Automated User Feedback Classification

Members: Berkay Alkan, Olcayto Türker, Sertay Akpınar
Advisor: Fatma Başak Aydemir
Department of Computer Engineering, Boğaziçi University

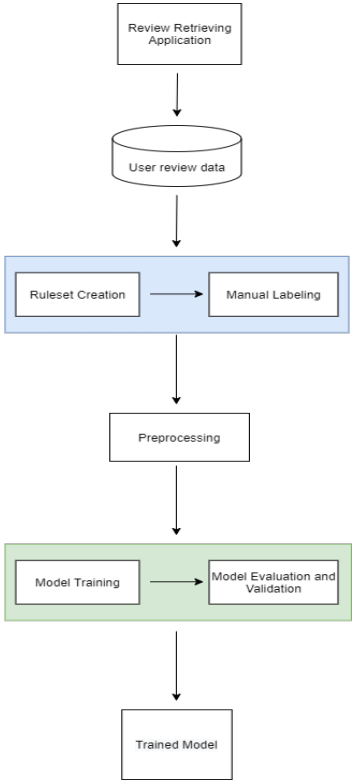
Introduction & Motivation

In app stores, there are thousands of user comments given for apps. Analysis of these user feedbacks has a huge importance. In this project with natural language processing and machine learning techniques, it is aimed to classify the user feedback with their sentiments.

By clustering user feedback from app stores, app companies could plan their future releases. User level of satisfaction from the app could be measured. Bugs that are not known could be fixed by taking account of the user comments.

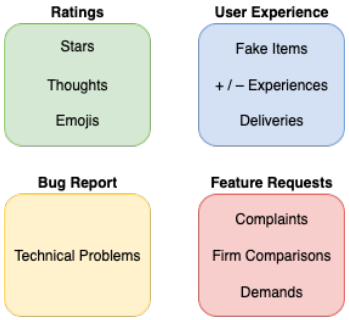
In addition, there is no scientific work about this subject in Turkish language. This was one of our main motivations.

Flow Chart



How to tag reviews?

- Total of 1058 reviews.
- To create a common mindset first labeled the reviews together, then separately.
- For assessing the reliability of agreement, Fleiss' Kappa score is used which is **85.05%**. Python's *statsmodels* package is used.



Example Reviews

- Ratings and Feature Request: 'Berbat kesinlikle tavsiye etmiyorum. Ne bir müşteri hizmetleri yardım ediyor ne de bir baskasi urunlerin hangi kargo ile gönderileceği yazmiyor bile.'
- Ratings: '❤️❤️❤️'
- Bug Report: 'Sepete eklediğim ürünleri sepetten kaldırınca önceden eklediklerim olarak ürünlerim yine sepete görünüyor. Bunu düzeltebilirseniz çok iyi olur.'
- User Experience: 'Reklam reklam reklam'
- Feature Request: 'Hepsiburada uygulaması daha iyi. Trendyol kargo ücreti çok alıyor.'

Model Used

- BERTurk, Turkish extension of the BERT model is used.
- BERT, a transformer-based machine learning technique for natural language processing.
- Utilized several Turkish NLP models created by MDZ Digital Library team at the Bavarian State Library.

Results

dbmdz/bert-base-turkish-128k-cased

Precision	Recall	F1	Accuracy
0.92	0.919	0.92	0.89

dbmdz/bert-base-turkish-cased

Precision	Recall	F1	Accuracy
0.903	0.910	0.90875	0.85

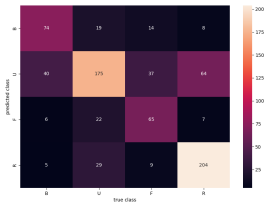
dbmdz/electra-small-turkish-cased-discriminator

Precision	Recall	F1	Accuracy
0.872	0.883	0.884	0.79

dbmdz/bert-base-turkish-128k-cased Test

Precision	Recall	F1	Accuracy
0.667	0.642	0.654	0.673

dbmdz/bert-base-turkish-128k-cased Train Confusion Matrix



dbmdz/bert-base-turkish-128k-cased Test Confusion Matrix