



Knowledge based topic retrieval for recommendations and tourism promotions

Ram Krishn Mishra*, J Angel Arul Jothi, Siddhaling Urolagin, Kayan Irani

Department of Computer Science, BITS Pilani, Dubai Campus, Dubai, United Arab Emirates

ARTICLE INFO

Keywords:

Tourism
Online reviews
Machine learning
Recommender system
Predicting ratings
Feedback model
Yelp
Restaurants

ABSTRACT

The Users regularly share their views on services through internet reviews. Digital tourism allows travelers to plan and manage their trips ahead of time using the internet. The proposed system will enable users to create recommender systems that can help to find new places as per individual interests. The portals such as Tripadvisor and Yelp assist travelers in decision-making and providing feedback. The data available in the form of reviews, blogs, and rating help the researcher build a better recommendation and promotion system. The recommendation system will help tourists discover restaurants often ignored due to a lack of promotion. This research proposes an automated system to perform three subtasks: predicting star ratings from reviews, a feedback model, and a knowledge-based recommender system. The experimental work results show that Random Forest and Decision Tree classifiers give the highest accuracy for predicting star ratings. The research utilizes clustering and topic modeling to identify the topic from the reviews. Moreover, this research uses salience and valence of the topic to generate the feedback model. Furthermore, a recommender system is developed using user preferences, restaurant knowledge, and the valence of each topic of each restaurant. Inverse Document Frequency and K-means clustering have been used to improve the recommendation quality and construct a CSR matrix. This system will link restaurants with similar reviews to ensure that customers find restaurants almost identical to the ones they enjoyed in the past.

1. Introduction

Over the years, several distinguished professionals have provided various definitions for tourism. Among the many definitions for tourism, the most appropriate for this research is provided by Oxford Dictionary: "The business activity connected with providing accommodation, services, and entertainment for people who are visiting a place for pleasure" (Chica et al., 2022). Today, there are several forms of tourism depending upon the purpose of the visit. Some sub-categories of tourism are cultural tourism, adventure tourism, and heritage tourism.

With the development of technology, we have observed an increase in social media (Grover et al., 2022) and review platforms (Sahu & Ahuja, 2016). These platforms now serve as electronic Word of Mouth (e-WOM) and are often used by users to express opinions via reviews and ratings. User reviews are an essential information source that provides more insight into businesses (Sharma et al., 2022). The reviews and ratings say the user's delight or discontent with the business. Moreover, reviews play an essential role in decision-making. In other words, users give great importance to online reviews when deciding which services to employ.

Tourism can impact a country in three categories: economic, socio-cultural (Gandhi & Kar, 2022), and environmental. Furthermore, each category can have positive or negative impacts. There are different types of economic impacts that tourism can have. For instance, tourism can impact employment in an area, as indicated by the findings from research on the economic impact of tourism in Central Finland. Tourism involves several sectors, such as accommodation, tourism services, transportation, food, and beverages (Kumar et al., 2022). The food and beverage (F&B) sector consists of restaurants and drinking establishments that provide meals and beverages to the consumer. Since the F&B is an indispensable aspect of tourism sector, it can mobilise the economy of a tourist site (Tohmo, 2018).

In the case of tourism, online platforms such as TripAdvisor and Yelp are used to gather information about tourism and its related businesses. Specifically, these online reviews play an essential role in decision-making during travel planning and choosing restaurants (Alnogaithan et al., 2019, Theses et al., 2017, Gretzel & Yoo, 2008). Also, tourism businesses can analyse these online reviews to improve their services and decrease customer dissatisfaction (Kumar et al., 2022) (Adomavicius et al., 2005). However, the sheer number of

* Corresponding author.

