



Categorization of hotel reviews into different customer values

Presented by Group 18



Meet the Group



Divyesh Rohit - 201801086

Ishani Bandyopadhyay - 201801102

Janvi Patel - 201801178

Tulsi Shah - 201801180

Skand Vala - 201801210

Introduction

- This system is used by several websites to identify the sentiments of the customers.
- To save the time of reading hundreds of reviews, we did some preliminary work.
- Implemented an algorithm that rates a review based on specific keywords to reflect the customer's sentiment



Motivation

General Overview

The main idea behind this topic came from trip planning while making an online hotel reservation.

Research

- Research Paper on "Sentimental Analysis of IMDb reviews" on the Stanford website.
- Shortcomings: Reviews that have a positive and negative sentiment, makes it difficult for the user to get an accurate feedback

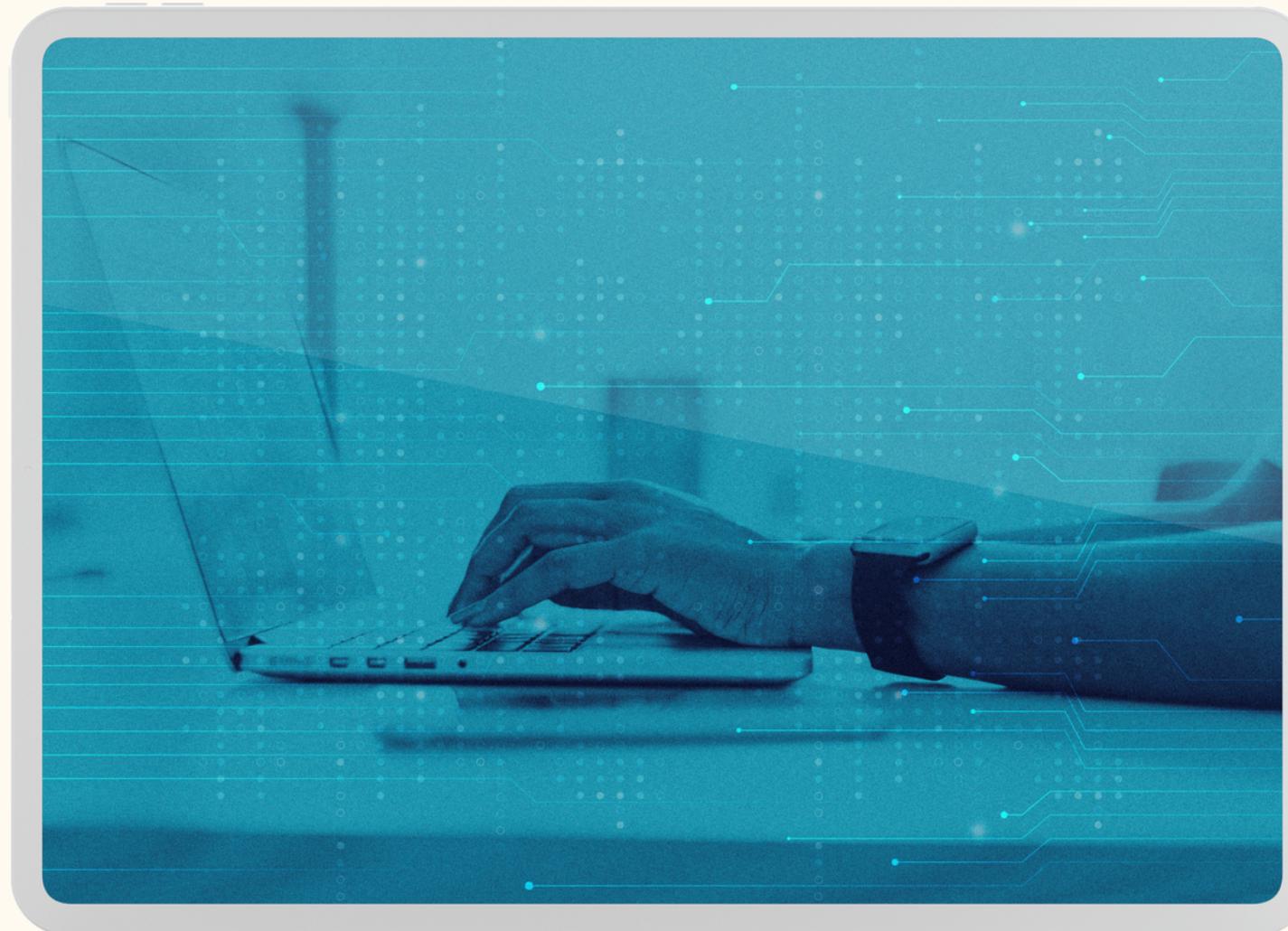


Problem Statement

Our aim is to classify hotel review *text* into different categories of importance and rate them whether, how positively or negatively has been the user about the hotel for different amenities.



