Identification and categorization of offensive language in German tweets

Kinga Gémes

kinga.gemes@tuwien.ac.at

12 May 2021
Warning!

The following presentation contains foul language.
What is toxicity and offensive language?

**Toxicity:** An extremely harsh, malicious, or harmful quality. (Merriam-Webster dictionary\(^1\))

**Offensive:** Something that is offensive upsets or embarrasses people because it is rude or insulting. (Collins dictionary\(^2\))

---

\(^1\)https://www.merriam-webster.com/

\(^2\)https://www.collinsdictionary.com/
Classic approaches

- Dictionaries:

  - Slang
  - Emotion
  - WordNet and other knowledge bases

Problems:

- l3375P3Ak (Leetspeak)
- Ever evolving language
- Sarcasm

Kinga Gémes (kinga.gemes@tuwien.ac.at)

Identification and categorization of offensive language in German tweets
Classic approaches

- Dictionaries:
  - Slang

Kinga Gémes (kinga.gemes@tuwien.ac.at)  Identification and categorization of offensive language in German tweets
Classic approaches

- Dictionaries:
  - Slang
  - Emotion

Kinga Gémes (kinga.gemes@tuwien.ac.at)
Identification and categorization of offensive language in German tweets
Classic approaches

- Dictionaries:
  - Slang
  - Emotion
  - WordNet and other knowledge bases
Classic approaches

- Dictionaries:
  - Slang
  - Emotion
  - WordNet and other knowledge bases

- Problems:
Classic approaches

- Dictionaries:
  - Slang
  - Emotion
  - WordNet and other knowledge bases

- Problems:
  - l3375P3Ak (Leetspeak)
Classic approaches

- Dictionaries:
  - Slang
  - Emotion
  - WordNet and other knowledge bases

- Problems:
  - l3375P3Ak (Leetspeak)
  - Ever evolving language
Classic approaches

- Dictionaries:
  - Slang
  - Emotion
  - WordNet and other knowledge bases

- Problems:
  - l3375P3Ak (Leetspeak)
  - Ever evolving language
  - Sarcasm
Datasets

- GermEval2018 - 5009 + 3398 German tweets
Datasets

- GermEval2018 - 5009 + 3398 German tweets
- GermEval2019 - 3980 + 3031 German tweets
Datasets

- GermEval2018 - 5009 + 3398 German tweets
- GermEval2019 - 3980 + 3031 German tweets
- HASOC2019 - 3819 + 850 German tweets
Twitter data processing

- @username can be masked
Twitter data processing

- @username can be masked
- numbers, urls, dates can be masked
Twitter data processing

- @username can be masked
- numbers, urls, dates can be masked
- typo correction
Twitter data processing

- @username can be masked
- numbers, urls, dates can be masked
- typo correction
- emoticons can be replaced by their textual representations

#ImportantHashtag - what should we do?
- cut it up by the camel case and remove the #
- leave it as is, but remove the #
- mask it completely
Twitter data processing

- @username can be masked
- numbers, urls, dates can be masked
- typo correction
- emoticons can be replaced by their textual representations
- #ImportantHashtag - what should we do?
Twitter data processing

- @username can be masked
- numbers, urls, dates can be masked
- typo correction
- emoticons can be replaced by their textual representations

#ImportantHashtag - what should we do?
- cut it up by the camel case and remove the #
Twitter data processing

- @username can be masked
- numbers, urls, dates can be masked
- typo correction
- emoticons can be replaced by their textual representations
- #ImportantHashtag - what should we do?
  - cut it up by the camel case and remove the 
  - leave it as is, but remove the 

Kinga Gémes (kinga.gemes@tuwien.ac.at)  Identification and categorization of offensive language in German tweets
Twitter data processing

- @username can be masked
- numbers, urls, dates can be masked
- typo correction
- emoticons can be replaced by their textual representations
- #ImportantHashtag - what should we do?
  - cut it up by the camel case and remove the #
  - leave it as is, but remove the #
  - mask it completely
Datasets

- GermEval2018 - 5009 + 3398 German tweets
- GermEval2019 - 3980 + 3031 German tweets
- HASOC2019 - 3819 + 850 German tweets
Datasets

- GermEval2018 - 5009 + 3398 German tweets
- GermEval2019 - 3980 + 3031 German tweets
- HASOC2019 - 3819 + 850 German tweets

