**Kl divergence:**   KL divergence is the measure of the relative difference between two probability distributions, P and Q. KL divergence is also known as relative entropy. Therefore, when two distributions are identical, their KL divergence is zero. We can also use KL divergence to determine whether a Poisson distribution or a normal distribution can better approximate the data

**Cross Entropy:**Cross entropy is often used as a loss function in machine learning. Cross entropy calculates the difference between two probability distributions and is closely related to, but distinct from, KL divergence. It calculates the total entropy between two probability distributions, whereas KL divergence can be thought to calculate the relative entropy between the distributions It can also be said that cross entropy is the average number of bits needed to encode data from a distribution P, if we use the q model.

**Conditional Entropy:**The conditional entropy of a random variable Y when another variable X is present, defines how much extra information one still needs to provide on average to express Y given X is known.

**Entropy:** Entropy is a way to measure the uncertainty/randomness of a random variable X. In other words, entropy measures the amount of information in a random variable. It is usually measured in bits. When the entropy becomes 0, the dataset has no contamination. Datasets with 0 impurities are not useful for learning. If the entropy is also 1, then this type of dataset is good for learning.