TOXIC COMMENT CLASSIFICATION

Mario Bravo

William Jessop

THE PROJECT

- Learn how to apply ML to text data
- Build and test a toxic comment binary classifier
- Deploy the trained classifier to a public website

OVERVIEW

- Identify the Dataset
- Clean the Dataset
- Train the Models
- Deploy the Best Model

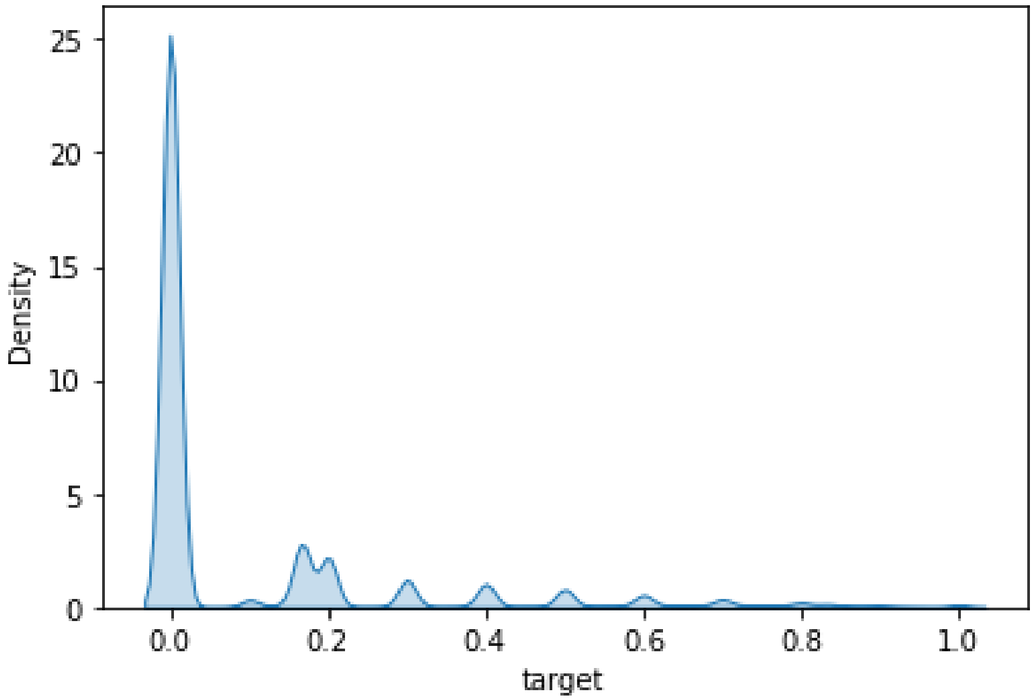
FINDING THE DATA

- We found the dataset on the Kaggle Platform
- The data is a large list of comments made on a public forum
- The data has 1.8 million samples of text with labels

