AI Assisted Sentiment Analysis of Hotel Reviews Using Topic Modeling

- Sushma Ghogale-190635387

MSc Artificial Intelligence

School of Electronic Engineering and Computer Science (EECS),

Queen Mary University of London, London, United Kingdom

- Email: s.ghogale@se19.qmul.ac.uk
- Supervisor-Prof. Mustafa Bozkurt

Index:

Introduction

- Heavy surge in Customers providing feedback on Hotels online in recent times.
- Hotel industry is heavily influenced by ratings and feedbacks provided by Customers.
- Hotels are continuously looking for Customer feedback to improve features.
- Tourists/Travellers count on reviews of the Hotel to judge the services.
- Challenge is to map the keyword of the reviewer to the aspects of the hotel.

  E.g. if review says "Easy access to nearest points" its pointing to location of the hotel.

- Ratings are easier to handle, but they don't help identify aspects, features and latent sentiments (positive, negative or neutral).
- Algorithms to segregate review comments by topics and to assess the sentiment are used.
- This experiment uses LDA Topic modelling algorithm to categorise feedback data into different subgroups, which is subjected to Sentiment analysis with Machine learning Classifiers to find the emotions.

Background

- Humans can relate a comment provided to features of a Hotel easily.

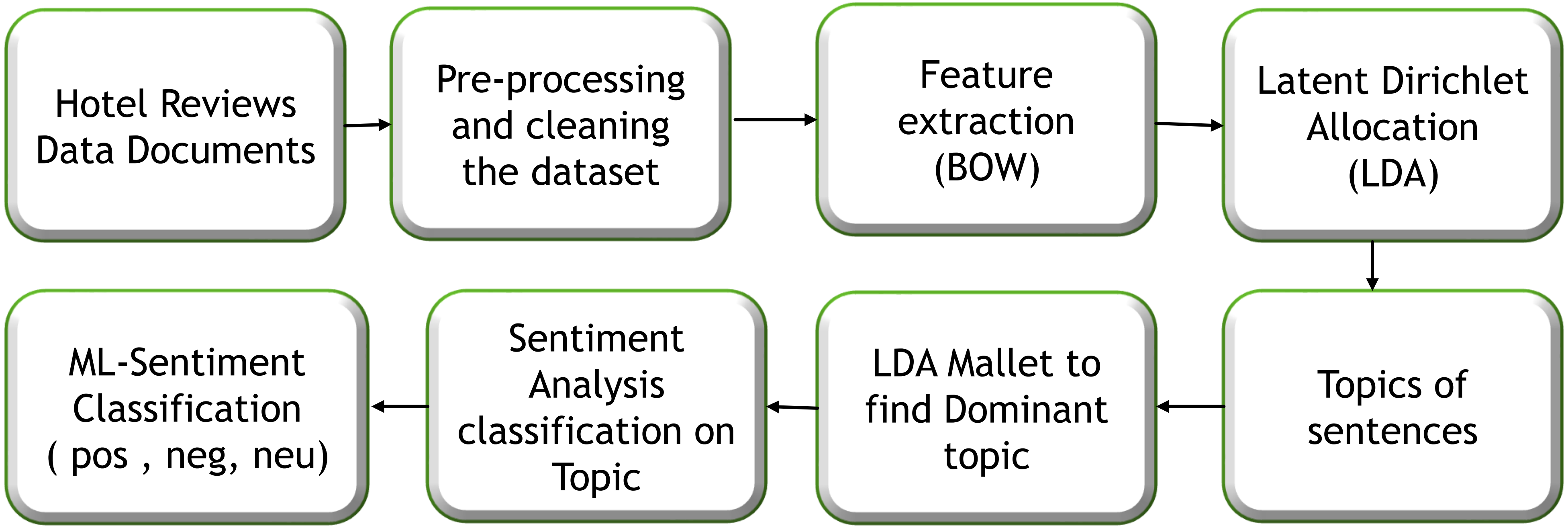
E.g. “The Hotel has a large swimming pool and clean water!” and Humans can easily detect the emotions behind the reviews, But machines can’t

- Lot of research has been done in this area to classify text in to features or topics and review them along with the sentiments associated with them
- Blei et al developed topic modelling technique, LDA to train the Machine to behave intelligently and detects the topics across the document
- Natural Language Processing (NLP) and Machine learning algorithms are helping analysts to classify sentiments
- Technology is scaling up to handle large amount of text data being provided continuously by the Customers/Users

