

Sarcasm Detection with NLP and Deep Learning

Elsa Scola Martín | Master in Big Data Analytics | July 2020

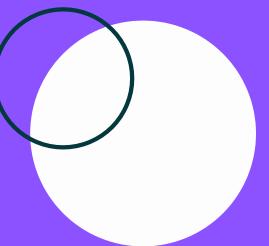
Advisor: Isabel Segura Bedmar

WHAT IS SARCASM?

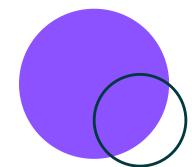
99

Saying the opposite of what is meant in order to hurt someone's feelings or to criticize something in a humorous way.

“

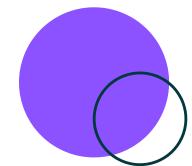


Why it's important to detect sarcasm



EXTRACT COSTUMER OPINION

"the speaker is so good that I can hear sounds from other galaxies! Totally worth the money"



IDENTIFY ABUSIVE CONTENT

A screenshot of a tweet from Justine Sacco (@JustineSacco). The tweet reads: "Going to Africa. Hope I don't get AIDS. Just kidding. I'm white!" It was posted on 12/20/13 at 10:19 AM from Hillingdon, London. There is a blue verification checkmark icon next to her name.

Her account got deactivated.

It was sarcasm.

Objectives

LEARN

the fundamentals neural networks and some deep learning models (LSTM and BERT).

REVIEW

the main approaches and datasets for sarcasm detection in text.

DEVELOP

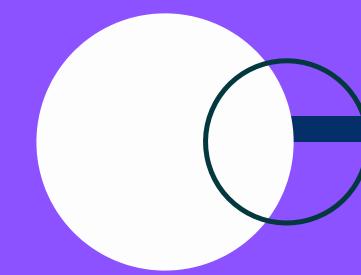
train and evaluate two deep learning architectures (LSTM and BERT) for the sarcasm detection task using the News Headlines dataset.

ANALYZE

both models's performance and compare it to the current state of the art.

Perform an error analysis of the models to identify their weaknesses and future lines of work.

State of the Art



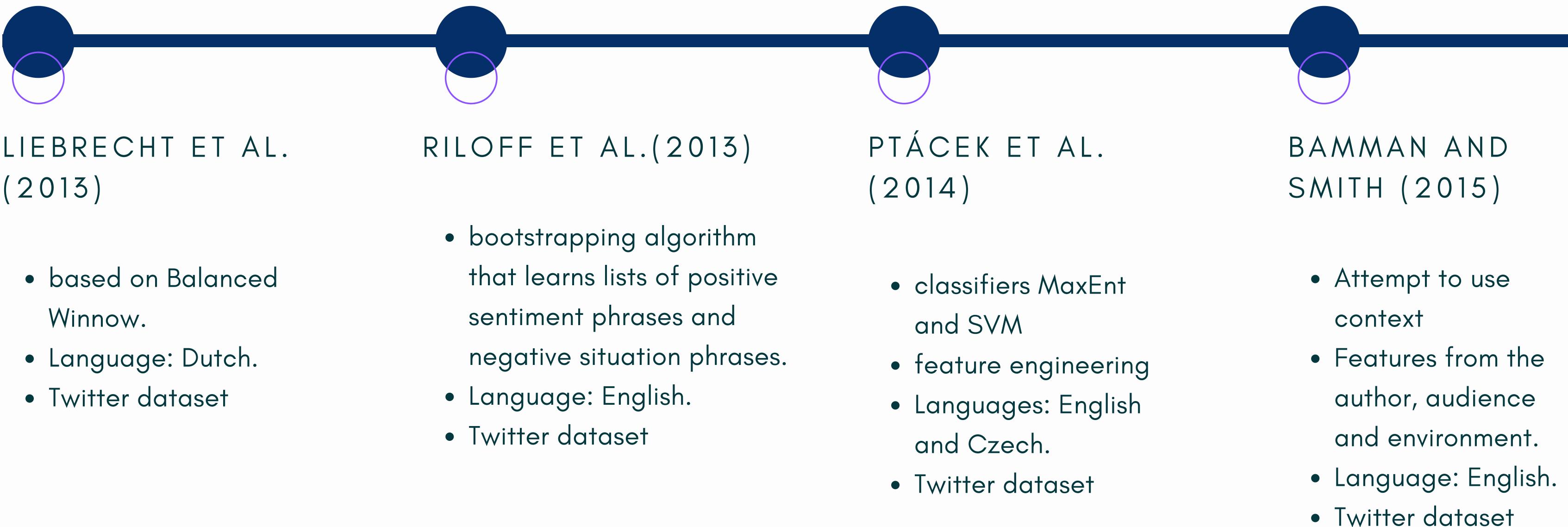
**TRADITIONAL MACHINE
LEARNING**

(2013-2015)

