
Online Review Spam Detection

Tirth Patel, Harsh Kachhadia, Piyush Biraje

Department of Computer Science, North Carolina State University, Raleigh, NC
tdpatel12@ncsu.edu, hmkachha@ncsu.edu, pbiraje@ncsu.edu

1 Project Idea

Many online retailers use fake reviews to make profits. Online review spam detection deals with classification of user reviews as spam or not-spam. Initially we will use a hotel review dataset provided by kaggle for training our different classifiers. These classifiers may be developed with neural network, Decision Tree, linear regression, etc. We will compare their performance and select the best classifier.

We are planning to create a dataset of our own using web-scraping. We then plan to use the trained model on this dataset.

2 Dataset

<https://www.kaggle.com/rtatman/deceptive-opinion-spam-corpus?select=deceptive-opinion.csv>

<https://s3.amazonaws.com/amazon-reviews-pds/tsv/index.txt>

3 Software

1. **Language:** Python
2. **Online IDE:** Google Colab
3. **Other Tools:** Numpy, Pandas, Matplotlib, TextBlob, NLTK, Tensorflow

4 Papers to read:

1. https://www.researchgate.net/publication/325075174_Detection_of_fake_online_hotel_reviews
2. <https://arxiv.org/pdf/1903.12452.pdf>
3. <https://www.cs.uic.edu/~liub/FBS/opinion-spam-WSDM-08.pdf>

5 Team and work division

1. Tirth Patel - Neural Network Classifier
2. Harsh Kachhadia - Decision Tree Classifier
3. Piyush Biraje - Linear Regression Classifier

6 Midterm milestone:

Currently we plan to deliver 3 models by mid-term (Oct 9). These models are: 1. Neural Network, 2. Decision Tree, 3. Linear Regression. We will perform classification experiment with compare all these 3 classifiers and select the one with with highest accuracy for further task.