Benefits

- Ability to classify the topic level sentiments of the reviewer
- Ability for Companies to relate review comments to features of the hotel
- Understand emotions for the features (Happy or not happy about the Hotel Facility, Service, Staff ,Location etc)
- Consolidated view of the Customers experience, and preferences related to Hotel services
- Opportunity for Hotels to improve their services to improve the ratings and hence the sale
- Tool for travellers to assess the Hotels based on parameters of their choice and able select the Hotels suitable to their needs
- Ability to handle large data set

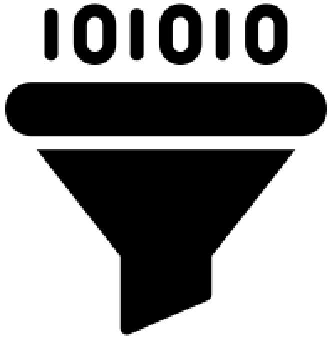
Approach



Hotel Reviews Data Documents
Pre-processing and cleaning the dataset
Feature extraction (BOW)
Latent Dirichlet Allocation (LDA)
Topics of sentences
LDA Mallet to find Dominant topic
Sentiment Analysis classification on Topic
ML-Sentiment Classification (pos , neg, neu)

Dataset

- The Dataset used in this experiment is Hotel Reviews.
- Gathered this dataset from Kaggle.com and dataworld.com websites.
- Dataset has more than 500k reviews.
- We are using only reviews and review ratings columns in this experiment
- No user identifiable data was used to ensure compliance with regulations (e.g. GDPR)
- Analysis followed ethics by avoiding any user related information like race, gender, geography etc..



Pre-processing:-

- It is a technique to remove all unwanted texts from the document
  - Tokenising:- Process of dividing document in to separate words.
  - Stop-word removal:- Removing of words which has no meaning
  - Stemming and Lemmatisation:-Reduces the data size and transforms all tenses to present.
  - Bigram and Trigram:- consequently occurring pair of word and group of three words

- Bag of words(Corpus):-

It is collection of pre-processed group of words and its value from the document

Latent Dirichlet Allocation (LDA)

LDA is one of the most used unsupervised text classification algorithm.

In this approach, to find aspects(topics) we are using LDA gensim along with Mallet's implementation.

LDA is a concept of probability, documents are probability distribution of relating topics and topics are probability distribution of words.

Each topic has its own characteristics and is unique. Topics are the probability of words appearing in a set of document.

