



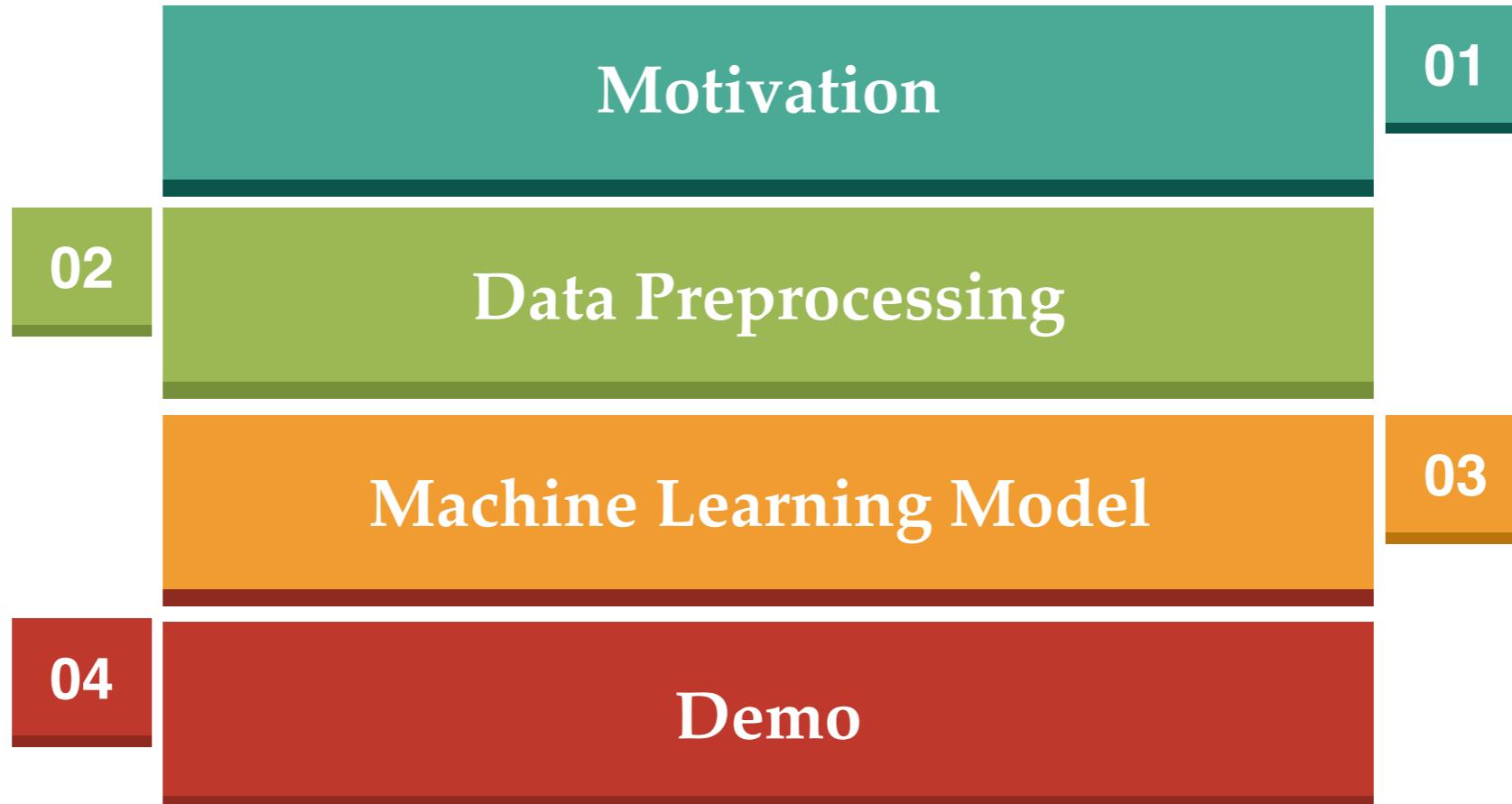
Comments Toxicity Detection

Dr. Baishali Dutta



Outline

2



Motivation

Life in 21st Century



Motivation

Negative Effects of Social Media

4

More children 'self-harming because of cyber-bullying'

New cyber-bullying weapon: Mobile phones

CCGS NEWS

'CYBERBULLYING RUINES LIVES'

Cyberbullying knows no age limit

Death by social media

Online abuse becoming part of kids' life

4 In 10 Parents Unable To Help; Schools Urged To Teach How To Tackle Cyberbullying

The images and text snippets highlight various ways that social media can negatively impact individuals, particularly young people, through cyberbullying, self-harm, and the broader issue of online abuse.



