



THE GOOD, THE BAD AND THE UGLY

Shakespeare Dataset



LAST NAME: RASHAD
NAME: AMR MOHAMED NAZIH MOHAMED
ID: 991043

Abstract

The aim of this project is to perform emotion detection techniques on the given Shakespeare dataset, observe the emotional evolution of each character throughout the play and build an emotional profile for each character that would describe the various types of dominant emotions a character feels. I have decided to focus on a specific play which is 'Antony & Cleopatra'. I have tried different emotion detection techniques in order to make comparisons between each technique and how can a character be emotionally represented in the play differently. I have also implemented social network analysis to describe the emotion each character emits when interacting with another person.

1. Introduction:

Emotion detection is the process of labeling a piece of text to as an emotional expression, for instance emotion detection can be used to describe the emotional context of a product/service review, this would give a better understanding of why a user has given a specific rating for the product/service received. In our case we are implementing emotion detection to describe the emotional profile of fictional characters in a play written by the famous English poet & writer. The dataset selected is a csv file, containing several plays in the form of rows and columns, where the rows represent each sentence or part of a sentence a character is reciting, and the columns considered in the csv file are the following:

- Player (i.e., character)
- Player Line
- Play (as mentioned earlier the play name is ‘Antony & Cleopatra’)
- ‘ActSceneLine’ (for example 1.2.3 means Act 1, Scene 2, Line 3)

This project can be broken down into 3 phases which are the following:

- Implementing Text2Emotion model (non-hugging face model)
- Implementing a variant of RoBERTa, a pre-trained transformer-based model, available through Hugging Face Transformers library, this variant is 'EmoRoBERTa'.
- Implementing Social Network Analysis (SNA) to describe the dominant emotion a character emits when interacting with another character in the play.

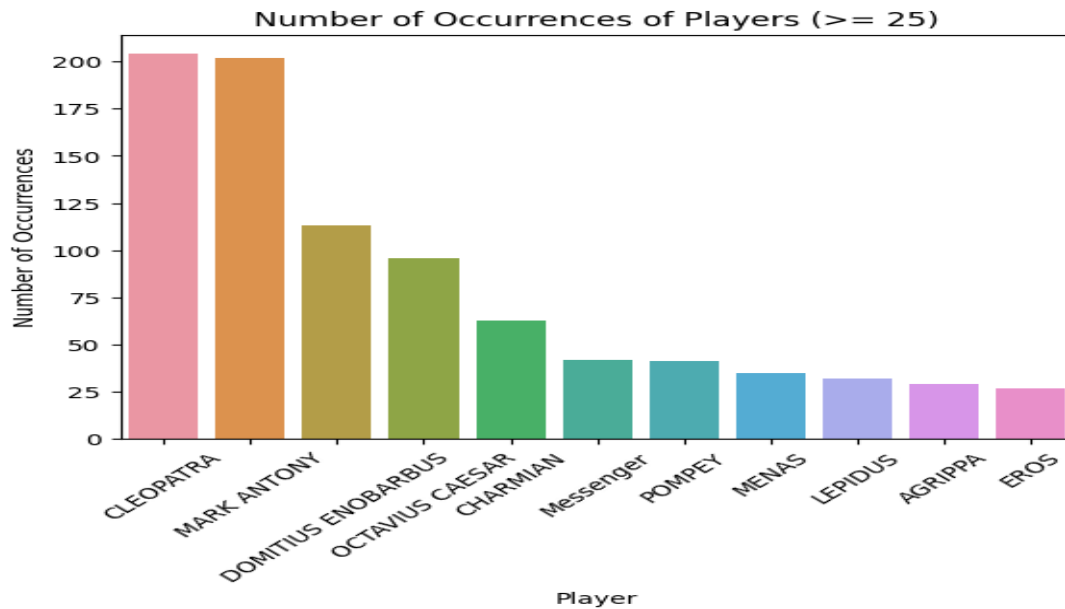
2. Emotion Detection in Text Mining:

Emotion detection, which is a specific type of ‘Text Classification’, can be implemented in various ways to determine the sentiment behind a specific text. A simple implementation of this technique involves binary classification, where a text is classified as either having a positive or negative sentiment. However, it can become more complex by classifying multiple texts into various types of emotions. Since we are working on theatrical play dataset, it should be appropriate to consider a more complex approach to describe the different emotions each character feels when reciting his/her lines and also to create a more complex emotional profile for each character. That is why I have decided to use two different models to perform emotional detection on this dataset and see which one gives a more appropriate output.

