

Quora Insincere Questions Classification

Springboard Capstone II

Overview

Quora is a service that helps people learn from each other by asking and answering questions - and a key challenge in providing this type of service is filtering out insincere questions. Quora is attempting to filter out toxic and divisive content to uphold their policy of “Be Nice, Be Respectful”.

What is an Insincere Question?

- **Non-neutral tone**

Uses exaggerated tone to underscore a point about a group of people

- **Disparaging or inflammatory**

Suggest discriminatory ideas or seeks stereotype confirmation

- **Not grounded in reality**

Contain absurd assumptions

- **Use sexual content for shock value**

Data Source

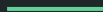
→ <https://www.kaggle.com/c/quora-insincere-questions-classification/data>

Goals

1. Identify and flag insincere questions using machine learning
 2. Maximize F1 Score by accurately predicting whether a question is sincere or not
-

Specialization

- Advanced NLP
- TensorFlow and Keras



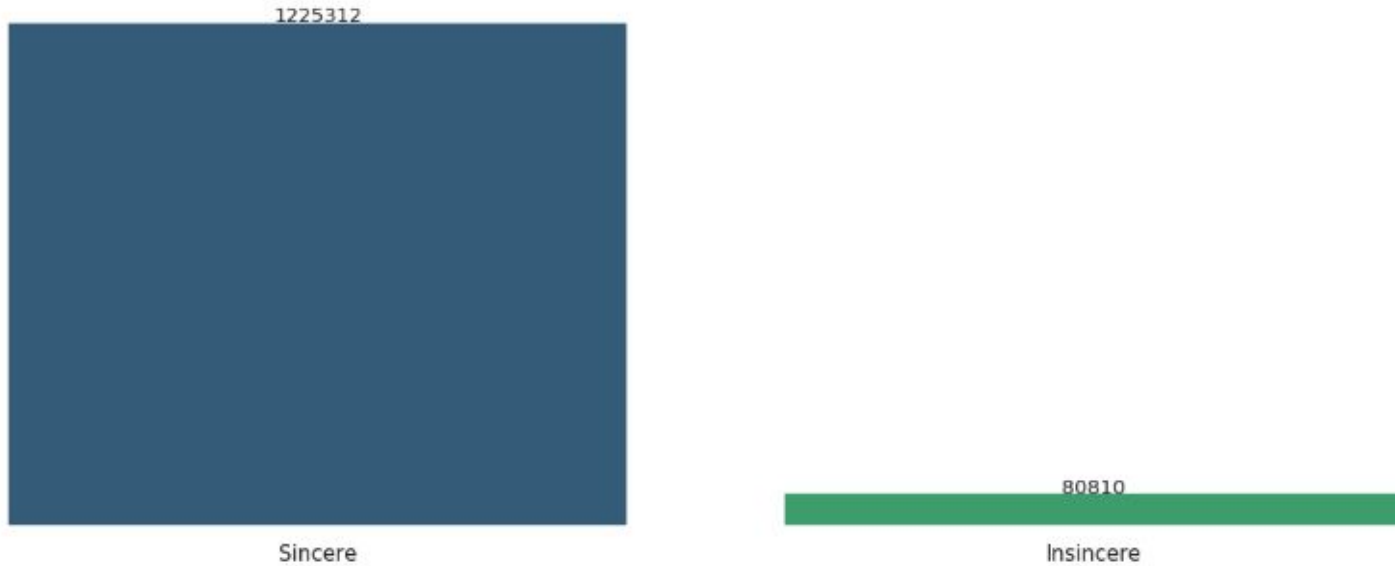
Value of Solution

An accurate solution can help Quora develop more scalable methods to detect toxic and misleading content and combat online trolls at scale

This solution will help Quora to uphold their policy of ‘Be Nice, Be Respectful’

