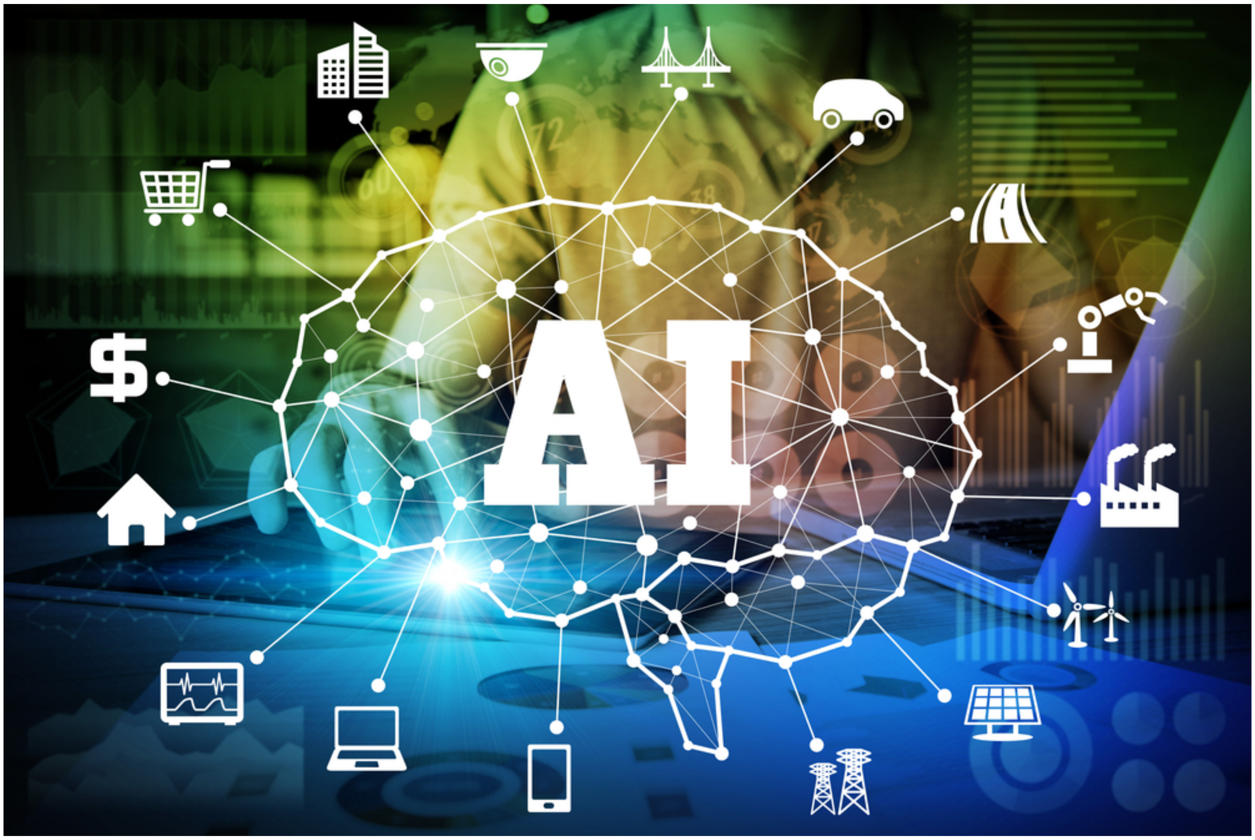
Project 3

AI IN HOSPITALITY INDUSTRY

HOTEL RECOMMENDATION SYSTEM



Contributions: Mohammed Shoaib Khan, Soumyajit Ghosh, Sarthak Gupta, Diksha Aggarwal, Rutuja Pabale

Prototype Selection

In the final product we transform implicit information provided by users into explicit features for the hotel recommendation system engine. There are two parts to this recommender engine using hotel attributes and reviews by users respectively to build two separate recommendation engines. With this recommendation system users can search for any similar hotel by inserting the place where they aint a similar hotel which they already know on the basis of features or by customer reviews.

A) Feasibility: This project can be developed and deployed by all the third party sites that help with the hotel booking or by the hotel chains to help the customer find the perfect room type.

B) Viability: As the retail industry grows in India and the world, there will always be small businesses existing which can use this service to improvise on their booking and data warehousing techniques. So, it is viable to survive in the long-term future as well but improvements are necessary as new technologies emerge.

C) Monetization: This service is directly monetizable as it can be directly released as a service on completion which can be used by businesses.

Abstract

Industries that are adamant on incorporating new technology revolutions are more likely to go backward in the 21st century. Businesses all around the world now understand how crucial it is to use modern digital technologies to promote continuous revenue development. In the realm of digital solutions, there have been amazing advancements and achievements during the past ten years. Artificial intelligence is one of these innovative technological advancements (AI). The goal of this article is to emphasize the contribution that robotics and artificial intelligence (AI) have made to the hospitality sector. the integration of multiple technologies to enhance customer experience and service in the travel industry. This study focuses on the next problems and changes that will affect hotels and restaurants.

Problem Statement

The hotel sector has seen tremendous transformation in the past 15 years due to the growth of internet technology and services. In this trillion-dollar sector, only a few businesses have developed monopolistic positions throughout the years. The most crucial components of the ecosystem—the value of host and guest— were removed in favor of complex, centralized ecosystems. Since there is no way to draw attention to particularly intriguing or distinctive aspects of their offering, several hotels fall short in terms of facility descriptions. This centralized, recycled strategy produces monotonous material and inhibits the sector from developing. The primary difficulties for service providers are to develop an exceptional and unforgettable experience in order to offer lodging and meals.

In this project, the objective is to transform implicit information provided by users into explicit features for the hotel recommendation system engine. There are two parts to this recommender engine using hotel attributes and reviews by users respectively to build two separate recommendation engines.

