Islamophobia Detector

The tasks we have accomplished so far are data collection, labeling, and model development. Each team member was assigned one of the following platforms: Reddit, YouTube and Twitter. 1,000 tweets that contained either “#Islam” or “#Muslim” from the past seven days were scraped with the Twitter API (the code that does this is here). For reddit, a list of relevant subreddits was gathered by searching for keywords and exploring topics related to Islam. The Python API PRAW, was used to develop a script which takes the URL of the subreddit and scrapes its comments and sub comments. The script results in an excel file with the comments, and the date of publication. We also created a form using Google Forms to collect data manually from YouTube. We searched for videos discussing topics related to Islam. Then, we examined the comment section and looked for comments that discuss islam in a positively or negatively. Each comment was manually examined to ensure only relevant comments are included in the dataset. Ironic text, comments that were vague and difficult to identify as Islamophobic, and poorly written comments were removed. A total of a 1,000 were labeled as either Islamophobic or not Islamophobic.

The files collected from each of the platforms will be combined into a dataset. The data will undergo preprocessing which includes removing unnecessary characters, hashtags, and usernames. Data will be exported to prometheus and loaded by grafana, to create visualizations. The dataset will be split into testing and training sets and loaded into the classifiers. The models will be tested and evaluated, the best performing model will be used as the classifier. A program will be developed that allows the user to enter text, which connects to the model and results in a classified output. A user interface will be developed using JavaScript and python. The interface will prompt the user to enter text, connect to the classifier and display the result of the model.

Data collection had a number of challenges, one of the hardest being examining the comments. It is challenging in some cases to determine when the sentiment of the comment shifts from a person sharing their opinion to prejudice. It took a substantial amount of time to examine each comment to gather a diverse dataset. In some cases the majority of the comments under a post did not meet the requirements. Which meant reading hundreds of comments to only have a small amount being fit for the dataset. Also, finding comments on YouTube was a time consuming process because we had to find videos and then scroll through the comment section and locate comments manually.

In addition to data collection, we also created code for the text classification models. The prediction is binary because it is either islamophobic or not.We have three different types of vector embeddings: Keras, BERT, and tf-idf. We also have three

different types of models: SVM, naive bayes, and logit. The code for these models can be found here. Once we finish deciding how to approach text pre-processing (removing hashtags from tweets, stemming, etc.), we will estimate each type of model and compare the metrics (f1-score, precision, recall). Whichever model performs the best will be the one we choose.

Members:

Nelofer	Arjumand	narjum2Illinois.edu	leader
Emily	Sallenback	emilygs2@illinois.edu	
Mohammed	Alabdullatif	maa27@illinois.edu	