

Title: Hinglish Products Review-Based Sentiment Analysis on Using ML Algorithms

Author: Sabbir Ahmed; Md. Sazzadur Ahmed

Abstract: In the realm of e-commerce, understanding consumer sentiment towards products is paramount for businesses to thrive. With the exponential growth of online shopping in India, an increasing number of customers express their opinions through a blend of Hindi and English, commonly known as Hinglish. This study delves into the domain of sentiment analysis, focusing specifically on Hinglish product reviews, and employs machine learning (ML) algorithms to decipher the sentiments expressed within these reviews. The research aims to develop a robust framework capable of accurately categorizing Hinglish product reviews into positive, negative, or neutral sentiments. The proposed framework leverages NLP to preprocess and analyze Hinglish text data. This preprocessing stage includes tokenization, stemming, and eliminating stop words in order to extract significant characteristics from the text. The system then utilizes machine learning algorithms including Support Vector Machines (SVM), Naive Bayes, and Random Forest to train sentiment categorization models. The training datasets for these models consist of labeled datasets, including Hinglish product reviews, each annotated with sentiment polarity. Iterative sentiment analysis applied to Hinglish product reviews has important ramifications for enterprises in the Indian market. Through effective identification of client attitudes, organizations can acquire substantial knowledge regarding customer preferences, degrees of satisfaction, and prospects for enhancement. As a result, businesses are better equipped to customize their product offerings, marketing campaigns, and customer support programs to better suit the demands and preferences of their target market. I used new conceptual data in the Hindi language but English alphabet format. I have used ten models here. I am trying to find out the best accuracy models among them. ExtraTree best performance at 93% accuracy, Random Forest 92.13% and SVM 91.71% model with accuracy.

Keywords: Sentiment Analysis , Machine Learning , Learning Algorithms , Support Vector Machine , Random Forest , Text Data , Online Shopping

DOI: <https://doi.org/10.1109/COMPAS60761.2024.10796995>

<https://ieeexplore.ieee.org/abstract/document/10796995>