

**Università degli Studi Milano**

Data Science and Economics - Text Mining and Sentiment Analysis

**How do you  
feel my dear?**

ANDREA PIO CUTRERA

# Emotion detection

## NLP TASK

"to aid computers understand  
and sometimes generate human  
languages"

We want to build a deep learning  
based system able to predict  
emotions from texts

# Emotion detection

We want to build a supervised classifier detecting accurately emotions from dialogues.

**DATASET**  
WASSA-2017



# WASSA challenge

- SHARED TASK ON EMOTION INTENSITY

“

## WASSA challenge

- DATA ARE ORGANISED IN 3 FOLDERS
- TRAIN, DEVELOPMENT AND TEST
- ANNOTATED TEXTS FROM TWEETS
- 4 POSSIBLE EMOTIONS

**DATASET**  
WASSA-2017

“

# Data Pre-processing

<b>id</b>	<b>text</b>	<b>sentiment</b>
11041	Nothing is more relentless than a dog begging ...	anger
40980	Ok it really just sunk in that I'm seeing the ...	sadness
41210	@MariamVeiszadeh #depressing it's so freaking ...	sadness
41365	I feel like a burden every day that I waste bu...	sadness
31085	@partydelightsUK it's 5679787. Cannot DM you a...	joy
30832	This tweet is dedicated to my back pain, which...	joy
20804	@StephenKing\n\nStephen King never once spoke ...	fear
30318	I found #marmite in Australia. `:)	joy

- 7,102 RECORDS
- RAW CORPUS TEXTS FROM TWEETS
- SENTIMENTS ANNOTATED

**DATASET**  
WASSA-2017

“

# Texts processing

```
def clean_emoji_output(text):
    return re.sub(":", " ", text)

def strip_lowercase(text):
    return text.strip().lower()

def clean_text(data):
    data['clean_text'] = data['text'].apply(nfx.remove_emails)
    data['clean_text'] = data['clean_text'].apply(nfx.remove_numbers)
    data['clean_text'] = data['clean_text'].apply(nfx.remove_urls)
    data['clean_text'] = data['clean_text'].apply(nfx.remove_userhandles)
    data['clean_text'] = data['clean_text'].apply(demoji.replace_with_desc)
    data['clean_text'] = data['clean_text'].apply(clean_emoji_output)
    data['clean_text'] = data['clean_text'].apply(nfx.remove_special_characters)
    data['clean_text'] = data['clean_text'].apply(nfx.remove_bad_quotes)
    data['clean_text'] = data['clean_text'].apply(nfx.remove_html_tags)
    data['clean_text'] = data['clean_text'].apply(nfx.remove_punctuations)
    data['clean_text'] = data['clean_text'].apply(nfx.remove_stopwords)
    data['clean_text'] = data['clean_text'].apply(nfx.remove_multiple_spaces)
    data['clean_text'] = data['clean_text'].apply(strip_lowercase)
```

**DATASET**

WASSA-2017



# Texts processing

*@myusername Today I'm VEry happy  
because my website 333876  
<https://myprettywebsite.com> HAS been  
'deployed'".*



*contact me if  
something ççççooo need to be"" fixed.....":  
usr@gmail.com*

# **DATASET**

## WASSA-2017

“

# remove emails

*@myusername Today I'm VEry happY  
because my website 333876  
<https://myprettywebsite.com> HAS been  
'deployed'". ☺☺☺ contact me if  
something çççç°°° need to be"" fixed.....":*

“

# remove numbers

*@myusername Today I'm VEry happY  
because my website  
<https://myprettywebsite.com> HAS been  
'deployed'". ☺☺☺ contact me if  
something ☺☺☺ need to be"" fixed.....":*

“

## remove urls

*@myusername Today I'm VEry happY  
because my website*

*HAS been*

*'deployed'". ☺☺☺ contact me if  
something ☺☺☺ need to be fixed....":*

“

# remove usernames

*Today I'm VEry happY  
because my website  
HAS been  
'deployed'". ☺☺☺ contact me if  
something çççç°°° need to be"" fixed.....":*

