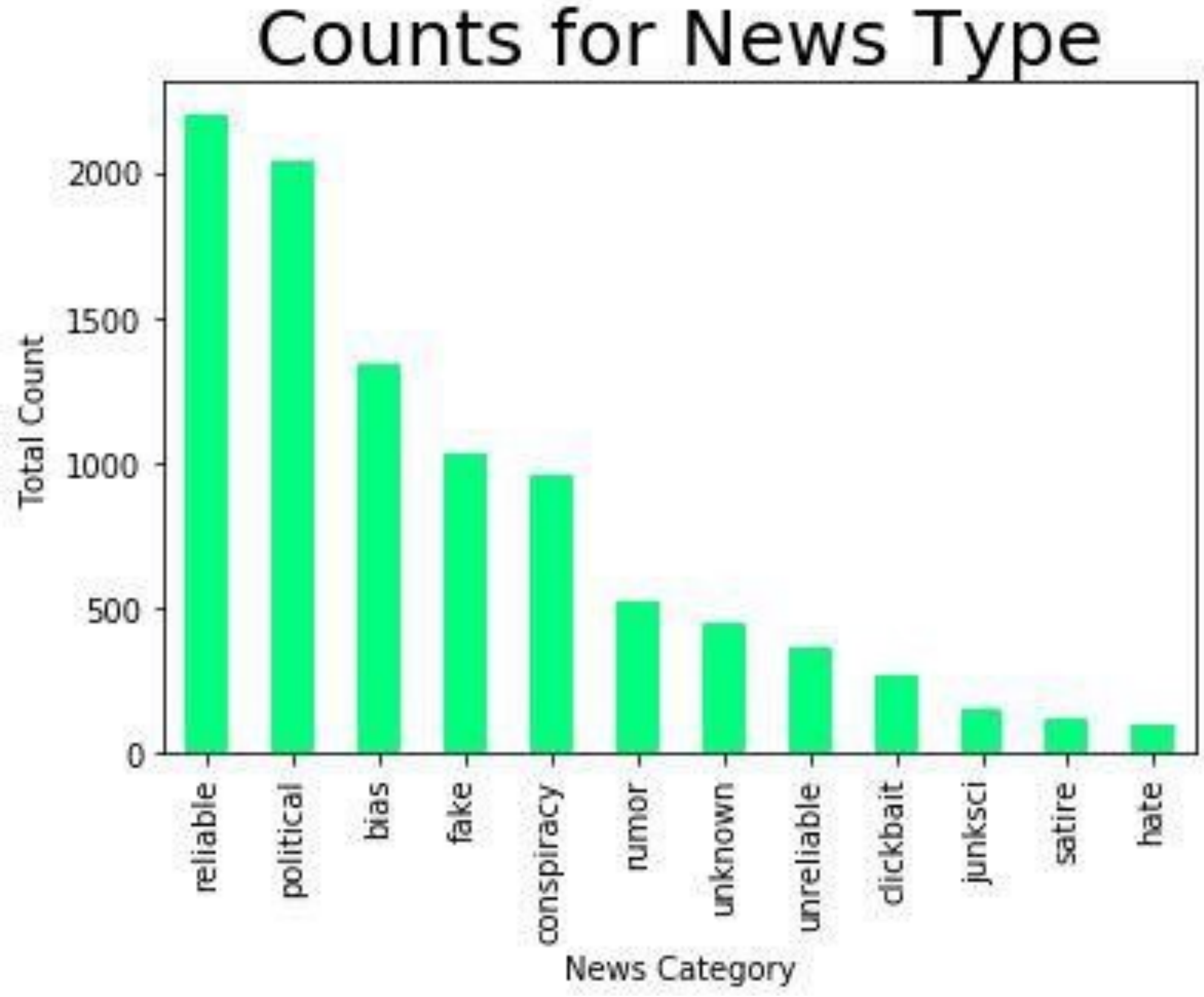
Fake News Project

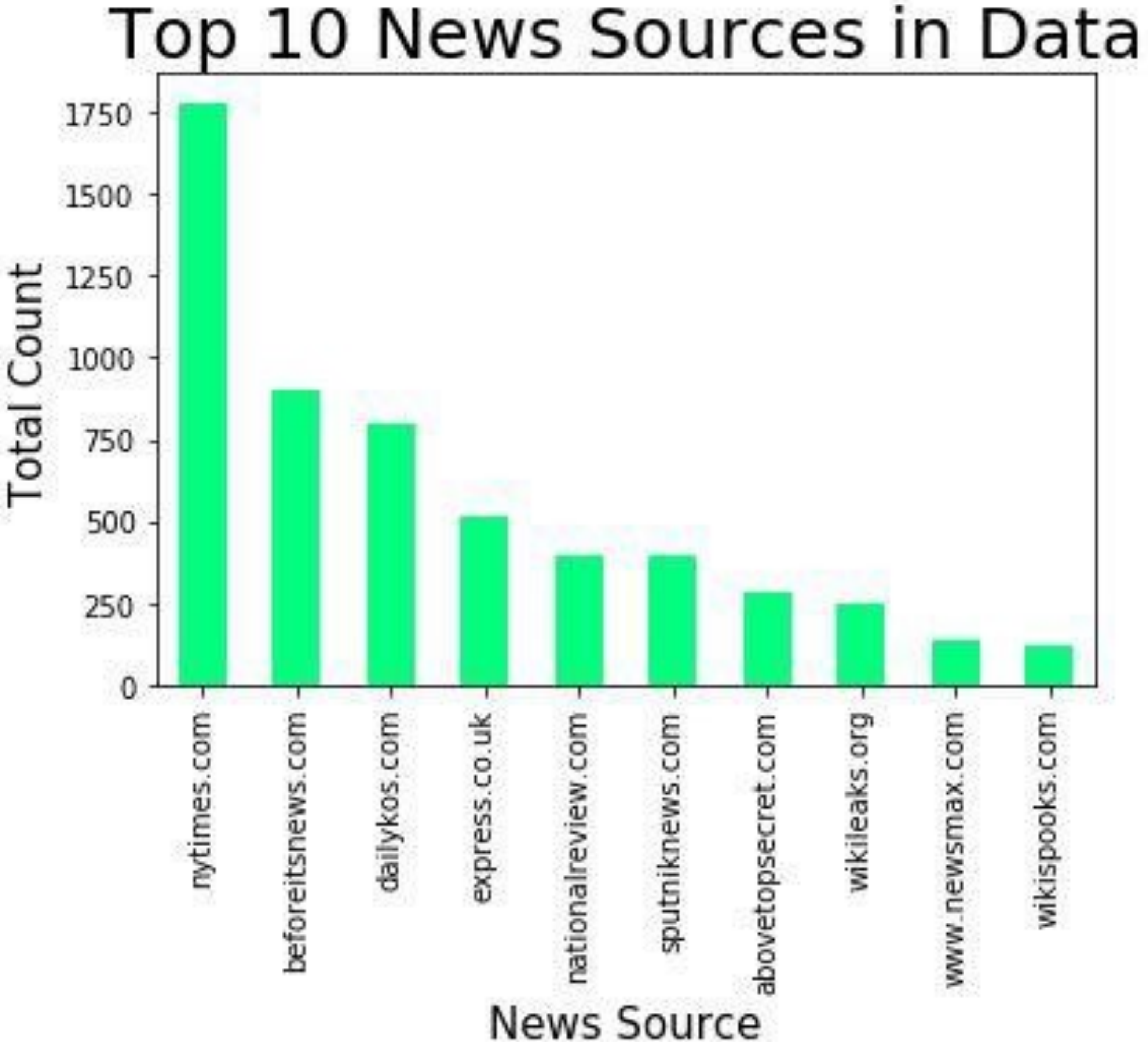
Introduction

- What can machine learning techniques tell us about fake news detection?
- How do different styles of sampling data influence results?

Category Counts in Initial Sampling

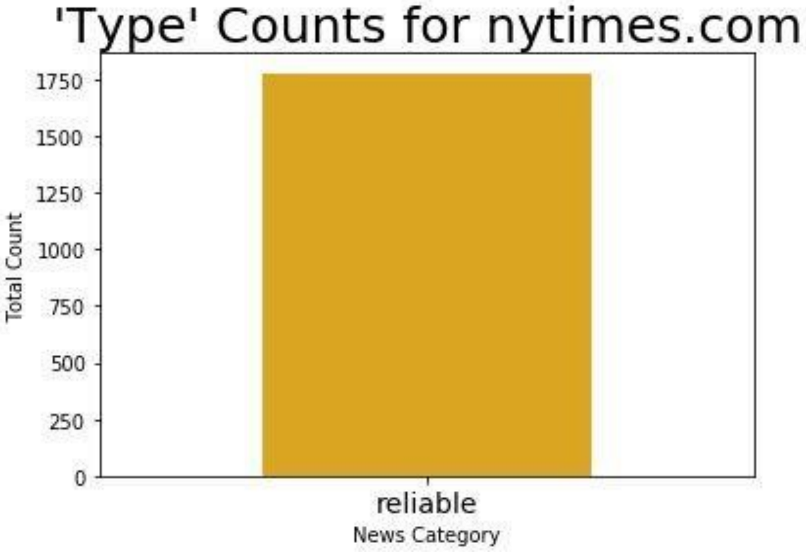


