



TripAdvisor Hotel Reviews Classification and Summarization

using IBM Granite

*Student Developer Initiative by
Hacktiv8 x IBM SkillsBuild*

next slide →

02



Data Source

Data dalam Project ini dapat diakses secara publik melalui tautan berikut.

https://www.kaggle.com/datasets/andrewmvd/trip-advisor-hotel-reviews?select=tripadvisor_hotel_reviews.csv

next slide →



Project Overview

Industri perhotelan merupakan salah satu sektor ekonomi krusial industri pariwisata dengan fokus pada penyediaan akomodasi, makanan dan minuman, serta layanan terkait bagi wisatawan. ***Keberhasilan industri perhotelan sangat bergantung pada ulasan dan penilaian tamu yang mencerminkan kualitas layanan, reputasi, dan citra hotel.*** Dengan kemajuan teknologi, saat ini tamu dapat dengan mudah memberikan ulasan dan penilaian melalui berbagai platform digital seperti TripAdvisor. Namun, ***ulasan sebagai data teks merupakan data yang tidak terstruktur, sehingga diperlukan analisis sentimen.*** Di sisi lain, penilaian (*rating*) sebagai data terstruktur hanya menggambarkan pengalaman tamu secara global, sehingga ***diperlukan analisis sentimen untuk memahami pengalaman tamu dalam menikmati kualitas layanan hotel secara komprehensif.***





Project Objective

Beberapa tujuan dari mengklasifikasikan ulasan tamu secara otomatis adalah sebagai berikut.

1. Mengukur kepuasan tamu secara otomatis berdasarkan ulasan yang masuk,
2. Mendeteksi adanya pola kepuasan dan keluhan yang disampaikan tamu,
3. Menyediakan metrik kuantitatif, misalnya melalui persentase tamu yang merasa puas dengan kualitas layanan hotel yang ditunjukkan dengan rating 4-5 dari skala 1-5.

Sementara itu, tujuan dari melakukan rangkuman dari ulasan tamu secara otomatis adalah sebagai berikut.

1. Mengidentifikasi kelebihan dan kekurangan dari hotel yang dirasakan oleh tamu,
2. Menarik intisari dari berbagai ulasan yang masuk, sehingga pihak hotel dapat mengambil kebijakan dan rekomendasi yang sesuai secara efisien,
3. Memperkuat strategi pemasaran, dengan melakukan branding pada ringkasan ulasan positif tamu.



Analysis Process

1

BRAINSTORMING

Penentuan topik yang digunakan dalam Capstone Project.

2

DATA COLLECTING

Pencarian data berbasis Teks yang bersifat publik (open-source)

3

DATA PREPROCESSING

Sampling data & rating categorized untuk digunakan sebagai Data Input pada Model

4

EXPLORATORY DATA ANALYSIS

Memvisualisasikan data ulasan yang terpilih menjadi Data Input dalam Wordcloud

5

SUBMITTING DATA TO THE IBM GRANITE MODEL

6

EVALUATE THE RESULTS

Mengevaluasi hasil klasifikasi yang dilakukan model dengan klasifikasi aktual pada dataset

Data Preprocessing

Sebelum masuk pada tahap analisis data, diperlukan beberapa langkah pra-pemrosesan data, sehingga analisis dapat dijalankan dengan efisien. Terdapat dua tahapan preprocessing data secara garis besar, yaitu *Data Sampling* dan *Categorized the Review*.

- *Data Sampling* dilakukan untuk menentukan data input secara efisien, sehingga pemrosesan pada model menjadi lebih optimal.
- *Categorized the Review* digunakan untuk mengategorikan *rating* (penilaian) dari tamu atau wisatawan sebagai label yang akan diprediksi model.

Data Preprocessing

Categorized the Review dilakukan untuk mengategorikan rating dari ulasan tamu sebagai berikut.

