

TOPIC MODELING ON HOTEL REVIEWS

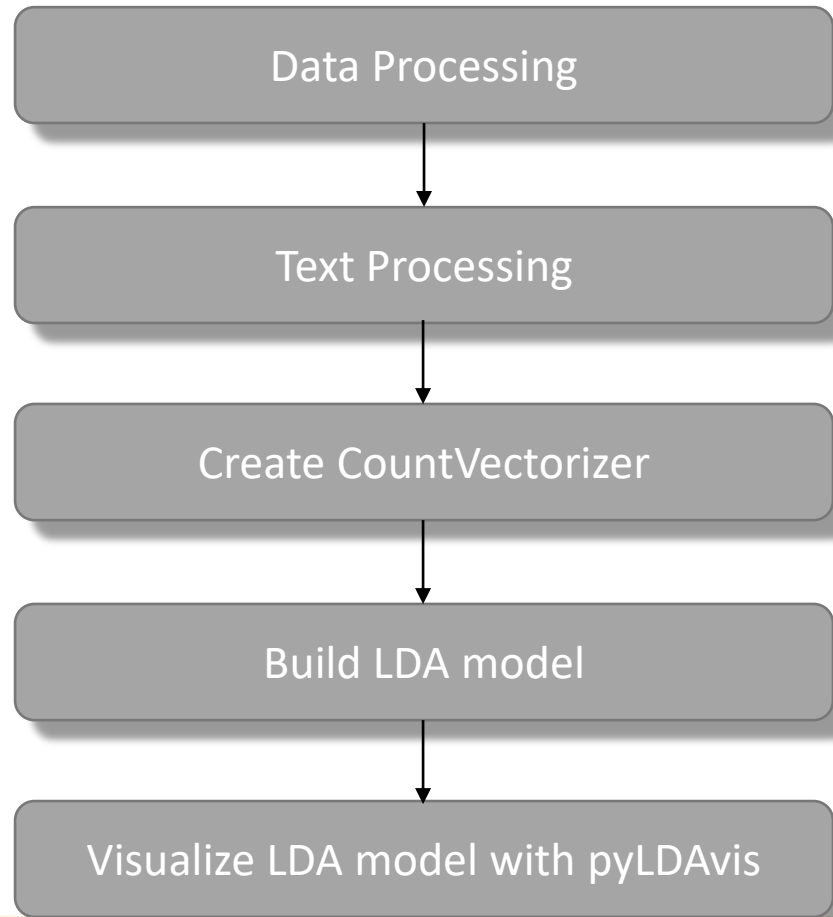


Prepared by Anița

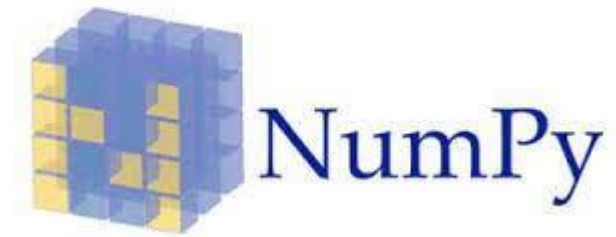
Objective

- To help hoteliers to improve the performance of their hotels based from the topic modeling of the negative reviews received about their hotels in 2017 and 2018.

Methodology



Tools Used



Data Collection

- From <https://www.kaggle.com/datafiniti/hotel-reviews>

Data Processing

- Obtain only negative reviews for year 2017 and 2018 from the dataset.
- Negative reviews are obtained for reviews with rating < 3 with the review rating scale ranging from 1 to 5.

Text Processing

- Text processing for the negative reviews is done by removing numbers, capital letters and punctuation.

Create CountVectorizer

- CountVectorizer is created for parsing/counting words.
- Created with unigrams and stop_words = 'english'

LDA Model (for negative reviews in 2018)

- Build LDA model with n_components (topics) = 4
- Top 15 keywords for each topic:

	Word0	Word1	Word2	Word3	Word4	Word5	Word6	Word7	Word8	Word9	Word10	Word11	Word12	Word13	Word14
Topic0	hotel	room	bed	did	stay	time	light	minutes	desk	staff	water	family	check	experience	door
Topic1	room	service	really	didn't	good	customer	sleep	just	experience	time	bad	shower	desk	better	told
Topic2	staff	really	stay	room	desk	rooms	time	hotel	location	care	member	customer	didn't	door	nice
Topic3	room	hotel	good	stay	old	bad	rooms	like	service	coffee	want	just	night	breakfast	business

LDA Model (for negative reviews in 2018)

- To see dominant topic in each document for the first 15 documents:

