Model Card - English to IsiZulu Machine Translation using a RNN

Model Details

- Recurrent Neural Network
  - Feed forward RNN vs Bidirectional RNN
  - Standard Tokenization vs Byte-Pair Encoding
  - With word embedding vs without word embedding
- Implemented using Keras

Intended Use

- Intended to test the quality of machine translation from English to isiZulu.
- Intended to be used to get a general understanding of the translation.
- Not intended to be used in a real setting where a correct translation is essential.

Factors

- The English sentences are sampled from the News Crawl dataset and therefore are skewed to young, white and male perspectives.

Metrics

- Bilingual Evalutation Understudy Score
- The BLEU score is a metric for comparing the quality of a generated sentence (predicted machine translation) to the reference sentence (ground truth translation).
- A score of 100 means the sentences match perfectly.
- A score of 0 means the sentences don't match at all.

Evaluation Data

- Sourced from Umsuka English - isiZulu Parallel Corpus 998 English sentences with 2 corresponding isiZulu translations each

Training Data

- Sourced from Umsuka English - isiZulu Parallel Corpus
- 4739 English sentences and their corresponding isiZulu translations.

Ethical Considerations

- Data should be gathered from originally Zulu sources, rather than English text which has been translated into isiZulu by human translators.

Caveats and Recommendations

- In future it would be beneficial to retrieve source (English) sentences from a wider range of sources, and ones which are unbiased.

Quantitative Analyses

Table 1: BLEU Scores

Model	**Training Score**	**Testing Score**
1	14.85	19.05
2	2.32	1.62
3	72.54	4.9
4	71.8	2.02
5	78.23	8.35
6	67.4	2.68
7	3.79	35.97
8	1.99	14.79