BLG4902 - Graduation Project

Deep Recommender System Using Aspect-Based Sentiment Analysis of Users’ Reviews

Ayşe Betül Çetin - 150180730

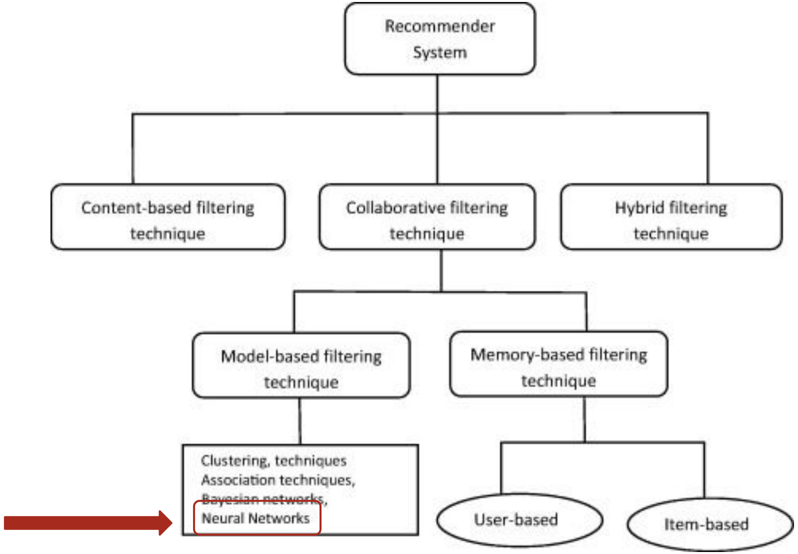
Bilgehan Emiral - 010160050

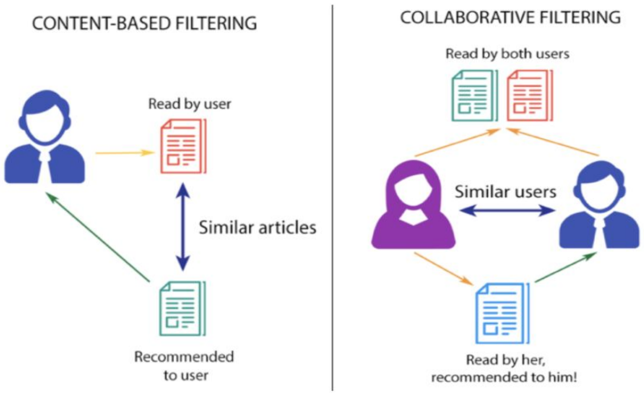
Contents

INTRODUCTION

- ❑ **Recommendation System Methods?**
- ❑ **Multi-Criteria Recommendation?**

We propose a novel solution for improving **collaborative filtering** model performance that combines **aspect based sentiment analysis** and **multi-criteria recommendation** with **deep learning**. We use **deep neural network** to discover the **<u>relationship between overall rating and sub-ratings with aspect polarities.</u>**





ABSA

Aspect-based sentiment analysis which is applied to the reviews
Polarities are extracted
Polarity features feed into the dataset and final dataset obtained

DNN

We get the author and hotel embeddings, concatenate them, and give them as input to a Polarities and Sub-ratings Deep Neural Network (DNN) model to predict the sub-ratings and polarities.

DNN

We give obtained criteria array as input to another DNN to predict the overall score.

Evaluate

The proposed model was evaluated with a multi-criteria rating and review dataset called HotelRec, based on TripAdvisor

DATASET & Preprocessing

HotelRec (Tripadvisor) Dataset

Data Cleaning

- ❑ Handle multiple reviews, untitled users
- ❑ Rows with user columns undefined or none were removed.
- ❑ Averaging multiple reviews (case of a user making more than one review on a hotel)

