**Problem:** Explain in detail the pre-process works you would do in order to a prepare a raw, unstructured text data into a form that Python and machine learning model can understand. In other words, what specific tasks will you perform to put your native data into a state that will be ready to be used as input into a machine learning system? Again, walk-us-through what you would do from beginning to end, in details, to clean up a raw data to put it into its final form so that it will be ready for machine learning operation. Your answer should be no shorter than one-page, single-line description, and an accompanying diagram. Page size limit does not include the diagram. The solution to this problem should be in your own words, and not cut-n-paste from the Internet. Include a bibliography of any sources that influenced your explanation. There should be ***no*** ***coding*** in your answer. You may consider reviewing what we have discussed in the past seven weeks.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| GRADING RUBRIC | | | | |
| Grading Criteria | **Exceeds**  *Excellent*  Epic Wow | **Meets**  *Satisfactory*  O.K. | **Partially Meets**  *Below Expectations*  Not Yet | **Does Not Meet**  *Unacceptable*  Fail |
| **Connection and comprehension** | +35 - You exhibit excellent comprehension on the topic of data preparation prior to machine learning operation | +27 - You showed satisfactory comprehension on the topic of machine learning pipeline. | +15 – The connection and comprehension on data pre-processing for machine learning is “not yet” ready. | Unfortunately, no evidence of comprehension on the topic of scrubbing data so it will be ready for machine learning. |
| **Presentation** | +10 - Documentation was robust, showed design details including diagrams, and the “hows” were explained. | +8 - The document is okay but has room for improvement. Design details and the “hows” is satisfactory. | +5 – Documentation and design details are below expectation. | Unfortunately, no documentation, or no design details. |
| **Time Management** | +5 - You used time effectively and stayed within the timeframe allotted. | -5 Unfortunately, the work was not submitted on time. | -10 - Time management is poor and below expectation. | -15 - Unfortunately, this work is extremely late. |

**Solution:** Place your solution below on a separate page, and make sure your name is on your work before submitting to Canvas.