Twitter Sentiment analysis

Prepared by: Alaa & Tasneem

Agenda

- The limitations of the dataset
- The methodology and approaches that you used
- The results
- The other approaches that might be used for the same problem

Limitations of the dataset

- The previous data set we used was too small (less than 5k record).
- After changing the dataset, the vectorization step causes a runtime error.

Methodology and approach used

RoBERTa (short for “Robustly Optimized BERT Approach”)

- It is a variant of the BERT (Bidirectional Encoder Representations from Transformers) model, which was developed by researchers at Facebook AI.
- RoBERTa is a transformer-based language model that uses self-attention to process input sequences and generate contextualized representations of words in a sentence.
- One key difference between RoBERTa and BERT is that RoBERTa was trained on a much larger dataset and using a more effective training procedure.
- RoBERTa uses a dynamic masking technique during training that helps the model learn more robust and generalizable representations of words.
- RoBERTa has been shown to outperform on a variety of natural language processing tasks, including language translation, text classification, and question answering.
- Overall, RoBERTa is a powerful and effective language model that has made significant contributions to the field of NLP and has helped to drive progress in a wide range of applications.

Other approaches that can be used

- Rule-based models
- Traditional machine learning models
- Deep learning models

References

https://www.geeksforgeeks.org/overview-of-roberta-model/

https://pub.towardsai.net/16-open-source-nlp-models-for-sentiment-analysis-one-rises-on-top-b5867e247116

Any questions??

Thank you