E-mail address: rkmishra@dubai.bits-pilani.ac.in (R.K. Mishra).

online reviews makes it challenging to discover valuable insights (Kushwaha et al., 2022). Consequently, this impacts the management or authorities in charge of the business and the tourists. As the number of online reviews grows, researchers have been trying to study these reviews to garner insights about tourists and tourism businesses. Over the years, the application of machine learning algorithms and techniques has been extended to online reviews by researchers. Researchers have used various methodologies to analyse online reviews by developing automated models. Additionally, online reviews are used to recommend new suggestions to users, such as restaurants or tourist places (Qin et al., 2022).

A recommendation system (RS) involves three major elements: users, items and transactions. Users interact with online e-stores by buying products, selling products, and acting as brokerage agents. Data released by the International Telecommunication Union (ITU) on June 2020 estimates that around 61% of people worldwide rely on the internet to support their regular activities ("Statistics." 2022). Amazon sells more than 12 million products to fulfil customers' daily needs through its eCommerce site ("Ecommerce 101: The History and Future of Online Shopping." 2022). RS is a dynamic field where many innovative machine learning algorithms are proposed, and a number of information filtering techniques are applied.

RS is an active research area where several machine learning algorithms, information filtering approaches, and data mining techniques are applied. RS analyses the characteristics of the user or item, or transaction for delivering better recommendations. Recommendations are provided to individual users by analysing their likings or on a general basis. Based on these criteria, a generalised recommendation system and a personalised recommendation system are built. A generalised recommendation system recommends a standard generic product to all users across the globe. It suggests the current market trends and new products introduced into the market. Based on the previous user activities, the recommender system provides customised product recommendations.

RS also plays a major role in analysing user preferences in detail, such as the likes and dislikes of each product.

This research focuses on online reviews for sentiment analysis or rating prediction. The analysis helps in understanding the shared sentiment about a product or a service. However, researchers are yet to take up the consumer sentiment for specific areas in a product. The primary objective of the proposed model is to use the reviews and identify the sentiments of users in different areas or dimensions of a restaurant and develop a customised model for every product or service.

This research utilises the Yelp reviews of restaurants in the United States of America (USA) in three approaches: firstly, the development of a feedback model that will assist the restaurant authorities in understanding the current status of their business and help in decision-making. Secondly, the prediction of ratings from reviews, and finally, building a knowledge-based recommender system using existing information about restaurants, user's responses to queries, topics of restaurant reviews and the valence of those topics.

The structure of this paper is organised as follows. Section 2 is a review of previous research work done in this field. In Section 3, the methodology is described. In Section 4, the experimental results are provided. Section 5 is a discussion of the inferences drawn from the results. Section 6 is the conclusion of the research.

2. Literature survey

Researchers have used several methods to analyse online reviews for various tasks like understanding the effects of reviews, sentiment analysis, recommendation systems, providing feedback, and rating prediction. The following paragraphs present the literature on analysing online reviews.

Digital Tourism refers to the practice of managing, organising, and planning tourist trips using digital devices. With the advent of the internet (Asrani and Kar, 2022), finding information and reviews about the

places, you plan to visit has become much easier. Due to the digital flood of information, it has now become possible for people (Sharma et al., 2022) to evaluate the locations they intend to see without even leaving their houses. While simple star ratings can indicate how good or popular a place is, text-based reviews (Kushwaha et al., 2022) allow us to get a more complete and nuanced assessment of any location. Digital tourism, combined with pictures and videos of the location, allows anyone in the world to have a good idea of what type of locations they plan to visit and the ones they wish to avoid. Digital data (Gandhi & Kar, 2022) (Gupta et al., 2022) also allowed recommender systems to take this information and create personalised trips suited to the user's needs based on their previous experiences and reviews. Digital tourism data has prompted research such as (Bawack et al., 2022) which details digital tourism's total uses and extent.

