The Moon is made of green cheese!

Screening for fake news using Journalism ethics.

Introduction

- There is a rise of false claims on the Web.
- These claims often have political, or financial agenda.
- What has been done so far to curb fake news?
  1. Manual verification
  2. Automated detection

How to screen for fake news in the absence of evidence or counter-evidence?

Problem Statement

Related work

Paper	Fact Checking	Linguistics Features	Source Credibility
Singh et al. (2018)	Y	Y	Y
Popat et al. (2017)	Y	Y	Y
Rashkin et al. (2017)		Y	
Potthast et al. (2017)		Y	

Solution Approach

- For a given news article, quantify how much it aligns with the principles of Ethical Journalism.

- Project-relevant principles of Journalism:
  1. Presentation of multiple viewpoints.
  2. Independence from political, financial, or cultural affiliation.
  3. Truth and accuracy.

Solution Approach

- These principles form the basis of features to be extracted from the news articles. (First stage). The features include the following:
  1. Propaganda Score (violates principles 1 and 2).
  2. Click bait Score (violates principle 2).
  3. Emotional language Score (violates principle 1).

- The features extracted are the input into a binary classifier. (Second stage)

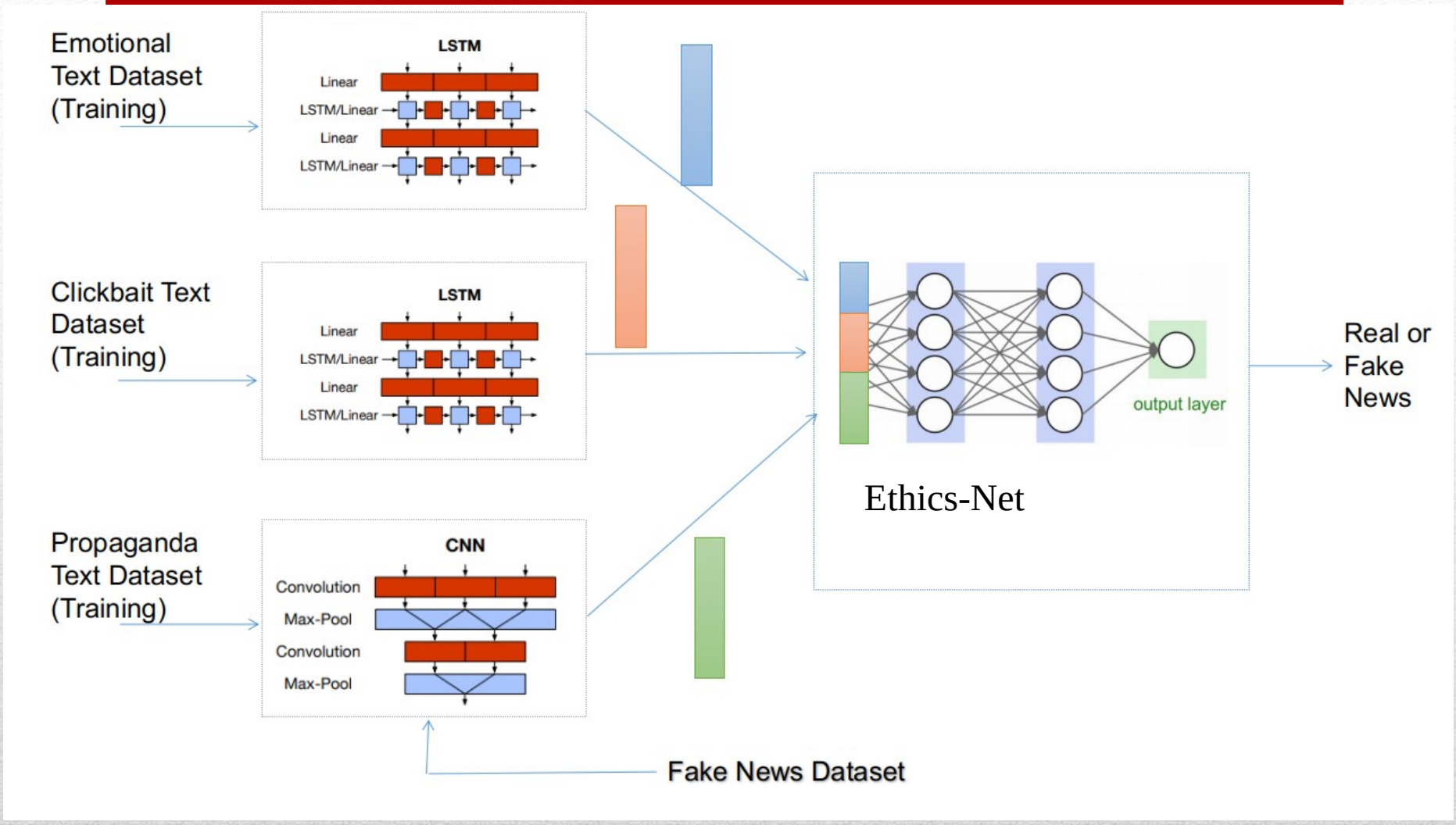
Solution Approach

Dataset	Number of articles
Clickbait Data	32,000
Propaganda Data	20,000
CrowdFlower Emotion in Text	40,000
Fake News Corpus	20,000

Datasets

Journalistic objectivity can help determine the veracity of news articles

Experiment Hypothesis



Solution Sketch

Emotion Detection

- **Aim**
  - Classification of a given piece of text to be **emotional** or **neutral**
- **Architecture**
