3000 reviews

Training 80% = 2400

Validation 20% = 600

Get most common words (value > 50)

Classify each word manually into 4 categories: food, service, decoration, cleanliness -> 8 value

Each review -> each sentence -> search for categories -> find sentiment value -> get highest absolute value

Final steps:

Each review -> sentiment categories -> business id

Business id -> 4 categories -> 8 value (positive, negative \* 4 categories) 🡺 final files

One sentence => one category, one sentiment value (multiple sentiment -> average and round)

Negative word