Data Filtering

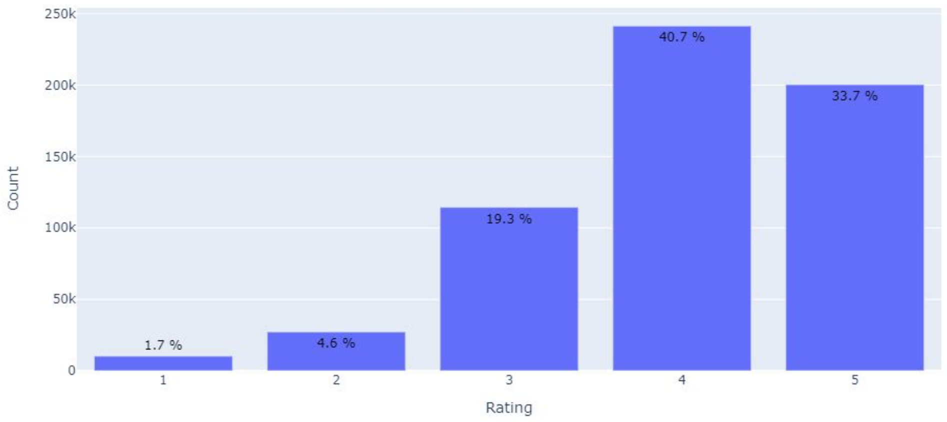
- ❑ Reviews made in the last ten years were selected.
- ❑ Filtering users and hotels: Users with ten or more reviews were selected for the model.
- ❑ After these operations, approximately 1% of the original data is selected for the model.

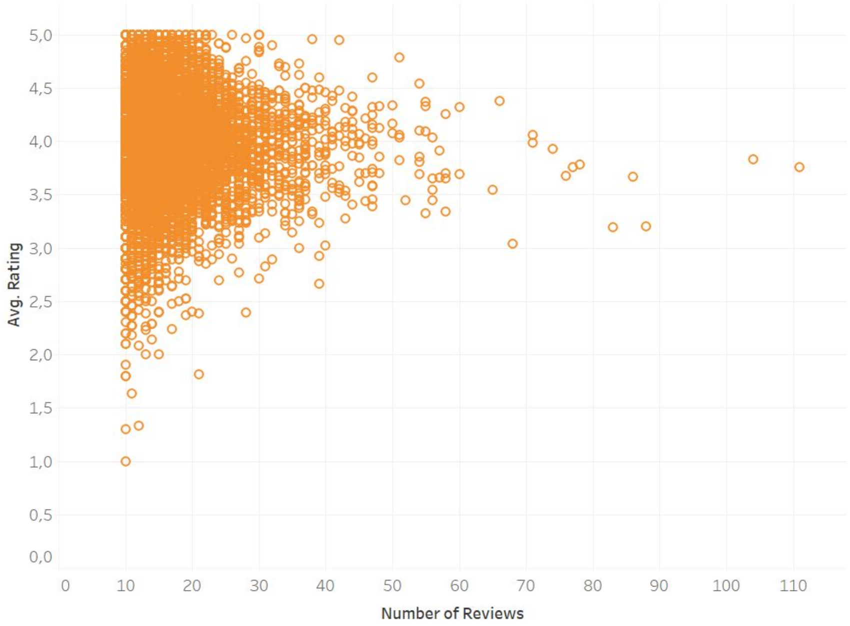
Feature Engineering

- ❑ Encoding the categorical data (user, hotel)
- ❑ Normalization of polarities and ratings
- ❑ Number of reviews column generated as a feature

- ❑ 593.000 rows
- ❑ 24.890 unique users
- ❑ 178.096 unique hotels
- ❑ 31.191 hotel locations

Overall Rating and Sub-Ratings & Number of Reviews





Sub-Ratings

	Overall	Service	Location	Value	Cleanliness	Sleep	Rooms
Null Rate:	0.0%	28,5%	57,5%	57,5%	42,5%	40,2%	58,4%
Mean:	4.0014	4.1617	4.2659	3.9971	4.2493	4.1198	3.9796
Normalized Mean:	0.7399	0.5946	0.3723	0.3479	0.3704	0.3406	0.3382

	Service	Location	Value	Cleanliness	Sleep	Rooms
1	1.5%	0.3%	0.9%	0.6%	0.8%	0.7%
2	2.7%	1.0%	2.1%	1.3%	1.6%	2.0%
3	11.3%	6.2%	8.8%	5.6%	6.6%	9.2%
4	23.3%	14.5%	15.2%	14.1%	14.2%	15.0%
5	32.7%	20.5%	15.5%	20.8%	17.1%	14.6%

ASPECT BASED SENTIMENT ANALYSIS

- ❑ **What is aspect?** attributes or components of an entity (meal, staff, surrounding..)
- ❑ **What is sentiment?** positive or negative opinions about a particular aspect
- ❑ **What is sentiment analysis?** method of natural language processing (NLP) that uses machine learning to analyze and classify the emotional tone of text data
- ❑ **What is polarity?** values that represent the user's positive (max: +1) or negative (min: -1) interaction for that aspect

Example:

"I love room" – this opinion unit is 'Positive" (sentiment) and is about 'room' (aspect) "but I wish location was more accessible" – this opinion unit is 'Negative' (sentiment) and is about 'location' (aspect)

Aspect-based sentiment analysis operations were performed for **each sentence** of **each review**.

