



Foodlytics

# Foodyltics

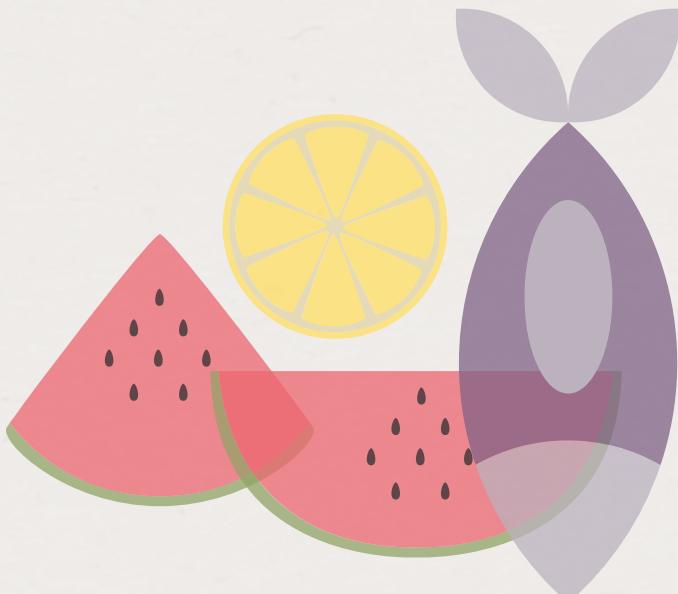
Trusted analysis to improve your restaurant!

---

T5 DATA SCIENCE BOOTCAMP

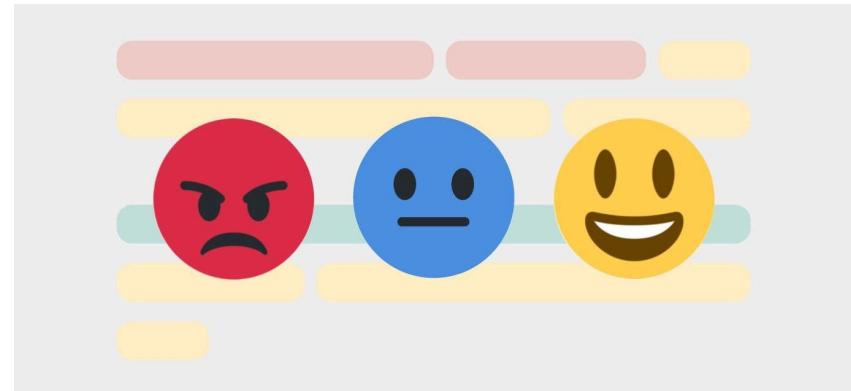
By:

Yazeed Musallam



# Introduction

- **What is Foodlytics?**
- **Problem Statement**
- **Value of Foodlytics company:**
  - Is the review positive or negative?
  - From user reviews, how we can enhance our restaurant?



