UNIVERSITY OF WESTMINSTER

INFORMATICS INSTITUTE OF TECHNOLOGY

BSc (Hons) Computer Science

Final Year Project

" Falso "

A fake review identification system for online products using supervised learning methods

By

Miss H. M. Nethmi Venu Jayawardena

2014107

w1534146

nethmi.2014107@iit.ac.lk

Supervised by

Ms. Abarnah Kirupananda

Signature of Supervisor: .....

April 2019

This report is submitted in partial fulfilment of the requirements for the BSc (Hons) Computer Science degree at the University of Westminster

ABSTRACT

Online shopping is the act of purchasing products or services through the internet. When shopping online, product reviews are one of the only methods of helping a buyer with their purchase decisions. Since many online sellers exist, the competition among them has led to the rise of fake online reviews to manipulate buyer purchase decisions. This is known as fake review spamming. By fake review spamming sellers can boost their own ratings to improve sales or degrade a competing seller's ratings to reduce their sales. Fake review spamming is a rising problem and many previous researchers have attempted to solve this problem using methods such as opinion mining, optimization algorithms and machine learning.

This research proposes a natural language processing and machine learning based approach to identify fake review spamming. Natural language processing is used for pre-processing textual inputs and extracting its features. Machine learning classifiers use these extracted features for identifying fake reviews. By comparing five supervised classification algorithms, it was possible to achieve a detection accuracy of 81% with a gradient boosting classifier called XGBoost. Using this classification approach, a fake review identification system for online products was developed under the name 'Falso'. This solution was designed to be used by the administrators of an online shopping application. It allows an administrator to view the authenticity of a single review or an entire dataset of reviews at once. Using Falso, the administrators will be able to control fake review spamming to make sure that buyers will not be tricked into incorrect purchase decisions.

Keywords: Machine Learning, Supervised Learning, Natural Language Processing, Boosting, XGBoost, Spam review detection