INDIAN INSTITUTE OF SCIENCE
भारतीय विज्ञान संस्थान

Sentiment analysis of mobile reviews

Department of Computational and Data Sciences

https://github.com/DSBA-IISc-Team8/sentiment-analysis-of-mobile-reviews

Team

1. Abhinav Nigam *(abhinavnigam@iisc.ac.in)*
2. Karthik S K *(karthiksk@iisc.ac.in)*
3. Sasikiran Karri *(sasikirankar@iisc.ac.in)*
4. Satyam Kumar *(satyamk@iisc.ac.in)*
5. Siva Kranthi Kumar Mallipeddi *(sivam@iisc.ac.in)*
6. Sreedhar Reddy Vundela *(sreedharv@iisc.ac.in)*
7. Sudhakar Kulkarni Mukayya *(sudhakark@iisc.ac.in)*
8. Sumit Sinha *(sumitsinha@iisc.ac.in)*

Problem Statement

“

Sentiment analysis of mobile reviews. The final model classifies the reviews as positive/negative/neutral

Motivation

- Lot of discussions about products on ecommerce platforms
- Business need to understand the product sentiments. Also aspects contributing to that
- Mobiles being focus of interest

Dataset

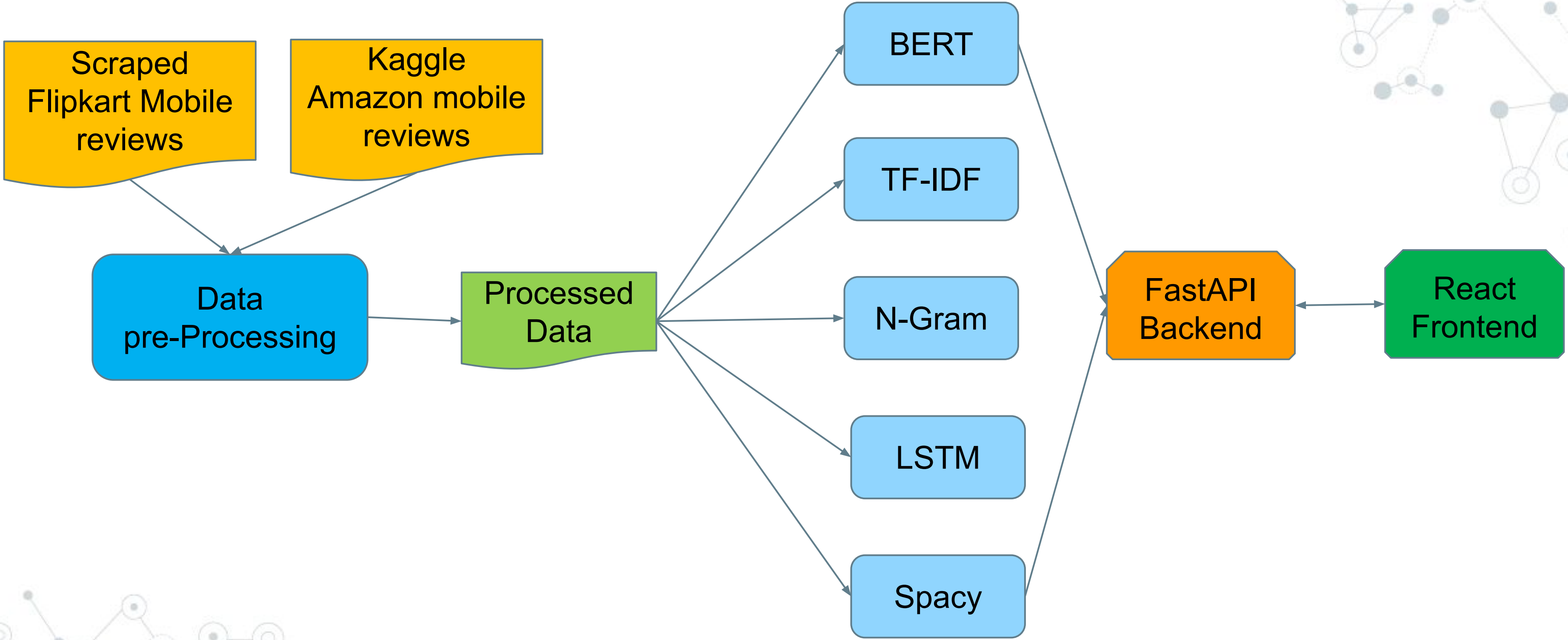
- ~1.2 Lakh Mobile reviews are scrapped from flipkart.com using scrapy python library
- Scrapped data contains mobile name, review text, rating, likes and dislikes
- ~1.5 Lakh Amazon reviews from [kaggle]

Process

We followed the below process while doing the sentiment analysis.

- Scrapping flipkart.com mobile reviews and merged with amazon mobile review datasets
- Data Cleaning, EDA & transformation of data : Punctuations, stop words, space removal, normalization of texts etc.
- Experiment with basic models: **TF-IDF** (LineraSVC), **N-Grams** (Multinomial NaiveBayes) & **LSTM**
- Experiment with state of the art Models: **BERT**
- Feature sentiment analysis through rule based dependency graphs using **SpaCy**
- Estimating the performance (**Accuracy & F1**) on unseen data

Process Flow



Challenges faced

- Data collection: Duplication of reviews, processing required
- Understanding the semantic meaning of the review
- Fine tuning BERT model on the customized dataset
- Handling sarcasm

Results

Model	Accuracy	Precision	Recall	F1	Hyperparameters	features
N-grams(1)	0.6	0.52	0.61	0.53	{'alpha': 1.0, 'fit_prior': True}	88280
N-grams(2)	0.6	0.51	0.69	0.51	{'alpha': 0.75, 'fit_prior': True}	1544246
N-grams(3)	0.6	0.51	0.60	0.49	{'alpha': 0.75, 'fit_prior': True}	4820742
TF-IDF(Logistic)	0.53	0.43	0.53	0.41	{'logistic__C': 10}	127874
LSTM (Binary)	0.8*	0.96	0.8	0.25	dropout:0.5, activation: sigmoid	
BERT	0.96* 0.53 **				Epochs:3, bacth_size: 4	

*: At sentiment level (_ve, -ve & neutral)
**: At rating level (1, 2, 3, 4 & 5)

LSTM (Binary): This was just trained on positive/negative reviews, hence have higher values but not preferred model

References

1. https://github.com/philipperemy/amazon-reviews-scraper
2. [2019] Sentiment Analysis of Amazon Products Using Ensemble Machine Learning Algorithm
3. https://www.kaggle.com/yaowenling/amazon-cell-phone-review-nlp
4. https://medium.com/analytics-vidhya/sentiment-analysis-on-smartphones-reviews-431793d8b91a
5. https://www.diva-portal.org/smash/get/diva2:1241547/FULLTEXT01.pdf
6. https://towardsdatascience.com/sarcasm-detection-with-nlp-cbff1723f69a

Thank You

Any Questions?