

Group 17

Classification of Fake and Real News

<https://github.com/StellaLukasser/AIR>

Armin Bodingbauer – preprocessing, model-design, visualization, integration

Nur El-Din El-Rez – model-design, slides

Stella Lukasser – preprocessing, model-design, visualization, integration, slides

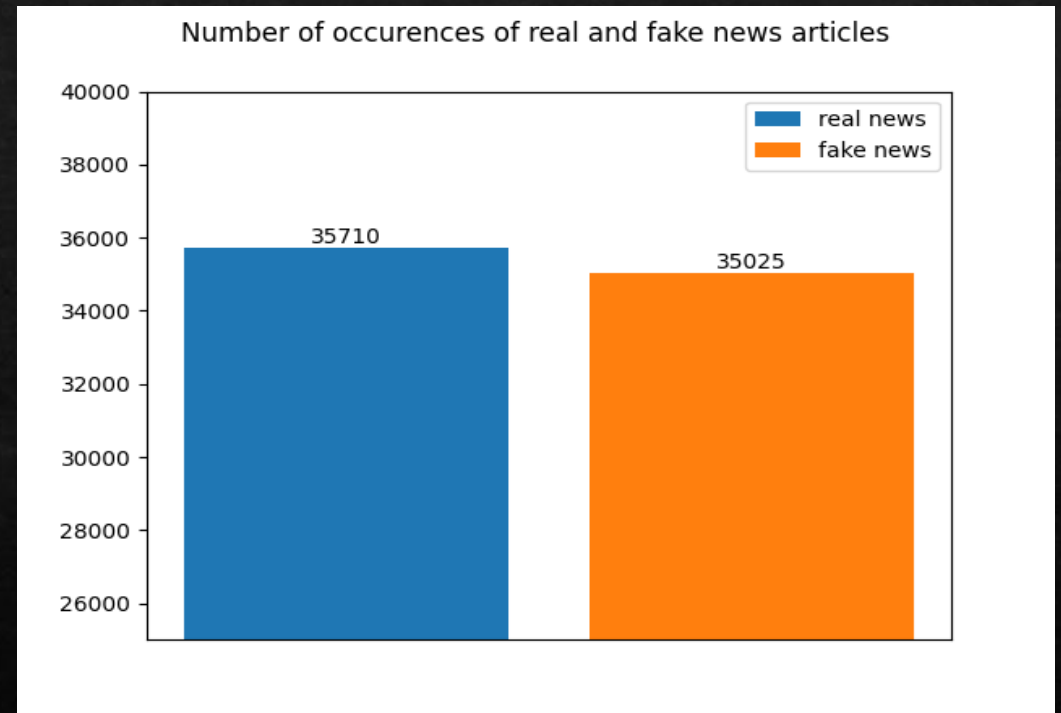
Frederik Wagner – preprocessing , integration

Motivation

Is it possible to classify fake and real news articles just based on their titles? How is the performance compared to using the whole news text as input?

Data Set

- ◆ 72,134 news articles
- ◆ Real 50.5 %
- ◆ Fake 49.5 %
- ◆ Merged from Kaggle, McIntire, Reuters, BuzzFeed Political