Subtasks

Kinga Gémes (kinga.gemes@tuwien.ac.at)
Identification and categorization of offensive language in German tweets
Datasets

- GermEval2018 - 5009 + 3398 German tweets
- GermEval2019 - 3980 + 3031 German tweets
- HASOC2019 - 3819 + 850 German tweets

Subtasks

- Binary classification - offense or not
Datasets

- GermEval2018 - 5009 + 3398 German tweets
- GermEval2019 - 3980 + 3031 German tweets
- HASOC2019 - 3819 + 850 German tweets

Subtasks

- Binary classification - offense or not
- Fine-grained classification - offense categories
Datasets

- GermEval2018 - 5009 + 3398 German tweets
- GermEval2019 - 3980 + 3031 German tweets
- HASOC2019 - 3819 + 850 German tweets

Subtasks

- Binary classification - offense or not
- Fine-grained classification - offense categories
- Binary classification - explicit or implicit

Kinga Gémes (kinga.gemes@tuwien.ac.at)
Abuse: The tweet does not just insult a person but represents the stronger form of abusive language. By abuse we define a special type of degradation. This type of degrading consists in ascribing a social identity to a person that is judged negatively by a (perceived)majority of society. The identity in question is seen as a shameful, unworthy, morally objectionable or marginal identity. E.g. Ich persönlich scheisse auf die grüne Kinderfickerpartei
Abuse: The tweet does not just insult a person but represents the stronger form of abusive language. By abuse we define a special type of degradation. This type of degrading consists in ascribing a social identity to a person that is judged negatively by a (perceived)majority of society. The identity in question is seen as a shameful, unworthy, morally objectionable or marginal identity. E.g. *Ich persönlich scheisse auf die grüne Kinderfickerpartei*

Insult: The tweet clearly wants to offend someone. E.g. *ein #Tatort mit der Presswurst #Saalfeld geht gar nicht #ARD*
Abuse: The tweet does not just insult a person but represents the stronger form of abusive language. By abuse we define a special type of degradation. This type of degrading consists in ascribing a social identity to a person that is judged negatively by a (perceived) majority of society. The identity in question is seen as a shameful, unworthy, morally objectionable or marginal identity. E.g. *Ich persönlich scheisse auf die grüne Kinderfickerpartei*

Insult: The tweet clearly wants to offend someone. E.g. *ein #Tatort mit der Presswurst #Saalfeld geht gar nicht #ARD*

Profanity: Usage of profane words, however, the tweet clearly does not want to insult anyone. E.g. *Juhu, das morgige Wetter passt zum Tag SCHEIßWETTER*
GermEval2018 data distribution

- Other: 66.3%
- Abuse: 20.4%
- Insult: 11.88%
- Profanity: 1.42%
GermEval2018 data distribution

Abuse: 21.69%
Insult: 10.83%
Profanity: 1.32%
Other: 66.16%

Identification and categorization of offensive language in German tweets by Kinga Gémes (kinga.gemes@tuwien.ac.at)
GermEval2019 data distribution

- Other: 67.79%
- Abuse: 12.71%
- Insult: 15.68%
- Profanity: 3.82%
GermEval2019 data distribution

- Other: 68%
- Abuse: 13.2%
- Insult: 15.14%
- Profanity: 3.66%
HASOC - Hate Speech and Offensive Content Identification in Indo-European Languages

- **Hate speech:** Describing negative attributes or deficiencies to groups of individuals because they are members of a group (e.g. all poor people are stupid). Hateful comment toward groups because of race, political opinion, sexual orientation, gender, social status, health condition or similar.
HASOC - Hate Speech and Offensive Content Identification in Indo-European Languages

- **Hate speech**: Describing negative attributes or deficiencies to groups of individuals because they are members of a group (e.g. all poor people are stupid). Hateful comment toward groups because of race, political opinion, sexual orientation, gender, social status, health condition or similar.

- **Offensive**: Posts which are degrading, dehumanizing, insulting an individual, threatening with violent acts.

Kinga Gémes (kinga.gemes@tuwien.ac.at)
Identification and categorization of offensive language in German tweets
HASOC - Hate Speech and Offensive Content Identification in Indo-European Languages

- **Hate speech**: Describing negative attributes or deficiencies to groups of individuals because they are members of a group (e.g. all poor people are stupid). Hateful comment toward groups because of race, political opinion, sexual orientation, gender, social status, health condition or similar.

