Insincere Question Detection
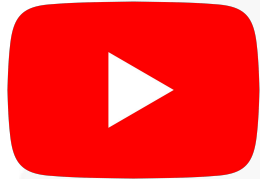
Aaron Reich
Kavin Krishnan
Thomas Hu

- Toxic Content Detection is an essential task for platforms
  - Toxic content can heavily influence perception of Brands
- Q & A Platform **Quora** faces this task in the form of classifying the <u>sincerity</u> of questions
  - Ideally questions should be legitimately requesting helpful answers
  - Example Insincere Questions:
    - "Why is European food so terrible?"
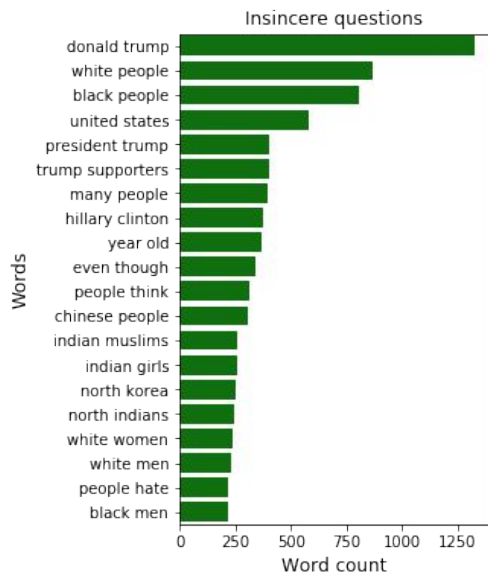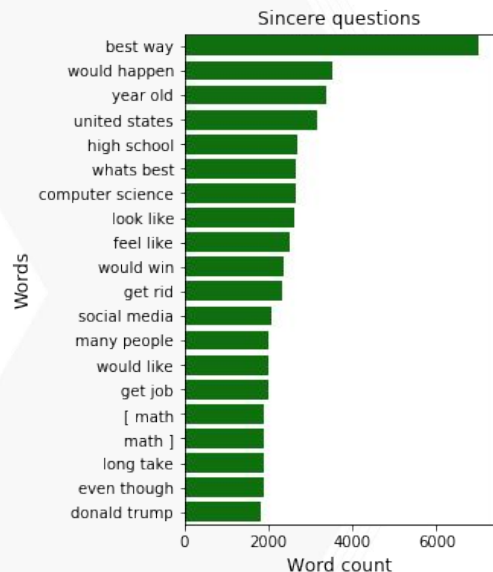   - "How is Trump so Incompetent at being President?"

Georgia
Tech
CREATING THE NEXT

Dataset

- Dataset comes from the "Insincere Questions Classification" Challenge hosted by Quora on Kaggle.com
- Consisted of CSV files for Training and Testing containing over 1.3 Million Questions for classification
  - Only minimal cleaning was required

Data Analysis

Sincere Questions
1,225,312

Insincere Questions
80,810

Vocabulary Size (Tokens)
508,823

Georgia
Tech
CREATING THE NEXT

Baselines:

- Support Vector Machine Model with GloVe Embeddings
- Bi-LSTM with Attention Model with GloVe Embeddings

Proposed Approach:

