Detecting Malaria Rubric

DS 4002 - Spring 2023 – Amshala Bharathan
Due: TBD
Submission format: Link to GitHub Repository (Canvas)

Individual Assignment

General Description: Submit to canvas assignments a link to your GitHub repository for this project.

Preparatory Assignments

- Class discussion and readings for the case study

Why am I doing this? The goal of this assignment is for you to gain practice in identifying what models to use given a problem, building models, and evaluating models. Having practice with training models and trying out different hyperparameters will help you practice your machine learning skills. The graphs you produce and present will help develop your visualization skills. This project is intended to put as much of your data science skills as possible to use.
- Course Learning Objective: Visualizing data
- Course Learning Objective: Building and Evaluating Models

What am I going to do? You will begin by downloading the provided dataset and reading the 'hook document'. After reading the hook document, you should start brainstorming ideas for a modelling approach and start researching which model would be most appropriate to use. Perform exploratory analysis on the data and upload any figures you produce to your git repo. Then, handle any preprocessing work necessary for the data and split the data into train, tune, and test sets. Afterwards, you will build a model using the modelling approach you have selected and try different combinations of hyperparameters to get the best model possible. Make sure to print out evaluation metrics and keep record of the different accuracy and f1 scores you get. Once you think you cannot improve the model any further, you will upload your code to your git repo and add all your figures to the repo as well. Write your analysis paper and submit the link to your git repo on canvas.

Tips for success:
- Make sure to thoroughly research before starting to code.
- Take breaks—if you feel stuck, some time away from the assignment may be extremely helpful.
- Remember you have limited time: since you only have three weeks to complete this assignment, you may not be able to do as well as you could if you had more time. Do the best you can given the time constraint.
- Ask your peers for advice. Although this is an individual assignment, talking with peers to get some ideas rolling around is encouraged.
- Use your resources: The professor and TA are here to help. Ask them any questions you may have.

How will I know I have Succeeded? You will meet expectations on Case Study: Detecting Malaria when you follow the criteria in the rubric below.

Spec Category	Spec Details
Formatting	- One Github Repository (submitted via link on collab) - The top level page should contain - A README.md file (which auto displays) - A LICENSE.md file (use MIT as default) - A SRC folder - A Analysis Paper - A FIGURES folder
README.md	- Goal: This file serves as an orientation to everyone who comes to your repository, it should enable them to get their bearings. - Use markdown headers to divide content - Make an H2 (##) section explaining the contents of the repository - SRC section - Make an H3 section for Installing/Building your code - Make an H3 section for Usage of your code - Analysis Paper - Write and executive summary describing what the analysis paper is - Include a link to the paper - FIGURES section - Include a table of contents that describes the figures produced and what you gleaned from them - Use markdown table formatting - REFERENCES section - All references should be listed at the end of the Readme.md file (Use IEEE Documentation style (link)) - Include any acknowledgements
LICENSE.md	- Goal: This file explains to a visitor the terms under which they may use and cite your repository. - Select an appropriate license from the GitHub options list on repository creation. - Usually, the MIT license is appropriate.
SRC folder	- Goal: This folder contains all code you used for this project - Include code for making your graphs - Include all preprocessing, splitting, training, and evaluating code - If you use some code from another source, cite it in your code block and thoroughly comment the code to explain what each part does
Analysis Paper	- Goal: Summarize your findings - Talk about the results and what you gleaned from them - Explain what went well in your project - Explain what you can do better on next time

	• Format: Two paragraphs
Figures folder	• Goal: This folder contains all of the figures generated by your project
	• This should include any exploratory plots you make
	• Include a chart depicting the accuracy rate and f1 score for each hyperparameter combination you try
	• Clearly label and explain all figures
References	• All references should be listed at the end of the document
	• Use IEEE Documentation style (link)