**ASSIGNMENT ( PART 2)**

**Question 1: Write the difference between the following:**

**i. Gaussian Naive Bayes**

**ii. Multinomial Naive Bayes**

**iii. Complement Naive Bayes**

**iv. Bernoulli Naive Bayes**

**v. Categorical Naive Bayes**

**vi. Out-of-core naive Bayes model fitting**

Answer:

One highly practical Bayesian learning method is the naïve Bayes classifier. The naïve Bayes classifier applies to learning tasks where each instance x is described by a conjunction of attribute values and where the target function f(x) can take on any value from some finite set V.

1. Gaussian Naïve Bayes: In this algorithm, the continuous variables associated with each feature are assumed to be distributed according to the Gaussian Distribution. When plotted, it gives a bell-shaped curve which is symmetric about the mean of the feature.
2. Multinomial Naïve Bayes: This algorithm is used for discrete counts. It is used to check how many times an outcome has occurred over the total number of trials. It is mostly used in text classification problems. For example, to find out how many times a particular word has occurred is the document.
3. Complement Naïve Bayes: This algorithm is used to work with imbalanced data. It works like Multinomial Naïve Bayes but instead of calculating the probability of item belonging to a certain class, we calculate the probability of an item, belonging to all classes, i.e. complement.
4. Bernoulli Naïve Bayes: This algorithm is used to give binary output. It is used to check whether a particular event has occurred or not, instead of checking for its frequency. Mostly used for Text Classification Problems.
5. Categorical Naïve Bayes: This algorithm is suited for discrete features that are categorically distributed. The features here are extracted from a categorical distribution.
6. Out-of-core Naïve Bayes Model fitting: This particular method is used when the dataset is so huge that even the complete training set won’t fit into memory, meaning that the classification problem is very huge. To handle this, online classifier can be used with partial\_fit method.

**Question 2: Define which text preprocessing and text transformation steps did you use for the above.**

Answer:

1. The Null Values were filled:
   1. Categorical with the Mode values
   2. Numerical Values with mean or median, whichever was fit
2. Normalization was also used to transform words from slangs to to their appropriate form.

**Question 3: What is Jaccard and Cosine Similarity?**

Answer:

1. Jaccard Similarity: This is basically the coefficient of similarity between 2 sets. Normally used for Checking number of similar words between 2 sets. The range is 0 to 100. The more the similar words, the more the percentage. The similarity is measured by taking the intersection of the two sets and dividing it by their union.
2. Cosine Similarity: In this method, all the words and sentences are converted to vectors and then the cosine angle between those vectors is checked by taking their dot product. The larger the angle, the lesser the similarity. This means that when the angle is 0, the similarity is 1.