

MACHINE LEARNING LAB

EXPERIMENT- 1

TOPIC CHOSEN: Predict the percentage of an student based on the no. of study hours.

DATA SET LINK: <https://www.kaggle.com/code/ameythakur20/tsfinternshiptask-1-supervised-learning>

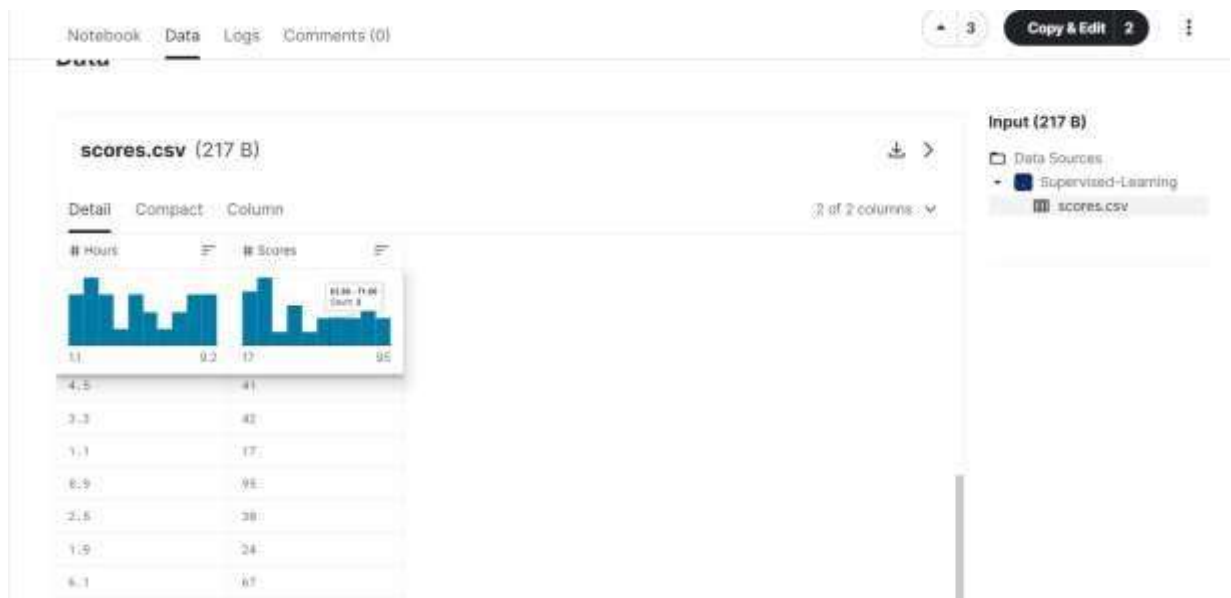
DATASET NAME: scores.csv

SOURCE: KAGGLE

DATA SET DETAILS:

There are two classes in the data collection, each with 50 instances, for a total of 150 instances. Each class describes the hours and scores.

DATASET:



1. SUPERVISED LEARNING:

The machine learns while being watched over in supervised learning. It includes a model that can make predictions using data that has been labelled. A dataset that has been labelled means that you already know the intended response.

Supervised learning can be further divided into two types:

- Classification
- Regression

DECISION TREE CLASSIFIER:

The general motive of using a Decision Tree is to create a training model which can be used to predict the class or value of target variables by learning decision rules inferred from prior data(training data).

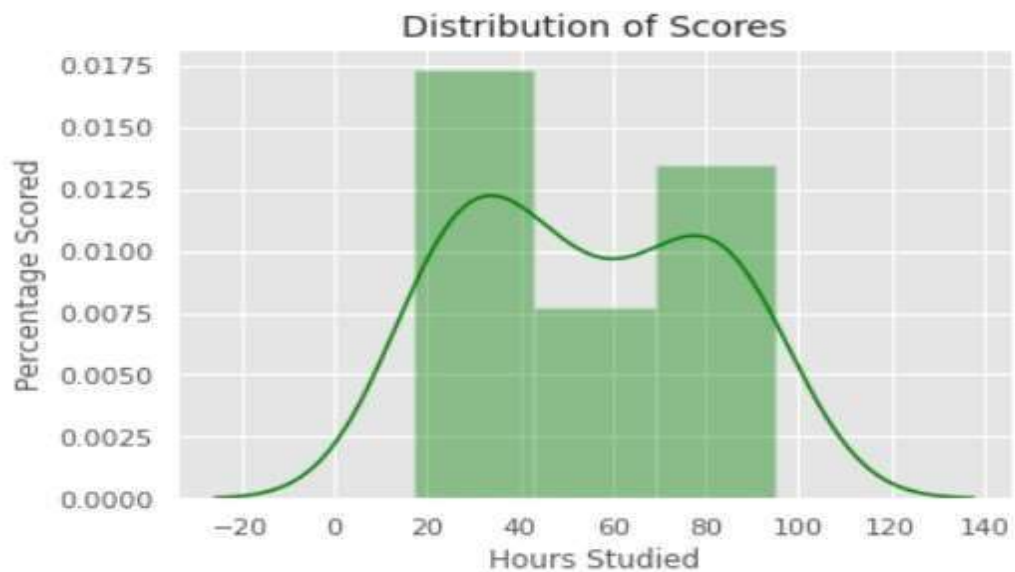
It tries to solve the problem, by using tree representation. Each internal node of the tree corresponds to an attribute, and each leaf node corresponds to a class label.

PREPROCESSING THE DATA:

```
%matplotlib inline import
pandas as pd import numpy
as np
import matplotlib.pyplot as plt
import seaborn as sns sns.set()
plt.style.use('ggplot') data =
pd.read_csv("scores.csv") X =
data.iloc[:, :-1].values y =
data.iloc[:, 1].values
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size= 0.2,
random_state= 0)
```

DISTRIBUTION PATTERN:

```
sns.distplot(y_train, kde=True, color='green',)
plt.title('Distribution of Scores') plt.xlabel('Hours
Studied')
plt.ylabel('Percentage Scored')
```



VISUALISING THE RESULTS:

```
sns.regplot(X_train, y_train, color='green', )
plt.title('Hours vs Scores') plt.xlabel('Hours Studied')
plt.ylabel('Percentage Scored')
```



2.UNSUPERVISED LEARNING:

Unsupervised learning is a type of machine learning in which models are trained using unlabeled dataset and are allowed to act on that data without any supervision.

The unsupervised learning algorithm can be further categorized into two types of problems:

- Clustering
- Association

K-MEANS CLUSTERING ALGORITHM:

The goal of the K-means clustering algorithm is to find groups in the data, with the number of groups represented by the variable K. The algorithm works iteratively to assign each data point to one of the K groups based on the features that are provided.

The outputs of executing a K-means on a dataset are:

- K centroids: Centroids for each of the K clusters identified from the dataset.
- Labels for the training data: Complete dataset labelled to ensure each data point is assigned to one of the clusters.

K-MEANS:

```
data.plot(x='Hours', y='Scores', color='green', style='*')  
plt.title('Hours vs Percentage') plt.xlabel('Hours  
Studied') plt.ylabel('Percentage Scored') plt.show()
```