3. Identifying The Main Characters:

The main characters have been identified by introducing a threshold, if certain character’s lines exceed this threshold, then s/he is to be considered as a main character. The value of this threshold is 25. In order to avoid having the same character counted twice if s/he has 2 or more consecutive lines, a new column has been introduced to only count the instance when a character starts reciting.

By Performing a bar plot, we can determine that we have 11 players that are considered as main characters.



4. Emotion Detection Techniques Implemented:

i. Text2Emotion:

Text2Emotion is a python package that is used to classify textual data into different types of emotions. The output of this package is in the form of a dictionary, where each emotion is a key, and the value is between 0 and 1. There are 5 basic emotion categories represented in this package, which are the following: Happy, Angry, Sad, Surprise and Fear. An example output could be: {'Angry': 0.12, 'Fear': 0.42, 'Happy': 0.04, 'Sad': 0.33, 'Surprise': 0.08}, we can see that the two dominant emotions are Fear and Sad.

In our case, we only consider the emotion with the highest value, which is 1.0, to be the dominant emotion that would represent the emotion of a given line. If there aren't any emotion represented with 1.0 as a value, then we would categorize the emotion of the given line as 'Neutral'. This is to avoid running into the situation where a line is represented by two or more emotions. Below is a representation of how the dataframe will look after implementing Text2Emotion.

	Play	Player	PlayerLine	Act	Scene	Line	StartOfSequence	Happy	Angry	Surprise	Sad	Fear
0	Antony and Cleopatra	CLEOPATRA	If it be love indeed, tell me how much.	1	1	15	True	1.0	0.0	0.0	0.0	0.0
1	Antony and Cleopatra	MARK ANTONY	There's beggary in the love that can be reckon'd.	1	1	16	True	1.0	0.0	0.0	0.0	0.0
2	Antony and Cleopatra	CLEOPATRA	I'll set a bourn how far to be beloved.	1	1	17	True	0.0	0.0	0.0	0.0	1.0
3	Antony and Cleopatra	MARK ANTONY	Then must thou needs find out new heaven, new ...	1	1	18	True	0.0	0.0	1.0	0.0	0.0
4	Antony and Cleopatra	MARK ANTONY	Grates me: the sum.	1	1	20	True	0.0	1.0	0.0	0.0	0.0

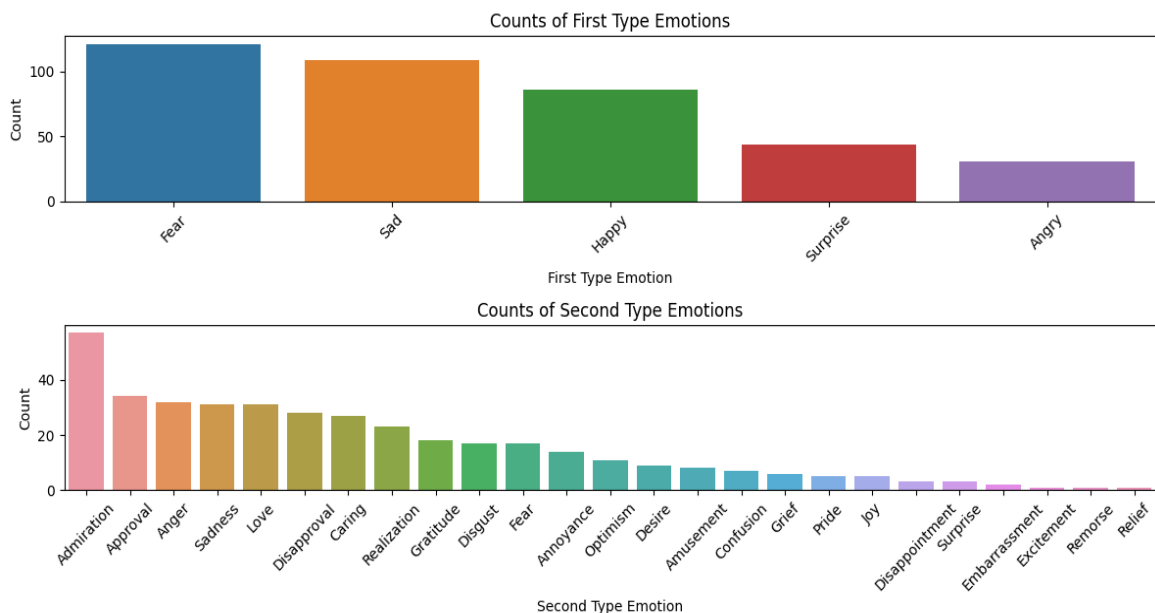
ii. EmoRoBERTa:

EmoRoBERTa is a variant of the RoBERTa model designed for emotion recognition. To understand this variant, we need to explain the RoBERTa model first. RoBERTa, which stands for ‘Robustly Optimized BERT Pretraining Approach’, is a pre-trained transformer-based model that is provided by Hugging Face Transformer’s library. It is a natural language processing model (NLP) that is built upon the transformer architecture. Moreover, The RoBERTa model is in fact an extension of the BERT (Bidirectional Encoder Representations from Transformers) model. The aim of it is to improve upon the BERT’s training methodology and to achieve better performance on several NLP tasks.

EmoRoBERTa model is well-suited for certain applications that would require analyzing the emotions conveyed in textual data. This is an example output of the EmoRoBERTa model: {'label': 'gratitude', 'score': 0.9964383244514465}. There are various types of emotions represented by this model, they are the following:

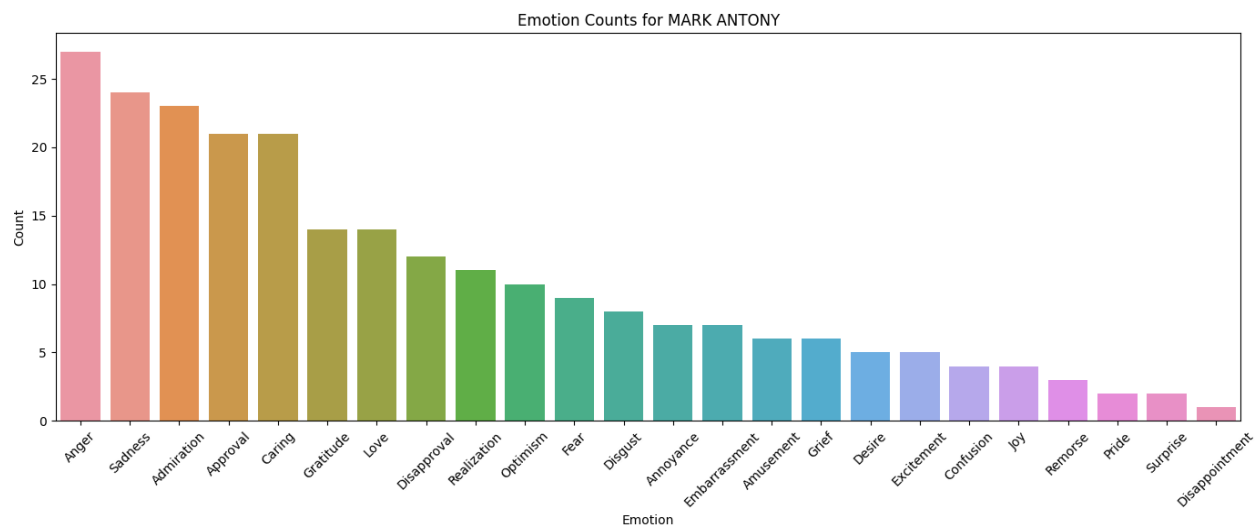
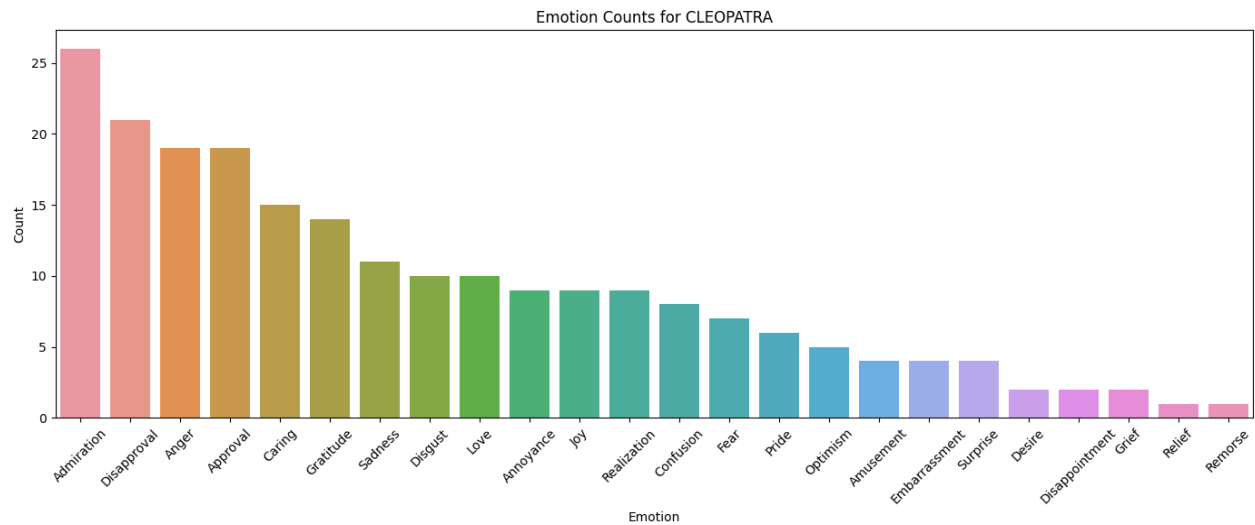
- admiration, amusement, anger, annoyance, approval, caring, confusion, curiosity, desire, disappointment, disapproval, disgust, embarrassment, excitement, fear, gratitude, grief, joy, love, nervousness, optimism, pride, realization, relief, remorse, sadness, surprise & neutral