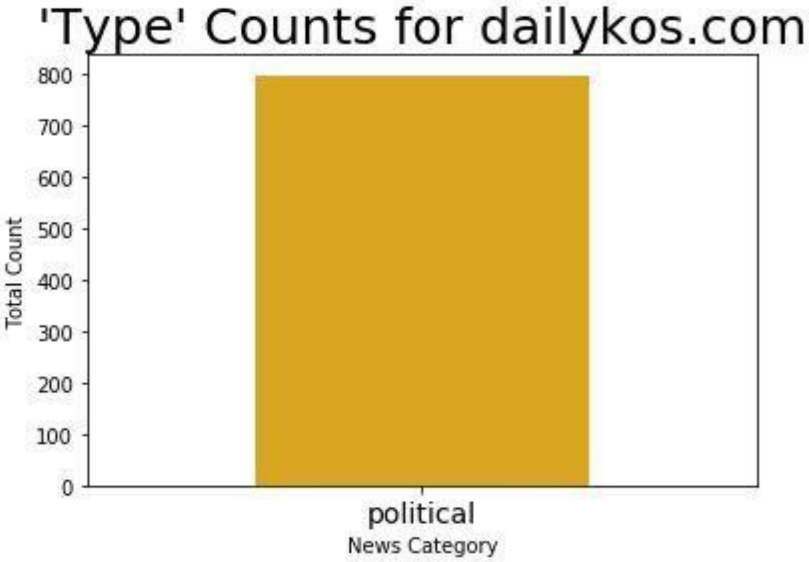
News Source Counts in Initial Sampling

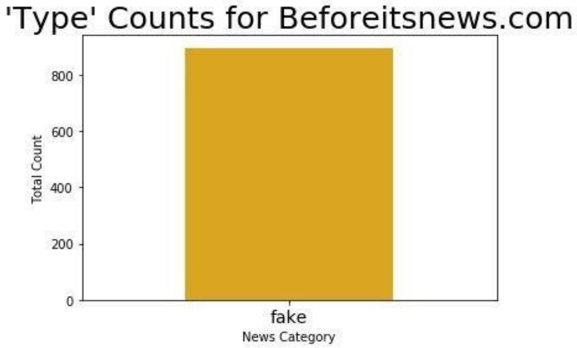


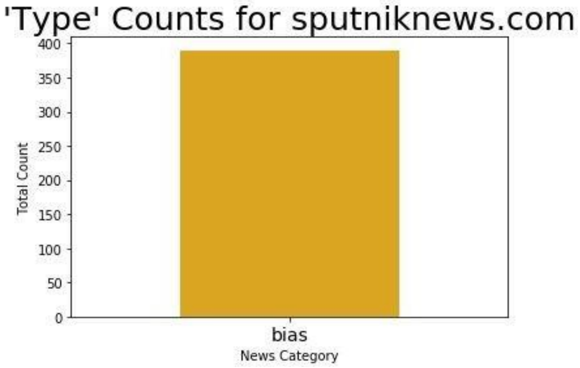
Every single article from a given source is in one category in the data

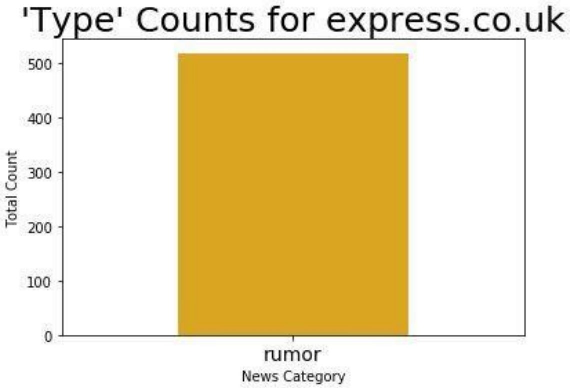
Is this problematic?



