**Question 6**

**Part1**

* Columns 0 – 150 represent input variable and ‘target’ represents target variable
* Goal is to use the input variables to correctly identify or predict target variable
* Usage of Python and Jupyter notebook for completing this exercise is preferred but if you are not comfortable, feel free to provide source code and summary in email
* Please summarize your thought and analytics under following sections
  + Any pre-analytics steps and your understanding of data before you perform the analytics
  + Your choice of analytical algorithms, various steps taken during analytics, any comparisons between other algorithms and your understanding of how this model is performing
  + Your final choice of model and summary

**Part2**

* The goal is to operationalize your model in Part 1 to run both inference and retraining pipelines
* You should use Python as main programming language to complete this exercise
* You may introduce various technology stack(s) to supplement your Python code as appropriately (as pseudocode), please share your reasons for doing so
* You can assume that the incoming new data will be available in the structured database readily
* Write the inference results in Part 1 into a structured database (preferably Microsoft SQL database)
* Please summarize your thoughts on the following:
  + Any anticipated pipeline exception/weaknesses and how you plan to handle these exceptions
  + Your proposed monitoring metrics for the analytical pipeline health, as well as potential troubleshooting steps
  + Any other pipeline components which will help to achieve MLOps

*Remarks:*

* *You should use Python as main programming language to complete this exercise*
* *Usage of Jupyter notebook for completing this exercise is preferred (Part 1) but if you are not comfortable, feel free to provide source code and summary in email*