Deception Detective NLP Project

Second Capstone Milestone Report

The problem

Identify Fake Opinion Spam

Consumers and businesses more and more rely on online reviews to make purchase and travel decisions. Positive online reviews can translate into significant gains in sales and consequently, profits. On the other hand, negative reviews can translate into lost sales, declining business, and decreased profits. Because of the high stakes involved for businesses, Deceptive Opinion Spam – “fictitious opinions that have been deliberately written to sound authentic, in order to deceive the reader”- is on the rise. Human readers are notoriously ill-equipped to discern these deceptive reviews, identifying fake reviews on a level with random guesses. Natural language processing algorithms are better equipped to correctly identify deceptive reviews.

This process will apply an end-to-end NLP flow to identify fake hotel reviews.

The Client

The clients are business owners who rely on accurate, trustworthy reviews to generate business leads and feedback for correction. No one profits from fake reviews, and they should be flagged and removed.

The Data

<https://www.kaggle.com/rtatman/deceptive-opinion-spam-corpus>

The corpus consists of 1600 truthful and deceptive hotel reviews of 20 Chicago hotels:

* 400 truthful positive reviews from TripAdvisor
* 400 deceptive positive reviews from Mechanical Turk
* 400 truthful negative reviews from Expedia, Hotels.com, Orbitz, Priceline, TripAdvisor and Yelp.
* 400 deceptive negative reviews from Mechanical Turk.

Adding Yelp Reviews

The final report will add additional positive and negative reviews from yelp. The model will be utilized to flag potential fake reviews in the yelp dataset.

Initial Findings

Initial findings achieved an 88% predictive accuracy using logistic regression and linear SVC models. The data was transformed into a TFIdf vector before running the models.