**Project-to-Understand-the-Principal-Component-Analysis**

Principal Component Analysis (PCA) is a linear dimensionality reduction technique that can be utilized for extracting information from a high-dimensional space by projecting it into a lower-dimensional sub-space. It tries to preserve the essential parts that have more variation of the data and remove the non-essential parts with fewer variation. Dimensions are nothing but features that represent the data. For example, A 28 X 28 image has 784 picture elements (pixels) that are the dimensions or features which together represent that image.

One important thing to note about PCA is that it is an Unsupervised dimensionality reduction technique, you can cluster the similar data points based on the feature correlation between them without any supervision (or labels)

Here in this Project we are using Dermotology Data Set

**Libraries Used**

1. Pandas
2. Numpy
3. matplotlib
4. seaborn
5. sklearn.preprocessing import StandardScaler
6. sklearn.decomposition import PCA
7. sklearn.model\_selection import train\_test\_split
8. sklearn.ensemble import RandomForestClassifier

**Steps**

1. Importing Essential Libraries
2. Loading the Data Set
3. Structure the Dataset
4. Check for Missing or Null Points
5. Exploration of the Dataset
6. Data Cleaning
7. Principal Component Analysis
8. Machine learning Model - Random Forest Classifier 9 Train, Test and Predict

**Summary**

We have successfully used the Dermatology data set and build PCA model with an accuracy of 95.45%