# *Investigating hybrid models*
# Machine Learning for Natural Language Processing 2022

**Clara Minel**
3ème année voie DSBD, MIE
clara.minel@ensae.fr

## 1 Problem Framing

While lexicon models are very useful for performing sentiment analysis on text data, they seem to perform poorly on analyzing unstructured social media content such as tweets. They consist in aggregating scores assign to a set of vocabulary that is able to capture sentiment polarity. However, static lexicons models fail at adapting to the context and the content of the text and consequently, they perform poorly when predicting sentiment polarity of each tweet. Tweets are increasingly used for assessing client satisfaction. In this project, we will try to assess the customer services of different airlines through social media content. Thus, the aim of this project is to combine regular lexicon methods with simple natural language processing algorithm to combine the convenience of lexicon models and the ability to adapt to the content of machine-learning models. By combining both types of models with a weighted average ensemble, we aim at increasing the performance of social media sentiment analysis as recently donc in the literature by Bashar(2022) (Bashar, 2022) and Muhammad(2014) (Muhammad et al., 2014).

## 2 Experiments Protocol

Firstly, I compute several descriptive statistics on the dataset itself, as presented in Appendix A1 and then on the tweet data, presented in Appendix A2. This enables us to understand the structure of our dataset. The experiment consists in implementing a static lexicon model on the tweet dataset and comparing standard NLP models such as TF-IDF. Then, I compute the BERT model to use a more advance approach on the dataset. Finally, I implement a hybrid approach, using a weighted average ensemble that combines results from the TF-IDF approach to the Lexicon approach.

## 3 Results

Results from the Afinn model, TF-IDF, BERT and the hybrid approach are presented in appendix A3 to A6. Overall the lexicon model seems overall to underperform relatively to other approaches when predicting labels. This mainly due to the fact that lexicon models poorly adapt to the context. Following the literature, I create a hybrid approach to solve the shortcomings of the lexicon models. This hybrid approach seem to overall underperform relative to the TF-IDF model, as shown in our code [1], however, when analyzing the confusion matrix, the model performs better for predicting negative and positive tweets. Thus, this model is quite interesting for sentiment polarization. [2], however, when analyzing the confusion matrix, the model performs better for predicting negative and positive tweets. Thus, this model is quite interesting for sentiment polarization.

## 4 Discussion/Conclusion

For future research it would be interesting to test the Hybrid Lexicon model with Bert. Concerning this dataset, we could also limit ourselves to a positive-negative polarity (by excluding neutral), or consider the confidence level of the sentiment for more complex models.
Let's see how our sentiment analysis on social networks can be insightful. In this case, an airline could compare itself to its competitors, better determine the expectations of its customers and thus improve its customer service. With an automatised analysis, a company would also be able to quickly identify changes in brand perception and quickly correct a problem before it becomes too serious.

[1] Google Colab : https://colab.research.google.com/drive/1Gs5LQo-uiz1Ce52CIqb-et45BBe6cJEE?usp=sharing
[2] Github : https://github.com/cminel23/NLP-Project

