1. If the linear regression coefficient of a predictor is 0.54 then what does it mean?

Ans: Linear Regression is a statistical model it works on regression coefficients and weighted averages

Regression coefficient explains change in the value of dependent variable corresponding to the unit change in the independent variable.

If the linear regression coefficient of a predictor is 0.54 indicates unit increase in independent variables increases 54 percent of target variable

Example : T.V ads –predictor

Sales ---target

If T.V ads coefficient is 0.54 means :

Unit increase in T.V ads increases 0.54 times sales

100$ spend on T.V ads increases 54 percent sales

1. How would you deal a data with Target class imbalance problem?

Ans : If Target class Imbalances are present in data set it leads to classification accuracy. Model will be biased to majority of the classes.

**Resampling Techniques helps** in maintaining balanced classes .The main objective of balancing classes is to either increasing the frequency of the minority class or decreasing the frequency of the majority class. This is done in order to obtain approximately the same number of instances for both the classes.

**Random Under-Sampling**

Random Undersampling aims to balance class distribution by randomly eliminating majority class examples.  This is done until the majority and minority class instances are balanced out.

#### Random Over-Sampling

Over-Sampling increases the number of instances in the minority class by randomly replicating them in order to present a higher representation of the minority class in the sample.

#### Cluster-Based Over Sampling

In this case, the K-means clustering algorithm is independently applied to minority and majority class instances. This is to identify clusters in the dataset. Subsequently, each cluster is oversampled such that all clusters of the same class have an equal number of instances and all classes have the same size

* Each of the methods has own pros and cons

1. You have built a classification model with 90% accuracy but your client is not happy because False Positive rate was very high then what will you do?

Ans: FPR : FP/FP +TN

False positive rate is nothing but actual value s zero(0) but model predicted as 1

If FPR is high then model is biased to positive classifiers we need to check classifier imbalances

And try to maintain balanced classifiers.

1. Does multicollinearity effects in Naïve Bayes? If yes/no then why?

Ans: Multicollinearity does not effect in Naïve Bayes.In Naïve Bayes each variable is consider as independent of each other effect of one variable on the other variable is zero. Hence Multicollinearity doesn’t effect.

1. If we do not define number of trees to be built in random forest then how many trees random forest internally creates?

Ans: The number of trees in the forest in default is 10

A random forest is a meta estimator that fits a number of decision tree classifiers on various sub-samples of the dataset and use averaging to improve the predictive accuracy and control over-fitting. The sub-sample size is always the same as the original input sample size but the samples are drawn with replacement if bootstrap=True (default)

n\_estimators : integer, optional (default=10)