Analytics Vidyha : Job-A-Thon

The following contains the approach used to solve the Problem : Health Insurance Lead Prediction

1. Since the given problem and data set consists of categorical target variable, I have used Classification based ML Algorithms which consists of multiple algorithms such as Logistic Regression, K-Nearest Neighbours(KNN) ,Random Forest and Naive Bayes Classifier.

2. After inspecting the data, primarily I have performed the step of indentifying the missing values and its treatment, and ensured that the data set does not have further null values.

3. This was followed by the task of converting the irregularities in the categorical data to get a standard form of numerical value for better model fitting. This was done based on the concept of Label Encoding.

4. For the final predictions done on the test data set I have considered the Random Forest Algorithm since I could get a better score for it compared to the other models.