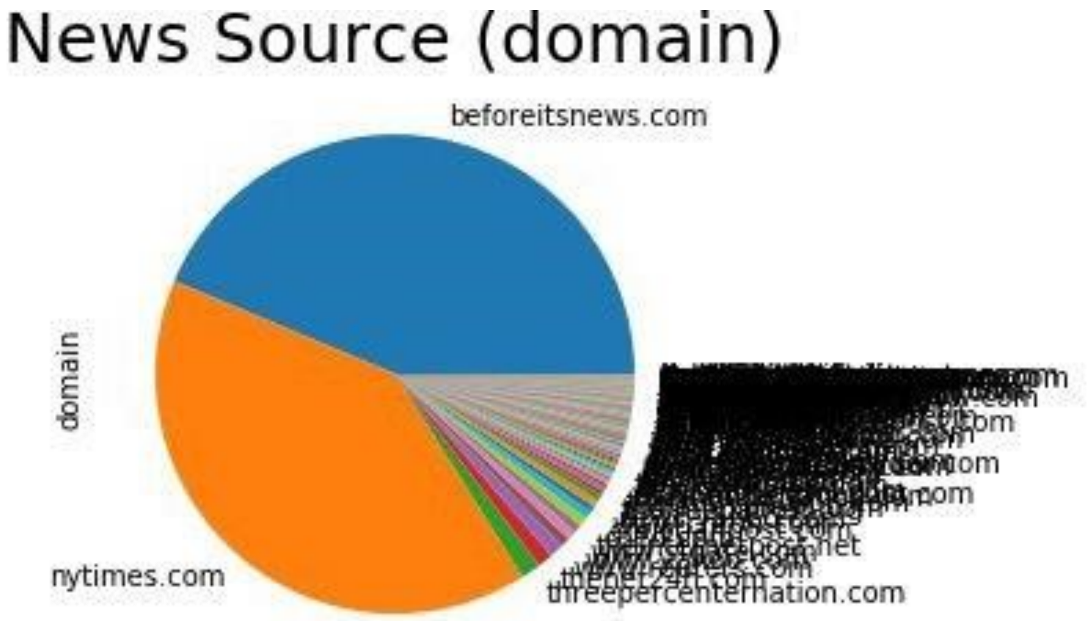


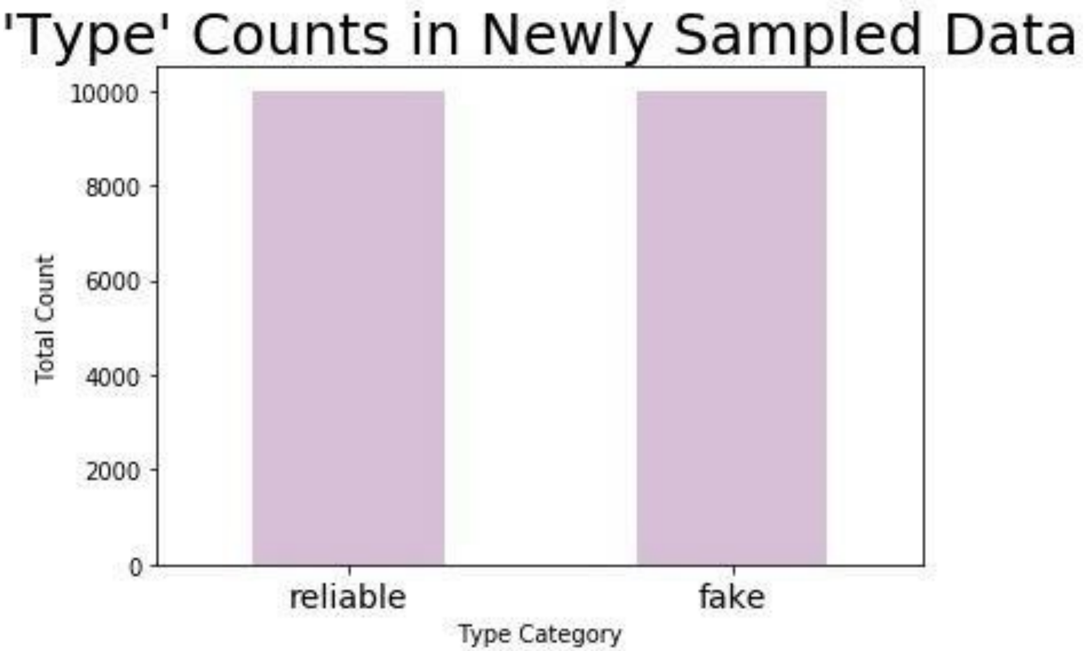






Initial Analysis – Reliable vs. Fake





Sentiment Analysis

	Label	SIA Polarity Score	TextBlob Polarity Score	TextBlob Subjectivity
Fake News	1	5468	983	1128
	-1	3013	67	5784
	0	5013	7944	3088
Reliable News	1	6721	1175	762
	-1	2581	73	6800
	0	698	8752	2438

Predictive Modeling – Reliable vs. Fake

Bag-of-words Vectorization

Predictions were 86.7% accurate with Multinomial Naïve Bayes.

87% accurate with LinearSVC().

87.5% accurate with XGBoost().

Predictive Modeling – Reliable vs. Fake

Tf-idf Vectorization

Predictions were 87.2% accurate with Multinomial Naïve Bayes.

90.8% accurate with LinearSVC().

89.0% accurate with XGBoost().

