Classification Modelling

1) **What is the advantage of performing dimensionality reduction before fitting on SVM?**
   SVM algorithm performs better in the reduced space. It is beneficial to perform dimensionality reduction before fitting on SVM. If the number of feature is large when compared to the number of observations.
2) **What is Logistic Regression?**
   It is a statistical technique in order to analyze a dataset and predict the binary outcome given a set of independent variables. The outcome has to be a binary outcome that is either zero or one, yes or no.
3) **Why is Naïve Bayes?**
   It is naïve because it uses conditional probability to make classifications, the algorithm simply assumes that all features of a class are independent. This is considered naïve Bayes, because in reality, it is not often the case. The upside is that the math is simpler, the classifier runs quicker, and the results are often quite good.
4) **Could you explain KNN imputation?**
   In KNN Imputation, the missing attribute values are imputed by using the attributes values that are most similar to the attribute, whose values are missing. By using a distance function, the similarity of two attributes is determined.
5) **What is confusion matrix?**
   A tabular display (2 x 2 in the binary case) of the record counts by their predicted and actual classification status.
6) **What is accuracy?**
   The Percent (or proportion) of cases classified correctly. Accuracy is measure of how good our model is. It is expected to be closer to 1; it means our model is performing well. Accuracy is the ratio of correct predictions and all total prediction.
7) **What is Precision?**
   The percent (proportion) of predicted 1's that are actually 1's. Precision is defined as how many selected items are relevant that is how many of the predicted one are actually correctly predicted.
8) **What is sensitivity or Recall?**
   The percent (or proportion) of 1's correctly classified. Recall tells how many relevant items we selected mathematically.
9) **What is Specificity?**
   The percent (or proportion) of 0's correctly classified.
10) **What is ROC curve?**
    A plot of sensitivity versus specificity.
11) **What is AUC curve?**
    AUC is believed to be one of the ways to summarize performance in a single number, AUC indicates that the probability of a randomly selected positive example will be scored higher by the classifier than a randomly selected negative example; if you have multiple models with nearly same accuracy, you can pick the one that gives higher AUC.
12) **What is lift?**
    A measure of how effective the model is at identifying 1s at different probability cutoffs.
13) **What is F-Score?**
    F-Score is a measure of accuracy technically it is harmonic mean of precision and recall.
14) **What is Class Imbalance?**
    Providing an equal sample of positive & negative instances to the classification algorithm will result in optimal result. Dataset that are highly skewed toward one or more classes have proven to be a challenge.
15) **How to detect Class Imbalance?**
    • make a cross table.
16) **How to fix the Class Imbalance?**

The class imbalance issue can be solved using the different types of sampling methods as follow

- Oversampling
- Under sampling
- Both Sampling

17) **Can you explain me what machine learning Algorithms are used in this project?**
Sure, we were mostly using classification algorithms like naïve Bayes, logistic regression for churn prediction, Recommendation algorithms like collaborative filtering to recommend advanced next best offers to customers, regression algorithms for campaign targeting and marketing mix modelling. Association rule algorithms for affinity Analysis.
18) **What is supervised and unsupervised Machine Learning?**
Supervised machine learning are used with labelled data whereas unsupervised machine learning used unlabeled data.
19) **Could you explain the difference between Generative & Discriminative Models?**
A generative model will learn categories of data, while a discriminative models will simply learn the distinction between different categories of data.
20) **What is the F1-score? How would you use it?**
It is a weighted average of the precision and recall of a model, with results tending to “1” being the best and those tending to “0” being the worst. You would use it in classification test, where true negative don’t matter much.
21) **What are parametric models?**
Parametric models are those with a finite numbers of parameters to predict new data. You need to know the parameters of the model to run the test example including linear regression, logistic regression and linear SVM.
22) **Do Gradient Descent Method always converge to same point?**
No, they don’t because in same cases it reaches a local minima or a local optima point. You don’t reach the global optima point. It depends on the data at its starting conditions.