Q1 The filter method filters out the irrelevant feature and redundant columns from the model by using different metrics through ranking. The advantage of using filter methods is that it needs low computational time and does not overfit the data.Filter methods measure the relevance of features by their correlation with dependent variable while wrapper methods measure the usefulness of a subset of feature by actually training a model on it. Filter methods are much faster compared to wrapper methods as they do not involve training the models.

Q2 The main differences between the filter and wrapper methods for feature selection are: Filter methods measure the relevance of features by their correlation with dependent variable while wrapper methods measure the usefulness of a subset of feature by actually training a model on it. Filter methods perform the feature selection independently of construction of the classification model. Wrapper methods iteratively select or eliminate a set of features using the prediction accuracy of the classification model. In embedded methods the feature selection is an integral part of the classification model.

Q3 In an embedded method, feature selection is integrated or built into the classifier algorithm. During the training step, the classifier adjusts its internal parameters and determines the appropriate weights/importance given for each feature to produce the best classification accuracy. The most Common embedded technique are the tree algorithm's like RandomForest, ExtraTree and so on. Tree algorithms select a feature in each recursive step of the tree growth process and divide the sample set into smaller subsets.

Q4

The filter method looks at individual features for identifying it's relative importance. A feature may not be useful on its own but maybe an important influencer when combined with other features. Filter methods may miss such feature.One downside of such methods is that they do not interact with the predictive model for feature selection. Another disadvantage is seen in the case of univariate filter methods where dependencies between features are normally ignored.The increasing overfitting risk when the number of observations is insufficient. The significant computation time when the number of variables is large

Q5 For large data you should use the Filter approaches because these approaches are rapid and for small size of data it is better to use Wrapper (KNN, SVM,...) approaches because they are slower than the Filter approaches. or you can combine the two approaches to have better results than the two approaches.

Q6 Telecom Churn Prediction is a machine learning project which focuses on predict customer churn in the telecom industry. Churn prediction is important for telecom companies because it helps them to retain their customers and reduce customer acquisition costs.Churn calculations are built on existing data – the number of customers who left your service during a given time period. A predictive churn model extrapolates on this data to show future potential churn rates. This helps you predict your revenue and avoid risks like overspending. Step 1: Problem Definition Step 2: Data Collection Step 3: Exploratory Data Analysis (EDA) Step 4: Feature Engineering Step 5: Train/Test Split Step 6: Model Evaluation Metrics Definition Step 7: Model Selection, Training, Prediction and Assessment Step 8: Hyperparameter Tuning/Model Improvement

Q7 Analyzing each team's strengths, weaknesses, current form, and past results can provide valuable insights into their potential performance. Similarly, studying key players' performance, injury status, and current form can help gauge their impact on the match outcome.

The first thing that any ML Engineer should do is to define the problem. For this project, the problem we are facing is accurately predicting the player ratings without having to hard code an algorithm, so building a workable ML model should be our end goal. In FIFA 20, a player’s overall score is defined by the score of each player attribute such as Pace, Shooting, Passing, Dribbling, and many others. We are trying to predict the overall score based on the player’s offensive and defensive attributes. The data already comes with labeled training examples which means we are dealing with a supervised learning problem.

Q8 To visualize the correlation between attributes, we can use the pandas wapper() method. This function plots every numerical attribute against every other numerical attributes. Let’s just look at the top five promising attributes to make things easier since there are 35 numerical attributes.Analyzing each team's strengths, weaknesses, current form, and past results can provide valuable insights into their potential performance. Similarly, studying key players' performance, injury status, and current form can help gauge their impact on the match outcome.