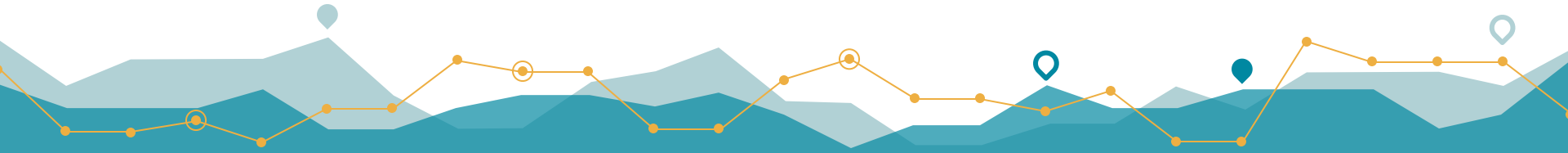


Group 33

FBI - FakeNews BERT Inspector

Sebastian Weidinger - Classification
Laura Weißl - Data Analysis

FBI GitHub Repository

MOTIVATION

- Large number of news articles contain misinformation
- How do Fake and No-Fake News articles differ due to e.g., word choice, topic and sentiment?
- Goals
  → Gather new information with text analysis
  → Classify type (Fake/No-Fake) with transformer-based model

MISINFORMATION & FAKE NEWS TEXT DATASET 79K (Kaggle)

‘Fake’

- 43.642 articles
- American right-wing extremist websites
- EUvsDisinfo project

‘True’

- 34.975 articles
- *Reuters, the New York Times, the Washington Post, …*

Misinformation Dataset Kaggle

METHODS

Data Analysis

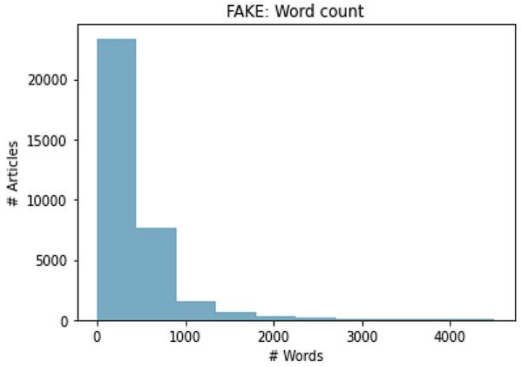
- Text Analysis (NLTK)
- NER (SpaCy)
- Topic Clusters (BERTopic)
- Sentiment Analysis
  - RoBERTa
  - DistilBERT

