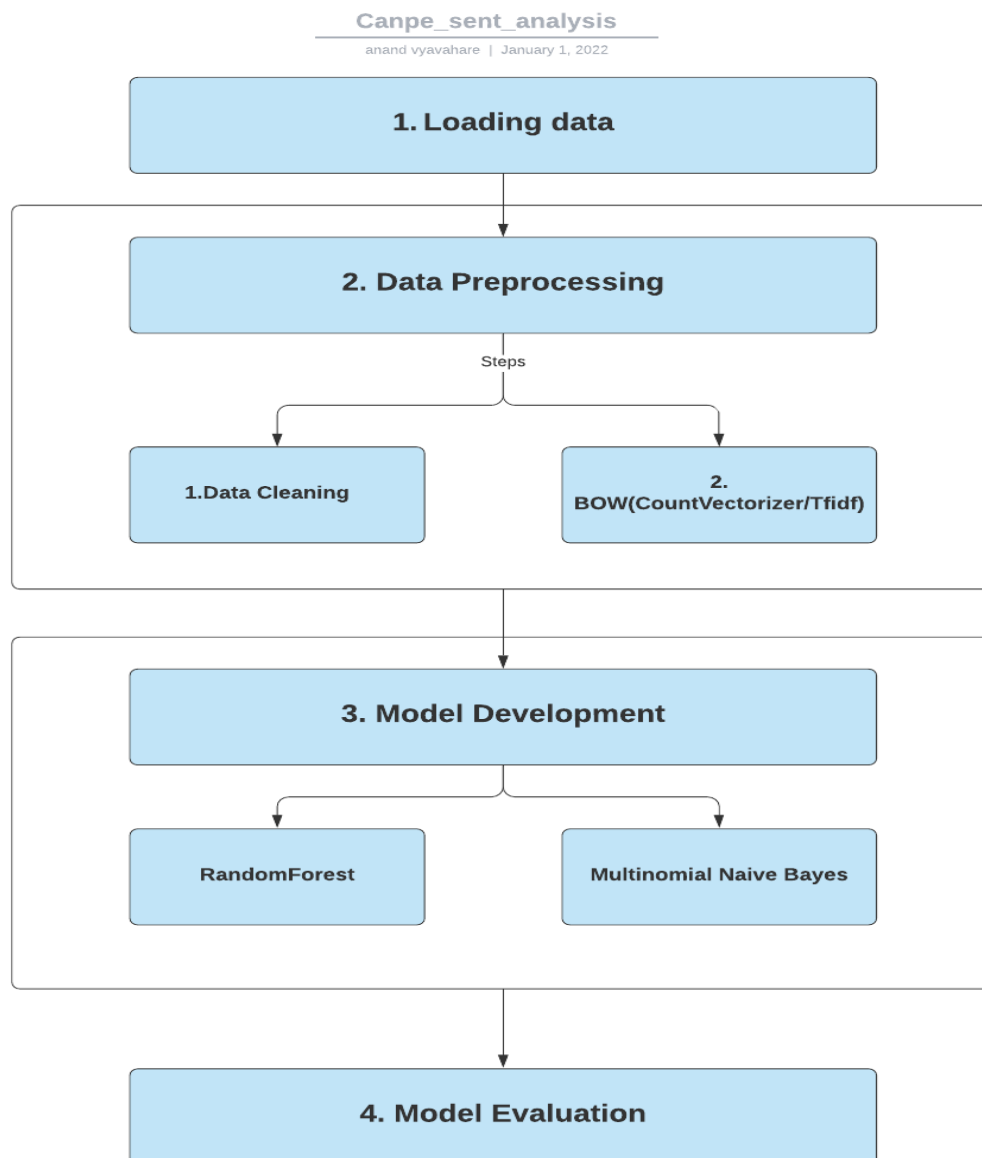# Sentiment Analysis

As the name suggests Sentiment Analysis is a technique used to determine the sentiments in a textual data. It is one of the Natural Language Processing (NLP) techniques.
Below is the flow chart of the methodology we have opted for this problem.

In our problem, we are working with F.R.I.E.N.D.S TV series dialogues and utterances to find the sentiments behind them.

Canpe_sent_analysis
anand vyavahare | January 1, 2022

1. Loading data

2. Data Preprocessing

Steps

1.Data Cleaning

2. BOW(CountVectorizer/Tfidf)

3. Model Development

RandomForest

Multinomial Naive Bayes

4. Model Evaluation

Results

We tried out different methods. Following are the methods and their respective evaluation metrics.

We are using the below evaluation metrics as precision and recall, and even Auc ROC are generally used when we have imbalanced data. And we are also using accuracy.

Sr. No	Methods\Evaluation Metrics	Accuracy	Precision	Recall
1	Countvectorizer + RF(GridSCV)	0.47	0.38	0.47
2	TFIDF(ngram 1,1) + RF(GridSCV)	0.52	0.62	0.52
3	TFIDF(ngram 1,2) + RF(GridSCV)	0.5	0.66	0.5
4	TFIDF(ngram 2,2) + RF(GridSCV)	0.49	0.69	0.49
5	TFIDF(ngram 1,3) + RF(GridSCV)	0.51	0.67	0.51
6	TFIDF(ngram 1,3) + MultinomialNB	0.52	0.6	0.52

But what are these metrics?
Recall: Tells what percentage of *positives* were correctly identified.
Precision: Proportion of *positive* results that were correctly classified.
Accuracy: Fraction of predictions our model got right

So what do the scores mean?
Ideally, all these scores should be high - close to one. But in reality, it is not only hard to achieve but there is also a trade off between precision and recall. The best model should come up with high scores, should be easy to implement (in terms of time and complexity).

Conclusion and Future Scope

As we can see, the results we have obtained aren't the best. But these models we tried out serve best as the baseline model. We can come up with even more data preprocessing techniques and try out some Neural Nets to train our model and improve the scores.