- **Offensive**: Posts which are degrading, dehumanizing, insulting an individual, threatening with violent acts.

- **Profanity**: Unacceptable language in the absence of insults and abuse. This typically concerns the usage of swearwords (Scheiße, Fuck etc.) and cursing (Zur Hölle! Verdammt! etc.) are categorized into this category.
HASOC data distribution

Figure: Train distribution
HASOC data distribution

Figure: Test distribution

- Other: 84%
- Profanity: 2.12%
- Offense: 9.06%
- Hate Speech: 4.82%
### Leader board on GermEval 2018

<table>
<thead>
<tr>
<th>Team</th>
<th>Other</th>
<th>Abuse</th>
<th>Insult</th>
<th>Profanity</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>uhhLT</td>
<td>84.85</td>
<td>53.25</td>
<td>39.46</td>
<td>29.63</td>
<td>52.71</td>
</tr>
<tr>
<td>TUWienKBS</td>
<td>85.8</td>
<td>52.4</td>
<td>43.71</td>
<td>20.34</td>
<td>51.42</td>
</tr>
<tr>
<td>uhhLT</td>
<td>84.26</td>
<td>51.96</td>
<td>40.18</td>
<td>15.58</td>
<td>48.44</td>
</tr>
<tr>
<td>uhhLT</td>
<td>82.88</td>
<td>46.1</td>
<td>21.12</td>
<td>3.92</td>
<td>43.04</td>
</tr>
<tr>
<td>InriaFBK</td>
<td>83.29</td>
<td>41.34</td>
<td>32.89</td>
<td>4.88</td>
<td>41.77</td>
</tr>
</tbody>
</table>
### Leader board on GermEval 2019

<table>
<thead>
<tr>
<th>Team</th>
<th>Other</th>
<th>Abuse</th>
<th>Insult</th>
<th>Profanity</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>upb</td>
<td>86.57</td>
<td>50.79</td>
<td>38.89</td>
<td>26.21</td>
<td>53.59</td>
</tr>
<tr>
<td>FoSIL</td>
<td>84.22</td>
<td>49.37</td>
<td>45.2</td>
<td>24</td>
<td>52.74</td>
</tr>
<tr>
<td>FoSIL</td>
<td>84.95</td>
<td>49.21</td>
<td>42.16</td>
<td>22.7</td>
<td>52.67</td>
</tr>
<tr>
<td>bertZH</td>
<td>86.66</td>
<td>50.07</td>
<td>44.37</td>
<td>28.27</td>
<td>52.64</td>
</tr>
<tr>
<td>upb</td>
<td>84.9</td>
<td>49.79</td>
<td>41.37</td>
<td>28.4</td>
<td>52.48</td>
</tr>
</tbody>
</table>

Kinga Gémes (kinga.gemes@tuwien.ac.at)  
Identification and categorization of offensive language in German tweets
# Leader board on HASOC 2019

<table>
<thead>
<tr>
<th>Team</th>
<th>Macro F1</th>
<th>Weighted F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSV-UdS</td>
<td>34.68</td>
<td>77.49</td>
</tr>
<tr>
<td>LSV-UdS</td>
<td>27.85</td>
<td>58.29</td>
</tr>
<tr>
<td>HateMonitors</td>
<td>27.69</td>
<td>75.37</td>
</tr>
<tr>
<td>3Idiots</td>
<td>27.58</td>
<td>77.79</td>
</tr>
<tr>
<td>Cs</td>
<td>27.4</td>
<td>75.7</td>
</tr>
</tbody>
</table>

Kinga Gémes (kinga.gemes@tuwien.ac.at)

Identification and categorization of offensive language in German tweets
Most popular and successive approaches

- uhhLT (Wiedemann et al., 2018): BiLSTM-CNN model
Most popular and successive approaches

- uhhLT (Wiedemann et al., 2018): BiLSTM-CNN model
- TUWienKBS (Montani, 2018): Word2vec and ensemble machine learning model
- upb (Paraschiv and Cercel, 2019): pre-trained BERT with last six layers replaced with an output layer after binary classification
- FoSIL (Schmid et al., 2019): FastText with SVM and radial kernel
- bertZH (Graf and Salini, 2019): pre-trained BERT classifier
- LSV-UdS (Ruiter, Rahman, and Klakow, 2019): 10 fold ensemble BERT classification after binary classification
- HateMonitors (Saha et al., 2019): SVM and Gradient boosted trees
- 3Idiots (Mishra, 2019): BERT classifier
Most popular and successive approaches