Missouri woman indicted in case involving MySpace-related suicide

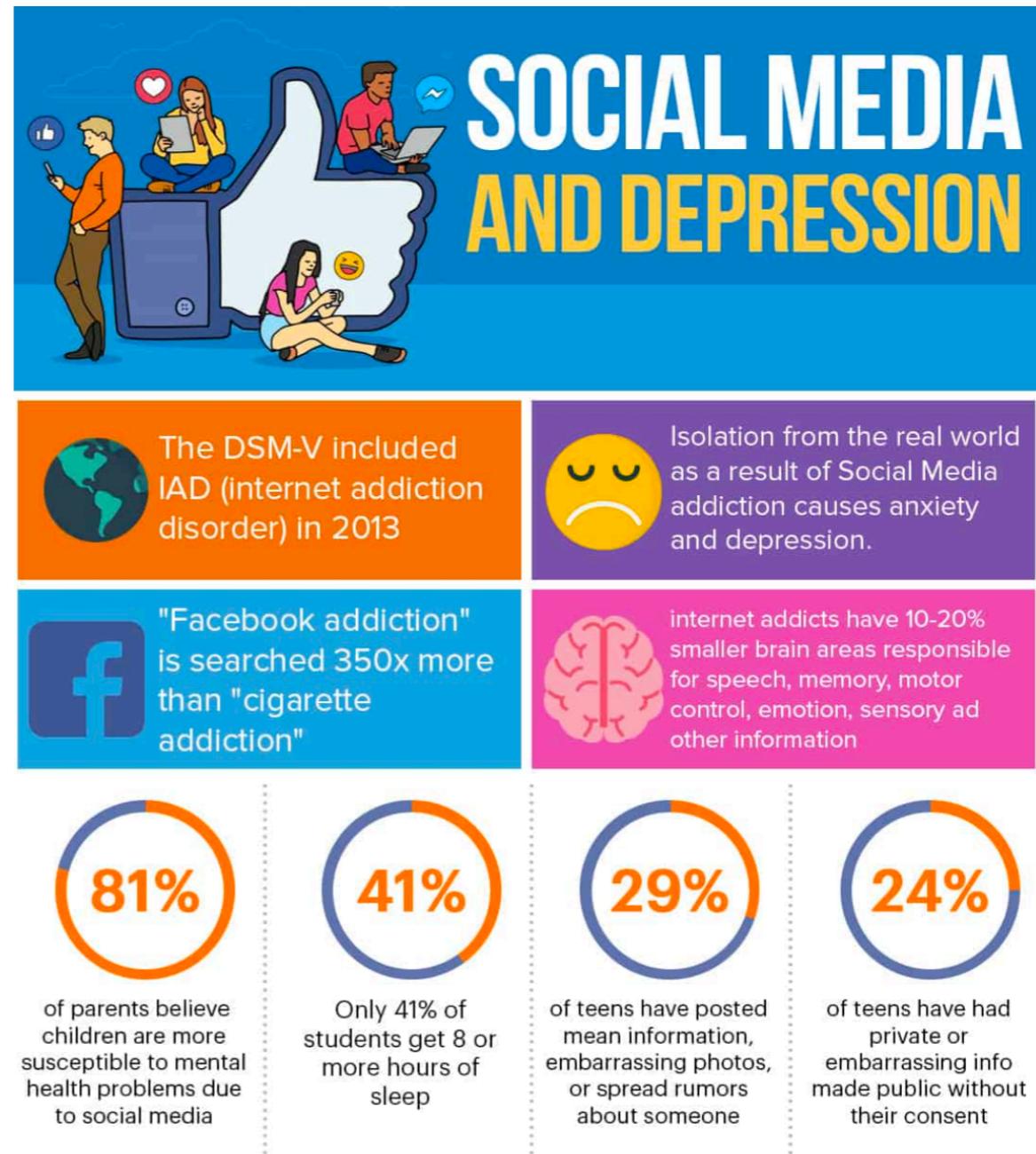
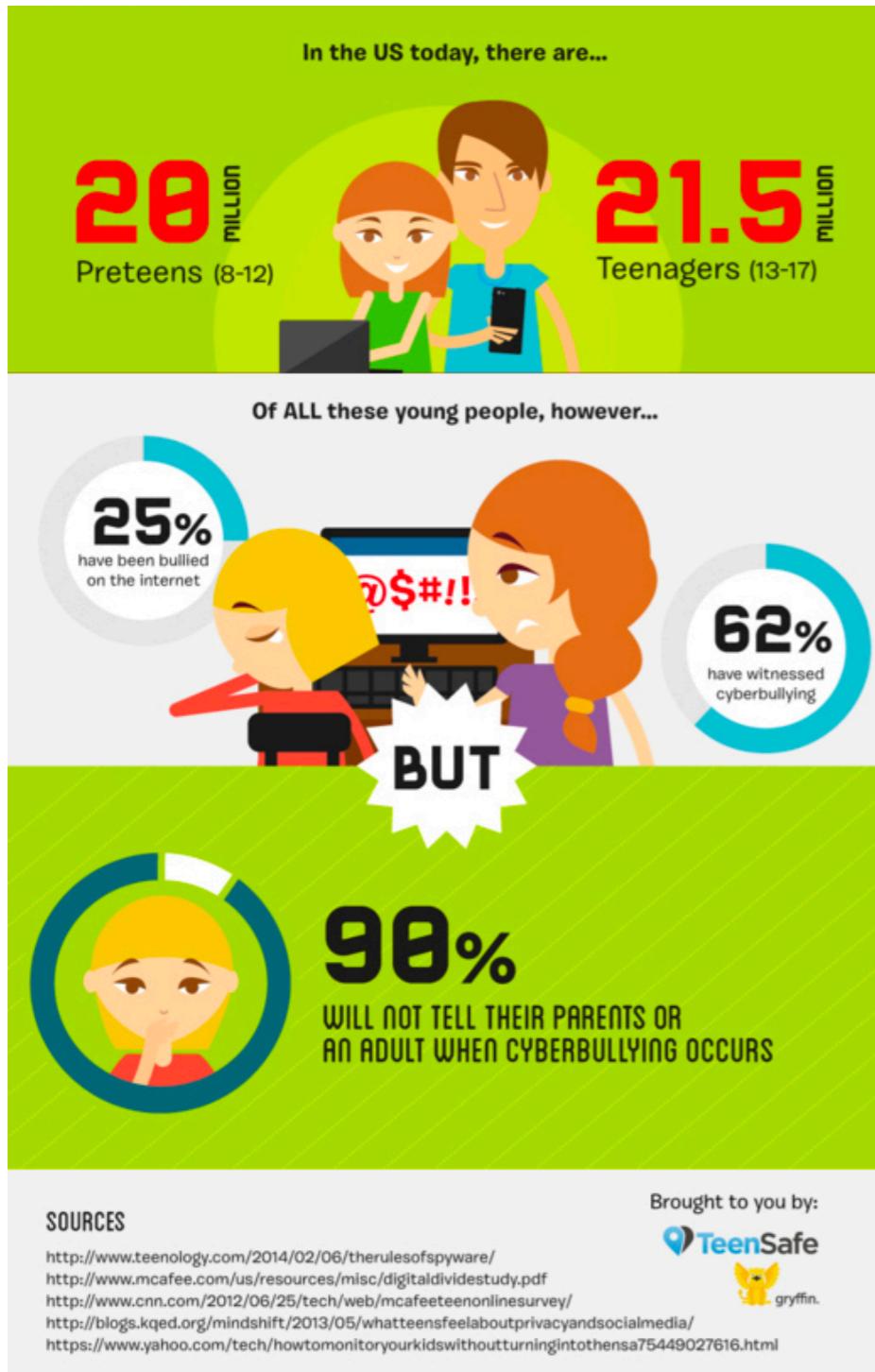
CYBER BULLYING AFFECTS REAL LIVES!

This section includes a photograph of a woman and several smaller portraits of other individuals, likely victims of cyberbullying.

