E-commerce proved its importance based on the fact where time is the essence. People are relying on e-commerce more than before. With e-commerce comes a huge amount of user feedback based on the products they buy. As the internet has become cheaper and easy to get, more people are getting connected through different social media and platform where they are expressing product-related feedbacks. With the rise of e-commerce, people are relying more on product reviews to get a clear view and user experience. But there is no convincing way to authenticate the reviews posted on products on e-commerce websites. To generate more revenue and fulfill some immoral benefits, some sellers are making investments and hiring people to post fake reviews. These fake reviews are generated to convince people to buy the product. To detect these fake reviews, several methodologies were introduced. Most of the models are supervised models which rely on pseudo fake reviews or large scale labeled dataset. In this paper, a model has been proposed with a new technique which combines two different types of learning methods (active and supervised) by creating a manually labeled dataset. This model has 4 different filtering phases that are based on TF-IDF, Countvectorizer
and n-gram features of the review content and then Principal Component Analysis to reduce the feature set. It achieves a very encouraging result while working on 2000 reviews from Amazon. In the best case precision, recall, and f-score are slightly above 91% and the accuracy achieved is up to 90%. After comparing the results with similar successful methods where PCA is used as a feature selection technique, it is quite clear that the proposed model is efficient and encouraging.

References


Index Terms

Computer Science

Information Sciences

Keywords

Review spam detection, Fake review, PCA, Active Learning, Machine Learning.