Data Foundations: Final Project

Team Name: Automat3rs
Team: Brenda Parnin, Rudy Martinez, Jose Fernandez, Arsalan Bhojani
Submission Date: May 7, 2021

Background
Today, social media has become a large part of society. At our fingertips, we can leverage social media platforms to share opinions, life updates, and a variety of other aspects of our lives. We are able to read or watch the news, participate in live events, and communicate with others who may be hundreds of thousands of miles away. Due to the connectivity social media provides, an unfathomable amount of data is produced, and it has the inherent potential to unite or destroy.

Describe Your Approach
For this project, our task is to create a model that accurately predicts the classification of offensive language in Twitter tweets. To initiate this process, we were provided two files: a *train.tsv* dataset and a *test.tsv* dataset. We used the train dataset to train our model while the test dataset contained test examples. The data in each of the datasets were as follows: *Twitter identification numbers*, *Text of Twitter tweets*, and *Labels that correspond to the tweet's classification*. The labels are categorized as:

- *Not Offensive (NOT)*: Posts that do not contain offense or profanity
- *Targeted Insult (TIN)*: Posts containing insult/threat to an individual, a group, or others
- *Untargeted (UNT)*: Posts containing non-targeted profanity and swearing. Posts with general profanity are not targeted, but they contain non-acceptable language.

To determine which offensive classification model to use, each group member tested LinearSVC, RandomForest, and RandomForest + Lexicon models. Each model that was tested contained different parameters and GridSearchCV values. For the RandomForest + Lexicon model, a negative word dataset taken from the Data Foundations course was used to create the Lexicon Features, and these features were incorporated into the RandomForest model to enhance its performance (Appendix A). Python script was developed to duplicate the test dataset, convert it into a data frame and replace the existing labels with the models' predictions. The updated data frame was then exported to CSV for viewing (Appendix B).

Model Selection Procedure
In our exploration of possible offensive language classification models, we determined that the Random Forest + Lexicon machine learning model yielded the best validation results. After experimenting with the parameters and Lexicon features detailed in the *Describe Your Approach* section, the below Random Forest + Lexicon model produced the highest F1 Micro and Macro scores of 0.8758 and 0.3113 respectively (Appendix C).

Random Forest + Lexicon Validation Scores

F1 Micro: 0.8758, **F1 Macro:** 0.3113

Random Forest Validation Scores

F1 Micro: 0.8667, **F1 Macro:** 0.3095

Final Model Parameters
The following parameters were incorporated into the final model:

```
n_estimators = [10, 30, 50, 70, 90, 100]
parameters = {'n_estimators': n_estimators}
```

The Random forest classifier was initialized, and the below gridsearchCV values were implemented to fit the model on X_train_w_lex:

```
rand_forest_lex = RandomForestClassifier()
clf_rand_lex = GridSearchCV(rand_forest_lex, parameters, cv = 10, verbose = 2)
clf_rand_lex.fit(X_train_w_lex, y_train)
```

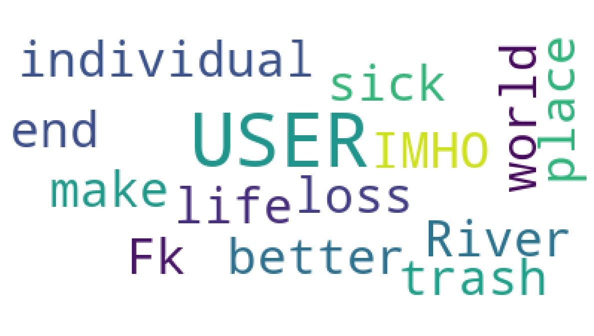
Using these features, a validation score of 0.7342 was produced and a best parameter value of n_estimators equal to 70 was selected.

Error Analysis on RandomForest + Lexicon
The first step in the error analysis process was to take a random sample of the RandomForest + Lexicon model results. Our sample included 30 observations. A detailed view of the selected observations are shown in Appendix D. Through a manual review of these observations individually, we concluded that our model predicted 80% of the observations correctly (24 out of 30). Out of the 6 observations that were classified incorrectly, 5 were found to be false negatives and 1 was found to be a false positive. This is likely due to the lack of specific negative words within this group. Moreover, it

seems that model didn't accurately take into account that the tweets were targeting another Twitter user or users. It's important to note that an increase or decrease in sample size may impact these results as well as the perspective of the annotator.

Word Clouds

The below word clouds are a visual representation of the targeted and untargeted text data in the RandomForest + Lexicon model results (Appendix E).

