**Arabic Sentiment Analysis**

**Team\_ID: 11**

**Team Members**

|  |  |  |
| --- | --- | --- |
| Name | ID | Department |
| أحمد حسن حسن خطاب | *20201700033* | Computer Science(CS). |
| انس سعيد صبري | *20201700159* | Computer Science(CS). |
| عبدالرحمن هاني خيري الدين | *20201700488* | Computer Science(CS). |
| أحمد هاني حامد محمد | *20191700875* | Computer Science(CS). |
| عمر رضا عبدالدايم عبدالعزيز | *20201701267* | Computer Science(CS). |

**Introduction**

This report will explain what algorithms for the project for preparing the text data an train it into different neural networks models.

**-------------------------------------------------------------------------------------------------------------**

**Data Preparation**

1. **Text Preprocessing**

* **Word Tokenization.**
* **Remove stop words, punctuations and Arabic diacritics.**
* **Lemmatizing the stemmed words using qalsadi.**

1. **Feature extraction**

* **Extracting numerical features to be understood for the model.**
* **Used the text\_to\_sequence for tokenized words then using the pad\_sequence for each generated sequence.**
* **Fits only on the training data and transform on both training and testing data.**

1. **Features Saving**

* **Saving the extracted features to be read directly next time.**

**Neural Network Models**

1. **Long Short-Term Memory [ LSTM ].**

* **First , An Embedding Layer is added which takes 3 parameters:**
  + **The input\_dim parameter is set to the maximum index of the words in the tokenizer increased by 1.**
  + **The output\_dim parameter is set to 200, representing the dimensionality of the embedding space.**
  + **The input\_length parameter is set based on the shape of the input features**
* **Second, A LSTM Layer is added to the model with 100 units**
* **Third****, A Dense layer is added to the model with 3 units and a softmax activation function which is used for multi-class classification problems as it converts raw scores into probabilities.**
* **Finally, The model is compiled using the Adam optimizer and Categorical cross entropy is chosen as the loss function, which is suitable for multi-class classification problems and accuracy as the metric for evaluation**
* **The model is trained within 1 epoch and achieved the highest testing accuracy (0.84337)**

1. **Convolutional Neural Network [ 1D CNN ]**

* **First , An Embedding Layer is added which takes 3 parameters:**
  + **The input\_dim parameter is set to the maximum index of the words in the tokenizer increased by 1.**
  + **The output\_dim parameter is set to 200, representing the dimensionality of the embedding space.**
  + **The input\_length parameter is set based on the shape of the input features**
* **Second , A 1D Convolutional Layer is added which has 100 filters each with kernel size of 5 and activation function used is RELU which introduces non-linearity to the model**
* **Third , A Global Max Pooling layer is added to the model to reduce the spatial dimensions of input data while retaining the most important information and this helps in extracting the most relevant features from the output of convolutional layer**
* **Fourth , A Dense layer is added to the model with 3 units and a softmax activation function which is used for multi-class classification problems as it converts raw scores into probabilities**
* **Finally , The model is compiled using the Adam optimizer, categorical cross entropy loss function, and accuracy as the metric for evaluation.**
* **The model is trained within 1 epoch and achieved testing accuracy (0.82831)**

1. **Transformer DNN**

* **--**