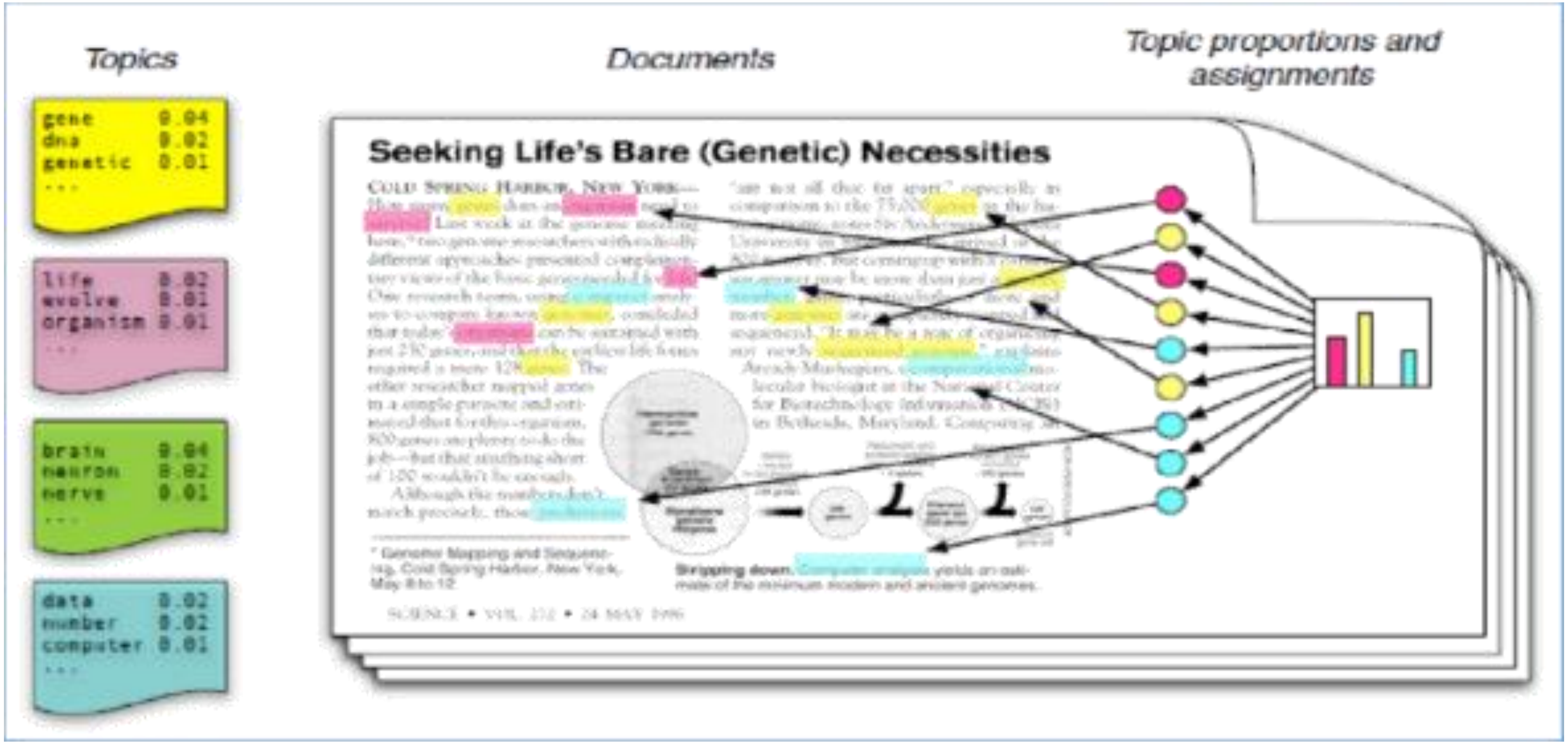
LDA can generate number of topics depending on the document. For this experiment we producing 12 topics and 12 words in each topic.

Each word in topic has its own weight, depending on the feedback of the