It contains the following steps.

ABSA Steps

1. Breaking the whole review into sentences by using the Natural Language Toolkit tokenization package and by using sent_tokenization method.
2. Splitting the sentences into opinion units: Words in a review contains different importance with each aspect. Most of the times a user's reviews contains multiple aspect for a specific item.
3. Word Extraction stage: Operations include removing stop words (e.g., at, on, in), removing words containing less than three letters, removing numbers written in a typeface.

For instance, we should perform the following operations for the sentence:

"I liked the room, but the food was horrible."

First of all, it is necessary to find the sentence's **predicate**. In this sentence, "like" and "was" are the verbs we want to find. There is the verb "like" at the beginning, but there is a verb in the continuation of the sentence, which is "was". If there is **more than one verb** as in the example, the model looks at the rest of the sentence to find it. Thus, it detects more than one emotion in a sentence. By looking at the properties and positions of the tokens, it determines the domain of verbs in the sentence. In this way, sentences are divided into judgments according to different **aspects**. These judgments are called **opinion units.**

ABSA Steps

4. Lemmatization stage: Process of distinguishing verbs and nouns by identifying a word's position in a sentence using NLTK

5. Detection of aspects phase: Aspects are found by identifying nouns in the sentence. The most frequently used nouns are designated as aspects.

6. The polarity was found using the SentimentIntensityAnalyzer method in the VaderSentiment library. For instance, this method helps classify polarity as positive if the word “great” is used for a particular aspect and negative if “bad” is used.

7. Adding polarity phase, polarities created and extracted feature column according to aspects.

For instance, if the **aspect** from the sentence is "Railway", "View", "Station", "Airport" or "Distance", the **polarity** of the "Location" **category** is obtained. To give another example, if the aspect of the sentence is "Drink", "Breakfast", or "Food", the polarity of the "Meal" category is obtained.

Aspect Categories:

Value	Location	Service	Meal	Facility	Room	Quality	Staff	Surrounding
Price	Railway	Desk	Drink	Pool	Bed	Satisfactory	Good	Landmark
Amount	View	Check-in	Breakfast	Spa	Dirty	Ample	Polite	Monument
Rate	Station	Check-out	Spicy	Wi-fi	Clean	Hygienic	Helpful	Temple
Cheap	Airport	Reliable	Food	Gymnasium	Toilet	Proper	Friendly	Mosque
Worth	Distance	Fast	Tasty	Gym	Bathroom	Ambience	Reliable	Church
Low	Far	Convenient	Tea	Internet	Shower	Odour	Quick	Restaurant
Money	Close		Buffet	Ample	Dryer	Smell		Diner
Economical	Convenient		Bar	Parking	Fridge		Mall	
Reasonable	Train		Restaurant	Wireless	View			Market