Targeted Insult Word Cloud	**Untargeted Insult Word Cloud**
	

Appendix A

Negative Word Lexicon Function

```
class LexiconClassifier():
    def __init__(self):

        self.negative_words = set()
        with open('negative-words.txt', encoding='iso-8859-1') as iFile:
            for row in iFile:
                self.negative_words.add(row.strip())

    def count_neg_words(self, sentence):
        num_neg_words = 0
        for word in sentence.lower().split():
            if word in self.negative_words:
                num_neg_words += 1
        return num_neg_words
```

Appendix B

Predictions Data Frame

```
rand_test_predictions_lex_df = pd.DataFrame(data = rand_test_predictions_lex)

# Write to a csv
with open('test.tsv') as file:
    test_file = csv.reader(file, delimiter = '\t', quoting = csv.QUOTE_NONE)
    rf_test_lex_df = pd.DataFrame(test_file)

rf_test_lex_df[2] = rand_test_predictions_lex_df[0]
rf_test_lex_df.columns = ['TWITTER_ID', 'TEXT', 'PREDICTIONS']
rf_test_lex_df.to_csv('00. RandomForest + Lexicon Prediction Results.csv', index =
False)
```

Appendix C

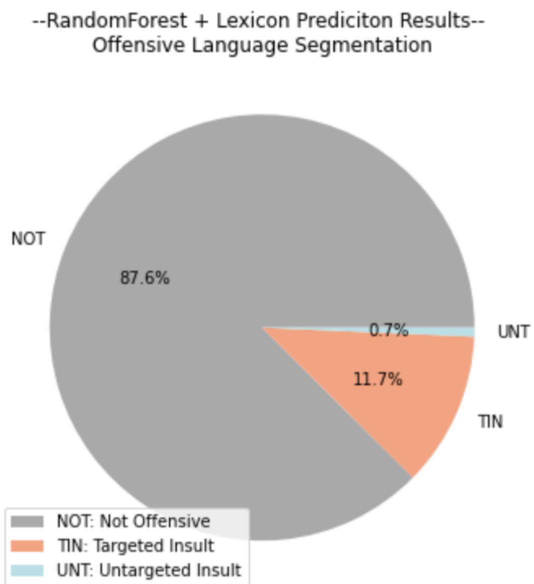
Random Forest + Lexicon Validation Score

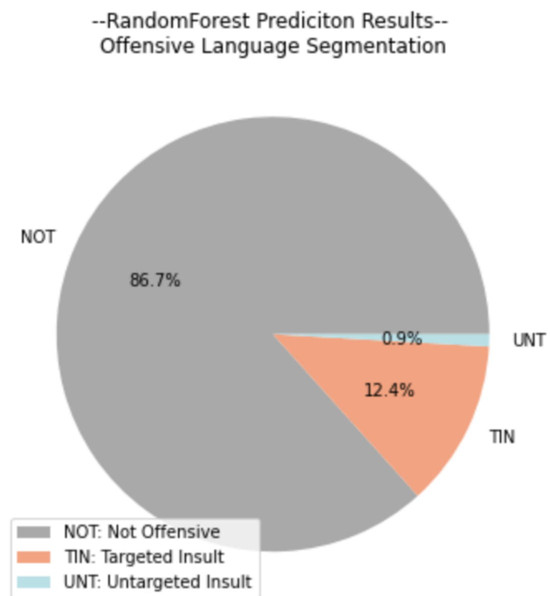
```
f1_micro_rand_lex = f1_score(y_test, rand_test_predictions_lex, average = 'micro')
f1_macro_rand_lex = f1_score(y_test, rand_test_predictions_lex, average = 'macro')
print("F1 Micro: {:.4f}".format(f1_micro_rand_lex))
```

```
print("F1 Macro: {:.4f}".format(f1_macro_rand_lex))
```

Random Forest Validation Score

```
f1_micro_rand = f1_score(y_test, rand_test_predictions, average = 'micro')
f1_macro_rand = f1_score(y_test, rand_test_predictions, average = 'macro')
print("**Scoring**\n")
print("F1 Micro: {:.4f}".format(f1_micro_rand))
print("F1 Macro: {:.4f}".format(f1_macro_rand))
```





Appendix D

Error Analysis

Tweet	Predictions	Error Analysis	Model Prediction
@USER What are you afraid of liberals. You screamed for transparency!Now that POTUS is doing just that you are screaming to keep it secret.The truth is there. You don‚Äôt want it seen.Liberals are hypocrites & the	NOT	TIN	Incorrect (False Negative)

alphabet agencies have some bad apples involved in a soft coup. Tick Tock			
90% of these conservatives" really aren't. They just found the liberal side of the aisle to crowded. URL	**NOT**	**NOT**	**Correct**
@USER @USER @USER That's literally in the Bible that conservatives love to thump but never read.	**NOT**	**UNT**	**Incorrect (False Negative)**
@USER lol fuck you	**TIN**	**TIN**	**Correct**
@USER @USER @USER @USER *accept. I don,Äôt hate men. I don,Äôt hate conservatives. I hate stupid. And I hate hypocrites. And the fact that you think being a conservative matters at all in this matter means to me that you are not so much into law and order as you think you are.	**TIN**	**TIN**	**Correct**
. . . And we wonder why conservatives who work for Twitter fear saying so . . URL	**NOT**	**NOT**	**Correct**
@USER @USER Just another Dem tactic. Leftist do that not good conservatives.	**NOT**	**NOT**	**Correct**
@USER @USER @USER @USER Lord Of The Pies should kept his c*ck holster shut. Nobody gives a toss what you Hollyweirdo's have to say. #MAGA	**NOT**	**TIN**	**Correct**
@USER shit doesn,Äôt feel real to me. then again... life is very whack and i don,Äôt like how it,Äôs set up to begin with so. idk.	**TIN**	**NOT**	**Incorrect (False Positive)**
@USER I think I would pick that just for the fun of itüòÇüòÇ	**NOT**	**NOT**	**Correct**
@USER That solved the problem in Chicago alright!! I wish these ,Äúleaders,Äù would think up something different to get attention. Gun control is getting old.	**NOT**	**NOT**	**Correct**
@USER @USER @USER @USER Evidence against Manafort was substantial and beyond reasonable doubt" on 8 counts. He was found guilty in court, any	**NOT**	**NOT**	**Correct**