Market Need Assessment

Hospitality firms must adapt to this transition as new technology and applications become crucial components of customer behavior. Today, the hotel industry, where comfort-defining advancements are most rapidly incorporated, has sophisticated its entire system with the adoption of many innovative methods used for providing satisfying customer service. Every trip or encounter with a tourist brand needs to be extremely personalized and distinctive in every way, according to the industry. The expectations of the customer are consequently paramount, and firms work to comprehend the sophisticated thought processes of travelers as well as how they respond to social media and technology. The hospitality sector must now put a lot of effort into making sure that customer wishes can be deciphered and understood.

Target Specifications

Target marketing has become more popular as a result of the realization that consumers' requirements, wants, resources, tastes, and purchasing habits vary. One of the key strategies in marketing that helps discover different customer groups is market segmentation. The requirements, wants, attitudes, buying patterns, media consumption, price sensitivity, and other traits shared by these groups are comparable. In order to meet consumers' requirements, wants, and preferences more precisely than a mass marketing approach could, segmentation aims to find homogenous groups of customers. This increases marketing efficiency and effectiveness. The appraisal of market segments and the choice of target market segments are steps in the market targeting process. Businesses should look at the market while considering different market sectors.

External Research

To Know more about AI in hospitality
https://marutitech.com/hotel-industry-ai-awesome-user-experience/

The dataset is the "515K Hotel Reviews Data in Europe" dataset on Kaggle **(https://www.kaggle.com/jiashenliu/515k-hotel-reviews-data-in-europe)**. The dataset is a csv file, containing most text. The positive and negative reviews are already in columns. The reviews are all in English, collected from Booking.com from 2015 to 2017.

The dataset contains 515738 reviews for 1492 luxury hotels in Europe.

Data files are not included in this repository because of the large size.

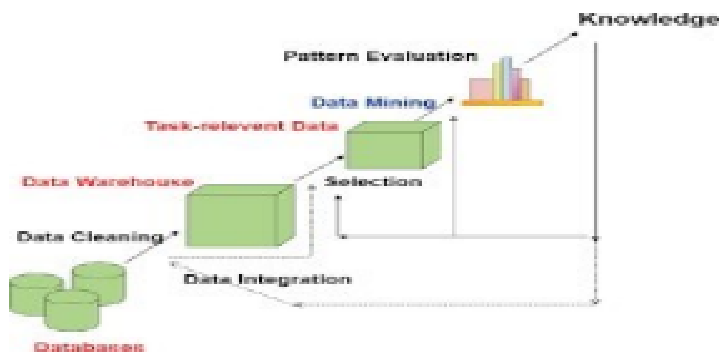
Concept Generation

Finding a success story of a hospitality firm X making a technological change Y that raised sales by Z percent while excluding the terms "machine learning" or "artificial intelligence" is becoming more and more difficult. These two ideas, though occasionally used synonymously, have distinct meanings. Computer science's large field of artificial intelligence investigates how to make machines think and behave like people. An important but not exclusive subset of AI is machine learning. In a word, machine learning is the process of creating models that, using the input data, accurately anticipate the end result. It makes use of statistical techniques to allow machines to become more accurate as more data is supplied into the system. The final output of machine learning models depends on the:

1. The more the diversity and richness of the data, the better the computer can identify patterns and the more accurate the outcome. Companies frequently have to go on a real quest for solid datasets due to the rising demand for high-quality data.

2. Features are useful inputs that are already present in the data, such as user gender, location, browser extension, etc. The key characteristics must be chosen since data typically contains more information than is required to create the model. Depending on how valuable they are for analysis, the characteristics are either chosen or rejected throughout this procedure by the analyst or the modeling tool.

3. A data analysis algorithm searches for patterns or trends before determining the best model parameters. Choosing the optimal algorithm to complete a task can be difficult since every algorithm produces a distinct outcome, and some algorithms provide more than one type of result. How a machine learning-powered model is created is as follows:



Machine learning models can outperform classical rigid business intelligence where business rules cannot capture the hidden patterns. Travel companies are actively implementing AI & ML to dig deep in the available data and optimize the flow on their websites and apps, and deliver truly superior experiences.

How can Artificial intelligence solve problems in the hotel industry?

Apple's Siri started providing voice-activated assistance to its mobile users to an extent that it has become almost a norm now. Amazon Echo and Alexa have also joined the race of creating a richer, more delightful customer experience using the power of machine learning of AI software.

Ever since the artificially intelligent system has crept into the hotel industry, the hospitality sector is abuzz with AI’s ability to learn about customers using its data analytics platform that helps hotel staff create a better frame of customers. Utilizing the full potential of AI software, they can capture a gamut of information about:

· Customer Purchases

· Travel choices

· Journey patterns and itinerary

· Location preferences

· Hotel rating inquiries

· Payment methods

The knowledge gathered thus can further be translated into providing insightful experience to hotel guests as they travel, inquire, stay and enjoy the luxurious hotel amenities.

Project Implementation

Dataset

	Hotel_Address	Additional_Number_of_Scoring	Review_Date	Average_Score	Hotel_Name	Reviewer_Nationality	Negative_Review	Review_Total_Negative_Word_Counts
0	s Gravesandestraat 55 Oost 1092 AA Amsterdam ...	194	8/3/2017	7.7	Hotel Arena	Russia	I am so angry that i made this post available...	397
1	s Gravesandestraat 55 Oost 1092 AA Amsterdam ...	194	8/3/2017	7.7	Hotel Arena	Ireland	No Negative	0
2	s Gravesandestraat 55 Oost 1092 AA Amsterdam ...	194	7/31/2017	7.7	Hotel Arena	Australia	Rooms are nice but for elderly a bit difficul...	42
3	s Gravesandestraat 55 Oost 1092 AA Amsterdam ...	194	7/31/2017	7.7	Hotel Arena	United Kingdom	My room was dirty and I was afraid to walk ba...	210
4	s Gravesandestraat 55 Oost 1092 AA Amsterdam ...	194	7/24/2017	7.7	Hotel Arena	New Zealand	You When I booked with your company on line y...	140

```
df.info() # Shows the datatype of the columns
```