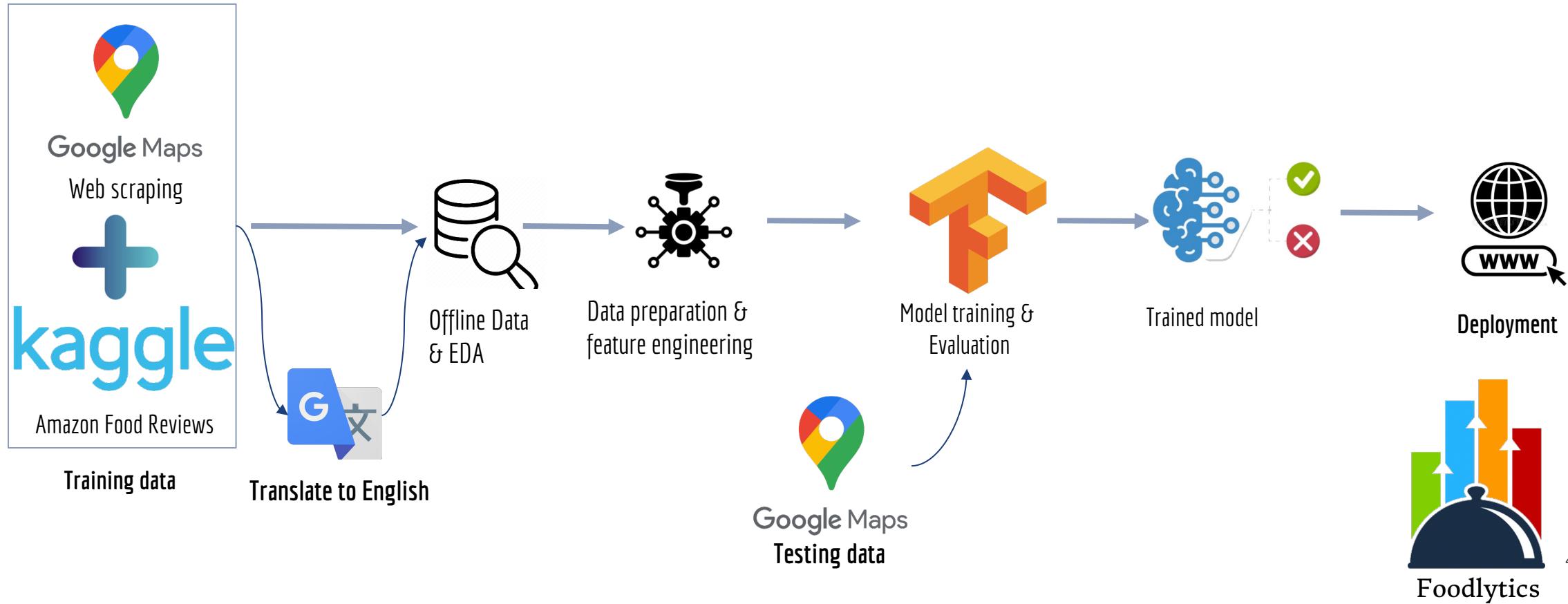
## Example of reviews

|                      | Rate: |                               |
|----------------------|-------|-------------------------------|
| .. افضل مطعم بالرياض | 5     | The best restaurant in Riyadh |
| الاكل غير جيد        | 2     | The food was bad              |
| لا اعرف المطعم       | 1     | I don't know the restaurant   |

Outlier!



# Foodyltics Architecture



## Dataset Description and

### Google Maps Dataset (scraped):

- 270 restaurants
- 100k reviews



Google Maps  
Web scraping

### Amazon Food Reviews (Kaggle):

- Food product reviews
- 100k reviews

kaggle

Amazon Food Reviews



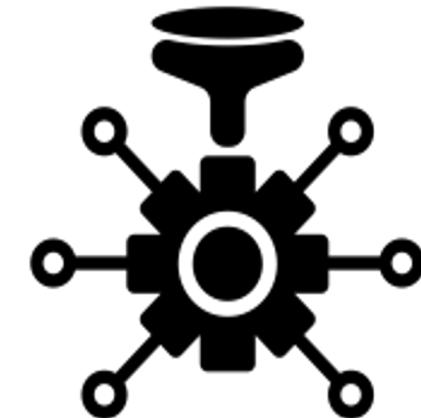
# Data Preparation & Cleaning

## Things removed:

- Duplicates: 5k
- Deconcatenation: won't ==> will not
- Special character: \$#@
- HTML tags: </br>
- Stop words: why, how ..