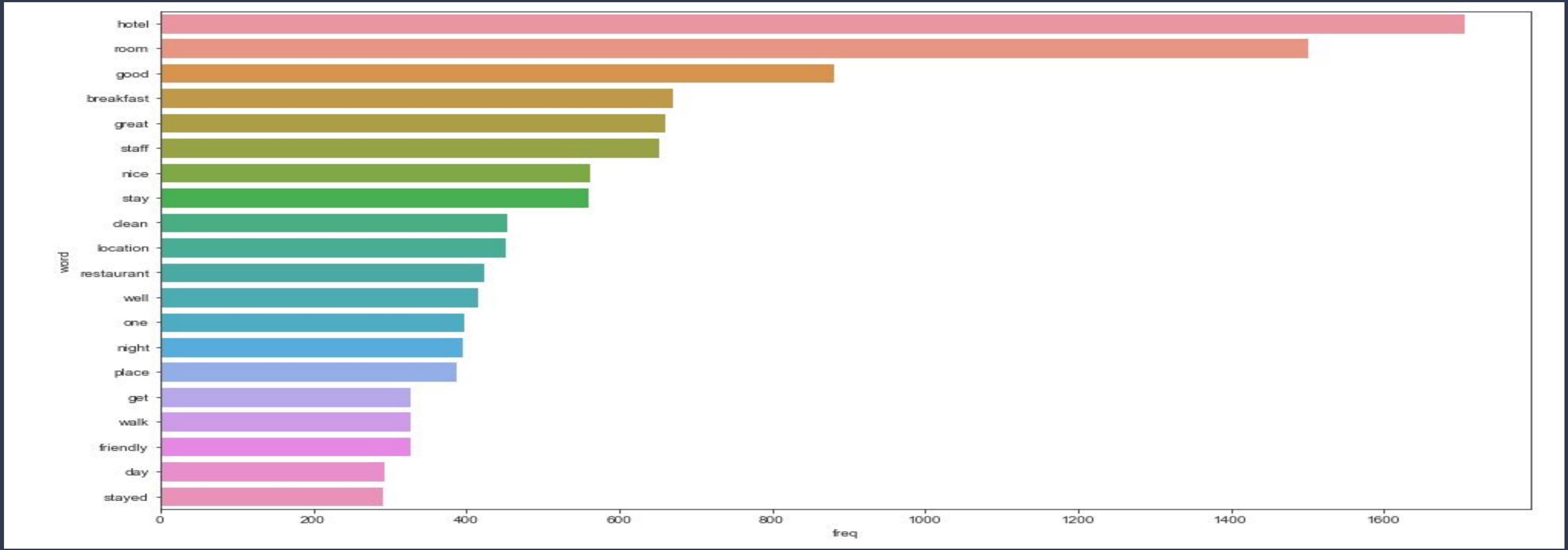
Review Text Example

“The property is surrounded by trees, which are teeming with birds and inquisitive miniature monkeys who venture onto the balconies looking for treats of papaya from the staff and guests. The rooms are large and comfortable with high ceilings, cable TV, and hot showers. Breakfast was included, with an amazing selection of fruits, bread, cereals and hot food like eggs and local sausage casserole.”

	Mean
Total Polarity	0.65327
Value Polarity	0.42399
Location Polarity	0.49672
Service Polarity	0.50180
Meal Polarity	0.47155
Facility Polarity	0.48178
Room Polarity	0.52255
Quality Polarity	0.46807
Staff Polarity	0.51647
Surrounding Polarity	0.48803

ABSA Sentence Examples

Sentence	Aspect	Polarity
Service is outstanding as well staff are all nice and very helpful.	Service	0.8999
Service was fast and efficient.	Service	0.6486
Service is nice and friendly.	Service	0.6908
Service is friendly.	Service	0.4939
Staff however were very friendly and accommodating.	Staff	0.5413
We were in Room 930 which is top floor, corner room.	Room	0.2023
Service and food at the beach was good though a bit pricey.	Service	0.4404
Rooms big but aircon noisy.	Room	-0.2617
Rooms always clean and great restaurant in the lobby.	Room	0.7845
Stayed in the G Deluxe Room (aka White Room).	Room	0.0
Location good.	Location	0.4404
Location great for tourist attractions, jonker st ect.	Location	0.7845
The Hotel is stunning The Location is perfect.	Location	0.743
Staff was always friendly and greeted me all the time.	Staff	0.6486
Service is excellence, and the staff truly cannot do enough for each and every guest, service is truly incredible!	Service	0.8805
Room was newly decorated, not amazing but nice with really comfortable bed and lovely toiletries.	Room	0.7938
Service is always impeccable from Genaro, Javier, Domingo, Luis, Mario, Benito, both Jorge’s, (pool bar and Palazzo) both Ricardo’s, (pool, Ventanas, Azia, and Palazzo) Eduardo, Erik, Miguel, Adan, Ilse, and of course Isis, I am very sure that there are many other wonderful RH staff members that we omitted, and we will be sure to give each of them the credit that is due in our next review.	Service	0.8805