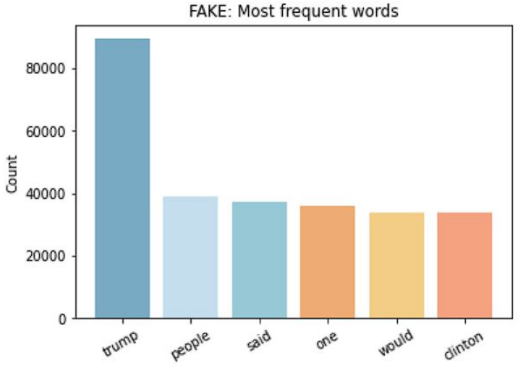
Fake News Classification

- Model Fine-Tuning
  - DistilBERT
  - Classification Layer

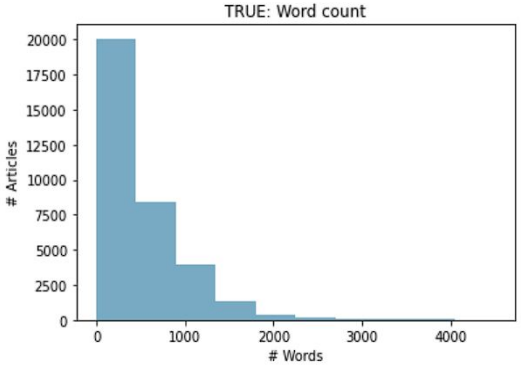
ANALYSIS

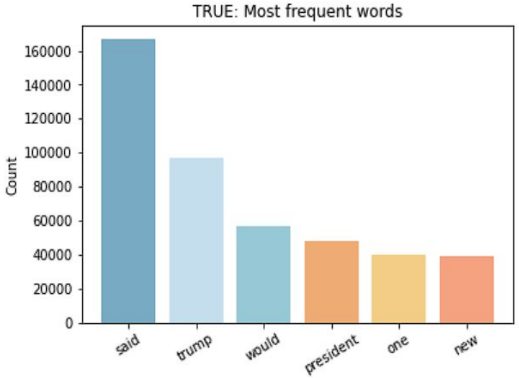
FAKE





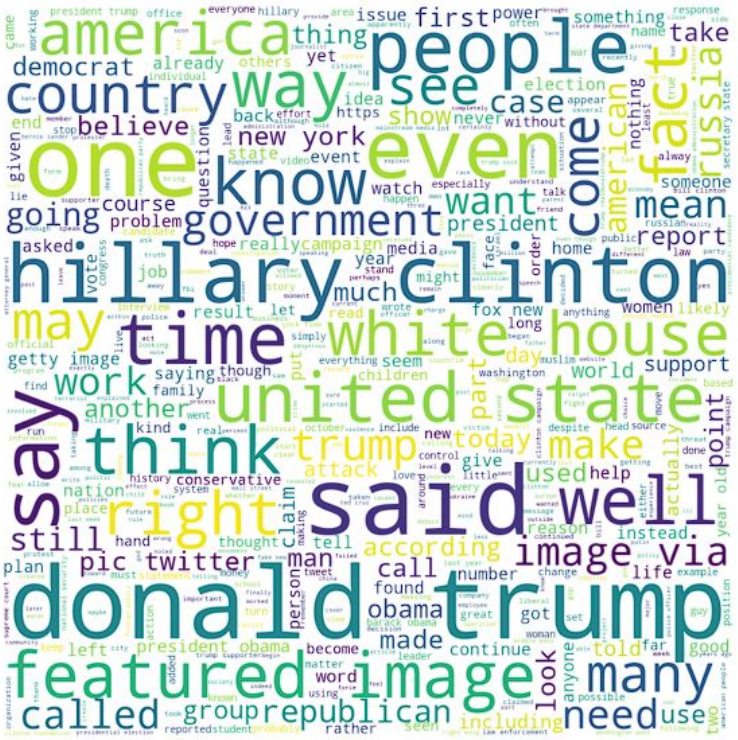
TRUE



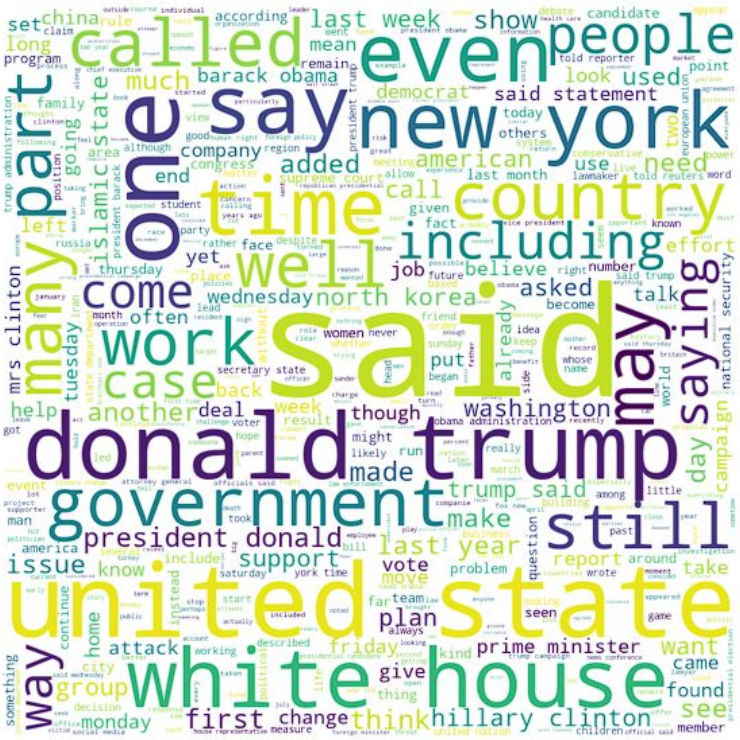


ANALYSIS

FAKE

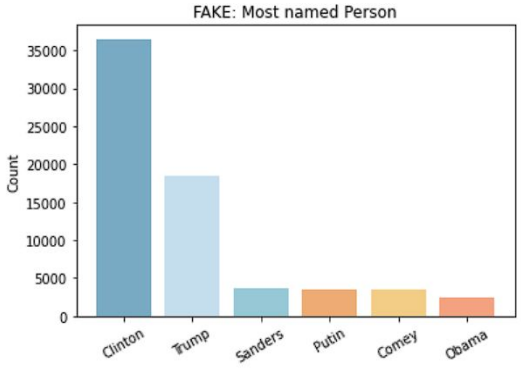


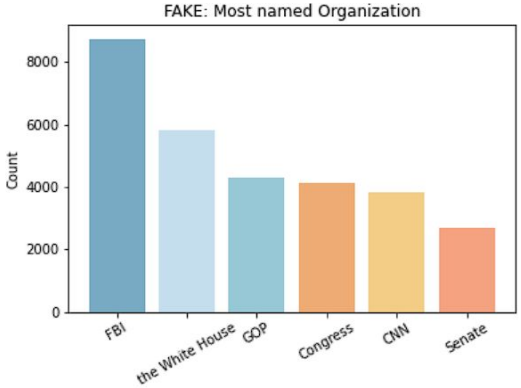
TRUE

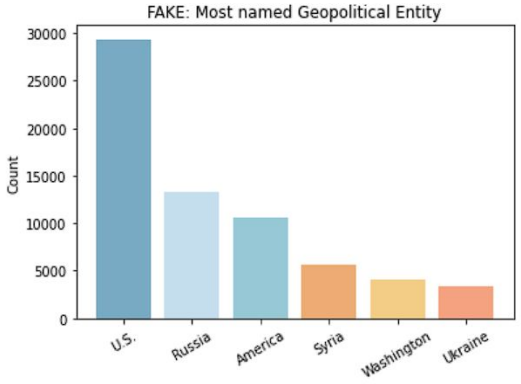


ANALYSIS

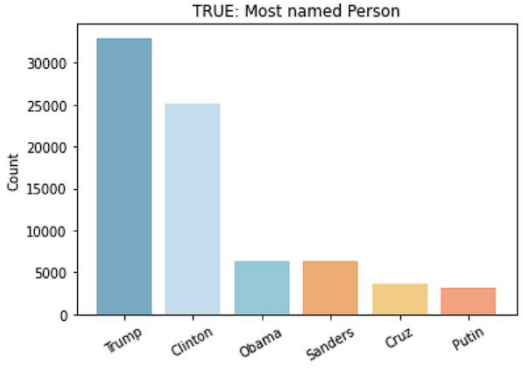
FAKE

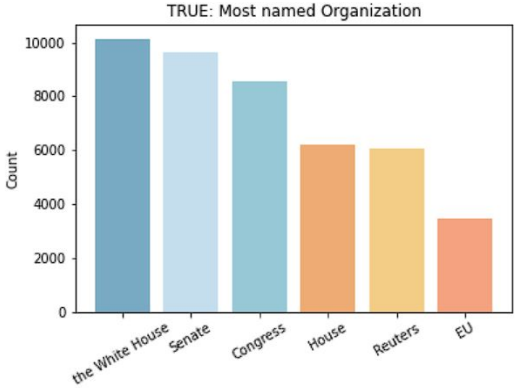


