

Flipkart reviews\readme.md

Customer Review Sentiment Analysis

Objective

The objective of this project is to classify customer reviews as positive or negative and understand the pain points of customers who write negative reviews. By analyzing the sentiment of reviews, we aim to gain insights into product features that contribute to customer satisfaction or dissatisfaction.

Dataset

We have utilized a dataset consisting of 8,518 reviews for the "YONEX MAVIS 350 Nylon Shuttle" product from Flipkart. The dataset includes features such as Reviewer Name, Rating, Review Title, Review Text, Place of Review, Date of Review, Up Votes, and Down Votes. This data was scraped by a team of Data Engineers from the Flipkart website.

Data Preprocessing

- Text Cleaning: Special characters, punctuation, and stopwords have been removed from the review text to improve analysis accuracy.
- Text Normalization: Lemmatization or stemming techniques have been applied to reduce words to their base forms.
- Numerical Feature Extraction: Techniques such as Bag-of-Words (BoW), Term Frequency-Inverse Document Frequency (TF-IDF), Word2Vec (W2V), and BERT models have been used for feature extraction.

Modeling Approach

- Model Selection: Various machine learning and deep learning models have been trained and evaluated using the processed text data.
- Evaluation Metric: The F1-Score has been chosen as the evaluation metric to assess the performance of the models in classifying sentiment.

Model Deployment

- Flask App Development: A Flask web application has been developed for user interaction. Users can input a review, and the app will generate the sentiment (positive or negative) of the review.
- Model Integration: The trained sentiment classification model has been integrated into the Flask app for real-time inference.

- **Deployment:** The Flask or Streamlit app has been deployed on an AWS EC2 instance, making it accessible over the internet for users to use and gain insights from customer reviews.