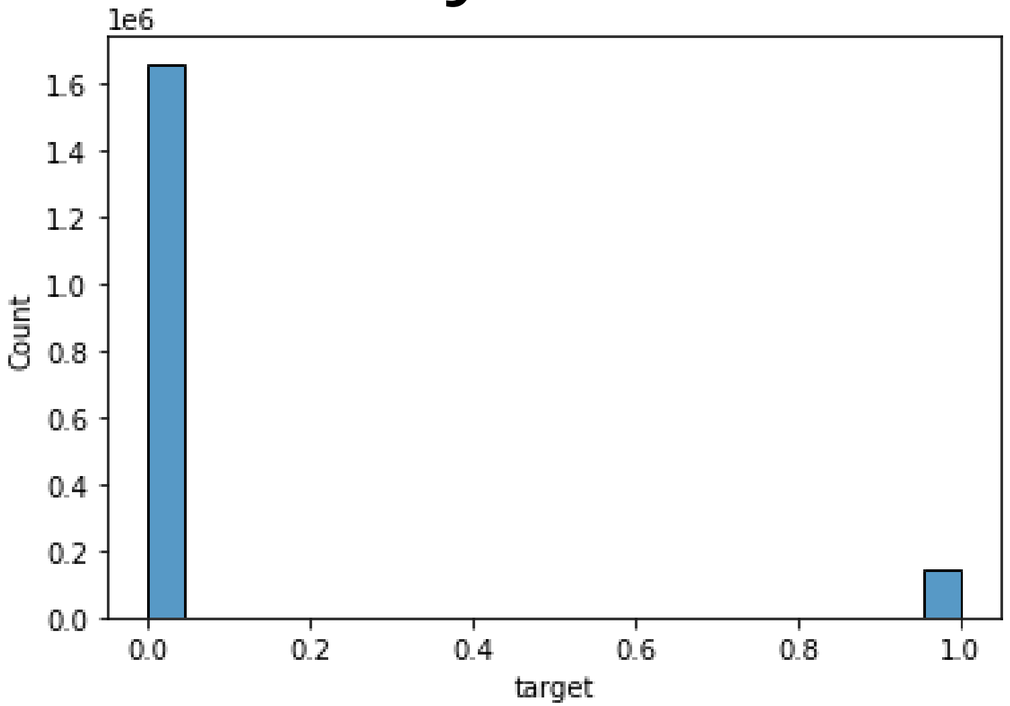
CLEANING THE DATA

Relabeling The Data

Original Labels



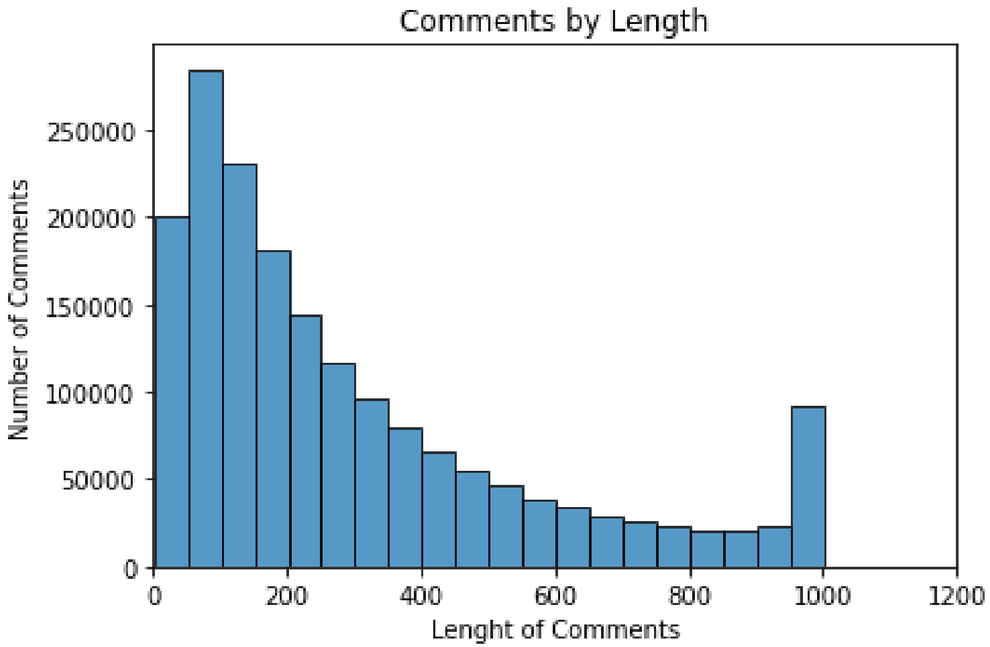
Binary Labels

