

**Predictive Analytics Term 4, 2020**  
**Associate Professor Ole Maneesoonthorn**

**Syndicate Task #5**

Let us reconsider the issue of wine quality scoring. Wines that are given quality scores of 7 or above are considered “good quality” wines. Explore the wine quality predictive problem in the binary classification context.

1. Redefine your output variable as a binary classification using the following code (assuming your original data is called “wine”):  
  

```
wine$QS<-as.factor(wine$QS>=7)
```

  

```
levels(wine$QS)<-c("No","Yes")
```

  

```
colnames(wine)[12]="GoodQS"
```
2. Explore the logistic regression and various machine learning methods in predicting the binary classification for a good quality wine. Your discussion of the predictive assessment MUST include:
  - a. The logistic regression model
  - b. At least four of the machine learning/ensemble methods covered throughout the subject.
  - c. Adapt the loss function that you have developed in Syndicate Task 3 to assess the predictive performance of your models in a binary classification setting.
3. Discuss the differences between the approach used in this syndicate task and the previous tasks. Compare the benefits and limitations of the two approaches and provide an argument as to which approach you prefer.

Produce a 5-page report that summarizes your analysis.

**This task is due at 6pm on Sunday 1<sup>st</sup> November 2020.**