Exploratory Data Analysis

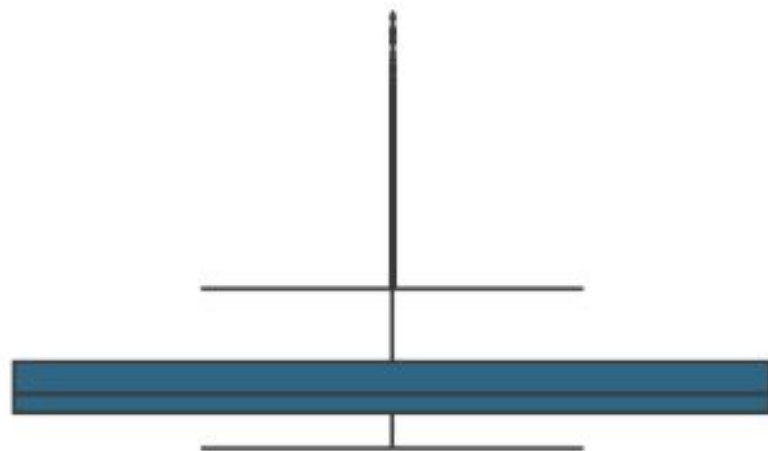
Distribution of Questions



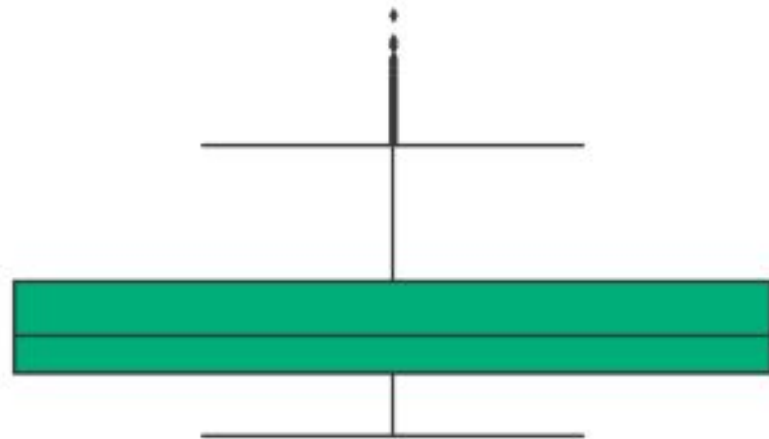
Sincere questions: 93.4%

Insincere questions: 6.6%

Number of Tokens per Question



Sincere



Insincere

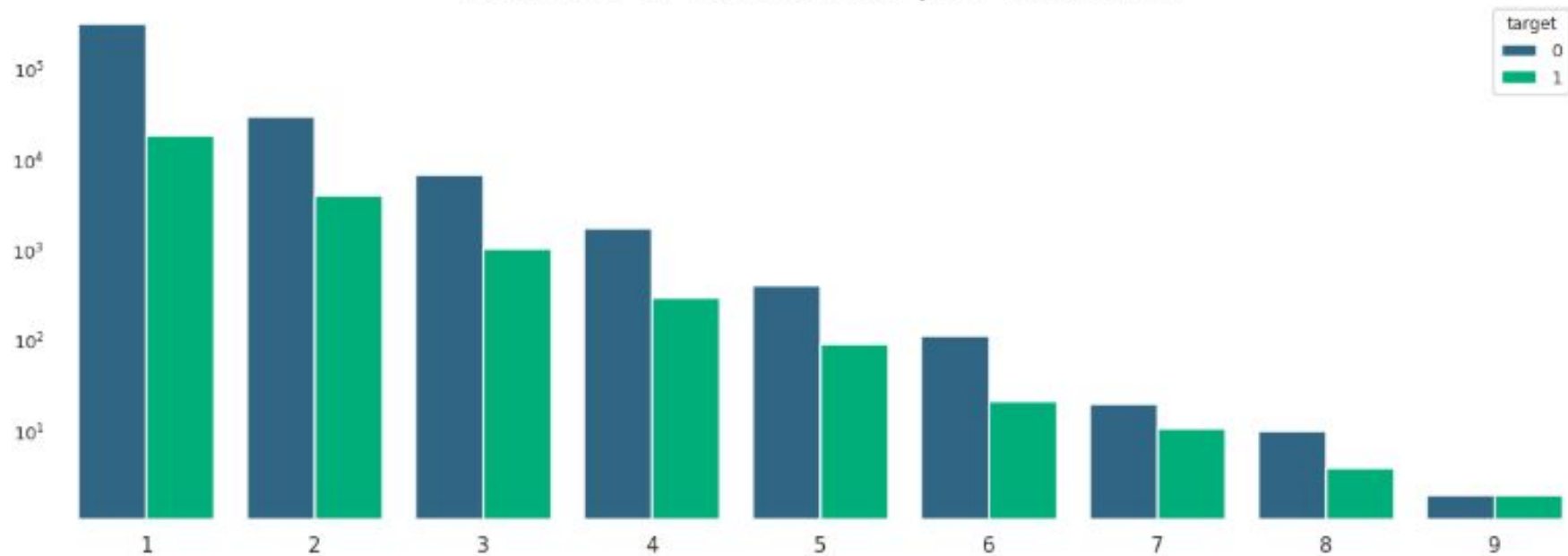
-106.72

T-Test Comparing Number of Tokens

