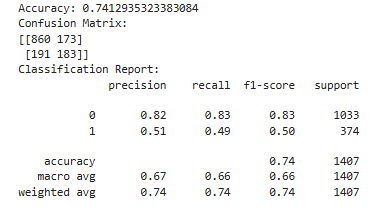
# Task\_2

Dataset: Customer Churn dataset is used for this task

Customer churn information:In Telco customer churn where customers discontinue their subscription services with a telecommunications company. Analyzing churn is crucial for these companies to understand why customers leave and to develop strategies to retain them. Common factors influencing churn include service quality, pricing, customer support, and contract terms. Data analysis and machine learning models can predict churn by examining customer behavior and characteristics, helping companies take proactive measures to reduce churn rates.

Steps :

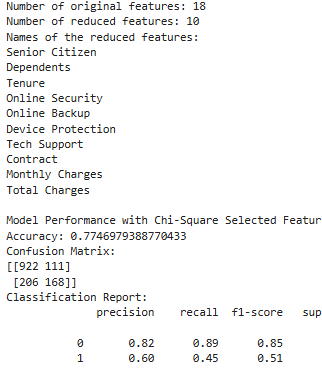
1. Load data: using Pandas library
2. Cleaning data : where missing values are handled using drop technique but I found there were not any missing values in data set and then drop unnecessary features in cleaning method and change the type of column which is **total charges.**
3. Encode features : change categorical columns into number here label encoder is used to encode data so that machine easily understand the values
4. Split data:Firstly separate target feature from data then data is separated into train and test .
5. Model Training:KNN(k-nearest neighbour ) model is used for this dataset to predict Customer Churn where the accuracy of using this model is this ..



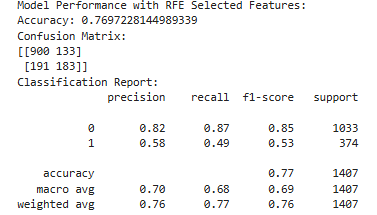
Without feature selection this is the accuracy score.

#### Feature Selection Categories:

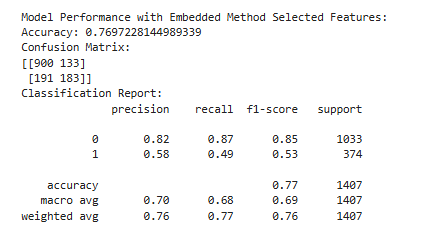
1. Filter Method :
   1. Chi-squre:Chi-square technique is used for categorical form of data by using library sklearn feature selection. Where kBest value is 10 means 10 best column will be selected from data set and I show the names of that features. After that standrize these feature and then I found the accuracy after feature selection method



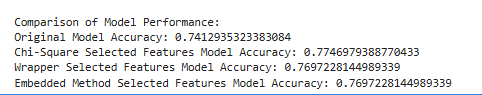
1. Wrapper Method:
   1. Recursive Eliminator feature : RFE technique is used in wrapper method where **estimator is KNN** according the requirement and 10 best features will be selected from this data set and unnecessary features will be dropped it do so to improve the performance of model. The performance of model after applying this technique is



1. Embedded Method:
   1. Random Forest: Selection of this technique to see which features are important, It evaluates how each feature contributes to reducing the impurity (e.g., Gini impurity or entropy) in the decision trees because low impurity level is good, identifying features that are informative for predicting churn. And handle linear and nonlinear and the interaction of features also. After applying this technique the accuracy is..



After applying these methods and techniques the comparison result is



Accuracy improve in filter method chi-square test.