# Appendix

## 4.1 A1 - Descriptive statistics regarding tweets sentiment and the distribution across airlines

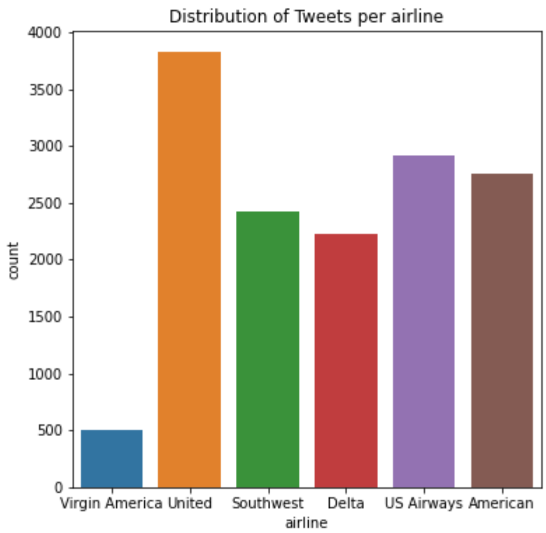


Figure 1: Distribution of airline tweets

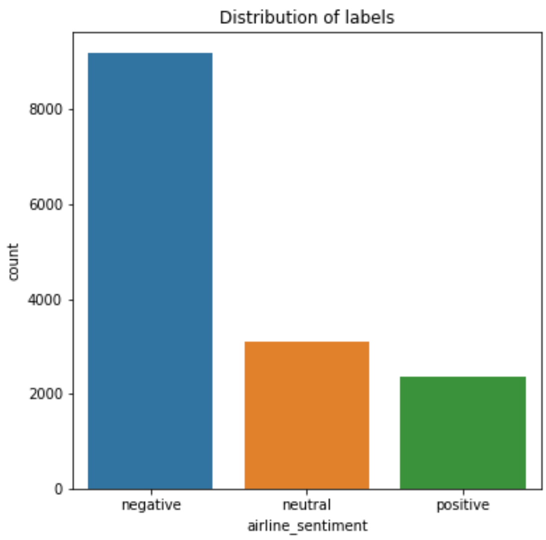


Figure 2: Distribution of tweets sentiment

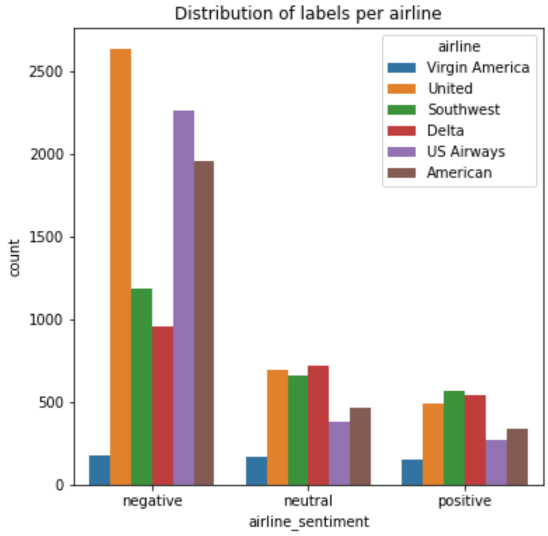


Figure 3: Distribution of labels per airline

## A2 - Descriptive statistics on the tweets data

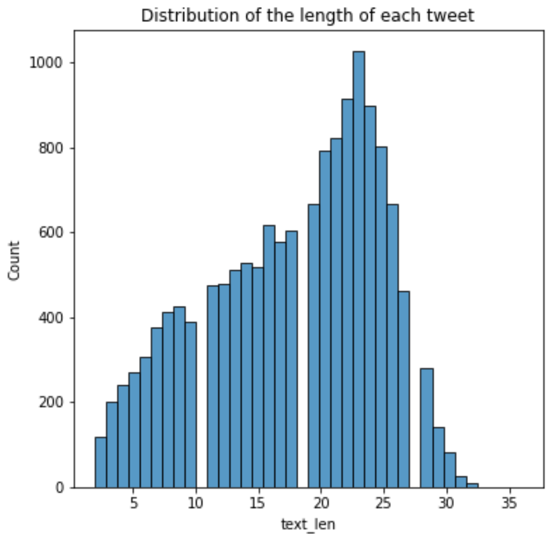


Figure 4: Tweet length

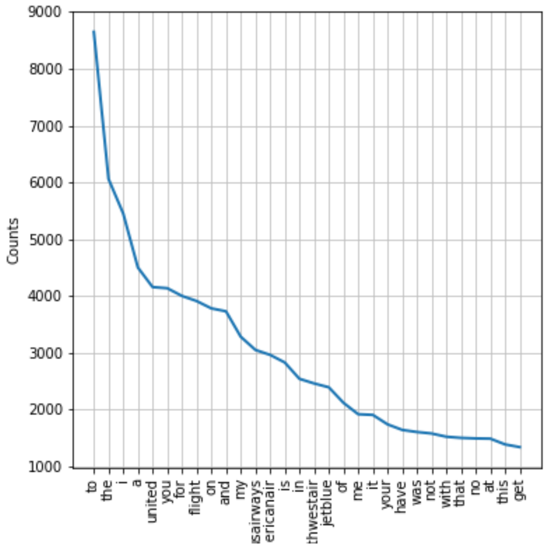


Figure 5: Distribution of the most common words of our dataset

Table 1: Most common words in the tweets

| | words | count |
|---|---|---|
| 0 | to | 8648 |
| 1 | the | 6060 |
| 2 | i | 5452 |
| 3 | a | 4501 |
| 4 | united | 4157 |
| 5 | you | 4137 |
| 6 | for | 4001 |
| 7 | flight | 3911 |
| 8 | on | 3782 |
| 9 | and | 3730 |
| 10 | my | 3284 |
| 11 | usairways | 3051 |
| 12 | americanair | 2964 |
| 13 | is | 2830 |
| 14 | in | 2542 |
| 15 | southwestair | 2461 |
| 16 | jetblue | 2393 |