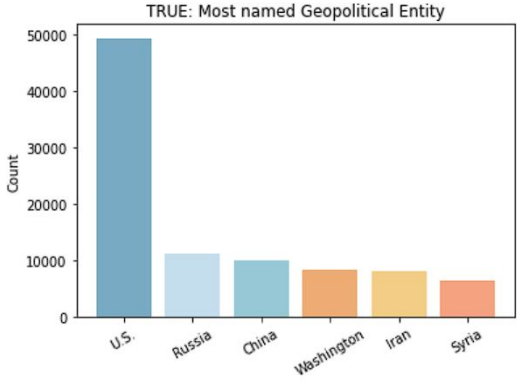




TRUE

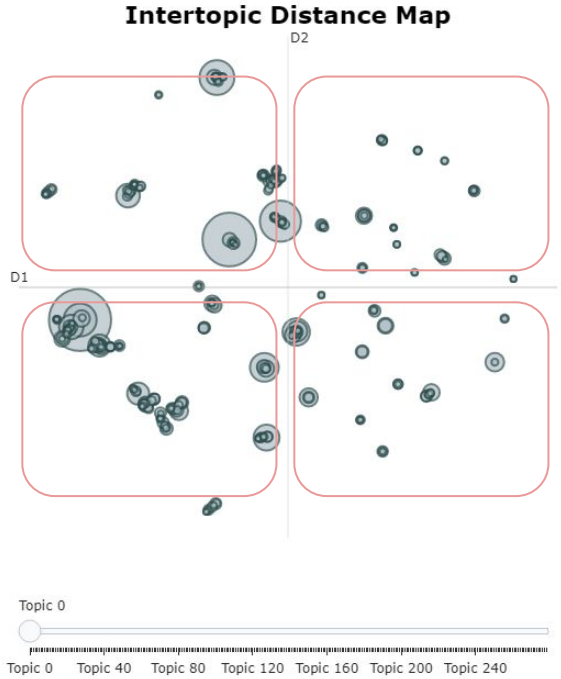






ANALYSIS

FAKE



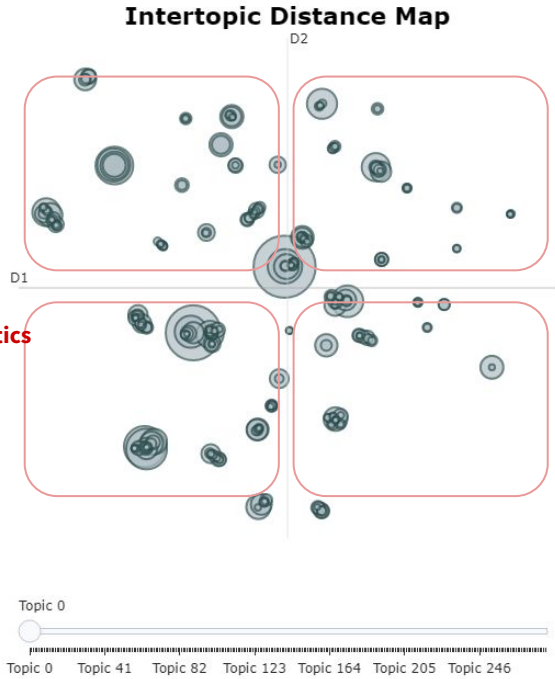
Political topics / Elections
ObamaCare, voters, …

Countries
Syria, Ukraine, Russia, …

Trump / Clinton
FBI, investigation, press, police,…

Law / Religion / Education
justice, students, school, ...

TRUE



Sports / Entertain-ment / Religion
movies, music, teams, ...

Countries \ Natural Disasters
Syria, Israel, hurricane, …

Internal politics / Environ-ment
ObamaCare, climate, tax, law, …

