# **Dataset Overview :**For this task, we’ll use the **Voice Dataset**, which contains various mathematical attributes like mean, skew, std etc., and a target label indicating whether an individual is male or female.

# The dataset includes the following features:

* **duration**: length of signal
* **meanfreq**: mean frequency (in kHz)
* **sd**: standard deviation of frequency
* **median**: median frequency (in kHz)
* **Q25**: first quantile (in kHz)
* **Q75**: third quantile (in kHz)
* **IQR**: interquantile range (in kHz)
* **skew**: skewness (see note in specprop description)
* **kurt**: kurtosis (see note in specprop description)
* **sp.ent**: spectral entropy
* **sfm**: spectral flatness
* **mode**: mode frequency
* **centroid**: frequency centroid (see specprop)
* **peakf**: peak frequency (frequency with highest energy)
* **meanfun**: average of fundamental frequency measured across acoustic signal
* **minfun**: minimum fundamental frequency measured across acoustic signal
* **maxfun**: maximum fundamental frequency measured across acoustic signal
* **meandom**: average of dominant frequency measured across acoustic signal
* **mindom**: minimum of dominant frequency measured across acoustic signal
* **maxdom**: maximum of dominant frequency measured across acoustic signal
* **dfrange**: range of dominant frequency measured across acoustic signal
* **modindx**: modulation index.

**Step 1: Loading and Exploring the Data:**

Import pandas' library ,load and read the dataset and explore the data using describe, info and columns function

# **Step 2: Data Preprocessing**

Machine learning models often perform better when the data is cleaned and preprocessed. We’ll take the following steps:

**Handle Missing Data**: If any missing values exist, impute or remove them. Use different pandas function like isnull, fillna and remove.

**Encode Categorical Features**: Many machine learning algorithms don’t work with categorical data directly, so we need to convert these into numerical representations. For this import Labelencoder library and use fit\_transform() function of Labelencoder.

**Feature Scaling**: Normalize or standardize numerical features. For this import StandardScaler library and use fit\_transform() function of StandardScaler. Before Scaling the data, split it into train and test set using train\_test\_split library and scale the train set.

# **Step 3: Model Selection and Training**

# Experiment with models such as Logistic Regression or Support Vector Machines (SVM). For this import sklearn library.

# **Step 4: Model Evaluation**

To evaluate the performance of our model, we’ll use common classification metrics such as accuracy, precision, recall, and F1-score.

# **Conclusion**

We successfully built a machine learning model using the Voice dataset to predict whether an individual is male or female. We concluded that Random Forest give better result than logistic regression ,SVM ,Decision Tree Classifier ,Kneighbour Classifier and GaussianNB.