Sam Miller

CS 584

HW<sub>3</sub>

My experimental setup for this homework involved using Python with sklearn to process the "20 news group" dataset and evaluate the performance of a Naïve Bayes classifier on differentiating between documents labelled as 'motorcycle' and 'auto.' The first step involved fetching the data from sklearn using the sklearn.datasets.fetch\_20newsgroups and sklearn.datasets.fetch\_20newsgroups\_vectorized functions. By supplying the former function with the list ['rec.autos', 'rec.motorcycles'] as the optional categories parameter, only documents that were labelled with categories were fetched. I then used these documents to identify which documents I needed to filter out from the vectorized data that is returned by the second sklearn function. Additionally, I supplied these functions with the subset parameter of 'train' or 'test' to retrieve the dataset already partitioned. Finally, using sklearn.naive\_bayes.MultinomialNB class, I fit a Naïve Bayes classifier to the training set and evaluated its performance over the testing set using F1 score as the main metric. The F1 score returned by this code was 0.8970518199910767 which indicates high accuracy and a balance between precision and recall for the Naïve Bayes classifier for this task.

In the second part of this assignment, I took the first four misclassified documents from part 1 and fed them to ChatGPT with the following zero shot prompt:

"You are an expert at multiclass classification tasks, specifically with text documents. You are analyzing the standard 20 news group dataset, specifically looking at documents that are classified as 'motorcycle' or 'auto.' You are asked to classify the following documents into one of these two categories, or say 'Could not classify' if you can't tell what it is."

The documents themselves were appended to the end of this prompt. They are, in order:

## Document 1:

Talk to Philip Greenspun. He took Ford to court recently and, despite much manouvering and trickery on Ford's part, he won! Well, actually I think Ford settled out of court on the provision he shut his mouth and stopped causing them trouble. I love it when the little guy wins. I don't have Philip's address anymore, but a "Philip, where are you" call may bring him out of hiding. Cheers, Paul.

## Document 2:

*>>>>>>>>	>
-----------	---

## Document 3:

I had the same inconvenience when I bought my EX-500; only I could at least go 50 mph during the initial break-in. Despite the high quality of motor oils and motorcycle engines these days, I would follow the instructions, and just keep to city street riding for the first 500 miles. Heck, those miles go by fast, and it's worth it to know you haven't possibly screwed up your engine. At the same time, this has all the makings of a 6-week-long thread debating the whole break-in topic.

## Document 4:

released here we go again... now these are just rumors.. so dont quote me. New Integra supposedly wedge shaped again. 175 hp and all-wheel drive in top models. Then a variant called the zx-r comes later. (roadster?), i think it gets unveiled at end of summer.

ChatGPT replied with the following answer:

"Given the content of each document, I'll classify them as follows:

- Document 1: This document discusses a legal matter involving an individual and Ford, which is an automobile company. The mention of Ford directly links the content to automobiles. Therefore, this document would be classified as 'auto.'
- Document 2: This document contains very little information and does not provide any context that can be directly linked to either motorcycles or automobiles. Without any relevant information, this document would fall under 'Could not classify.'
- Document 3: The discussion revolves around the initial break-in period for an EX-500 motorcycle, mentioning motor oils, motorcycle engines, and specific riding advice. This document is clearly related to motorcycles, so it would be classified as 'motorcycle.'
- Document 4: This document talks about the New Integra, mentioning horsepower and all-wheel drive, which are characteristics related to automobiles. The document also hints at a future unveiling, which is common for new car models. Therefore, this document is classified as 'auto.'"

Therefore, ChatGPT was able to correct the mistakes of the Naïve Bayes model by accurately labelling documents 1 and 4 as 'auto' and document 3 as 'motorcycle'. It also could not identify which category document 2 should belong to, which makes sense upon close inspection of the content of that document.

I also asked Claude the same question with the same prompt. Its response is as follows:

"Document 1: Could not classify. This document seems to be discussing legal issues with Ford, but does not clearly relate to motorcycles or automobiles.

Document 2: Could not classify. This is just a short textual expression without any clear relation to motorcycles or automobiles.

Document 3: Motorcycle. This document discusses breaking in a new motorcycle engine (EX-500 model) and following the recommended procedures during the initial mileage.

Document 4: Auto. This document appears to be discussing rumors about a new Acura Integra model, which is an automobile, not a motorcycle. It mentions details like horsepower, all-wheel drive, and a potential roadster variant."

Claude agreed with ChatGPT on documents 2, 3, and 4. For documents 3 and 4, it correctly identified the proper label. But for document 1, unlike ChatGPT, it could not ascertain that this document belonged to the 'auto' category. Based on these results, ChatGPT may be slightly better suited for this task than Claude. Both LLMs were able to label these 4 test documents more accurately than the Naïve Bayes model, but this sample size is likely not sufficient to say that they are definitively more accurate than the Naïve Bayes model for this task.