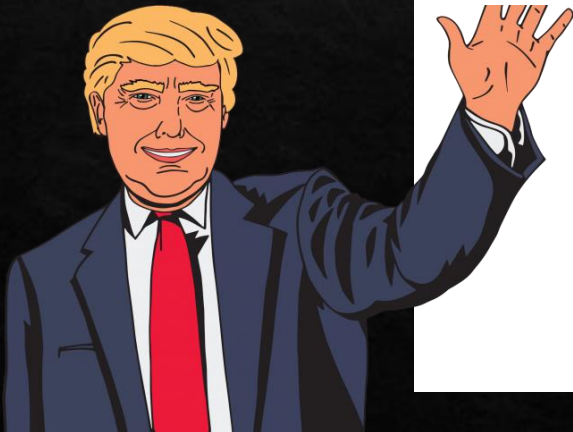
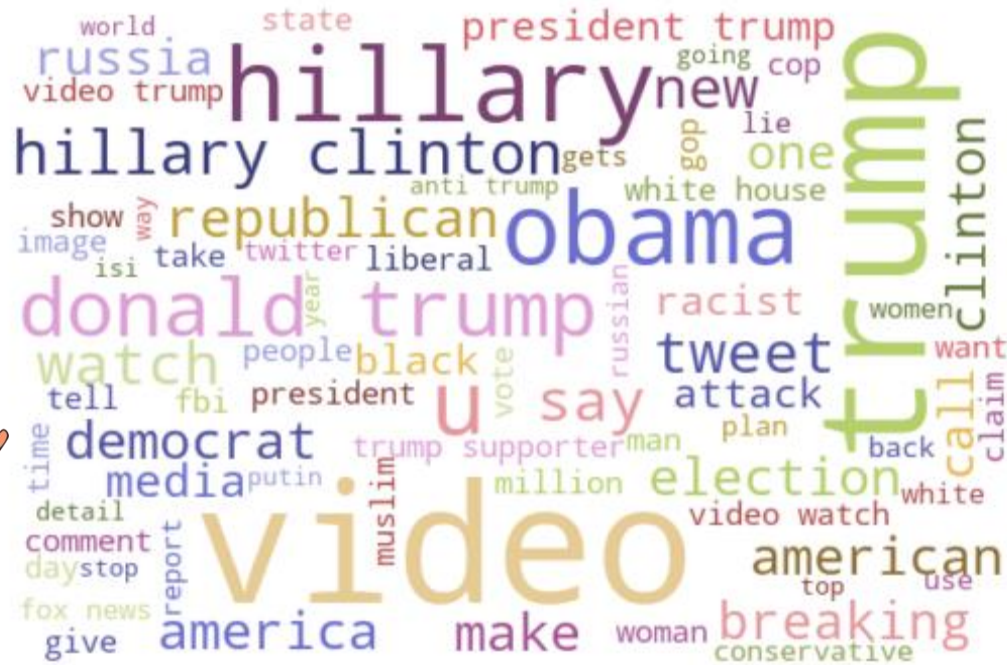