Sincere vs Insincere

P-Value = 0

Number of Sentences per Question



-56.09

T-Test Comparing Number of Sentences

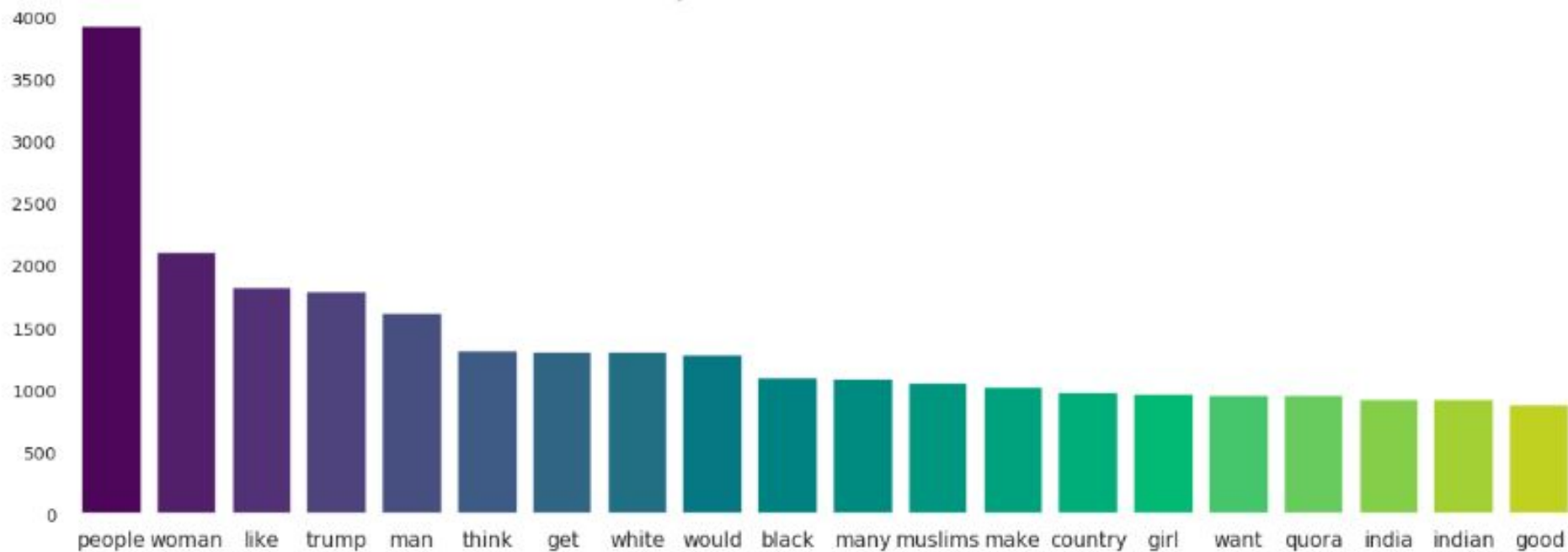
Sincere vs Insincere

P-Value = 0

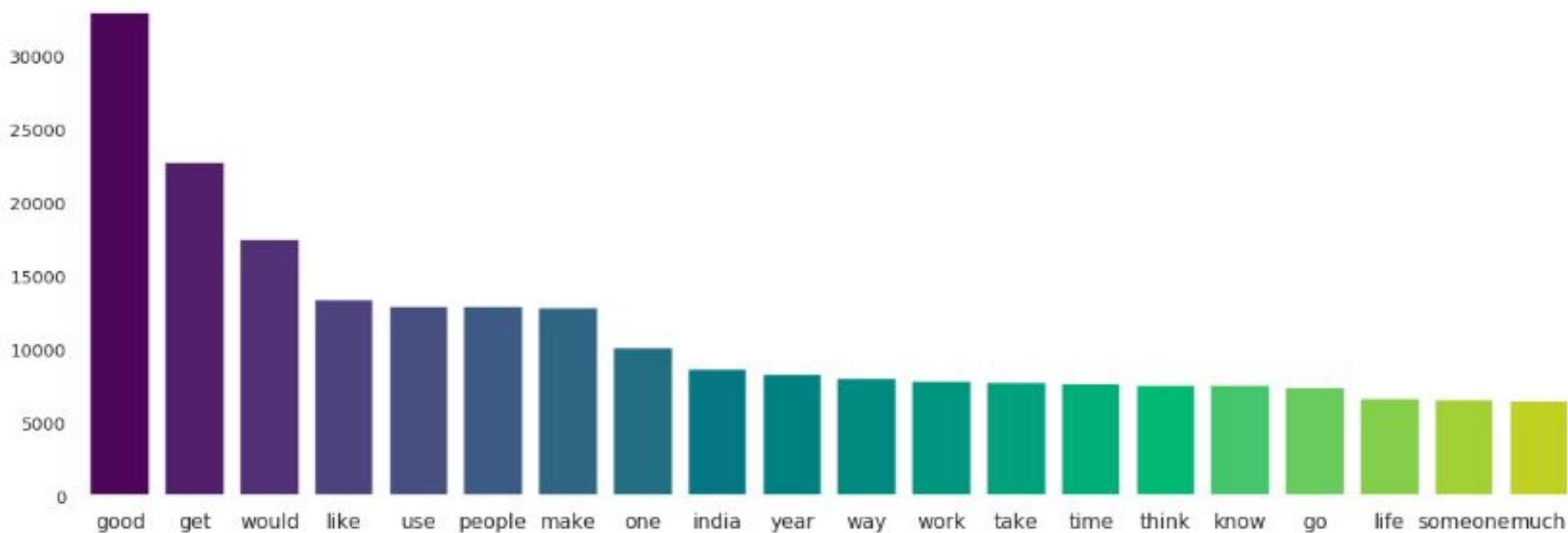
**Why do Chinese
hate Donald
Trump?**

**Do Americans that
travel to Iran have a
mental illness?**

Insincere Questions Common Words



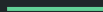
Sincere Questions Common Words



Models

Baseline Models

- Logistic Regression
- Naive Bayes
- XGBoost
- Voting Classifier



0.483

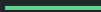
Baseline Ensemble with Downsampling F1 Score