	Topic0	Topic1	Topic2	Topic3	Dominant Topic
Doc0	0.42	0.02	0.02	0.54	3
Doc1	0.97	0.01	0.01	0.01	0
Doc2	0.01	0.01	0.01	0.96	3
Doc3	0.01	0.01	0.01	0.96	3
Doc4	0.04	0.04	0.04	0.87	3
Doc5	0.01	0.83	0.01	0.16	1
Doc6	0.02	0.02	0.02	0.93	3
Doc7	0.01	0.96	0.01	0.01	1
Doc8	0.01	0.01	0.96	0.01	2
Doc9	0.00	0.99	0.00	0.00	1
Doc10	0.00	0.00	0.00	0.99	3
Doc11	0.99	0.00	0.00	0.00	0
Doc12	0.01	0.01	0.96	0.01	2
Doc13	0.00	0.00	0.99	0.00	2
Doc14	0.00	0.99	0.00	0.00	1

LDA Model (for negative reviews in 2018)

- What is topic0?
 - Doc1 talks 97% about topic0 rather than the other topic.
 - Topic0: hotel, room, **bed**, did, stay, time, light, minutes, desk, staff, water, family, check, experience, door.
 - Doc1: bad the place was falling apart the tub was cracked and broken the bathroom shelf was barely holding on **the bed was extremely tough**.
 - We can say topic0 is regarding the bed.

LDA Model (for negative reviews in 2018)

- What is topic1?
 - Doc7 talks 96% about topic1 rather than the other topic.
 - Topic1: **room**, **service**, really, didnt, good, customer, sleep, just, experience, time, **bad**, shower, desk, better, told.
 - Doc7: **bad room service was slow** have hidden pet fees and rules and the pipes werent working.
 - We can say topic1 is regarding room service.

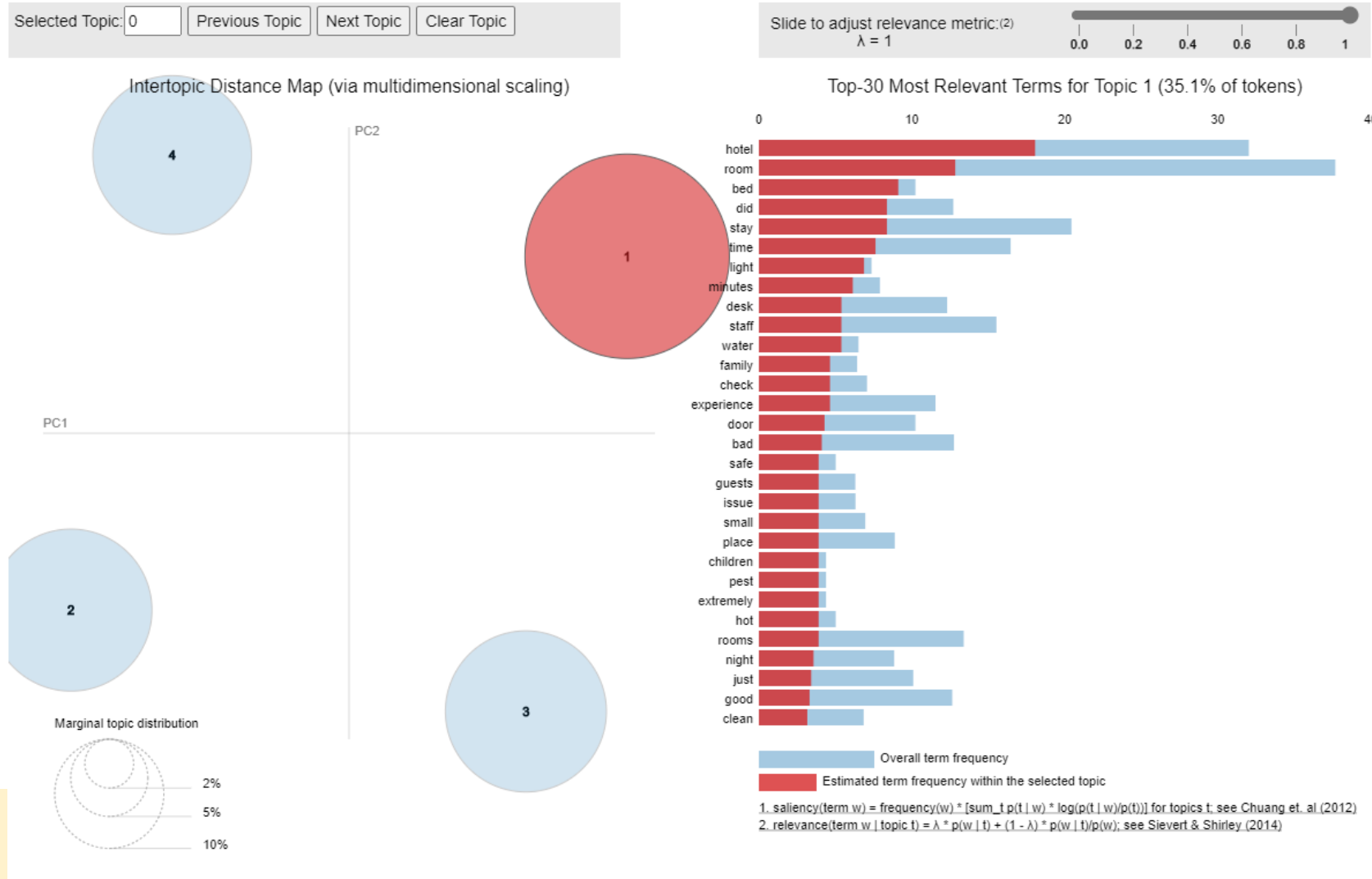
LDA Model (for negative reviews in 2018)