Word cloud <Fake>

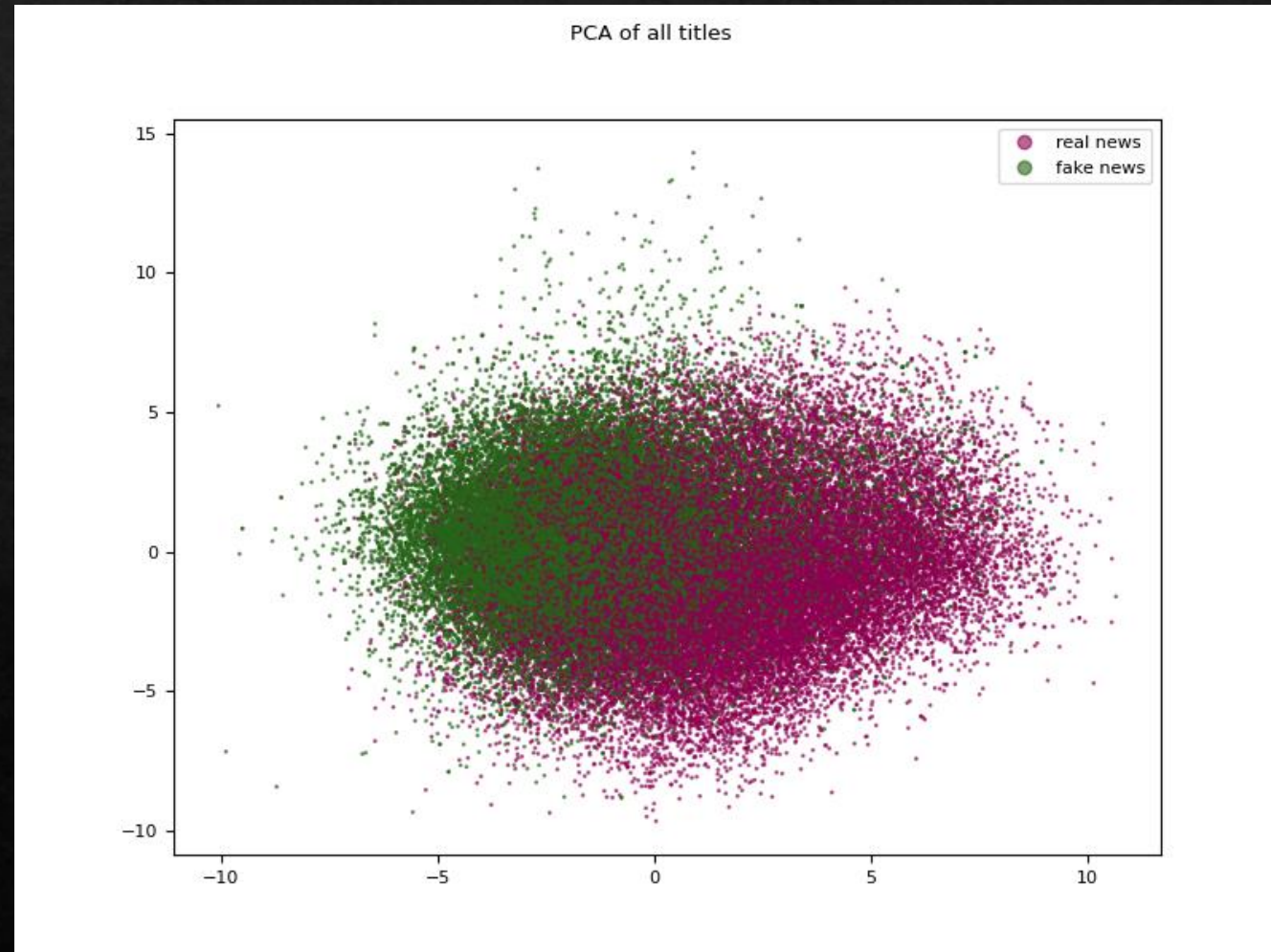


ex white house state face times u
call russian bill eu plan vote
report win new fight deal see
party china rule iran talk source say may
breitbart north korea back
end democrat support turkey protest syria president
year law times
trump say poll obama clinton government donald trump
leader case republican former want official hillary clinton u n
take russia election times trump supreme court
first islamic state meet american attack congress one

