Data Science Research Project

The Data:

1. This data represents a technology which is in a remote location. It sends data periodically on its status – failure or not. It is very expensive to fix these machines because they are in remote areas and its failure is hard to predict. It is your job as the researcher to help us predict failures which may occur in the future
2. Historical file : predict\_failure.csv

The Task

1. Research : Please create a predictive model using a jupyter notebook with the python 3 language.
   1. Highlight your thought process in the notebook as you explore the data
   2. Showcase your best predictive model – and highlight why it’s the best
   3. Evaluate your model
   4. *Please only show relevant items in the notebook – not super long please!*
   5. Please additionally provide a PDF file from the notebook
2. Coding : Please package your best model into one ***class*** (or more) that will
   1. Train the model using the data provided
   2. Make predictions on a similar CSV input file
   3. Return the entered CSV file with probability of failure column appended.
   4. Also
      1. Provide your dependencies
      2. If using anaconda please provide conda yml file
      3. If using virtual env please provide requirements.txt

If you have any questions, please send them to christopher.brossman@therealreal.com