Predictive Modeling – Reliable vs. Fake

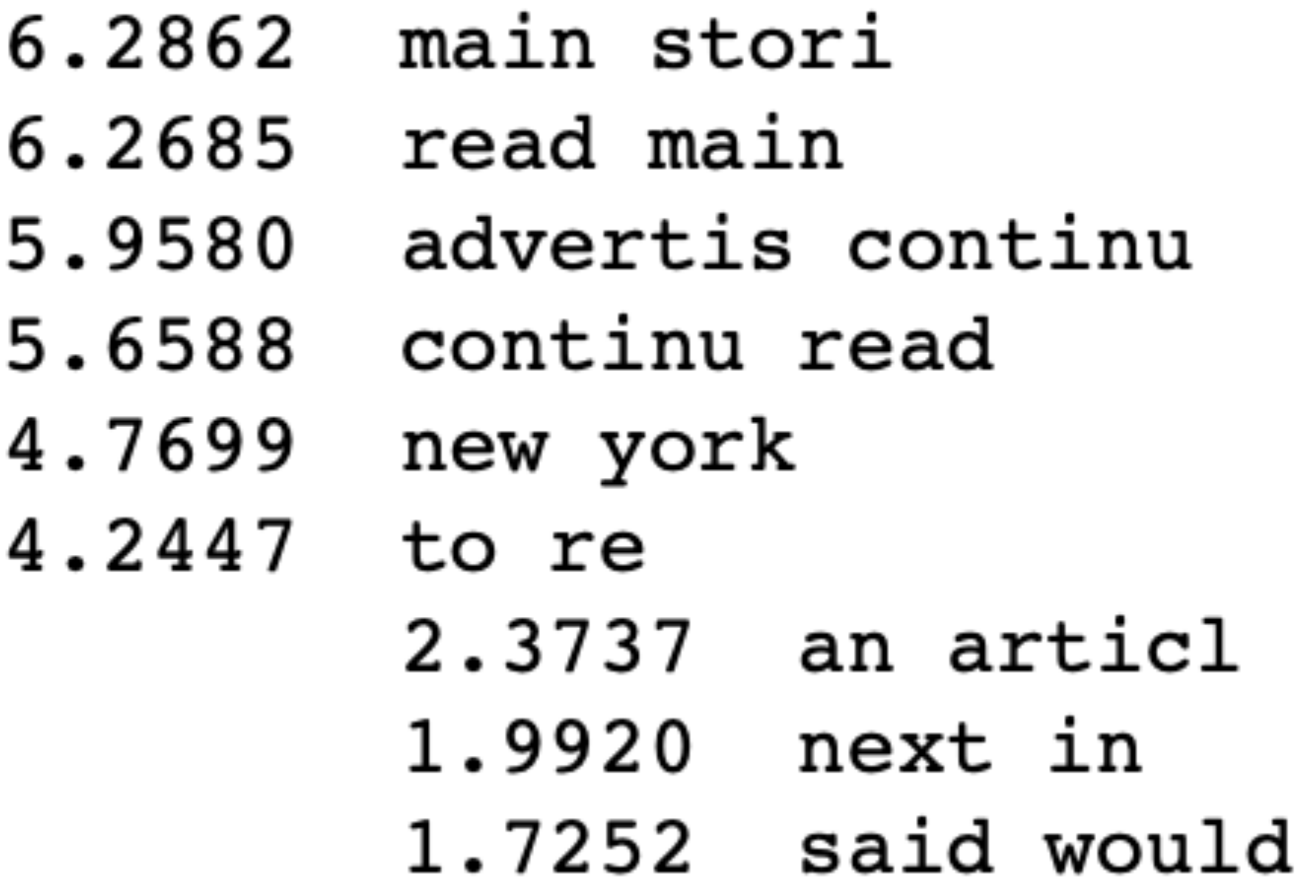
Tf-idf Vectorization with Bigrams

Predictions were 90.4% accurate with Multinomial NB.

91.7% accurate with LinearSVC().

83.3% accurate with XGBClassifier().

Most Predictive features for Initial Reliable vs. Fake Analysis

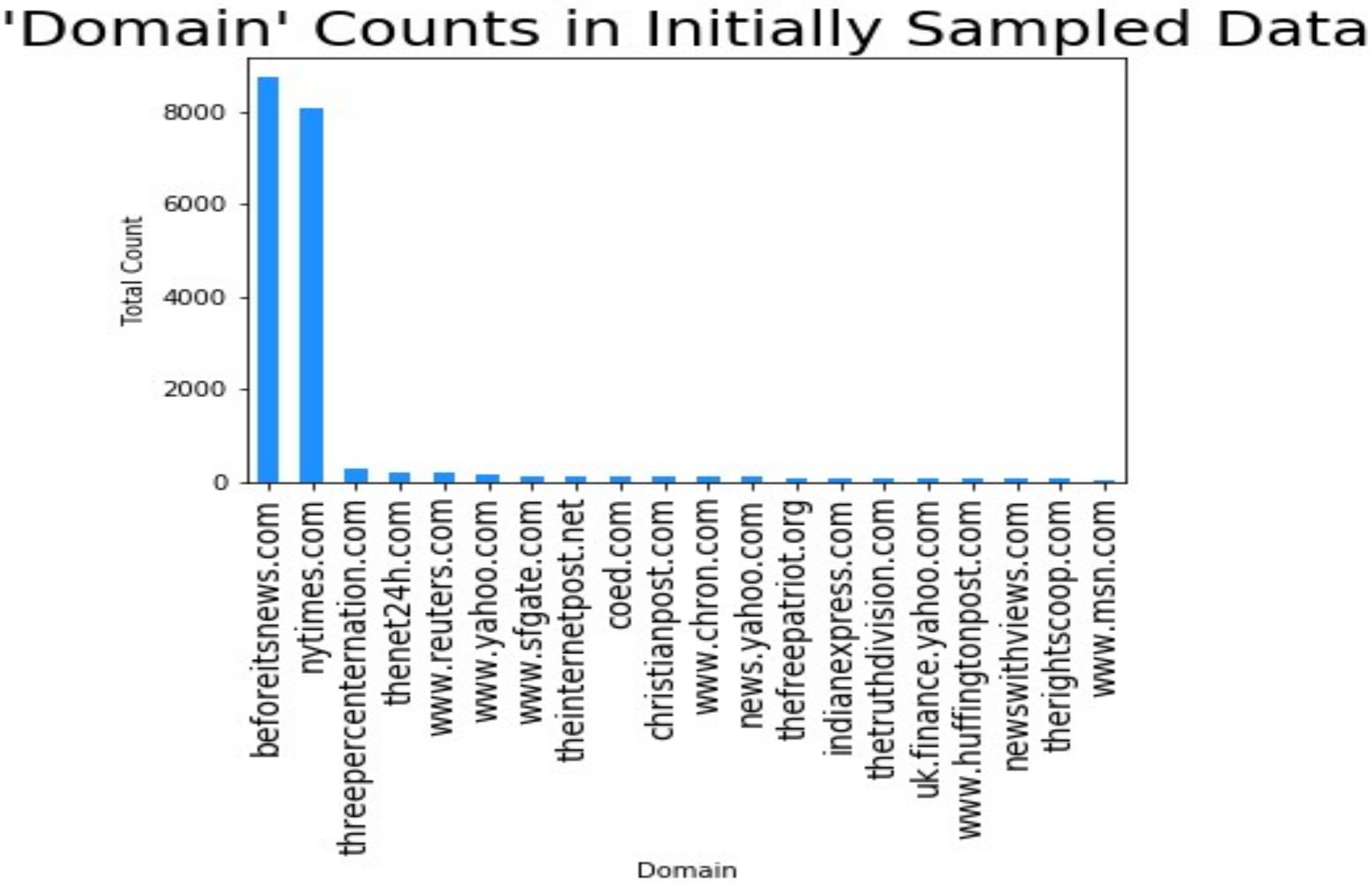


```
6.2862   main stori
6.2685   read main
5.9580   advertis continu
5.6588   continu read
4.7699   new york
4.2447   to re
         2.3737   an articl
         1.9920   next in
         1.7252   said would
```