One of the most popular research projects performed on online reviews is sentiment analysis (Kushwaha et al., 2021). Sentiment analysis aims to determine the opinion about something from a piece of text. Sentiment analysis in business can provide helpful feedback to businesses and help them improve their products or service. Al Amrani et al. proposed a hybrid approach to perform sentiment analysis on online reviews and classify reviews into positive and negative classes. The research used two machine learning algorithms, Random Forest and Support Vector Machine, to build its hybrid model (RFSVM) and compared the performance of the hybrid model against the individual performance of Random Forest and Support Vector Machine. From the research, it was observed that RFSVM gave higher accuracy for sentiment analysis (al Amrani et al., 2018). Researchers have compared the performance of existing machine learning algorithms when used for sentiment analysis.

Moreover, studies have been conducted to improve the accuracy of current machine learning classifiers. The authors used genetic algorithms for feature selection and improved the accuracy of the Naive Bayes classifier when Naive Bayes was used for sentiment analysis (Masripah et al., 2020). Another research used Chinese reviews and developed a framework which identified aspect-opinion pairs and predicted the sentiment polarity of the pairs. Chen et al. has enhanced the neural networks (Kumar et al., 2021) by using external knowledge provided by a sentiment knowledge graph. The knowledge enhanced neural networks were used for sentiment analysis. The research studied the impact of the sentiment knowledge graph and observed that the sentiment knowledge graph improved the performance of neural network architectures in sentiment classification (Goel et al., 2021), (Jain et al., 2021) (Chen and Huang, 2019).

Recommendation system (RS) (Amzad and Vijayalakshmi, 2022) in general is classified into three categories: collaborative recommendation system (Mustafa et al., 2017), content-based recommendation system (Pazzani & Billsus, 2007) and hybrid recommendation system (Li et al., 2020). Other recommendation systems like knowledge-based recommendation systems, ontology-based recommendation systems, and instance-based recommendation systems are applied to specific applications. A collaborative recommendation system (CRS) works by grouping customers with similar opinions. CRS identifies a group of people who share the same opinion (Ezin et al., 2019). With the assumption that people with a similar opinion have similar likeability, it recommends the product purchased by person A to person B. It works directly on the ratings provided by the user. Ratings can be implicit or explicit. Explicit ratings range on a scale of 1-5 obtained directly from the user using feedback forms (Kyaw & Wai, 2019). Implicit ratings are extracted indirectly using a specific algorithm or approach example: items added to a cart or wish list are considered to be likeable by a user and given five ratings. These ratings corresponding to each user are put in a matrix where rows represent the users and columns represent items.

On the other hand, a content-based recommendation system recommends (Gulati & Bhagat, 2021) products by analysing the customer's past purchase history. Hence, the customer (Chatterjee et al., 2021) may have a common choice of products, which would be similar to his purchased products. A combination of both recommenda-

tions mentioned above is applied in a hybrid. Online e-stores like eBay (“Electronics et al.,” 2022), Amazon (Yelp Dataset, 2022), and Flipkart (“Online Shopping Site for Mobiles, Electronics, Furniture, Grocery, Lifestyle, Books and More. Best Offers!” 2022) uses hybrid recommendation system and strives hard to retain existing customers and attract new customers.

Researchers use online reviews to build recommender systems. A recommender system is an information filtering system that predicts the user’s preference for items and suggests things to the user accordingly. A recommender system helps an organisation to understand what the user wants (Mishra et al., 2020). Consequently, organisations can create or improve business services or products to meet user needs (Aggarwal, 2016). Lin et al. used online reviews to build a content-based recommendation system. The proposed method used topic modelling to derive product features and consumer preferences. The results were later used to provide recommendations to users. Nilashi et al. used online TripAdvisor’s hotel ratings to build a collaborative filtering recommendation system (Nilashi et al., 2018). The proposed method used multi-criteria ratings to improve the accuracy of the recommender system. The research used Self-Organising Maps (SOM) and Expectation Maximisation (EM) algorithms and Classification and Regression Trees (CART). Research has been done on predicting ratings from online reviews. Reddy et al. trained several classifiers to predict the ratings of restaurant reviews. The study inferred that the Bigram-Trigram model takes less time for computation and the performance is almost similar to that of the Random Forest classifier (Reddy et al., 2017).