  - Bi-LSTM + Scaled dot product Attention
- **Data pre-processing**
  - Replaced urls with <URL> and user names with <USER> tokens
  - Spacy for word tokenization
  - Bucketing strategy

Emotion Detection

- **Initial Attempts**

Number of classes	Model	Accuracy
13	GRU	~21
8	Bi-LSTM + Glove	~33

Emotion Detection

- **Training**

Architecture	Accuracy
Bi-LSTM + Glove	63.2
Bi-LSTM + Glove + Attention	77.6

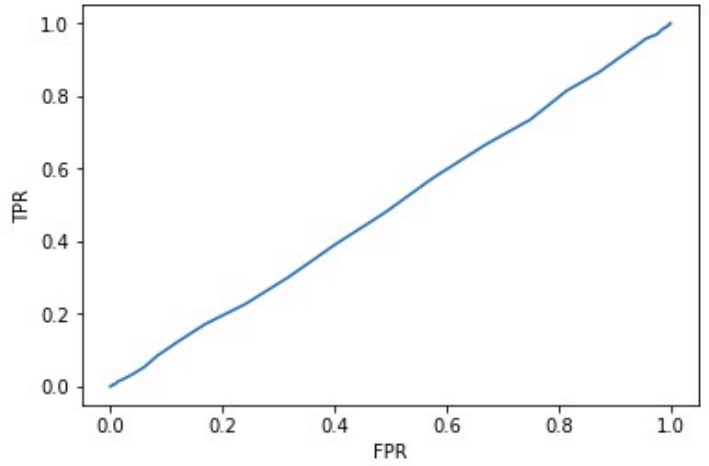
- **Metrics**

Metric	Score
Precision	0.78
Recall	0.91
F1-score	0.84

Performance of the model on out-of-domain dataset.

```
In [162]:
1  plt.plot(tpr,fpr)
2  plt.xlabel('FPR')
3  plt.ylabel('TPR')
```

```
Text(0, 0.5, 'TPR')
```



Emotion Detection

Propaganda Detection

Overview

- Analysed linguistic patterns across the news articles by characterizing the content using various lexical resources.
- Tokenize the text into a 3-D matrix whose rows are the sentences and columns, words.
- Each word is represented as a 71-D vector encoding membership in various lexicons.
- Model architecture is a Convolutional Neural Network to detect indicative n-grams.

Lexicons used

1. Linguistic Inquiry and Word Count (LIWC), a lexicon widely used in social science studies.
2. Subjective words often used to dramatize or sensationalize a news story.
3. Hedge words that indicate vague or obscuring language.
4. Intensifying lexicons used to enliven and attract readers.

Propaganda Detection

Model architecture and result

The hyperparameters of the best performing CNN model are:

- 16 kernels
- Kernel size of 1x3
- 3 convolution layers
- 1 max pool layer
- 1 dropout layer (rate of 0.5)

The model achieved an accuracy of **90%** in its in-domain test set.

