Team Roles:

In approaching the project, our team divided our responsibilities to maximize efficiency by playing to each team member's strengths. Minh took the lead in sourcing and preprocessing the data, ensuring its quality and relevance for our machine-learning model. After he found lots of data for us to consider, we discussed the best choice as a team and decided that we could make a functional and quality model using the Goodreads data. He also performed the EDA work for the model and discovered duplicates in the data. Allie focused on the writing aspects of the project, analyzing the data and model by creating our model card and datasheet, and READ ME file. She created some of the initial functions, variables, and testing for our model, adding notes throughout. Minh continued with the Model building until it was functional, and he also performed tests to ensure its accuracy. Gabe overviewed all aspects of the project, adding notes to the model, EDA work, and model card. He also completed the "READ ME" portion of our model.

Major Hurdles:

The project presented several challenges, each of which we tackled collaboratively. After performing EDA work, we discovered duplicate books assigned to different genres and summaries in our data. We had a team discussion on what to do with the data and eventually decided to keep the duplicates. We decided that the decision on which duplicates to delete was too subjective, and keeping them all would keep our model as accurate as possible.

Our team had to navigate busy schedules due to finals. We found it difficult to find times to discuss our project, and it was difficult to complete each aspect of the project during finals week. To create an open line of communication, we created a discord server where we could share calls, new info, and ideas. We also gave ourselves tasks and deadlines for different aspects so we could pressure each other to finish on time. Because of this proactive planning and strict timelines, we managed to stay on track and meet project deadlines.

Becoming familiar with the code provided in class posed an initial hurdle. None of our team members had ever created a model before. To familiarize ourselves with

model creation, we reread the 5.1 assignment multiple times. Using information from this notebook gave us helpful insight into perfecting our model and its accuracy.

Data Insights:

The project provided us with valuable insights that extend beyond technical aspects. I n our virtual classroom, data breakdowns were provided in our notebooks by Dr. Lindgren. During this project, we had to independently explore and understand the data. This process helped us appreciate the importance of tailoring our model to the specific characteristics of the dataset. This was especially important during the EDA work on our model, and the Data analysis.

The inclusion of multiple labels for the same book prompted a realization: our goal wasn't just about achieving the "perfect" classification but ensuring accuracy. This insight emphasized how important it is to customize your objectives to your data. The nuances of the data played a big role in our team's decision to keep duplicates and remain accurate to Goodreads information. Working with one's data and its flaws was a key lesson that will shape our future coding projects.

Coordinating on GitHub during a busy period required lots of digital communication. Through the creation of our discord and scheduled virtual meetings, we honed our asynchronous collaboration skills, enhancing our ability to work together efficiently despite our difficult schedules.

In conclusion, our team project served not only as a technical endeavor but also a conduit for personal and collective growth. Each of us played a vital role, and the challenges we faced together strengthened our collaborative skills. The insights we gained into digital media will help us not only in the realms of data science but analysis and digital communication as a whole.