Although the effect of the doc2vec model is not as good as the model using CountVector. We still see the potential of using this model to analysis the sentiment of the news titles.

We tried three different Doc2Vec models, the model trained on the training set we have, a pre-trained model based on Associated Press News data, a pre-trained model based on Wikipedia Text data. The last two datasets are much larger than our title dataset.

The result shows the pretrained models perform better than the model trained on our title dataset. We can infer that pretrained models bring in other information contained in the text of news and Wikipedia. Therefore, the vector space has better ability to represent the information contained in the titles.

Chart, line chart

Description automatically generated