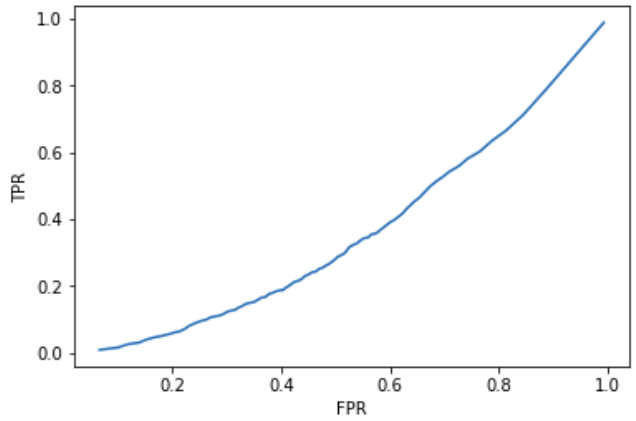
Propaganda Detection

Propaganda Detection

Performance of the model on out-of-domain dataset.

```
In [164]:
1  plt.plot(tpr,fpr)
2  plt.xlabel('FPR')
3  plt.ylabel('TPR')
```

```
Text(0, 0.5, 'TPR')
```

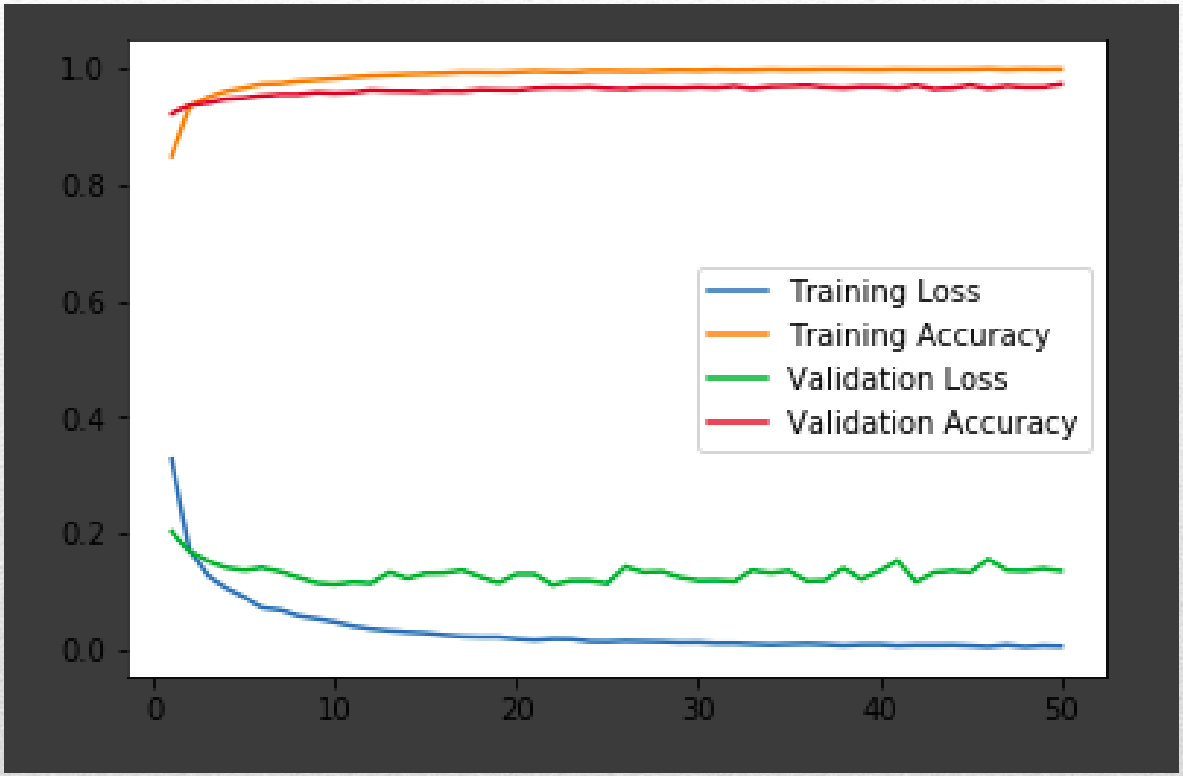


Clickbait Detection

- Clickbait detection was done by using Recurrent NNs specifically Bi-LSMTs
- **Data Pre-Processing**
  - Word tokenization - Spacy tokenizer
  - Unknown words - <UNK> in the vocabulary
  - Bucketing strategy - prepare batches with similar lengths so that minimal padding is added
- **Training**
  - Various methodologies with various hyperparameters were searched
  - Model with the best accuracy and F1-Score was chosen

	Loss	Accuracy
RNN	19	93
LSTM	15.5	94
Bi-LSTM	15	94.5
LSTM + Glove	14	95
Bi-LSTM +Glove	**11.0**	**96.84**

Training and Validation Accuracy over the epochs
To get better results, dropout is applied and number of LSTM layers =2