Foreign politics
foreign conflicts, foreign political parties, …

ANALYSIS

FAKE

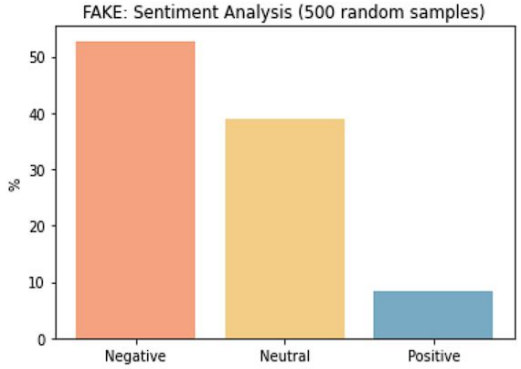
Count	Name
21356	-1_trump_people_said_one
1289	0_fbi_comey_emails_clinton
961	1_vote_voting_election_voters
563	2_mosul_syrian_aleppo_syria
404	3_ukraine_ukrainian_russia_russian

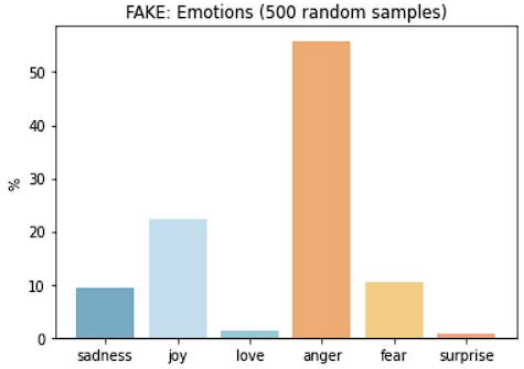
TRUE

Count	Name
16974	-1_trump_said_people_president
1207	0_korea_north_china_korean
950	1_intelligence_comey_russian_flynn
531	2_obamacare_bill_healthcare_senate
520	3_coal_climate_energy_oil

