

DETECTION D'ÉMOTIONS DANS UN TEXTE

Marius ADJEVI
Richard ALLADO

Encadré par **Ayoub Esswidi**



CONTEXTE

Détection des émotions d'un texte grâce au
Machine Learning et au Natural Language
Processing (NLP)



DOMAINES D'APPLICATION

- Analyse des réseaux sociaux (les commentaires sur un produit par exemple)
- Santé Mentale (pensées suicidaires, stress émotionnel...)
- Chatbots empathique



APPROCHE

1. Data cleaning and Preprocessing
2. Construction des modèles
3. Evaluation et analyse des performances



GOEMOTIONS

GoEmotions est un corpus de 58 000 commentaires soigneusement sélectionnés, extraits de Reddit, et annotés par des humains selon 27 catégories d'émotions ou la catégorie Neutre.

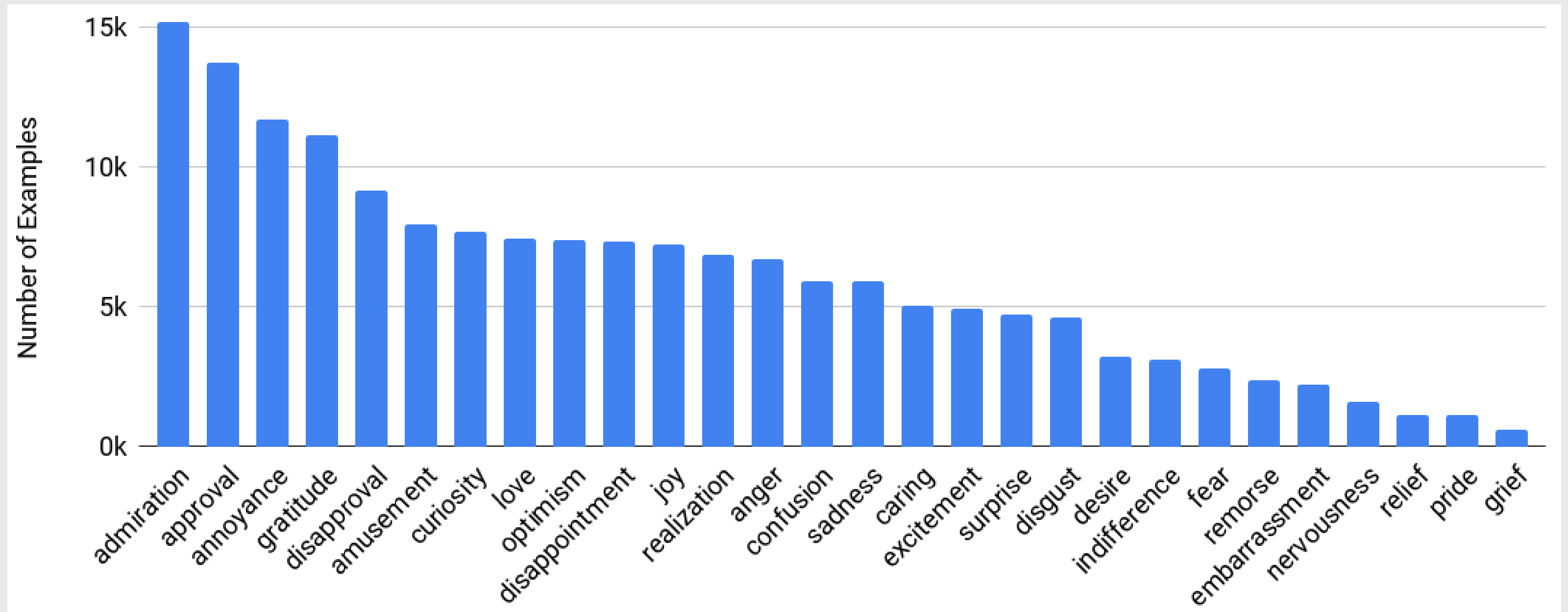
Il a été construit par une équipe de recherche de **Google**



LIMITES DE LA DATASET

- **Inégalité des classes** : ~30% des entrées ont la classe “neutre”
- **Multi-label** : nous avons jusqu’à 5 emotions pour une seule entrée

GOEMOTIONS





DATA CLEANING

- **Emojis.** Ex :
- **Contractions et Acronymes** . Ex : we're → we are, cuz → because
- **Punctuation**
- **Majuscules**
- etc



MODÈLES

- **LogisticRegression** : 3%
- **SGBClassifier** : 24%
- **LinearSVC** : 32%

MODÈLES : BERT

BERT est un modèle de langage basé sur l'architecture Transformer, développé par Google en 2018.

Points forts :

- **Bidirectionnel** : comprend le contexte complet autour d'un mot.
- **Transférable** : le pré-entraînement sur un large corpus (Wikipedia + BookCorpus) permet de fine-tuner facilement sur diverses tâches.
- **Performant**

MODÈLES : BERT

Points faibles :

- **Taille et ressources**
- **Limites contextuelles** : Longueur de séquence, Pas de compréhension du langage humain
- **Pré-entraînement et biais** : pré-entraîné sur zikipédia et BookCorppus

Résultat : 45~58%



APPLICATION



THANK YOU