Clickbait Detection

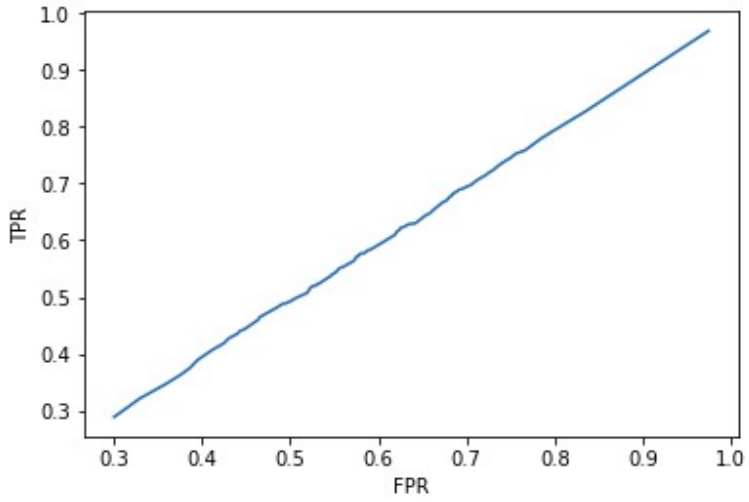
```
In [20]: 1 predict_clickbait(model, "A school girl gave her lunch to a homeless man")

0.9992517828941345

In [21]: 1 predict_clickbait(model, "This enormous predator was recently uncovered in Patagonia")

0.0599878653883934

In [32]: 1 predict_clickbait(model, "Big predator found in Patagonia, you won't believe the last pictur

0.7575799226760864

In [31]: 1 predict_clickbait(model, "You won't believe how these 9 shocking clickbaits work")

0.9999940395355225
```

Clickbait Detection

AUC-ROC Curve : Using clickbait probabilities to predict Fake News. Shows one feature not enough.

```
In [157]:
1  plt.plot(tpr,fpr)
2  plt.xlabel('FPR')
3  plt.ylabel('TPR')
```

```
Text(0, 0.5, 'TPR')
```

