Jedha

Data Science Bootcamp

“How do you feel today ?”

GoEmotions : text-based emotion detection

Ibrahim Benjelloun

Florian Akretche

Perrine Panisset

MOTIVATION

CONTEXT

- One of the many applications of **Natural Language Processing (NLP)**
- Text-based emotion detection is generally limited to a **small number of emotions** (6 to 12)
- **Difficulty of interpretation** due to subjectivity (sarcasm, irony…)

FIELDS OF APPLICATION

- **Social media analysis** in different areas (*product/brand reviews, hate speech, etc.*)
- **Mental health** *(emotional distress, suicidal thoughts, etc.)*
- **Personalized customer services**
- **Empathetic chatbots**

AMBITION

Building a **text classification model** that detects **one or multiple emotions** on a **large spectrum of emotions**

APPROACH

- **Data selection and exploration**
- **Data cleansing**
- **Building classification models**
- **Evaluation and performance analysis**

GOEMOTIONS DATASET

INTRODUCTION

Built by a **Google Research** team (subject of a research paper)

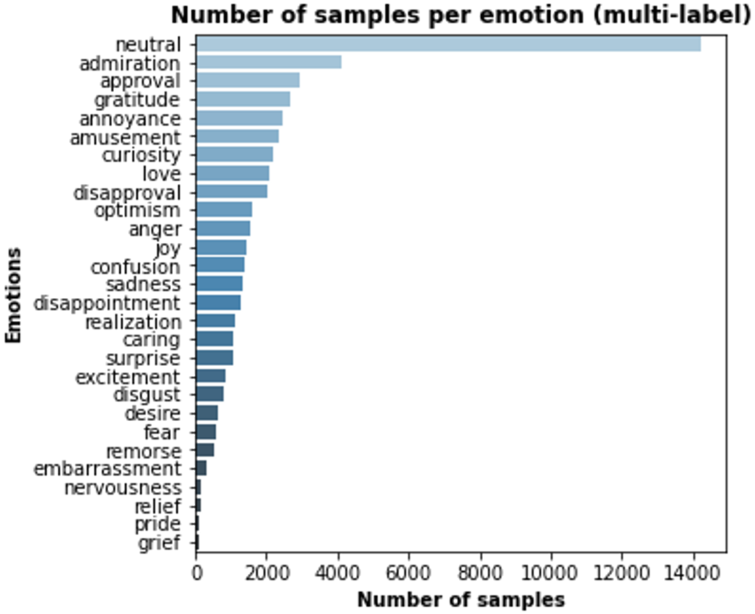
Gathers more than **58K Reddit comments (English)**