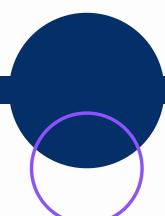
**DEEP LEARNING
APPROACHES**

(2016-2020)

TRADITIONAL MACHINE LEARNING (2013-2015)

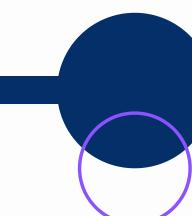


DEEP LEARNING APPROACHES (2016-...)



AMIR ET AL.(2016)

- Avoid laborious feature engineering by only using the text of previous posts of the users to create the user embeddings.
- CNN to extract relevant features.
- Language: English.
- Twitter dataset



GHOSH ET AL.
(2017)

- Focuses on conversational context.
- Conditional LSTM and LSTM with sentence level attention.
- Language: English.
- Discussion Forum data and Twitter dataset.



MISRA AND
ARORA (2019)

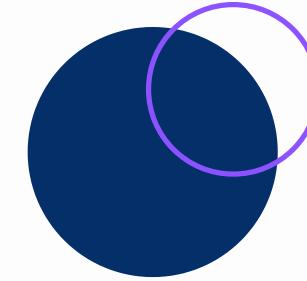
- Hybrid neural network based on Bi-LSTM.
- Language: English.
- News Headlines dataset.
 - TheOnion
 - HuffPost



KHATRI ET AL. (2020)

- Traditional algorithms are used but embeddings are created using BERT and GloVe.
- Classifiers: SVM, Logistic Regression, Gaussian Naive Bayes and Random Forest.
- Language: English.
- Best result: GloVe + Logistic Regression.
- Twitter dataset.

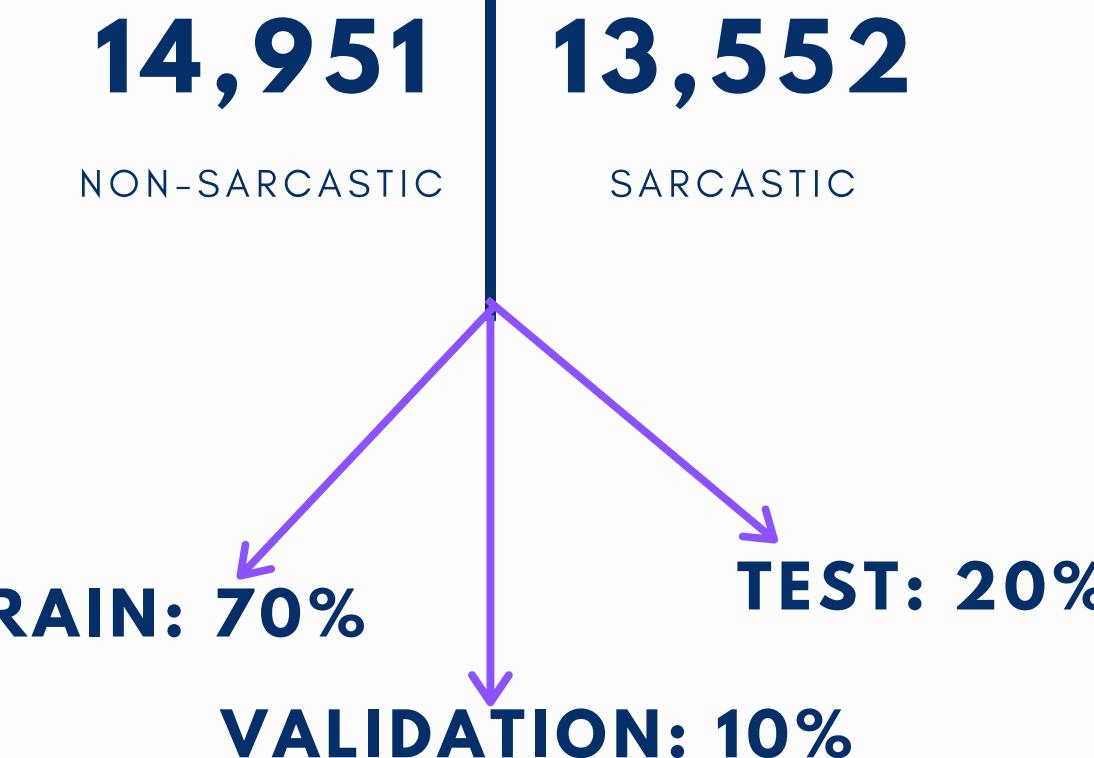
NEWS HEADLINES DATASET



Sarcastic instances from:



28,503 HEADLINES



Non-sarcastic instances from:



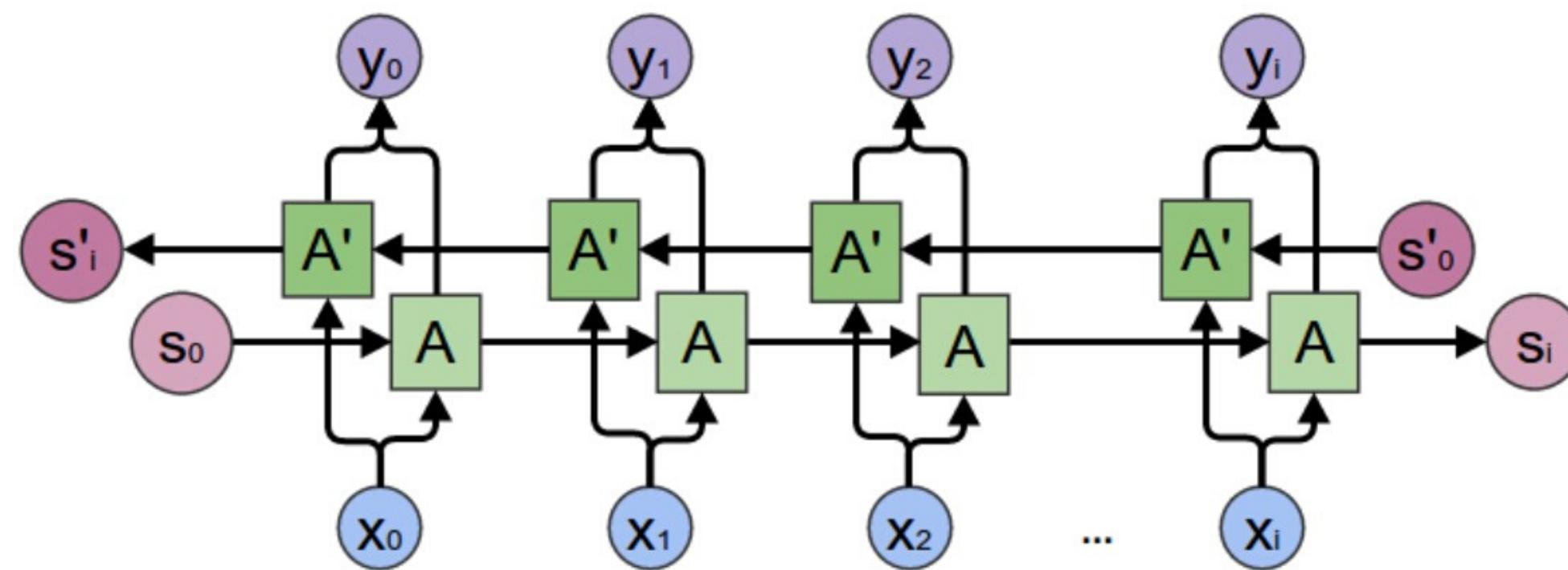
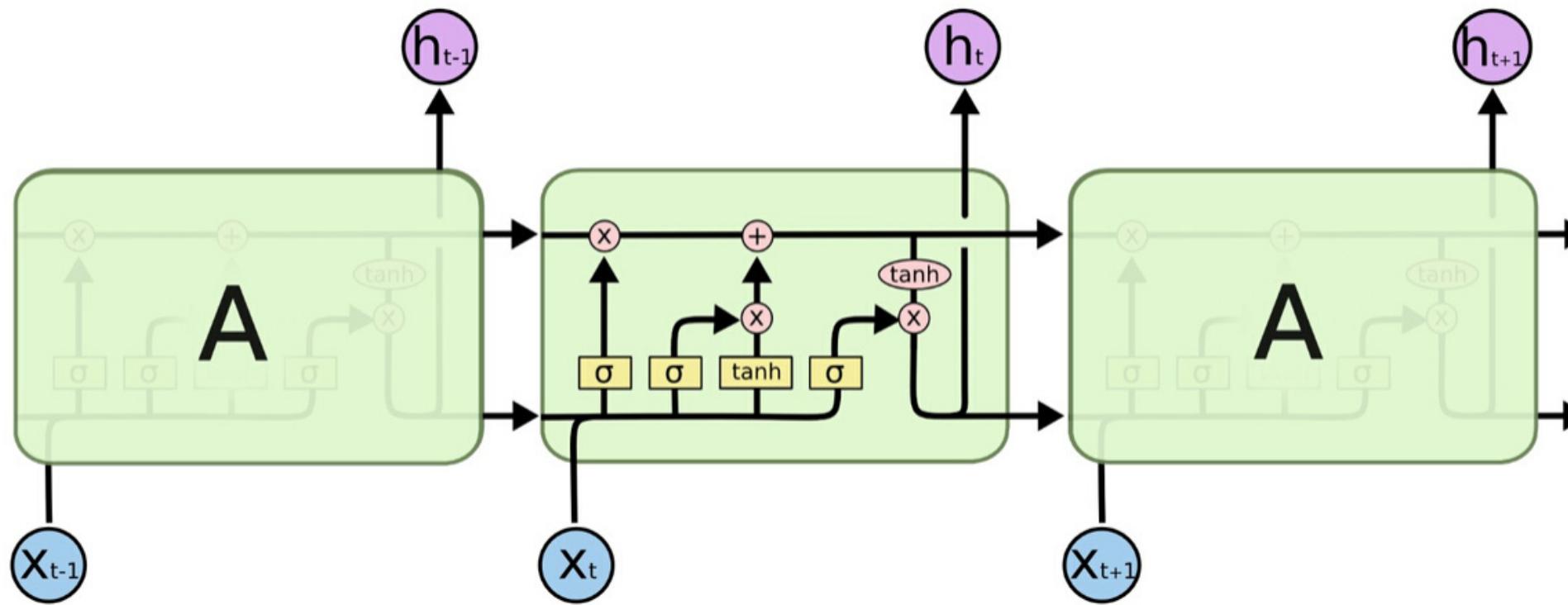
(all with balanced classes)