- What is topic2?
 - Doc8 talks 96% about topic2 rather than the other topic.
 - Topic2: **staff**, really, stay, room, desk, rooms, time, hotel, location, care, member, customer, didnt, door, nice.
 - Doc8: bad within the first minutes of being there we got a complaint from a neighbor for having music too loud the speaker was not even up half way and it was oclock in the afternoon on a saturday at the end of the night **staff were rude**.
 - We can say topic2 is regarding the staff.

LDA Model (for negative reviews in 2018)

- What is topic3?
 - Doc3 talks 96% about topic3 rather than the other topic.
 - Topic3: room, hotel, good, stay, **old**, **bad**, **rooms**, like, service, coffee, want, just, night, breakfast, business.
 - Doc3: **bad old outdated rooms** even though its a classic hotel does not mean they dont have to put forth effort to modernize rooms.
 - We can say topic3 is regarding the rooms.

pyLDAvis graph (for negative reviews in 2018)



LDA Model (for negative reviews in 2017)

- Build LDA model with n_components (topics) = 4
- Top 15 keywords for each topic:

	Word0	Word1	Word2	Word3	Word4	Word5	Word6	Word7	Word8	Word9	Word10	Word11	Word12	Word13	Word14
Topic0	room	hotel	bad	good	did	staff	stay	desk	old	motel	night	just	breakfast	didnt	tv
Topic1	room	good	bad	hotel	small	location	wasnt	nice	card	hours	phone	staff	got	pool	reservation
Topic2	room	hotel	staff	bad	work	good	manager	laptop	stay	bar	bed	experience	time	lobby	available
Topic3	bad	room	good	staff	hotel	time	bed	location	experience	access	horrible	service	coffee	floor	guest

LDA Model (for negative reviews in 2017)

- To see dominant topic in each document for the first 15 documents:

	Topic0	Topic1	Topic2	Topic3	Dominant Topic
Doc0	0.02	0.94	0.02	0.02	1
Doc1	0.97	0.01	0.01	0.01	0
Doc2	0.91	0.03	0.03	0.03	0
Doc3	0.01	0.01	0.01	0.96	3
Doc4	0.00	0.00	0.00	0.99	3
Doc5	0.96	0.01	0.01	0.01	0
Doc6	0.92	0.03	0.03	0.03	0
Doc7	0.00	0.99	0.00	0.00	1
Doc8	0.01	0.01	0.01	0.98	3
Doc9	0.02	0.95	0.02	0.02	1
Doc10	0.01	0.01	0.01	0.97	3
Doc11	0.03	0.03	0.43	0.50	3
Doc12	0.13	0.13	0.13	0.61	3
Doc13	0.01	0.97	0.01	0.01	1
Doc14	0.93	0.02	0.02	0.02	0

LDA Model (for negative reviews in 2017)

- What is topic0?
 - Doc1 talks 97% about topic0 rather than the other topic.
 - Topic0: room, hotel, bad, good, did, staff, stay, desk, old, motel, night, just, **breakfast**, didnt, tv.
 - Doc1: bad the restaurant was closed for a refurb and as a result we had to eat our **breakfast** in the lobby on our knees we had tiny plastic plates and plastic cutlery and a **poor food choice**.
 - We can say topic0 is regarding breakfast.

LDA Model (for negative reviews in 2017)

- What is topic1?
 - Doc0 talks 94% about topic1 rather than the other topic.
 - Topic1: room, good, bad, hotel, small, location, wasnt, nice, card, hours, phone, **staff**, got, pool, reservation.
 - Doc0: **staff werent helpful** and no one tried to come and clean our room on several nights.
 - We can say topic1 is regarding the staff.

