The Metrics for the different feature subset selection methods is different because each method uses their unique scoring function to select the features In our case Anova uses f-statistic value An [F-statistic](https://en.wikipedia.org/wiki/F-test), or F-test, is a class of statistical tests that calculate the ratio between variances values, such as the variance from two different samples or the explained and unexplained variance by a statistical test, like ANOVA. The ANOVA method is a type of F-statistic referred to here as an ANOVA f-test. ANOVA is used when one variable is numeric and one is categorical, such as numerical input variables and a classification target variable in a classification task. It uses f\_classif and selectKbest to select top k features from dataset. PCM uses Correlation coefficient as a scoring function. **Correlation coefficients** are used to measure how strong a relationship is between two [variables](https://www.statisticshowto.com/probability-and-statistics/types-of-variables/) that means it calculates correlation coefficient value of the each feature with the target feature and gives the list we selected top k features with correlation coefficient values.Chi-square it uses chi-square value as a scoring function to select top features. A chi-square test is used in statistics to test the independence of two events. Given the data of two variables, we can get observed count O and expected count E. Chi-Square measures how expected count E and observed count O deviates each other. In feature selection, we aim to select the features which are highly dependent on the target feature. We use chi2() function and SelectKbest to select top k features from the dataset based on chi2 value.Information gain uses information gain value as a scoring function to select top features. Information gain can also be used for feature selection, by evaluating the gain of each variable in the context of the target variable. In this slightly different usage, the calculation is referred to as mutual information between the two random variables.We used mutual\_info\_classif() function and selectKbest to select top k features from the dataset based on mutual information value. Wrapper methods uses the p-value as a scoring function to rank the features. The p-values are a decimals between 0 and 1 that represent the probability that the data given to us occurred by chance under the hypothesis test. Simply put, the lower the p-value, the better the chance that we can reject the null hypothesis. For our purposes, the smaller the p-value, the better the chances that the feature has some relevance to our response variable and we should keep it. Sequential forward selection goes forward one by one and select the features based on p-value and and sequential backward goes from backward and eliminates the feature with low p-value these wrapper methods do these on any one algorithm So based on these scoring functions of different feature subset selection methods we are getting different metric values for different methods and those are mainly depends on the data we used and the algorithm we used. We came to know about the algorithm and feature selection method compatibility results are different with different features and algorithms.