# Fake News Detection - Health Domain

Ojaswi Binnani 20161006
Abheet Sharma 20161091

# Why do we fall for it?

- Savvy fake news publishers target their content, knowing that when a story matches our own opinions, we tend to take it as true; our Confirmation Bias plays a pivotal role sometimes. “If a lie is telling you something you want to hear, you’ll more likely think it’s true” -Professor Sharon Kaye, John Carroll University. And if it’s something you “want to hear”, you will more likely than not share it.
- The formal structure and format of news articles tends to throw us off our guard. Using links and references(even though it is fake) tends to make it seem more believable, without us checking.
- Moreover, we are not bothered enough to fact-check EVERY news article we come across.

# Motivation

- Health Domain: Harder to fact-check. Requires very technical expertise sometimes. Fake news about health seems to be more pervasive and harder to weed out.
- Information—and misinformation—about your well-being is likely to feel more high-stakes than information about the business world, or celebrities, because Health in general tends to be very personal for many people.
- The scientific process takes a long time, which means new developments happen very slowly. Science takes a long time to answer questions; findings get refuted; the accepted wisdom changes. And some climate change deniers and vaccination skeptics have started sowing doubt by saying the science on these issues is unsettled, that there are still open questions we have to investigate.

# Motivation

- Fake News is rampant in health-related new articles
- Health Articles are harder to fact-check- requires technical expertise
- Information - and misinformation- about health and well-being is higher-stakes than information about business or celebrity because health in general tends to be more personal
- We observe that the misinformation in these sources is typically of the kind where scientific claims or content from social media are exaggerated or distilled either knowingly or maliciously (to attract eyeballs).

# Dataset

- Fake Health corpus - which has the original news article, the review of said article scrapped from HealthNewsReview.org, the Social Engagement/discussions on Twitter about the article, and the user network.
- Approximately 2100 news articles in the Health domain which have been reviewed by experts doctors according to 10 criterion.
- Articles that received less than 3 out of 5 are classified as fake, while the rest as true. We did a separate test with threshold set as less than 2

# Creating emotionized text representations

- NRC-emotion-intensity-lexicon
    - Data is in a three tuple format (w, e, s)
    - W is word
    - E is emotion associated with the word
    - S in the intensity of the emotion
- Every word in a document and its emotion (if it is above a specified threshold. In our project it is 0.6) is added to create a D' so emotion representation is included in our model

# Model1

- Trained two SVMs, one with and one without emotional cognizance as a feature. We prove that emotional cognizance of the article body does help the accuracy of a model.

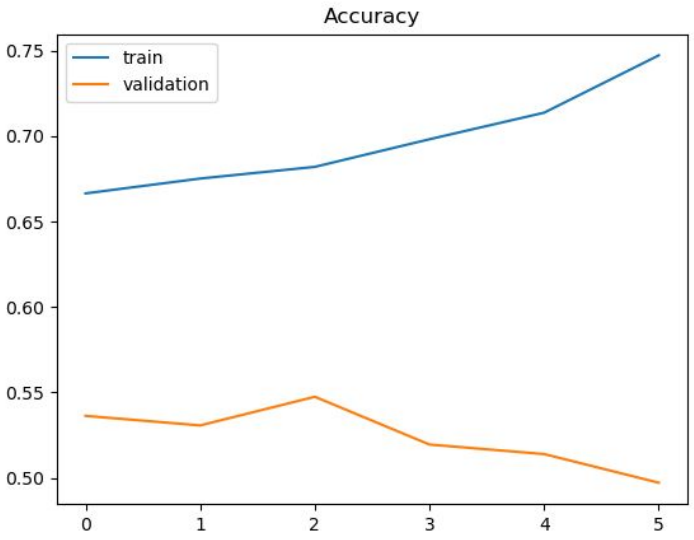
Threshold = 2: Accuracy: 91.3% | 90.37%
Threshold = 3: Accuracy: 62.43% | 60.74%