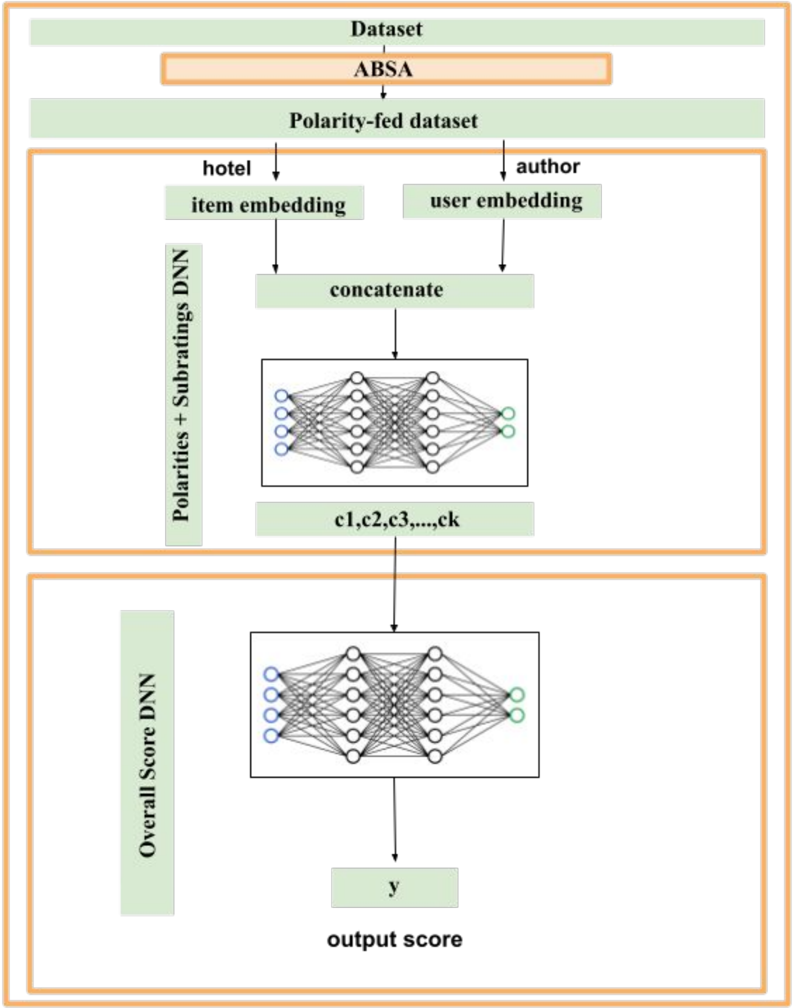


	author	hotel_url	numReview	rating	service	location	value	cleanliness	sleep	rooms	...	valuePol	locationPol	servicePol	mealPol	facilityPol
0	-Global-Galivanter-	Hotel_Review-g255068-d9980609-Reviews-Summer_H...	1	1.00	0.0	0.0	0.0	0.0	0.0	0.0	...	0.423108	0.489669	0.495914	0.471109	0.481706
1	-Global-Galivanter-	Hotel_Review-g293916-d2180563-Reviews-Studio_S...	1	0.75	0.0	0.0	0.0	0.0	0.0	0.0	...	0.423108	0.489669	0.495914	0.471109	0.481706

MODEL

The model is composed of 3 parts.

1. ABSA
2. We get the concatenated author and hotel **embeddings** and feed them as input to the Polarities and Sub-ratings DNN.
    a. Since only IDs of users and items are used as a feature, this model uses **pure collaborative filtering** (not content based).
    b. The embedding vectors of users and items are first set to random values, and then the values are changed during model training to minimize the loss function.
    c. A deep neural network is used to **predict** a user's criteria ratings and polarities on an item in this step.
3. We give obtained **criteria array (c1, c2, c3,..,ck)** as input to Overall Score DNN to predict the overall score (y) that represents the probability of the user choosing the hotel.



Settings & Experiments

- We used **Keras** with **TensorFlow** for the neural network model implementation. On the other hand, we used **Natural Language Toolkit (NLTK)** for aspect-based sentiment analysis.
- **The learning rate** is a hyperparameter that controls how much to change the model in response to the estimated error each time the model weights are updated.
- We use **Rectified Linear Unit (ReLU)** in model that is a single-point function that gives 0 when the input value is lesser than the set value and gives the linear multiple if the input is given is greater than the set value.
- The **batch size** defines the number of samples that will be propagated through the network.
- In Sub-ratings and Polarities DNN, **the output layer has 20 neurons** that is equal to the number of the sub-ratings plus polarities and the number of reviews. In the Overall score DNN, **the output layer has one neuron** for the overall score representing the value between 0 and 1, showing the estimation of the probability of the user choosing the hotel.
- Model: **Keras-Sequential**

