REPORT
PROJECT NLP I AUTOMATED CUSTOMER REVIEWS

Kerem Senler & Julia Nuss

OUR STEPS:

1. Data Exploration (Shape of dataset, null values,…)
- Removing null values and unnecessary rows

2. Preprocessing
- Converting text to lowercase
- Removing whitespaces, special characters, emojis and stopwords
- Tokenization, Stemming, Lemmatizing, TF-IDF-Vectorization

3. Model Training

Models we tried:
- Logistic Regression
- Multinomial Naive Bayes
- Support Vector Machines
- RandomForest
- Neural Network
- BERT base uncased

4. Model Evaluation

Evaluated by:
- Test accuracy
- Precision, recall, f1-score and support for every sentiment

Rating:
1. BERT base uncased
2. Neural Network
3. Random Forest
4. SVM
5. Multinomial Naive Bayes

5. Model Refinement

Worked further on BERT-model and Neural Network (adding amount of epochs, layers, changing activation functions, learning rate,…) -> It wasn't possible to improve more than around 2 % through our refinements

6. Conclusion

It's hard to reach a high accuracy when it comes to text and with all the refinements done improve the model to a noticeable extent. Anyways we learned what models work best in NLP and how you can compare pre-trained models with traditional models. Of course since they've been trained a lot they perform mostly better on language processing.

25.10.2024