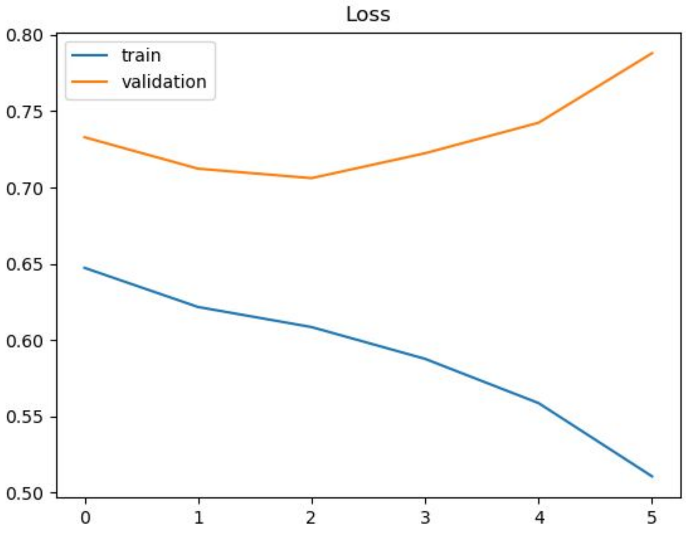
# Model 2

Base LSTM with dropout pretrained on Glove weights

```
model = Sequential()
model.add(Embedding(vocabulary_size, 100, input_length=MAX_SEQUENCE_LENGTH, weights=[embedding_matrix], trainable=True))
model.add(SpatialDropout1D(0.2))
model.add(LSTM(100, dropout=0.2, recurrent_dropout=0.2))
model.add(Dense(64, activation='relu'))
model.add(Dense(2, activation='softmax'))
model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])
```

# Statistics


Accuracy
train
validation
0.75
0.70
0.65
0.60
0.55
0.50
0
1
2
3
4
5


Loss
train
validation
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0
1
2
3
4
5

# Threshold = 2

Training Accuracy: 91.18% | 90.12%
Validation Accuracy: 85.47% | 83.77%
Testing Accuracy: 90.47% | 87.10%

# Threshold = 3

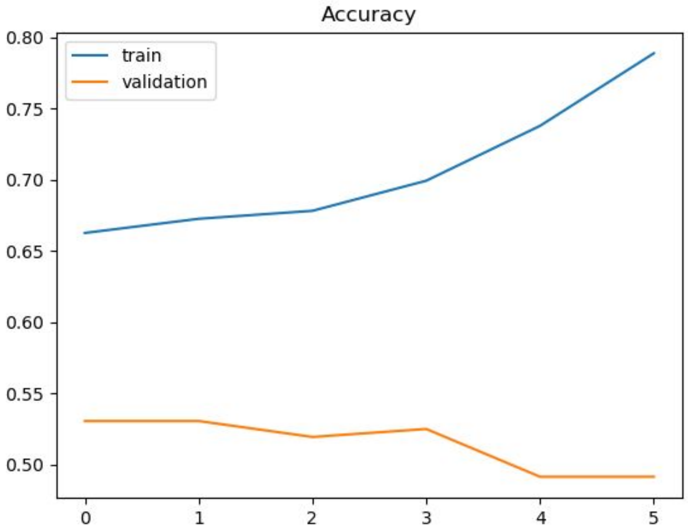
Training Accuracy: 78.84% | 77.13%
Validation Accuracy: 62.31% | 58.59%
Testing Accuracy: 63.4% | 57.82%

