ASSIGNMENT – NLP

Develop a Deep Learning-based approach to classify the Portuguese Tweets for Sentiment Analysis dataset available at https://www.kaggle.com/augustop/portuguese-tweets-for-sentiment-analysis

- The instructions on how to download the dataset and initial preparations are contained in the following notebook: https://colab.research.google.com/drive/1PF7X3rSnFmnC-5-c43ZKolv3jWcERSC3?usp=sharing
- In this assignment, the files Train3Classes.csv and Test3classes.csv will be used. It contains portuguese tweets divided in positive, negative and neutral classes for sentiment polarity classification.
- The training set contains 100k tweets, and the testing set contains 5k tweets. You can create a validation set from the training set.
- Try to use different architectures, such as 1DConvolutional Neural Networks or Recurrent Neural Networks (LSTM/GRU, Bidirectional-LSTM/GRU);
- Modify architecture, network parameters, and data to improve the classification performance.
- Insert regularization (Dropout, L2, etc); add/remove layers; modify optimizer and other parameters in general.
- Try to initialize the embedding layer with a pre-trained model (Glove, Fasttext, Word2Vec, etc..).
- Plot the loss and performance metric curves of the training and test set.

Write a report containing the learning curves, metrics, and confusion matrix of your best final model on the training and testing set. Also include the architecture and parameters used in the final model, as well as an analysis of the results obtained and the impacts related to the variation in network architecture/parameters and data. Also, briefly describe the experiments performed that did not achieve the best result.