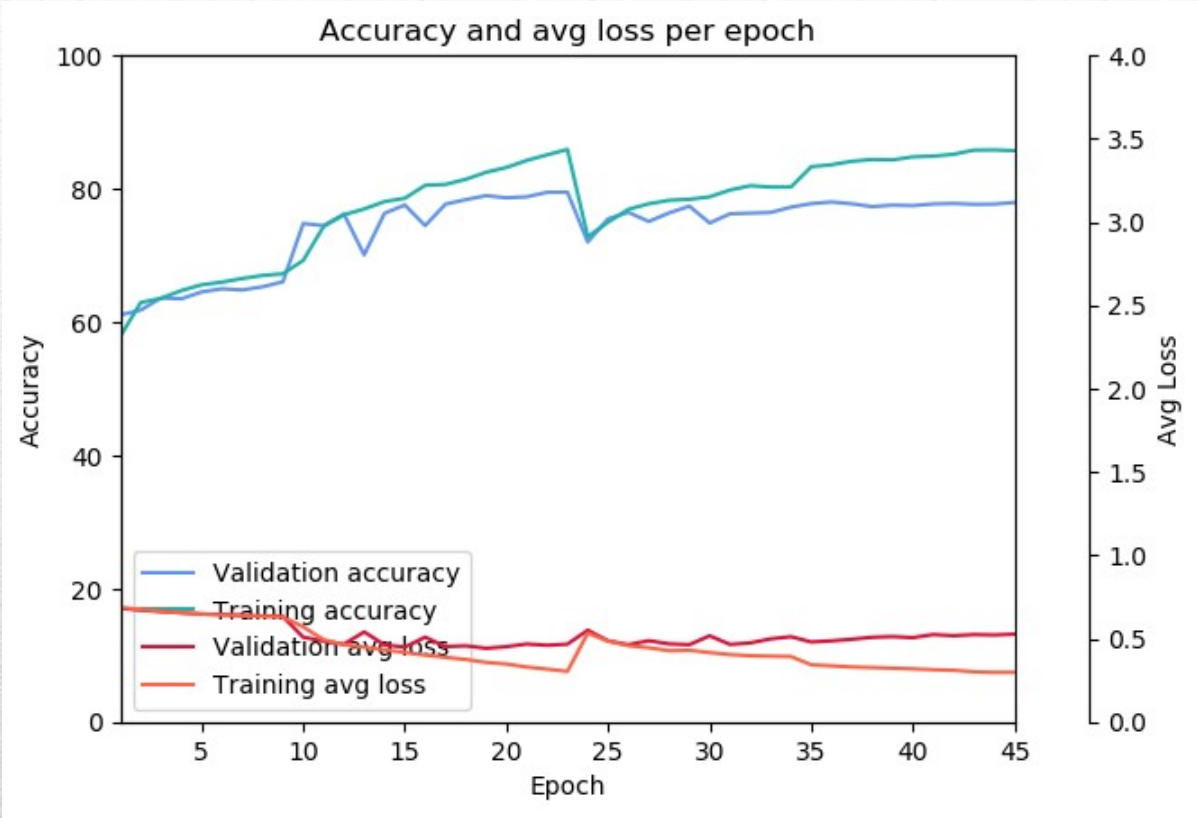


Clickbait Detection

- After the individual models are trained, they are frozen
- A hybrid model - is created that takes as inputs feature vectors from individual models
- This hybrid model is a neural network with two hidden layers and all layers are fully connected
- Output of the hybrid model is one value which says whether news is fake or not
- Based on the ground truth, weights of ONLY this model are trained - to maintain interpretability

Combining them all

Accuracy / Loss of Ethics-Net over epochs



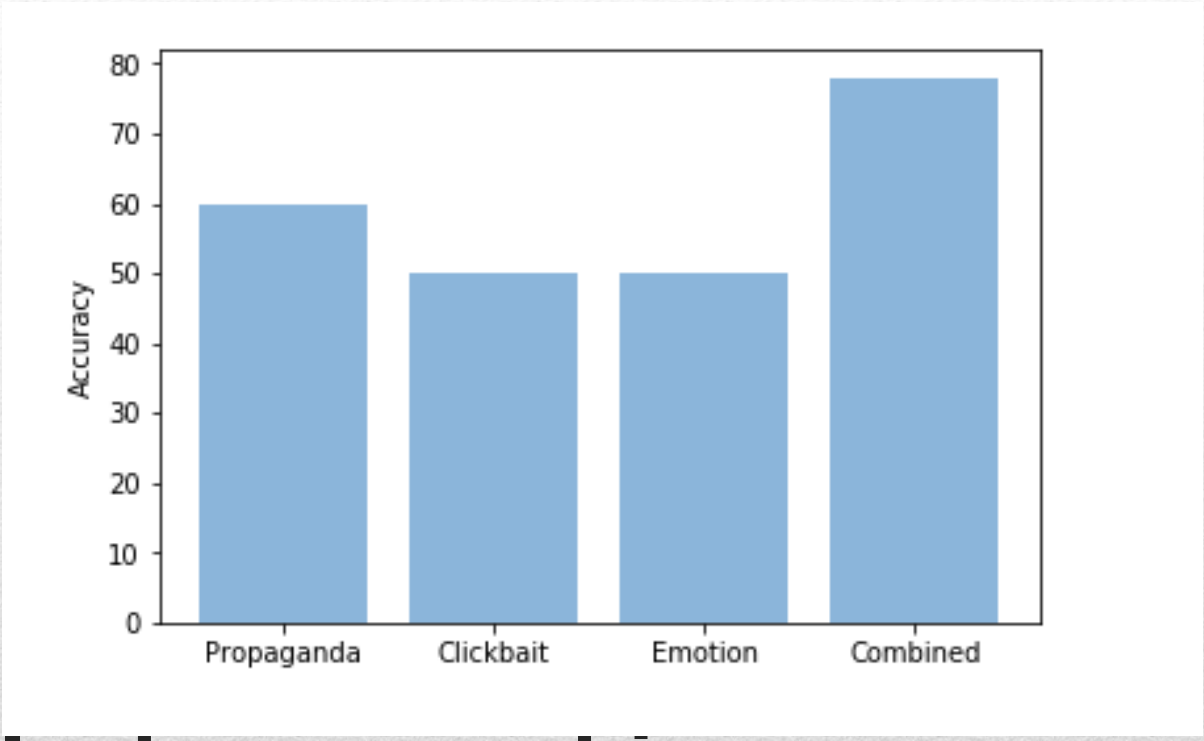
Evaluation Metrics

- Precision:80.722%
- Recall:76%
- F1-Score:78.2%
- Accuracy:78.9%
- Confusion Matrix

	Predicted Fake News	Predicted Real News
Actual Fake News	1520	480
Actual Real News	363	1637

Evaluation Metrics

- Below: Accuracy of each of the models+combined model for detecting Fake News
- Shows that combining features is better



Interpreting Results

Conclusion

- Feature detectors are trained over completely different dataset and tested on a different dataset with different objective
- Although Transfer Learning is difficult in textual domain, we acheived an accuracy of ~78%
- Thus we see combining journalistic feature detectors is a promising way of Fake News Detection