3. Methodology

The framework of recommendation and tourism promotions are depicted in Fig. 1. Work proposes an automated system to analyse the reviews of restaurants in the USA. The methodology has been performed in three stages, first, predicting star ratings from reviews; second, a feedback model based on salience and valence calculation and finally, the knowledge-based system based on the salience and valence calculation.

3.1. Dataset

Yelp.com is a well-known review site founded in 2004 and expanded throughout Europe and Asia between 2009 and 2012. Yelp offers crowd-sourced comments on local businesses and officially provides the dataset (Yelp Dataset, 2022) for the research work. The dataset comprises 6 JSON files: business, review, user, check-in, tip and photo. This research considers reviews regarding restaurants in the USA. The dataset consists of 2,494,488 reviews about 21,160 restaurants in the USA. Each row in the Yelp dataset has information about a restaurant, such as a name, business id, city, state, cuisine, latitudes, longitudes, attributes, and a review and rating of the restaurant. Each rating is provided on a scale of 1 to 5.

3.2. Data pre-processing and vectorization

The pre-processing step aims to improve the quality of the reviews to achieve better results. As shown in Fig. 1, the reviews are pre-processed by removal of HTML tags, special characters, Unicode characters, and digits, conversion of uppercase letters to lowercase letters, removal of stop words and stemming. The pre-processing steps applied in this research are shown in the architecture and Table 1.

Vectorisation converts a pre-processed review into a set of linear features called vectors. This research uses Term Frequency-Inverse Document Frequency (TF-IDF) vectoriser to convert a given review into a vector (Mishra et al., 2019). Given a collection of reviews, the TF-IDF vectoriser is a statistical approach that captures a word’s importance in a review. It is a weighting factor that is applied in text mining, information retrieval and the semantic web.

Let R denote a review, $\{w_1, w_2, w_3, \dots, w_n\}$ denotes the set of n words in a review, and w_i denotes a word. The process of converting a review into a vector using the TF-IDF approach is as follows:

- Compute the term frequency $TF(w_i)$ for all the words in the review using Eq. (1). Term frequency represents the count of the number of times the term occurs in a review.
- Compute the inverse document frequency $IDF(w_i)$ for all the words in the review using Eq. (2). IDF is a measure of the information provided by a word across all reviews. It is generally used as an inverse of logarithmically scaled value.
- Compute the TF-IDF values for all the words in the review using Eq. (3)
- Concatenate the computed TF-IDF values into a vector

$$TF(w) = \frac{\text{number of times a word } w_i \text{ appears in the review}}{\text{Total number of words in the review}} \quad (1)$$

$$IDF(w_i) = \frac{\log(\text{Number of times a word } w_i \text{ appears in the review})}{\text{Total number of reviews with the word } w_i} \quad (2)$$

$$TF - IDF(w_i) = TF(w_i) * IDF(w_i) \quad (3)$$

The output obtained from the vectorisation step is called the review-term matrix. The number of rows in the review term matrix is the number of reviews, while the number of columns is the vocabulary size (i.e. the total number of unique words found in all the reviews).

3.3. Predicting star ratings and identifying the topics from the reviews

As already mentioned, each row in the Yelp dataset has information about a restaurant, such as a name, city, state, cuisine etc., along with a review and rating of the restaurant. Unlike Yelp Dataset, other datasets often do not have a rating associated with a review. However, working with datasets with only reviews hinders the data analysis processes, which involve gaining insights from reviews and ratings. Moreover, sometimes the rating provided by the user may not match the review. For example, a review could convey a positive sentiment about a restaurant, but the rating could be 2 or 3. Hence, the user can use a machine learning model to accurately allocate a rating for the review he chooses to provide. The above two reasons are the motivating factors for predicting star ratings from reviews.