“

## replace emoji with description

*Today I'm VEry happy  
because my website  
HAS been  
'deployed'". :smiling face heart eyes :  
:beaming face smiling eye::beaming face  
smiling eye: contact me if something ççççooo  
need to be"" fixed.....":*

**DATASET**

WASSA-2017

“

## clean emoji description

*Today I'm VEry happy  
because my website  
HAS been  
'deployed'". smiling face heart eyes  
beaming face smiling eye beaming face  
smiling eye contact me if something ççççooo  
need to be"" fixed.....":*

**DATASET**

WASSA-2017

“

## remove special characters

*Today I'm VEry happY  
because my website  
HAS been  
'deployed'". :smiling face hearteyes  
beaming face smiling eye beaming face  
smiling eye contact me if something  
need to be"" fixed.....":*

**DATASET**

WASSA-2017

“

## remove bad quotes

*Today I'm VEry happy  
because my website  
HAS been  
deployed . smiling face hearteyes  
beaming face smiling eye beaming face  
smiling eye contact me if something  
need to be fixed.....:*

**DATASET**

WASSA-2017

# remove punctuations

# DATASET

## WASSA-2017

“

## remove stopwords

*Today I'm happy because  
<https://my.pretty.website.com>*

*hearteyes beaming face smiling eye beaming  
face smiling eye contact me if something çççç°°  
° need to be fixed*

“

## remove multiple spaces

*Today Im happy website deployed smiling face  
hearteyes beaming face smiling eye beaming  
face smiling eye contact need fixed*

“

## strip and lower case

*today im happy website deployed smiling face  
hearteyes beaming face smiling eye beaming  
face smiling eye contact need fixed*

“

## Initial text

*@myusername Today I'm VEry happy because my website 333876  
https://myprettywebsite.com HAS been 'deployed'". 😊@@@  
contact me if something çççç°°° need to be"" fixed.....": usr@gmail.com*

## Cleaned text

*today im happy website deployed smiling face hearteyes beaming face  
smiling eye beaming face smiling eye contact need fixed*

“

# Tokenization

Cleaned texts are tokenized

# Stemming

Tokens are substituted by the lemma where they come from (through **WordNetLemmatizer()**)

# De-tokenization

Lemmas are then joined to get the final cleaned and standardised texts (through **TreebankWordDetokenizer()**).

**DATASET**

WASSA-2017

# Emotions



FEAR



ANGER



SADNESS



JOY

“

# Numerical Encoding

Emotional status class is encoded numerically:

- Anger: -----0,
- Fear: -----1,
- Joy: -----2,
- Sadness: -----3.

# One-Hot-Encoding

Numerical labels are one-hot encoded to make functioning the binary classifiers.



# **Train-Validation-Test split**

## **DATA PREPARATION**

- 5313 training cleaned texts with corresponding 5313 training labels;
- 865 validation cleaned texts with corresponding 865 validation labels;
- 924 testing cleaned texts with corresponding 924 testing labels