- uhhLT (Wiedemann et al., 2018): BiLSTM-CNN model
- TUWienKBS (Montani, 2018): Word2vec and ensemble machine learning model
- upb (Paraschiv and Cercel, 2019): pre-trained BERT with last six layers replaced with an output layer after binary classification
Most popular and successive approaches

- uhhLT (Wiedemann et al., 2018): BiLSTM-CNN model
- TUWienKBS (Montani, 2018): Word2vec and ensemble machine learning model
- upb (Paraschiv and Cercel, 2019): pre-trained BERT with last six layers replaced with an output layer after binary classification
- FoSIL (Schmid et al., 2019): FastText with SVM and radial kernel
Most popular and successive approaches

- uhhLT (Wiedemann et al., 2018): BiLSTM-CNN model
- TUWienKBS (Montani, 2018): Word2vec and ensemble machine learning model
- upb (Paraschiv and Cercel, 2019): pre-trained BERT with last six layers replaced with an output layer after binary classification
- FoSIL (Schmid et al., 2019): FastText with SVM and radial kernel
- bertZH (Graf and Salini, 2019): pre-trained BERT classifier
Most popular and successive approaches

- uhhLT (Wiedemann et al., 2018): BiLSTM-CNN model
- TUWienKBS (Montani, 2018): Word2vec and ensemble machine learning model
- upb (Paraschiv and Cercel, 2019): pre-trained BERT with last six layers replaced with an output layer after binary classification
- FoSIL (Schmid et al., 2019): FastText with SVM and radial kernel
- bertZH (Graf and Salini, 2019): pre-trained BERT classifier
- LSV-UdS (Ruiter, Rahman, and Klakow, 2019): 10 fold ensemble BERT classification after binary classification
Most popular and successive approaches

- uhhLT (Wiedemann et al., 2018): BiLSTM-CNN model
- TUWienKBS (Montani, 2018): Word2vec and ensemble machine learning model
- upb (Paraschiv and Cercel, 2019): pre-trained BERT with last six layers replaced with an output layer after binary classification
- FoSIL (Schmid et al., 2019): FastText with SVM and radial kernel
- bertZH (Graf and Salini, 2019): pre-trained BERT classifier
- LSV-UdS (Ruiter, Rahman, and Klakow, 2019): 10 fold ensemble BERT classification after binary classification
- HateMonitors (Saha et al., 2019): SVM and Gradient boosted trees
Most popular and successive approaches

- uhhLT (Wiedemann et al., 2018): BiLSTM-CNN model
- TUWienKBS (Montani, 2018): Word2vec and ensemble machine learning model
- upb (Paraschiv and Cercel, 2019): pre-trained BERT with last six layers replaced with an output layer after binary classification
- FoSIL (Schmid et al., 2019): FastText with SVM and radial kernel
- bertZH (Graf and Salini, 2019): pre-trained BERT classifier
- LSV-UdS (Ruiter, Rahman, and Klakow, 2019): 10 fold ensemble BERT classification after binary classification
- HateMonitors (Saha et al., 2019): SVM and Gradient boosted trees
- 3Idiots (Mishra, 2019): BERT classifier
Why is BERT so popular? (Vaswani et al., 2017)
### Why is BERT so popular? (Devlin et al., 2019)

Why is BERT so popular? (Devlin et al., 2019)

Kinga Gémes (kinga.gemes@tuwien.ac.at)

