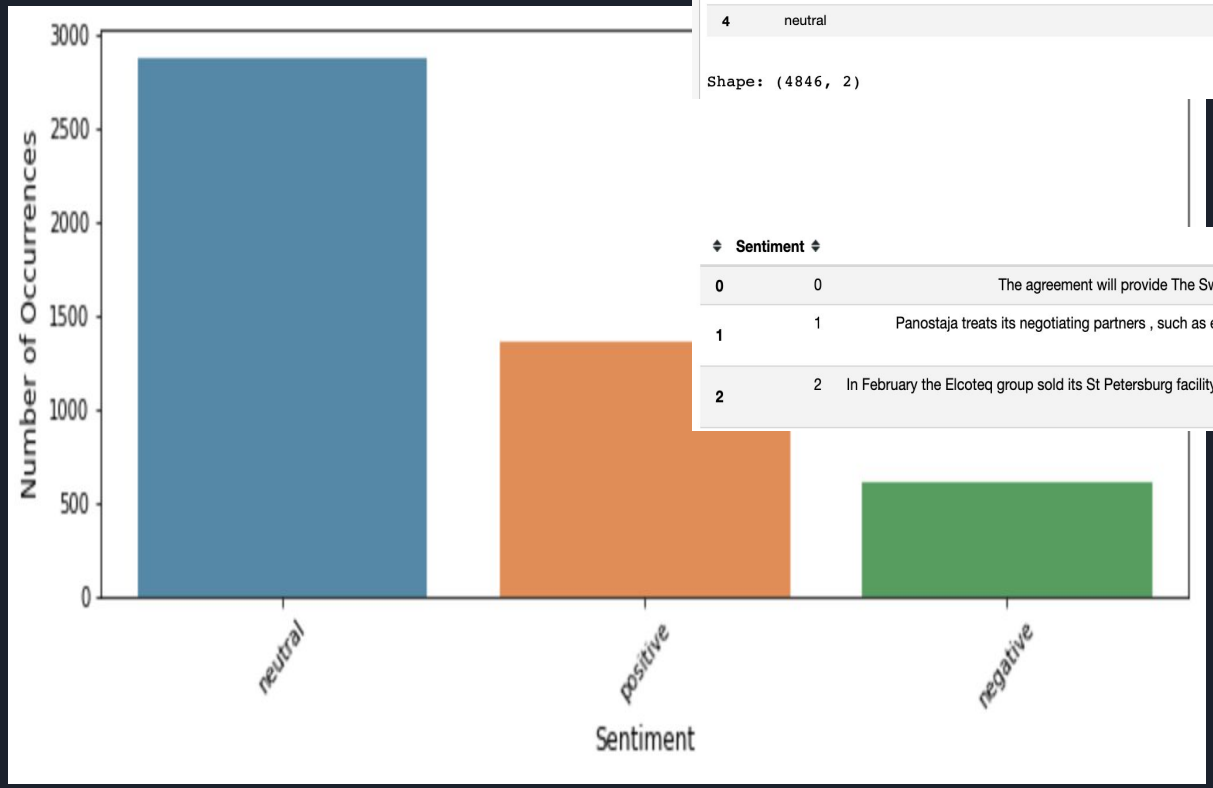
Financial News Sentiment Analysis Using Neural Networks

Arash Peimani
Flatiron Capstone
1/27/20201

The Data

	Sentiment	Headline
0	positive	The agreement will provide The Switch with double the converter capacity , whilst opening up further manufacturing locations in China .
1	neutral	Panostaja treats its negotiating partners , such as entrepreneurs selling their life s work and newcomer entrepreneurs seeking a career , fairly , justly and according to the law .
2	negative	In February the Elcoteq group sold its St Petersburg facility ; according to unconfirmed information the reason could have been supply problems because of the Russian customs service .
3	positive	Rinkuskiai raised the sales by 18.1 percent , to 1.37 million liters , while the sales of Kauno Alus grew by 14.3 percent , to 960,000 liters .
4	neutral	Unit prices for straddle carriers vary between EUR700 ,000 and EUR900 ,000 , the company added .

```
Shape: (4846, 2)
```



