



# Sentiment Analysis Tunisian Series

---

Supervised By: Mr.Wael Ouerda

# Teammates



Hedi Aloulou



Med Aymen Alimi

# Problematic

**Most Tunisian people, especially during Ramadan, are obsessed with TV series.**

**Unfortunately, the majority of producers and production companies nowadays are focusing only on:**

- Making controversial media
- Creating the 'buzz'
- Ignoring people's interests

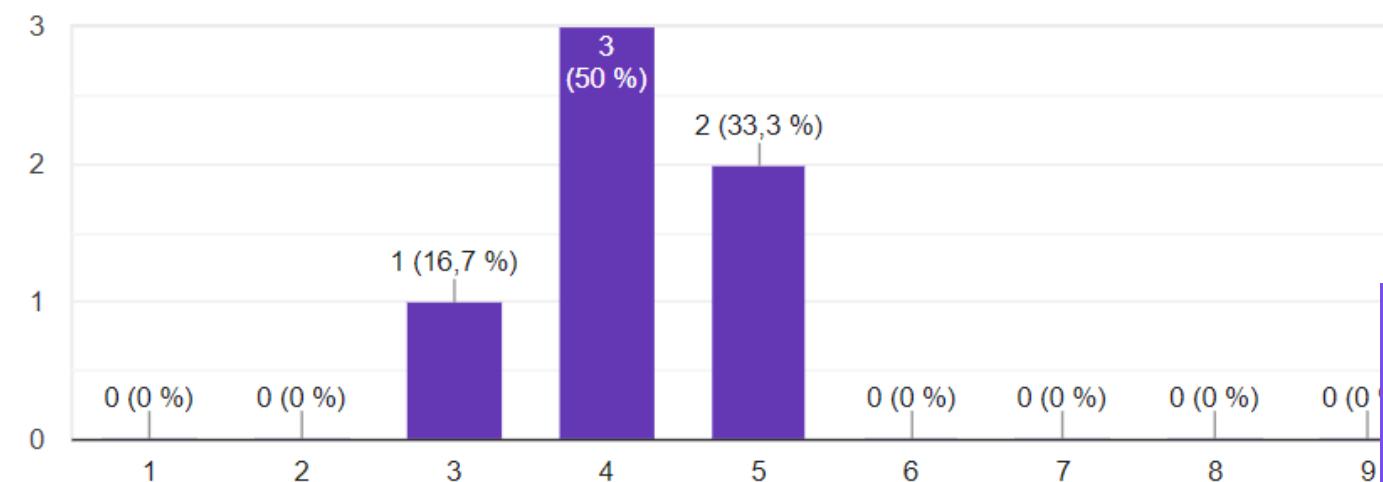


# Statistics

Please rate your satisfaction

Copier

6 réponses



Do you have any tips or comments for Tunisian producers in general about their work?

6 réponses

They should improve their scenarios , and produce more series during the whole year not only in ramadan

be more creative

Ragouj is the best , although i gave a chance to other series but they are so boring.

bad tunisian producers (sawsan jemni...)