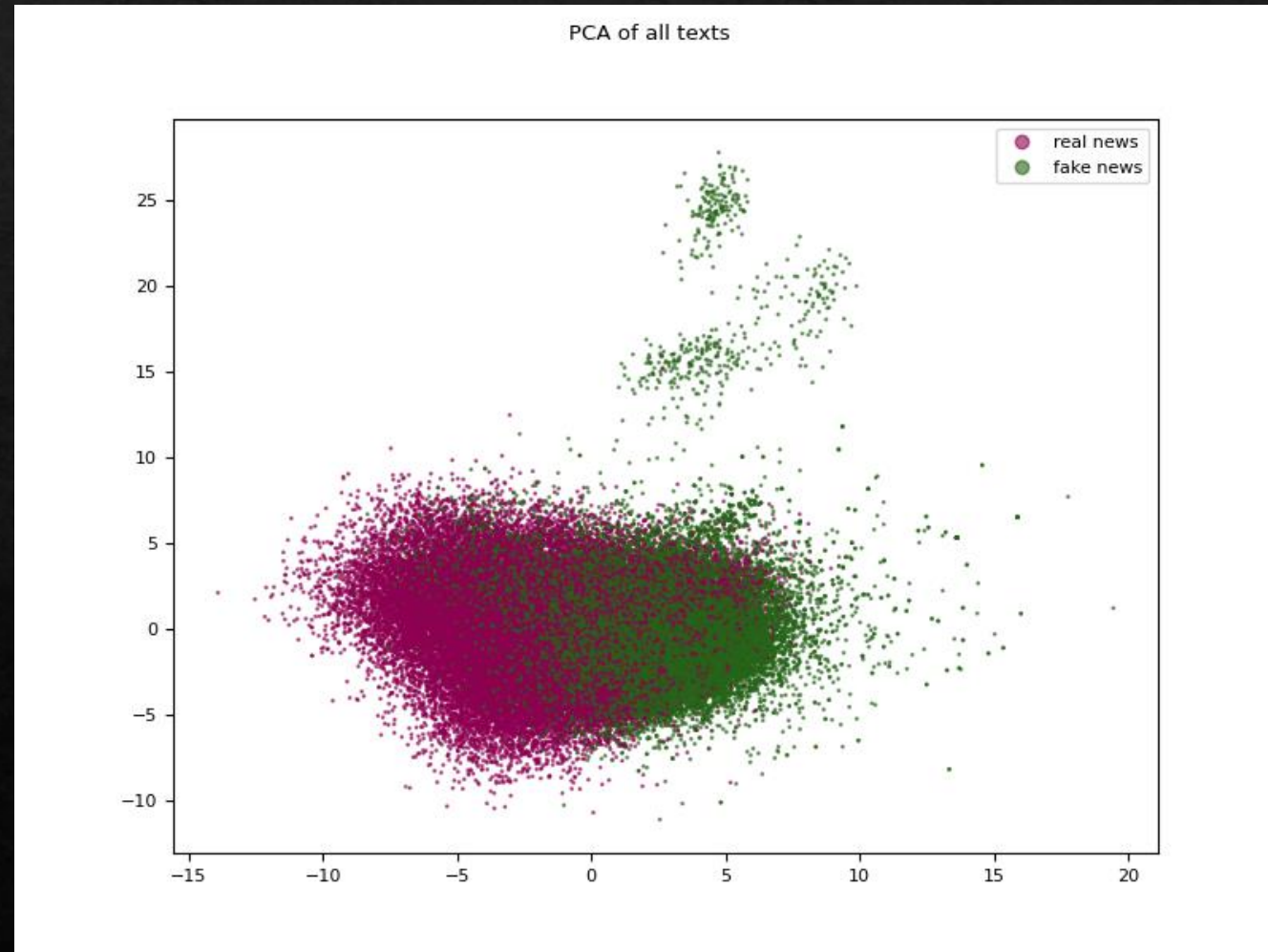
Word cloud <Real>



Distribution of real/fake news by TITLE



Distribution of real/fake news by TEXT



Preprocessing

- ❖ Skip short lines (length ≤ 10)
- ❖ Tokenization with NLTK
- ❖ Stopword removal
- ❖ Stemming