reviewer. Words appear in high weighted to low weighted sequence (descending order).

The keywords for the topic labelling are chosen from the combination of highest weighted words.

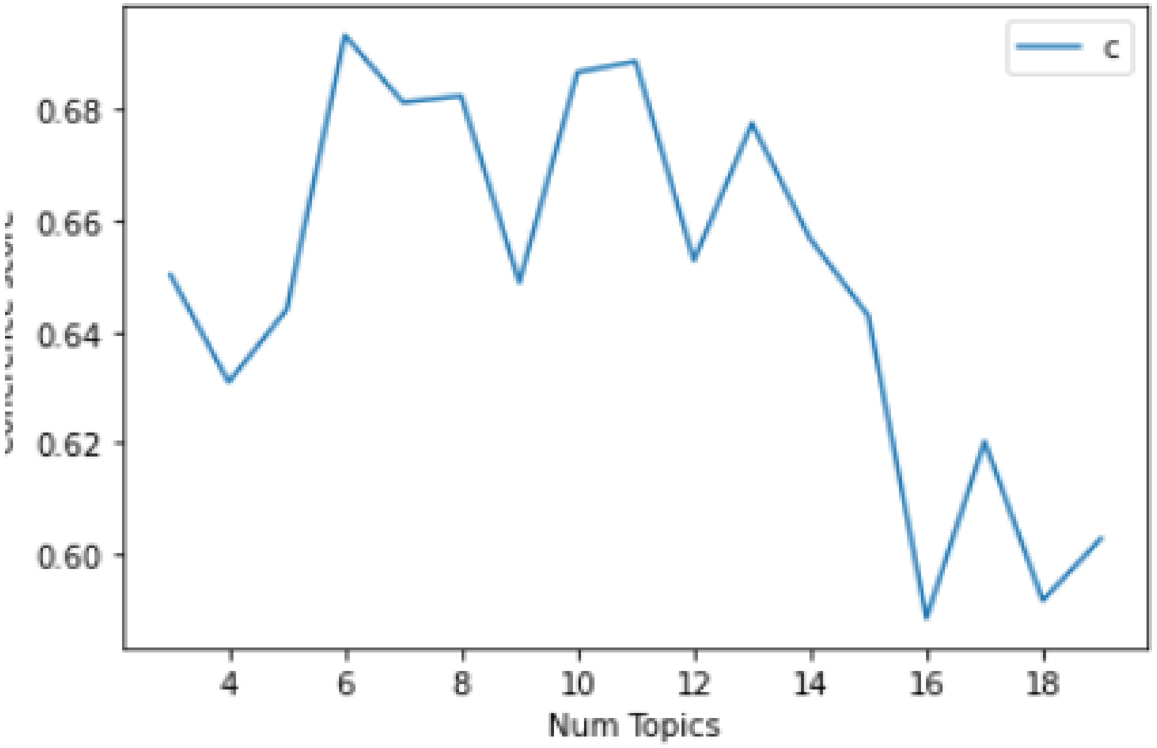
Latent Dirichlet Allocation (LDA)