Datasets

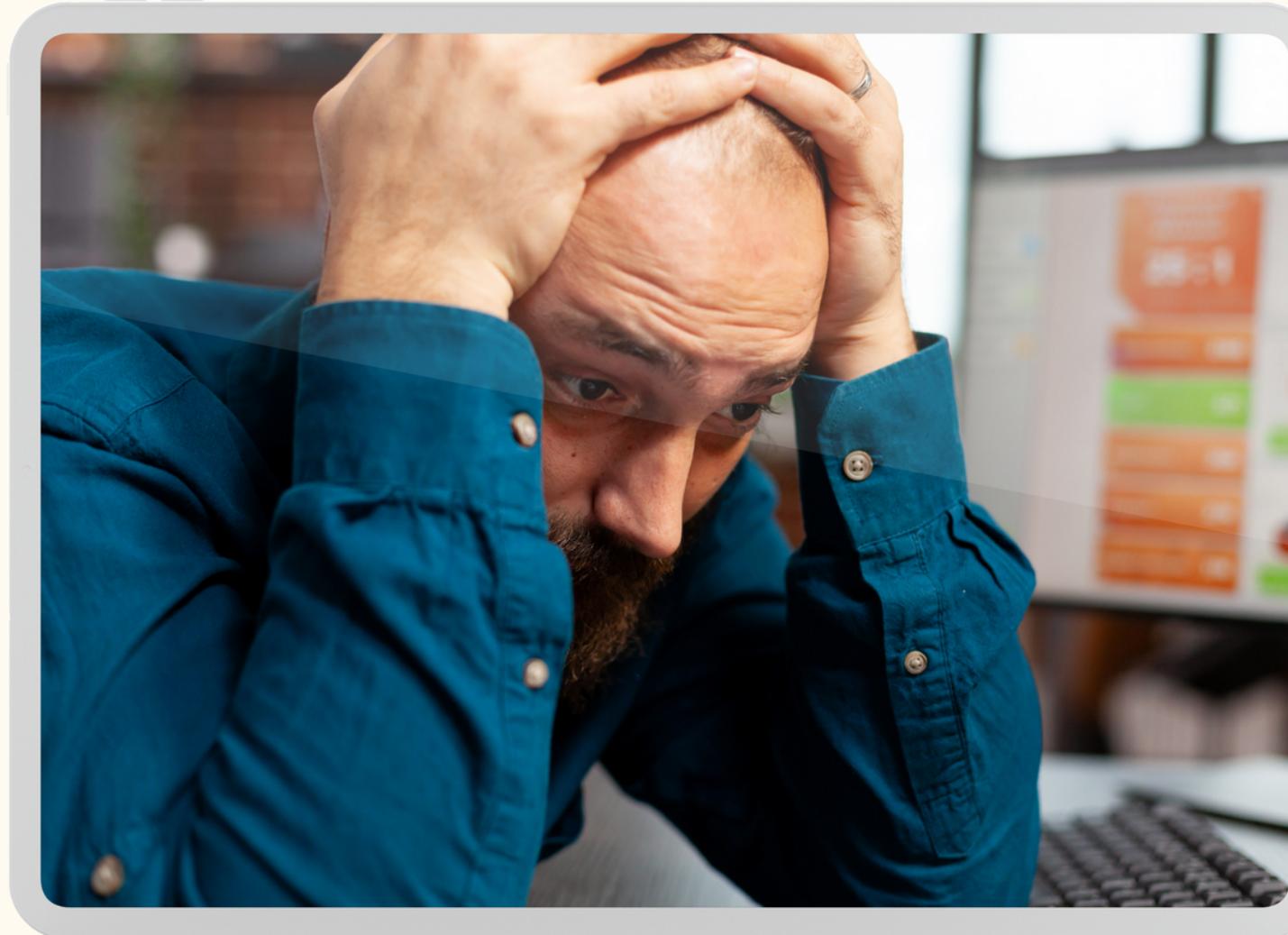


Deciding key categories: Analyzed and scoured through around 10-20 websites containing hundreds of reviews on each website.

We built our own dataset: around 1600 reviews, which were manually reviewed, rated and classified categorically by us.

5 key rating categories: food, safety and hygiene, location, value for money, and hospitality.

Datasets



Category rating:
Bad- 0, Average-1, Good-2

Tedious and time-consuming: Each review needed to be classified into five categories.

Dataset accuracy: the slightest error in rating the reviews categorically, will give an incorrect output.

Methodology



Data Cleaning

Dropping special chars and #'s

removing full stops, commas, ?, #, numbers etc.

EG: Wow....Loved this place.

---> Wow Loved this place

Converting to small-case

---> wow loved this place

Dropping Stopwords (i.e. common words)

---> wow loved place

Stemming (words --> their root)

---> wow love place

Naive Bayes

effective and simple algorithm with reasonable accuracy. Eg: spam filter classification problem

Logistic Regression

the classic method for classification problems

KNN Algorithm

assumes the similarity between new cases and the existing cases. Classification is done based on the similarity