# Model

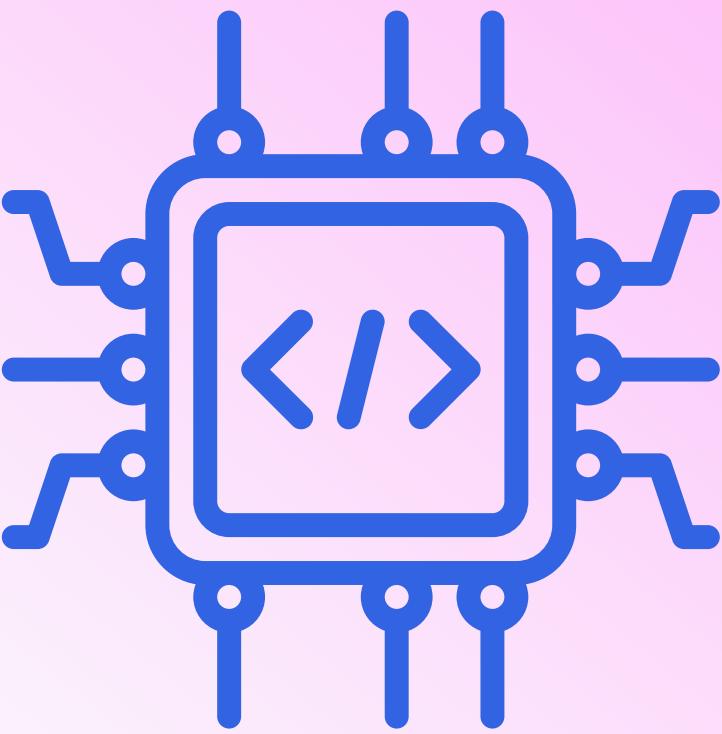
## RECURRENT NEURAL NETWORK

The architecture has few main features which allows us to make the learning from the sequences (of words) meaningful and memory preserving.

# Embedding

## TEXT VECTORIZATION

Get a vectorial representation of our tokenized padded sequences. Pre-trained model GloVe has been used for this layer.



# CudNNLSTM

## **MEMORY PRESERVING**

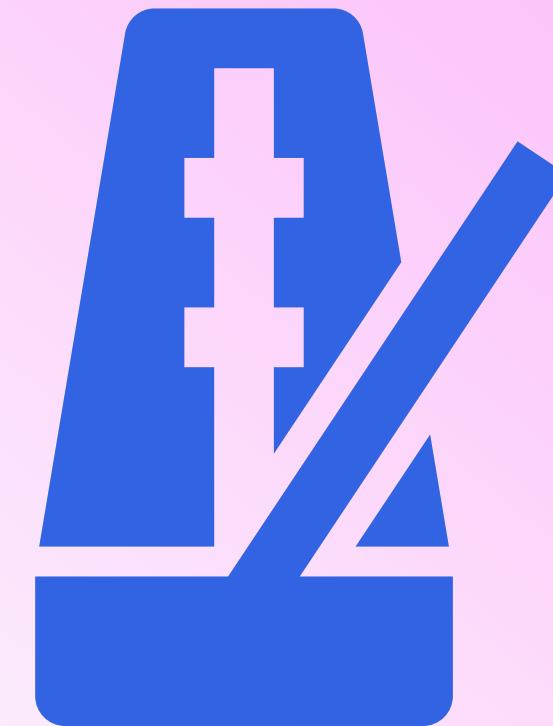
Long-Short Term Memory layers which propagate the signal forward and backward trying not to lose the interdependence which exist among words in a proposition.



# Dropout

## REGULARIZATION

Layers which prevent overfitting by stopping the updating of some weights of input neurons.