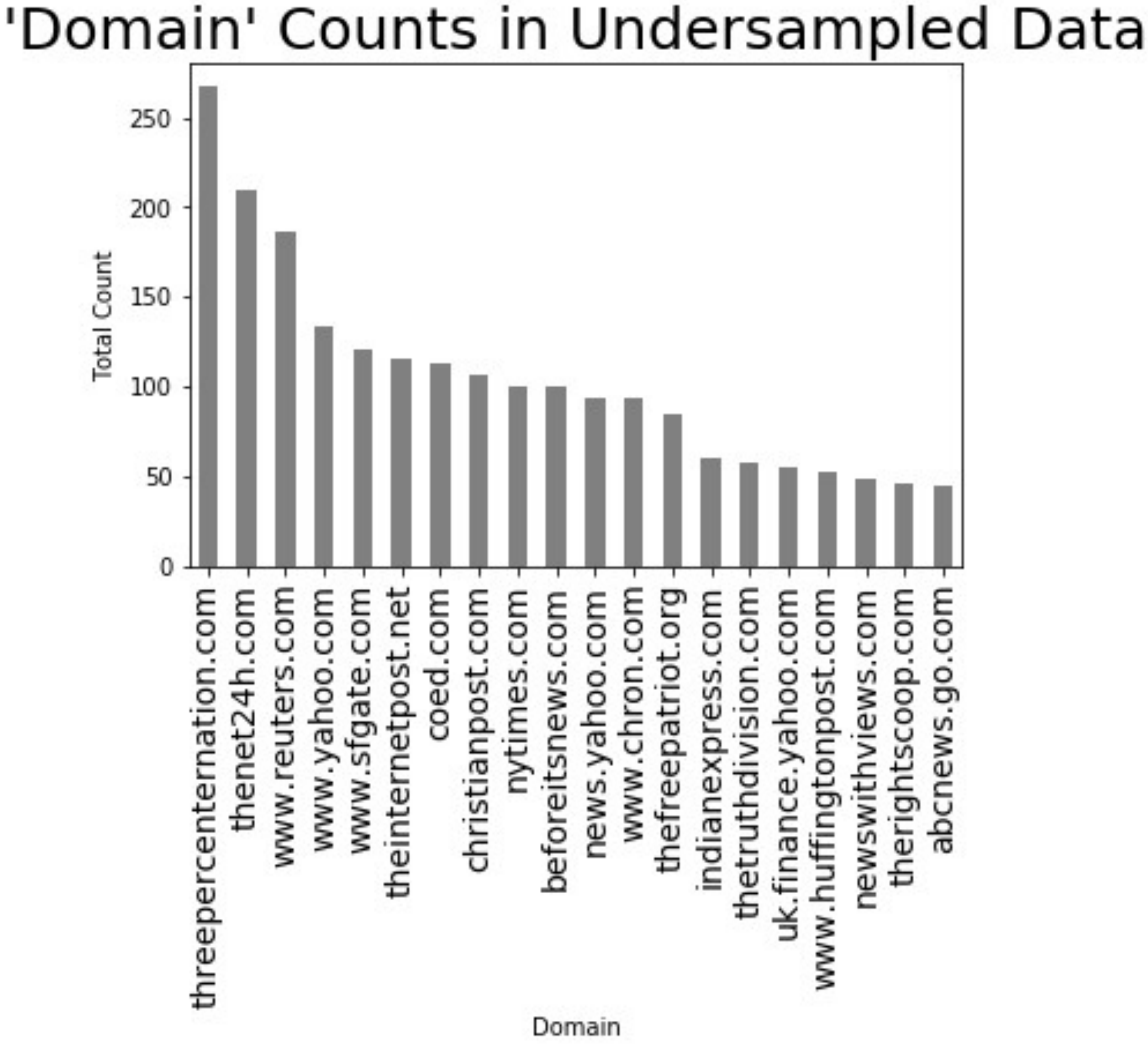
- These results indicate that the fact that a given article is from the New York Times is more predictive than anything else in the data. Because of this, the data was resampled (see following slide)

Resampling

The New York Times and beforeitsnews.com were vastly overrepresented in the initial sample:



Resampling – data was resampled for better balance across domain

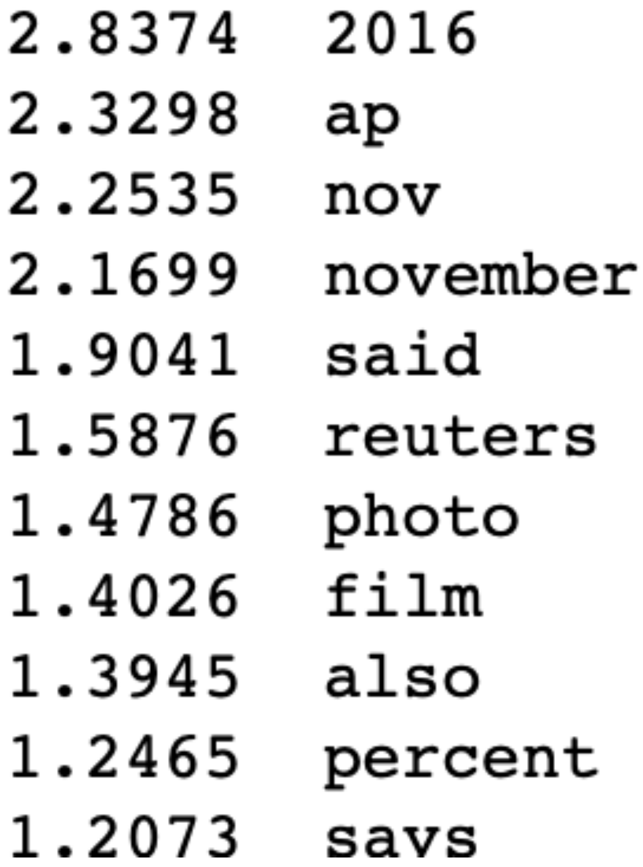


Predictive Accuracy

BAG-OF-WORDS	Tf-idf	Tf-idf with two bigrams
Predictions were 78.4% accurate with Multinomial NB.	Predictions were 63.9% accurate with Multinomial NB.	Predictions were 66.0% accurate with Multinomial NB.
80.7% accurate with LinearSVC().	85.6% accurate with LinearSVC().	82.1% accurate with LinearSVC().
79.8% accurate with XGBClassifier().	84.6% accurate with XGBClassifier().	78.7% accurate with XGBClassifier().