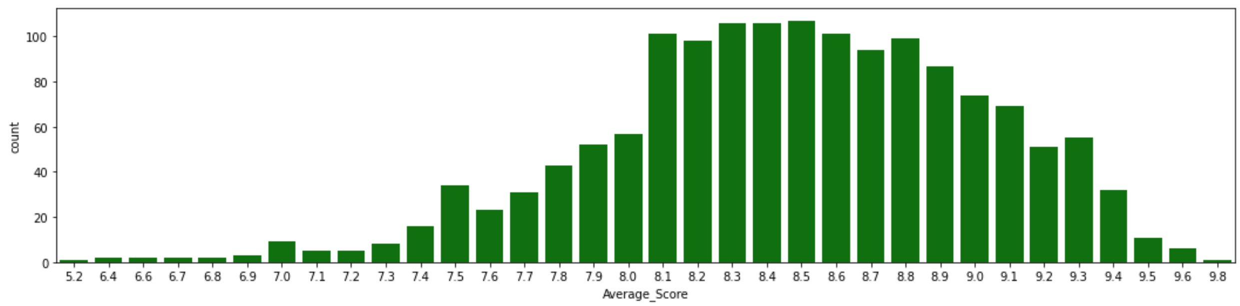
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 515738 entries, 0 to 515737
Data columns (total 17 columns):
 #   Column                                      Non-Null Count   Dtype  
---  ------                                      --------------   -----  
 0   Hotel_Address                               515738 non-null  object 
 1   Additional_Number_of_Scoring                515738 non-null  int64  
 2   Review_Date                                 515738 non-null  object 
 3   Average_Score                               515738 non-null  float64
 4   Hotel_Name                                  515738 non-null  object 
 5   Reviewer_Nationality                        515738 non-null  object 
 6   Negative_Review                             515738 non-null  object 
 7   Review_Total_Negative_Word_Counts           515738 non-null  int64  
 8   Total_Number_of_Reviews                     515738 non-null  int64  
 9   Positive_Review                             515738 non-null  object 
 10  Review_Total_Positive_Word_Counts           515738 non-null  int64  
 11  Total_Number_of_Reviews_Reviewer_Has_Given  515738 non-null  int64  
 12  Reviewer_Score                              515738 non-null  float64
 13  Tags                                        515738 non-null  object 
 14  days_since_review                           515738 non-null  object 
 15  lat                                         512470 non-null  float64
 16  lng                                         512470 non-null  float64
dtypes: float64(4), int64(5), object(8)
memory usage: 66.9+ MB
```

The csv file contains 17 fields. The description of each field is as below:

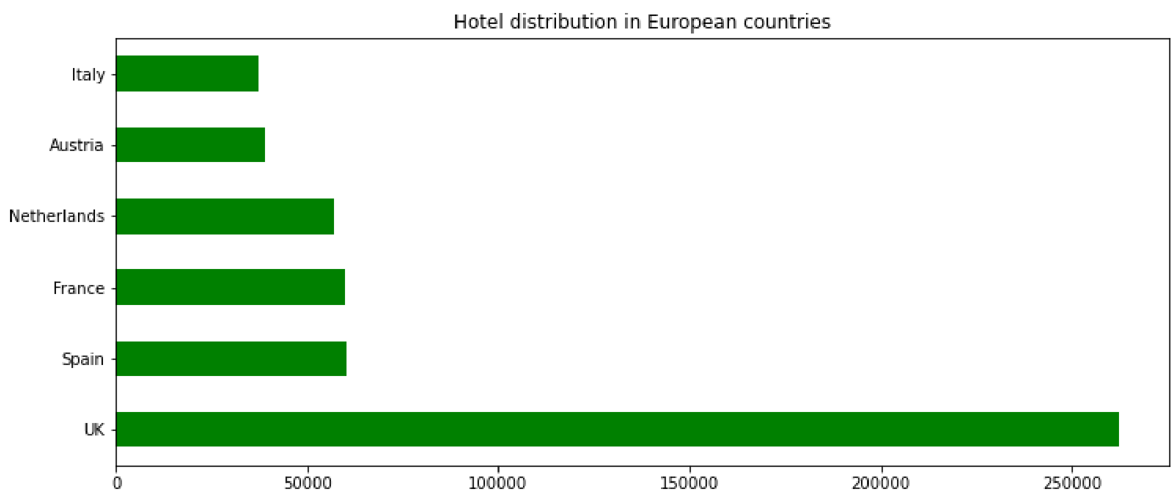
1. **Hotel_Address**: Address of hotel.
2. **Review_Date**: Date when reviewer posted the corresponding review.
3. **Average_Score**: Average Score of the hotel, calculated based on the latest comment in the last year.
4. **Hotel_Name**: Name of Hotel
5. **Reviewer_Nationality**: Nationality of Reviewer
6. **Negative_Review**: Negative Review the reviewer gave to the hotel. If the reviewer does not give the negative review, then it should be: 'No Negative'
7. **ReviewTotalNegativeWordCounts**: Total number of words in the negative review.
8. **Positive_Review**: Positive Review the reviewer gave to the hotel. If the reviewer does not give the negative review, then it should be: 'No Positive'
9. **ReviewTotalPositiveWordCounts**: Total number of words in the positive review.
10. **Reviewer_Score**: Score the reviewer has given to the hotel, based on his/her experience
11. **TotalNumberofReviewsReviewerHasGiven**: Number of Reviews the reviewers have given in the past.
12. **TotalNumberof_Reviews**: Total number of valid reviews the hotel has.
13. **Tags**: Tags reviewer gave the hotel.
14. **Days since review**: Duration between the review date and scrape date.
15. **AdditionalNumberof_Scoring**: There are also some guests who just made a scoring on the service rather than a review. This number indicates how many valid scores without review there.
16. **lat**: Latitude of the hotel
17. **lng**: longitude of the hotel

Data Exploration

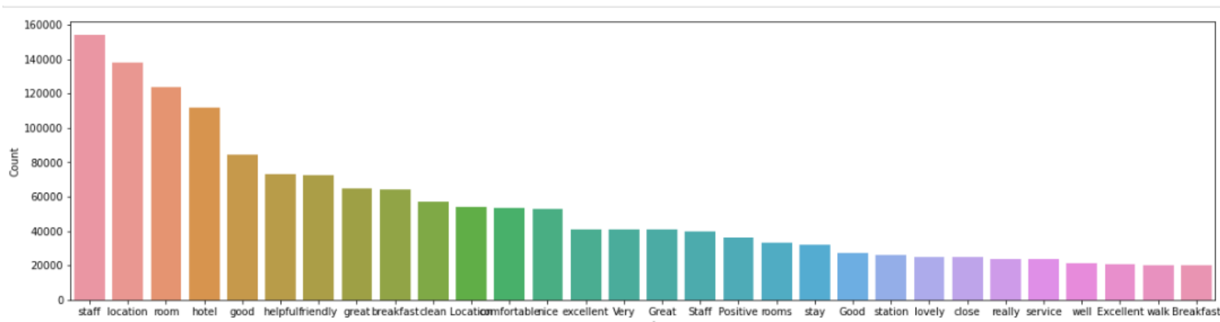
1. Graph of counts vs Average score



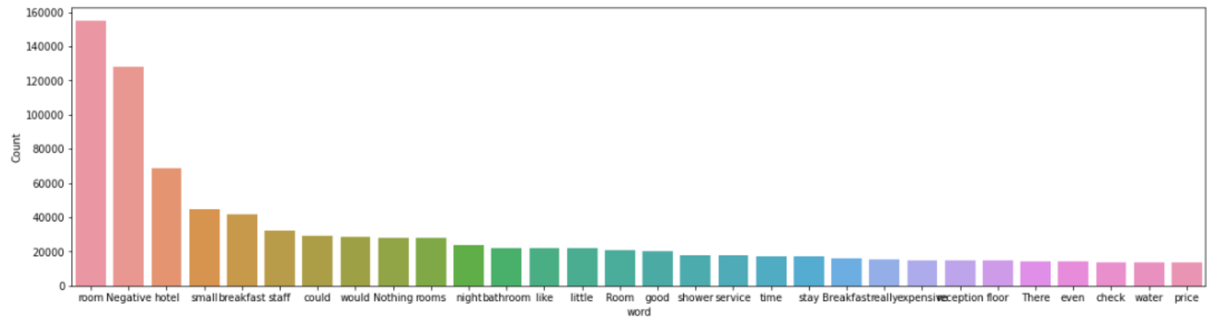