Using different ML concepts

Front-end

Hotel Reviews		Naive Bayes	Review Analysis				Search...
HOTEL NAME	FOOD		LINE	LOCATION	VALUE FOR MONEY	HOSPITALITY	RATINGS
Hotel Russo Palace	3.0	Naive Bayes k-Nearest Neighbors Logistic Regression	3.5	2.9	3.8	 3.4	
Americas Best Value Inn	3.1	2.9	3.7	2.4	2.8	 3.0	
Intermezzo	1.0	1.0	1.0	1.0	1.0	 1.0	
Studio 6	1.0	1.5	1.0	1.0	1.5	 1.2	
Little Paradise Hotel	3.7	4.1	4.0	4.0	4.4	 4.0	

Front-end

Enter your review

The staff at this property are all great! They all go above and beyond to make your stay comfortable. Please (HN) give your staff awards!

Apply

MODEL NAME	FOOD	SAFETY AND HYGIENE	LOCATION	VALUE FOR MONEY	HOSPITALITY	RATINGS
Naive Bayes	3	5	3	5	5	4.2
K-Nearest Neighbours	3	5	5	3	5	4.2
Logistic Regression	3	5	5	3	5	4.2

Results

1

Naive Bayes

- Food -> 0.774
- Safety and Hygiene -> 0.765
- Location -> 0.734
- Value for money -> 0.777
- Hospitality -> 0.783

2

K nearest neighbour

- Food -> 0.594
- Safety and Hygiene -> 0.643
- Location -> 0.634
- Value for money -> 0.548
- Hospitality -> 0.676

3

Logistic regression

- Food -> 0.898
- Safety and Hygiene -> 0.899
- Location -> 0.913
- Value for money -> 0.870
- Hospitality -> 0.930



Thank you!

Do you have any questions for us?