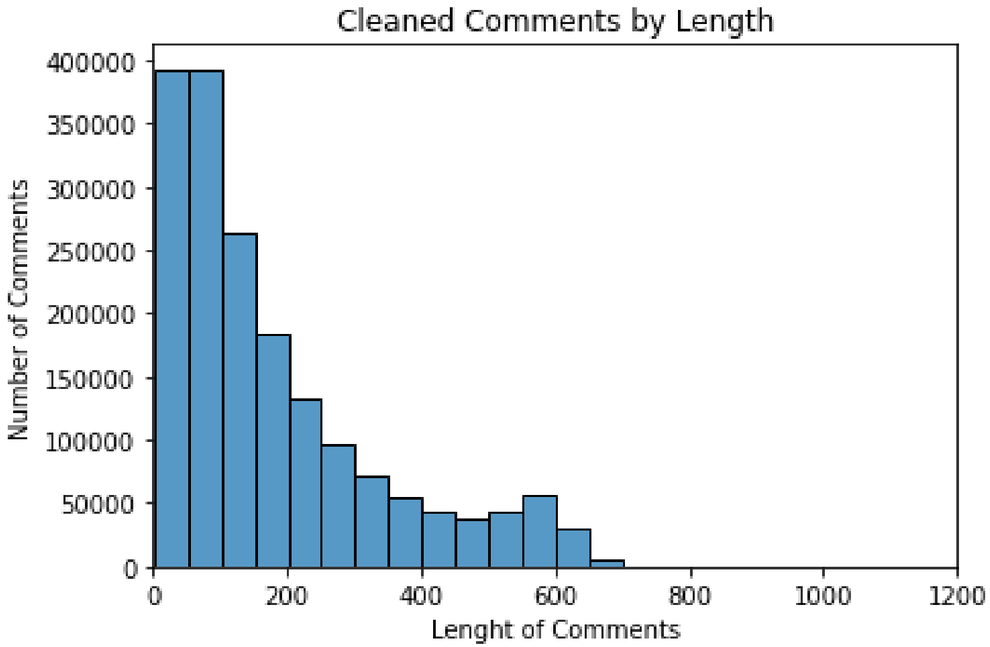


Cleaning the Text

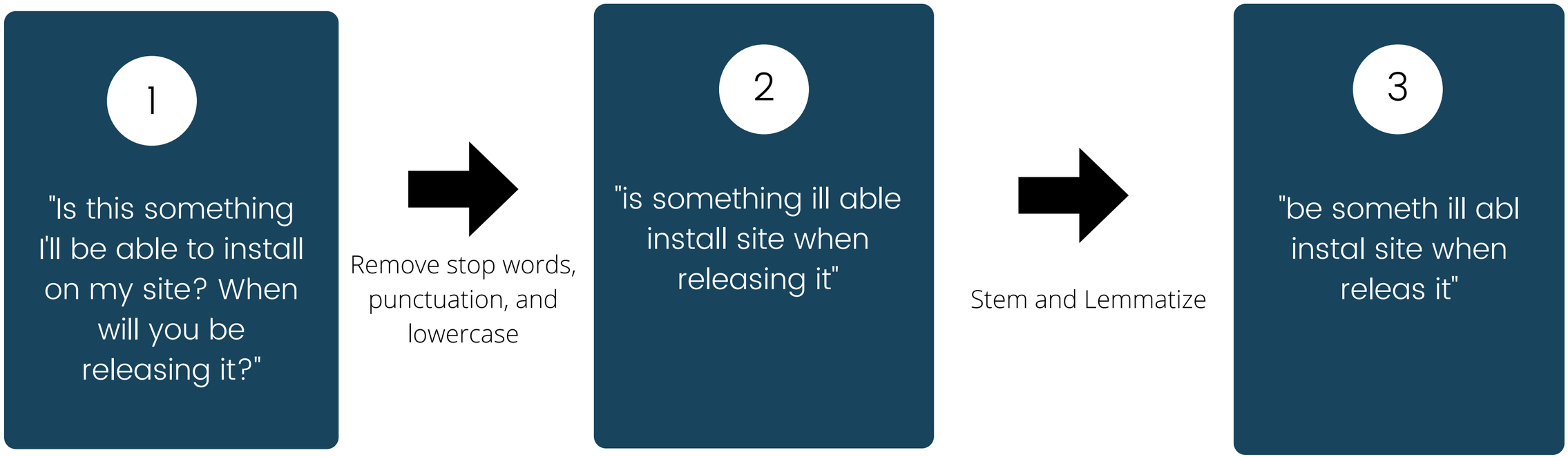
- Removed stop words (e.g. the, of, ...)
- Removed punctuation
- Lowercased all text