- BERT Model
- DistilBERT Model

Georgia Tech
CREATING THE NEXT
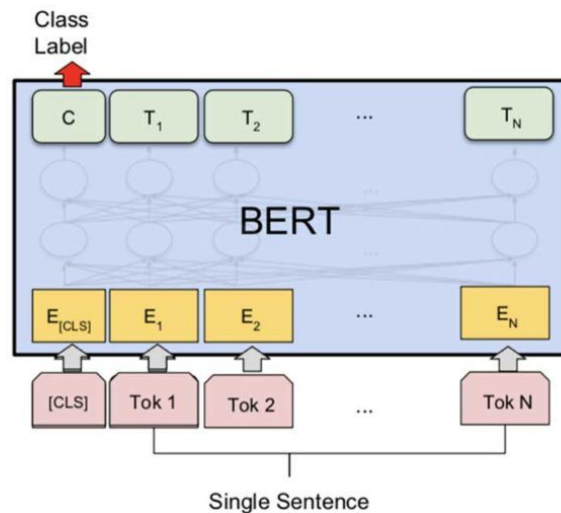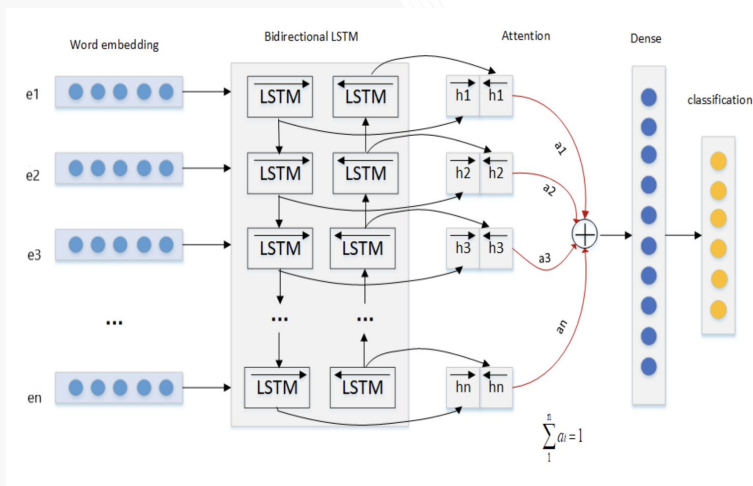
Experiment

- Google Colab - Tesla P100 16 GB GPU and 38GB of RAM
- Training Set- 75%, Test Set- 25%

- SVM - 10-Fold Cross Validation
- Validation Set - 10% of Training Set
- F1 Score, Precision, and Recall
- BERT Model -2 Epochs
- DistilBert Model- 6 Epochs

Results

Baseline and Proposed Method Results on Validation Set

Model	Cased Or Uncased	Precision	Recall	F1 Score
SVM	Cased	0.3134	0.7217	0.4373
Bi-LSTM	Uncased	0.6282	**0.7524**	0.6847
	Cased	0.6269	0.7178	0.6693
DistilBERT	Uncased	**0.7325**	0.6976	**0.7146**
	Cased	0.7264	0.6143	0.6657
BERT	Uncased	0.6849	0.6905	0.6877
	Cased	0.7314	0.6068	0.6633

Baseline and Proposed Method Results on Test Set

Model	Cased Or Uncased	Precision	Recall	F1 Score
SVM	Cased	0.3130	**0.7470**	0.4412
Bi-LSTM	Uncased	0.6152	0.7467	0.6746
	Cased	0.6235	0.7313	0.6731
DistilBERT	Uncased	0.7173	0.6745	**0.6953**
	Cased	0.7214	0.6092	0.6606
BERT	Uncased	0.6897	0.6782	0.6839
	Cased	**0.7272**	0.6176	0.6679

Future Work

Results for Models pretrained over BOOKCORPUS and Wikipedia for 125k steps

Model		Parameters	SQuAD1.1	SQuAD2.0	MNLI	SST-2	RACE	Avg	Speedup
BERT	base	108M	90.4/83.2	80.4/77.6	84.5	92.8	68.2	82.3	4.7x
	large	334M	92.2/85.5	85.0/82.2	86.6	93.0	73.9	85.2	1.0
ALBERT	base	12M	89.3/82.3	80.0/77.1	81.6	90.3	64.0	80.1	5.6x
	large	18M	90.6/83.9	82.3/79.4	83.5	91.7	68.5	82.4	1.7x
	xlarge	60M	92.5/86.1	86.1/83.1	86.4	92.4	74.8	85.5	0.6x
	xxlarge	235M	**94.1/88.3**	**88.1/85.1**	**88.0**	**95.2**	**82.3**	**88.7**	0.3x