A Tool for Businesses for Analysing Online Reviews

Author: Ben Ryles

Supervisor: Dr William J. Teahan

Aim and Objectives

The primary aim is to develop a tool that performs sentiment analysis on online reviews that is easy to use for non-specialists. To achieve this aim, the following objectives need to be fulfilled:

- **Develop** a web scraper to scrape online reviews.
- **Develop** sentiment analysis software for analysing online reviews and classifying customer feedback with a user friendly graphical interface.
- **Implement** a tool using Python that integrates the web scraper and sentiment analysis software.
- **Evaluate** the sentiment analysis software that the tool uses against standard datasets.

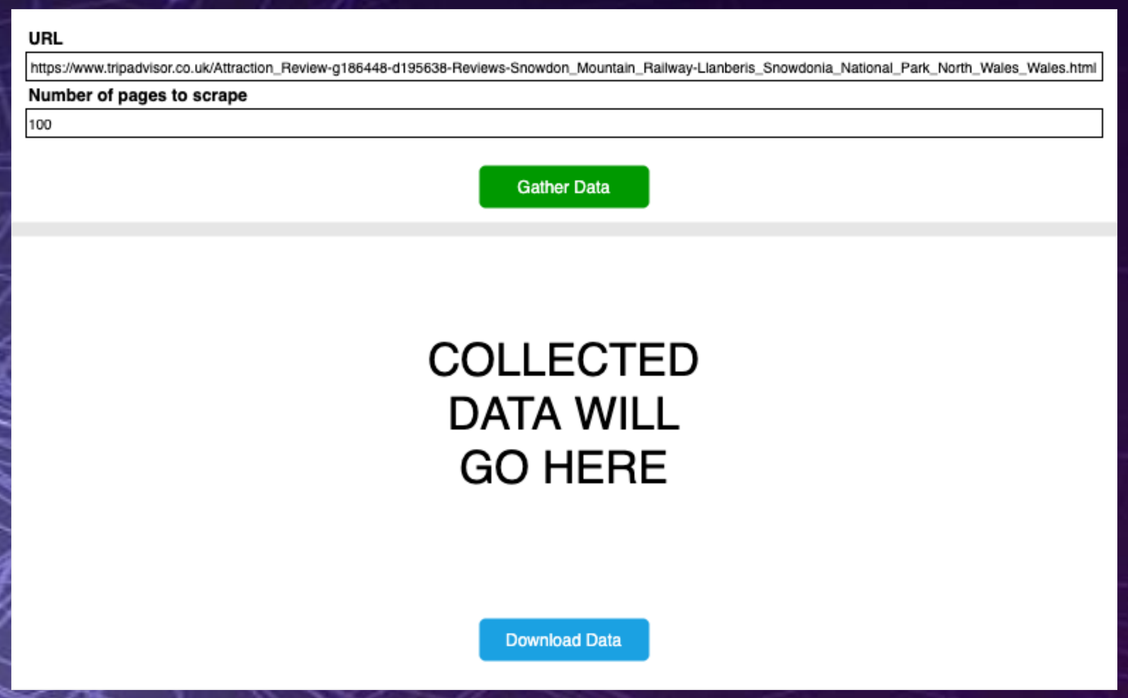
Background

Sentiment analysis is a technique used to detect opinions in text. Sentiment analysis is an interesting intersection between natural language processing and information retrieval.

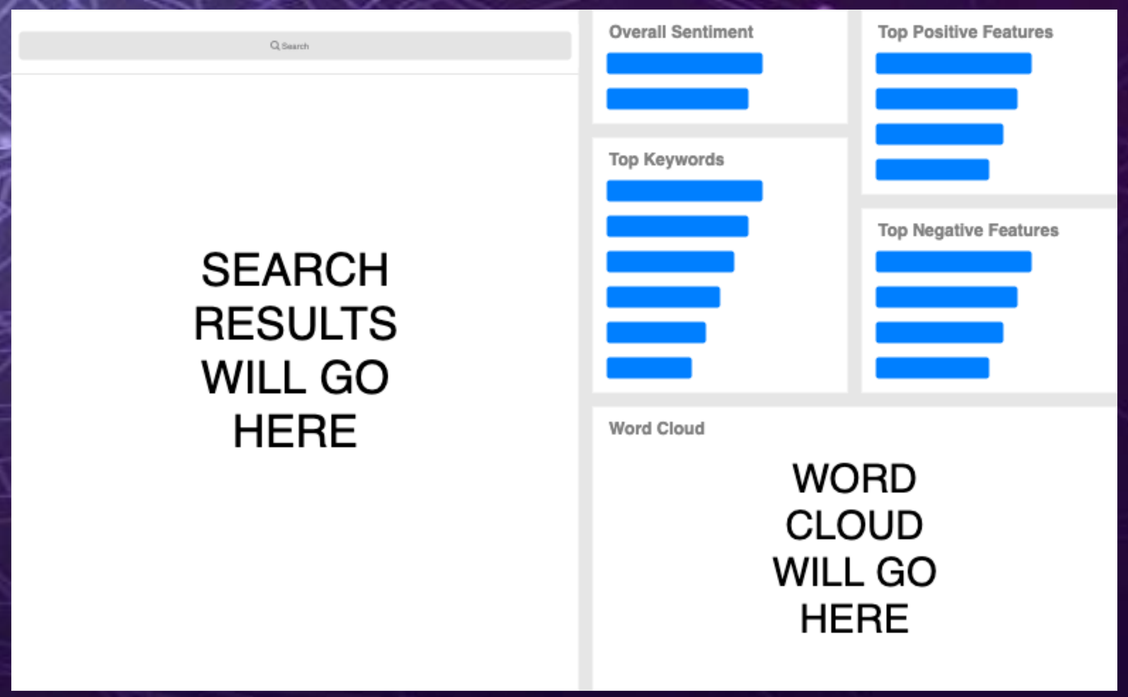
Classifying a text as either positive or negative is the most basic task of sentiment analysis. More fine-grained approaches allow for determining the sentiment of specific features.

Methods

The web scraper was built using Python and Selenium. The scraper extracts online reviews from TripAdvisor. Classifiers from Scikit-learn are trained and tested using the Yelp dataset. The Yelp dataset contains many reviews of businesses, therefore providing a good corpus to train on. The classifiers are then applied to the scraped data.



Mockup of the web scraper



Mockup of the sentiment analysis tool

Initial results from training and testing various classifiers

Experimental Results

Experiments were conducted using a sample of 10,000 reviews from the Yelp dataset. Classifiers from Scikit-learn were trained and tested using this sample. TF-IDF was used to create feature vectors. Precision, recall, f1-score, and accuracy are the metrics used to measure the performance of the classifiers. The results provide an accuracy ranging from 78% to 93%. These initial results indicate a good level of performance, the classifiers should perform well when applied to the scraped data.

Work Completed So Far

At this stage, the web scraper is functional, it can extract data from TripAdvisor pages and save data as a CSV file. The next step would be to add a graphical user interface. Various machine learning classifiers have been trained, tested and are ready to be used with data extracted by the web scraper. Initial results are promising, the classifiers should perform well on the extracted data.

Future Work

Implementing real-time analysis would make a great addition to the software. Users would be able to track sentiment over time and notice any changes. Additional web scrapers to extract data from more websites would allow for even better analysis. Interactive visualizations could also be added to further enhance the analysis provided to the user. Ensemble learning methods can also be used to obtain better performance.