- Applied the port stemmer
- Applied a lemmatizer
- Discard empty comments





Example Text Transformation



TRAINING THE MODELS

- We selected these models
  - Stochastic Gradient Decent (SGD)
  - Naïve Bayes
  - Logistic Regression
- The 2 best were hyper parameter optimized

Best Model Scores

SGD

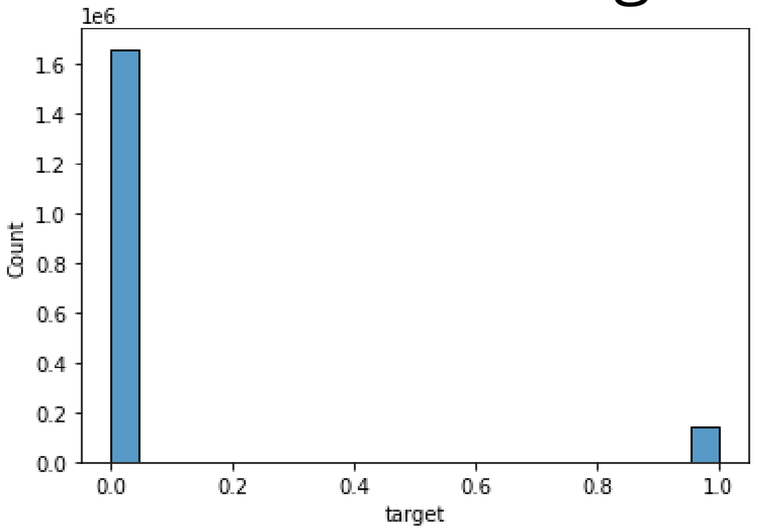
- Overall f1-score: 0.94
- Non-Toxic Recall: 100%
- Toxic Recall: 22%

Logistic Regression

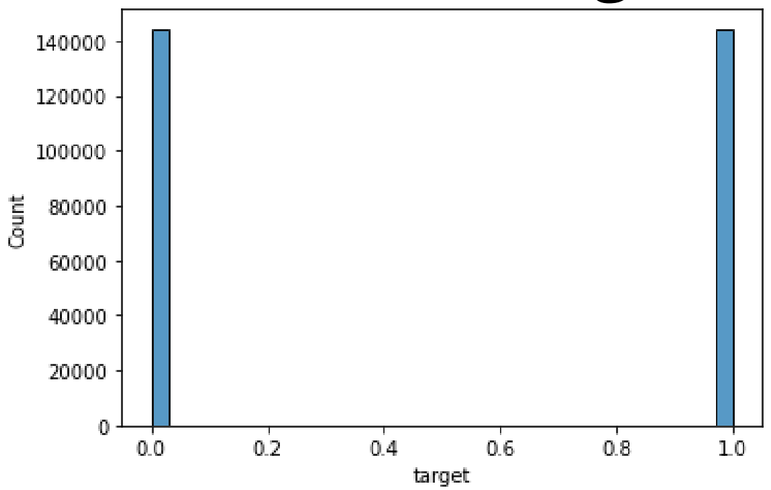
- Overall f1-score: 0.94
- Non-Toxic Recall: 99%
- Toxic Recall: 44%

- These models are very overfitted for classifying **Non-Toxic** comments
- The model has high overall accuracy due to data imbalance

Before Balancing



After Balancing



BALANCING THE DATA

- Our original data was very biased toward Non-Toxic comments
- We balanced the data by randomly removing Non-Toxic comments

Best Model Scores (Balanced Data)

SGD

- Overall f1-score: 0.85
- Non-Toxic Recall: 91%
- Toxic Recall: 80%

Logistic Regression

- Overall f1-score: 0.85
- Non-Toxic Recall: 89%
- Toxic Recall: 81%

- These models have a lower overall accuracy but are not overfitted
- The models were trained with less data overall

DEPLOYMENT

- We used python pickles to store the trained SGD model
- We built a simple app and deployed it to Heroku

https://cmu-toxic-comment-classifier.herokuapp.com/


THANK YOU