OUR PROPOSALS

BI-LSTM

BERT

Bi-LSTM



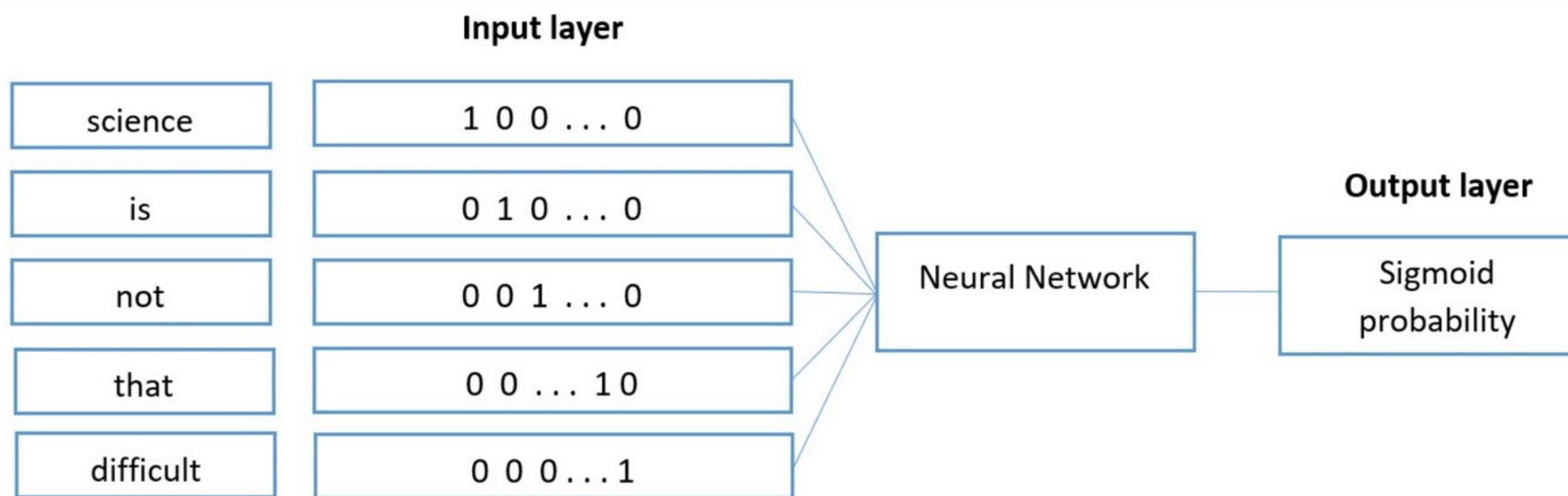
Bi-LSTM

GloVe pre-trained word vectors: [glove.6B.200d](#)