The vectorised reviews and their associated ratings are given as input to a classifier model to predict star ratings from the reviews. In this work, several different classifiers are trained to predict the star ratings, and their performance is compared. The other classifiers used are Random Forest (Tripathi et al., 2021), Decision Tree (Neogi et al., 2021), Logistic Regression (Sufi, 2022), K-Nearest Neighbors (K-NN) (Altman, 1992) and Support Vector Machine (SVM) (Herath & Mittal, 2022). The choice of the classifiers is because of their widespread use in the literature for similar tasks.

After converting the reviews into vectors, the next step is to identify the topics from the reviews. A topic can be defined as a matter that is described in a review and can be anything like the nature of the food, staff, service time etc. A clustering algorithm in tandem with the elbow method is used to determine the optimum number of topics. The clustering algorithm (Malhotra et al., 2022) will group the reviews into clusters/topics based on the underlying terms, and the elbow method finds the optimum number of clusters/topics.

3.4. Clustering and elbow method

Clustering is an unsupervised machine learning approach in which similar objects are grouped together to form a cluster such that intra-cluster similarity is less and inter-cluster similarity is more. One notable

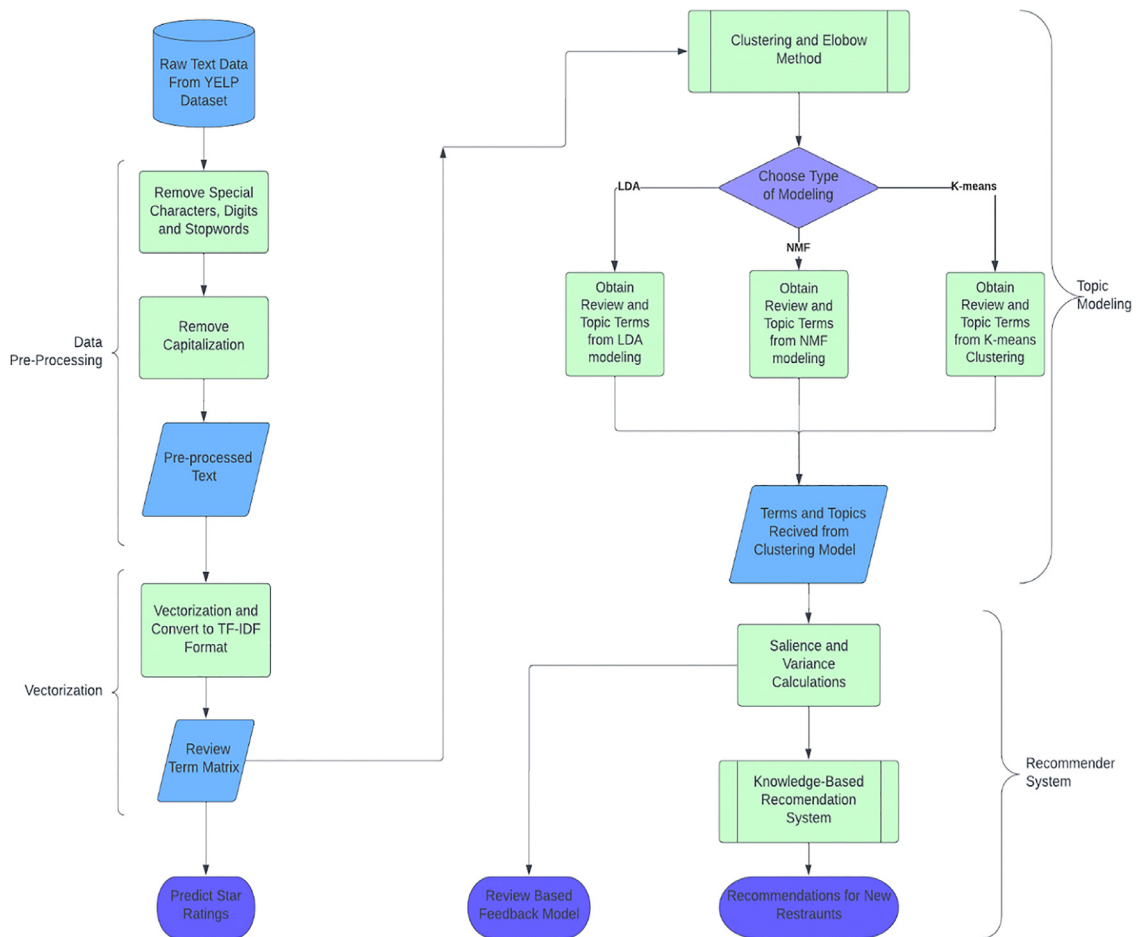


Fig. 1. Architecture of proposed system.

Table 1
Review before and after pre-processing.

Before Pre-processing	After Pre-processing