Largest manually labeled dataset

CHALLENGES

Class imbalance: ~30% of "neutral" samples

Multi-label: Up to 5 emotions for a single comment



EMOTIONS WHEEL

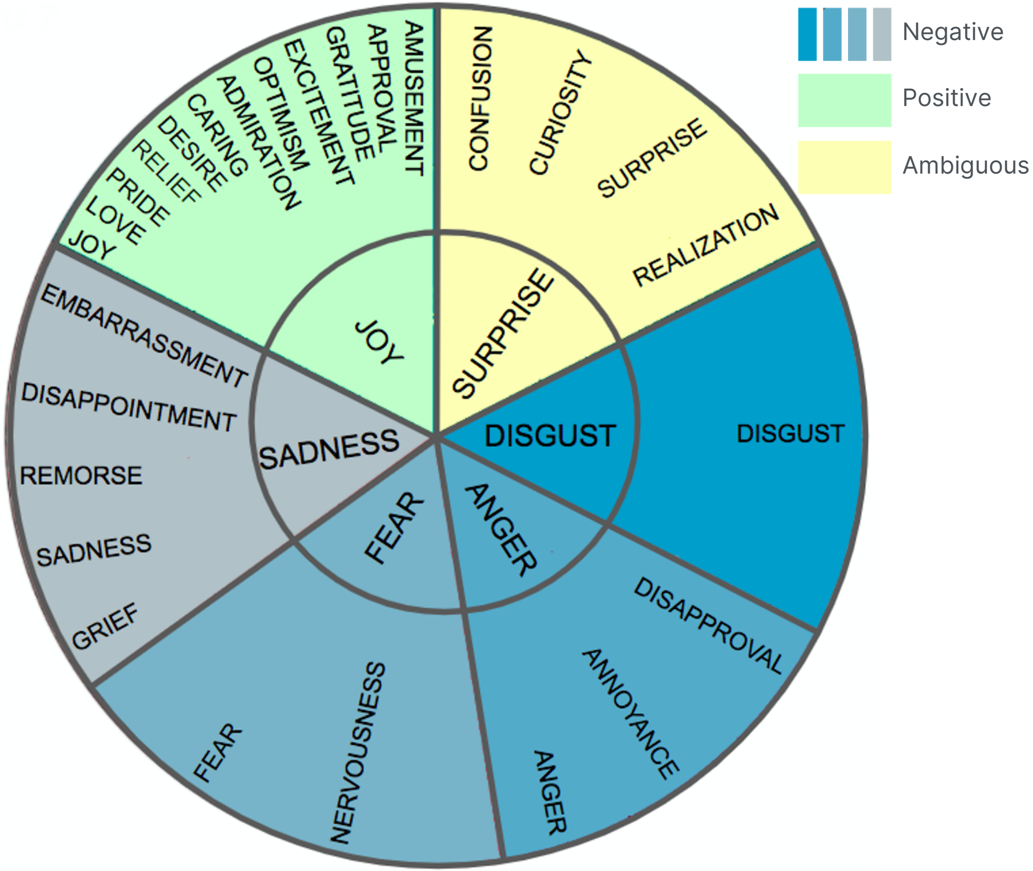
2 TAXONOMIES

GoEmotions *(27 emotions) + “neutral”*

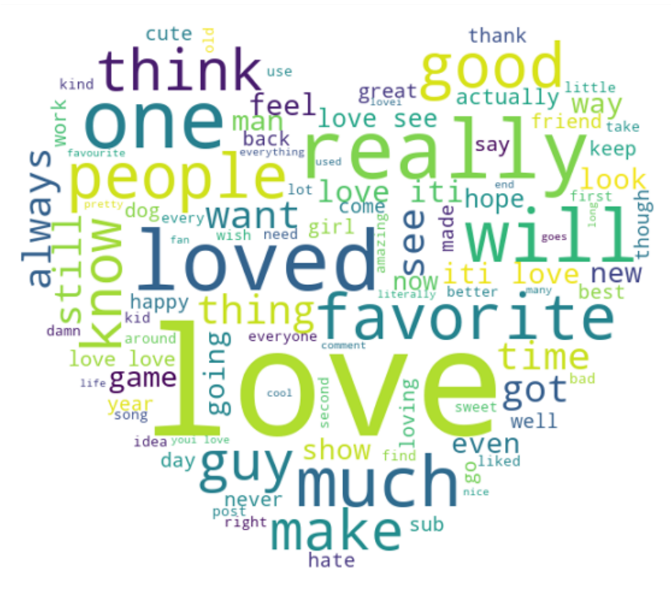
Ekman *(6 emotions) + “neutral”*

SCOPE OF STUDY

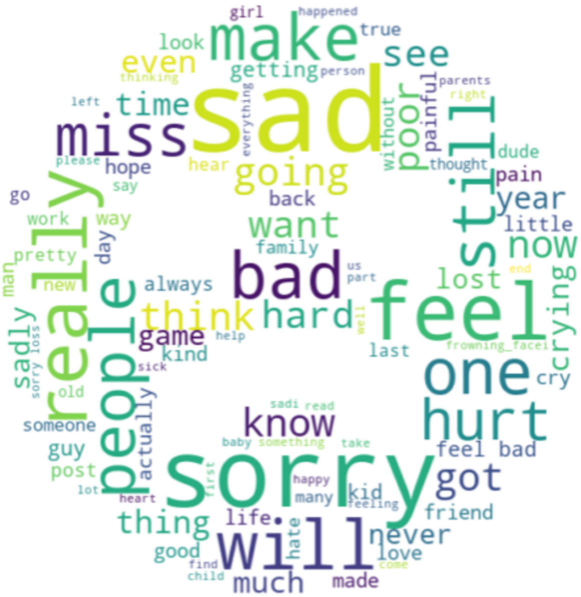
Emotions analysis (vs Sentiment analysis)
(Focus on GoEmotions taxonomy)