Motivation

Surveys

5



Motivation

6

Prevent Cyberbullying

Reduce Toxicity



Motivation

Preventive Measures

7



STUDENTS TAKE A STAND AGAINST CYBERBULLYING

Mismerd Cry is the collective effort of young people, all of who are in their late teens and in the first leg of schooling, coming together to take a stand against Cyber Bullying.



POST. DON'T ROAST.

THINK BEFORE YOU TYPE.. STOP CYBER BULLYING!

DON'T BE MEAN, BEHIND THE SCENE.

46433559

JOIN OUR CAMPAIGN!
STOPCYBERBULLYING.COM

A graphic featuring a woman's hands clasped together. To the right, there is text for "Tikinagan Child & Family Services" and "RED ALERT, BULLYING HURTS!". Below this, it says "Bullying & Cyberbullying Campaign for First Nations Children & Youth". At the bottom, there are logos for "Our Partners" and several other organizations.

"SCARED TO SHOW MY face"

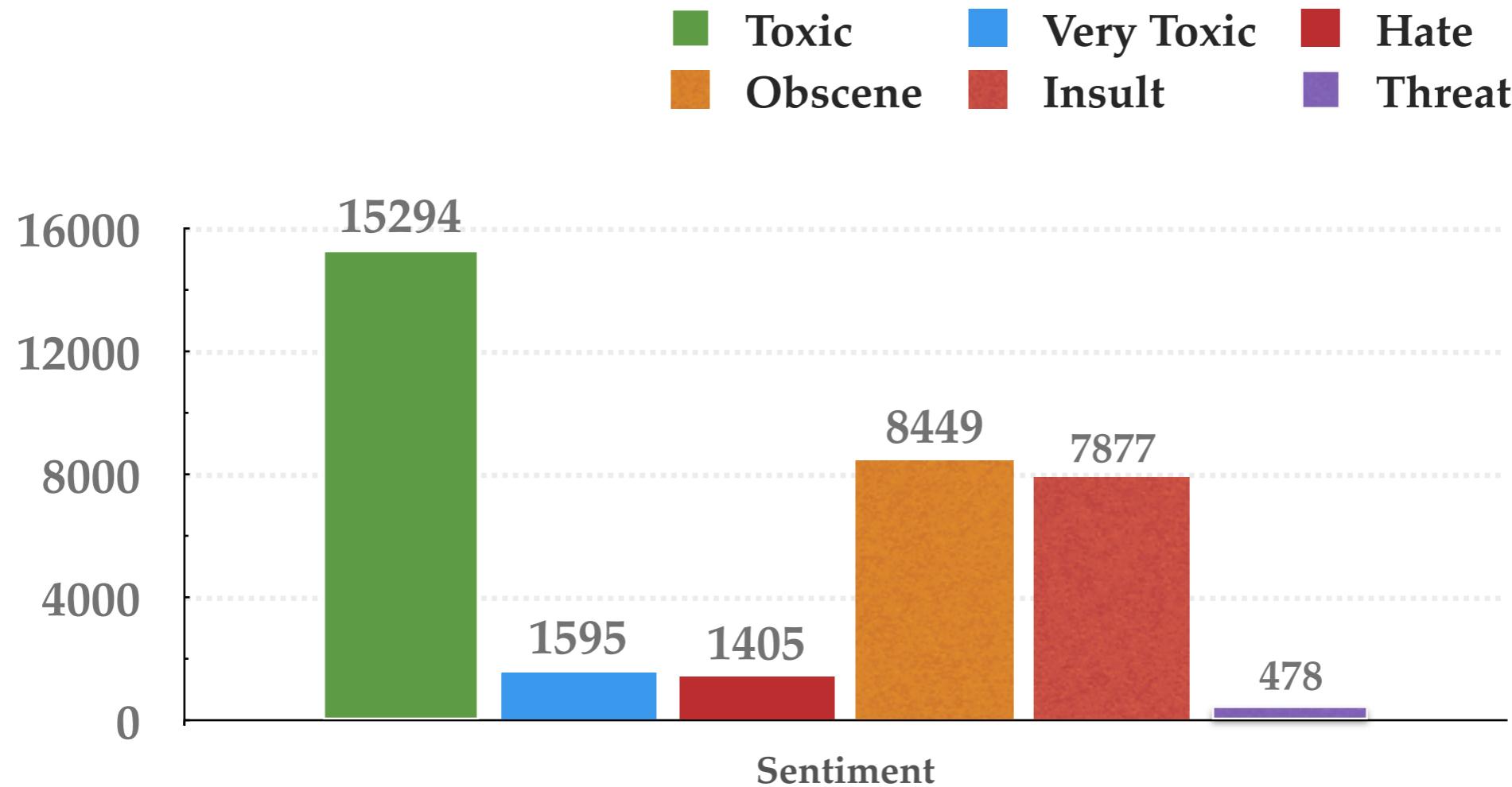
An illustration of a person sitting alone, looking sad, with the text "SCARED TO SHOW MY face" in large blue letters.