Identification and categorization of offensive language in German tweets

<table>
<thead>
<tr>
<th>Input</th>
<th>[CLS]</th>
<th>my</th>
<th>dog</th>
<th>is</th>
<th>cute</th>
<th>[SEP]</th>
<th>he</th>
<th>likes</th>
<th>play</th>
<th>#ing</th>
<th>[SEP]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Token Embeddings</td>
<td>$E_{[CLS]}$</td>
<td>$E_{my}$</td>
<td>$E_{dog}$</td>
<td>$E_{is}$</td>
<td>$E_{cute}$</td>
<td>$E_{[SEP]}$</td>
<td>$E_{he}$</td>
<td>$E_{likes}$</td>
<td>$E_{play}$</td>
<td>$E_{#ing}$</td>
<td>$E_{[SEP]}$</td>
</tr>
<tr>
<td>Segment Embeddings</td>
<td>$E_{A}$</td>
<td>$E_{A}$</td>
<td>$E_{A}$</td>
<td>$E_{A}$</td>
<td>$E_{A}$</td>
<td>$E_{A}$</td>
<td>$E_{B}$</td>
<td>$E_{B}$</td>
<td>$E_{B}$</td>
<td>$E_{B}$</td>
<td>$E_{B}$</td>
</tr>
<tr>
<td>Position Embeddings</td>
<td>$E_{0}$</td>
<td>$E_{1}$</td>
<td>$E_{2}$</td>
<td>$E_{3}$</td>
<td>$E_{4}$</td>
<td>$E_{5}$</td>
<td>$E_{6}$</td>
<td>$E_{7}$</td>
<td>$E_{8}$</td>
<td>$E_{9}$</td>
<td>$E_{10}$</td>
</tr>
</tbody>
</table>
Why is BERT so popular? (Devlin et al., 2019)

- WordPiece tokenizer and embeddings with 30,000 token vocabulary
Why is BERT so popular? (Devlin et al., 2019)

- WordPiece tokenizer and embeddings with 30,000 token vocabulary
- Architecture: multi-layer bidirectional Transformer encoder
Why is BERT so popular? (Devlin et al., 2019)

- WordPiece tokenizer and embeddings with 30,000 token vocabulary
- Architecture: multi-layer bidirectional Transformer encoder
  - Large model: 24 Transformer layer
- Pre-training on un-labeled data
  - Masked LM
  - Next Sentence Prediction
Why is BERT so popular? (Devlin et al., 2019)

- WordPiece tokenizer and embeddings with 30,000 token vocabulary
- Architecture: multi-layer bidirectional Transformer encoder
  - Large model: 24 Transformer layer
  - Base model: 12 Transformer layer

Identification and categorization of offensive language in German tweets

Kinga Gémes (kinga.gemes@tuwien.ac.at)
Why is BERT so popular? (Devlin et al., 2019)

- WordPiece tokenizer and embeddings with 30,000 token vocabulary
- Architecture: multi-layer bidirectional Transformer encoder
  - Large model: 24 Transformer layer
  - Base model: 12 Transformer layer
- Pre-training on un-labeled data

Kinga Gémes (kinga.gemes@tuwien.ac.at)
Identification and categorization of offensive language in German tweets
Why is BERT so popular? (Devlin et al., 2019)

- WordPiece tokenizer and embeddings with 30,000 token vocabulary
- Architecture: multi-layer bidirectional Transformer encoder
  - Large model: 24 Transformer layer
  - Base model: 12 Transformer layer
- Pre-training on un-labeled data
  - Masked LM
Why is BERT so popular? (Devlin et al., 2019)

- WordPiece tokenizer and embeddings with 30,000 token vocabulary
- Architecture: multi-layer bidirectional Transformer encoder
  - Large model: 24 Transformer layer
  - Base model: 12 Transformer layer
- Pre-training on un-labeled data
  - Masked LM
  - Next Sentence Prediction
Why is BERT so popular? (Devlin et al., 2019)
Why is BERT so popular?

- BERT (and its relatives) proves to be a strong model on sequence classification and sequence tagging problems

3 https://huggingface.co/bert-base-multilingual-cased
4 https://huggingface.co/bert-base-german-cased
Why is BERT so popular?

- BERT (and its relatives) proves to be a strong model on sequence classification and sequence tagging problems
- bert-base-multilingual-cased\(^3\)

\(^{3}\)https://huggingface.co/bert-base-multilingual-cased
\(^{4}\)https://huggingface.co/bert-base-german-cased
Why is BERT so popular?

- BERT (and its relatives) proves to be a strong model on sequence classification and sequence tagging problems
- `bert-base-multilingual-cased`³
- `bert-base-german-cased`⁴

³https://huggingface.co/bert-base-multilingual-cased
⁴https://huggingface.co/bert-base-german-cased

Kinga Gémes (kinga.gemes@tuwien.ac.at) Identification and categorization of offensive language in German tweets
Why is BERT so popular?

- BERT (and its relatives) proves to be a strong model on sequence classification and sequence tagging problems
- `bert-base-multilingual-cased`\(^3\)
- `bert-base-german-cased`\(^4\)
  - Published on Jun 14th, 2019

\(^3\)https://huggingface.co/bert-base-multilingual-cased
\(^4\)https://huggingface.co/bert-base-german-cased
Why is BERT so popular?