# Model



# Hyperparameters optimization

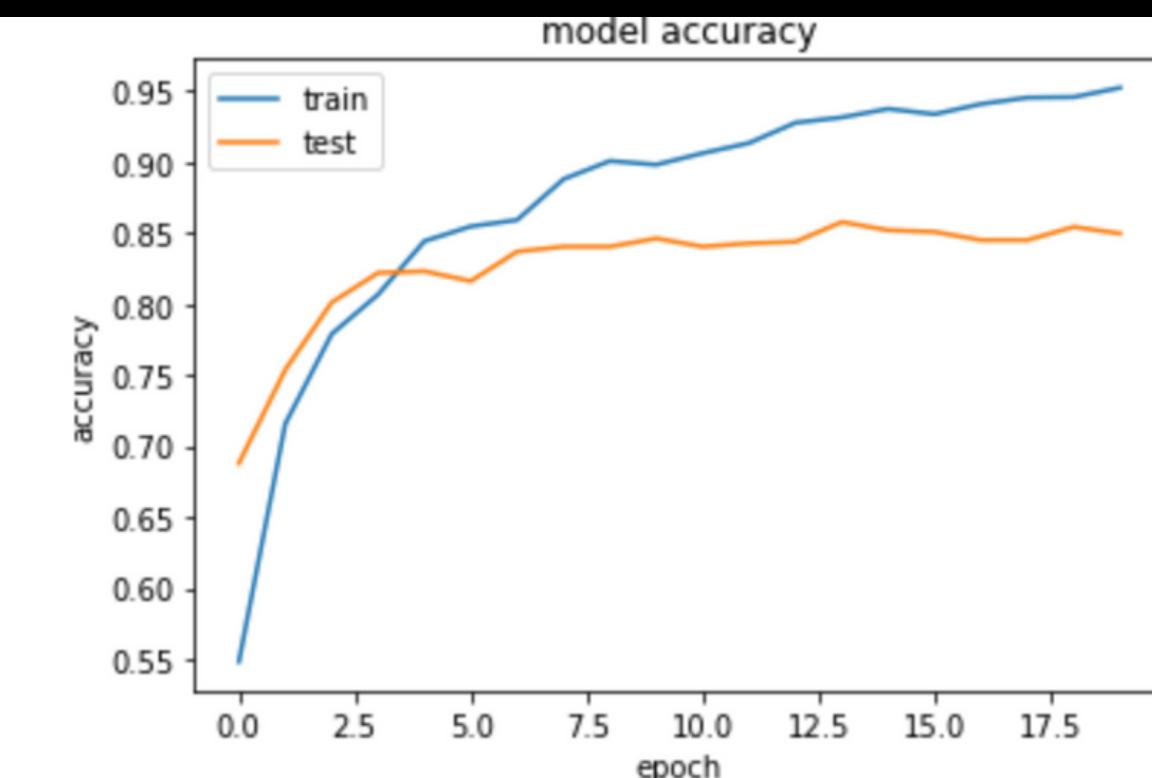
## TUNING

- cardinality of first LSTM layer: 50
- cardinality of second LSTM layer: 150
- cardinality of third LSTM layer: 50

