Analytics Project
SCMA 648
Virginia Commonwealth University
Fall 2021

1 Summary

This project will give you the opportunity to apply the methods we have covered in class to a problem involving hospital operations. The project is intended to give you experience in applying analytics methods for data analysis in the context of a real world case, using analytics software, and communicating results.

You are to form two-person teams with another member of the class. Send an email with your team members to the instructor at `jpbrooks@vcu.edu` by November 10th or indicate that you have no partner.

Begin by reading the case in the file ited.2021.0251cs.pdf.

The project consists of two central tasks:

- Clustering patients
- Prediction of which patients "flip" from observation to inpatient status.

Your team will submit a final report with the following structure: executive summary, problem introduction, description of preprocessing and methods, results, conclusions, and appendices.

The report should be written from the perspective of Dr. Kelly and her academic advisor (the data scientists in the case). The audience for the report are the project sponsors: the chief of medicine, the chief operating officer, and the chief medical officer/VP of the hospital system.

The executive summary should be no more than one page and should introduce the problem and the people involved (in the context of the case), give an overview of the approach taken, summarize results, and make recommendations. Resist the temptation to save major results for later in the report - the executive summary should clearly and concisely state all major findings. The function of the remainder of the report is to provide further details supporting the recommendation in the executive summary.

The report should be a narrative written in sentence-paragraph form and not simply an itemized list. The report is limited to 10 pages (double spaced, 12 point font). The report should be written using complete sentences and correct grammar. Submit your reports in pdf format via Canvas. Turn in your code and data as separate files.

2 Analytics Tasks

2.1 Patient Clustering

Cluster the patients into three groups using the data provided. Using principal component analysis (PCA), plot the clusters and describe the differences in values between each pair of clusters (e.g., "cluster x patients tend to have a higher pulse than those in cluster y"). Provide a visualization of the clustering using PCA.

2.2 Prediction of "Flipped" Patients

Create models to predict whether patients will "flip" from the observation unit (OU) to an inpatient unit or whether they are discharged from the OU. Be sure to apply appropriate treatment of categorical variables and to use proper practices in the empirical evaluation of predictive methods.

Apply at least two methods for the classification (e.g., classification trees and SVM) and compare the methods. The methods that you use do not necessarily need to be methods that we covered in class. You must use appropriate cross-validation techniques. Tune the parameters for at least one of the techniques. The methods should be tested using the same holdout data. Report prediction performance using confusion matrices and ROC curves.

Report on which features are important for prediction for at least one of your classification models. Create visualizations with at least two important features to demonstrate whether they are associated with determining whether a patient will flip or not.

Apply your model to the prediction data provided. Submit your predictions as a comma-delimited text file in the following format:

```
ObservationRecordKey,prediction
443621z2,0.75
131565z1,0.55
438080x1,0.76
244227z1,0.24
...,...
```

The number in the second column indicates the probability that the patient will flip.

3 Important dates

Due Date	Task
November 10th	Project teams
December 9th	Final Report