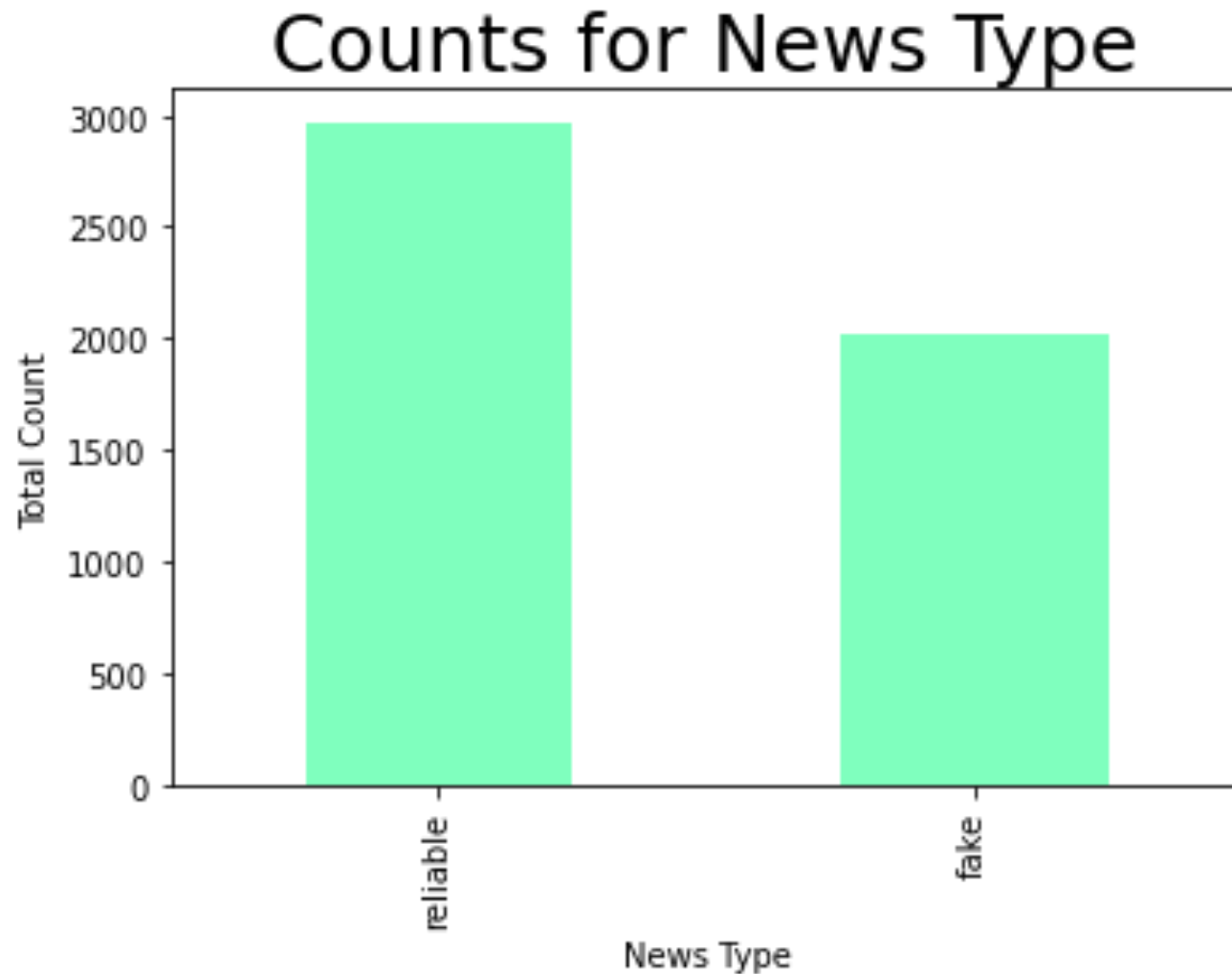# Fake News Prediction

Predicting if the given news article is fake or realistic using text classification

Class Labels

- About 5000 rows
- Fake/ Realistic class labels
- Even distribution between the two class labels

Class label distribution

Text Classification

- Count vectorizer

- Tfidf vectorizer

- Preprocessed text using stop words

Algorithms

- Applied naïve bayes and svm
- Issue : The features were in form of numbers
- To fix that did further preprocessing

Preprocessing text

- Removed punctuation
- Porter stemming
- Applied algorithms
- Issue : The number of features exceeded the number of rows by a large amount
- Applied different algorithms

Naïve Bayes

Classification Report:

	precision	recall	f1-score	support
0	0.73	0.90	0.81	983
1	0.77	0.50	0.61	663
accuracy			0.74	1646
macro avg	0.75	0.70	0.71	1646
Wght. avg	0.75	0.74	0.73	1646

Support vector machine

Classification Report:

	precision	recall	f1-score	support
0	0.75	0.80	0.77	983
1	0.67	0.60	0.63	663
accuracy			0.72	1646
macro avg	0.71	0.70	0.70	1646
Wght. avg	0.72	0.72	0.72	1646

XGBoost

Classification Report:

	precision	recall	f1-score	support
0	0.73	0.88	0.80	983
1	0.75	0.52	0.61	663
accuracy			0.74	1646
macro avg	0.74	0.70	0.71	1646
Wght. avg	0.74	0.74	0.72	1646

Logistic Regression

Classification Report:

	precision	recall	f1-score	support
0	0.73	0.88	0.80	983
1	0.75	0.53	0.62	663
accuracy			0.74	1646
macro avg	0.74	0.70	0.71	1646
Wght. avg	0.74	0.74	0.73	1646

Decision Trees

Classification Report:

	precision	recall	f1-score	support
0	0.75	0.75	0.75	983
1	0.63	0.63	0.63	663
accuracy			0.70	1646
macro avg	0.69	0.69	0.69	1646
weighted avg	0.70	0.70	0.70	1646