2. Hotel distribution by countries



3. Frequency of words in positive reviews after removing stop words



4. Frequency of words in negative reviews after removing stopwords



5. Word Clouds representation of positive reviews

6. Word Clouds representation of negative reviews

Data Preprocessing

1. · Installed NLTK Natural language processing toolkit.
2. · Checked the most frequent words
3. · Checked the positive and negative reviews and their counts.
4. · Convert positive and negative reviews into word clouds.

Feature Engineering :

1. Implemented feature engineering and made dummy-variables.
2. Count the number of same words and remove a few of them.
3. Converting the data into lowercase.

4. Identification of type or rooms in the dataset.
5. Identification of types of hotel characteristics.

Modeling:

1. Subsetting the data into useful characteristics.
2. Grouping the data into similar categories for recommendation.
3. One hot encoding of new attributes.
4. Finding cosine similarity between data point vectors.
5. Predicting the recommendation based on cosine similarity.

Business Model

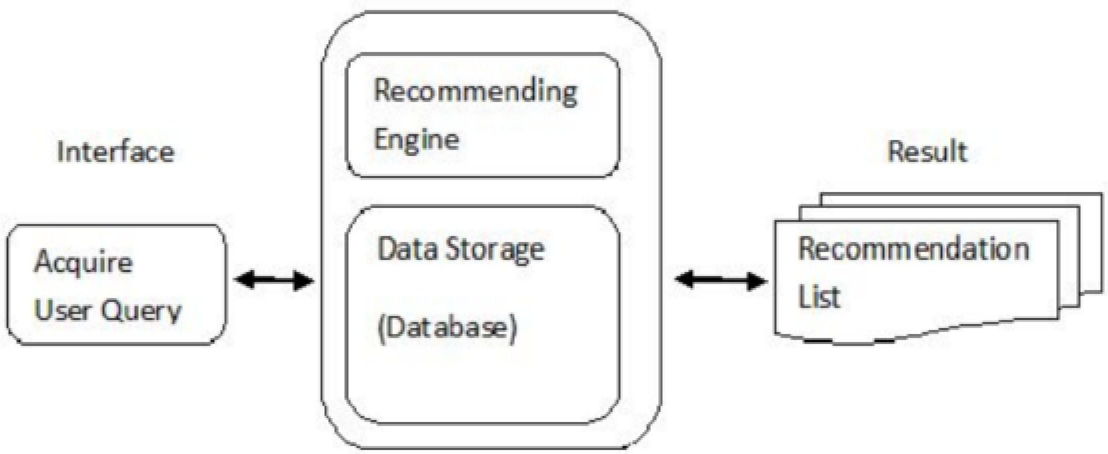
In this part of the report, we will look at the business model suggested for the idea presented earlier. There are many business models available but we have chosen the 'SaaS or Software as a Service business model' which is the one suited for our idea.

SaaS Business Model

SaaS or Software as a Service business model is a centrally-hosted software that is hosted on a cloud infrastructure. Customers pay a subscription fee to utilize the software.

This is a customer service hotel recommender system and can be utilized by small, medium, and even large businesses. It provides a better customer service experience.

This is a classic example of something used on mass markets with a high margin of return.



Mission

To transform implicit information provided by users into explicit features for the hotel recommendation system engine.

Vision

To provide a recommendation system aimed at suggesting relevant hotel to users based on their recommendations and choices.

The Executive Summary

In a very general way, recommender systems are algorithms aimed at suggesting relevant items to users (items being movies to watch, text to read, products to buy or anything else depending on industries).

There are two main data selection methods:

Collaborative-filtering: In collaborative-filtering items are recommended, for example hotels, based on how similar your user profile is to other users', finds the users that are most similar to you and then recommends items that they have shown a preference for. This method suffers from the so-called cold-start problem: If there is a new hotel, no-one else would've yet liked or watched it, so you're not going to have this in your list of recommended hotels, even if you'd love it.

Content-based filtering: This method uses attributes of the content to recommend similar content. It doesn't have a cold-start problem because it works through attributes or tags of the content, such as views, Wi-Fi or room types, so that new hotels can be recommended right away.

The point of content-based is that we have to know the content of both user and item. Usually you construct user-profile and item-profile using the content of shared attribute space. For example, for a movie, you represent it with the movie stars in it and the genres (using a binary coding for example).