Quebec cyber-bullying, the social media is a place where most school children and teenagers become the victim of extremely anti-social behavior.
Stop Cyber Bullying. Use social media. Don't Abuse It.

act
anti-cyberbullying.ca

Exploratory Data Analysis (EDA)

8



- ✓ Training Data: Total: 159571 entries, Neutral: 124473 (none of the toxicity labels assigned)
- ✓ Huge unbalanced dataset, no null entries and empty strings present
- ✓ Multi-Label Classification Problem, target labels are not mutually exclusive, i.e. More than one right answer

Data Preprocessing

9

Lower Case All Words

- ▶ lower, camel or upper case written words treated as a same word
- ▶ all, All, ALL converted to all

Contraction Mapping

- ▶ Contracted words are expanded
- ▶ Example : aren't : are not, can't: cannot etc.

Fixed Misspelled Words

- ▶ Corrected with TextBlob library

Text Processing

Removed Punctuations

- ▶ . , ; : etc.
- ▶ Do not add any helpful information

Removed Emojis

- ▶ Removed emoticons, symbols, flags etc.
- ▶ Could be helpful for certain cases of sentiment analysis

Removed Stopwords

- ▶ *the, in, among, for, where* etc.
- ▶ Do not add any helpful information

Lemmatisation

- ▶ Finds the base form of words (spacy)
- ▶ Debatable in Sentiment Analysis as it can reaks parts-of-speech tagging and alters polarity of a word
- ▶ Reduces the word corpus