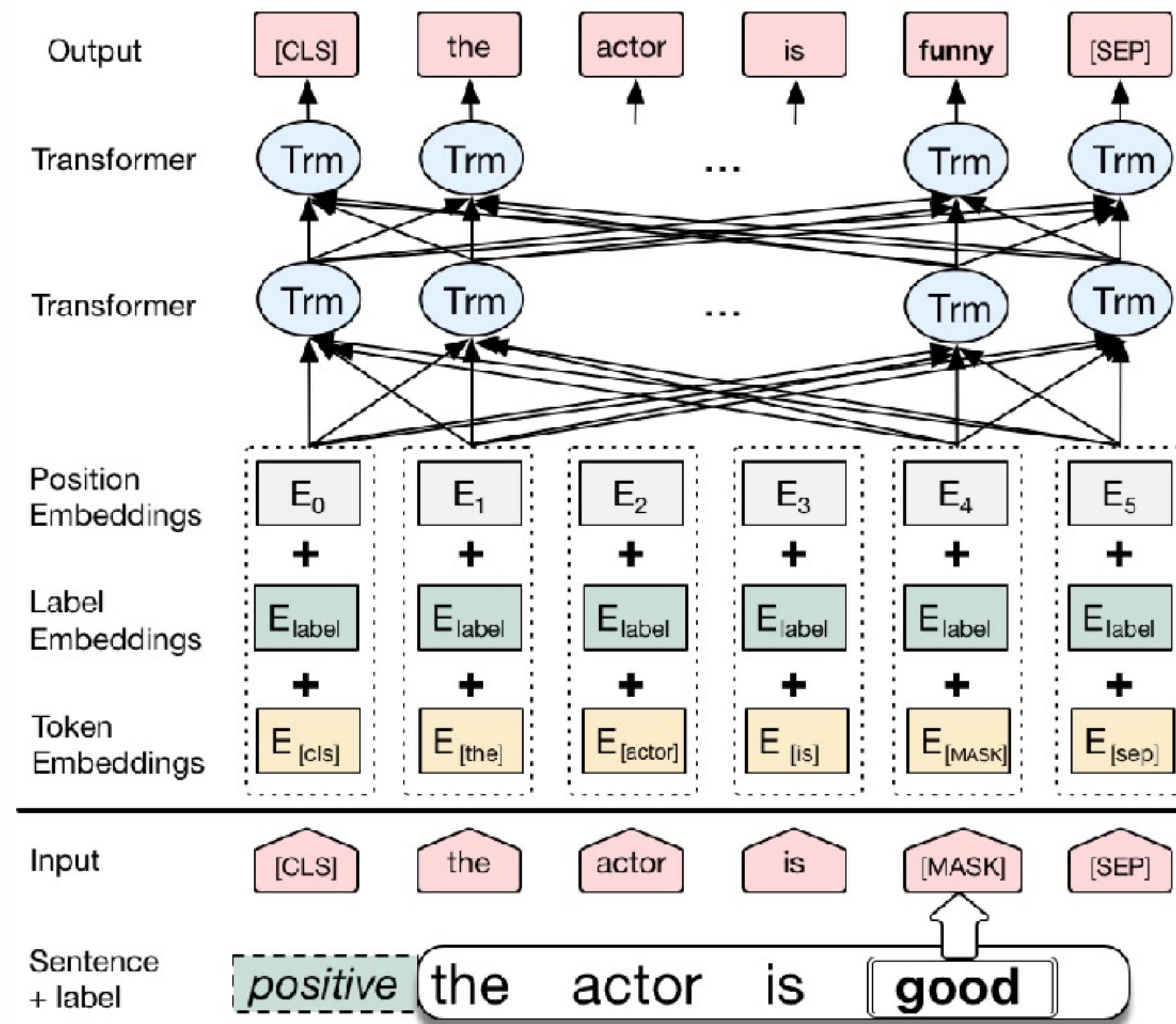


Wikipedia 2014 + Gigaword 5 corpora

6 Billion words with each word being a 200 dimensional vector.



BERT



RESULTS

F-SCORE

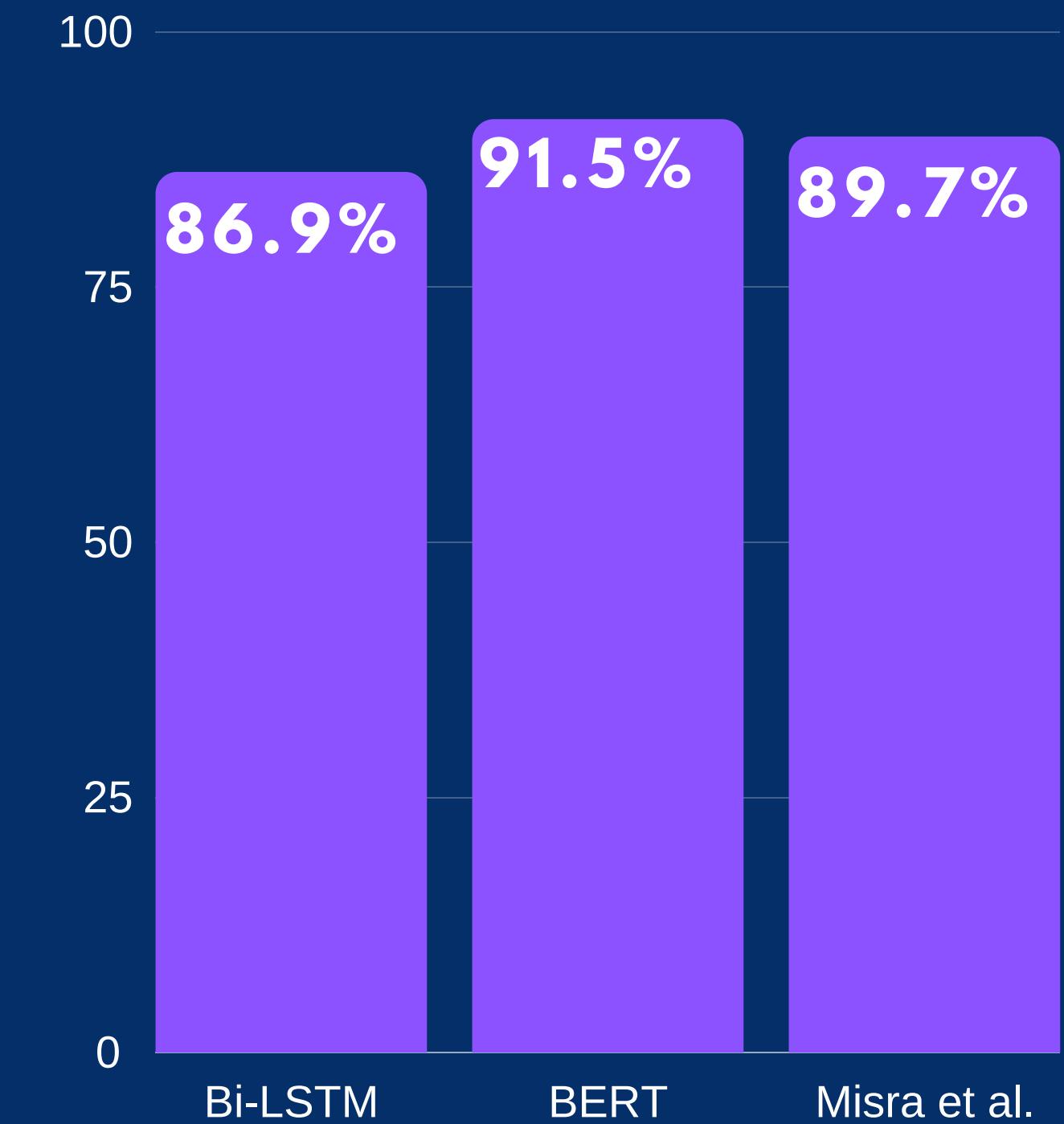
86.2%

Bi-LSTM

90.9%

BERT

ACCURACY %



$$accuracy = \frac{TP + TN}{TP + FN + TN + FP} \quad Recall = \frac{TP}{TP + FN} \quad Precision = \frac{TP}{TP + FP} \quad F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