Methods

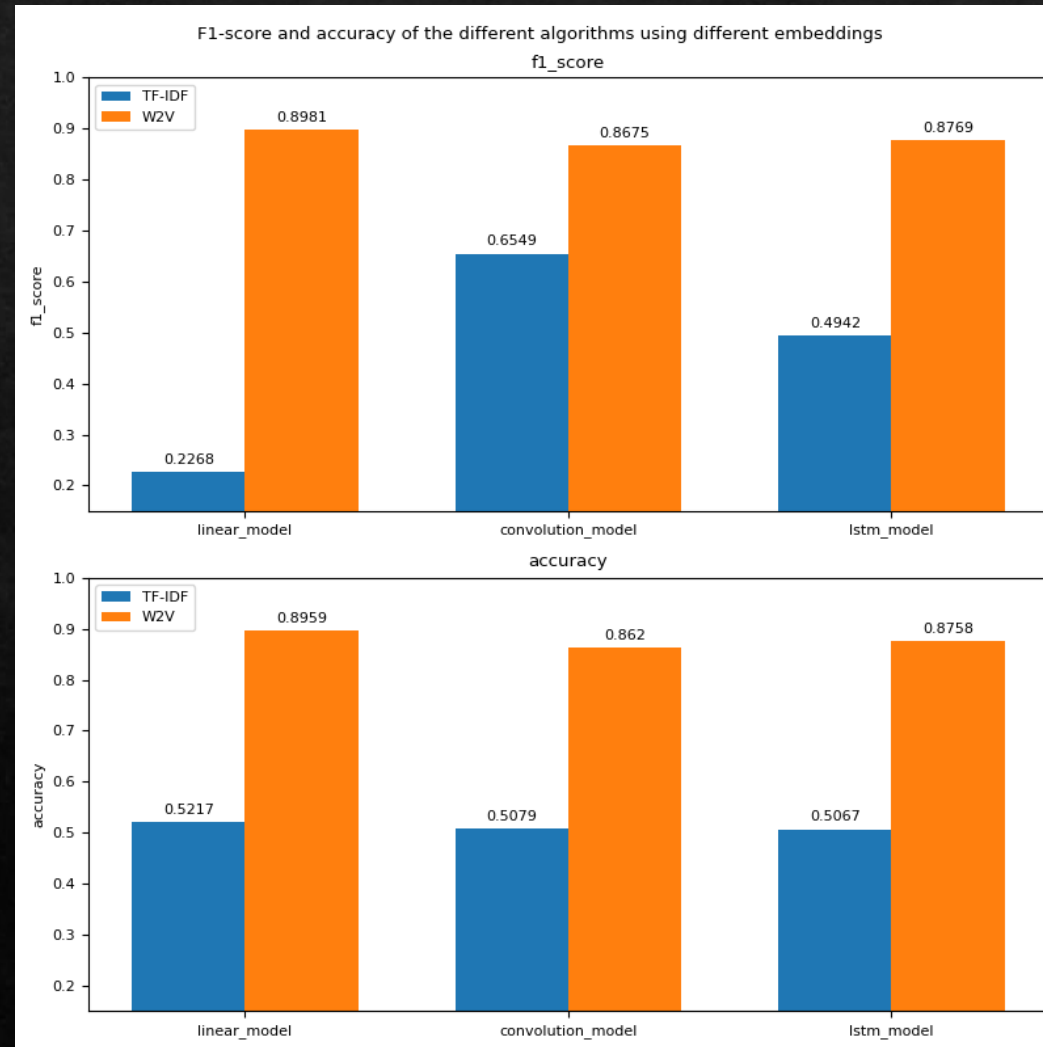
Word Embeddings

- ◆ TF-IDF
- ◆ Word2Vec

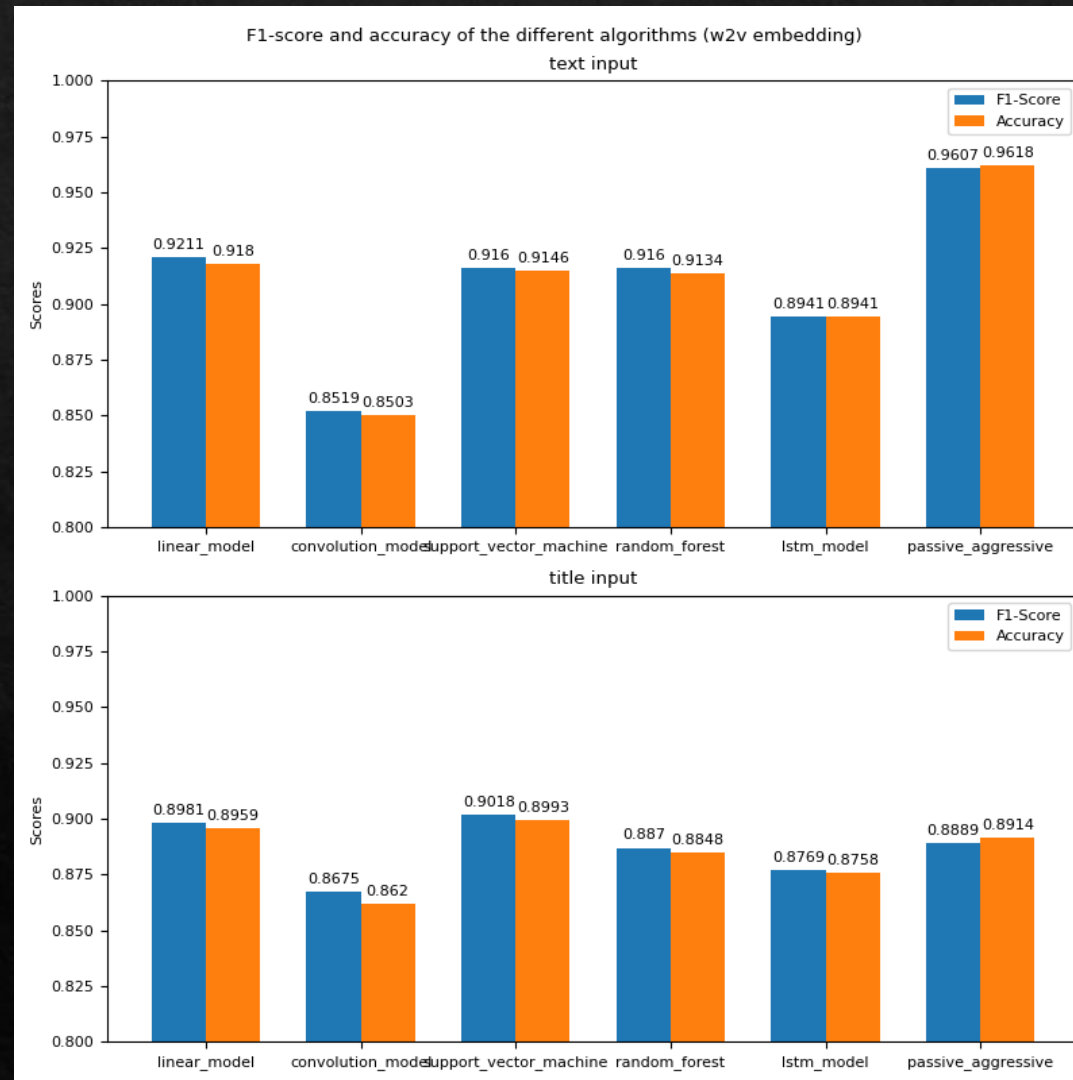
Models

- ◆ Linear Regression
- ◆ Convolutional model
- ◆ Bidirectional LSTMs
- ◆ Support Vector Machines
- ◆ Random Forests
- ◆ Passive Aggressive Classifier