How LDA assigns words to topics



Coherence and perplexity

- Optimal topics are the group of words which have high coherence value and less perplexity.
- We choose optimal topics by finding the coherence value and model finds the number of topics with high coherence score.
- We got perplexity score of -7.054584201409449
- Figure shows number of topics which has high coherence values, for example topic 5 to 8 has coherence score more than 0.68 and topics 10 and 11 also has high coherence values. On the other hand topic 4 ,9 ,12 has least values



Sample Output

Reviews from Documents	Staff were friendly and helpful	Free breakfast very limited. Just toast coffee cereal waffles, meat eggs hash. They said they only those weekends. Room very outdated.
LDA generated Topics	‘0.057*”staff” + 0.027*”friendly” + 0.027*”helpful”	0.102*"coffee" + 0.078*"breakfast" + 0.065*"poor" + 0.029*"cold" + ' '0.028*"fresh”
Manually Labelled Topics	**Staff**	**Food**

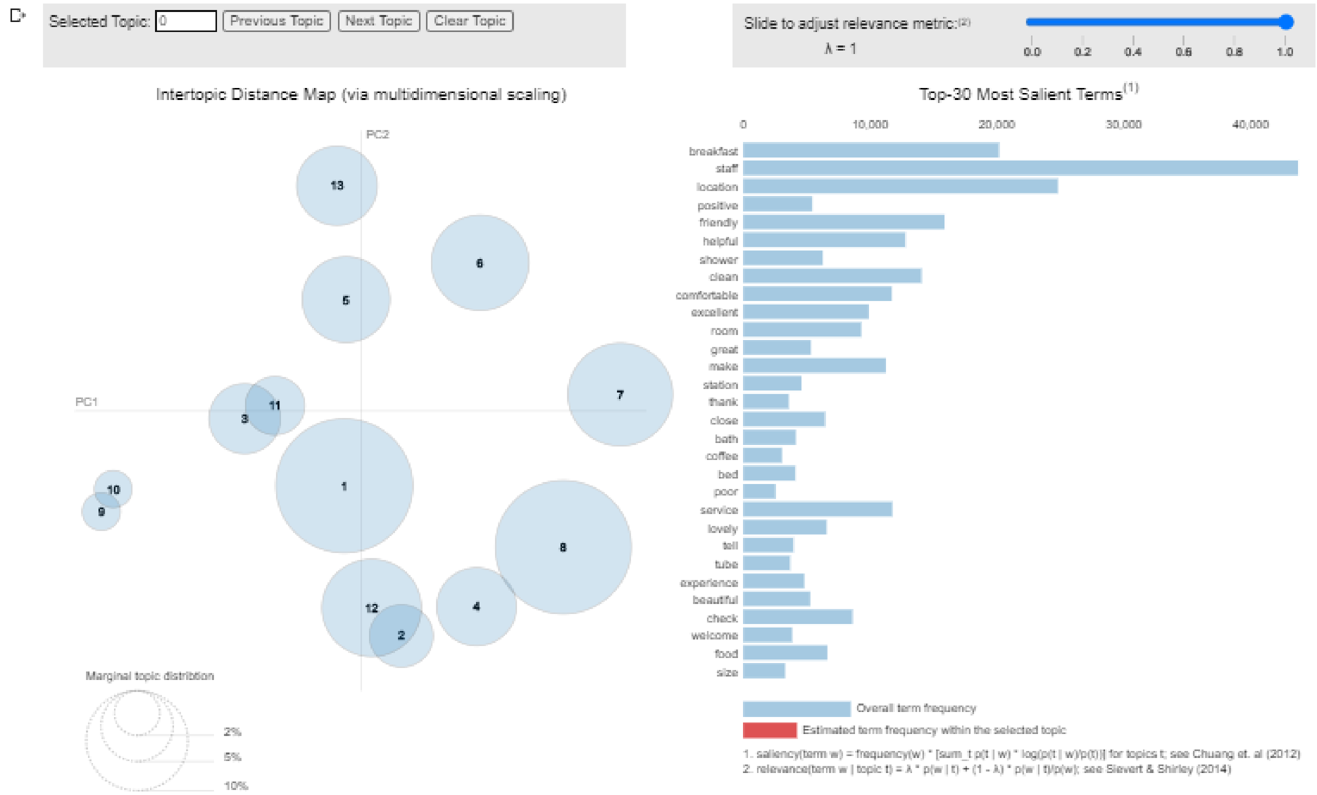
Dominant Topic

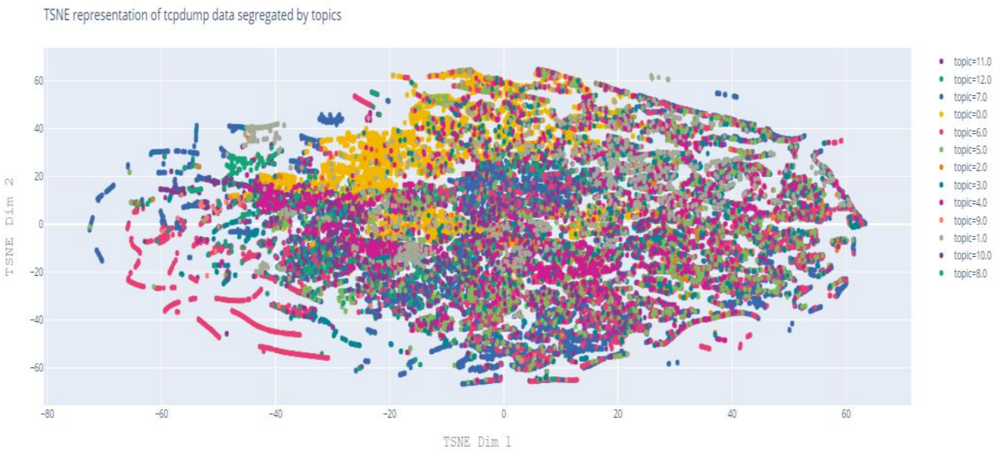
- We use LDA mallet to choose dominant topics across the document. Manually labelled topics according to Keywords extracted from Reviews

	Document_No	Dominant_Topic	Topic_Perc_Contrib	Keywords	Review_Title	Review Rating	dominant_topic_Label
	0	10.0	0.2093	walk, close, station, easy, minute, train, cit...	pleasure this nights recently. This perfect ev...	5.0	Facilities,Location
	1	11.0	0.3978	amazing, definitely, would, recommend, wonderf...	very lovely first visit this iconic bar! Wonde...	5.0	Service
	2	5.0	0.3910	staff, location, breakfast, friendly, helpful,...	Rhodes Hotel nights, location taking Paddingto...	4.0	Staff
	3	6.0	0.1827	feel, make, always, special, smile, home, welc...	Form moment arrived until left experienced abs...	5.0	Food, Service
	4	0.0	0.6617	check, book, give, reception, night, ask, take...	Well strange London's 5star when comes along e...	1.0	Facilities

