**Feature Scaling**

**What is Feature Scaling?**

“Scale Data for better performance of Machine Learning Models. “

When you get a large amount of data, you have to do a lot of Processing.

Feature Scaling is one of the steps of processing that helps you build Machine Learning Model.

There are many Independent features present in your data, now to standardize those independent features present in the data in fixed range we use a technique called Feature Scaling.

Feature Scaling helps you build a strong machine learning algorithm.

Feature Scaling helps to bring all independent features in same standing or basically we can say that it normalizes the data within a particular range.

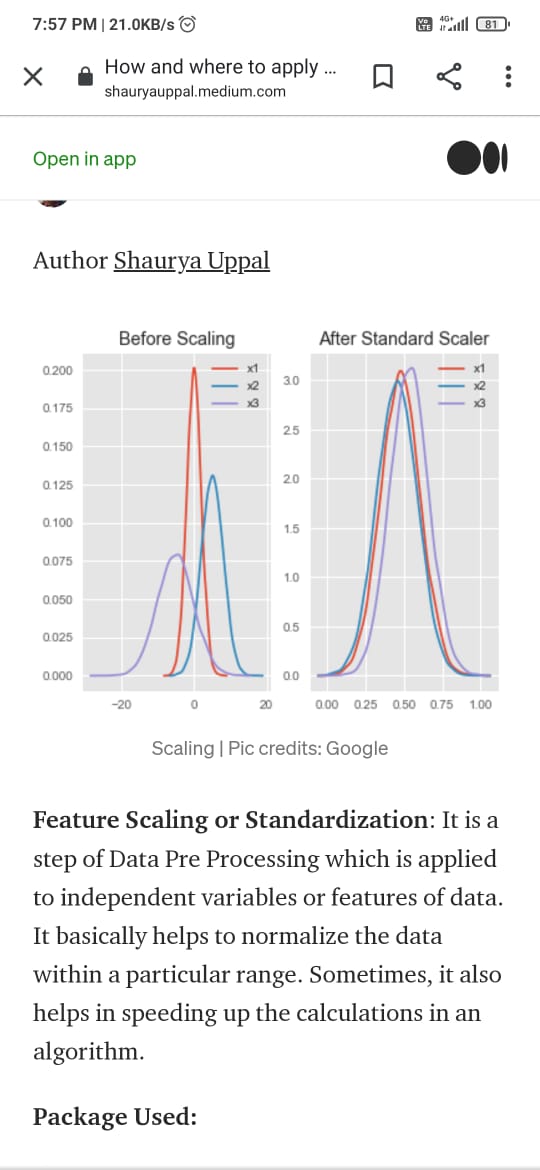
Sometimes, Feature Scaling also helps by increasing the speed of calculations in an algorithm.

Package Used:

sklearn.preprocessing

Import:

from sklearn.preprocessing import StandardScaler



Algorithms where Feature Scaling Matters:

1.K-Means

2.K- Nearest - Neighbours

3.Gradient Descent

4.Principal Component Analysis (PCA)

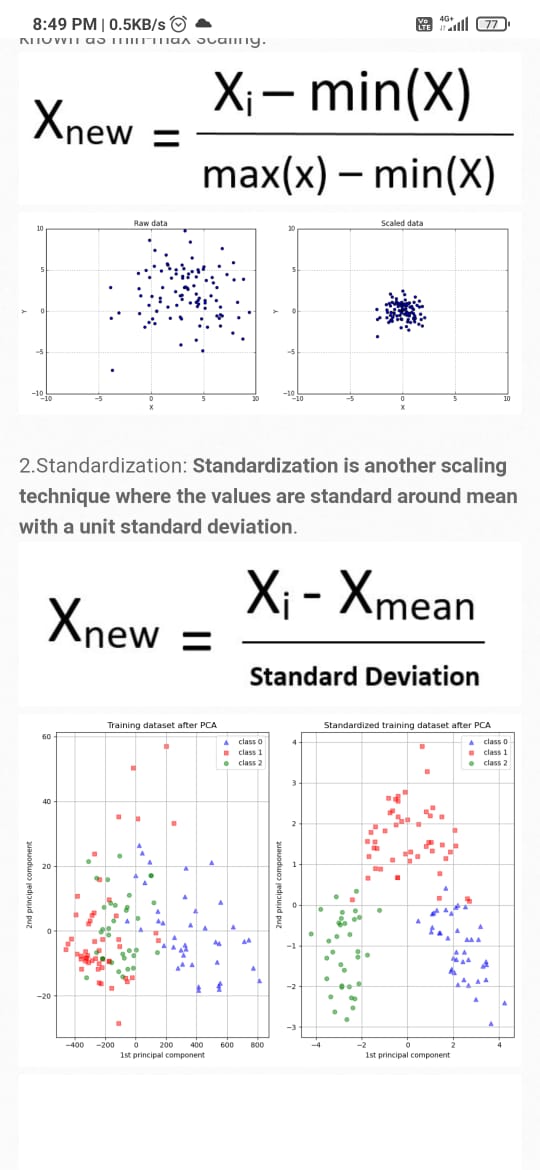
Now, since you have got an idea of what is Feature Scaling . Let us explore what methods are available for doing Feature Scaling:

**1. Normalization**:

Normalization is also known as min-max scaling or min-max normalization.

It is one of the simplest method used for scaling. So, Normalization is used to rescaled the range of variables/ features to scale the range in [0,1].

Formula for Normalization is given as:



Here, max (x) and min (x) are the maximum and minimum values of the features.

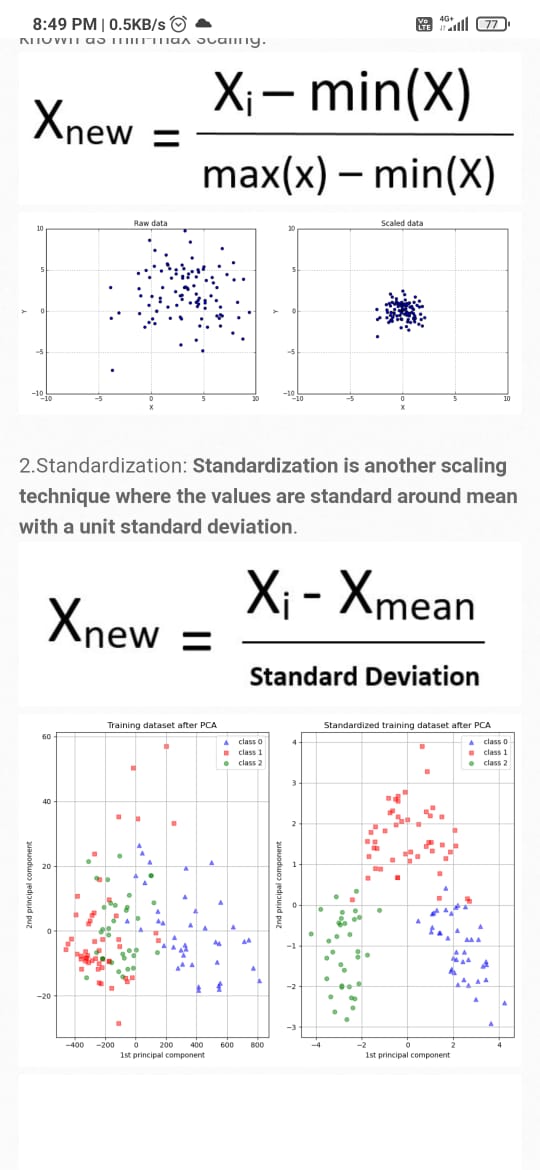
Normalization is helpful in cases when you know the distribution of data does not follow gaussian distribution.

Normalization ranges between 0-1.

Popular example of date Normalization is Image Processing, where pixel intensities have to be normalized to fit within a certain range.

**2.Standardization**:

Standardization is another scaling technique where the values are standard around mean with a unit Standard Deviation.



\*standard deviation of feature vector.

x is the average of the features .

Standardization is helpful in cases where data follows a Gaussian distribution. Standardization does not have specific ranges.

**Why are we using Feature Scaling?**

Now let us understand that why features need to be scaled.

As I already told you, we need Feature Scaling to standardize Independent features or to put those features in the same range or same scale, So that no variables is dominated by the other.

Sometimes when features are not in same range, we can get ambiguous result. Some Machine Learning algorithm like KNN are sensitive to features, so to avoid ambiguity , we can use Feature Scaling.

Some algorithms like:

Linear Regression, Logistic Regression, Neural Networks etc. use gradient descent for optimization. So, these algorithms need their data to be in the same range or we can say that they need to be scaled. The presence of feature value will affect the step size of gradient scale.

**How Feature Scaling works?**

Let's consider you are an intern and your age (18-100) Years, your stipend is (25,000-50,000)Euros and your height is (1-2) Meters. These are all the independent variables or features of the given dataset of an Intern.

These independent features have different ranges, feature scaling would help these features by bringing them all in the same range, for example: centered around 0 or in the range(0,1) depending on the scaling techniques.

**Learn more by implementing Feature Scaling in Python.**

**Happy Learning.**