KU LEUVEN

Benchmark Study in Flemish Twitter Sentiment Analysis

Outline

- Introduction
- Research Questions
- Methodology
- Results
- Discussion
- Conclusion

Introduction

Twitter

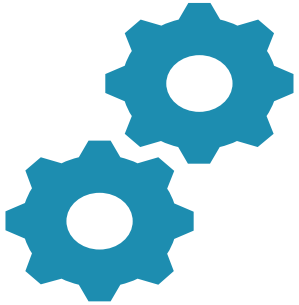
- Over 370 million users
- More than 500 million tweets a day
- Massive pile of data

Sentiment Analysis

- Research field that detects people's sentiment, opinions, emotions in a written text

Introduction

Lexicon-Based Methods



Traditional Machine Learning Models

Neural Networks

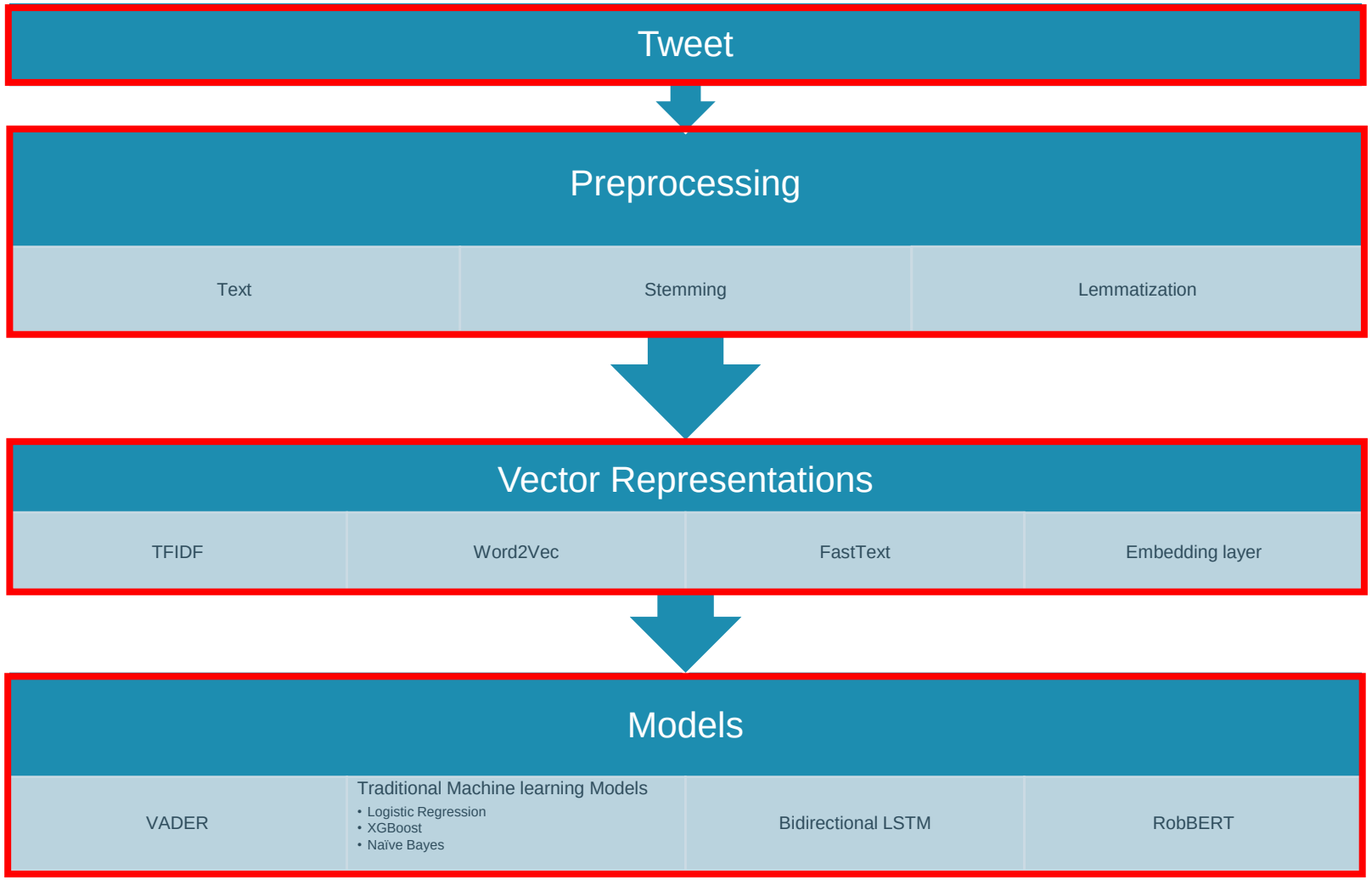
Attention-Based Models

Research Questions

How are the results affected by using different models and is increasing complexity beneficial?

How do different preprocessing methods influence the results of sentiment analysis?

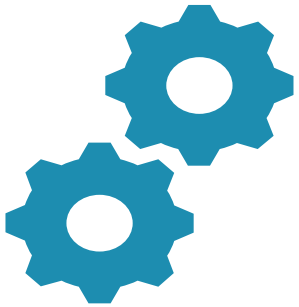
Methodology: Pipeline



Methodology

- Dictionary
- Interpretable
- Simple method

Lexicon-Based Methods



Traditional Machine Learning Models

- Often less of a black box
- Not computationally expensive

- Many parameters
- High computation power
- Black box
- Good results

Neural Networks

Attention-Based Models

- Many parameters
- Computationally expensive
- State-of-the-art

Methodology: evaluation

Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$
Macro F1	$\frac{1}{\lvert T \rvert} \sum_{t \in T} \frac{2P_t R_t}{P_t + R_t}$
Macro Precision	$\frac{\sum_{t \in T} TP_t / (TP_t + FP_t)}{\lvert T \rvert}$
Macro Recall	$\frac{\sum_{t \in T} TP_t / (TP_t + FN_t)}{\lvert T \rvert}$

Results

Model	Accuracy	Macro F1	Macro precision	Macro recall
VADER (lemmatized)	0,4283	0,3215	0,5198	0,3870
LogReg TFIDF (lemmatized)	0,6091	0,6030	0,6114	0,6012
XgBoost Word2Vec pretrained (lemmatized)	0.5908	0.5848	0.5878	0.5846
Multinomial Naïve Bayes TFIDF (lemmatized)	0,6041	0,5957	0,6084	0,5956
Bernoulli Naïve Bayes TFIDF (Stemmed)	0,6037	0,5976	0,6021	0,5968
Bidirectional LSTM	0,6053	0,5985	0,6024	0,5986
Finetuned RobBERT	**0,6559**	**0,6507**	**0,6537**	**0,6507**

Conclusion

- How are the results affected by using different models and is increasing complexity beneficial?
  - RobBERT outperforms all other models
  - Lexicon-based model is very bad
- How do different preprocessing methods influence the results of sentiment analysis?
  - Not much difference; lemmatization seems better than stemming