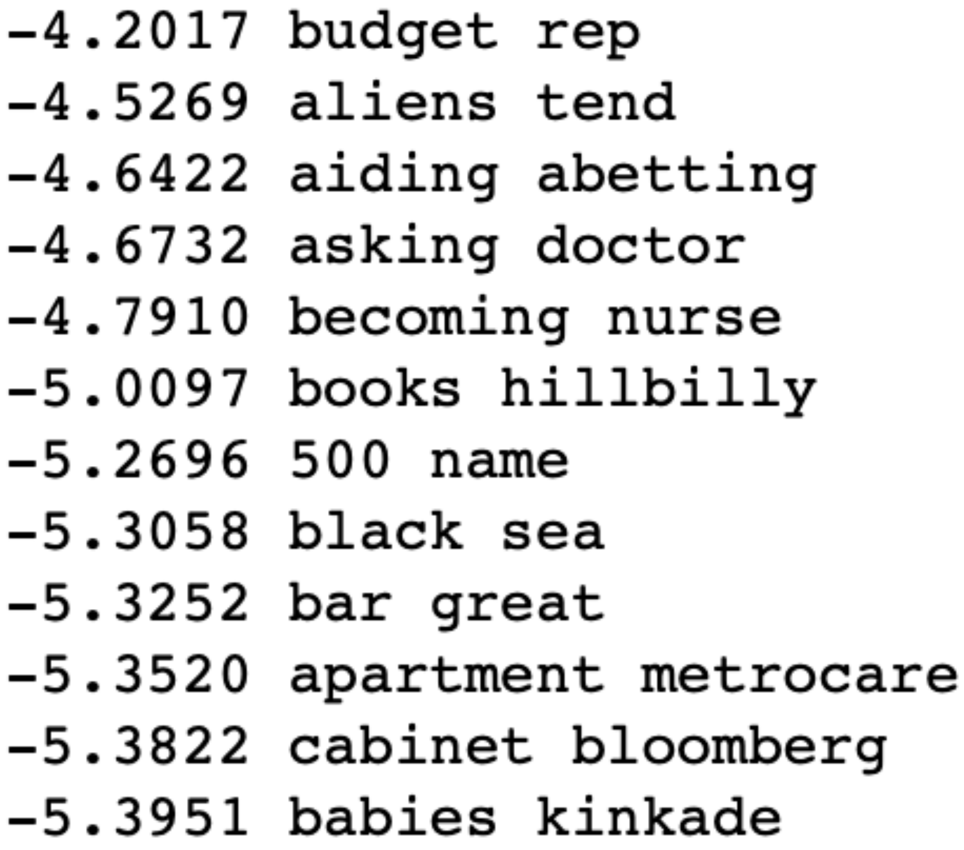
Predictive Features

Unigrams

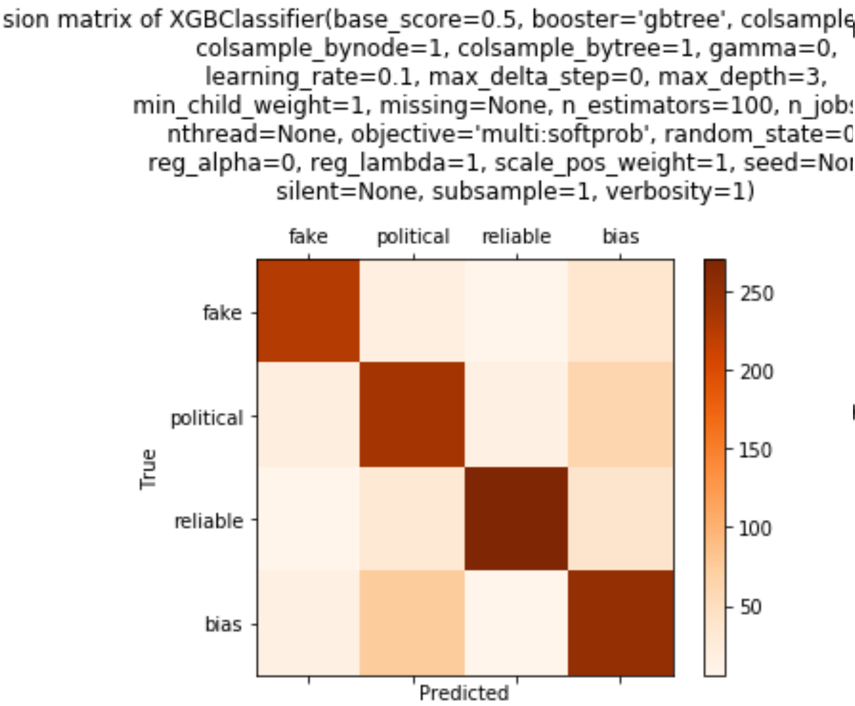


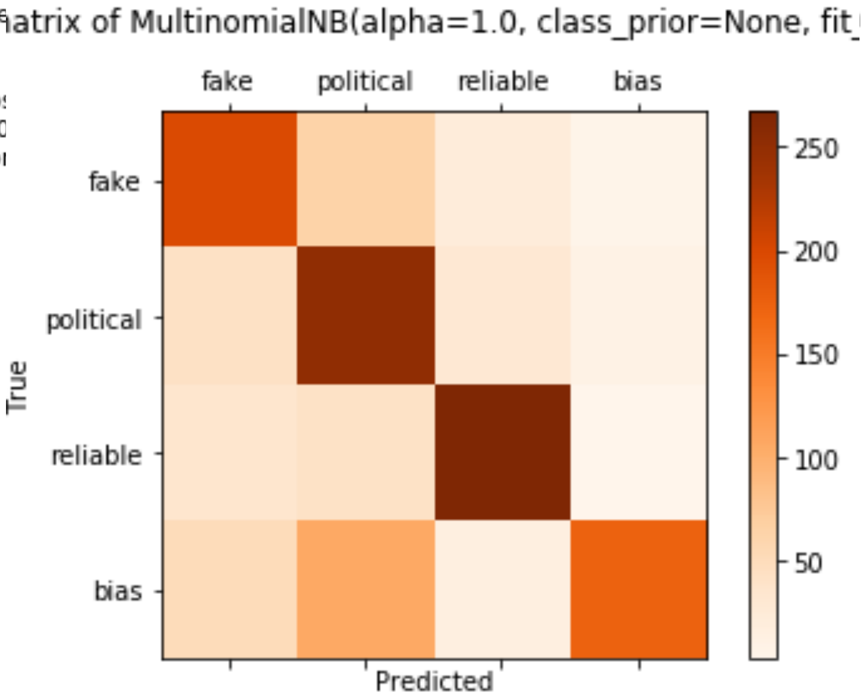
```
2.8374   2016
2.3298   ap
2.2535   nov
2.1699   november
1.9041   said
1.5876   reuters
1.4786   photo
1.4026   film
1.3945   also
1.2465   percent
1.2073   says
```

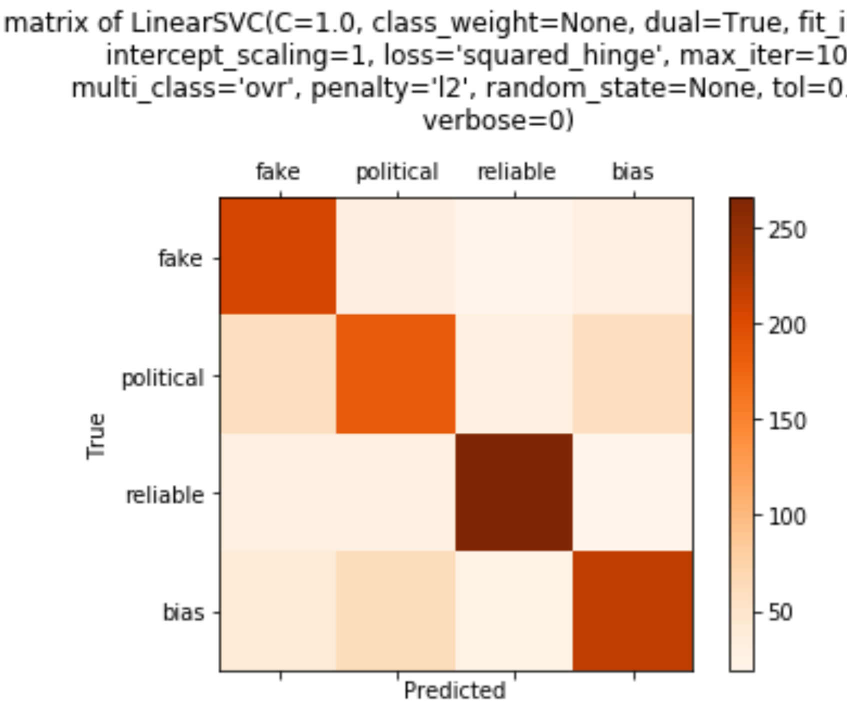
Bigrams



```
-4.2017 budget rep
-4.5269 aliens tend
-4.6422 aiding abetting
-4.6732 asking doctor
-4.7910 becoming nurse
-5.0097 books hillbilly
-5.2696 500 name
-5.3058 black sea
-5.3252 bar great
-5.3520 apartment metrocare
-5.3822 cabinet bloomberg
-5.3951 babies kinkade
```



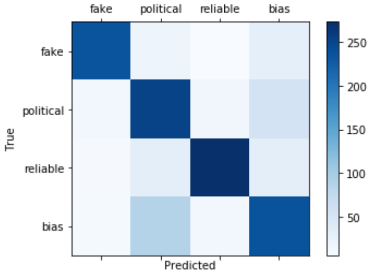




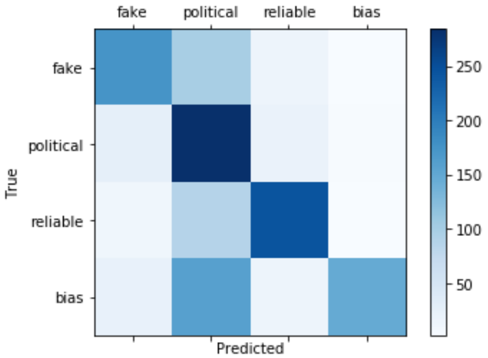
Multiclass Classification with count vectorization

