

**In-depth Product Reviews**



**By:**

Parul Bhutani wadhwa

snehal kathiriya

**How It Works:**

**NLP:** NLP is an area of artificial intelligence that deals with how computers and humans interact with each other through natural language. NLP uses a variety of techniques and algorithms to understand, comprehend, and create human language in a useful way.

**In-depth Product Reviews:** Our project aims to analyze both good and bad reviews. More specifically, we want to gain insight from a business point of view, looking at the reasons for positive or negative reviews for the mobile phones. Our analysis will go deep into the factors that contribute to the reviews, giving us an in-depth insight into the main reasons for each kind of review.

There are usually several stages involved in our project implementation:

1. **Text Preprocessing:** The first step in text preprocessing is to clean and prepare the text data. This may include things like removing unique characters(stopword), breaking the text down into individual words or sentences(tokenization), and stemming (shrinking words down to their root form).
2. **TextBlob for Sentiment Analysis:** The preprocessed text data is sent to TextBlob’s trained sentiment analysis model. In TextBlob, each word in the text is given a sentiment score (called a ‘polarity’) and the overall sentiment score is determined by the lexicon of the text. Word polarity is a numerical score that indicates the sentiment of a text on a scale from -1 meaning the most negative sentiment in the text to 1 meaning the most positive sentiment in the text. It also gives a neutral sentiment for polarities equal to 0. But since neutral reviews will not be useful in our analysis, we are discarding neutral and only considering “positive” and “negative” sentiments.
3. **Topic Modelling:** We have used LDA (Latent Dirichlet Allocation) for topic modelling. As part of our negative review analysis, LDA can find common themes or issues that customers are complaining about. By presenting each negative review as a mix of themes, we can gain insight into common problems such as battery issues, heating problems, customer service issues, etc. And similarly for positive review.
4. **Interpreting Mobile Reviews:** We have divided entire dataset into the negative and positive reviews in the total of 11 topics.

**Topic 0:** Lenovo - overall disappointment,

**Topic 1:** Samsung - Poor camera quality,

**Topic 2:** iPhone - screen issues/broken,

**Topic 3:** Lenovo - poor charging and heating issues,

**Topic 4:** Product doesn't meet expectation.

A close-up of words

Description automatically generated

**Topic 5:** Best battery and excellent camera.

**Topic 6:** Samsung good condition.

**Topic 7:** Lenovo heat charging issue.

**Topic 8:** Worth the price.

**Topic 9:** Smooth touch overall nice.

**Topic 10:** Superb display good for game.

A close-up of words

Description automatically generated

1. **Machine learning:** Using this topic modelling output dataset, we have created machine learning model so that this topic modelling can be used for any other product in future.

Here is the list of all the model with the accuracy score:

Naive Bayes - 82%

Gradient Boosting Classifier - 92%

Random Forest Classifier - 93.7%

Logistic Regression - 93.1%

SVM - 93.6%

Since Random Forest Classifier gave the highest accuracy, we chose it as the final model.

**Dataset:**

We have downloaded two datasets from the Kaggle and combined it to get the better Results.

**Link:** [Mobile Reviews (kaggle.com)](https://www.kaggle.com/datasets/prakharprasad/mobile-reviews)

<https://www.kaggle.com/datasets/grikomsn/amazon-cell-phones-reviews>

This dataset consists of customer reviews related to the mobile phone and our final dataset consists of total of 82661 reviews.

**Contribution:**

**Snehal Kathiriya:** NLP & Data preprocessing

**Parul Bhutani Wadhwa:** Topic generation & Machine Learning

**References**

*pandas documentation — pandas 2.1.3 documentation*. (n.d.). <https://pandas.pydata.org/docs/>

*NLTK :: Natural Language Toolkit*. (n.d.). <https://www.nltk.org/>

*scikit-learn: machine learning in Python — scikit-learn 1.3.2 documentation*. (n.d.)

<https://scikitlearn.org/stable/index.html>

TensorFlow. (n.d.). *TensorFlow*. <https://www.tensorflow.org/>

Da Costa-Luis, C. (n.d.). *tqdm documentation*. <https://tqdm.github.io/>

TextBlob: Simplified Text Processing — TextBlob 0.16.0 documentation. (n.d.).<https://textblob.readthedocs.io/en/dev/>

*Gensim: topic modelling for humans*. (n.d.). <https://radimrehurek.com/gensim/>

Amueller. (n.d.). *GitHub - amueller/word\_cloud: A little word cloud generator in Python*. GitHub. <https://github.com/amueller/word_cloud>

*Matplotlib — Visualization with Python*. (n.d.). <https://matplotlib.org/>

*seaborn: statistical data visualization — seaborn 0.13.0 documentation*. (n.d.). <https://seaborn.pydata.org/>

*NumPy*. (n.d.). <https://numpy.org/>