BRAC UNIVERSITY

Inspiring Excellence

Online Sexism Detection

Maruf Morshed

20101299

Md. Danial Islam

23241067

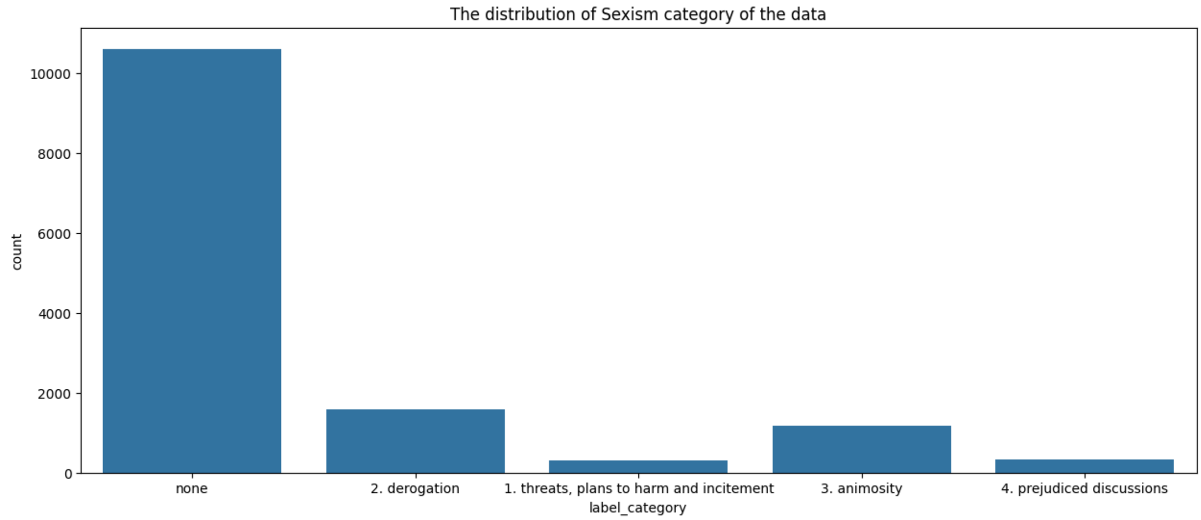
Introduction:

In this era of digital communication online sexism has become a very common thing among us. Nowadays, the importance of sorting out these comments has become an important task for us. The rise of the internet has also led to increased online sexism, which poses significant negative impacts on individuals and society. In this project, we aim to identify online sexism using simple LSTM, bi-LSTM and GRU. We will train our LSTM model on a dataset containing both sexist and non-sexist comments, using it thereafter to detect sexist language in new comments. Our task was divided into two subparts. They are Binary Sexism and Categorical Sexism. The first task involves classifying comments into two categories: sexist or not sexist. The second task further breaks down the sexist comments into four categories: Threats, Derogation, Animosity, and Prejudiced Discussion. We will evaluate whether our LSTM model can effectively identify sexist language with high accuracy.

Datasets description:

In the datasets, there are collections of sexist and non-sexist comments. These comments are compiled from Gab and Reddit, which were then processed for analysis. In task A, we utilized the 'label_sexist' column, which comprises 14,000 data points, including 3,398 labeled as sexist and 10,602 as non-sexist. However, there is an imbalance where there are 7,204 more non-sexist than sexist entries.

In task B, main focus was given to classify the types of sexism within the 3,398 sexist entries: 310 are categorized as threats, 333 as prejudiced discussions, 1,165 as animosity, and 1,590 as derogation. Here, the categories of animosity and derogation significantly outnumber threats and prejudiced discussions, indicating another layer of imbalance in the dataset.



Model Description:

Simple LSTM: LSTM is a type of RNN which is designed to solve the problem of vanishing gradients that can occur in traditional RNNs. This model is chosen because of its specialized architecture, which includes memory cells and multiple gates—input, output, and forget gates—that regulate the flow of information. These attributes of LSTM helps to retain important information and get rid of irrelevant information which is important for our task and this dataset. The LSTM layer has lstm_out units and a dropout rate of 0.2. The model is compiled using binary cross-entropy, Adam, and accuracy. Then it finds the accuracy and loss for training and test data.

Bi-LSTM: This is an enhancement of traditional LSTM which improves its context understanding by processing data in both directions(forward and backward). This operation style helps to retain some losses in one way data processing. This is specially used in NLP fields like sentiment analysis. Using this in our datasets is a good option for its characteristics.

GRU: Gated Recurrent Unit is a type of RNN which simplifies the standard LSTM architecture by combining the forget and input gate into a single ‘update gate’. The use of GRU for this task is particularly suitable for it’s characteristics.

Results:

TaskA: For simple LSTM structure, we have set the epochs value 20 and after 20 epochs, we are getting Training Accuracy: 0.9762, and Training Loss: 0.0820. And for the test result, Testing Accuracy: 0.8150, and Testing Loss: 0.6440. Next, for the bi-LSTM model, the epoch size is 10 and the Training Accuracy is 0.9812, Training Loss is 0.0644 and the Testing Accuracy is 0.7979, Testing Loss is 0.6002. Again for the GRU model, the epoch size is 10 and the Training Accuracy is 0.9302, Training Loss is 0.1770 and the Testing Accuracy is 0.8064, Testing Loss is 0.4900.

TaskB: For simple LSTM structure, we have set the epochs value 20 and batch size of 32, and after 20 epochs, we are getting Training Accuracy: 0.9439, and Training Loss: 0.1743. And for the test result, Testing Accuracy: 0.7529, and Testing Loss: 1.0033. Next, for the hypertuned LSTM model, the epoch size is 20, batch size is 64, and the Training Accuracy is 0.85, Training Loss is 0.42 and the Testing

Accuracy is 0.78, Testing Loss is 0.71. Again for the GRU model, the epoch size is 20 and the Training Accuracy is 0.95, Training Loss is 0.17 and the Testing Accuracy is 0.75, Testing Loss is 0.97.

References:

1) CodaLab - Competition. (n.d.).

https://codalab.lisn.upsaclay.fr/competitions/7124