Case study: Text Classification Problem

Problem statement and Domain Knowledge:
XYZ Health Services is a top ranked Health care provider in USA with stellar credentials and provides high quality-care with focus on end-to-end Health care services. The Heath Care Services range from basic medical diagnostics to critical emergency services. The provider follows a ticketing system for all the telephonic calls received across all the departments. Calls to the provider can be for New Appointment, Cancellation, Lab Queries, Medical Refills, Insurance Related, General Doctor Advise etc. The Tickets have the details of Summary of the call and description of the calls written by various staff members with no standard text guidelines.

The challenge is, based on the Text in the Summary and Description of the call, the ticket is to be classified to Appropriate Category (out of 5 Categories) and Subcategories (Out of 20 Sub Categories).

Few words:

- Read all the way through the instructions. Code must be submitted in R/Python and write-ups in a PDF.

- Feature engineering is very important that helps us to evaluate your creativity and model performance as well. Clean the data with whatever method(s) you believe is best/most suitable. Explore all preprocessing techniques

- Visualizations is very important. Data scientist must speak in terms of numbers of visualizations.

- Please build two distinctly different machine learning/statistical models in R/Python to predict. When writing the code associated with each model, please have the first part to develop model and save the model, followed by a second part that loads and applies the model on test data.

- Submit your work: Please submit all of your code for cleaning, prepping, and modeling your data, your "Results" file, a brief write-up comparing the pros and cons of the two modeling techniques you used. You do not need to submit the original data back to us. Your work will be scored on techniques used (appropriateness and complexity), model performance, an understanding of the two techniques you compared in your write-up, and your overall code.

- Apart from accuracy and predicted values, extract some insights from data and build some recommendations.

Good Luck!!!