- BERT (and its relatives) proves to be a strong model on sequence classification and sequence tagging problems
- `bert-base-multilingual-cased`³
- `bert-base-german-cased`⁴
  - Published on Jun 14th, 2019
  - Trained on German Wikipedia, OpenLegalData and News data

³https://huggingface.co/bert-base-multilingual-cased
⁴https://huggingface.co/bert-base-german-cased
Why is BERT so popular? - Syntactic knowledge (Rogers, Kovaleva, and Rumshisky, 2020)

- BERT representations are hierarchical rather than linear, like syntactic trees
Why is BERT so popular? - Syntactic knowledge (Rogers, Kovaleva, and Rumshisky, 2020)

- BERT representations are hierarchical rather than linear, like syntactic trees
- BERT embeddings encode information about pos, syntactic chunks and roles
Why is BERT so popular? - Syntactic knowledge
(Rogers, Kovaleva, and Rumshisky, 2020)

- BERT representations are hierarchical rather than linear, like syntactic trees
- BERT embeddings encode information about pos, syntactic chunks and roles
- BERT does not store this information in its self-attention weights, but it can be recovered from the token representations
Why is BERT so popular? - Syntactic knowledge (Rogers, Kovaleva, and Rumshisky, 2020)

Kinga Gémes (kinga.gemes@tuwien.ac.at)  Identification and categorization of offensive language in German tweets
Why is BERT so popular? - Syntactic knowledge (Rogers, Kovaleva, and Rumshisky, 2020)
Why is BERT so popular? - Semantic knowledge (Rogers, Kovaleva, and Rumshisky, 2020)

- BERT encodes information about semantic roles, entity types, relations
Why is BERT so popular? - Semantic knowledge
(Rogers, Kovaleva, and Rumshisky, 2020)

- BERT encodes information about semantic roles, entity types, relations
- BERT struggles with numbers; it does not form a good representation of floating point numbers and fails to generalize
Why is BERT so popular? - Semantic knowledge (Rogers, Kovaleva, and Rumshisky, 2020)

- BERT encodes information about semantic roles, entity types, relations
- BERT struggles with numbers; it does not form a good representation of floating point numbers and fails to generalize
- BERT does not form a generic idea of named-entities
Why is BERT so popular? - World knowledge (Rogers, Kovaleva, and Rumshisky, 2020)

- BERT can be competitive with methods relying on knowledge bases for some relation types.
Why is BERT so popular? - World knowledge
(Rogers, Kovaleva, and Rumshisky, 2020)

- BERT can be competitive with methods relying on knowledge bases for some relation types
- BERT struggles with pragmatic inference, role-based event knowledge, and abstract attributes of objects
Why is BERT so popular? - World knowledge
(Rogers, Kovaleva, and Rumshisky, 2020)

- BERT can be competitive with methods relying on knowledge bases for some relation types
- BERT struggles with pragmatic inference, role-based event knowledge, and abstract attributes of objects
- BERT cannot reason based on world-knowledge

“Dante was born in [MASK].”

Neural LM Memory Access  →  Florence
Twitter data processing for BERT

- @username → [USER]

numbers → [NUM], urls → [URL], dates → [DATE]

emoticons should be replaced by their textual representations because of the WordPiece tokenizer

#ImportantHashtag cut it up by the camel case and remove the #

Kinga Gémes (kinga.gemes@tuwien.ac.at) Identification and categorization of offensive language in German tweets
Twitter data processing for BERT

- @username → [USER]
- numbers → [NUM], urls → [URL], dates → [DATE]
- Emoticons should be replaced by their textual representations because of the WordPiece tokenizer.

#ImportantHashtag → cut it up by the camel case and remove the #
Twitter data processing for BERT

- @username → [USER]
- numbers → [NUM], urls → [URL], dates → [DATE]
- emoticons should be replaced by their textual representations because of the WordPiece tokenizer
Twitter data processing for BERT

- @username → [USER]
- numbers → [NUM], urls → [URL], dates → [DATE]
- emoticons should be replaced by their textual representations because of the WordPiece tokenizer
- #ImportantHashtag → cut it up by the camel case and remove the #


Mishra, Shubhanshu (2019). “3Idiots at HASOC 2019: Fine-tuning Transformer Neural Networks for Hate Speech Identification in Indo-European Languages”. In: FIRE.