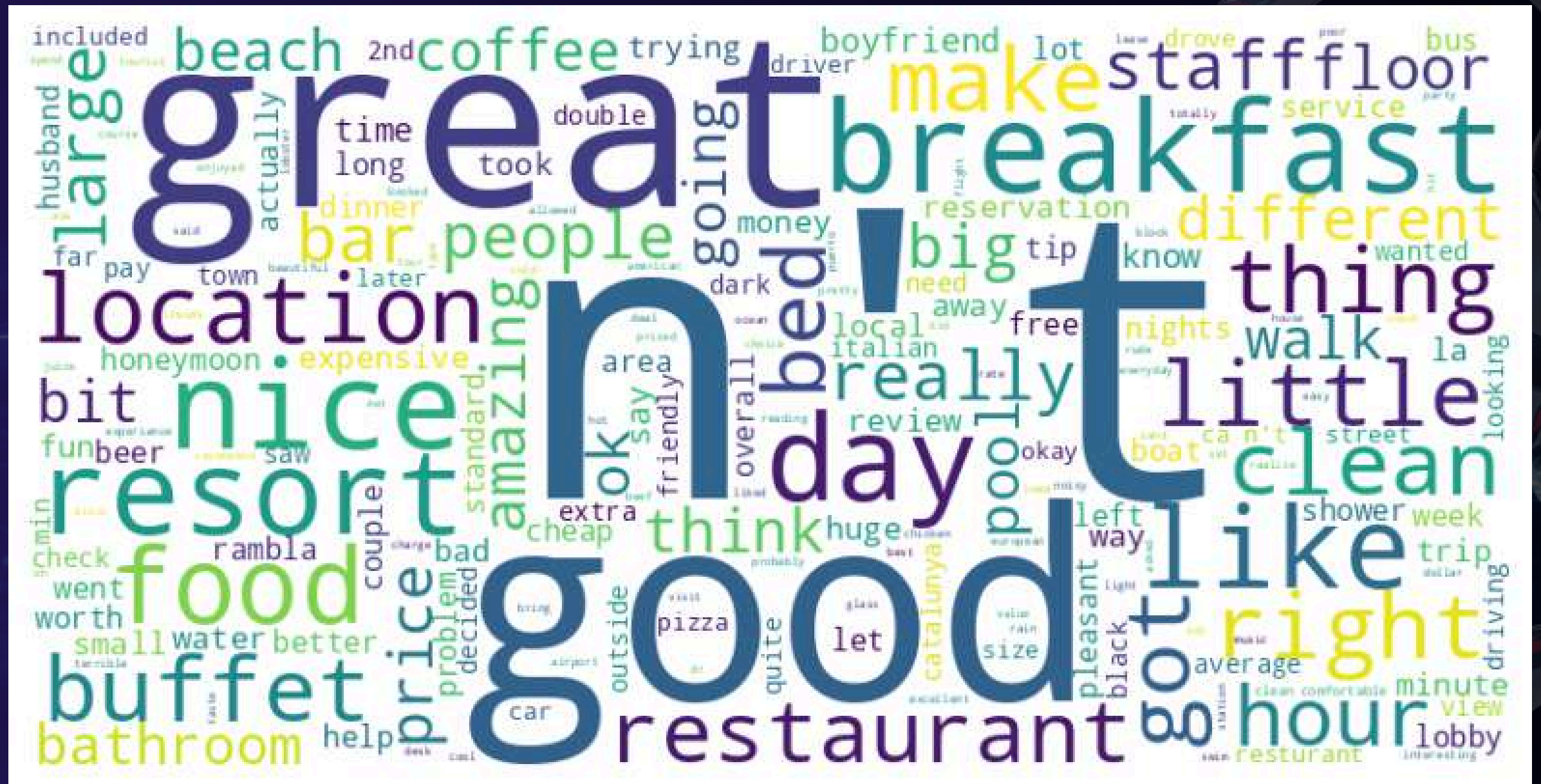
- 1-2: Negatif.
- 3: Netral.
- 4-5: Positif.

Word Cloud pada ulasan dengan Label **Aktual Positif**, terlihat bahwa kata dengan kemunculan tertinggi adalah “Good”, “Nice”, dan “Great”.



Insight & Findings (1): Word Cloud

Word Cloud pada ulasan dengan Label **Aktual Netral**, terlihat bahwa **kata** dengan **kemunculan tertinggi** adalah **"n't"** yang merupakan singkatan dari **"don't"**, **"can't"**, dan **"won't"**.



Word Cloud pada ulasan dengan Label **Aktual Negatif**, terlihat bahwa kata dengan **kemunculan tertinggi** adalah **"Bad", "Staff" "Know", dan "n't"**.



Insight & Findings (1): Word Cloud

- Berdasarkan tiga gambar *Word Cloud* yang disediakan, tampak bahwa ***ulasan tamu bersifat saling tumpang tindih***. Hal ini terlihat dari kemunculan kata atau tokenisasi berupa “**Good**” pada *Word Cloud* dengan Label Aktual Negatif, ataupun tokenisasi berupa “**n’t**” yang muncul pada *Word Cloud* dengan Label Aktual Positif.
- Hal ini menjadi tantangan tersendiri, terutama dalam menangkap *insight* bagi pihak pengelola hotel. *Word Cloud* hanya menunjukkan frekuensi kata, bukan hubungan antar kata atau sentimen sebenarnya.
- Kata yang sering muncul belum tentu paling penting atau paling mewakili keluhan/pujian utama.

Insight & Findings (2): Model Accuracy

Setelah dilakukan prediksi kelas menggunakan IBM Granite Model pada 150 sampel ulasan tamu dari *dataset* yang diperoleh, dapat dibentuk *Confusion Matrix* berukuran 3 x 3 sebagai berikut.