We can clearly see that this model offers a much wider range of emotions compared to the Text2Emotion package. That's why this model appears to be more suitable for implementation on textual data from a theatrical play, as it can provide a more detailed emotional profile for each character in the play. Below is a plot that represents the difference between the two models, it's worth noting that the lines labeled with Neutral emotion have been omitted.

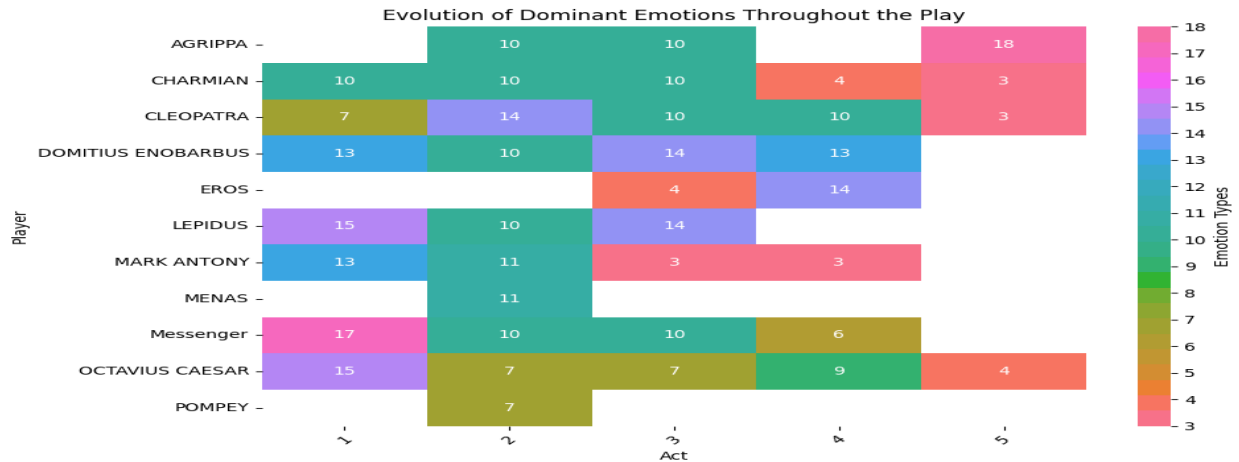


5. Character Emotional Profiles in the Play:

The process of creating an emotional profile for any character was done by implementing the EmoRoBERTa model and considering the dominant emotion expressed in the recited line. For instance, I was interested in viewing the different emotions expressed by each character. Below are two good examples which are Cleopatra and Antony.



We can see that both characters express a variety of emotions, however some of the main emotions that are expressed are: Admiration, Anger, Approval & Caring. This leads to another observation that needs to be made. Since the main characters express numerous types of emotions, I decided to investigate the dominant emotion of each character in each act and observe his/her emotional evaluation throughout the play. Below is a heatmap, with dictionary legend underneath, that represents the dominant emotion a character expresses in each act.