Word-cloud of topics

Visualisation of topics by pyLDAvis and TSNE graph

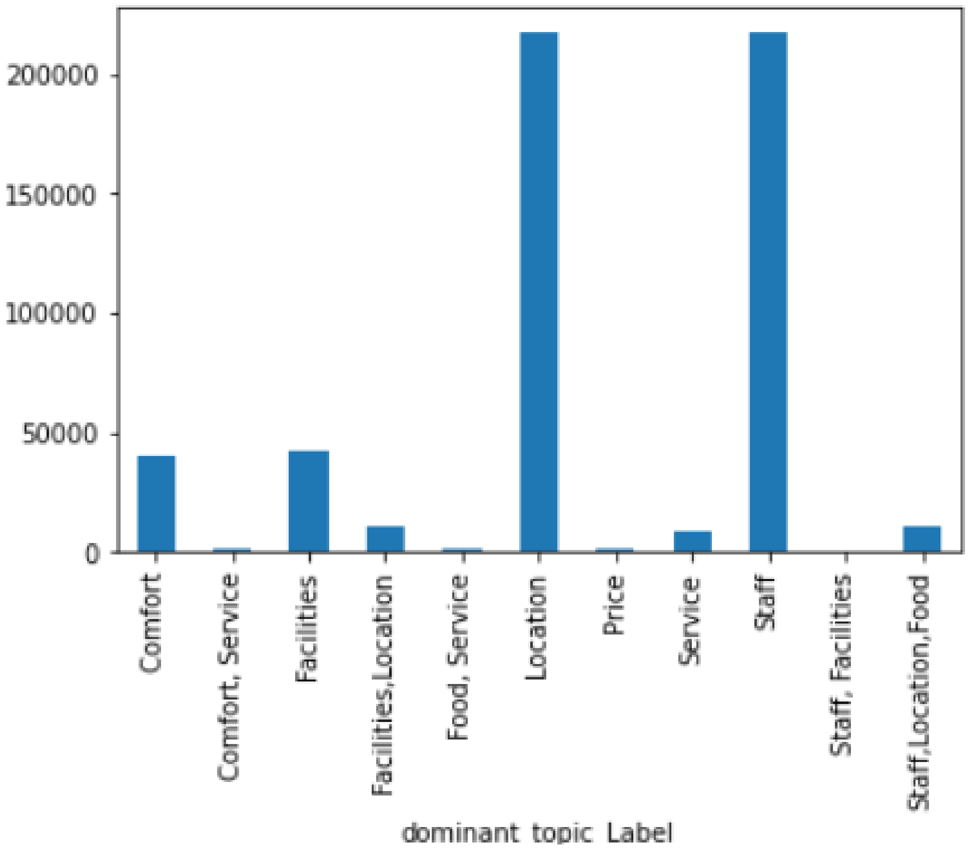


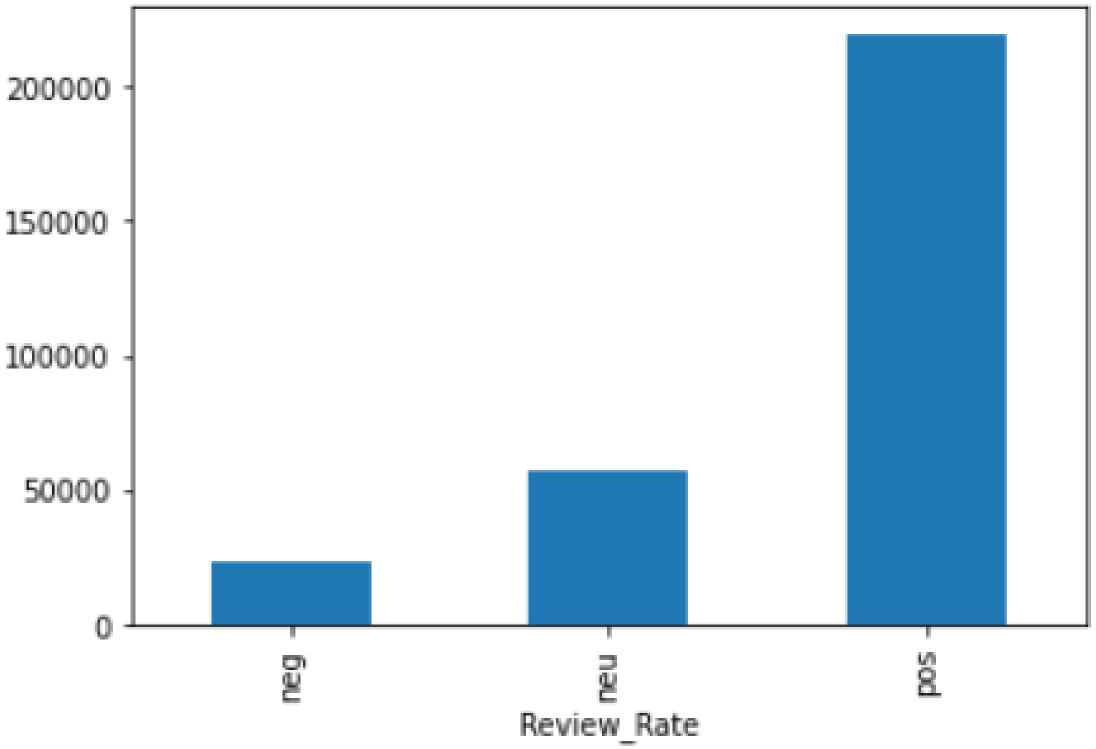


Sentiment Analysis

- Sentiment Analysis is a process of finding emotions out of review text provided by the Customer
- To do this on aspect level, we use output of LDA dataset. i.e. Topic level Sentiment analysis is done using data generated from LDA.
- In this experiment we have taken the review rating belongs to each review and categorised them in to Positive, Negative and Neutral with respect to reviewer ratings.
- Review rating >=4 is Positive
- Review rating <=2 is Negative
- Review rating >2 and <4 is Neutral

Dominant Topics and Classified Reviews Based on Reviewers Rating





Machine learning Classifiers NB,SVM,LR and DT

- To classify Sentiments on topic level we used Machine learning techniques.
  - Naïve Bayes classifier (NB)
  - Support Vector Machine (SVM)
  - Logistic Regression (LR)
  - Decision Tree classifier (DT)
- We pre-process the dataset same way as LDA pre-processing technique and Machine learning classifier on topic level.
- For Evaluation of topic level through four classifiers, we divided the dataset in to training and testing with 80% and 20% respectively for each topic.
- We extract each topic and evaluate it by different classifiers, and check which classifier model gives high accuracy.

Evaluation

To find how accurate these classifiers perform on the model, We find Accuracy, F1-score, Precision and Recall.

True reviews

Predicted reviews

True Positive (TP)	False Positive (FP)
False Negative (FN)	True Negative (TN)

$$\text{Accuracy} = \frac{Tp+Tn}{Tp+Tn+Fp+Fn}$$

$$\text{Precision} = \frac{Tp}{Tp+Fp}$$

$$\text{Recall} = \frac{Tp}{Tp+Fn}$$

$$\text{F1-score} = 2 * \frac{Precision \ x \ Recall}{Precision+Recall}$$

Evaluation

- Evaluation is depending on how accurately the model has classified the text. for calculating, we need to classify True positive (Tp), True negative (Tn), False positive (Fp) and False negative (Fn) values.