Sentiment Column Transformed for Processing

	Sentiment	Headline
0	0	The agreement will provide The Switch with double the converter capacity , whilst opening up further manufacturing locations in China .
1	1	Panostaja treats its negotiating partners , such as entrepreneurs selling their life s work and newcomer entrepreneurs seeking a career , fairly , justly and according to the law .
2	2	In February the Elcoteq group sold its St Petersburg facility ; according to unconfirmed information the reason could have been supply problems because of the Russian customs service .

```
neutral     2879
positive    1363
negative     604
```

Clean the Data

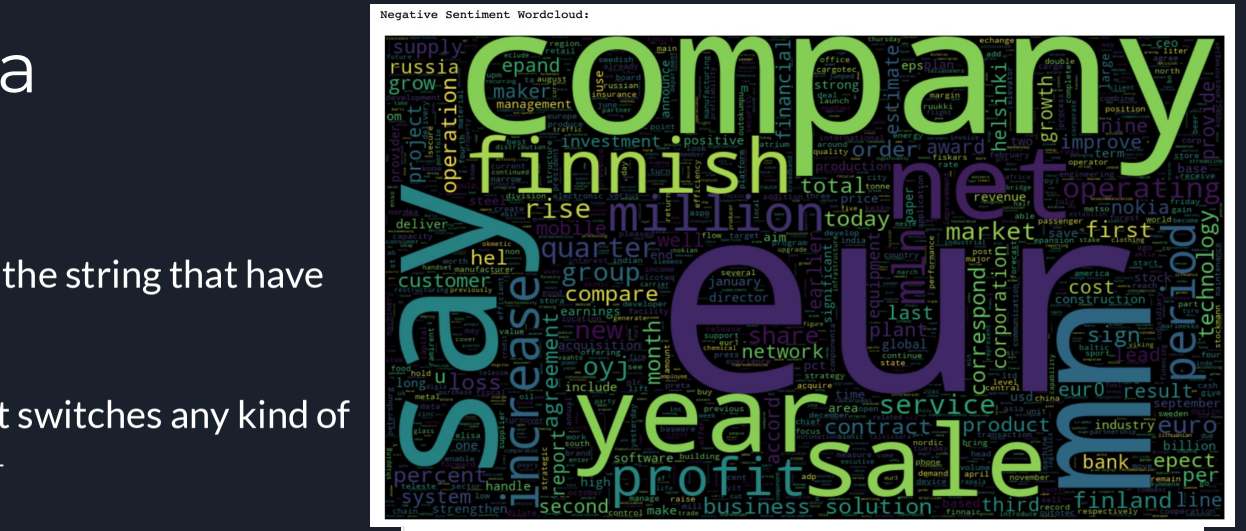
Stopwords: Remove words in the string that have no unique values.

Lemmatization: a method that switches any kind of a word to its base root mode.[1]

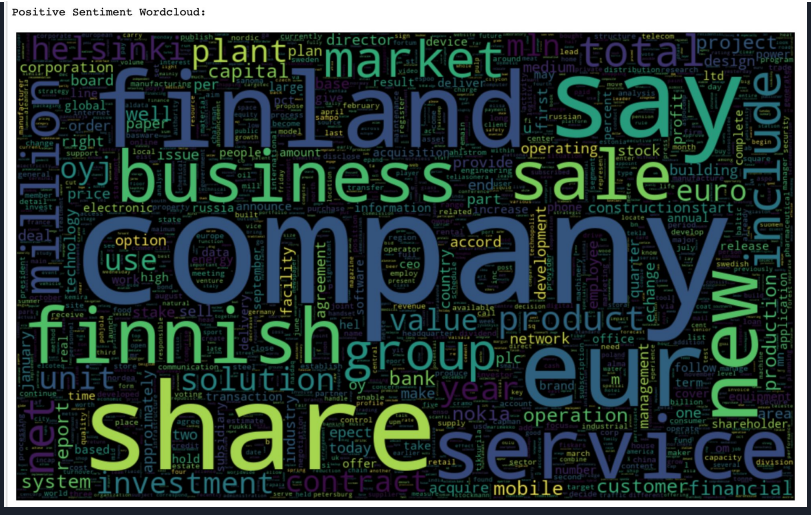
Beautiful Soup: used to strip any HTML tags and metadata