Husband was craving Chicken Teriyaki & gyoza, so we found Musashi. I was very unimpressed. We started with gyoza and edamame. Neither were anything special. We then ordered a chicken teriyaki plate and a few sushi rolls. The chicken teriyaki was nothing more than some boiled chicken smothered in teriyaki sauce. Was not good at all. The sushi was mediocre at best. While they were friendly, and the service was pretty good - I will not be back	husband crave chicken teriyaki gyoza found musashi unimpress start gyoza edamam neither anyth special order chicken teriyaki plate sushi roll chicken teriyaki noth boil chicken smother teriyaki sauc good sushi mediocr best friendli servic pretti good back

approach is K-means (Mishra et al., 2022), a vector quantised method that groups n objects into K clusters. Every object I belongs to at least one cluster, and each cluster will contain at least one object.

The K-means algorithm (Jain et al., 2021) is run by varying the number of clusters (K value), and the Within-Cluster Sum of Squares (WCSS) (Chen and Huang, 2019) value is noted. The point where there is a drop in the sum of squared errors value is emphasised as the elbow angle, and the corresponding K value gives the optimum number of topics. It can be noted in Fig. 2 that the optimum number of topics is 3 for the given dataset.

3.5. Topic modelling

Topic modelling algorithms aim to identify a set of topics and the underlying terms of the topics. The optimum value of K obtained from the clustering step, and the review-term matrix obtained from the vectorisation step are input to the Non-Negative Matrix Factorisation (NMF) algorithm for further analysis. The NMF factorises the input matrix into the review-topics matrix (W) and the topics-term matrix (H). Apart from NMF, we also analyse the topics generated using K-means and Latent Dirichlet Allocation (LDA) (Blei et al., 2003). LDA fixes a set of topics