ERROR ANALYSIS

FALSE POSITIVE

***"man apparently opens beer with butt,
inspires bartenders everywhere"***

The screenshot shows a news article from The Huffington Post. The header includes the site's logo, navigation links for NEWS, CORONAVIRUS, POLITICS, 2020 ELECTIONS, ENTERTAINMENT, LIFE, PERSONAL, VIDEO, and SHOPPING, and options to Log In or Join HuffPost Plus. The article is categorized under WEIRD NEWS and was published on 01/26/2015 at 01:29 pm ET, last updated on Jan 26, 2015. The main title is "Man Appears To Open A Beer With His Butt, Inspires Bartenders Everywhere". It features a video thumbnail showing a shirtless man opening a beer bottle with his butt, with the caption "Yo Soy; 'El Destapador Humano'". Below the video are social sharing icons for LinkedIn, Facebook, Twitter, Pinterest, Email, and Print. The video player shows it's 0:14 / 0:41 long. To the right, there's a "TRENDING" sidebar with articles about Tucker Carlson, Hillary Clinton, Grant Imahara, and Naya Rivera.

WEIRD NEWS 01/26/2015 01:29 pm ET | Updated Jan 26, 2015

Man Appears To Open A Beer With His Butt, Inspires Bartenders Everywhere

By Ron Dicker

Yo Soy; "El Destapador Humano"

Ver más ta... Compartir

MÁS VÍDEOS

0:14 / 0:41

TRENDING

- Tucker Carlson Announces 'Vacation' After Top Writer's Racist, Sexist Posts Revealed
- Hillary Clinton: Trump Revealed The 1 Thing He's Fearful About With Roger Stone Clemency
- 'Mythbuster' Grant Imahara Dead At 49
- 'Glee' Star Naya Rivera's Body Found At Lake Piru In California

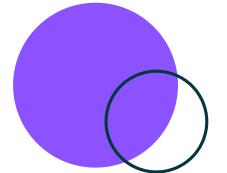
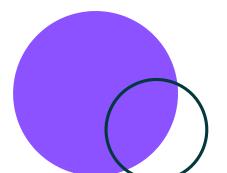
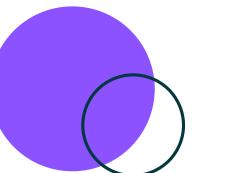
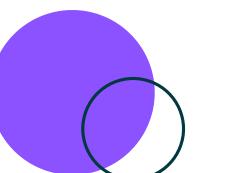
ERROR ANALYSIS

FALSE NEGATIVE

"angelina jolie coming for your baby"



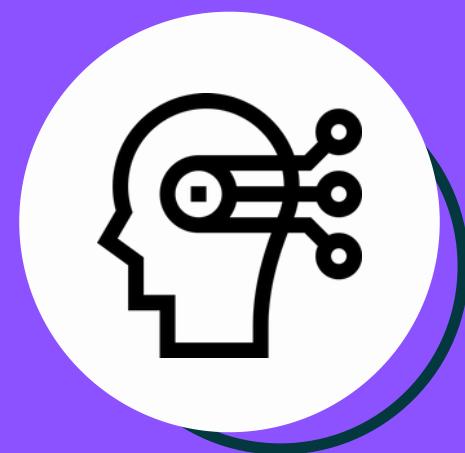
CONCLUSIONS

-  **Review of the main related work and publicly available datasets for sarcasm detection.**
-  **Error analysis is performed, displaying the weaknesses of each approach.**
-  **BERT approach overcomes the hybrid neural architecture described in Misra et al.**
-  **First study that applies BERT to detect sarcasm in texts that are not extracted from social media.**

FUTURE WORK



**Evaluate models
using other
datasets**



**Study
how to encode
knowledge of the world**



**Explore multimodal sarcasm
detection accompanying
texts with audio**



**Create Spanish dataset
for the sarcasm
detection task**

GITHUB REPOSITORY

**THE CODE OF THIS PROJECT IS
AVAILABLE ONLINE**

<https://github.com/ElsaScola/Sarcasm-Detection-with-Natural-Language-Processing-and-Deep-Learning>

BERT ARCHITECTURE

