Resolving Complaints

Much of a persons opinion about a company if formed by how the company responds to complaints about a product. The speed at which the complaints can be resolved depends, in part, on getting the complaint to the right person who can resolve the issue. The Consumer Complaint Database is a collection of complaints about consumer financial products and services that we sent to financial companies for response. The dataset CompanyComplaints.csv is a subset of that database (there are about 1.5 million complaints in the original database which you can download if you are interested) and contains complaints sent to companies via email along with which department the complaint was ultimately sent to to be resolved. Ordinarily, companies will hire specialists to sort through the complaints and get them to the right person but in this analysis you will be using machine learning to try to deduce which department the complaint should go to just based on the text of the complaint.

Notice that the dataset does not contain any "explanatory variables" as we are used to. Rather, you will need to build your own set of explanatory variables based on key words or characteristics of the complaint text. Be creative. There is no lower (or upper) limit on the amount of explanatory variables you are allowed to create. Further, you will NOT be graded based on the accuracy of your predictions. Rather, in the "describe methods/models" section of your report you will need to describe and justify what explanatory variables you created. If I feel you did not put much effort into building explanatory variables you will be docked points accordingly.

In your analysis, make sure you answer the following questions:

1. How accurately are you able to classify complaints?
2. If your chosen method allow for it, what key words or symbols or explanatory variables were useful in classifying complaints?
3. Are there departments that are commonly confused? That is, do you commonly classify complaints as going to department "A" when it should go to department "B"?
4. The file WhichDepartment.csv contains information on a few complaints that recently came in. What department(s) does your model think they should be directed to?

Audience: Head of data science and head of customer care (so feel free to speak technically AND know the audience really cares about implementation/accuracy/etc.)