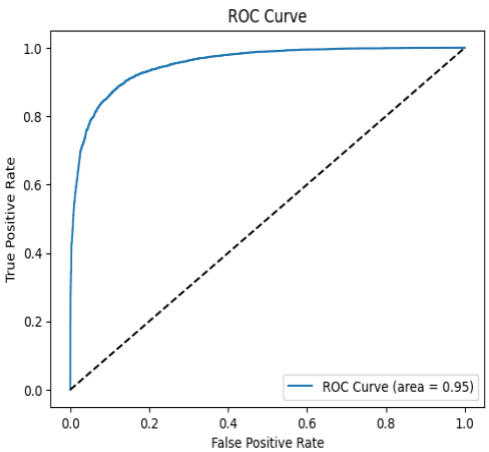
Supplementary material

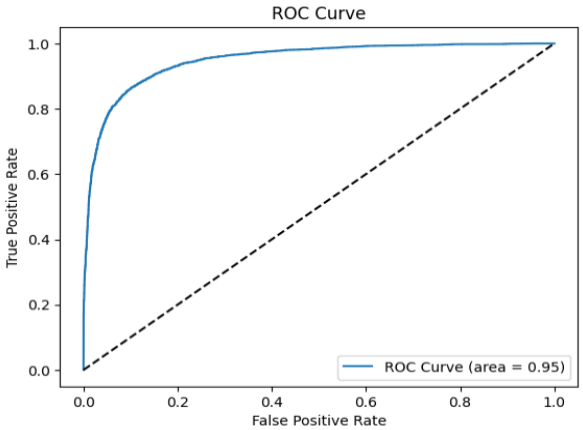
The supplementary material contains additional experimental results that could not be included in the main paper due to space limitations. Specifically, we provide additional performance metrics such as precision, recall, and F1-score for each model on the test set. We also include additional visualizations, including confusion matrices and ROC curves, to provide a more detailed model performance analysis.

Data cleaning and pre-processing:
Depending on the review, the IMDb movie review dataset contains 50K MOVIE reviews, labelled as positive and negative. The first thing we did is converted the positive and negative labels as 0 and 1 so they could be integers. In our dataset, there are 101895 unique words. We did a code so we could print the most and least common words:
Top 10 most common words: ['the', 'and', 'of', 'to', 'is', 'br', 'it', 'in', 'this', 'that']
Top 10 least common words: ['ünel', 'ünfaithful', 'üvegtigris', 'üzümcü', 'ýs', 'þorleifsson', 'þór', 'żmijewski', 'ברמון' ,'יגאל']
Our dataset contained some other language words, even though this dataset belongs only to UK movie reviews,but we removed all of these.

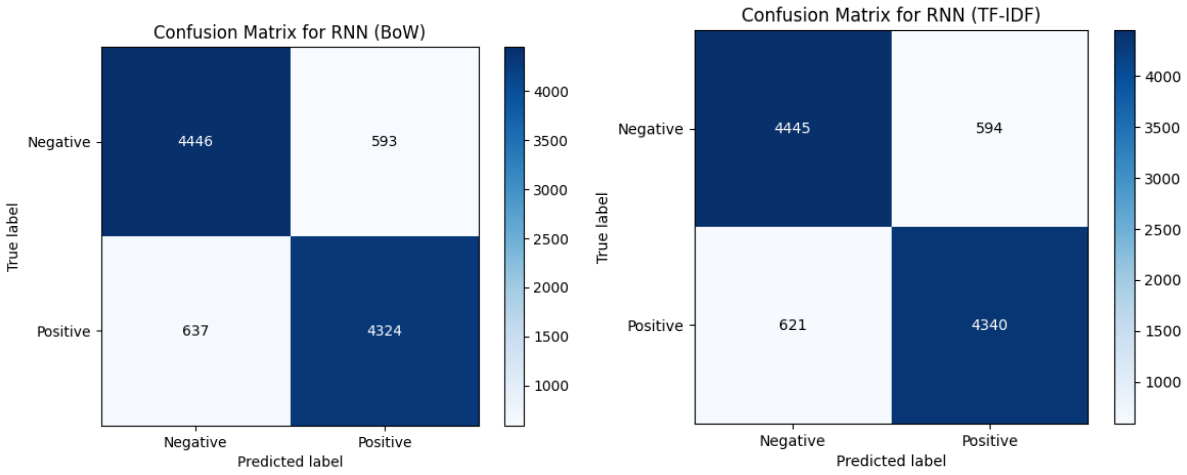
Logistic regression model:

We applied the logistic regression model with TF-IDF and BoW as a baseline, with an accuracy of 89%. After we studied the confusion matrix and ROC curve, we concluded that the TF-IDF was the best choice for the LR. Below we are going to include also the ROC curve for each of them:
The accuracy under the ROC curve is the same for both models, so based on the ROC curve we cannot determine which feature extraction is the best for our baseline model. But we did a classification report to get a better picture, showing that LG with TF-IDF is slightly better than with the method of BoW feature extraction.





RNN MODEL WITH TWO DIFFERENT FEATURE EXTRACTION (BoW +LSTM)



Based on the confusion matrix, the confusion matrix of the RNN +BoW performed better than the other model.