Data Preprocessing

10

Embedding Vectors

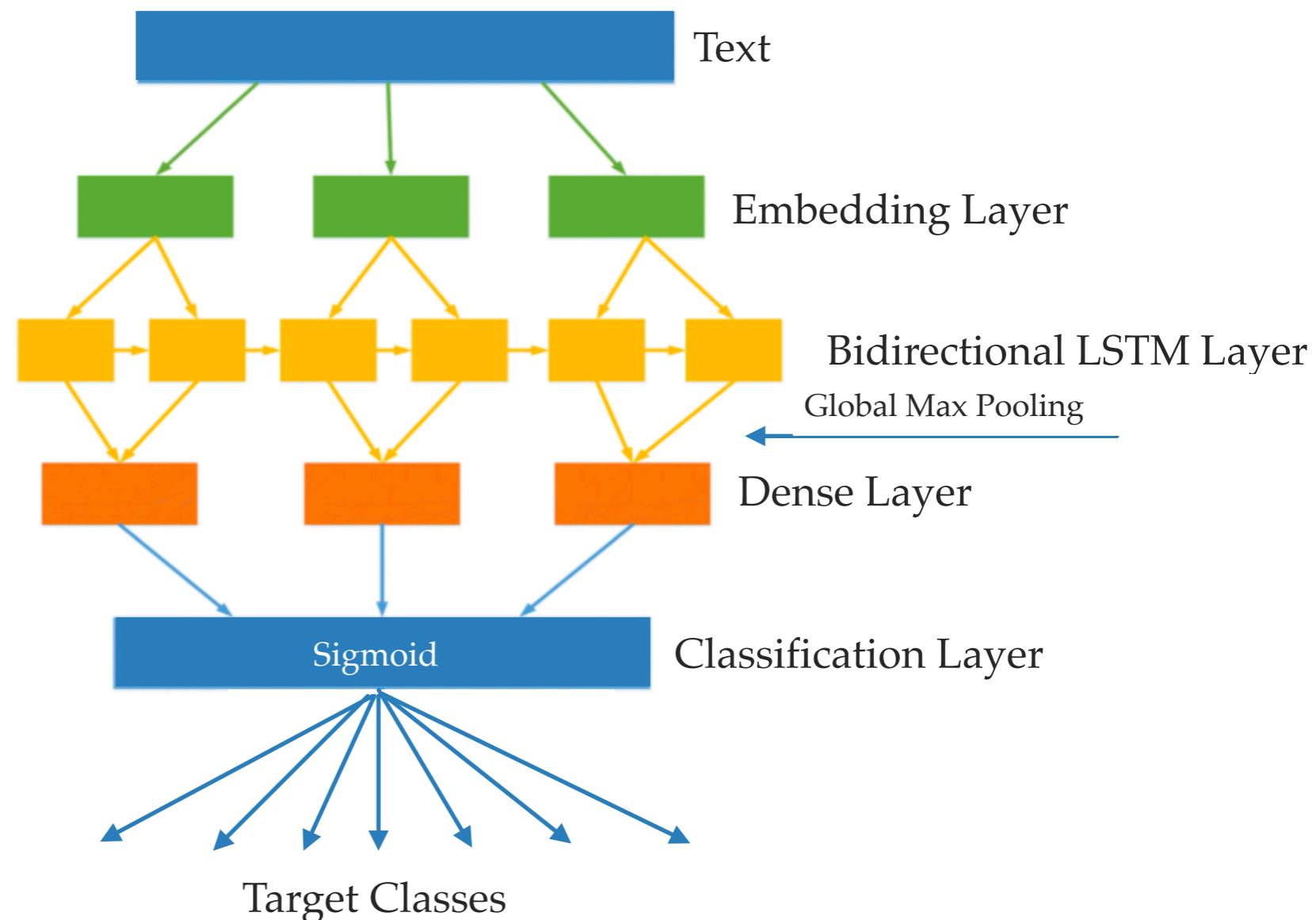
- ▶ Pre-trained word embedding vectors [Glove.6B](#) is used

Tokenizer

- ▶ Tokenizer (from keras) is used to tokenize the text with max vocab size of 20000
- ▶ Each word is assigned to the corresponding feature vector from the Glove.6B

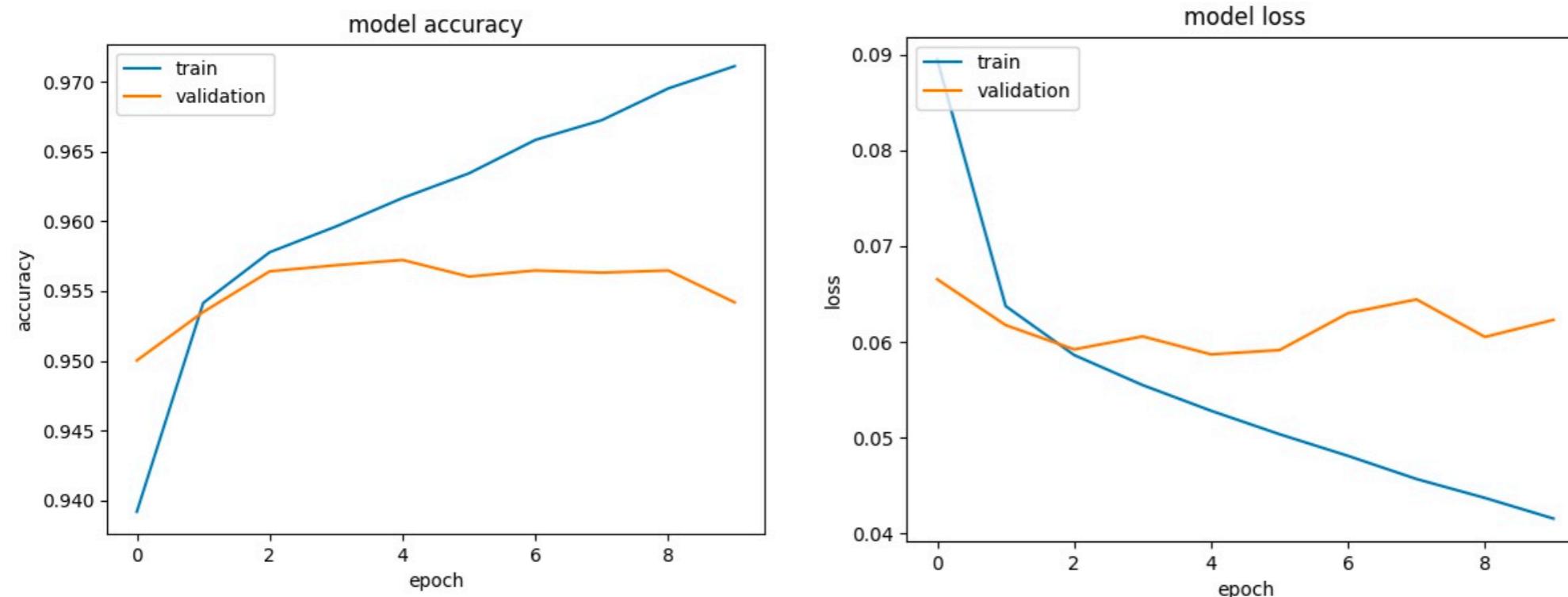
Deep Learning Model (Bi-directional LSTM)