There are a number of popular encoding schemes but the main ones are:

· One-hot encoding

· Term frequency–inverse document frequency (TF-IDF) encoding

· Word embeddings

In this project, we will be discussing content-based filtering of the recommender engine, turning implicit attributes into explicit features for hotel recommender engines.

Situation

With the evolution of new web technologies, the recommender systems (RS) are getting significant attention by the business people as well as customers due to its role in better e-commerce, refined business strategy, improved customer's satisfaction, etc.

Tourism is one of the most famous and powerful industries in the world which has a huge impact on the world's total GDP or employment. Tourism is linked with hotels because tourists always wanted to know about the hotels where they are going to stay in their tour. In recent years, online hotel booking has become one of the primary choices of the hotel customers.

So, this seems to be an excellent in demand prototype.

Value proposition

Everyone likes their products to be personalized and behave the way they want them to. Given a user, recommender systems aim to model and predict the preference of a product.

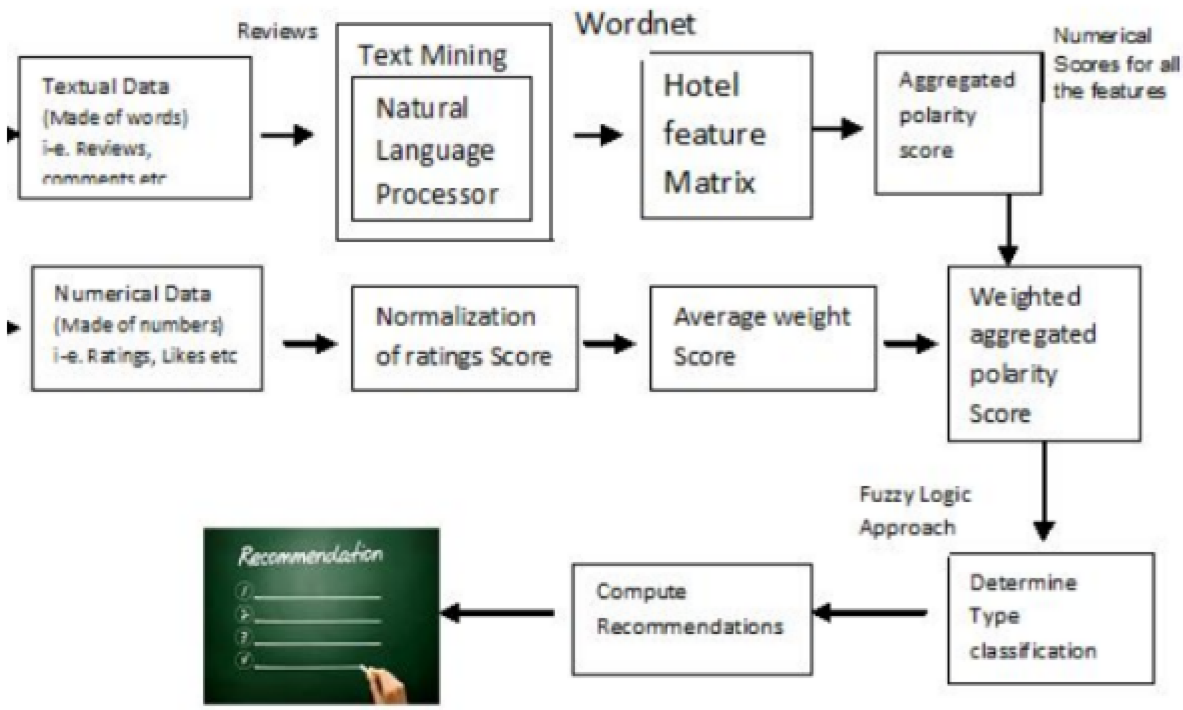
The proposed system uses the heterogeneous nature of data (textual, numerical) crawled in from World Wide Web (www). Data is obtained from the selected hotel websites (data sources) containing the keywords present in the active user search query. The data is usually found in the form of numeric (such as votes, ranks and number of video views) and text (such as reviews and comments). To get true recommendations our system has used ranks, votes and reviews data to extract Hotel features.

License and Permit

- Shops and Establishment Registration
- Tax Registration
- Website registration
- In Future – Trade License

Implementations setup

System Overview



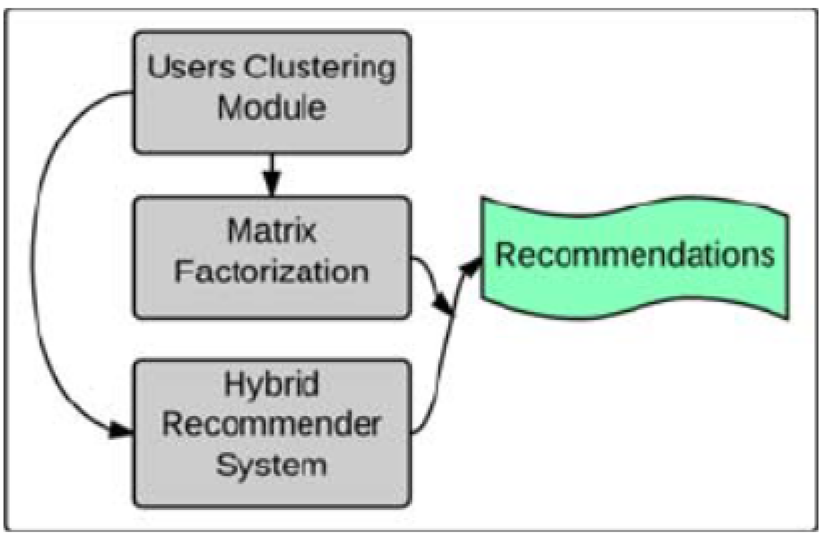
Marketing Strategies

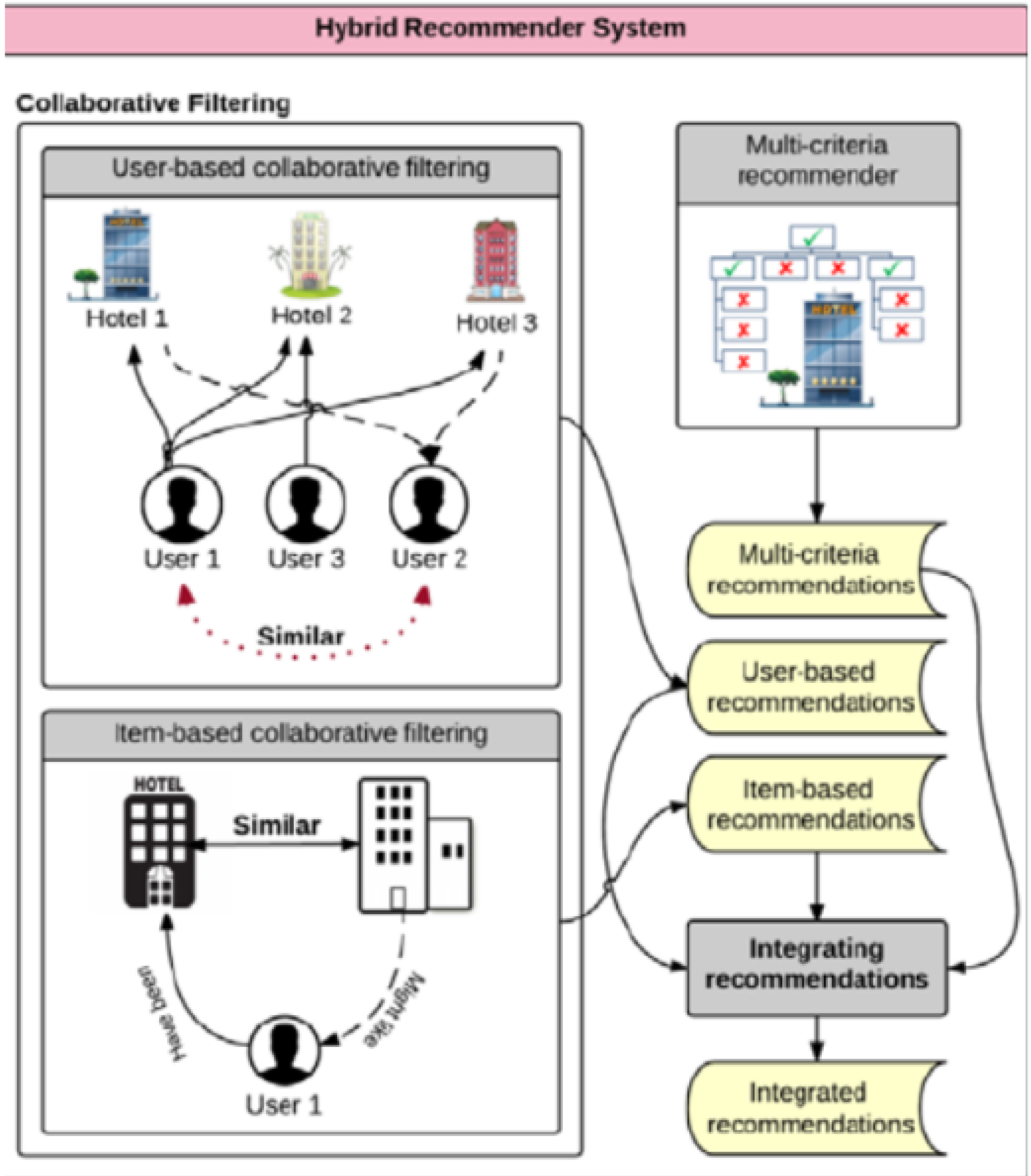
- List your hotel on Google
- Have your website
- SEO: The backbone of websites
- Ads, ads and more ads!
- Go social the social media way
- Tell your story through video
- Don't forget email marketing
- Take advantage of influencer marketing
- Blog it right
- Practice the trend: Chatbots
- Maintain your online reputation

The Recommendation System

The recommender engine contains three main modules, namely user clustering, matrix factorization, and the hybrid recommender system. Users are first clustered based on various features. The selected cluster is then fed into the matrix factorization module and the hybrid recommender system. The output of the mentioned two modules forms the final set of recommendation

The recommender engine contains three main modules, namely user clustering, matrix factorization, and the hybrid recommender system. Users are first clustered based on various features. The selected cluster is then fed into the matrix factorization module and the hybrid recommender system. The output of the mentioned two modules forms the final set of recommendations.





Performance Evaluation

