NLP-Project

**Story –**

Two techniques were applied on the dataset.

1. Bag of words (BoW)

The BoW is used to model the data for further prediction and classify comments as positive or negative.

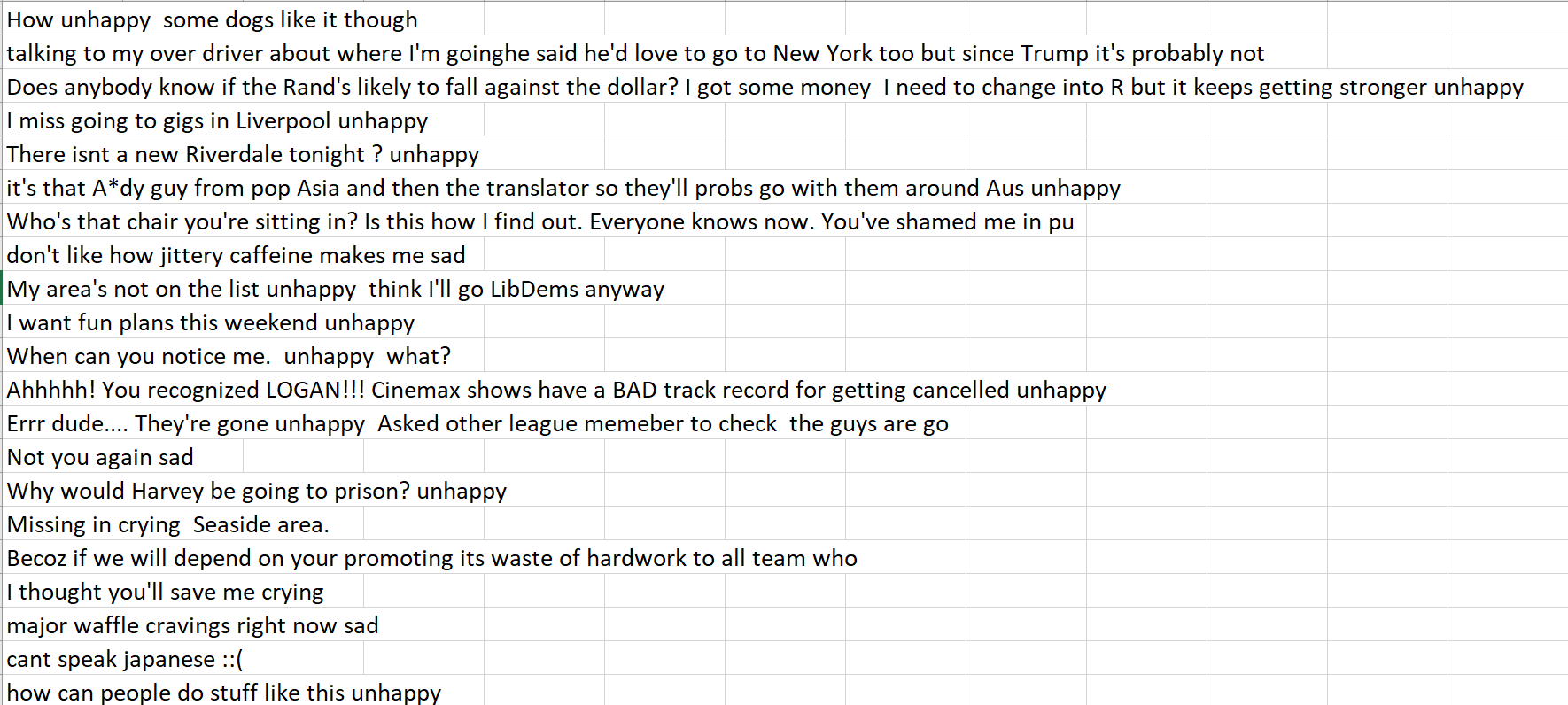
1. Latent Dirichlet Allocation (LDA)

This technique is used for topic modeling. In this dataset 5 different topics were extracted.

**Data Description:**

This dataset consists of two separate csv files for Positive comments and Negative comments. The first file includes 1117 positive comments, and the second file includes 1184 negative comments.

The following is the snippet of the transit dataset.



Data is available at:

<https://www.kaggle.com/shashank1558/preprocessed-twitter-tweets>

**Final result –**

The output of BoW is a prediction model that could predict that a comment is positive or negative. And also, LDA model, could show the five important topics in the comments.

**GitHub Repository –**

This is the link to our GitHub repository:

<https://github.com/spdzh/NLP-PROJECT/tree/main>