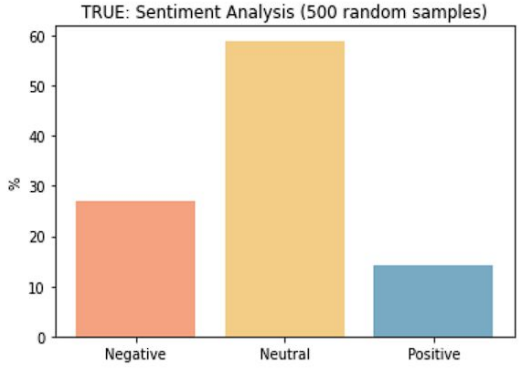
ANALYSIS

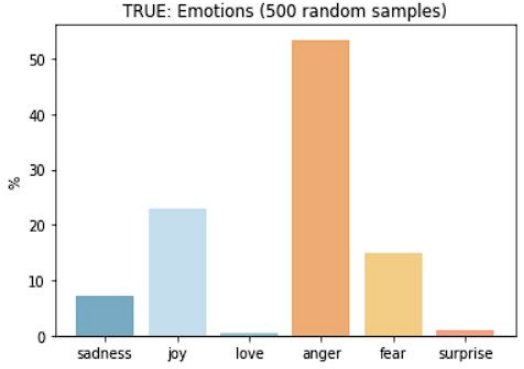
FAKE





TRUE

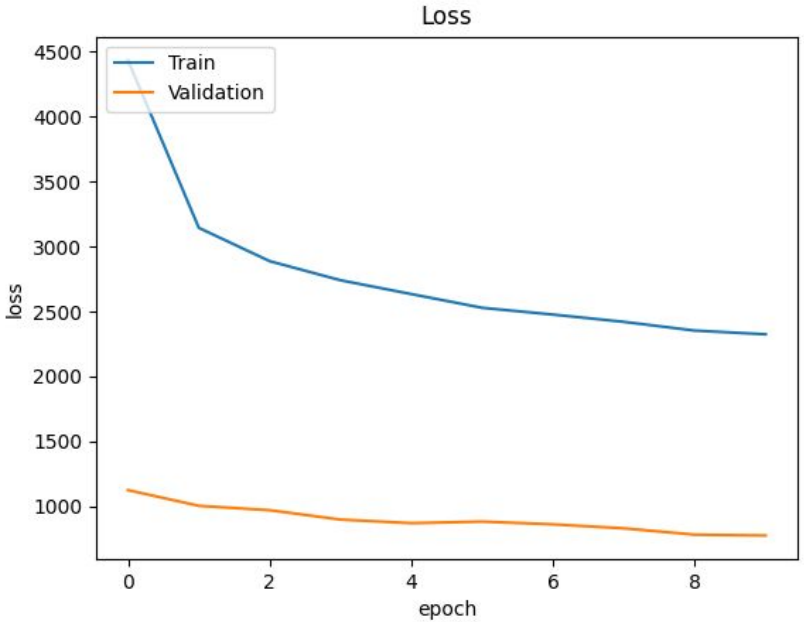


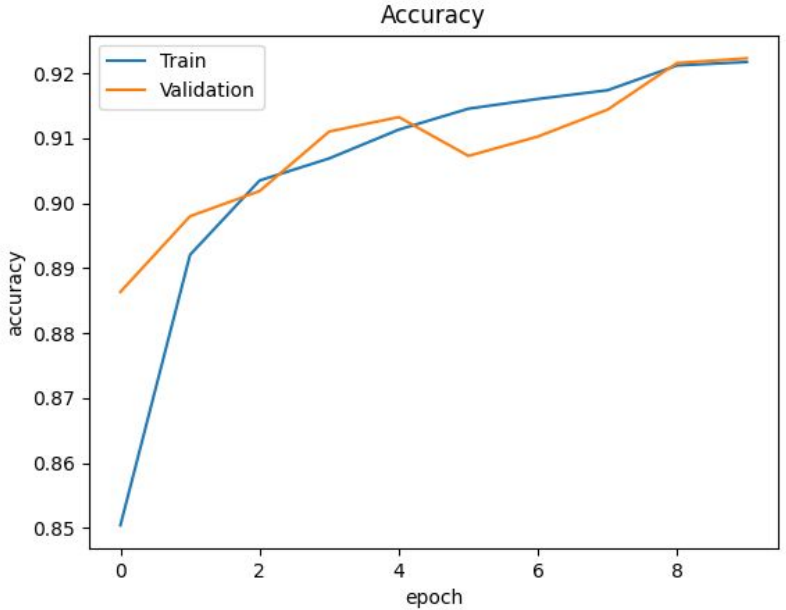


CLASSIFICATION - ARCHITECTURE

```
==========================================================================================================================
Layer (type:depth-idx)                                  Output Shape              Param #
==========================================================================================================================
FakeDetectorModel                                       [1, 1]                    --
├─DistilBertModel: 1-1                                  [[1, 512, 768]]           --
│    └─Embeddings: 2-1                                  [1, 512, 768]             --
│    │    └─Embedding: 3-1                              [1, 512, 768]             (23,440,896)
│    │    └─Embedding: 3-2                              [1, 512, 768]             (393,216)
│    │    └─LayerNorm: 3-3                              [1, 512, 768]             (1,536)
│    │    └─Dropout: 3-4                                [1, 512, 768]             --
│    └─Transformer: 2-2                                 [[1, 512, 768]]           --
│    │    └─ModuleList: 3-5                             --                        (42,527,232)
├─Linear: 1-2                                           [1, 512]                  393,728
├─ReLU: 1-3                                             [1, 512]                  --
├─Dropout: 1-4                                          [1, 512]                  --
├─Linear: 1-5                                           [1, 1]                    513