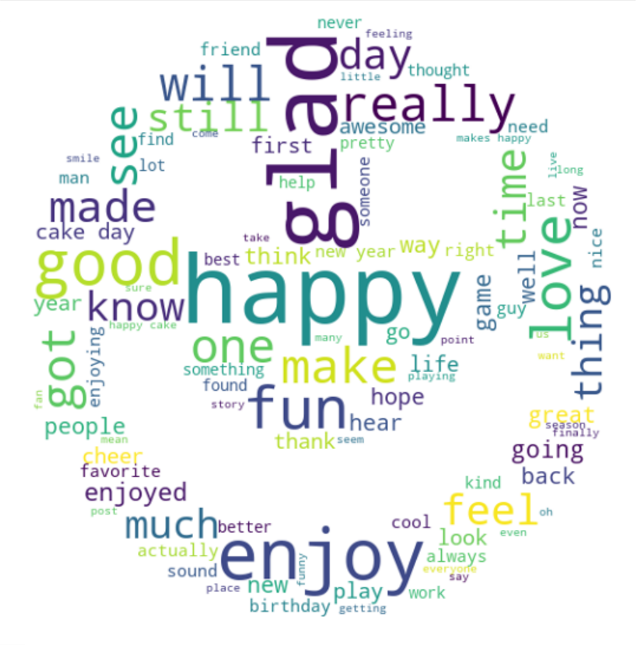
GOEMOTIONS DATASET - WordClouds



LOVE



SADNESS



JOY

STEP 1

spaCy

DATA CLEANING

DATA CLEANSING

- **EMOJIS**

  “demojize”

  👍 → *:thumbs_up:*

- **SPECIAL CHARACTERS / NUMERICAL**

  (#, @, … except “?” and “!”)

  !! 😠 VS !!!! 😡

  ?! 🤨 VS ??! 😱

- **CONTRACTIONS**

  Separate contractions

  We’re → *We are*

- **ACRONYMS**

  cuz → *because*

STEP 2

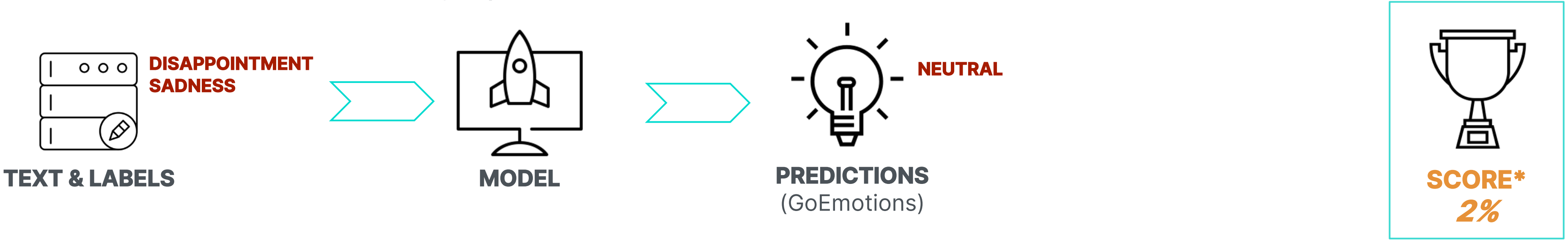
MODELING & EVALUATION

1ST Model	**Dummy model**
2nd Model	**Baseline model**
3rd Model	**BERT**

MODELING - Dummy & Baseline models

Comment sample : *“No one cares my guy”*

DUMMY MODEL - Always predicts “neutral”