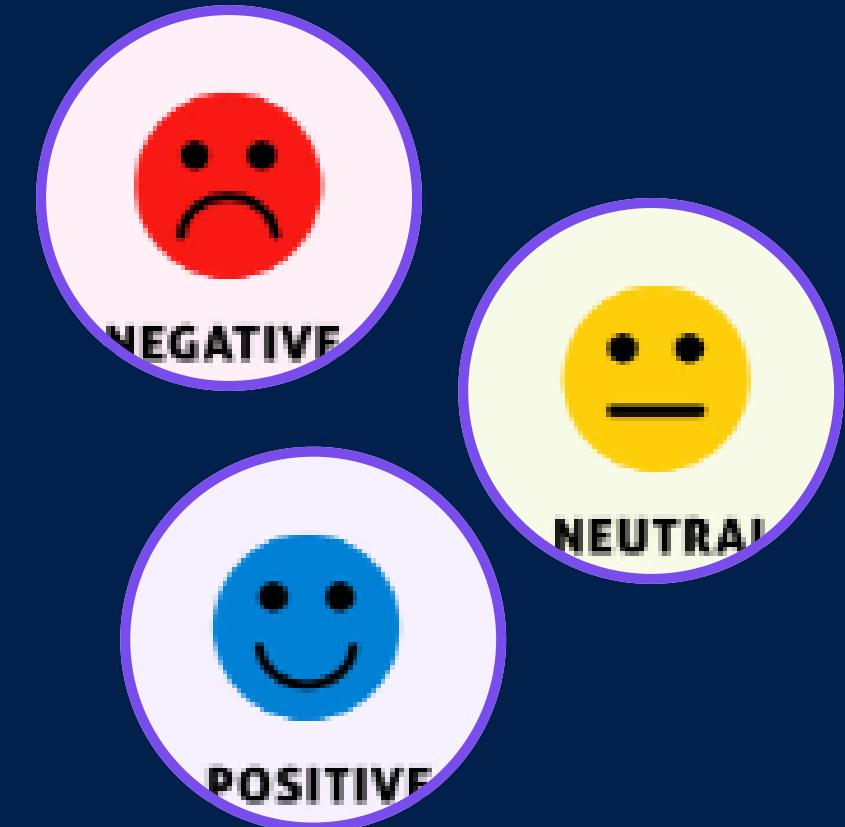
They have to get more creative with scenarios so the shows stays interesting until the end and be more realistic

We need more tv shows like ken ya makenech

# Solution

---

- One way to address the perceived decline in audience satisfaction with Tunisian series is by analyzing user feedback.
- By using sentiment analysis techniques, we can extract valuable insights from these comments.



- This will help us understand viewers opinions and identify areas for improvement.
- Ultimately, this feedback can empower series creators to craft more engaging and meaningful content that resonates with Tunisian audiences.

# Solution Process

---

# 1. Data Extraction

# 1. Data Extraction



# Youtube Comments extraction

```
df_t.to_csv('data_f.csv', index=False, encoding='utf-8')

print("Data downloaded successfully.")

Data downloaded successfully.
```

# put it into CSV

## Results

# 1. Data Extraction steps in details

Selection of YouTube videos that serves our problematic which is Tunisian television series.

We only kept the column of comments; we dropped the other columns like authorDisplayName, publishedAt ...

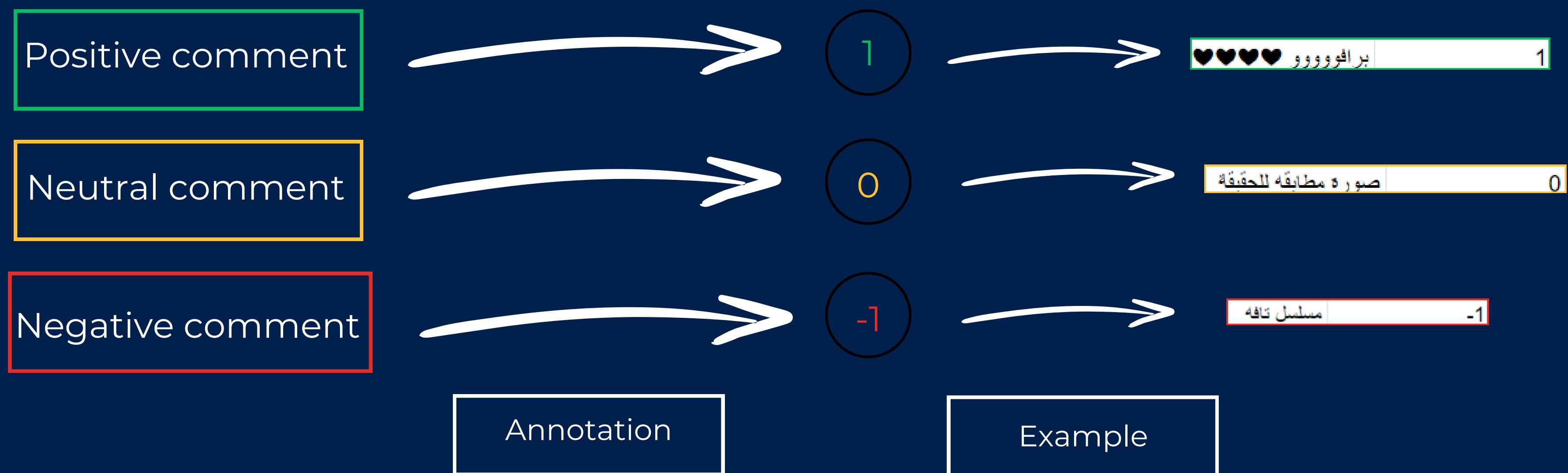
The deletion of non beneficial words, including all emojis or repetitive words like 'سفيان', 'يرحم', 'سفيان', 'حسونة', 'يرحمو', and comments that contain only digits