| 17 | of | 2120 |
| 18 | me | 1919 |
| 19 | it | 1907 |

Table 2: Bigrams of our dataset

| | bigrams | count |
|---|---|---|
| 0 | (on, the) | 533 |
| 1 | (to, get) | 477 |
| 2 | (for, the) | 446 |
| 3 | (I, have) | 435 |
| 4 | (on, hold) | 400 |
| 5 | (customer, service) | 358 |
| 6 | (to, be) | 357 |
| 7 | (my, flight) | 355 |
| 8 | (in, the) | 355 |
| 9 | (Cancelled, Flightled) | 350 |
| 10 | (for, a) | 342 |
| 11 | (to, the) | 323 |
| 12 | (I, was) | 302 |
| 13 | (@united, I) | 296 |
| 14 | (trying, to) | 272 |
| 15 | (a, flight) | 259 |
| 16 | (out, of) | 256 |
| 17 | (hold, for) | 242 |
| 18 | (need, to) | 234 |
| 19 | (at, the) | 234 |

### 4.2 A3. Results AFINN model

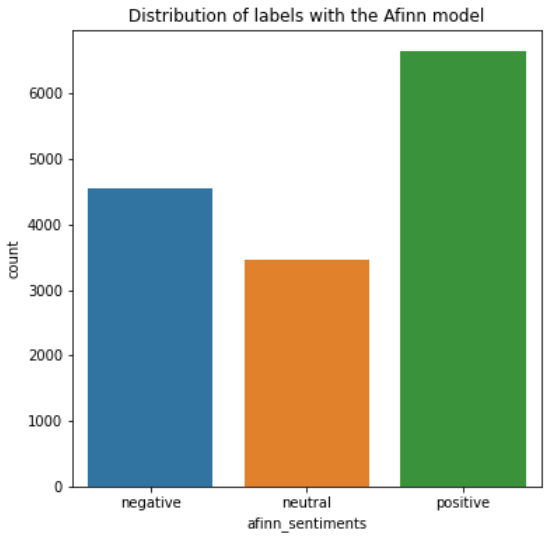


Figure 6: Distribution of labels predicted by the AFINN model

| afinn label<br>True label | negative | neutral | positive |
|---|---|---|---|
| negative | 0.444541 | 0.220092 | 0.335367 |
| neutral | 0.122298 | 0.389480 | 0.488222 |
| positive | 0.035971 | 0.097334 | 0.866695 |