- ✓ Recurrent Neural Network (RNN) is used as it was primarily built to tackle NLP problems (sequential data, sentences are sequence of words)
- ✓ In particular, Bidirectional Long Short-Term Memory (LSTM) RNN model is chosen
- ✓ LSTM can remember longer sequences than regular RNN
- ✓ Making it bidirectional, helps in a way that it can see at a given sequence both previous and next sequences in the text / sentence



Model Evaluation

- ✓ 20% of training dataset is kept for validation and rest used for training
- ✓ Accuracies on both training and validation sets and loss in each epoch (used early stopping monitoring validation loss)



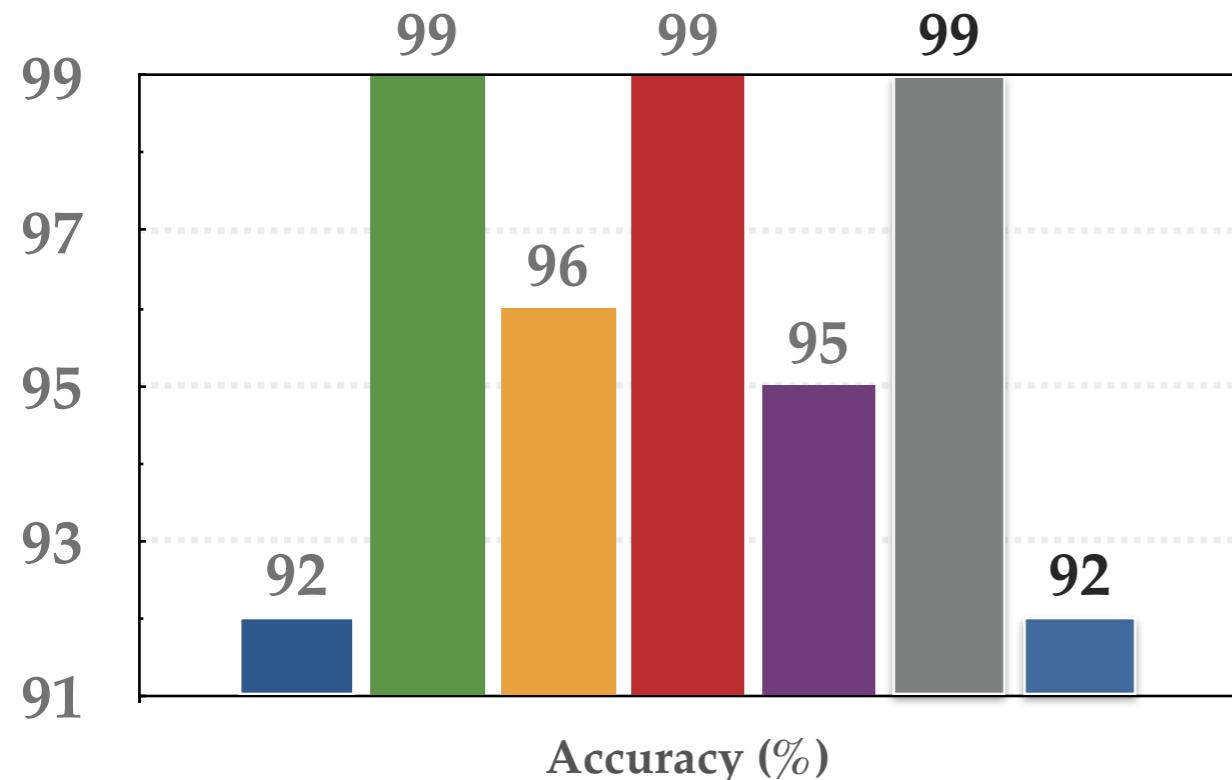
- ✓ Epoch = 4 is chosen as the best validation loss, patience = 5 indicates another 5 epochs that were run to confirm if the validation loss improves
- ✓ Next step is to evaluate the model on the test data (499 data points)