other spin is useless. This is how our system works. If Manafort had problems with verdicts he can appeal them, he is not."			
@USER @USER @USER I have to believe this is a code for signaling pedos & rapists world wide to do the Antifa peaceful protest.	**NOT**	**TIN**	**Incorrect (False Negative)**
@USER @USER That's how liberals interoperate the law! One more reason to make sure we flood the courts with conservative Judges.	**NOT**	**NOT**	**Correct**
@USER You Conservatives want to talk about scandals! Really? Let,Äôs get started. If you have the balls. Let,Äôs get busy you chickens.	**NOT**	**TIN**	**Incorrect (False Negative)**
@USER @USER I kind of feel like AntiFa are terrorists.but maybe I'm overstating it lol ü§∑‚Äç‚ôÇÔ∏è	**NOT**	**NOT**	**Correct**
...[bring] our railways back into public ownership, as franchises expire or, in other cases, with franchise reviews or break clauses. We will introduce a Public Ownership of the Railways Bill to repeal the Railways Act 1993 under which the Conservatives privatised our railways.""	**NOT**	**NOT**	**Correct**
@USER There,Äôs false reports of an antifa professor who was assaulted 36 ish (she can,Äôt quite recall) years ago by an upstanding SCOTUS nominee. You should look into that.	**NOT**	**NOT**	**Correct**
LOL Ted Cruz thinks #Texas voters are stupid. Says Beto O'Rourke will ban barbeque from the state of Texas. #resist #MAGA #Ridiculous URL	**TIN**	**TIN**	**Correct**
@USER Don,Äôt watch moron movies	**NOT**	**TIN**	**Incorrect (False Negative)**
@USER Its clear you wouldn't listen anyways since you are more interested in arguing over the non existent	**TIN**	**TIN**	**Correct**

racism of that tweet rather than calling out the ignorant white girl on display			
@USER @USER A liter of filter semi-skimmed milk costs 85p per ltr retail in Tesco School children drink about 35 ltr of milk per annum That works out to ¬£30 pa at retail prices - repeat RETAIL There are 1 million kids allowed school milk Parents are being ripped off to the tune of ¬£10 mil	**NOT**	**NOT**	**Correct**
@USER @USER @USER @USER @USER @USER @USER @USER @USER @USER @USER @USER @USER @USER @USER need to fund a lot things at public schools.... music and art class too. but conservatives always want to pay less taxes and every time they get the chance the cut funding for such things.	**NOT**	**NOT**	**Correct**
@USER @USER Every identity deals with their rep differently. But to say that Conservatives aren't ostracized on Canadian campuses for their views is pretty dumb. Have you never seen a Conservative booth at a uni society expo? Let alone even try to organize something as a conservative?	**TIN**	**TIN**	**Correct**
@USER @USER You do know what the chant Puto means don't you? Fined three times means its not acceptable both are being bullies. One was done by one player the other done by the culture of a country.	**NOT**	**NOT**	**Correct**
@USER its okay but you are the emo too i'm not LOOK TO YOUR ICON AND USERNAME JUST LOOK	**NOT**	**NOT**	**Correct**
@USER üòÇüòÇüòÇüòÇüòÇüòÇ if I say you are mad now you will say I'm tired of you.	**NOT**	**NOT**	**Correct**
@USER China's theft of IP. Is between 400 and 600 billion in economic value. Stop that for MAGA.	**NOT**	**NOT**	**Correct**
@USER @USER @USER @USER So where are all the British builders then?	**NOT**	**NOT**	**Correct**

@USER @USER And he is signing briefs on behalf of Special Counsel Mueller.	**NOT**	**NOT**	**Correct**

Appendix E

Targeted Insult Word Cloud

```
text = neg_df[0]
word_cloud = WordCloud(max_font_size=50, max_words=100,
background_color="white").generate(text)
plt.imshow(word_cloud, interpolation="bilinear")
plt.axis("off")
plt.show()
word_cloud.to_file("00. Targeted Insult Wordcloud.png")
```

Untargeted Insult Word Cloud

```
text = neg_df[0]
word_cloud = WordCloud(max_font_size=50, max_words=100,
background_color="white").generate(text)
plt.imshow(word_cloud, interpolation="bilinear")
plt.axis("off")
plt.show()
word_cloud.to_file("00. Untargeted Insult Wordcloud.png")
```