- Accuracy: The sum of correctly classified results for the data set
- Precision: Precision is the ratio of correctly classified positive reviews against total classified positive reviews
- Recall: Recall is the ratio of correctly classified positive reviews against the total actual positive reviews
- F1-score is derived by combining precision and recall

Results

Comparison of Accuracy Result

	Service	Staff	Location	Food	Comfort	Facilities	Price
Classifier							
Logistic Regression	0.844	0.76	0.744	0.90	0.493	0.631	0.62
Naïve Bayes	0.844	0.76	0.742	0.90	0.525	0.636	0.63
Support Vector Classifier	0.833	0.75	0.713	0.90	0.448	0.608	0.55
Decision Tree	0.73	0.64	0.612	0.85	0.400	0.499	0.48

Comparison of Accuracy Results between various classifiers for sentiment Analysis on topic

Table shows Naïve Bayes and Logistic regression produce high accuracy. Overall, Naïve Bayes classifier produces high accuracy compared to other classifiers.

Detailed Result

Topic	Evaluation	Logistic Regression (LR)	NaiveBayes (NB)	Support Vector Machine (SVM)	Decision Tree (DT)	Dominant classifier
Staff	Accuracy	0.76	0.76	0.75	0.64	NB & LR
	Precision	0.99	0.99	0.98	0.67	
	Recall	0.76	0.76	0.75	0.64	
	F1_score	0.86	0.86	0.85	0.66	
Location	Accuracy	0.74	0.74	0.71	0.61	NB & LR
	Precision	0.99	0.98	0.89	0.63	
	Recall	0.74	0.74	0.71	0.61	
	F1_score	0.85	0.85	0.79	0.62	
Food	Accuracy	0.90	0.90	0.89	0.85	NB & LR
	Precision	1.00	1.00	1.00	0.87	
	Recall	0.90	0.90	0.90	0.85	
	F1_score	0.95	0.95	0.95	0.86	
Service	Accuracy	0.84	0.84	0.83	0.73	NB & LR
	Precision	1.00	0.99	0.97	0.74	
	Recall	0.84	0.84	0.83	0.73	
	F1_score	0.91	0.91	0.89	0.73	
Comfort	Accuracy	0.49	0.52	0.44	0.40	NB
	Precision	0.79	0.97	0.54	0.41	
	Recall	0.49	0.52	0.44	0.40	
	F1_score	0.59	0.68	0.48	0.40	
Price	Accuracy	0.62	0.64	0.55	0.48	NB
	Precision	0.95	1.00	0.68	0.50	
	Recall	0.62	0.64	0.55	0.48	
	F1_score	0.75	0.77	0.60	0.49	
Facilities	Accuracy	0.63	0.64	0.60	0.50	NB
	Precision	0.94	0.99	0.85	0.51	
	Recall	0.63	0.64	0.60	0.50	
	F1_score	0.50	0.77	0.70	0.49	