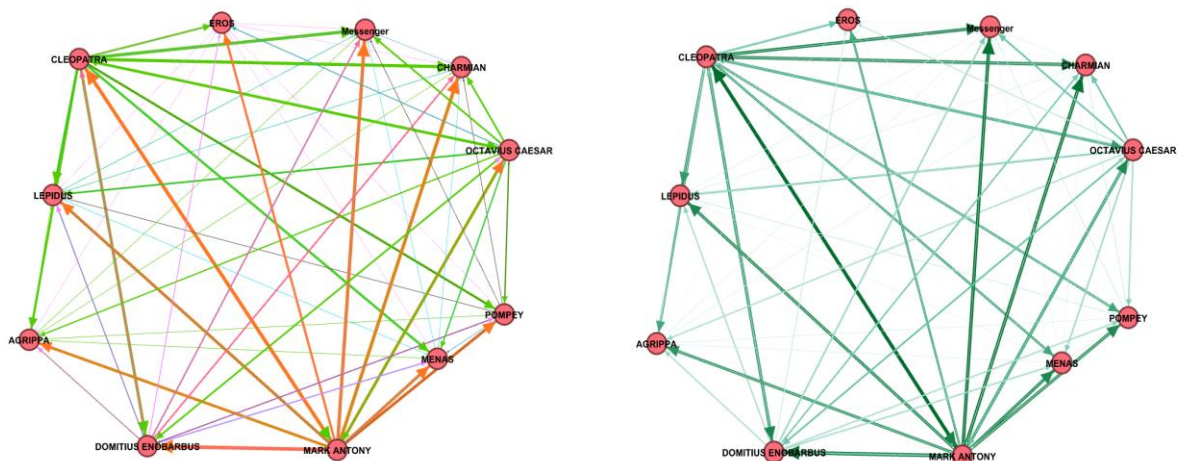


{'Love': 1, 'Grief': 2, 'Anger': 3, 'Caring': 4, 'Embarrassment': 5, 'Annoyance': 6, 'Disapproval': 7, 'Confusion': 8, 'Amusement': 9, 'Admiration': 10, 'Gratitude': 11, 'Optimism': 12, 'Sadness': 13, 'Approval': 14, 'Realization': 15, 'Disgust': 16, 'Fear': 17, 'Surprise': 18, 'Joy': 19, 'Remorse': 20, 'Desire': 21, 'Pride': 22, 'Disappointment': 23, 'Excitement': 24, 'Nervousness': 25, 'Relief': 26}

We can observe that 'Admiration' is the most frequently expressed emotion by all of the main characters throughout the play. While this may not provide a comprehensive representation of each character's emotional profile, it does offer insight into the primary emotions they express throughout the entire play. Additionally, if we are familiar with the play's storyline, we can make further observations. For example, in the final act, when Antony's death is known, we can see that the characters in this act experience shock and anger in response to his demise.

6. Analyzing Character Interactions and Emotions:

To represent the emotion expressed by one character when interacting with another, I had to create a weighted directed graph using the Networkx library. A character is linked to another character if they have interacted with one another. The interaction in this case is considered to be when two characters are present in the same scene. For example, characters X, Y & Z are present in Act 1 Scene 2, if X says a line, then X is linked with Y and Z. The more two characters interact with each other, the more weighted their edge is.



Admiration	(41.12%)
Approval	(16.82%)
Gratitude	(9.35%)
Anger	(9.35%)
Disapproval	(7.48%)
Caring	(7.48%)
Sadness	(3.74%)
Fear	(3.74%)
Annoyance	(0.93%)

The above graphs display the different dominant emotions each character expresses with another one and the level of interaction, for example we can see that Cleopatra and Antony share two edges where each one is dark green colored meaning, they both interact with one another a lot during the play, in this case they are together in a lot of scenes. This solution can be further improved in order to actually represent interactions between characters more accurately.

7. Conclusion & Final Thoughts:

This project serves as a foundational exploration of emotion detection in textual data. It focuses on three key aspects:

- Comparing the Text2Emotion and EmoRoBERTa models.
- Utilizing the EmoRoBERTa model to depict the emotional profiles and evolution of characters throughout the play.
- Representing character interactions through network graphs, visualizing the dominant emotions expressed during these interactions.

Further improvements can be implemented to better represent the emotional profiles and the accuracy of interactions between characters.