

# Fundamentals of Data Science Course

Faculty of Information Engineering, Informatics and Statistics,  
Sapienza University of Rome

Academic Year 2022-2023

## Sentimental Analysis Project



SAPIENZA  
UNIVERSITÀ DI ROMA

Group Participants:

Federico Barreca 1736423

Laura Mignella 1920520

Antonio Ratti 1700089

Maria Vittoria Vestini 1795724

Cosmin Zaharia 1805207



## Introduction

- Task Statement and Definitions
  - Analytical project for the building of a classification model through text analysis
- Data collecting and classification
  - Collecting Information :
    - The aim is to categorize each text with labels (Positive, Negative ) then perform an analysis of the tonality of text.
    - This will help to interpret better future comments or statements associating them with emotions

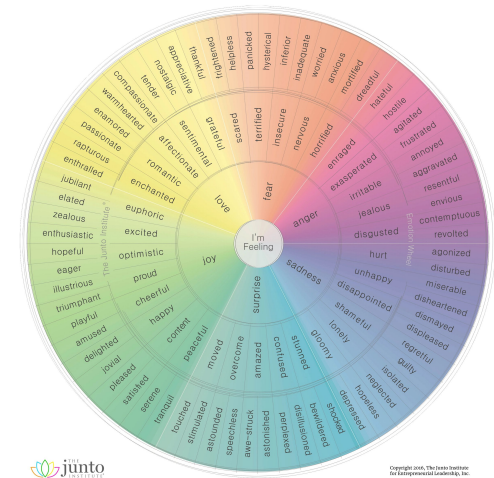


## Models, Tools, Data Manipulation

- Used Models
  - Logistic Regression
  - K-Nearest neighbors
  - Multinomial Classification (SoftMax method)
- Used Tools
  - Data Cleaning (Stopwords, demojize etc.)
- Data Manipulation
  - Grouping: Positive, Negative feelings
  - Binarizing Groups and labelling them
  - Obtaining X (text) and Y (labels)

# Development

- First Grouping
  - Mapping the starting 27 emotions in to 6 groups
  - Problems: low accuracy 50%

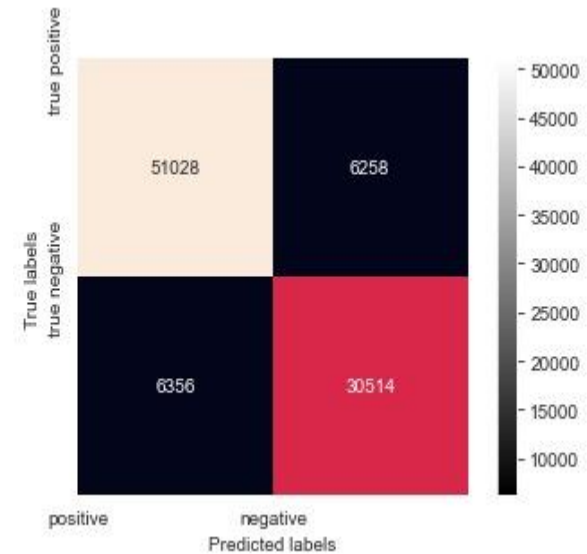


- Second Grouping
  - Mapping in to 3 groups positive, negative, ambiguous feelings
  - Problems: ambiguous reactions were too few respect the other two groups to be detected



## Analysis

- Benchmark
  - Logistic Regression:
    - Accuracy: 83%
  - K-Nearest neighbors:
    - Best result n=3
    - Accuracy: 79%
  - SoftMax method:
    - Accuracy: 82%





## Future work and References

- An increase in the number of samples could lead to a reevaluation of the case with 6 emotions.

## Link

- Dataset:  
<https://www.kaggle.com/datasets/shivamb/go-emotions-google-emotions-dataset>
- Human emotions:  
<https://www.visualcapitalist.com/a-visual-guide-to-human-emotion/>