**FEATURES & THE PROCESSES INVOLVED**

Features, basically means the characteristic property of data obtained from various processes that serves as input to our model and give the predicted/required results for further study.

One important term which also comes along is Domain. It means the category under which the model would operate, and hence features which comply under these domain would be used for serving the required model (Serves for feature classification).

Data that is being obtained (referred as raw data) usually consists of large amount of information of random data, from this large set taking out and categorising data into such groups that it gives us the relevant information which would help in analysing.

This is achieved after collecting the raw data and filtering it out from the unnecessary data and then extracting the required features.

For example, Consider a dataset obtained from a school, it consists of all data from surveys that could be obtained. Now after obtaining these raw data, first aim is to remove unnecessary ones and then group them properly in order to achieve useful information from it.

Now these features vary upon different types in general like

- Numerical: No.of students/classrooms etc.

- Categorical: Color of building blocks, subjects etc.

- Binary: (In yes/no) like whether they are provided with sufficient break time etc.

With respect to ML, features serve an important purpose as input to the models in order to analyse the function of models and predict the outcomes.

From the raw data, these useful information are categorised under features, this method is termed as feature extraction (extracting the useful data from the large data set).

For images, few important features would be contours, colors, edges, texture etc.

After extracting them the next important thing is selecting the most accurate one in order to input it (feature selection). And hence inputting these info into the model and get our desired results.

**AUDIO & SIGNAL FEATURES**

Audio is associated for the data obtained of sound which is basically in the range of human hearing and is represented as waveform.

Signal is associated with data which varies with time and is obtained from various forms as optical/electrical/electromagnetic etc and represented on various domains as time/frequency etc.

Audio features:

Depending upon the domain we have got the following features,

* Time-domain features:
* Amplitude: Magnitude of the sound signal.
* Zero Crossing Rate: Signals varying from +ve to -ve and vice-versa.
* Frequency-domain features:
* Spectral centroid: Deals about finding out where the most energy of the signal is present.
* Spectral Bandwidth: Width of the range of frequencies of the signals.
* Cepstral features:
* Mel-freq Cepstral Coefficient(MFCC): Short-term power spectrum signals.
* Delta and Delta-Delta MFCC: Consists of first and second order derivative of MFCC.

Signal features:

Similarly depending upon domain, we got the feature classification,

* Time-domain features:
* Mean/Variance/Skewness/Kurtosis: These provide the statistical information of the data obtained (since the data can be plotted on graph and we can find them).
* Peak-to-Peak: Gives the difference between the maximum and minimum values of the signal.
* Frequency-domain features:
* Power Spectral Density: Shows the distribution of power over the frequency ranges of the signals.
* Frequency Bands: Classifying the frequencies on various ranges
* Time-frequency domain features:
* Short time fourier transform: Represents data in both time and frequency domains.
* Wavelet transform: For non-stationary signals.

These are the various categories upon which the received data is to be set in order to feed into the model.