Emotion Evaluator

Demo

Table of contents

Intro

- Github repo https://github.com/Jve-git/emotionevaluator

Data

- IMDB data labelled either as positive or negative

Pre-trained transformer models

- Transformers package
  - Hugging Face's sentiment analysis pipeline
- Three models
  - distilbert-base-uncased-finetuned-sst-2-english
  - siebert/sentiment-roberta-large-english
  - aychang/roberta-base-imdb
- Binary classification
  - positive
  - negative
- Each pre-trained on different datasets

distilbert-base-uncased-finetuned-sst-2-english

- Default model of the sentiment-analysis pipeline
- DistilBERT-base-uncased finetuned on sst2 (Stanford Sentiment Treebank v2) corpus (movie reviews)
- DistilBERT is a transformers model, smaller and faster than BERT

siebert/sentiment-roberta-large-english

- fine-tuned checkpoint of RoBERTa-large
- trained on diverse text sources over different text types such as reviews and tweets
- case sensitive
- on average, outperforms a DistilBERT-base model which is only trained on sst-2 dataset

aychang/roberta-base-imdb

- simple **RoBERTa** transformer model
  - RoBERTa (Robustly Optimized BERT Approach) — a more refined version of BERT
- trained on imdb dataset
  - https://huggingface.co/datasets/imdb
- biased towards movie reviews

Preprocessing

- All three mentioned pre-trained models already preprocess input text
  - one is case sensitive, the other puts everything in lower case
  - truncation is done at length of 512 tokens
  - stemming and lemmatization not done
    - transformers learn from the context not from root words
- Only preprocessing before passing to transformers
  - stripping
  - removal of HTML tags
  - improved the performance for all three models a little bit

Results (1 of 3)

- benchmark report
  - accuracy
  - precision
  - recall
  - F1 score

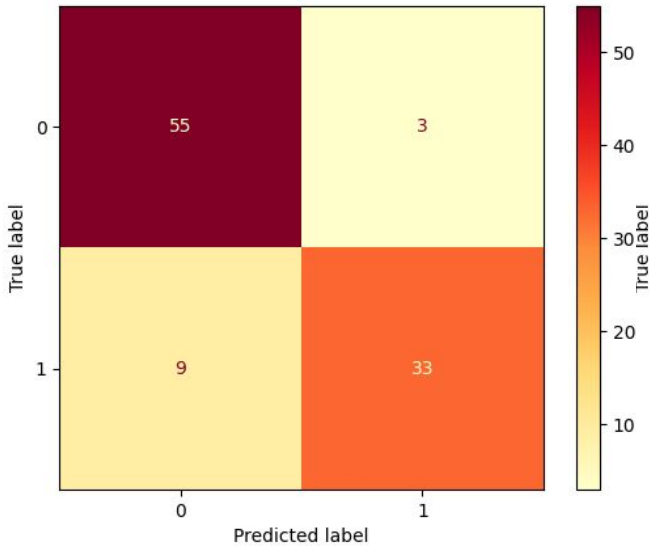
output > benchmark_report.md

model	accuracy	precision	recall	f1
distilbert-base-uncased-finetuned-sst-2-english	0.88	0.916667	0.785714	0.846154
siebert/sentiment-roberta-large-english	0.95	0.95122	0.928571	0.939759
aychang/roberta-base-imdb	0.99	0.976744	1	0.988235

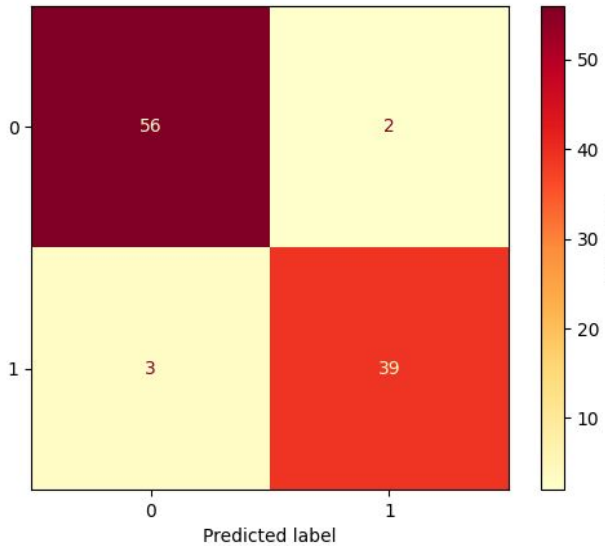
Results (2 of 3)

