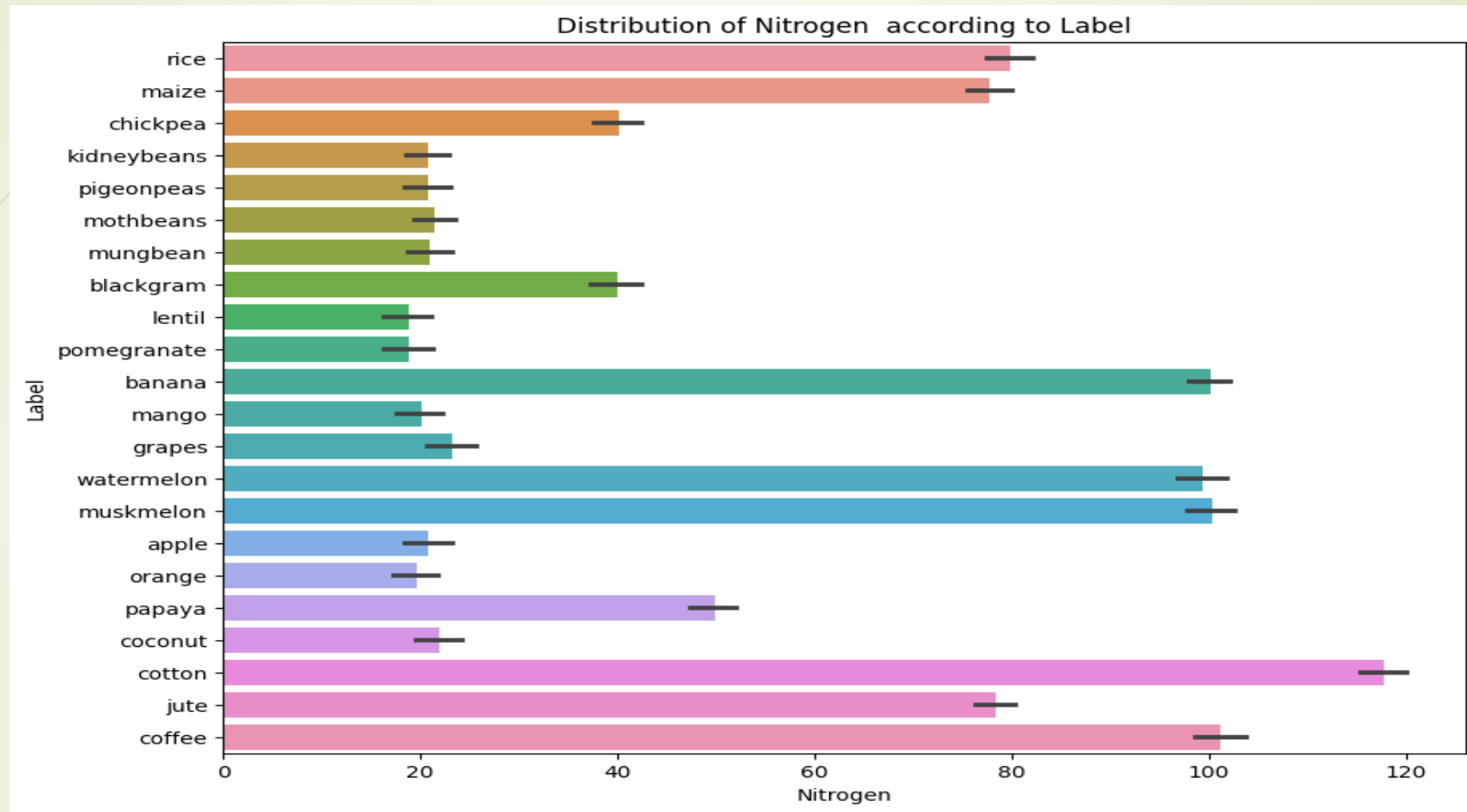
Visual 1

Knowing the needs of the crops and the reserves of the soil is very important for the success of the crop and to guarantee an excellent yield and subsequently economic profitability we will discuss two main factors for plant production which are water(rainfall)and nitrogen (N)



The bar plot above represents the distribution of rainfall according to Labes. in other words the variation in water needs depending on labels. This barplot shows us that certain label need a high water quantity (water demanding plants) such as rice, coconut and jute while the other label have moderate to low water need.

Visual 2



Nitrogen is an essential element for plant growth and production. The bar plot above represents the distribution of nitrogen according to Labels. In other words, the variation in nitrogen needs depending on labels. This barplot shows us that certain labels need a high nitrogen quantity such as rice, maize, banana, watermelon, muskmelon, and cotton, while the other labels have a low nitrogen need.

Strengths and limitations of your model

High Accuracy (0.99):

This means that it correctly classifies instances with a high degree of accuracy(99%), indicating that it can reliably recommend the suitable crops .

High Precision (0.99):

This means that when the model predicts a crops to grow it is correct 99% of the time.

High Recall (0.99):

With a recall of 99%, the model is capable of identifying the suitable crops to grow up correctly. This is essential for ensuring we don't grow other crop to optimize harvest yields .

High F1-Score (0.99):

The F1-Score of 0.99 is a robust measure of the model's overall effectiveness. It balances precision and recall.

Recommendation

To have best harvest yield and economic profitability Farmer should :

- Get a clear idea of field environmental conditions (temperature ,rainfall) and soil resources (N , P, K , ph)
- Choose crops according to model predictions
- Choose crops that do not require water for dry areas and crops that do not require mineral elements(N,P,K) given the excessive cost of inputs