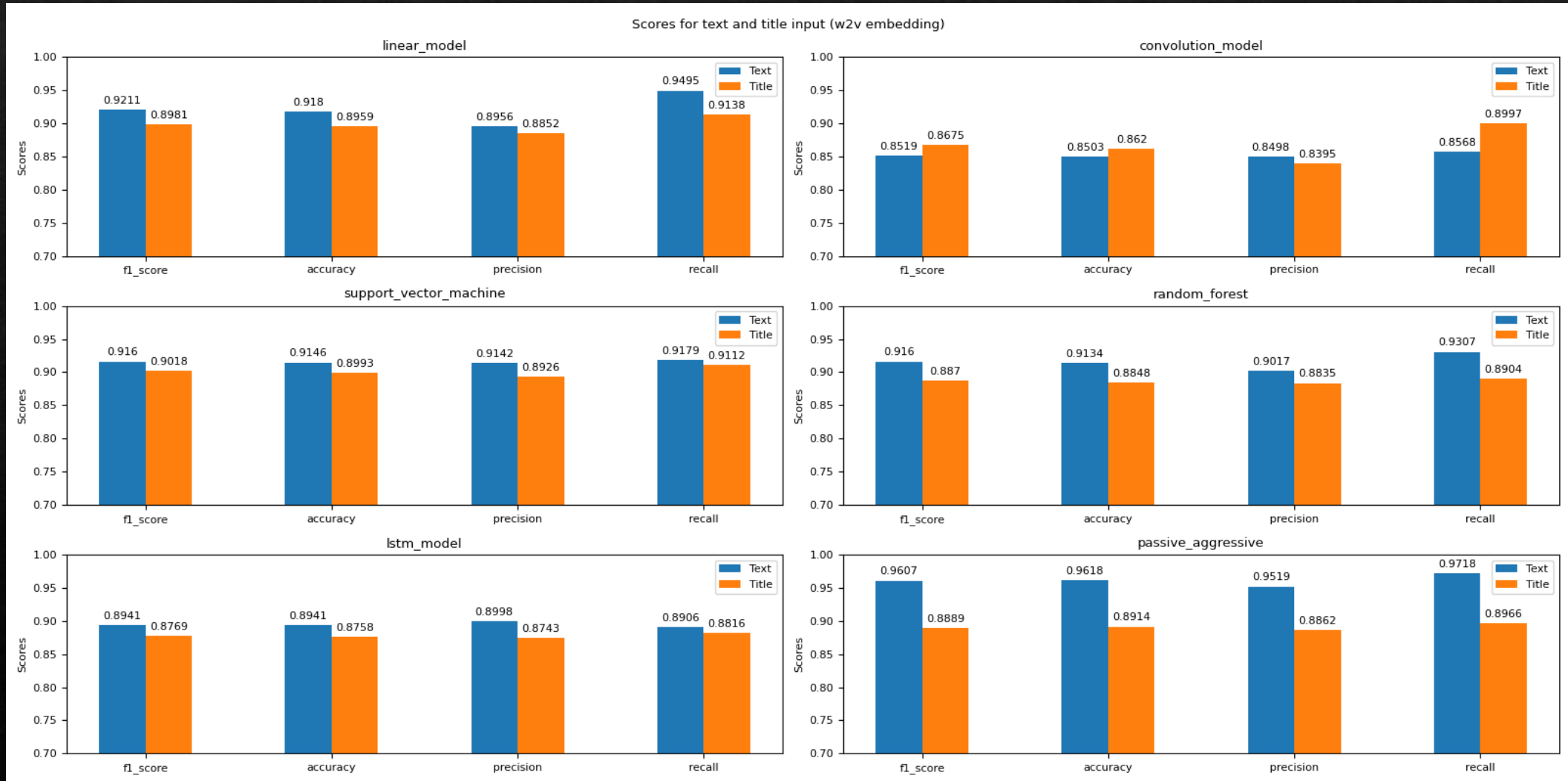
Comparing word embeddings



Comparing different models



Performance comparison TEXT and TITLE



Conclusion

- ◆ Overall classification on text better performance
- ◆ Performance difference negligible for most tasks
- ◆ Difference only significant for passive aggressive classifier

Group 17

Classification of Fake and Real News

<https://github.com/StellaLukasser/AIR>

Armin Bodingbauer – preprocessing, model-design, visualization, integration

Nur El-Din El-Rez – model-design, slides

Stella Lukasser – preprocessing, model-design, visualization, integration, slides

Frederik Wagner – preprocessing , integration