**Elimination  
of  
duplication.**

**Adding  
Sentiment  
Column  
'Target'**

# 2. Data Annotation

## 2. Data Annotation



# 3. Data Analysis

### **3. Data Analysis**

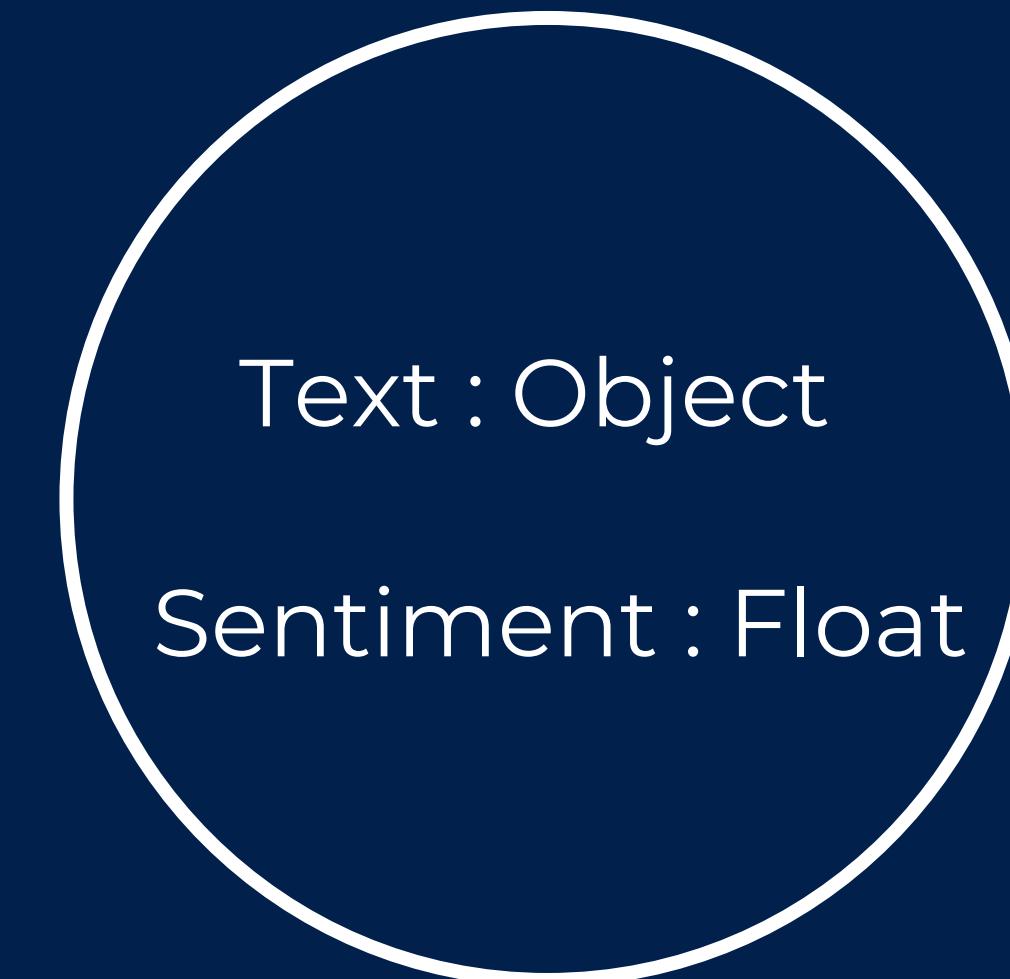
---



Features : 2

Samples : 9318

Shape

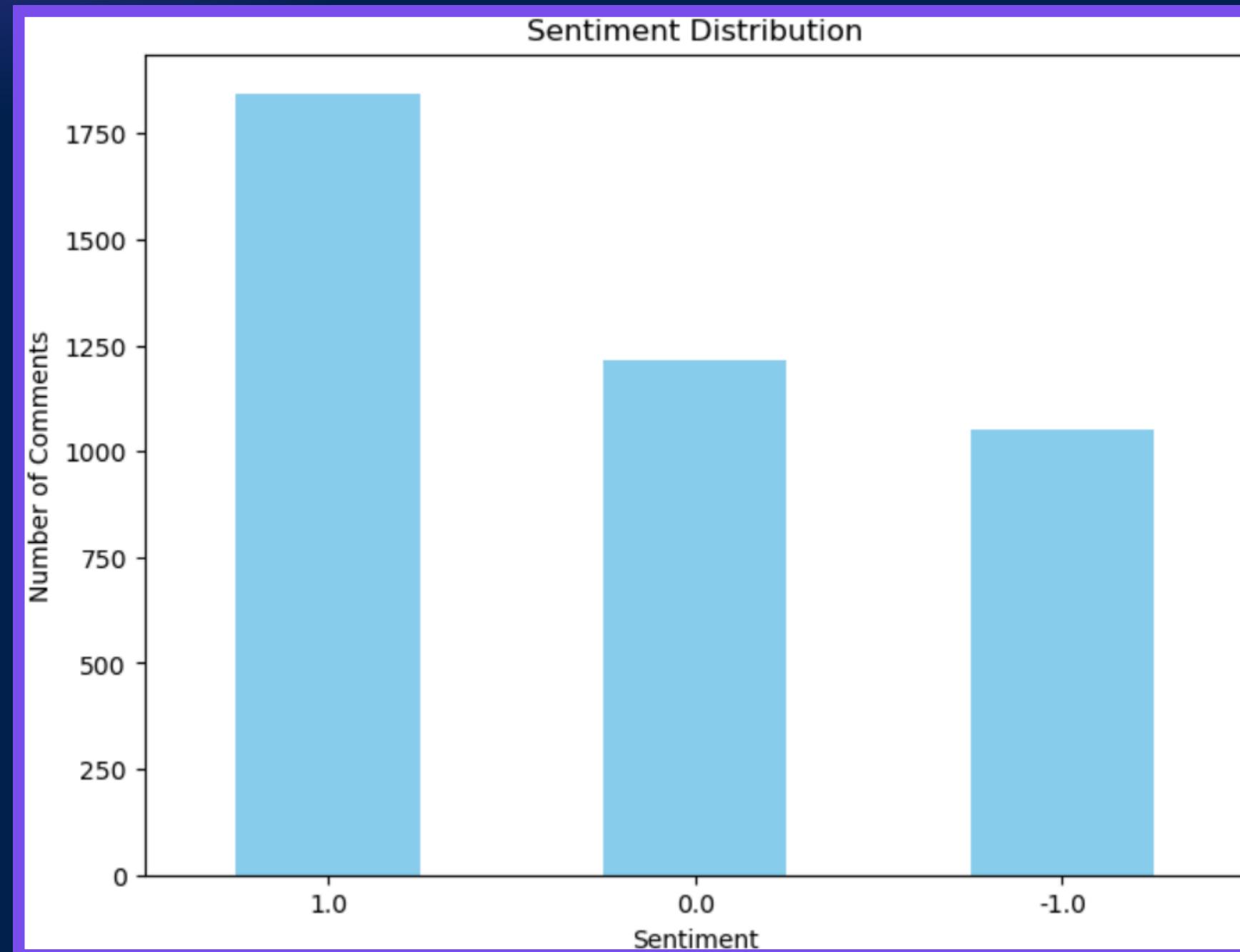


Text : Object

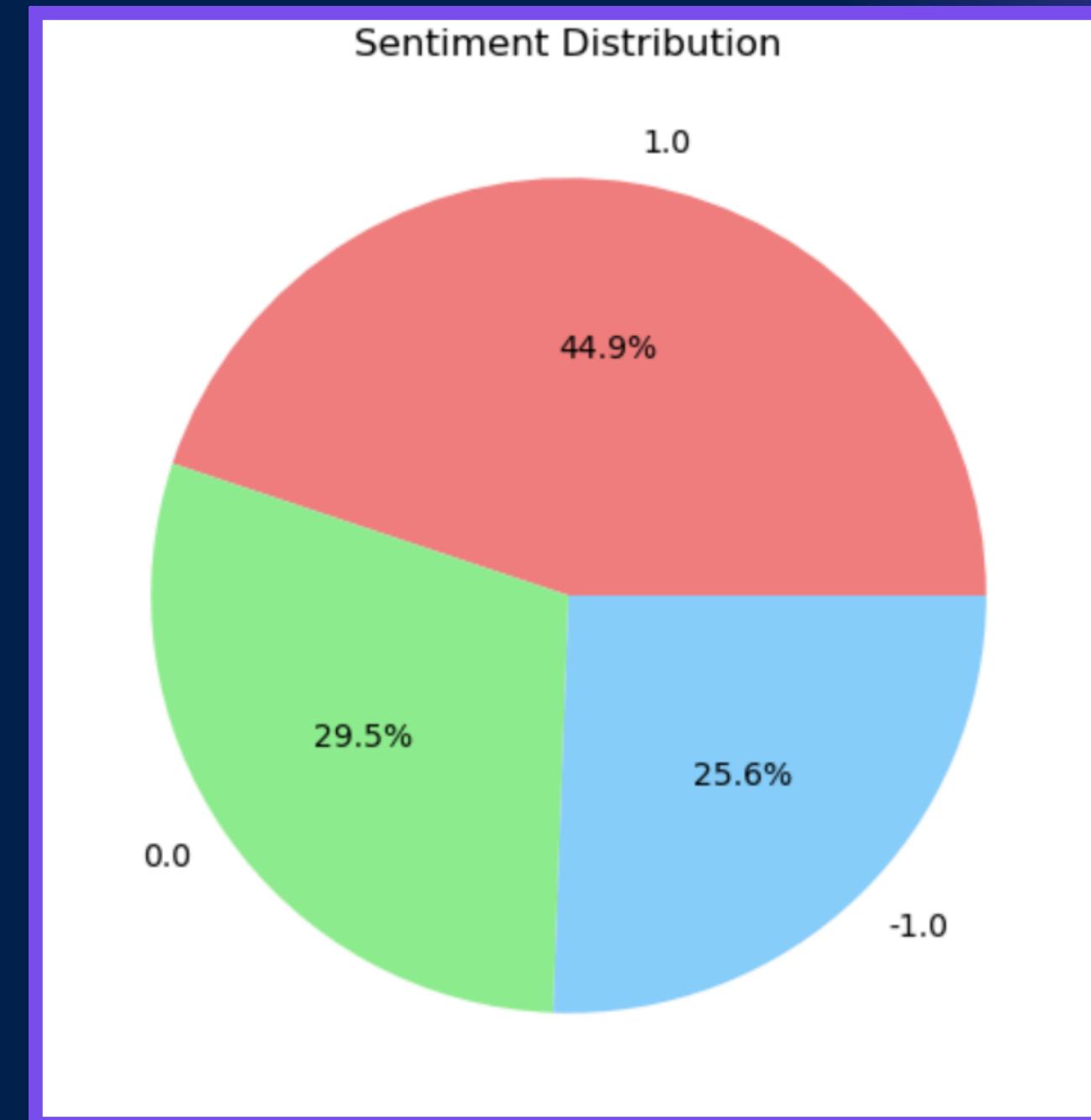
Sentiment : Float

Type

# 3. Data Analysis



Bar chart



Pie chart

# 4. Data Preprocessing

# . Data Cleaning

---

Remove non-arabic characters

Remove HTML tags

Remove Multiple white space

Remove single characters

Drop duplicate rows