DISAPPOINTMENT
SADNESS

TEXT & LABELS

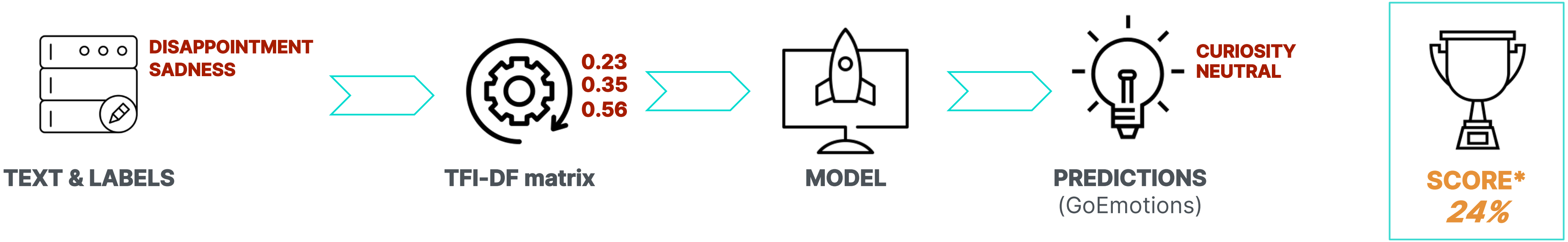
MODEL

NEUTRAL

PREDICTIONS
(GoEmotions)

SCORE*
2%

BASELINE MODEL - Machine Learning (Ridge Classification)



DISAPPOINTMENT
SADNESS

TEXT & LABELS

0.23
0.35
0.56

TFI-DF matrix

MODEL

CURIOSITY
NEUTRAL

PREDICTIONS
(GoEmotions)

SCORE*
24%

**on test data*

MODELING - BERT *(General information)*

PRESENTATION

BERT *(Bidirectional Encoder Representations from Transformers)*

Deep Learning model developed by Google for NLP tasks

Pre-trained on data extracted from **BooksCorpus** (800M words) and **English Wikipedia** (2,500M words)

Based on the **attention mechanism** (word contextualization)

