YELP Sentiment Model

Krishna Thakar

# Objective:

The primary goal of this sentiment analysis is to extract meaningful insights from customer reviews to help businesses improve their services. Specifically, this analysis aims to:

* Identify Customer Sentiment: Classify reviews into positive, neutral, or negative categories.
* Analyze Common Complaints: Discover recurring negative feedback that affects customer satisfaction.
* Highlight Positive Aspects: Identify what customers love about certain restaurants, such as food quality, service, or ambiance.
* Provide Data-Driven Insights for Business Growth: Help businesses make informed decisions to enhance customer satisfaction and loyalty.

# Data Collection:

1. Source: Official Yelp website - [Open Dataset | Yelp Data Licensing](https://business.yelp.com/data/resources/open-dataset/)
2. Volume: 7 million reviews
3. Timeframe: 2005 - 2022

# Preprocessing:

1. Text Cleaning:
   1. Remove stop words, punctuation, and special characters.
   2. Convert text to lowercase.
   3. Handle contractions (e.g., "don't" → "do not").
2. Tokenization: Split text into individual words or phrases.
3. Lemmatization/Stemming: Reduce words to their base or root form.
4. Handling Negations: Address negations (e.g., "not good" → "not\_good").
5. Vectorization: Convert text into numerical formats (TF\_IDF)

# Modeling:

1. Traditional Methods:
   1. Logistic
   2. SVM
   3. Naïve Bayes
   4. Random Forest
2. Deep Learning Based:
   1. Embeddings + LSTM
   2. Embeddings + CNN
3. Transformers Based - BERT