We did several experimental evaluations to assess the performance of the proposed recommender system. The leave- one-out cross validation (LOOCV) approach was selected for validating the results. In LOOCV with n data points, 1 observation (data point) is considered as the validation set in each run, while the remaining data points form the training set. The procedure is repeated n times, taking all data points as the validation set once. A set of decision-based error measures, We did several experimental evaluations to assess the performance of the proposed recommender system. The leave-one-out cross validation (LOOCV) approach was selected for validating the results. In LOOCV with n data points, 1 observation (data point) is considered as the validation set in each run, while the remaining data points form the training set.

The procedure is repeated n times, taking all data points as the validation set once. A set of decision-based error measures, as three prediction-based error metrics, i.e. mean absolute error (MAE), mean squared error (MSE), and root-mean- square error (RMSE), were calculated for evaluating the proposed recommendation system. as three prediction-based error metrics, i.e. mean absolute error (MAE), mean squared error (MSE), and root-mean- square error (RMSE), were calculated for evaluating the proposed recommendation system.

We did several experimental evaluations to assess the performance of the proposed recommender system. The leave-one-out cross validation (LOOCV) approach was selected for validating the results. In LOOCV with n data points, 1 observation (data point) is considered as the validation set in each run, while the remaining data points form the training set. The procedure is repeated n times, taking all data points as the validation set once. A set of decision-based error measures, i.e. accuracy, specificity, sensitivity, and informedness, as well as three prediction-based error metrics, i.e. mean absolute error (MAE), mean squared error (MSE), and root-mean-square error (RMSE), were calculated for evaluating the proposed recommendation system.

Financial Modeling

This model could be successfully implemented in hotel booking apps, travel websites, etc. Some of the stats representing this market along with their sources are given below.

Getting potential guests to notice you

According to these seven statistics, most people make their travel plans on the Internet, and a large part of their research involves reading a lot of online reviews. Therefore, if your hotel has a weak online presence, with few to no reviews, then potential guests will overlook you in favor of hotels with more reviews. Some ways to get more reviews include claiming your profile on review sites, routinely asking for reviews, and making it as easy as possible for your guests to leave reviews.

1. 70.9% of travelers say online content influences their choice of where to stay. (RMS)
2. 81% of people frequently or always read reviews before booking a hotel. (TripAdvisor)
3. 52% of individuals would never book a hotel that had zero reviews. (TripAdvisor)
4. 55% of travelers read several pages worth of reviews to get a better sense of public opinion. (TripAdvisor)
5. 66% of travelers plan to spend more time reading reviews about destinations. (TripAdvisor)