Model Evaluation

13

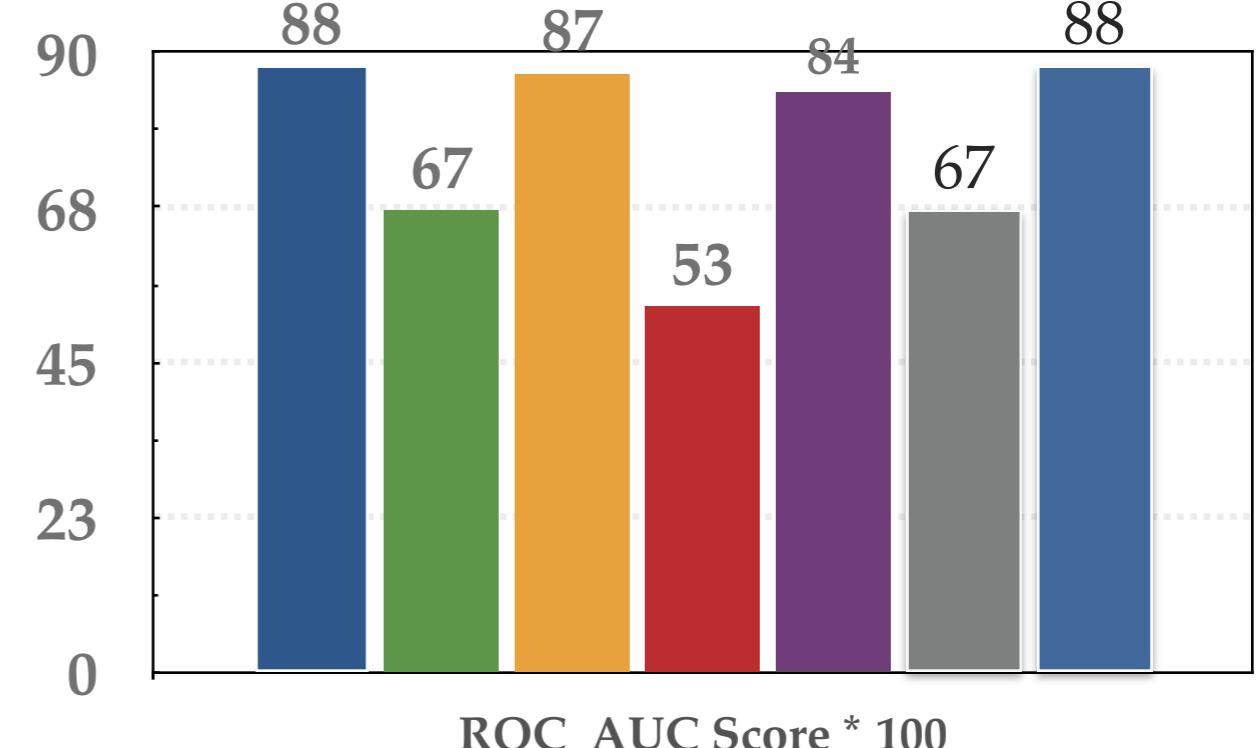


Accuracy (%) for different classes



Mean Accuracy: 96.3%

ROC_AUC Scores for different classes



Mean ROC_AUC Score: 0.76

DEMO