

SARCASM DETECTION using deep learning

Hareetima Sonkar (Group-7)





Problem Description

- Sarcasm is a complex sentiment expressed using intensified positive or positive words typically intended to convey a negative connotation.
- The challenge lies in interpreting sarcastic content correctly, especially in Natural Language Processing (NLP) and sentiment analysis.
- Our goal is to build a model that can recognize and understand sarcastic behaviour and patterns in text

1. Business Problem

Misinterpreting sarcasm can lead to inaccurate sentiment analysis, which can adversely affect businesses in various ways:

- Customer Feedback
- Public Relations
- Brand Reputation Management
- Social Media Monitoring
- Ad Campaign Analysis

2. Solution Description

1. Data Collection, Preprocessing & Cleaning:

- Data Loading
- Duplicate Removal, Removing null values etc.
- Data Summary
- removing punctuations, stopwords, etc.
- Visualizations
- Descriptive Statistics

2. Text Tokenization and Padding

- TensorFlow's Keras
- Padding type = 'post'

3. Model Training

- Model Architecture
- Data Splitting
- Training

4. Model Evaluation

- Performance Metrics
- Visualizations

5. Prediction

- The dataset contains text comments labelled as *sarcastic (1)* and *non-sarcastic (0)*.

- Each comment is a short piece of text sourced from social media or other platforms [Reddit].

- **Dataset size:** 1million (approx.)

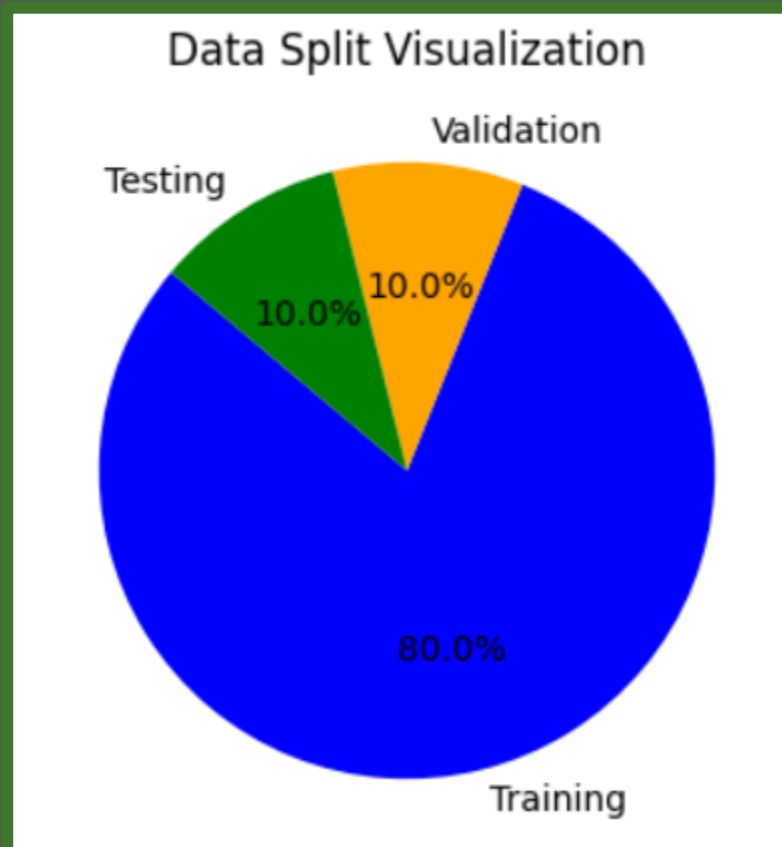
Examples:

label	comment																		
0	Wow what happened here																		
1	Forgot the																		
1	too bad thatll be the last card youll ever have in your entire life Enjoy your new pooping money sucking vampire you created																		
0	why did you include her address																		
1	I guess they dont get waves in Japan then																		
0	One less euphonium in the concert band																		
1	Too bad they dont have a player that can make some big plays and get the team hyped																		
0	Joey Butler did make history though																		
0	bruh																		
0	Do you still discard your hand																		
1	Worse product ever I hate these so much, I'm team mossy oak																		
1	Leet sKills bruh																		