Drop null rows

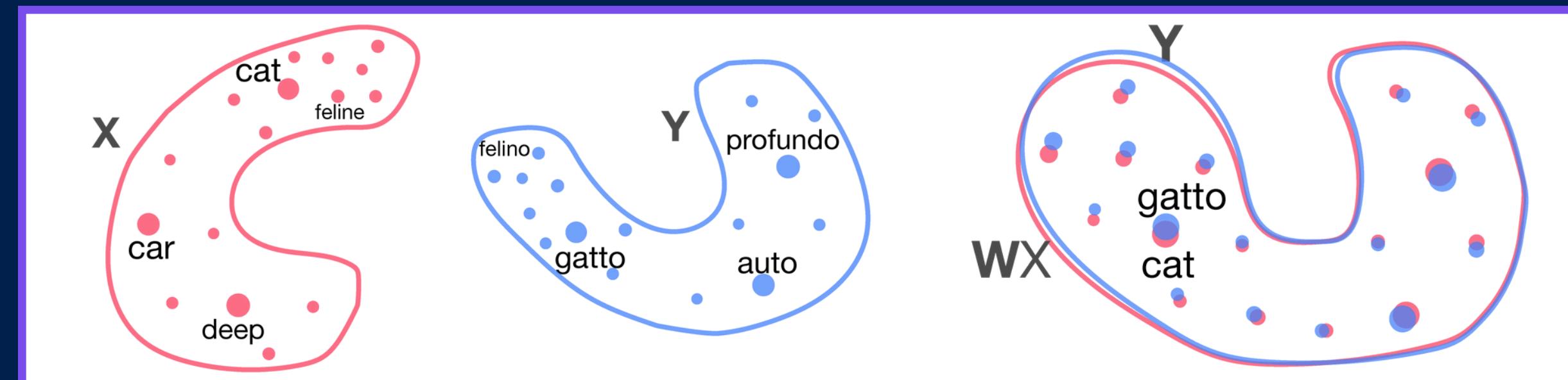
# • Word Embedding related to machine learning

- TF-IDF will transform the text into meaningful representation of integers or numbers which is used to fit machine learning algorithm for predictions.
- TF-IDF Vectorizer is a measure of originality of a word by comparing the number of times a word appears in document with the number of documents the word appears in.



# • Word Embedding related to Lstm

- Pre-trained Word Embeddings: The script uses FastText, a library developed by [Facebook's AI Research \(FAIR\) lab](#), for learning word embeddings and text classification.
