Approach :

Data visualization (Understanding the data since it supervised learning)

Preprocessing (Columns manipulated, Graphs plotted, Normalizing the data)

Applying the code (as understood from the preprocessing and some basic code were applied)

Ensembling the result from the different techniques/codes

Calculating Mcc and accuracy

Data Visualization

Visitation the data through naked eyed i.e. if possible checking for any relations

Plotting heat maps for the data to check for any relationships in the data

As no relationships were visualized from the heat maps, so we plotted box plots for different congestion type and different features present in the data

Preprocessing

Columns signifying year, month and cell name were removed because of the  insignificance in prediction of the final class of congestion

For dimensional reduction of data tSNE (t-Distributed Stochastic Neighbor Embedding) was plotted for the data

Data was normalized to make it a better fit for possible models.

Applying the code

A bit of correlation was observed from tSNE, here we can see that the points containing NC (No Congestion) , 3G\_BACKHAUL\_CONGESTION, 4G\_RAN\_CONGESTION, 4G\_BACKHAUL\_CONGESTION were segregated .

So we thought of using SVC as it also works on same way.

( A Support Vector Machine (SVM) is a discriminative classifier formally defined by a  separating hyperplane. In other words, given labeled training data (supervised learning),  the algorithm outputs an optimal hyperplane which categorizes new examples. In two  dimensional space, this hyperplane is a line dividing a plane into two parts wherein each  class dimensional lay in either side)

We also thought of a multi-layered approach for the data, as the the data had a large no of columns.

(A ​multilayer perceptron​ (MLP) is a class of feedforward artificial ​neural network​. An MLP  consists of, at least, three layers of nodes: an input layer, a hidden layer, and an output  layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation  function​.)

Next, we also tried Naïve Bayes.

(Initial intuition behind using Naive Bayes was the conditional independence of the  available features. This was implied from the ​correlation plot ​between all the  features)

Ensembling

(After getting a (near satisfactory) results from the above 3 we tried a hybrid approach.)

In order to capture various features such as non-dependency (​Naive Bayes​) as well as subtle  features from weaker algorithms ( ​SVC and MLP Neural Networks​) and various nuances from  different activation functions such as ReLU ​ensembling ​was incorporated.