Result

- Overall, we see from the tables that Naïve Bayes classifier produces the high accuracy and F1-score among SVM, LR, and DT
- This experiment has produced the results at 90% accuracy for topic “Food”, and the lowest accuracy produced by DT is 40% on topic “Comfort”.
- SVM and LR have also performed well. Produced good accuracy on topics service, staff, location and Food.
- SVM,LR and NB performs well in linear classification, but Decision tree is considered as non linear did not perform well in this model.

Limitations

- Accuracy could have been more with larger data set
- Manual labelling is time consuming
- The Data analysed had more positive reviews (could bring bias)
- Some reviews with generic text could not be classified in to subgroups
- Accurate labelling is required for topic classification and opinion mining
- Only dominant topics were classified in this experiment
- Mixed topics (combination of more than one topics in a subgroup) were not classified in this experiment
- LDA may take longer time to train depending on the size of the data
- Extending this approach to other languages requires respective libraries

Conclusion and Future Work

- Several researches are being done in this area due to demand
- This area still in nascent stage owning to complexities in deciphering user generated content
- This approach has worked successfully providing topics with high accuracy
- Further work is suggested to increase the accuracy of all topics
- Further work recommended on using Deep learning and neural network techniques on the text data
- Recommended to research on using different topic models like NMF, LSA etc…

Thank you....

References

1. Lu Y, Mei Q and Zhai C 2010 Investigating task performance of probabilistic topic models: an empirical study of PLSA and LDA Information Retrieval 14 178-203 R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
2. Akhtar, N., Zubair, N., Kumar, A. and Ahmad, T., 2017. Aspect based Sentiment Oriented Summarization of Hotel Reviews. Procedia Computer Science, 115, pp.563-571
3. Blei, D. M., Ng, A. Y., and Jordan, M. I. 2003. Latent dirichlet allocation. Journal of Machine Learning Research,pp 993-1022.
4. Brownlee, J., 2020. A Gentle Introduction To The Bag-Of-Words Model. [online] Machine Learning Mastery. Available at: <https://machinelearningmastery.com/gentle-introduction-bag-words-model/> [Accessed 17 August 2020].
5. Appel O, Chicalana F, Carter J, Fujita H 2016 .A hybrid approach to the sentiment analysis problem at the sentence level. Knowl Based Syst 108:110–124
6. Pranckevičius, T. and Marcinkevičius, V., 2017. Comparison of Naive Bayes, Random Forest, Decision Tree, Support Vector Machines, and Logistic Regression Classifiers for Text Reviews Classification. Baltic Journal of Modern Computing, 5(2).
7. ResearchGate. 2012. (PDF) Classification Of Customer Reviews Based On Sentiment Analysis. [online] Available at: <https://www.researchgate.net/publication/252067764_Classification_of_Customer_Reviews_based_on_Sentiment_Analysis> [Accessed 12 August 2020].
8. Grubert, E., 2016. Implicit prioritization in life cycle assessment: text mining and detecting metapatterns in the literature. The International Journal of Life Cycle Assessment, 22(2), pp.148-158.
9. Jones, Z. and Wallach, H., 2016. Inference on the Effects of Observed Features in Latent Space Models for Networks. SSRN Electronic Journal
10. H X Shi and X J Li 2011."A sentiment analysis model for hotel reviews based on supervised learning," in in International Conference on Machine Learning and Cybernetics China
11. D. M. Blei, Apr. 2012. ‘‘Probabilistic topic models,’’ Commun. ACM, vol. 55, no. 4, pp. 77–84,