ON3 Technologies – Data Science Project

**Problem Statement**

The UCI machine learning repository has a large number of public data sets available for analyses. We would like you to analyze and build a model for one of these datasets – **Wine Quality Data Set**. Please see the following site to download and read about this data set:

<http://archive.ics.uci.edu/ml/datasets/Wine+Quality>

We would like you to build the best possible (given the time constraint) classification model for predicting the wine quality based on the provided physicochemical tests. Please employ standard data science techniques to partition your dataset and evaluating the classifier performance. You can use any modeling approach and import any standard library within R or Python for this purpose. Do provide some discussion and insight into such choices in your report. Outline next steps and approaches you might adopt given more time to improve upon your initial effort. How do you see the wine producer and/or consumer communities possibly using such a model?

**General Stipulations**

* You have a full day (24 hours) to complete the project and turn in your artifacts. Please make sure that your email response is within 24 hours of receipt of the email request. Please coordinate with ON3 as to the best day and time to deliver the project for you to work on. However, we do not expect you to spend more than 5 to 8 hours on the problem.
* You can choose to complete this project using either R or Python. Please use the latest versions of any of these languages.
* Please submit your results as a Data Science report along with an R script or Python notebook (and README file) so that we can run your code and verify / duplicate your findings and insights.
* There are no right or wrong answers. We simply want to evaluate your approach to the Data Science problem and ability to write a preliminary report synthesizing your work.
* Please ensure that the submitted report and code is entirely your own effort.