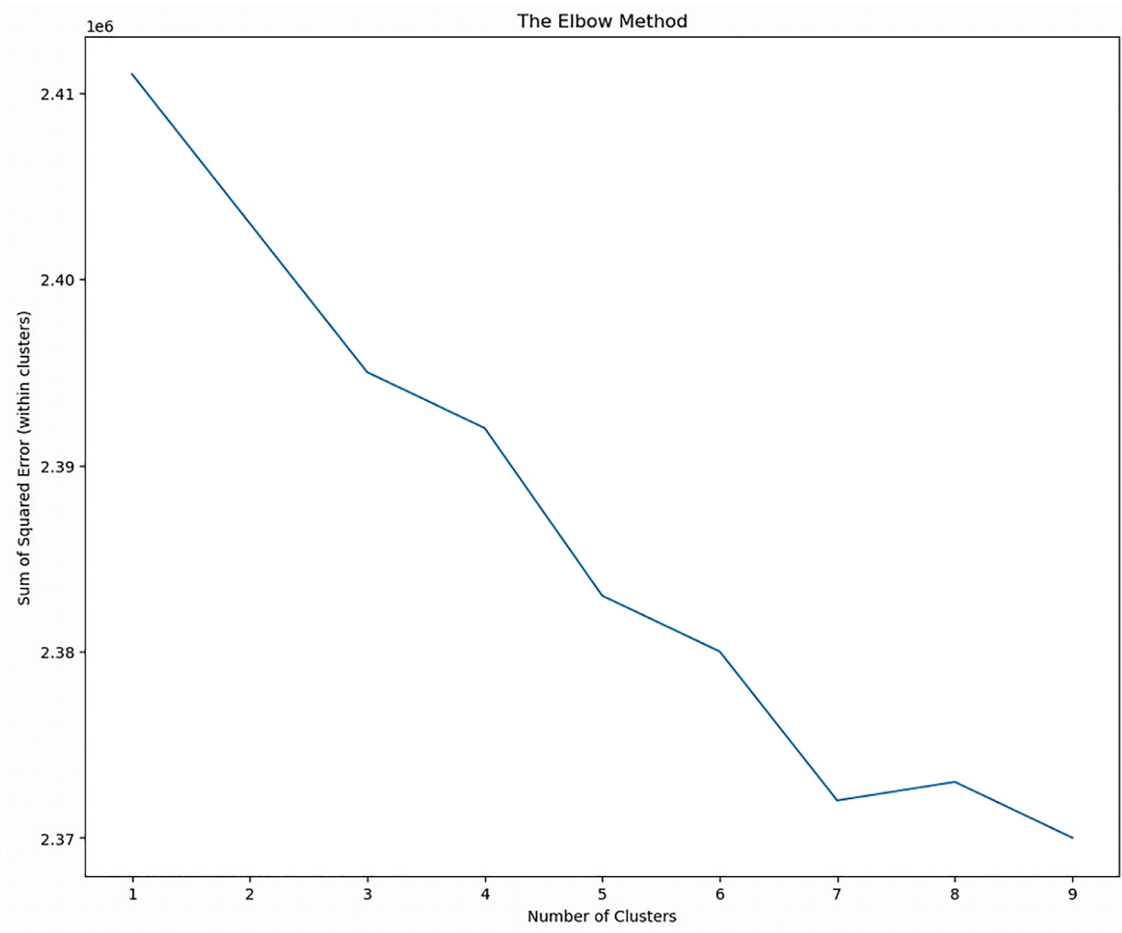


Fig. 2. Graphical representation of the elbow method.

where each topic contains a collection of words. It maps every document to a suitable topic such that the contents of the document and the collection of words in the topic are almost similar.

3.6. Salience and valence calculation

The salience and valence calculation for a topic has been provided by (Chang et al., 2022), (Neidhardt et al., 2017) and is used in this research. The salience of a topic is the ratio of the number of reviews on the topic to the total number of reviews and is given by Eq. (4). The greater the number of reviews in a topic, greater is the salience of that topic and it is considered an important topic for the tourists. Consequently, the topic can impact the likelihood of customers visiting the restaurant.

$$\text{Salience of a topic} = \frac{\text{number of reviews in a topic}}{\text{Total number of reviews}} \quad (4)$$

The valence of a topic is the ratio of the difference between the observed number of positive reviews (ONPR) and the expected number of positive reviews (ENPR) to the total number of reviews on the topic (TNRT) and is given by Eq. (5). It shows 'How positive or negative the topic is?'

$$\text{Valence of a topic} = \frac{(\text{ONPR} - \text{ENPR})}{\text{TNRT}} \quad (5)$$

A review is considered positive if the rating associated with the review is greater than or equal to 4 and vice versa. In Eq. (5), the total number of positive reviews on a topic is the observed number of positive reviews. To obtain the expected number of positive reviews, we first compute the percentage of positive reviews in the entire dataset. The product of the percentage and the number of reviews on the topic is the expected number of positive reviews. The valence of a topic can

be greater than 0 if the observed number of positive reviews is greater than the expected number of positive reviews on a topic. The higher the valence of a topic, the more positive reviews it has gained.