Multiclass Classification with tf-idf vectorization

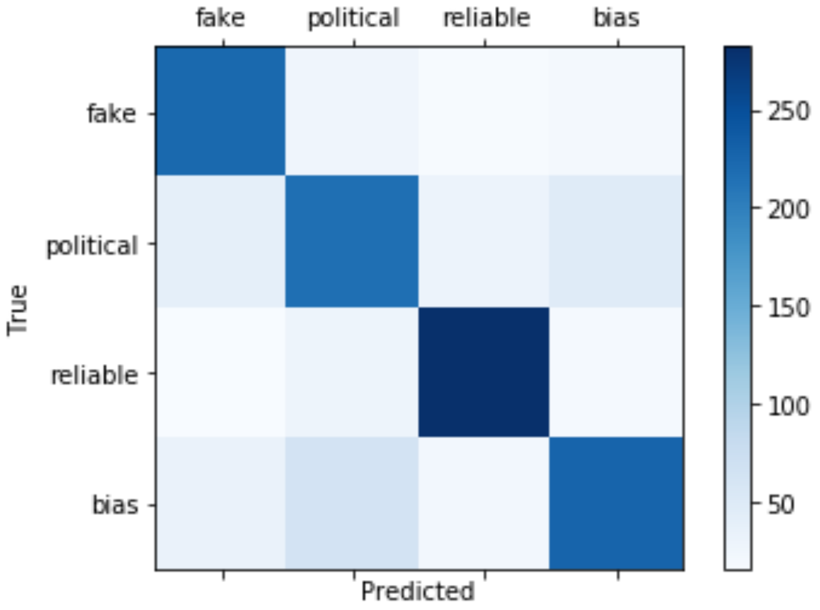
Confusion matrix of XGBClassifier(base_score=0.5, booster='gbtree', colsample_bylevel=1, colsample_bynode=1, colsample_bytree=1, gamma=0, learning_rate=0.1, max_delta_step=0, max_depth=3, min_child_weight=1, missing=None, n_estimators=100, n_jobs=1, nthread=None, objective='multi:softprob', random_state=0, reg_alpha=0, reg_lambda=1, scale_pos_weight=1, seed=None, silent=None, subsample=1, verbosity=1)



Confusion matrix of MultinomialNB(alpha=1.0, class_prior=None, fit_prior=True)

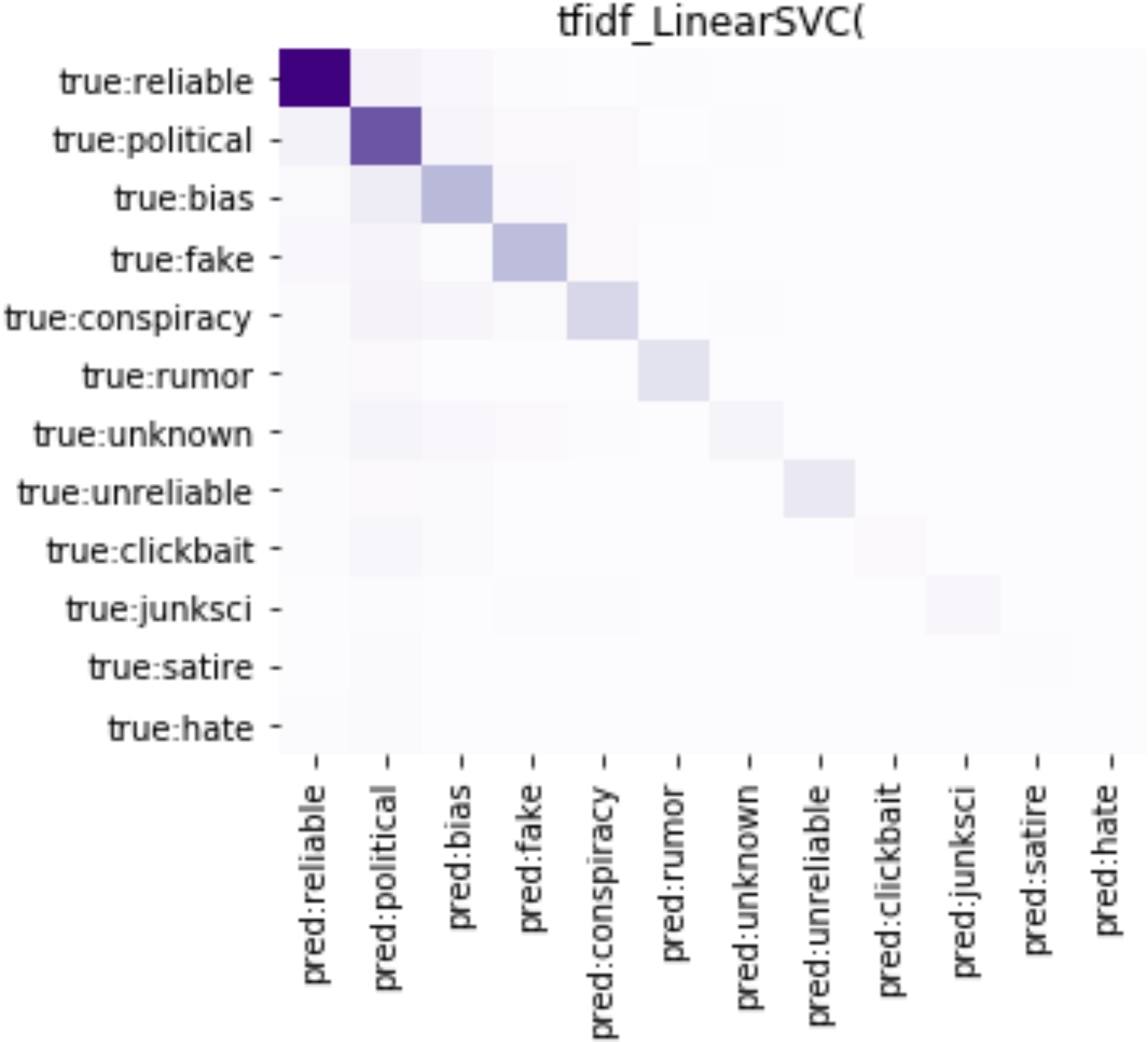


Confusion matrix of LinearSVC(C=1.0, class_weight=None, dual=True, fit_intercept=True, intercept_scaling=1, loss='squared_hinge', max_iter=1000, multi_class='ovr', penalty='l2', random_state=None, tol=0.0001, verbose=0)



More Classes

- The same six pairs of vectorization and classification method were applied to the data, but for all of the classes (instead of just the four largest ones). There were too few observations of the the smaller classes for such an analysis to be very useful. Results were similar to the analysis with just the four largest classes.



Takeaway Points

- Machine learning classification can, with a significant degree of accuracy, predict fake news.
- Prediction seems to be inherently tied to source, as some of the most informative features were direct references to the site or online newspaper that articles came from.
  - This highlights the problem of designating all articles from a given source with the same reliability category.
- This analysis was conducted with a smaller subset of data from the very large dataset. Conducting the analysis with a larger amount of data may yield more interesting results.