Settings & Hidden Layer Experiments

- 7 hidden layer for both DNN’s

Hidden Layers	MAE	Elapsed Time (s)
[16, 8]	0.11192	195.5212
[32, 16, 8]	0.10101	217.5914
[64, 32, 16, 8]	0.10161	217.5524
[128, 64, 32, 16, 8]	0.18150	277.5289
[256, 128, 64, 32, 16, 8]	0.26197	238.8719
[512, 256, 128, 64, 32, 16, 8]	0.17999	277.4658
[1024, 512, 256, 128, 64, 32, 16]	**0.09948**	**343.1469**
[2048, 1024, 512, 256, 128, 64, 32]	0.12090	337.6681

Hyperparameter Experiment Results

Best Hyperparameters

Hyperparameter	Search Space	Selected Value
Learning Rate	[0.05-0.001]	0.001
Epoch	[8-128]	64
Optimizer	[Adam, Adamax, Rmsprop]	Adam
Embedding Size	[32-512]	256
Batch	[8-128]	32
Epoch	[8-128]	32

Embedding Size	MAE	Elapsed Time (s)
32	0.14229	277.6943
64	0.10668	277.6408
128	0.18006	277.6465
256	**0.10290**	**277.8883**
512	0.22208	266.2203

Epoch	MAE	Elapsed Time (s)
8	0.30479	46.1332
16	0.14823	76.2778
32	**0.10063**	**258.5229**
64	0.10246	136.4422
128	0.10175	525.5475

Batch Size	MAE	Elapsed Time (s)
8	0.10312	938.02425
16	0.10425	517.74439
32	**0.10107**	**157.65366**
64	0.14004	256.11178
128	0.10413	81.61894

Learning Rate	MAE	Elapsed Time (s)
0.0001	0.100110	252.4514
0.0005	0.105321	255.2264
0.001	**0.09969**	**257.0208**
0.002	0.10106	277.6118
0.005	0.10996	267.9195

Accuracy Metrics

where r__ui' is predicted score of user (u) for item (i), r__ui is the actual score, and |N| is size of test set.

$$MAE = \frac{1}{|N|} \sum_{r'_{ui} \epsilon N} |r_{ui} - r'_{ui}|$$

NOTE: A small MAE is a good indicator, while a large F1 is good.

Precision is calculated as recommended and relevant items divided by recommended items. **Recall** is calculated as recommended and relevant items divided by relevant items. **The definition of relevant item is based on its actual rating which is greater than a specific threshold.** If the estimated rating is greater than the given threshold in its k-neighbours then that item is considered as a recommendation to the user. We set the threshold 0.5 since predictions in interval 0-1.

$$F1 = 2 \times \frac{(precision \times recall)}{(precision + recall)}$$

Compared Methods & Results

SVD: Probabilistic Matrix Factorization.

KNNBaseline: a basic collaborative filtering algorithm taking into account a baseline rating.

NormalPredictor: predicts a random rating based on the distribution of the training set and it is one of the most basic algorithms that does not do complex work.

The table shows that multi-criteria ratings and ABSA features are better than the other models.

To directly predict the score a singular neural network is used with only multi-criteria ratings to show the difference between the version with added polarities.

Method	**MAE**	**F1-Score**
BaselineOnly	0.164395	0.728009
SVD++	0.168171	0.723819
KNNBaseline	0.171455	0.730498
KNNWithZScore	0.175365	0.753533
KNNWithMeans	0.175581	0.752269
KNNBasic	0.177417	0.819961
SVD	0.177629	0.723647
NormalPredictor	0.244793	0.639061
CoClustering	0.403692	0.759395
Single Rating DNN	0.154805	0.774335
Multi-criteria Ratings DNN	0.112469	0.794236
Proposed model	**0.099484**	**0.823616**

Conclusion & Future Work

A novel deep multi-criteria collaborative filtering approach using aspect-based sentiment analysis of users' reviews was developed for hotel recommendation.
The proposed model has demonstrated
better results as the novel architecture is implemented that integrates aspect-based sentiment analysis and multi-criteria ratings.
The proposed model discovers the relationship between overall rating and sub-ratings with aspect polarities.

- Future work will improve the dataset with **location** and season information. Location information can be extracted and classified from the hotel URL. However, there are problems in extracting the locations since the place names in the data are not very proper. In addition, users can be categorized as **domestic - international travelers** according to their activities. **Traveler-type** information such as single, couple, family, business, and friends could benefit.

Thank you for listening