- These pre-trained word embeddings are loaded from a file called `cc.ar.300.vec.gz`, which were computed using machine learning algorithms implemented in FastText.
- These algorithms were trained on extensive text data, resulting in embeddings that represent words as continuous vectors in a multi-dimensional space.



# 5. Data Processing

# 5. Data Processing

Gnb  
poly\_svm  
sgd\_svm  
rbf\_svm

Machine learning models

Lstm Model

Deep learning Model

# 6. Evaluation

# Lstm Model

```
# Build LSTM model
model = Sequential()
model.add(Bidirectional(LSTM(units=128, dropout=0.2, recurrent_dropout=0.2, input_shape=(max_length, 300))))
model.add(Dense(units=3, activation='softmax')) # 3 classes, so softmax activation
```

## HyperParameter set

```
Epoch 20: val_accuracy did not improve from 0.51681
101/101 [=====] - 79s 783ms/step - loss: 23.6163 - accuracy:
0.2974 - val_loss: 4.3197 - val_accuracy: 0.3176
26/26 [=====] - 3s 103ms/step
```

The Best Accuracy in the lstm model is: **0.51681**

# LSTM Evaluation

1/1 [=====] - 0s 141ms/step

Comment: مسلسل 'وادي الذئاب' من أفضل المسلسلات التونسية التي شاهدتها.

