REVIEW OF RESEARCH JOURNAL

Business Intelligence (AMI23B)

DALARNA
UNIVERSITY

Critical Review Paper: Fake Reviews Detection through Analysis of Linguistic Features

- Faranak Abri, Luis Felipe Gutierrez, Akbar Siami Namin1 , Keith S. Jones2 , and David R. W. Sears

Reviewed By:

Maha Vajeeshwaran

h21mahna@du.se

SUMMARY:

The success and failure of the business are largely determined by online reviews. Customers first read reviews made by prior customers before purchasing goods and services. Posting fake and fraudulent reviews, on the other hand, might help or affect certain firms superficially. This research paper investigates how to detect fake reviews using natural language processing.

The authors emphasize that it is preferable to use fewer linguistic features in the classification problem in order to simplify the classification model, reduce the amount of noise and uncertainty introduced by less important features, and ignore correlated features and their potential interactions in classification modelling.

The authors investigated the efficiency of 15 linguistic factors in detecting fake online reviews in this article. More specifically, by using RF feature importance, RFE, Boruta, and ANOVA, they could be able to identify consistently the number of adjectives, redundancy, and pausality as the most important features in their research. The authors also conducted classification experiments using seven classifiers such as Decision Tree (DTC), Random Forest (RF), Support Vector Machine (SVM), Extreme Gradient-Boosting Trees (XGBT), Multilayer Perceptron (MLP), Logistic Regression (LR), and Naive Bayes (NB). By implementing the linguistic features using the spaCy 2.2.3 library), the MLP reports the highest accuracy with 79.09% with only 4 features.

OBJECTIVES:

The primary goal of this article is to investigate the effectiveness of linguistic cues in the context of fake reviews detection and to measure the significance and interactions between these linguistic cues and identify less important features by using the “Restaurant Data Set”.

CONTRIBUTION:

This paper helps merchants enhance their business in the E-commerce business world by providing valuable insights by identifying fake reviews. In contrast to the previous research, this paper studies the effectiveness of linguistic cues in transcribed textual data. Their results indicate that fake reviews tend to include more redundant terms and pauses, and generally contain longer sentences. The application of several machine learning classification algorithms revealed that it could be able to discriminate fake from real reviews with high accuracy using 15 linguistic features.

STRENGTHS:

The writers reviewed several previous research publications and conducted an appropriate study for this issue in order to fill in the gaps. This paper gives the proper reasons for the use of different machine learning methods and feature selection methods to perform the research. Furthermore, it used real-time reviews from legitimate users from online resources. Also, the work is well-organized and written in a clear and engaging way. The future research, visuals and results tables shown in the research paper give more information to the readers to perform the further research related to this topic. These elements helped in understanding the points of the paper better and made it more enjoyable to read.

WEAKNESS:

Firstly, this research paper did not address us well how their method can be utilized in different languages. Second, this report fails to identify techniques for using the research findings to assist the user to achieve their financial goals. Thirdly, in the discussion part with reference to the previous works, it mentioned that the highest performance and accuracy can be achieved by more features, but the authors performed with only 15 features. I feel more in-depth research is to be carried out to deploy their models. The highest accuracy shown by the MLP model is not sufficient for the deployment with only 4 features.

CONCLUSION:

Overall, this article examines fake reviews by analysing the textual linguistic features. This paper addresses a gap in many studies by providing results that have practical implications. However, by focusing on post-deployment stages and how the model may be distributed, this approach has several drawbacks overlooked in the world of E-commerce business. Even though the authors compared their model performance with the existing work, still further research on this topic is needed for effective utilization for the business. Many companies use fake reviews with the help of third-party agencies to promote their business and also to damage their competitor’s business, e.g., sellers from Amazon, Flipkart, and Ebay. It will add great value for many companies if authors discover the proper models to detect these kinds of fake reviews in different languages.