1.) Write a brief outline of your research plan:

The main idea of the project is detecting controversy in news articles by leveraging online comments and tweets. A method is already implemented for scoring sentences based on the likelihood of being controversial, as well as a simple web interface for presenting a list of predefined topics and their respective new articles, ranked based on controversy scoring and accompanied with relevant tweets.

The next step would be to improve the web interface, so that users can log in, perform keyword search and create a timeline for the chosen keyword. The system will crawl relevant news articles and tweets for each day, perform controversy scoring and present the most controversial news for each day, as well as a timeline that will show daily controversy for each keyword searched.

The system will also facilitate user feedback, so that users will indirectly will be able to label news articles as controversial or not, by clicking on a news article. Such feedback will facilitate an online learning back-end procedure to increase the accuracy of the scoring functions.

If time permits, we will also work on improving the controversy scoring function, either by implementing more features or developing a novel scoring function.

2.) What do you hope to learn and achieve by the end of the semester?

By the end of the semester, I will be able to learn basic text mining methods, understand how statistical metrics are used in text mining, how to leverage online social media for novel applications, the differences between unsupervised and supervised learning, as well as basic feature extraction techniques.