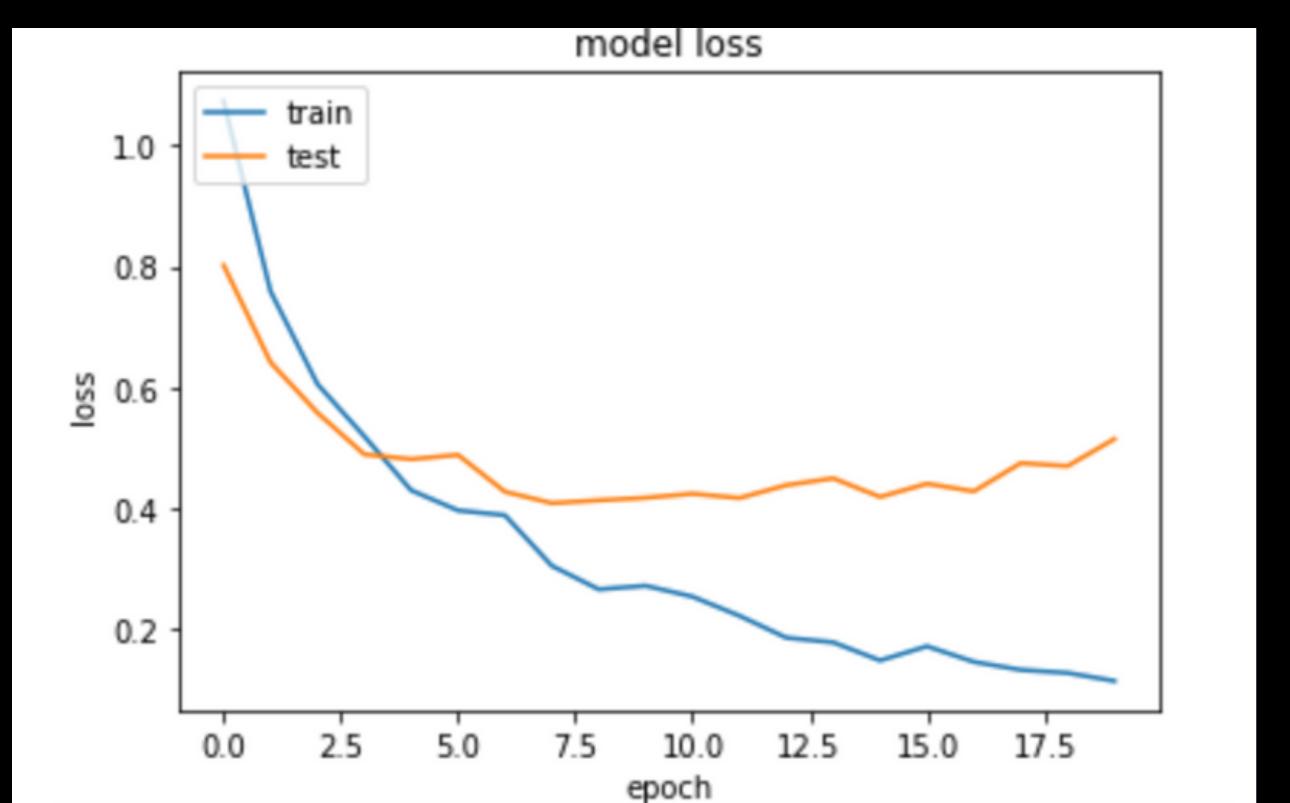
# training phase

OVERFITTING AFTER 10 EPOCHS

5-FOLD CROSS VALIDATED ACCURACY: 80%



STABLE 85% TEST ACCURACY AFTER 7 EPOCHS

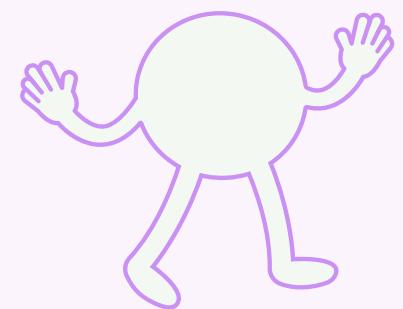


# Cornell Movie-Dialogs Corpus Data

## MOVIE DATABASE



220,579 conversational exchanges between 10,292 pairs of movie characters



9,035 characters from 617 movies



304,713 total utterances

# MAIN CHARACTERS

- Don Vito Corleone, the Godfather;
- Tom Hagen, the counselor (i.e. consigliere);
- Michael Corleone, the son of the Godfather approaching to become the heir of the family;
- Peter Clemenza, the main henchman of the family;
- Sonny Corleone the choleric son of the Godfather.



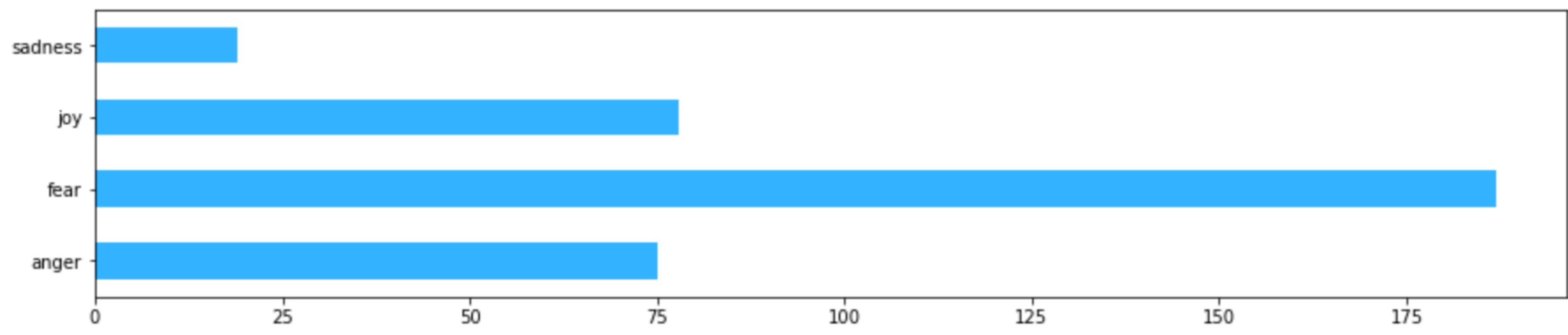
# The Godfather

## UTTERANCES

	<b>id</b>		<b>text</b>	<b>character</b>	<b>conversation_id</b>	<b>reply_to</b>
0	L593445	You owe the Don a service. He has no doubt th...		HAGEN	L593443	L593444
1	L593443	This is Tom Hagen; I'm calling for Don Corleon...		HAGEN	L593443	NaN
2	L593425		Yes.	HAGEN	L593422	L593424
3	L593423	Good. He never doubted you.		HAGEN	L593422	L593422
4	L593421	You owe the Don a service. In one hour, not b...		HAGEN	L593420	L593420