## Scoring:

- More than 3: Positive review
- less than 3: Negative review

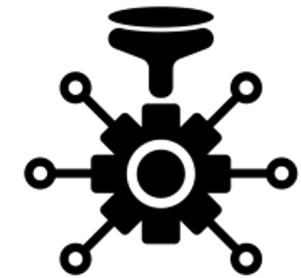
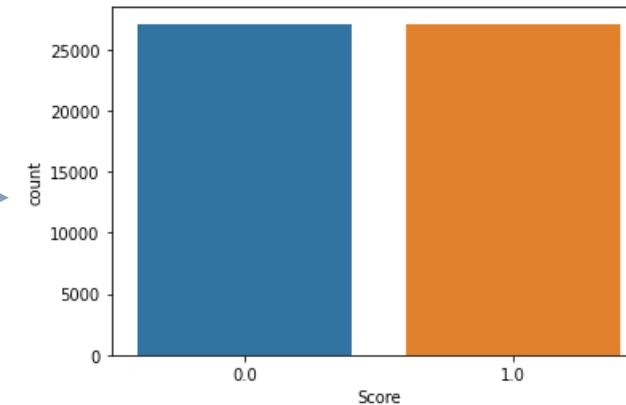
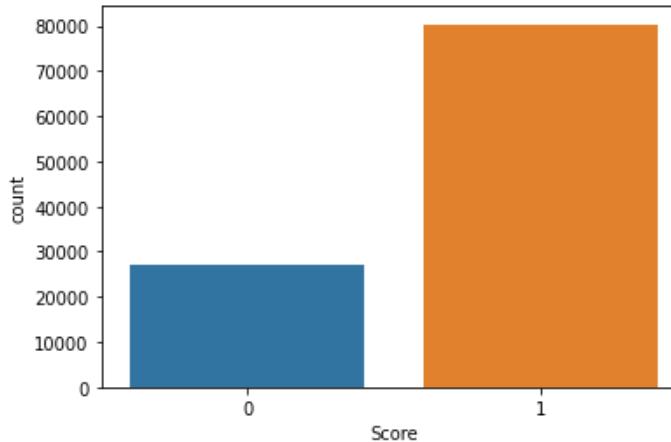


# Feature Engineering & Data Balancing

**Features added:**

- **Number of:** words, unique words, characters ...

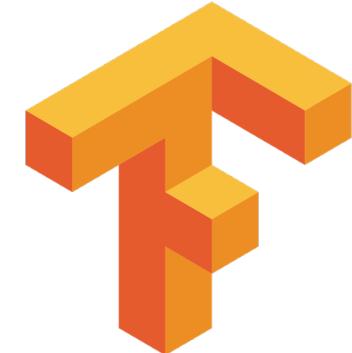
**Data Balancing:**



# Model Training & Evaluation

## Data split:

- **Training:** 65% of Google maps reviews + Kaggle dataset
- **Validation:** 15% of Google maps reviews
- **Testing:** 20% of Google maps reviews



# Model training & Evaluation

Tested models:

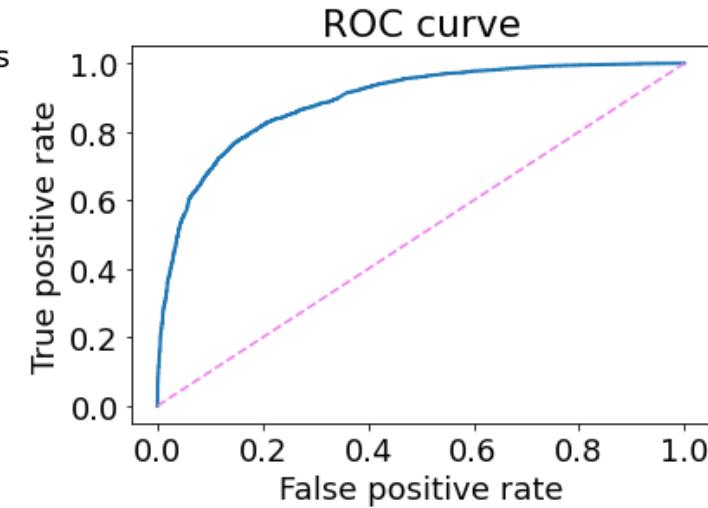
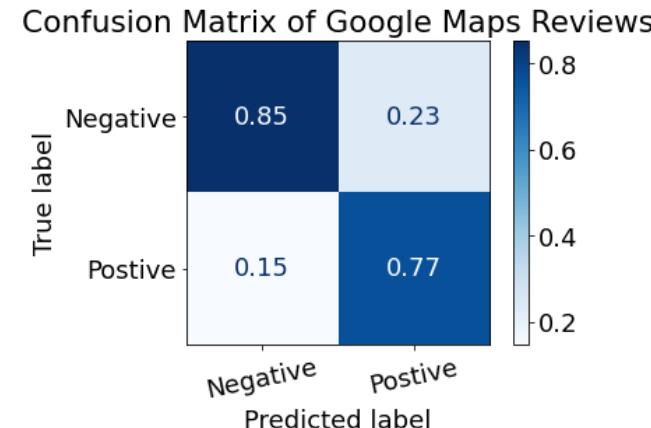
| Model                     | F1-SCORE  |
|---------------------------|-----------|
| Naive Bayes<br>(Baseline) | 81        |
| Linear regression         | 81        |
| Decision Tree             | 74        |
| RandomForest              | 82        |
| XGboost                   | 81        |
| Long short-term<br>memory | 82        |
| <b>BERT (The best)!</b>   | <b>83</b> |



