

WEEK 7

Using the following data set:

<https://www.kaggle.com/code/cemalefetezcan/imdb-review-sentiment-classification>

divide the data into train/validation data sets, build 2 models to perform movie review sentiment analysis

Model 1:

1. Perform required text pre-processing – lowering text, removing URLs, punctuation , stop words and correct spelling .
2. Perform tokenization and lemmatization on cleaned data .
3. Visualize the most frequent words and bigrams
4. Visualize the practical words that represent positive and negative sentiment in the dataset.
5. Create an embedding layer and build a 15 layer LSTM and a 20 layer BidRNN for predicting the sentiment.
6. Build your own test dataset with 10 movie reviews and tabulate accuracy.