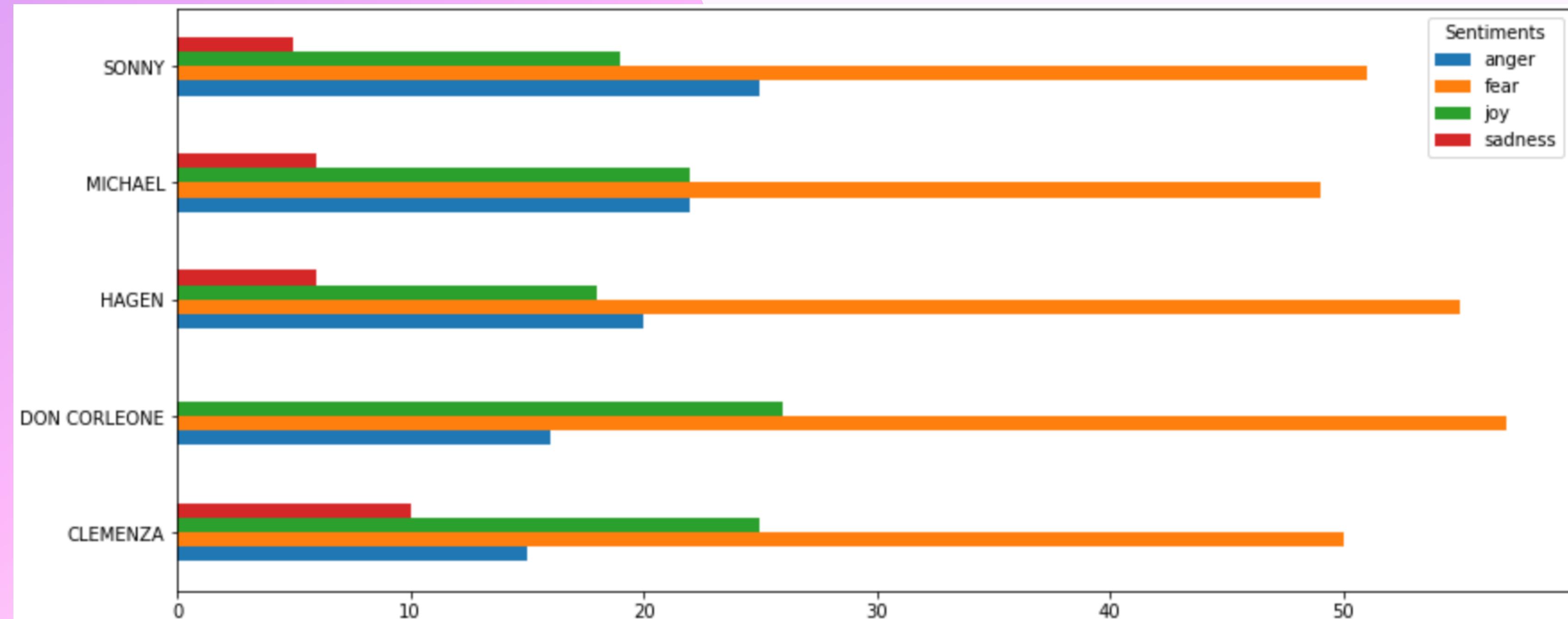
# The Godfather

## EMOTIONS



# The Godfather

## EMOTIONS BY CHARACTER



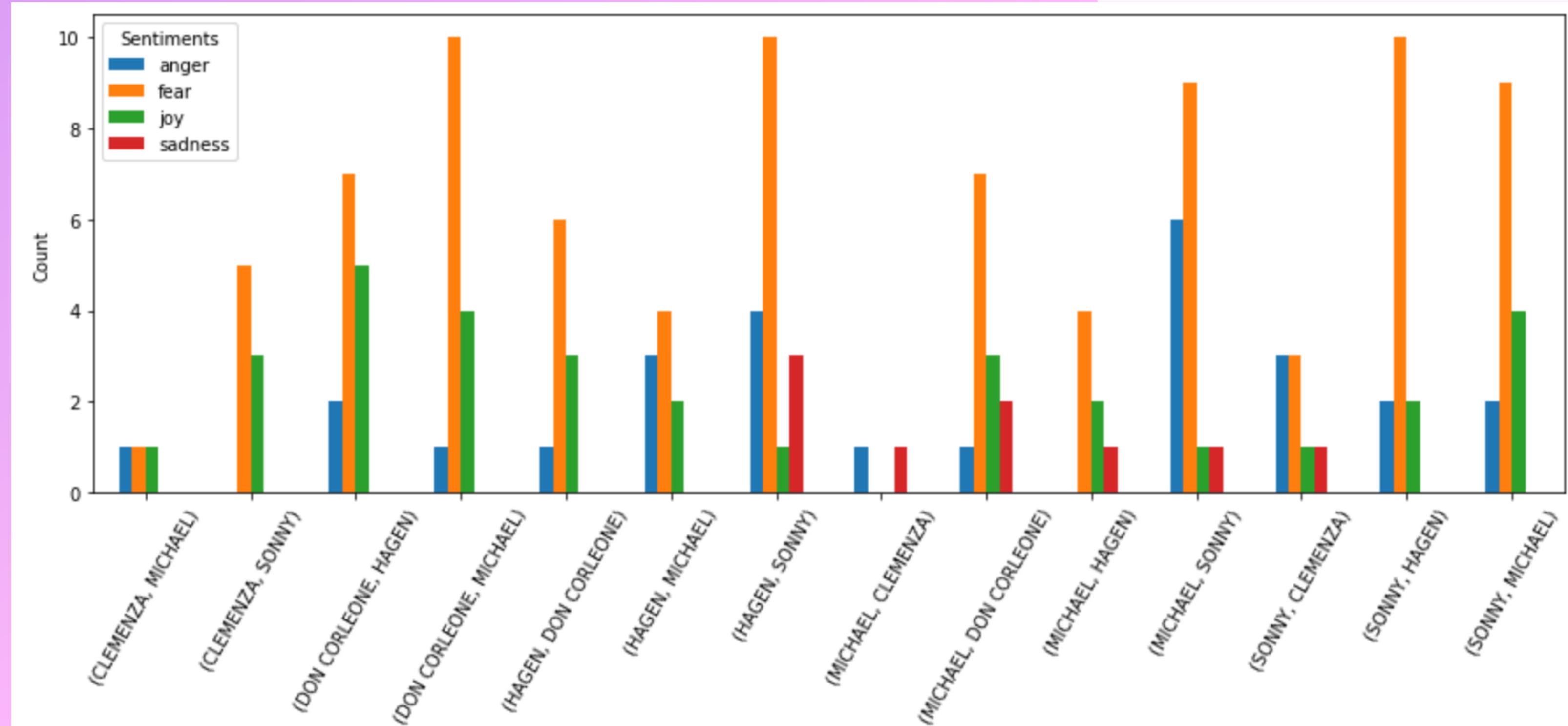
# The Godfather

## EMOTIONS BY SPEAKER-RESPONDENT COUPLES

conversation_id	character Speaker	text Speaker	prediction Speaker	character Respondent	text Respondent	prediction Respondent
147	L592889 DON CORLEONE	Is it necessary?	fear	HAGEN	You understand him better than anyone.	fear
55	L592928 HAGEN	Hello Kay. Your father's inside, doing some b...	anger	MICHAEL	Thanks Tom.	fear
65	L592936 DON CORLEONE	I'm sure it's the most generous gift today.	joy	HAGEN	The Senator called-- apologized for not coming ...	fear