ADVANTAGES

Very efficient

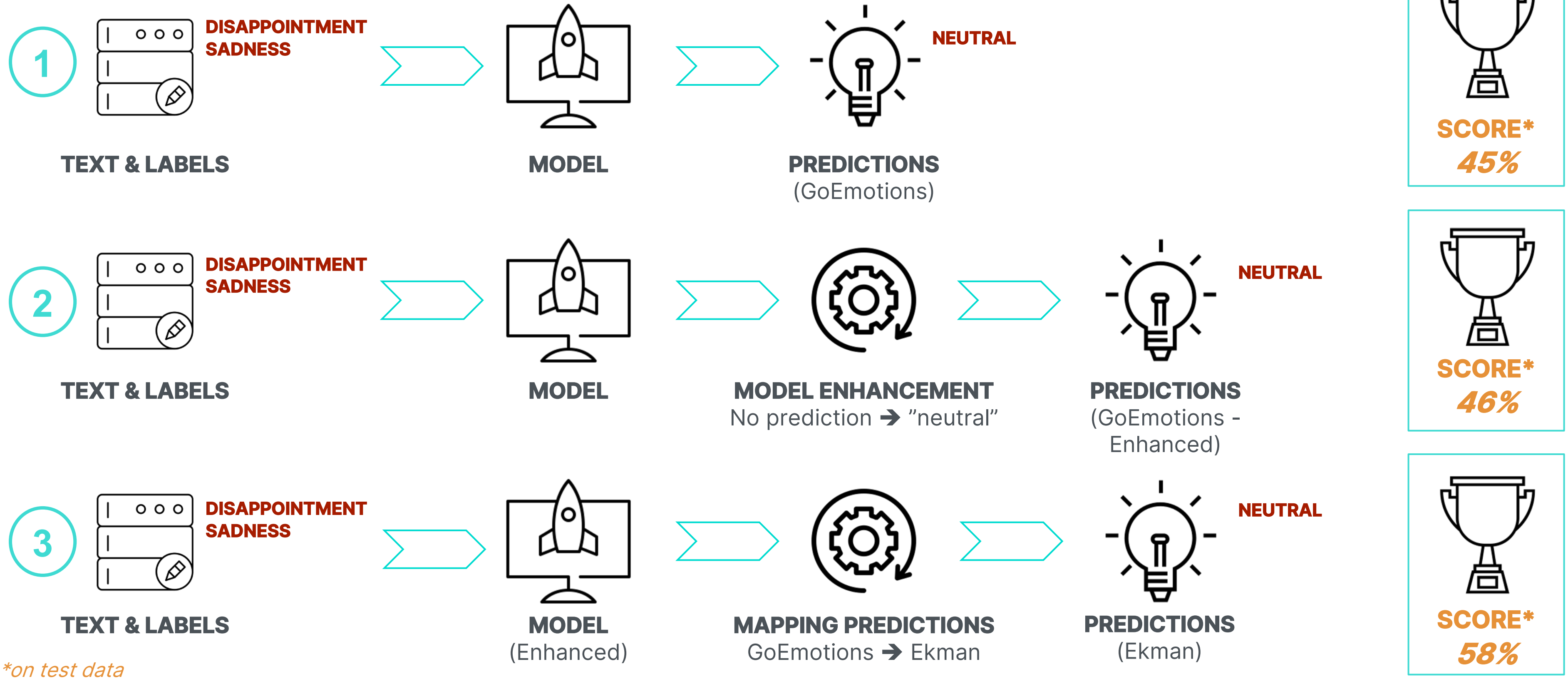
Keeps the **meaning of a sentence**

DISADVANTAGES

More than **100M trainable parameters** (base model)

MODELING - BERT *(Experiments)*

Comment sample : *"No one cares my guy"*



1

DISAPPOINTMENT
SADNESS

TEXT & LABELS

MODEL

NEUTRAL

PREDICTIONS
(GoEmotions)

SCORE*
45%

2

DISAPPOINTMENT
SADNESS

TEXT & LABELS

MODEL

MODEL ENHANCEMENT
No prediction ➔ "neutral"

NEUTRAL

PREDICTIONS
(GoEmotions - Enhanced)

SCORE*
46%

3

DISAPPOINTMENT
SADNESS

TEXT & LABELS

MODEL
(Enhanced)

MAPPING PREDICTIONS
GoEmotions ➔ Ekman

NEUTRAL

PREDICTIONS
(Ekman)

SCORE*
58%

**on test data*

MODELING - BERT *(Garbage in ... Garbage out ?)*

PROBLEM

"How was the problem resolved??? Having the same issue????"

NEUTRAL

TEXT & LABELS

VS

CONFUSION
CURIOSITY

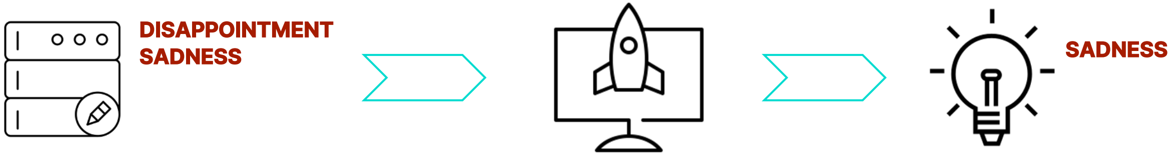
PREDICTIONS

INTERPRETATION

- The **"neutral" class** was used as a **"garbage" class** at the time of labeling
- The "neutral" class **adds noise to our data**

SOLUTION

Comment sample : *"No one cares my guy"*



DELETING NEUTRAL SAMPLES
(-30% of samples in the dataset)

MODEL

PREDICTIONS

SCORE*
53%

*on test data

CONCLUSION

SUBJECTIVITY BIAS

- In the **expression** and **interpretation** of emotions in a text
- In the **labelling**
- In **our evaluation of the detected emotions**

PERFORMANCES

INITIAL SCORE*
46%

- **Exceeded our expectations**
- Similar score to Google's research paper
- A large potential for improvement

THE CHERRY ON TOP

Training on "non-neutral" samples allows to

- **Better distinguish emotions**
- **Detect a "neutral" emotion a posteriori**

SCORE*
53%

**on test data*

POTENTIAL IMPROVEMENTS

- **Enhance the data cleaning** phase

- **Review training labels** *(False "neutral" samples, mislabeled samples, etc.)*

- Find **more data**, more **diversified** and more **representative of the general population**

- **Try other algorithms** *(GPT-2, RoBERTa, XLNet...)*

ÉTAPE 3

DÉMONSTRATION

(Web app: My Annoying Shrink)

Jedha

Thank you !