**Artifact 1 – Software Design and Engineering**

**Artifact Description**

This is a Jupyter Notebook in which I built an AI that uses natural language processing (NLP) techniques to be able to read a tweet and classify whether the tweet is about a disaster or not.

**Selection Justification**

I selected this artifact as NLP implementation is a very in demand skill in the world of machine learning and artificial intelligence. In this notebook I was able to utilize my skills in NLP to create an AI that had a high accuracy rate, and the ability to do this is a great skill to have for any future machine learning engineer.

This software design and engineering enhancement was focused on improving this software and expanding its complexity. This was achieved where I implemented data preprocessing, visualization, and model selection/tuning. Data preprocessing in this project involved cleaning and shaping the data to get it ready for machine learning models. This relates to how when designing software, data must be structured in way that maximizes performance, scalability, and maintainability. I implemented an algorithm to tokenize, transform, clean, and batch the training data in the most efficient and maintainable way, just as when a software engineer is designing algorithms to create the most efficient software possible. This improved the software by preparing the data in a way that decreased the training time needed for the models and improved their overall accuracy.

Regarding visualization, I created graphs to give a visual representation of the data’s features, and in the enhancement, I created visual cards that highlighted tweets which were classified correctly, which tweets were classified incorrectly, and how confident the model was in its selections. These visualizations relate to the field of software design and engineering as software developers very commonly employ visualizations in their work. They improved the project by making the actions being taken and assessment of the performance of the models much more apparent and easier to understand.

The process of enhancing the project through these visualizations also expanded the project’s complexity. Before the enhancement, the notebook primarily just contained the training of the models and exporting of the results. Now, new functions were created to fully understand the model’s performance. The new analyze\_predictions function separates the predictions into lists of true positives, true negatives, false negatives, and false positives. In addition, the visualize\_prediction\_card function created detailed visual cards of the now organized predictions utilizing the matplotlib library. This organization and clear visualization of the predictions made the project more complex while adding vital insight on why the model was getting different predictions wrong. It was through this that errors in the training and testing data were discovered which open the door for improving the model’s performance further through cleaning of this data.

Finally, the model selection and tuning which was done in this notebook runs parallel to a software developer’s use of employing libraries for specific tasks. Just as a software developer selects different coding libraries based on their suitability for the problem’s domain, code scalability, and ease of integration, I did the same when selecting the models used in the transfer learning section of this notebook.

Overall, this project and the enhancement to it showcased by skills in deep learning, data visualization, use of diverse coding libraries, and writing scalable code.

**Itemized List of Specific Enhancements**A) Data Exploration

* Created read\_csv\_with\_check function to validate existence of CSV files before importing data frames.
* Added markup text analyzing Count of Target Values graph.

B) Preparing Data

* Added markup text analyzing Distribution of Sequence Lengths graph.

C) Analyzing Metrics (new section)

* Created graphs analyzing accuracy metrics of the trained models with analysis text.

D) Analyzing Predictions (new section, target enhancement)

* Created a analyze\_predictions function to separate predictions into true positives, true negatives, false positives, and false negatives.
* Created visualize\_prediction\_card function to create detailed visual cards of the now organized predictions.
* Added analysis of most/least confident predictions in each of previously mentioned categories. Printed these predictions and hypothesized why the model was getting these wrong, added input for possible future improvements of training data.

E) Conclusion (new section)

* Added final remarks summarizing what was found in notebook.

**Reflection**

Enhancing and modifying the artifact was a great experience! Creating the visual cards gave me a lot more practice with the matplotlib library, and adding additional docstrings helped me feel like I was writing more professional and easier to understand code. The creation of the cards was the biggest challenge as I had to get more familiar with the matplotlib library to get them looking like how I had them in mind. Through doing this, I was able to get the tweets to have a blue background which I previously was unfamiliar with. I also got more practice spacing elements out on a card in matplotlib. Matplotlib is a fun library to experiment with and I got to see how the options for what can be created are quite vast.

The course objectives which I planned to meet for this assignment were to “employ strategies for building collaborative environments that enable diverse audiences to support organizational decision making in the field of computer science” and to “design, develop, and deliver professional-quality oral, written, and visual communications that are coherent, technically sound, and appropriately adapted to specific audiences and contexts”. I feel I met the first objective through the addition of my enhancement as well as through the further analysis and explanation given to the already existing visuals. These additions made the notebook much more accessible to others as it made it made it much more apparent what the training data’s features were, as well as the strengths and weaknesses of the final model. This would allow other developers to collaborate with the project and easily be able to identify where improvements and suggestions can be made.

I believe these same improvements also helped me reach the second previously mentioned objective. The enhancement and additional analysis allow others, including those not trained in machine learning, to be able to make sense of the model’s input and the training process using detailed visualizations. The new conclusion section also allows another person to easily tell which model ended up being the best to use and what the notebook accomplished in a written format. While someone who is familiar with machine learning may get more out of the notebook, it is now more accessible to an audience without training in this area. This allowed me to exhibit the ability to tailor presentations to both technical and non-technical audiences. The remaining three course objectives will be addressed by the two other artifact enhancements.