- confusion matrix

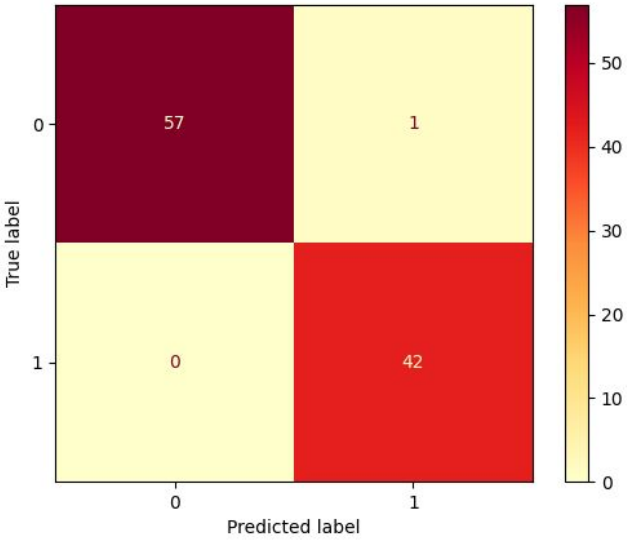
Distilbert (SS2-T)



Siebert (roberta-large)



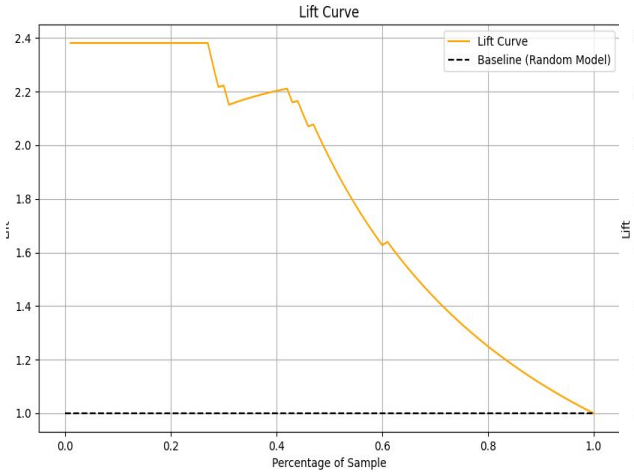
aychang (IMDB)



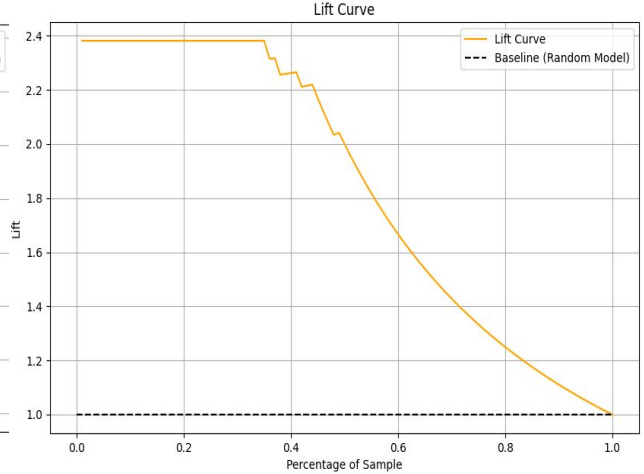
Results (3 of 3)

- Lift curve

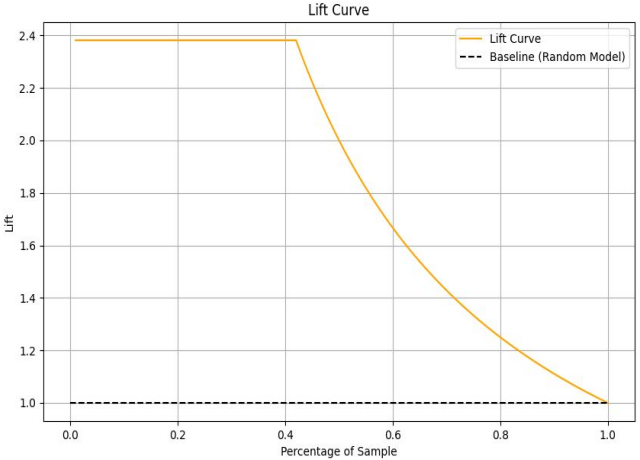
Distilbert (SS2-T)



Siebert (roberta-large)



aychang (IMDB)



Conclusion

- aychang (IMDB) best at performing on the IMDB reviews (54.95 seconds)
- distilbert is worst at performing on the IMDB reviews, however it is by far the quickest predictor (23.32 seconds)
- siebert is taking by far most time to predict the 100 IMDB reviews (166.70 seconds), however, as it is more generalised it is preferred to use for more general texts

Demo

Questions?