SENTIMENT ANALYZER PROJECT

Sebastian Escobar
Gabriel Restrepo
Juan Camilo Gonzalez

CONTEXT

Supervised Neural Networks are useful for study and predict some behaviors, in this case we have reviews taken from three web pages: product reviews from Amazon, movies reviews from IMDB and Restaurants reviews from Yelp, all classified in positive or negative, so we want to train a Supervised Neural Network to predict if the review received is positive or negative

OBJECTIVES

SPECIFIC

Analyze correctly the data given for plain preprocessing and the processing

Clean the data if the analysis determines it necessary

Implements a RNN and LSTM to train it and predict if the review is positive or not

GENERAL

Implements Neural Networks to detect patterns and predict output values given a set of entries

Take advantage of Natural Language Processing benefits to analyze sentiments present in some text

Implement Neural Network model tuning methods to get better results

DATASET

We received 3 **.txt** files, where each line has a review written by a user and, at the end, a tab **(\t)** and a number 1 or 0.

We put headers to each file **review** and **positive**.

Before making the datasets we clean the reviews by tokenizing the data and deleting the **stopwords** by using **NLTK**.

ARCHITECTURE AND TRAINIG PROCESS

RESULTS

(WITHOUT TUNING)

RNN

```
Accuracy: 0.67
Precision: 0.6794871794871795
Recall: 0.6838709677419355
F1 Score: 0.6816720257234727
Kappa Score: 0.3391188251001336
```

LSTM

```
Accuracy: 0.7366666666666667
Precision: 0.7516556291390728
Recall: 0.7322580645161291
F1 Score: 0.7418306535947712
Kappa Score: 0.4732162702822849
```

RESULTS

(AFTER TUNING)

RNN

```
Accuracy: 0.7066666666666667
Precision: 0.768
Recall: 0.6193548387096774
F1 Score: 0.6857142857142857
Kappa Score: 0.4165745856353592
```

LSTM

```
Accuracy: 0.7916666666666666
Precision: 0.7846153846153846
Recall: 0.8225806451612904
F1 Score: 0.8031496062992126
Kappa Score: 0.5821727019498608
```

CONCLUSION

CONCLUSION

Neural Networks + Natural Language Processing

DATA PROCESSING

The success of the Sentiment Analyzer was the combination of data processing, model selection, hyperparameter tuning and the evaluation of multiple indicators and metrics

THANKS