Table 3: Confusion matrix for the AFINN model

### 4.3 A4. Results TF-IDF model

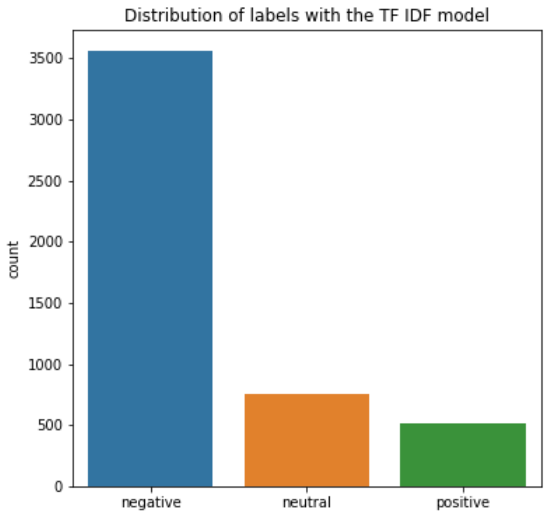


Figure 7: Distribution of labels predicted by the TF-IDF model

| Predicted label<br>True label | negative | neutral | positive |
|---|---|---|---|
| negative | 0.937439 | 0.048622 | 0.013938 |
| neutral | 0.446138 | 0.496951 | 0.056911 |
| positive | 0.297510 | 0.149410 | 0.553080 |

Table 4: Confusion matrix for the TF-IDF model

### 4.4 A5. Results from the BERT model

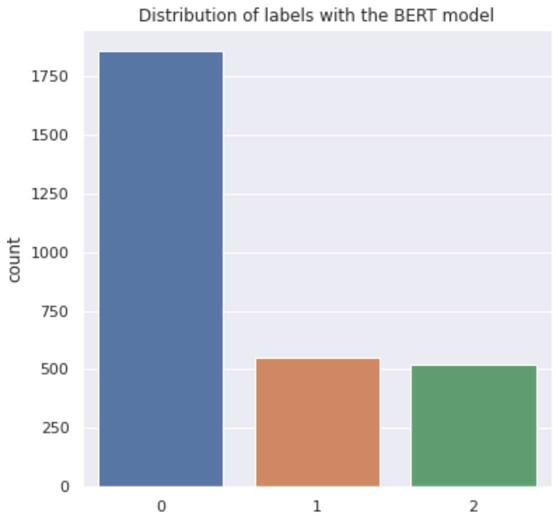


Figure 8: Distribution of labels predicted by the BERT model (0 is negative, 1 is neutral and 2 is positive)

| Predicted label<br>True label | negative | neutral | positive |
|---|---|---|---|
| negative | 0.912445 | 0.060573 | 0.026982 |
| neutral | 0.244300 | 0.640065 | 0.115635 |
| positive | 0.100402 | 0.094378 | 0.805221 |

Table 5: Tweets sentiment with hybrid model with equal weights

## 4.5 A6. Results hybrid models

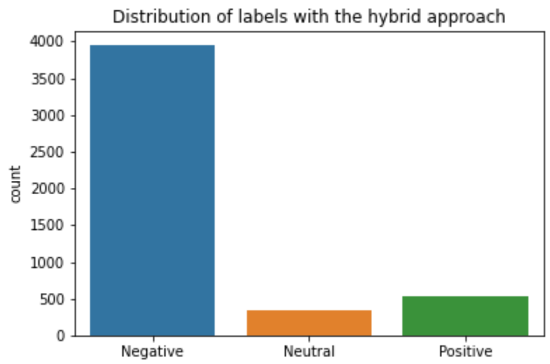


Figure 9: Distribution of labels predicted by the hybrid model

| Predicted label<br>True label | negative | neutral | positive |
|---|---|---|---|
| negative | 0.974068 | 0.011994 | 0.013938 |
| neutral | 0.699187 | 0.252033 | 0.048780 |
| positive | 0.340760 | 0.086501 | 0.572739 |

Table 6: Tweets sentiment with hybrid model with equal weights

## References

Aminu Muhammad, Nirmalie Wiratunga, and Robert Lothian. 2014. A hybrid sentiment lexicon for social media mining. In *2014 IEEE 26th International Conference on Tools with Artificial Intelligence*, pages 461–468. IEEE.

Md Khayrul Bashar. 2022. A hybrid approach to explore public sentiments on covid-19. *SN Computer Science*, 3(3):1–19.