6. 65% of people gain their travel inspiration from online searches. (Google)
7. 72% of travelers frequently or always read reviews before deciding where to stay or what to do. (TripAdvisor)

Gaining consumers' trust

These statistics reveal that reviews are an important form of social proof. People feel they can trust you if you garner high star ratings and offer courteous responses to guest reviews, as doing so demonstrates that you care about giving your customers the best possible experience. These things also highlight your hotel's transparency and authenticity.

8. When deciding between two similar properties, 79% of consumers are more likely to reserve a room at the hotel with a higher rating. (TripAdvisor)
9. Customers will value guest ratings over a hotel's brand 72% of the time. (Expedia)
10. Over 40% of travelers will leave a review if they have a positive experience at a hotel. (ReviewTrackers)
11. 48% of guests will leave a review after a bad hotel experience. (ReviewTrackers)
12. Travelers read an average of 9 reviews before deciding to book a hotel. (TripAdvisor)

13. 91% of travelers want property owners to respond to negative reviews. (Expedia)

Helping you earn more from each booking

This group of statistics reveals the relationship between your reputation and your bottom line. Unsurprisingly, consumers are willing to pay more for a quality hotel experience, and the amount they are willing to pay directly corresponds with the level of star rating a hotel has. As such, you should make sure potential guests can see a surplus of positive online reviews. One way to do this is to add a link to your reviews on your website. You can also share impressive customer feedback on your social media channels. This way, when people search for your hotel's name, they can easily find lots of positive word of mouth about your property.

14. Guests will pay 24% more for a hotel with a 3.9 rating over one that's rated 2.4. (Expedia)
15. Guests will pay 35% more for a hotel with a 4.4 rating over one with a 3.9 rating. (Expedia)
16. A 1-star rating increase leads to a 2.2–3.0% rise in hotel monthly revenues. (Yong Chen and ATM Sayfuddin)
17. Travelers spend an average of 71% of their time researching their trip online. (Google)
18. There is an average 27.8% higher demand for hotels rated one star higher on all review platforms. (Lewis and Zervas)
19. Properties see 10% more net room nights when guest reviews show how properties handle COVID-19 mitigation. (Expedia)
20. Over 90% of travel and hospitality business owners think online reviews are among the three most important factors affecting the future of their industry. (TripAdvisor)

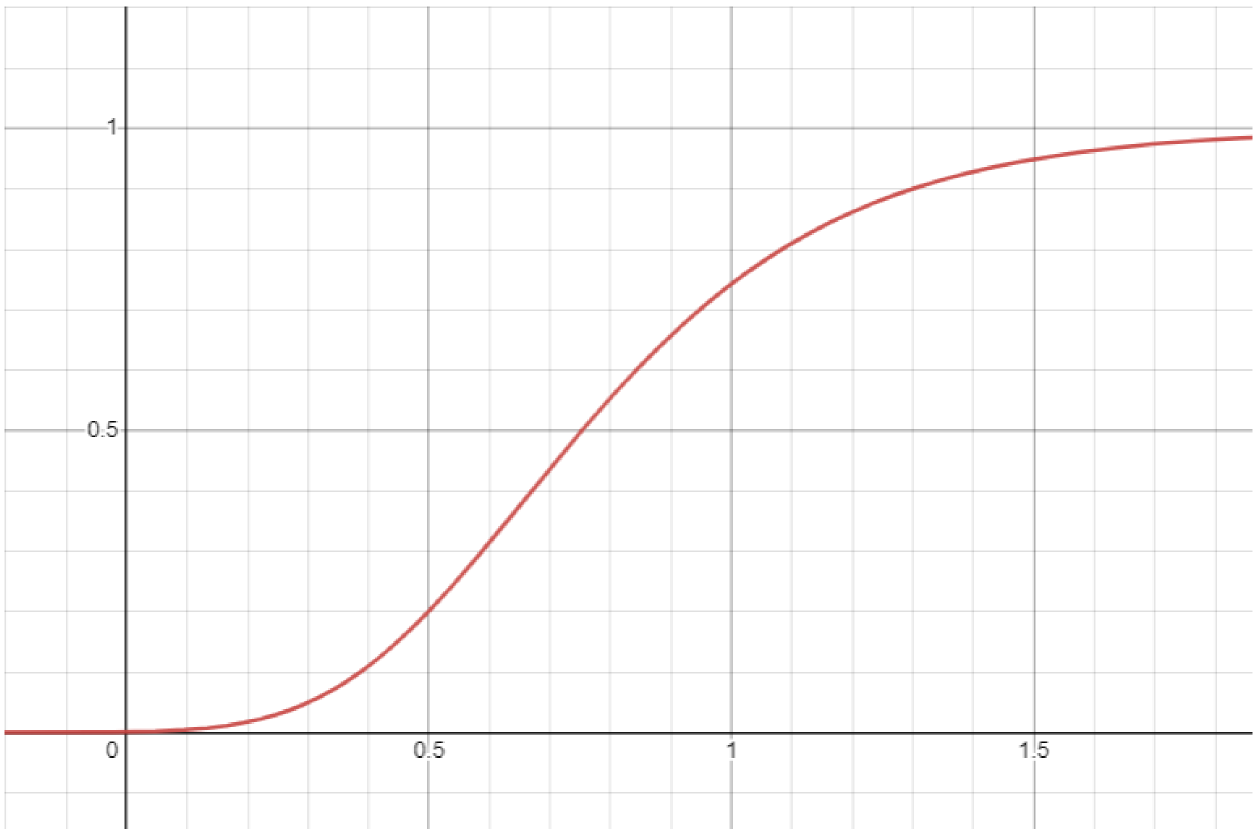
Hotel booking trends

1. The most valuable travel and tourism brand in the world, according to Brand Finance's 2021 list, isn't an accommodation company or an airline. It's Booking.com, which has a brand value of $8.9B. Hoteliers that want to reach a wide audience of potential guests can take advantage of Booking.com's massive marketing power.
2. Travelers visit an average of 38 websites before making a booking, an Expedia survey found in 2015.
3. According to Kalibri Labs, 27% of hotel bookings are made through the property directly, 25% through the hotel's own website, and 16% through online travel agencies in 2019.