Predicted Sentiment: Neutral

1/1 [=====] - 0s 165ms/step

Comment: لقد استمتعت كثيراً بمتابعة مسلسل 'علاء الدين'.

Predicted Sentiment: Neutral

1/1 [=====] - 0s 154ms/step

Comment: ما رأيك في مسلسل 'حكايات تونسية'؟

Predicted Sentiment: Neutral

Below average

Below what is expected

# LSTM Evaluation

might not capture  
the semantic  
nuances

Data mismatch

LSTM are sensitive to

- batch size
- Learning rate
- Number of units

Hyperparameter Tuning

# Machine learning algorithms Evaluation

***** Performance of neural network *****				
	precision	recall	f1-score	support
-1.0	0.79	0.62	0.69	188
0.0	0.59	0.62	0.60	255
1.0	0.74	0.80	0.77	360
accuracy			0.70	803
macro avg	0.71	0.68	0.69	803
weighted avg	0.70	0.70	0.70	803

# Machine learning algorithms Evaluation

***** Performance of linear svm *****				
	precision	recall	f1-score	support
-1.0	0.83	0.69	0.75	188
0.0	0.65	0.66	0.66	255
1.0	0.77	0.83	0.80	360
accuracy			0.74	803
macro avg	0.75	0.73	0.74	803
weighted avg	0.75	0.74	0.74	803

# Machine learning algorithms Evaluation

***** Performance of rbf *****				
	precision	recall	f1-score	support
-1.0	0.87	0.51	0.64	188
0.0	0.67	0.59	0.62	255
1.0	0.68	0.88	0.76	360
accuracy			0.70	803
macro avg	0.74	0.66	0.68	803
weighted avg	0.72	0.70	0.69	803

# Machine learning algorithms Evaluation

***** Perfomance of sgd *****				
	precision	recall	f1-score	support
-1.0	0.85	0.67	0.75	188
0.0	0.65	0.69	0.67	255
1.0	0.78	0.83	0.80	360
accuracy			0.75	803
macro avg	0.76	0.73	0.74	803
weighted avg	0.75	0.75	0.75	803