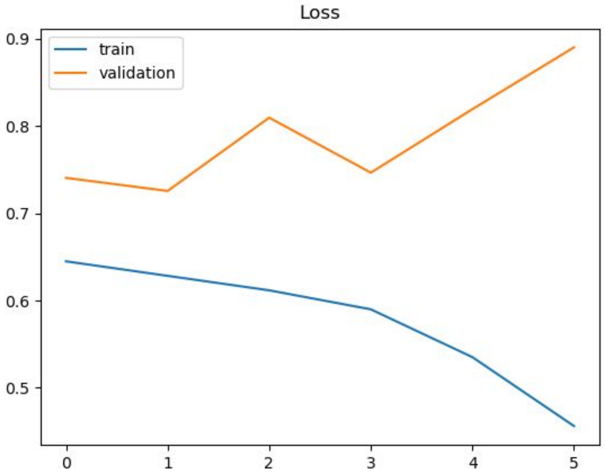
Epochs: 6
No. of Training Data: 1610
No. of Testing Data: 448

# Model 3

BiDirectional LSTM pretrained on Glove weights along with dropout was implemented to see if we can boost accuracies.

```
model = Sequential()
model.add(Embedding(vocabulary_size, 100, input_length=MAX_SEQUENCE_LENGTH, weights=[embedding_matrix], trair
model.add(SpatialDropout1D(0.2))
model.add(Bidirectional(LSTM(100, dropout=0.2, recurrent_dropout=0.2, return_sequences=True)))
model.add(Bidirectional(LSTM(50, dropout=0.2, recurrent_dropout=0.2)))
model.add(Dense(64, activation='relu'))
model.add(Dense(2, activation='softmax'))
model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])
```


Accuracy
0.80
0.75
0.70
0.65
0.60
0.55
0.50
train
validation
0
1
2
3
4
5


Loss
0.9
0.8
0.7
0.6
0.5
train
validation
0
1
2
3
4
5

## Threshold = 2

Training Accuracy: 92.44% | 91.76%
Validation Accuracy: 84.83% | 84.2%
Testing Accuracy: 91.68% | 89.03%

## Threshold = 3

Training Accuracy: 79.26% | 77.64%
Validation Accuracy: 62.37% | 56.41%
Testing Accuracy: 64.04% | 58.74%

Epochs: 6
No. of Training Data: 1610
No. of Testing Data: 448

# Additional

- In our dataset, tweets about various articles were given
- We assume that the majority of people tweeting their opinion on a news article while share a common emotion. If many people are suspicious, then possibly the news article is fake
- By using the tweets of a certain article (if they exist and sufficient discussion surrounds the article) we can use a sort of human evaluation to add data into our model

# Results with human evaluation

| Model | Threshold =  2 | Threshold = 3 |
|---|---|---|
| LSTM | 90.47 | 63.4 |
| BiLSTM | 91.68 | 64.04 |
| LSTM + SA | 90.56 | 63.27 |
| BiLSTM + SA | 91.53 | 63.92 |
| LSTM + Human Tweets | 92.15 | 64.8 |
| BiLSTM + Human Tweets | 92.31 | **65.47** |
| LSTM + Human Tweets + SA | 92.24 | 64.73 |
| BiLSTM + Human Tweets + SA | **92.45** | 65.62 |

# Interpretation

- Emotion Cognizance proved to help improve detection of fake news.
- Sentiment Analysis does not seem to play a role in helping accuracies.
- Analysis and mining emotion in user tweets has improved accuracies and detection of fake news.
- Emotion Cognizance is widely applicable and can help other fake news detection models in other domains.
- User reviews tend to have a lot of emotion embedded in the text. It represents how personal our Health is to us.

# Future Work

- Incorporate individual criterion scores (C1-C8) into the training model as well.
- Each news article has its own review in our dataset. We could try to match the emotions in the full fledged review to that of the user reviews directly for better detection.
- Utilize deep neural structures to fully take advantage of our dataset. The base idea "Emotion Cognizance" is widely applicable.
- There are other metadata marked with each news article such as Links, references etc.

# Thank you!

Our video presentation is on the link:
https://web.microsoftstream.com/video/1852242b-7fda-4915-9cd7-e068d3119a6e

The time stamp for our presentation is 30:46