**Loan Eligibility Prediction Tutorial | Python Projects For Data Science | Python Training**

**Machine Learning:**

* Branch of study that uses algorithms for the computer to learn from data and improve accuracy by itself through experience.

**Types of Machine Learning:**

1. Supervised learning: Train machine using labeled data

2. unsupervised learning train with unlabeled data, no answer provided for the machine to learn from. Here model works on its own to discover a pattern that was previously not discovered.

3. Reinforcement learning: Learned from experience or previous mistakes.

**Supervised Machine Learning:**

* This project uses Supervised Machine Learning.
* The machine is trained using labeled data.
* Algorithms:

1. Classification:

Decision Tree

Random forest

Naïve Bayes

1. Regression:

Simple linear regression

Multiple linear regression

Polynomial regression.

**Algorithms used in this project:**

* Since we need to classify if a person is eligible for a loan or not, we will be using classification algorithms.
* We use 2 algorithms and decide which is better:
  1. Decision tree classification algorithm
  2. Naïve Bayes classification algorithm

**Decision tree classification algorithm:**

* Used or both classification and regression but mainly used for classification.
* The decision tree forms a tree-like structure with the non-terminal, root node, and internal node. An internal node represents the attribute of the data.
* Branches represent the outcome of rules like yes or no and the leaf node is the final class or decision made.

To predict loan eligibility we have 2 attributes:

1. Credit history
2. Annual income

First, the best attribute is selected based on the below which sorts the best attribute to be placed on the tree.

1. Entropy: the measure of randomness in information being processed(low)
2. Information Gain: how well attribute separates training data.(high)
3. Gini index
4. Gain Ratio

Forms condition for dividing data:

1. Credit history>0
2. Annual income>4 lakh

**Naïve Bayes classification algorithm:**

* Simple, fast, reliable, accurate
* Uses Bayes theorem.
* The assumption made by naïve algorithm:

1. Each feature is independent of the other
2. Every feature contributes equally to the outcome.

**Libraries used in the project:**

1. Pandas: Powerful, fast, flexible library for data analysis, manipulation, and filtering

2. NumPy: To perform mathematical and logical operations on arrays

3. Matplotlib: comprehensive library to create static, interactive, animated visualization

4. Scikit learn : Tools for machine learning and statistical modeling

**Coding:**

1. Importing libraries we commonly use:

1.**import pandas as pd:** This line imports the Pandas library and gives it the alias 'pd.' Pandas is a popular data manipulation and analysis library in Python. Importing it this way allows you to refer to functions and objects from the Pandas library using the 'pd' prefix. For example, you can use 'pd.DataFrame' to create a DataFrame.

2.**import numpy as np:** This line imports the NumPy library and gives it the alias 'np.' NumPy is another popular library in Python used for numerical and mathematical operations. Importing it this way allows you to refer to functions and objects from the NumPy library using the 'np' prefix. For example, you can use 'np.array' to create NumPy arrays.

3.**import matplotlib.pyplot as plt:** This line imports the 'pyplot' module from the Matplotlib library and gives it the alias 'plt.' Matplotlib is a widely used library for creating static, animated, and interactive plots and visualizations in Python. Importing 'pyplot' allows you to use Matplotlib for creating charts and graphs. The 'plt' alias makes it convenient to access 'pyplot' functions.

4**.%matplotlib inline:** This line is a special command used in Jupyter Notebook to display Matplotlib plots directly within the notebook, rather than in a separate window or file. It's not a standard Python command but is used within the Jupyter Notebook environment to ensure that any plot generated using Matplotlib will be shown within the notebook.

1. Will be using 2 dataset:
   * 1. Train dataset: This will have all the labels assigned to it. Develop a model and test its accuracy.
     2. Test dataset: no predefined label, using the model built