Introduction
Sustainability and social responsibility are becoming increasingly important concerns for investors. While the industry has no consensus on how to define sustainable and socially-responsible investments, they generally fall under three categories: environmental, social, and governance (ESG). Important environmental concerns include greenhouse gas emissions, carbon footprint, and pollution. Social concerns include health and safety, workers’ rights, and, for software companies, data security. Governance concerns include CEO pay and corruption.

This project consists of a natural language processing approach to measuring ESG performance using annual corporate 10K reports filed with the SEC. It compares 10Ks to standards for reporting on sustainable business practices. It calculates positive and negative ESG scores for each (10K, reporting standard) pair.

Materials and Methods
The Global Reporting Initiative publishes a set of reporting standards that describe how organizations should report their impact in areas related to sustainable investing, such as the economy, the environment, and society. The SEC makes annual 10K filings available online in text format through the Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system. I used these 10K reports and the 2018 GRI Reporting Standards as input to the ESG scoring system.

The program uses a TF-IDF vectorizer model to extract feature vectors for calculating document similarity. It splits the 10K into sentences, extracts feature vectors, and calculates the similarity between the sentences and the GRI reporting standards. It then performs sentiment analysis on the similar sentences to determine how frequently the 10K speaks positively and negatively about the topics covered in each reporting standard.

Results
Textron, the munitions manufacturer, scored twice as high on positive occupational health as on negative occupational health. In the case of Tesla, its positive energy score was nine times as large as its negative energy score. Similarly, Tesla’s positive emissions score was twice as large as its negative counterpart. For both of these companies, the positive and negative ESG scores in relevant areas track the company’s expected performance.

However, the energy and emissions scores for FirstEnergy Corp., the coal company, complicate the proposition that we can simply compare a company’s positive and negative scores to each other to estimate its ESG performance. A comparison of positive scores to negative would indicate that FirstEnergy performs well in energy and emissions, the opposite of what I expected from a blacklisted coal company.

As an alternative, I calculated the ratio between FirstEnergy’s scores and Tesla’s. While FirstEnergy scores higher than Tesla for positive energy, it scores comparatively higher for negative energy. This is as expected.

Conclusion
The combination of TFIDF similarity and basic sentiment analysis produced a model for roughly capturing a firm’s ESG performance. More work is needed to refine the design, specifically in synthesizing the data produced by the text-analysis model into a useful metric. Potential developments include using more sentiment categories than just “positive” and “negative”, incorporating semantic analysis into the sentiment scorer, and compiling a set of baseline companies against which to compare scores in each category.

Acknowledgement
Project advisors: Dragomir Radev and Bryan Kelly

Figure 1. Positive and Negative ESG scores for 10Ks from 2017.

Figure 2. Text from the GRI reporting standard on water and effluents.