# Machine learning algorithms Evaluation

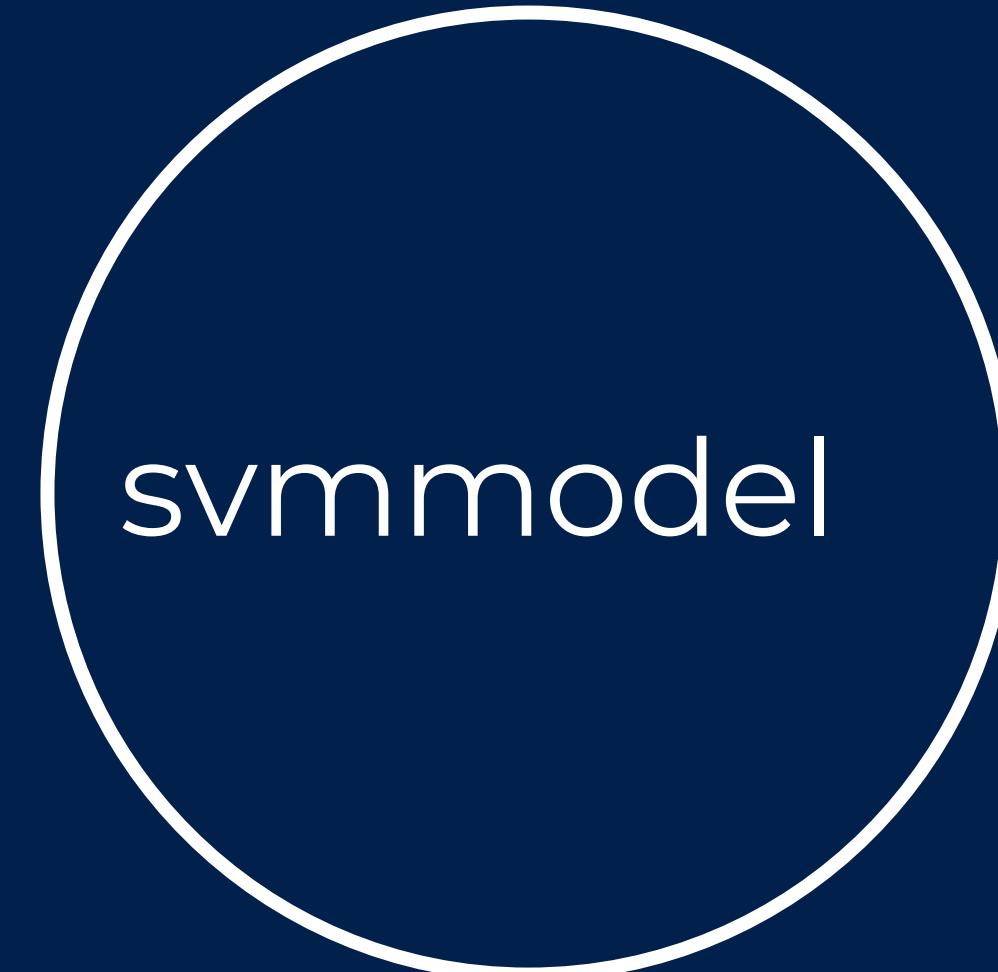
***** Performance of poly *****				
	precision	recall	f1-score	support
-1.0	0.95	0.49	0.65	188
0.0	0.72	0.42	0.53	255
1.0	0.59	0.92	0.72	360
accuracy			0.66	803
macro avg	0.75	0.61	0.63	803
weighted avg	0.72	0.66	0.64	803

# Machine learning algorithms Evaluation

***** Perfomance of gnb *****				
	precision	recall	f1-score	support
-1.0	0.73	0.69	0.71	188
0.0	0.46	0.68	0.55	255
1.0	0.75	0.51	0.61	360
accuracy			0.61	803
macro avg	0.64	0.63	0.62	803
weighted avg	0.65	0.61	0.61	803

# 7. Deployment

# 7. Deployment



Local URL: <http://localhost:8502>  
Network URL: <http://192.168.100.11:8502>

Local URL: <http://localhost:8501>  
Network URL: <http://192.168.100.11:8501>

# 7. Deployment

Leave your comment

Predict

Positive

## Sentiment Analysis LSTM

Leave your comment

Predict

Positive



# 7. Deployment

## Sentiment Analysis sigmoid

Leave your comment

مسلسل تافه

Predict

Negative

## Sentiment Analysis LSTM

Leave your comment

مسلسل تافه

Predict

Negative



NEGATIVE

# 7. Deployment

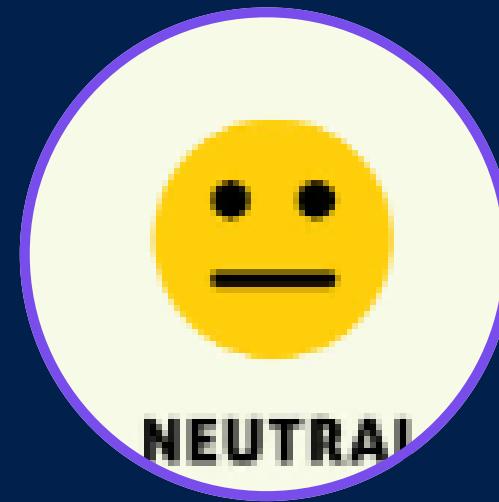
**Sentiment Analysis sigmoid**

Leave your comment

صورة مطابقة للحقيقة

Predict

Neutral



**Sentiment Analysis LSTM**

Leave your comment

صورة مطابقة للحقيقة

Predict

Neutral

**Thank You  
For Your Attention**