

## [NN'23] Arabic Sentiment Analysis (NLP)

|  |  |  |
| --- | --- | --- |
| **Team members** | **IDs** | **Department** |
| محمد ايمن رشدي عيد العظيم | 20201700674 | CS |
| عبدالرحمن محيي محمد رمضان | 20201700487 | CS |
| محمد عصام علي مصطفي صبري | 20201700711 | CS |
| فؤاد احمد سامي فؤاد الدهان | 20201700411 | CS |
| زياد سيد عوض محمد | 20201700307 | CS |

**Contents**

**1** **Introduction**

1.1 State Topic

1.2 Goals

**2 Data Definition**

2.1 Dataset

**3 Data Cleansing**

1. Data Preprocessing.
   1. Data Cleansing.
      1. Remove punctuation
      2. Remove Duplicated letters
      3. Remove non Arabic letters and emojis
      4. Join the words into a single string
   2. Tokenization
   3. Stemming

**4-Neural Network Models**

* 1. LSTM.
  2. Transformer.

**5 Save Models**

**6 Conclusion**

1

Introduction

* 1. State Topic

Arabic sentiment analysis using various advanced Deep Neural Network (DNN) models, including RNNs, LSTMs, and transformers. With a provided dataset, we're diving into data preparation, fine-tuning our models for the nuances of Arabic text.

In a competitive twist, we're taking our models to a Kaggle competition, where our project grade hinges on their performance. This isn't just about building models; it's a friendly showdown with other teams to see who can nail sentiment analysis the best.

* 1. Goals
* Apply the appropriate data preparation steps. (Preprocessing, Tokenize, stopwords, Cleaning…)
* Build multiple appropriate models and understand each part (LSTM, transformer, etc.)

2

Data Definition

2.1 Dataset info



4

Data Cleansing

A-Data Cleansing

4.1 Remove punctuation

This step involves eliminating any punctuation marks or symbols from the text, such as periods, commas, question marks, and exclamation marks.

* 1. Remove Duplicated letters

In this step, any duplicated letters within a word are removed. For example, if a word contains consecutive identical letters, only one of them will be kept.

4.3 Remove non Arabic letters and emojis

This step involves eliminating any characters that are not part of the Arabic alphabet and removing any emojis or emoticons present in the text.

4.4 Join the words into single string

After the previous steps, the remaining words are concatenated together to form a single string, where the words are separated by spaces.

B-Tokenization

 Stemming is the process of reducing words to their base or root form, known as the stem. It involves removing any suffixes or prefixes from the words to obtain the core meaning.

C-Stemming

 Stemming is the process of reducing words to their base or root form, known as the stem. It involves removing any suffixes or prefixes from the words to obtain the core meaning

10

Neural Network models

* 1. LSTM
     1. Prediction Code

A computer screen shot of a code

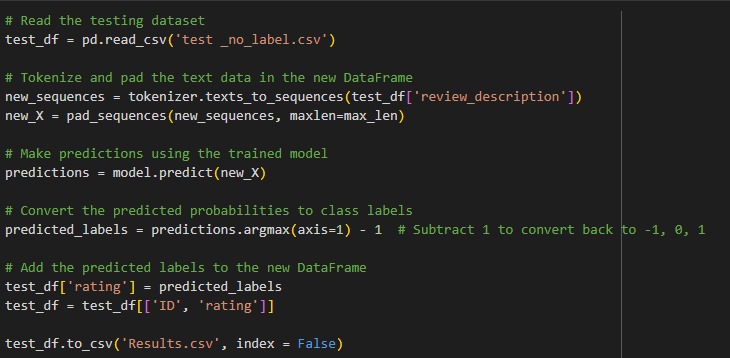
Description automatically generated

* + 1. Accuracy

A table of numbers and text

Description automatically generated with medium confidence

* 1. Transformer
     1. Transformer testing



* + 1. Accuracy

A screenshot of a computer

Description automatically generated

13

Save models

to save all models and save that.

A computer screen shot of a computer code

Description automatically generated

14

Conclusion

Our journey concludes with a successful performance in the Kaggle competition, showcasing the strengths of our DNN models in Arabic sentiment analysis. Our detailed report serves as a playbook, sharing our strategies, challenges, and victories with the NLP community. As we reflect on the experience, we recognize our contribution to the dynamic field of NLP and anticipate its continued evolution.