Tokenizer: Splitting sentences and words from the body of the text

Negative Sentiment Wordcloud:



Positive Sentiment Wordcloud:



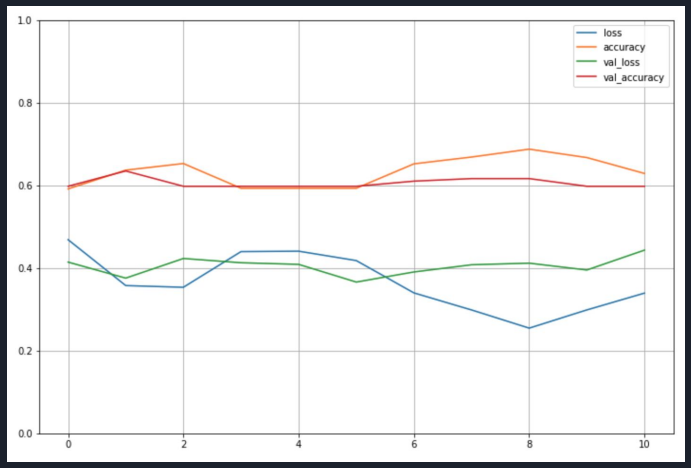
The Model and It's Evaluation

1. Hidden Layer

```
model = Sequential()
model.add(Embedding(num_words, embedding_dim, input_length=64))
model.add(Dense(200, activation='relu'))
model.add(Dropout(0.2)) # Optional Regularization
model.add(Dense(1, activation='sigmoid'))
```

2. LSTM (Long Short Term Memory)

```
model.add(Embedding(num_words, embedding_dim, input_length=max_length))
model.add(LSTM(64, dropout = 0.1))
model.add(Dense(32, activation='relu'))
model.add(Dense(1, activation='sigmoid'))
```




```
Train Loss: 30.066%
Train Accuracy: 60.502%
Test Loss: 40.929%
Test Accuracy: 59.922%
```

Testing The Model With Our Own Headlines

Scoring: 1=Positive, .5=Neutral, 0=Negative

```
Easier Headlines:
stock rise rapid for tesla
apple factory shut down, lay off hundred employee
average day of return for microsoft
WARNING:tensorflow:Model was constructed with shape (None, 64) for input Tensor("embedding_3_input:0", shape=(None,
64), dtype=float32), but it was called on an input with incompatible shape (None, 50).
[[[0.9118428 ]
  [0.13337815]
  [0.95264506]
```

```
Realistic Headlines:
snowflake more  double  debut  wall street embrace tech ipos
europe economic revival imperil raising the specter grind downturn
fed debate next step after shifting approach to rate setting
[[[0.85545903]
  [0.85545903]
  [0.03135353]
```

Future Work:

- Develop a Web Scraper to take Headlines from Financial News. Example: Reddit Financial Sites
- Fine Tune Neural Network for More Accurate Results.
- Cross Reference Sentiment Results with Stock Prices for that Day to Check Correlation.

THANK YOU

To the Flatiron Cohort and Staff for providing support and information to help and guide me through this project.

References

1. Analytics Steps
https://www.analyticssteps.com/blogs/what-stemming-and-lemmatization-nlp