==========================================================================================================================
Total params: 66,757,121
Trainable params: 394,241
Non-trainable params: 66,362,880
Total mult-adds (M): 66.76
==========================================================================================================================
```

CLASSIFICATION - TRAINING



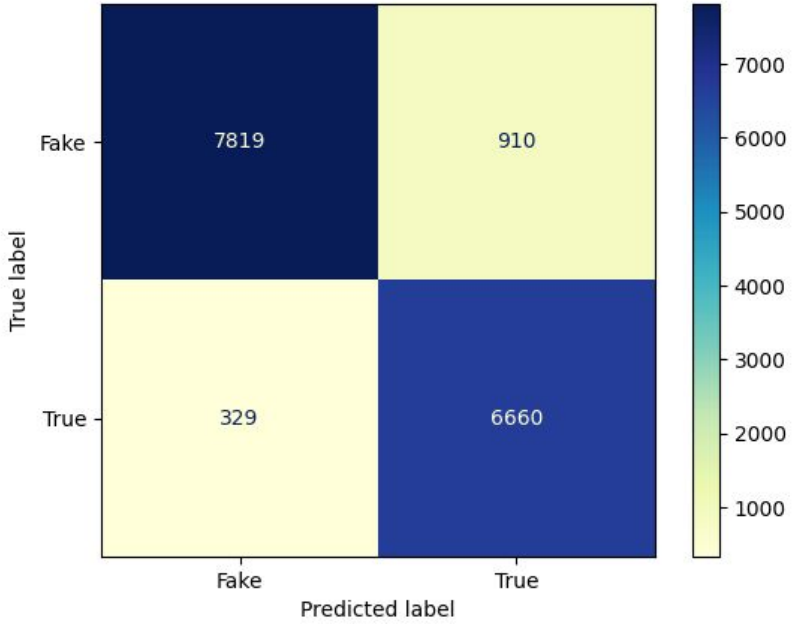


CLASSIFICATION - EVALUATION

	precision	**recall**	**f1-score**	**support**
Fake	0.96	0.90	0.93	8729
True	0.88	0.95	0.91	6989
accuracy			**0.92**	15718

Comparison model Word2Vec + LSTM: accuracy ~96%

CLASSIFICATION - CONFUSION MATRIX



CLASSIFICATION - SENTIMENT ANALYSIS

FAKE

Whole Dataset

Negative:	53%
Neutral:	39%
Positive:	8%

True Negative

Negative:	53%
Neutral:	39%
Positive:	8%

False Positive

Negative:	42%
Neutral:	47%
Positive:	11%

TRUE

Whole Dataset

Negative:	27%
Neutral:	59%
Positive:	14%

True Positive

Negative:	35%
Neutral:	53%
Positive:	11%

False Negative

Negative:	43%
Neutral:	44%
Positive:	13%

ANALYSIS

	precision	recall	f1-score	support
Fake	**0.97**	**0.92**	**0.95**	**5394**
True	0.83	0.92	0.87	2087
accuracy			**0.92**	7481
Fake	0.95	0.84	0.89	2991
True	**0.91**	**0.97**	**0.94**	**4625**
accuracy			**0.92**	7616
Fake	0.92	0.89	0.90	344
True	0.87	0.91	0.89	277
accuracy			**0.90**	621

Negative Messages

Neutral Messages

Positive Messages

CONCLUSION

- Fake and Non-Fake news differ especially in:
  - Topics
  - Sentiments
- Classification might be influenced by sentiment
- Correlation:
  - Fake news tend to have a negative sentiment
  - Non-Fake news are written more neutral

Q & A