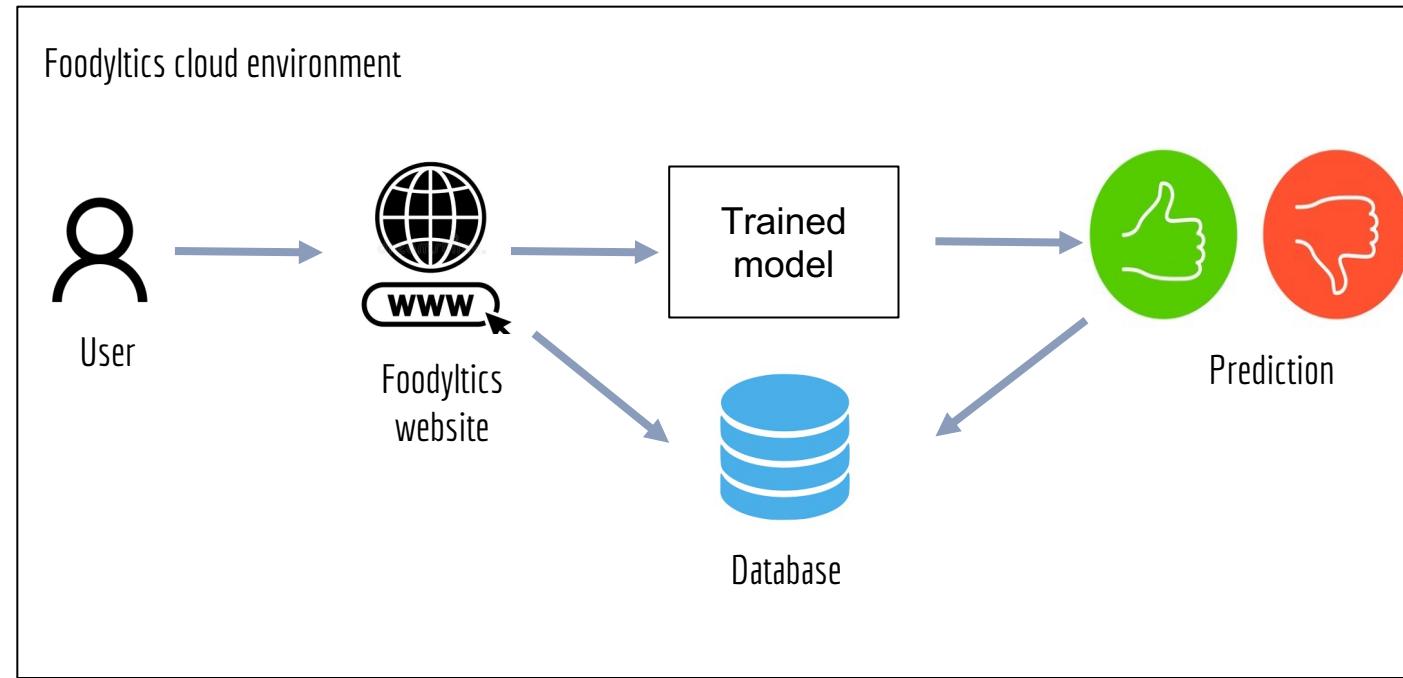
# BERT

## Results:

- **Training accuracy :** 80%
- **Validation accuracy :** 81.3%
- **Testing accuracy :** 81.2%
- **F1 Score:** 83%
- **ROC-AUC score:** 89



# Foodyltics



# Demo



\*Please try it with us

## Future works

- 1- Connect to Twitter, Hungerstation, etc..**
- 2- Enhance the performance**



\*Please try it with us

# Thank you !

## Any Questions?