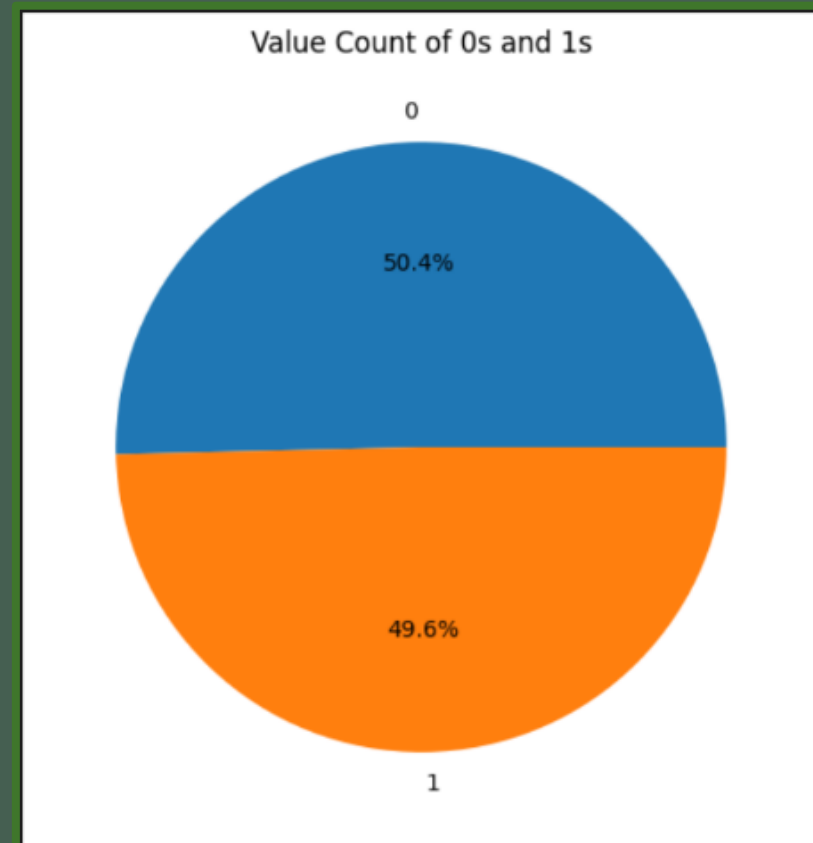
3. Dataset Description

4. Data Visualization

- Dataset Split:



- Label-wise Split:



5. Data Preprocessing

Text to lowercase; removing punctuations, duplicates, null values; new column for cleaned text; etc.

6. Tokenization and Embedding Techniques

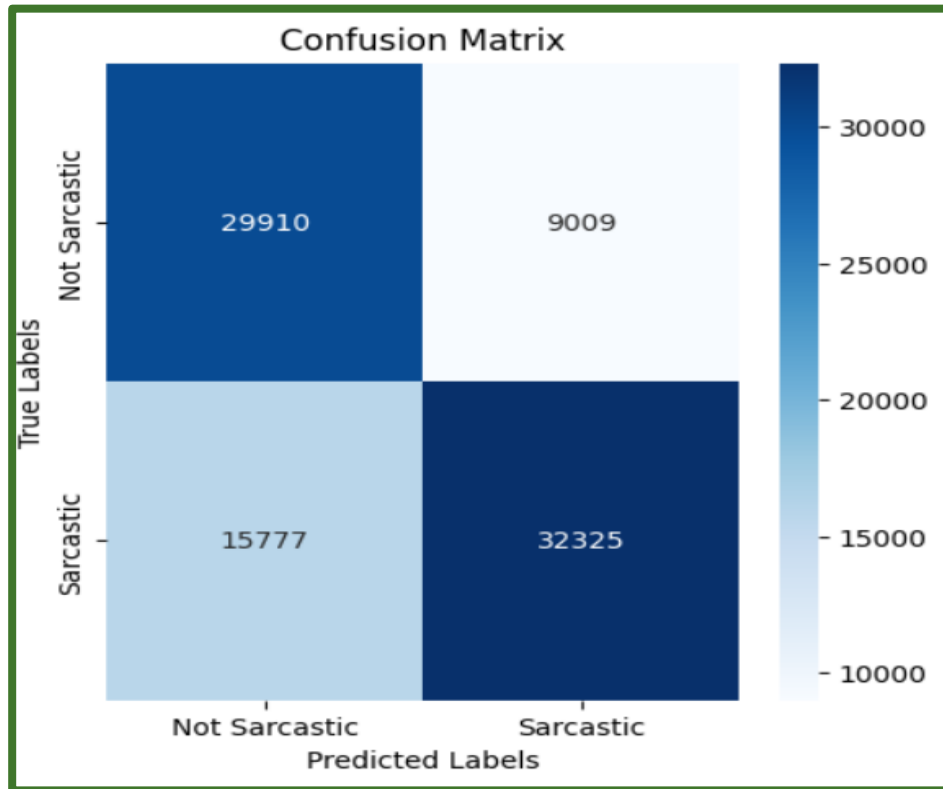
- Tokenization: The text data is tokenized using the **Keras Tokenizer**.
- Embedding: The model uses an **Embedding layer** to convert tokens into dense vectors

7. Modelling

Model: "sequential_2"

Layer (type)	Output Shape	Param #
embedding_2 (Embedding)	(None, 60, 200)	2,000,000
global_max_pooling1d_2 (GlobalMaxPooling1D)	(None, 200)	0
dense_8 (Dense)	(None, 40)	8,040
dropout_6 (Dropout)	(None, 40)	0
dense_9 (Dense)	(None, 20)	820
dropout_7 (Dropout)	(None, 20)	0
dense_10 (Dense)	(None, 10)	210
dropout_8 (Dropout)	(None, 10)	0
dense_11 (Dense)	(None, 1)	11

Confusion Matrix:



Classification Report:

Classification Report:				
	precision	recall	f1-score	support
Not Sarcastic	0.65	0.77	0.71	38919
Sarcastic	0.78	0.67	0.72	48102
accuracy			0.72	87021
macro avg	0.72	0.72	0.71	87021
weighted avg	0.73	0.72	0.72	87021



THANK YOU