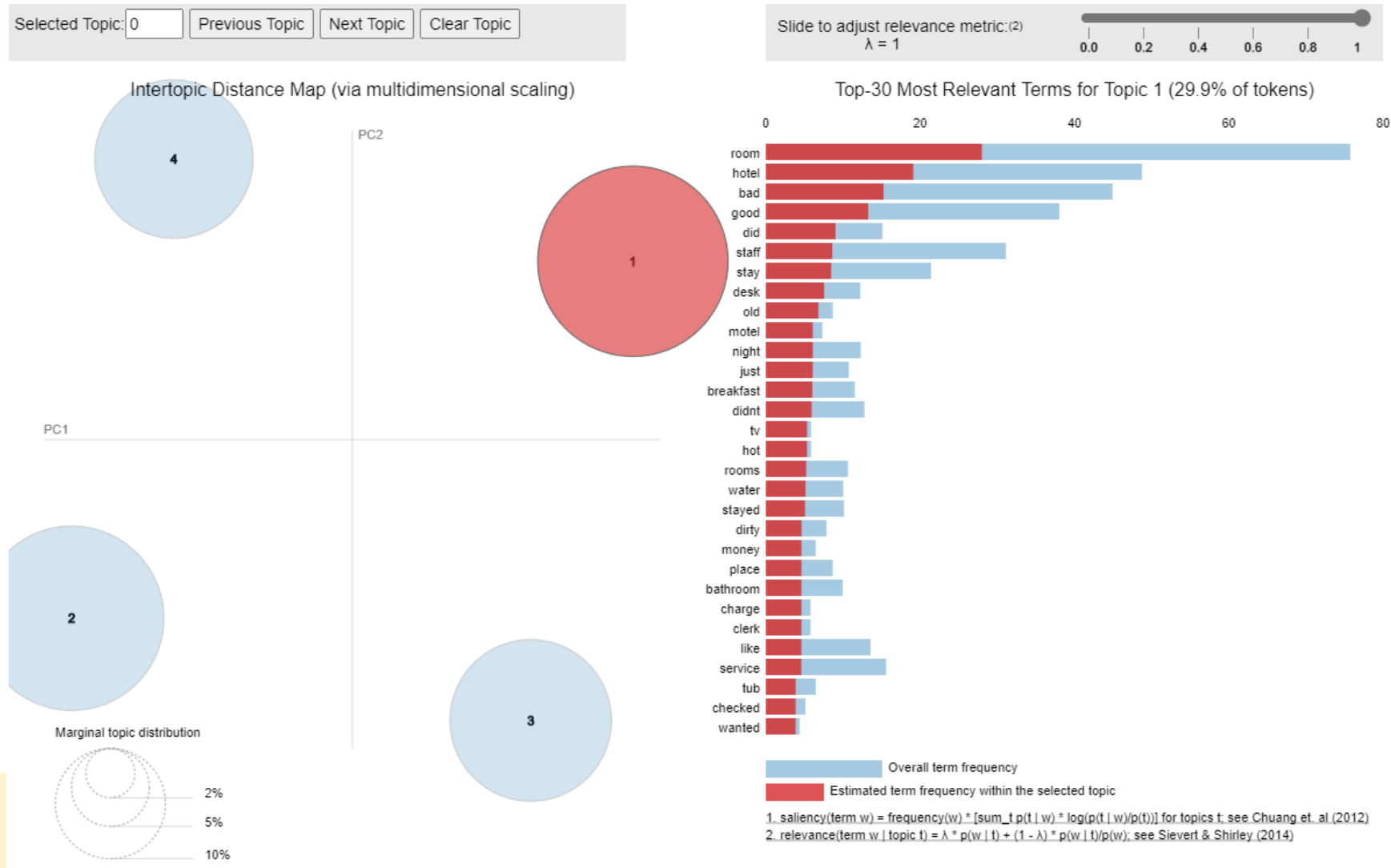
LDA Model (for negative reviews in 2017)

- What is topic2?
 - Doc24 talks 94% about topic2 rather than the other topic.
 - Topic2: **room**, hotel, staff, **bad**, work, good, manager, laptop, stay, bar, **bed**, experience, time, lobby, available.
 - Doc24: **bad** prepaid for a premium **room** with two double beds but was given a disability access **room** with a king **bed**.
 - We can say topic2 is regarding bedroom allocation.

LDA Model (for negative reviews in 2017)

- What is topic3?
 - Doc10 talks 97% about topic3 rather than the other topic.
 - Topic3: bad, room, good, staff, hotel, time, bed, location, experience, access, horrible, service, **coffee**, floor, guest.
 - Doc10: bad it was hard to get into the parking area because of traffic an event was going on **no coffee pot in the room**.
 - We can say topic3 is regarding complimentary facilities in the room.

pyLDAvis graph (for negative reviews in 2017)



Conclusions

- Topics from negative reviews received in 2018 were regarding:
 - Bed
 - Room service
 - Staff
 - Rooms
- Topics from negative reviews received in 2017 were regarding:
 - Breakfast
 - Staff
 - Bedroom allocation
 - Complimentary facilities in the room
- Hoteliers should improve on the above mentioned topics in order for their hotels to perform better.

THANK YOU