



Working Model Description

In the realm of e-commerce, online reviews play a pivotal role in shaping consumer opinions and decisions. However, the prevalence of fake reviews poses a significant challenge, necessitating the development of robust systems to identify and mitigate their impact. Our Fake Review Detection System is an amalgamation of advanced Natural Language Processing (NLP) techniques and web scraping, aiming to provide an efficient solution for distinguishing genuine from deceptive reviews.

Data Collection and Preprocessing:

The foundation of our system lies in a curated dataset of product reviews. Leveraging pandas, NumPy, and seaborn, we explored and visualized the data to gain insights into its distribution. The dataset comprises labeled reviews, with each review tagged as genuine or fake.

To prepare the data for model training, we performed extensive preprocessing using NLTK (Natural Language Toolkit) and Scikit-learn. This involved removing punctuation, tokenization, and handling stopwords. Stopwords are common words like "the," "and," and "is," which contribute little to the sentiment of the text and can be safely removed.

Feature Extraction:

The key to effective NLP models is extracting meaningful features from the textual data. We employed the CountVectorizer from Scikit-learn, transforming the raw text into a bag-of-words representation. This conversion captures the frequency of each word in the document, creating a numerical feature matrix suitable for machine learning algorithms.

Model Development:

The heart of our system is a Support Vector Classifier (SVC) trained on the preprocessed and vectorized reviews. The SVC excels in binary classification tasks and, in our case, categorizes reviews as genuine or fake based on learned patterns from the training data.

Web Scraping for Real-Time Data:

Recognizing the dynamic nature of online reviews, we integrated a web scraping script using the requests and BeautifulSoup libraries. This script extracts reviews from an Amazon product page, providing a real-time stream of data for analysis and model evaluation.

API Deployment with Flask:

To make the model accessible and user-friendly, we employed Flask to deploy a simple API. This API allows users to input a URL of an Amazon product page, triggering the model to predict the percentage of positive reviews. The web interface enables seamless interaction, making the system practical for users without technical expertise.

Challenges and Future Enhancements:

While our system showcases proficiency in identifying fake reviews, there are challenges such as adapting to diverse review formats and handling nuanced language. Future enhancements could involve exploring more sophisticated models, considering sentiment analysis, and incorporating user feedback to continually refine the model.

Conclusion

In conclusion, our Fake Review Detection System is a comprehensive solution addressing the pressing issue of deceptive online reviews. The utilization of NLP techniques, coupled with web scraping for real-time data, demonstrates the adaptability of the system to the dynamic nature of e-commerce platforms. As consumer trust continues to be a cornerstone of online transactions, our system contributes to fostering a more transparent and trustworthy digital marketplace.

Github Link:

<https://github.com/JoshuaR26/TrusTorch.Ext>