Ulasan Aktual	Hasil Prediksi Ulasan			Total
	Negatif	Netral	Positif	
Negatif	22	2	0	24
Netral	4	5	7	16
Positif	4	5	101	110
Total	30	12	108	150

Insight & Findings (2): Model Accuracy

Berdasarkan Confusion Matrix pada slide sebelumnya, dapat diperoleh metrik Evaluasi berupa Akurasi, *Precision*, *Recall*, dan *F1 Score* melalui *output* hasil *running syntax Google Colab* sebagai berikut.

```
Accuracy: 0.8533333333333333
```

```
Classification Report:
```

	precision	recall	f1-score	support
Negative	0.73	0.92	0.81	24
Neutral	0.42	0.31	0.36	16
Positive	0.94	0.92	0.93	110
accuracy			0.85	150
macro avg	0.70	0.72	0.70	150
weighted avg	0.85	0.85	0.85	150

Insight & Findings (2): Model Accuracy

- Berdasarkan hasil *Confusion Matrix*, tampak bahwa ***ulasan tamu dengan kelas aktual Netral*** memiliki kriteria kebaikan model yang lebih rendah dibandingkan dengan kelas lain. Hal ini dapat terjadi sebagai akibat dari kerancuan ulasan dari tamu, di mana ulasan dengan kategori netral dapat berisi kata atau frasa yang mengarah pada sentimen tertentu, baik negatif maupun positif, sehingga berakibat pada penurunan kriteria.
- Kriteria kebaikan model pada kelas aktual Netral yang lebih rendah dibandingkan dengan kelas lain tidak disebabkan karena *imbalanced data*, karena dalam proyek ini tidak dilakukan *training* model.

Text Summarization

Berikut adalah hasil Text Summarization dari ulasan dengan kelas Positif:

- *Customers appreciate the hotel's location, often praising its proximity to major attractions, shopping districts, and public transportation.*
- *The cleanliness and comfort of the rooms are frequently highlighted, with mentions of spaciousness, modern decor, and comfortable beds.*
- *Staff are consistently praised for their friendliness, helpfulness, and willingness to assist with recommendations for local restaurants, activities, and sightseeing.*
- *The quality of food, particularly breakfast, is often commended, with mentions of a wide variety of options and good taste.*
- *Some hotels are appreciated for their unique features, such as rooftop bars, pools, and spa services.*
- *Positive reviews often mention the value for money, with customers feeling they received excellent service and amenities for the price.*



Text Summarization

Berikut adalah hasil Text Summarization dari ulasan dengan kelas Netral:

- *Customers appreciate the hotel's location, often praising its proximity to major attractions, shopping districts, and public transportation.*
- *Some customers express mixed feelings about the hotel's design or decor, noting that while it's stylish, it might not be to everyone's taste.*
- *There are occasional complaints about minor inconveniences, such as slow elevators, lack of air conditioning, or noise from nearby streets or construction.*
- *Some customers mention that while the hotel is generally clean, there might be minor issues like a burnt-out light bulb or a slightly outdated bathroom.*
- *A few reviews mention that while the hotel is well-located, it might not be in the most vibrant or lively areas of the city.*



Text Summarization

Berikut adalah hasil Text Summarization dari ulasan dengan kelas Negatif:

- *The most common complaints are about poor service, with customers mentioning rude or unhelpful staff, slow response times, and a lack of flexibility.*
- *Some customers have issues with overbooking, leading to inconvenience and dissatisfaction.*
- *There are complaints about the quality of food, particularly at buffets, with comments about monotony, lack of variety, and poor presentation.*
- *A few reviews mention issues with cleanliness, such as roaches in the room or dirty bathrooms.*
- *Some customers express disappointment with the hotel's facilities, such as malfunctioning elevators, broken appliances in the room, or a pool that's not well-maintained.*
- *There are occasional complaints about the hotel's pricing, with customers*





thynk unlimited

18

Conclusion

Beberapa kesimpulan yang dapat ditarik berdasarkan analisis yang telah dilakukan adalah sebagai berikut.

- Kekuatan utama hotel terletak pada lokasi, kebersihan, kenyamanan kamar, serta keramahan staf.
- Namun demikian, masih terdapat sejumlah tantangan dan inkonsistensi dalam pelayanan, fasilitas, dan kualitas makanan.
- Beberapa aspek teknis dan estetika (seperti desain dan kondisi fasilitas) menjadi sumber komentar netral atau negatif.



Recommendation

Secara garis besar, berikut rekomendasi yang dapat diberikan kepada pihak hotel, terutama dalam hal perbaikan beberapa aspek yang dinilai masih menjadi kekurangan berdasarkan ulasan tamu.

- Peningkatan konsistensi *hospitality* (pelayanan) staf.
- Perbaikan dan perawatan fasilitas secara berkala, termasuk penjagaan kebersihan.
- Peningkatan kualitas makanan.
- Evaluasi masalah *Overbooking* yang menimbulkan ketidaknyamanan serius.
- Meminimumkan gangguan lingkungan, seperti akibat kebisingan konstruksi.



AI Support

Proyek ini melibatkan penggunaan beberapa AI sebagai berikut.

1. ChatGPT, sebagai sarana *brainstorming* ide, eksplorasi penggunaan sintaks, dan evaluasi luaran (*output*) dari IBM Granite.
2. IBM Granite untuk melaksanakan task *Text Classification and Summarization*.

next slide