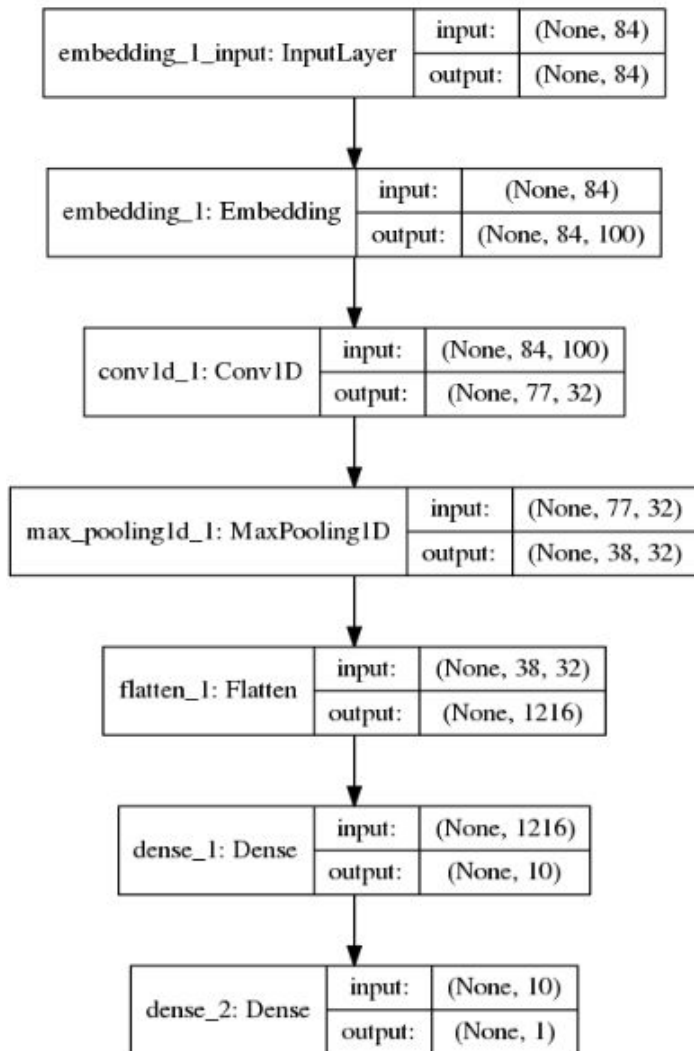
0.170

Baseline Ensemble with Upsampling F1 Score

Deep Learning Models

- Convolutional Neural Network
 - ◆ Glove Word Embedding
- Long Short Term Memory Network
 - ◆ Glove Word Embedding





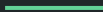
CNN Model

0.60

CNN Glove Embedding F1 Score

LSTM Model

```
-----  
Layer (type)                Output Shape                Param #  
-----  
input_1 (InputLayer)        (None, 114)                 0  
-----  
embedding_1 (Embedding)     (None, 114, 300)           54828600  
-----  
bidirectional_1 (Bidirection (None, 114, 256)           440320  
-----  
bidirectional_2 (Bidirection (None, 114, 128)           164864  
-----  
flatten_1 (Flatten)         (None, 14592)              0  
-----  
dense_1 (Dense)             (None, 64)                 933952  
-----  
dense_2 (Dense)             (None, 1)                  65  
-----  
Total params: 56,367,801  
Trainable params: 56,367,801  
Non-trainable params: 0  
-----
```



0.65

LSTM + Glove Embeddings F1 score

Challenges

Imbalanced Dataset

Solution: Resampling Techniques

The dataset is highly imbalanced, with only 6% of samples belonging to the target (insincere) class

Maximizing recall, or true positive rate, presents a difficulty here due to the small number of insincere samples

Large Dataset

Solution: Subsampling Data and
Model Optimization

The training data has over 1 million rows

Preventing memory errors and excessive training times presents challenges

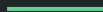
Production and Beyond

Production Environment

- Evaluate new questions as submitted
 - Sincere questions can be posted immediately
 - Insincere questions can be withdrawn by the user or submitted to an administrator for approval
 - The training set will be updated with new questions as they're asked
-

Future Model Improvements

- Data Augmentation
- Attention Layer
- Additional Pre-Trained Embeddings
- Topic Modeling as input to Neural Network



Questions?
