

## A Sentiment Analysis Research Project:

---

### *An assessment of NLP Models:*

As the final project for the Fall 2022, Intro to Python course at the Hertie School, our team compared various NLP models to determine the accuracy and f1-score of said models.

To rank these models we ran comparisons on an aggregated data set collected via polling our friends and families on questions concerning wedding RSVPs - Data hyperlinked [HERE](#)

### **Background:**

---

Planning a wedding is highly challenging and costs a lot of time. To make the planning process a bit easier for couples, we wanted to have a sentiment analyzer in place to automatically analyze the responses to wedding invitations. For this, we made a research project to test five different pre-trained analyzers on a testing set with responses to wedding invitations and check their accuracy. This research project strives to answer the question: *Which is the best Sentiment Analyzer Tool to analyze wedding invitation responses?*

### **Process:**

---

- Collect Data: Asking friends about responses to a wedding invitation
- Analyzing the responses by using five sentiment analyzers
- Input: all responses
- Output: “pos” / “neg”
- Comparing the results to the true situation by using measurement tools
- Getting the best sentiment analysis tool for our purpose

*Our script had a total of six functions; 5 NLP models for sentiment analysis and one function for accuracy and an F1 score.*

### **How to Run:**

---

The file sentiment analysis in the [intro2python](#) folder contains the script necessary to run the application. Functions needed to run the file are saved in and imported. The functions, and applications necessary for download are as follows;

#### **Sentiment Analyzers**

- HappyTransformer
- Flair\_Analysis
- HuggingFace\_Bertweet
- TextBlob\_Analysis
- Vader\_Analysis

#### **Import**

- pandas
- Openpyxl
- TextClassifier

- NLTK\_data - Vader\_Lexicon
- Senten
- Split\_Single
- HappyTextClassification
- F1\_score
- Accuracy Score

### Final Output:

---

	model	accuracy	f1_score
3	happytransformer_prediction	0.800000	0.833333
0	huggingface_prediction	0.775000	0.816327
1	flair_prediction	0.750000	0.800000
2	textblob_prediction	0.650000	0.720000
4	vader_prediction	0.600000	0.714286

### Possible future improvements and extensions:

---

Collecting more data, and working towards tuning the model would aid further in ranking the NLP models an exciting extension to this project would be an API that could run sentiment analysis on emails.

### Authors:

*Anna Morhon, Irene Quadrelli, Maurice Rios*