Project Report GRADED

Assignment | Scale: 100 pts | Weight: 25.00% | Submitted on: 10:24 AM CDT · Sat Mar 27, 2021 | Grades

Faculty Feedback

Score

100 / 100 (100%)

Your Feedback

Enjoy a well earned break!!

Student: Michael, Charles, Aiden

Presentation :

Topic: **Error Cases**
Slides:
NICE Slides !! Organization and clarity are key.
Motivation / Intro: **Great motivation here – *challenging and clear application!!***
YOU DID THIS – VERY WELL MOTIVATED!
Clear overview and flow:
Overall well presented!!
Good -- EDA and investigation!!
Good explanation of results and conclusions!
Subtotal: 25 / 25

Report:

Problem Statement: A VERY challenging and ***"real world" Project!!***
Data and Cleaning: Great choice of data set … it hard to find good (and free) data**!**
EDA: EDA is a very important part of any data driven process. Blindingly throwing machine learning solutions at data can lead to ill-conceived interpretations and recommendations.
GREAT work with your EDA – very EXTENSIVE. AWESOME SUPPORTING VIZ
Methods / Analyses:
Great NB, ARM, DT, RF, SVM and viz!!
Great investigation and experimentation of a variety of parameters for the models – AWESOME!!
Great word clouds and ROC curves (my fav)!

Conclusions: This is the end result. Key decisions and recommendations are often gleaned from the

conclusions! What did you learn? What recommendations / results were made and observed? Clarity and accuracy are key here.
Good clear conclusions! This was truly an excellent effort and well executed! An amazing overall submission – thank you!!

Subtotal: 50/ 50

Code: Code was provided and written by student following the instructions given.

Subtotal: 25 /25

GRAND TOTAL: 100 /100

GENERAL NOTES

Analysis
Subsection 1: The Data
This is where you talk all about the data, the variables, the cleaning, measures of cleanliness, etc.
Even if your data is clean, you must write code to clean it (pretend you cannot see that it is clean).

Minimally: NAs, changing things to factors as needed, outliers, discretization, incorrect values, such as .23 for age, etc.

Have a sub-subsection for each variable and show that you looked at all the criteria noted above.

Show AND MEASURE the before and after. For example, if you find that a variable has 10 missing values and you update these with the mean, then the before is the mean before, and the after is the mean after. In this case, you should also include the variance before and after. The measure is *very* dependent on what you clean and how you clean it. So this will be for you to think about.

Subsection 2: EDA: statistical and visual. Create a vis and/or table for each variable in the dataset.

Subsections 3 - n:
Here you run all methods learned in this class. One subsection for each. Include tuning and/or different options as applicable - such as different kernels and C for SVM, different k for kmeans, etc.
Explain as you go. Pretend you are writing a tutorial paper.

Extra Grading Notes for Analysis:
If there is not a clear subsection (and sub-subsections) that are all about the data, dataset, data cleaning, variables, EDA, visual EDA, prep, etc. (up to -15)
If there is not a subsection dedicated to each required methods noted above. (up to -10 for each missing method).
If there is not at least one vis for each required method in the Analysis area and within the subsection for that method. (-5 per missing vis)
If the model, model parameters (such as kernel for SVM), etc. are not well explained points can be lost.

If a model or method is not correctly applied or gives results that should not occur, points can be lost.

Results
Subsections 1 – n
You will have and will discuss results, issues, and limitations for all the analysis. You will note which ones worked well, and why, which ones did not, and why, etc.

GRADE NOTES:
If any method is missing. (-10 per missing method)
At least one results-supporting vis for each method. (Missing a vis – 5). These are not the same as the visualizations required in Analysis.
The Results section will have a subsection for each method. Each method subsection will be different in the sense that each method is different, has different parameters, etc. For example, when you do kmeans, you will talk about the different values of k, how each performed, what each revealed, and which was selected and why. Results talk about what happened in each method, but technically.

Conclusions:
3-4 paragraphs - NON-TECHNICAL. What was the outcome - what did you find, discover, predict, classify? WHy does your topic matter to humans? What do the results mean to humans?

GRADE NOTES
Less than 3 paragraphs (-10)
Any mention of technical results (which belong in the Results above) -10
A lack of flow or clarity about what actual findings were, why they matter, who they matter to, how they can improve things, etc…

A note about communication
1) Effective communication - via presentation and writing - is critical.
2) In life, we are often judged and measured through these avenues.
3) When writing technical, academic, and/or professional papers (assignments and projects):
- Avoid speaking in the first person. Avoid “I”, “you”, “me”, “we”, etc.
- Include important R code and visualizations throughout the paper. This does not mean that you should show every line of your R code (please do not), but rather you can include critical lines that show models and methods.
- Include R results in your Assignments and Project that support your models, methods, analysis, and results. Do this smartly.
4) Proof-read all submission out loud first. Be sure your assignments/project have a clear flow and are easy to follow and understand.