BENCHMARKING DEEP LEARNING AGAINST TRADITIONAL MACHINE LEARNING APPROACHES FOR SENTIMENT ANALYSIS IN TWITTER DATA: A COMPARATIVE STUDY

Nabgouri Othmane & Keouda Etienne

INTRODUCTION

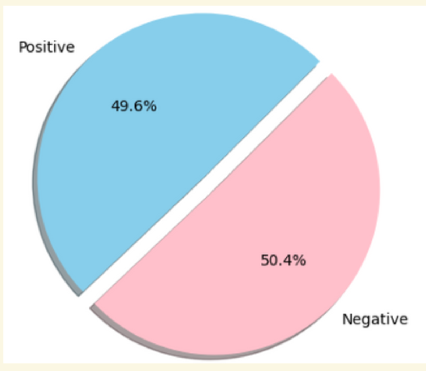
In the realm of social media, Twitter serves as a vibrant platform for expressing opinions, thoughts, and sentiments. Analyzing tweet sentiment has become crucial for various purposes, from understanding public opinion to market analysis. In this project, we aim to compare the efficacy of deep learning methods with traditional machine learning algorithms for sentiment analysis on French Twitter data.

To address this, we conducted a comparative study using deep learning techniques based on Transformers and traditional algorithms like Support Vector Machines (SVM). Our dataset comprised 100,000 labeled French tweets, partitioned into 80% for training, 10% for validation, and 10% for testing. This meticulous partitioning ensures robust model evaluation across different stages of development.

Through rigorous evaluation utilizing standard metrics, we seek to determine which approach yields superior performance in terms of accuracy, computational efficiency, and scalability. Ultimately, this project aims to provide insights into the suitability of deep learning versus traditional machine learning for sentiment analysis tasks on French Tweets, aiding practitioners in making informed decisions regarding model selection for similar applications.

1

Repartition of tweets



2

Most frequent words

Positive words

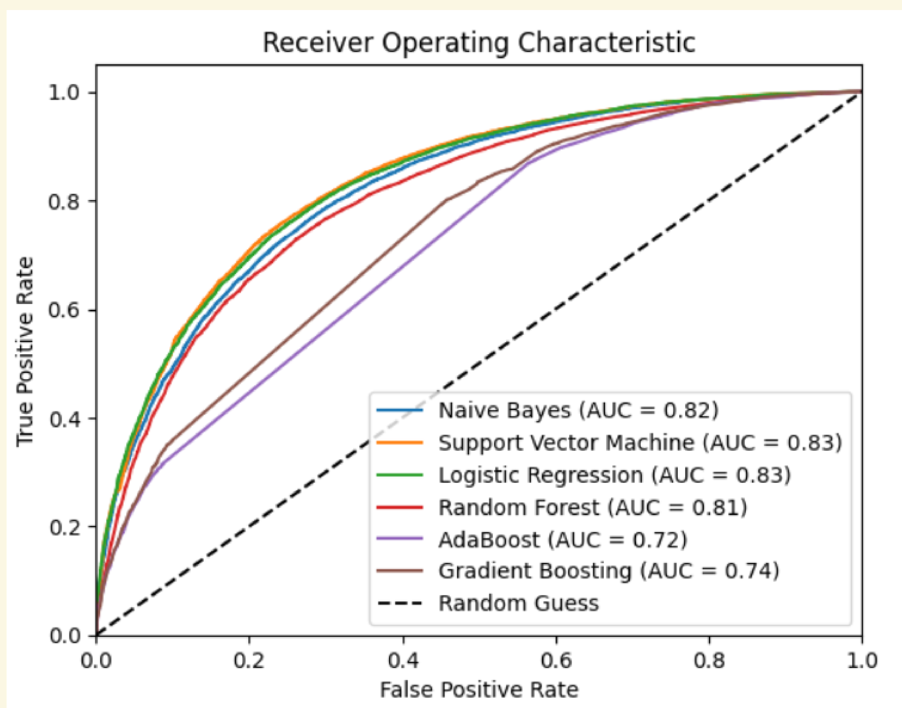
Negative words

3

MACHINE LEARNING APPROACH

We have chosen to test various machine learning algorithms to predict our labels. Our choices were focused on these 6 algorithms: Naive Bayes, SVM, Logistic Regression, Random Forest, AdaBoost and Gradient Boosting. Here are the results obtained:

Model	Data Cleaning	Precision	Recall	Accuracy	Execution Time(minutes)
Naive Bayes	No	0.75	0.0.75	0.748	1
Naive Bayes	Yes	0.745	0.745	0.757	1
SVM	No	0.76	0.765	0.76	16
SVM	Yes	0.76	0.755	0.76	16
Logistic Regression	No	0.76	0.76	0.758	1
Logistic Regression	Yes	0.75	0.755	0.75	1
Random Forest	No	0.735	0.735	0.734	3
Random Forest	Yes	0.73	0.735	0.73	3
AdaBoost	No	0.69	0.65	0.65	1
AdaBoost	Yes	0.69	0.67	0.67	1
Gradient Boosting	No	0.69	0.67	0.67	1
Gradient Boosting	Yes	0.73	0.72	0.72	1



4

DEEP LEARNING APPROACH

CamemBERT

•Camembert [2], inspired by the Roberta architecture, encompasses various specialized versions tailored for distinct natural language understanding tasks in French

•The training process for Camembert involved leveraging vast amounts of French text data, coupled with state-of-the-art optimization techniques, resulting in models that exhibit exceptional fluency and accuracy in understanding the intricacies of the French language.

FlauBERT

•Flaubert [3], leveraging the Bert architecture, encompasses various specialized versions tailored for diverse natural language processing tasks in French, akin to Camembert in its design philosophy.

•The training of Flaubert involved harnessing vast amounts of French text data, coupled with cutting-edge optimization techniques, resulting in models that exhibit remarkable fluency and accuracy in comprehending the complexities of the French language.

5

Model Performance Comparison

We fine-tuned Camembert and Flaubert for our task .The table below represents the various fine-tuned models along with their respective characteristics and performance.

Model	Data Cleaning	Parameters	Training Data(GB of text)	Precision	Recall	Accuracy	Execution Time(minutes)
Camembert-base	No	110M	OSCAR (138)	0.82	0.83	0.83	61
Camembert-base	Yes	110M	OSCAR (138)	0.815	0.81	0.815	61
camembert-base-oscar-4gb	No	110M	OSCAR (4)	0.83	0.83	0.83	60
Flaubert-base-Cased	No	138M	71	0.815	0.815	0.815	64
Flaubert-base-Uncased	No	138M	71	0.81	0.82	0.815	52
Flaubert-large	No	373M	71	0.815	0.815	0.815	150

6

Hyper-Parameters

In the hyper-parameter tuning phase of our models, we employed the Adam optimizer and experimented with various batch sizes. Our findings revealed that a batch size of 32 yielded the most optimal performances. Additionally, we explored different numbers of epochs and concluded that utilizing 2 epochs was optimal, as higher numbers led to overfitting issues. The figure below visually illustrates this :

7

CONCLUSION

In conclusion, our study underscores the superior performance of deep learning methods based on Transformers compared to classical machine learning algorithms for sentiment analysis on French Twitter data. Despite their effectiveness in capturing nuanced sentiments, it's essential to acknowledge the computational demands inherent to these advanced models. While they outshine traditional approaches in accuracy and efficacy, the reliance on substantial computational resources and GPUs for efficient execution poses challenges, potentially limiting their widespread adoption. Nevertheless, as technology progresses and efforts to optimize computational efficiency intensify, the promise of deep learning in sentiment analysis remains substantial, offering valuable insights into public opinion.

8

REFERENCES

[1]: KAGGLE , WWW.REALLYGREATSITE.COM
[2] : MARTIN, L., MULLER, B., ORTIZ SU´AREZ, P. J., DUPONT, Y., ROMARY, L., VILLEMONTE DE LA CLERGERIE, E., SEDDAH, D., AND ´ SAGOT, B. (2019). CAMEMBERT: A TASTY FRENCH LANGUAGE MODEL. ARXIV E-PRINTS, PAGE ARXIV:1911.03894, NOV
[3]:] LE H, VIAL L, FREJ J, SEGONNE V, COAVOUX M, LECOUTEUX B, ALLAUZEN A, CRABB´E B, BESACIER L, SCHWAB D. FLAUBERT: UNSUPERVISED LANGUAGE MODEL PRE-TRAINING FOR FRENCH. IN: PROCEEDINGS OF THE 12TH LANGUAGE RESOURCES AND EVALUATION CONFERENCE; 2020. P. 2479–90