4. The average booking window, or lead time, for hotel bookings in the US is about 25 days, according to Kalibri Labs. The average length of stay is 1.8 nights.
5. A 2015 study by the Global Business Travel Association found that people took 1.3 million business trips in the US each day.
6. Historically, about two-thirds of hotel industry revenue was driven by business travel. A 2021 study by Bloomberg found that 84% of global CEOs planned to spend less on business travel post-pandemic.
7. Millennials, especially millennial leisure travelers, are slightly less likely to be part of hotel loyalty programs than their older counterparts, according to a 2019 PwC study. Millennials were found to be part of 3 loyalty programs compared to 3.6 program memberships of older travelers.

Financial Equation

With advances in aviation and boost in the economy in the last few decades, the travel industry has been evolving as the people now have money to spend. This growth has yet to reach its peak as India is still a developing country. It is predicted that India would be the second largest economy in the world by 2050. More economic growth leads to more income which leads to more travel. We can see that the previous generations didn't travel a lot because of money and time constraints. But, people are traveling more and more now, this will continue to grow with the growing economy.

So, the hotel market will continue to grow with time. However after a certain point it will reach its peak and come to a saturation point but, that doesn't mean that there is not much capital involved even if it reaches its peak the the money from the hospitality industry will remain huge.



The equation representing the above curve is as follows

$$y = \left(\frac{1}{1 + e^{-kx}} \right)^{a}$$

y: represents growth

x: represents time

k: is the growth rate

a: represents the starting position of curve

Conclusion

Big data analysis is changing the operating mode of the global tourism economy, providing tourism managers with deeper insights, and infiltrating into all aspects of tourist travels, while driving tourism innovation and development . Tourism text big data mining techniques have made it possible to analyze the behaviors of tourists and realize real-time monitoring of the market.Both machine learning and current deep learning with high achievements have been greatly applied in NLP.

With the increasing number of applications on the Internet, the source of data is getting richer and richer. Therefore, the various factors in the new data bring new challenges. It is also a chance to create novel methods to achieve better recommendation results. Social networks are still the focus of the recommendation research, integration methods and new algorithms will continue to appear in the future. The sound, location and other user preference information are receiving more and more attention. I believe the future of the recommender system will be a hot area of innovation and research.

References

Li, J.; Xu, L.; Tang, L.; Wang, S.; Li, L. Big data in tourism research: A literature review. Tour. Manag. 2018, 68, 301–323.

Qin Li 1,2, Shaobo Li 3,4,* , Sen Zhang 1,2, Jie Hu 5 and Jianjun Hu 3,A Review of Text Corpus-Based Tourism Big Data Mining

Github Link

https://github.com/skhan4784/Hotel-Recommender