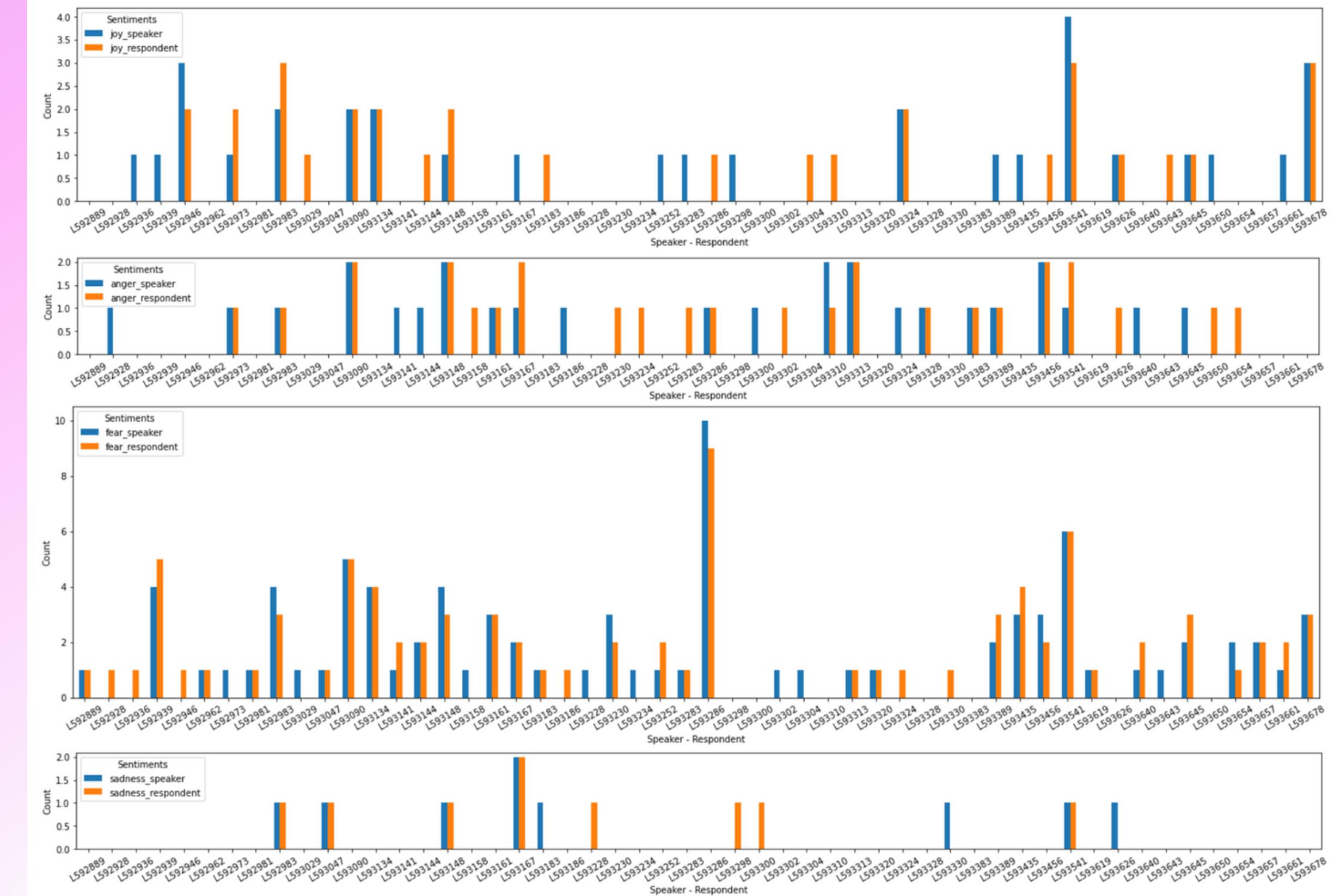
# The Godfather

## EMOTIONS BY SPEAKER-RESPONDENT COUPLES



# The Godfather

## EMOTIONS EVOLUTION



# Thank you

MATRICOLA

965591

EMAIL

[andrea.cutrera@studenti.unimi.it](mailto:andrea.cutrera@studenti.unimi.it)