The salience and valence values are plotted on a mixed graph which serves as the feedback model.

3.7. Knowledge-based recommender system

A knowledge-based recommender system generates product recommendations based on the user's knowledge about the user and the products (Silveira et al., 2019), ("Full article: Knowledge-based recommendation system using semantic web rules based on Learning styles for MOOCs." 2022). In this research work, a knowledge-based recommender system is developed that can be beneficial to both tourists and managers. The input to the recommender system algorithm is the knowledge about the restaurants provided by the Yelp dataset and the valence of topics, service time, staff, and food for each restaurant. The topics, 'Service Time' (ST), 'Staff' and 'Food' are extracted by the NMF algorithm. The output of the recommender system is the restaurant recommendations list.

First, the user is asked a set of questions such as 'What type of cuisine are you looking for?' or 'In which city would you like to find a restaurant?'. This is used to obtain the user preferences. With user preferences, the recommender system identifies the relevant topic. The recommender system takes the topics as input and outputs a recommendation list of restaurants sorted according to the valence of the topic for each restaurant. For example, if the user chooses 'Food' as the topic which is important for decision making then, the user will obtain a list of restaurants

Table 2
Topics and underlying words generated by K-means clustering, LDA and NMF for 3 optimum topics

K-means			LDA			NMF		
Service Time (ST)	Staff	Food	Service Time (ST)	Staff	Food	Service Time (ST)	Staff	Food
Place	order	Great	great	order	ramen	drink	good	tasty
Good	food	Food	Love	pizza	service	get	staff	place
Love	time	service	Food	time	dog	ask	amaze	sauce
Food	good	Good	Good	wait	like	food	friendly	burger
chicken	us	Place	Place	table	definitely	minute	always	really
Best	like	friendly	delicious	go	really	table	love	tried
Tried	get	staff	Best	come	best	time	place	fried
always	place	amazing	amazing	like	tried	wait	service	like
Sushi	wait	atmosphere	Friend	drink	always	order	food	chicken
Taco	go	price	Chicken	take	good	us	great	good

Table 3
Topics and underlying words generated by NMF for 5 topics

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
Drink	Good	Sauce	Thin	Favorite
Get	Staff	Flavor	Pepperoni	Tuna
Ask	Amaze	Taco	Order	Ayc
Food	Friendly	Burger	Best	Eat
Minute	Always	Really	Slice	Best
Table	Loved	Tried	Salad	Fish
Wait	Place	Fried	Cheese	Fresh
Order	Service	Like	Wing	Place
Us	Food	Chicken	Crust	Roll
Time	Great	Good	Pizza	Sushi

where the first restaurant recommended has the most positive valence value for food topic.

4. Experimental approach and result

The experimental investigation was conducted on a powerful server with an Intel® Xeon® Silver 4210R CPU operating at 2.40 GHz and 2.39 GHz and 32 GB of installed RAM. Anaconda and Visual studio were part of the Python code editor for development since these tools provide an integrated development environment for managing the many dependencies and python packages. A handful of the well-known Python packages, such as pandas and NLTK, were crucial in processing many Yelp reviews. Pandas libraries were utilised in cleaning, processing and handling data more effectively.

4.1. Topic modelling

The optimum number of topics value obtained from the clustering and the elbow method is 3. Hence, in this research, we generate three topics from the restaurant reviews using different topic modelling techniques like K-means clustering, Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorisation (NMF). It is evident from Table 2 that the topics provided by NMF are clearer and more distinct. Hence, the topics generated by NMF are named as 'Service Time' (ST), 'Staff' and 'Food'.

4.2. Results of NMF algorithm for five optimum topics

The NMF algorithm generates 5 topics for further experimentation, and the results are tabulated in Table 3. While observing the results in Table 3, it can be noted that the NMF algorithm degenerated the 'Food' topic and created additional topics which cannot be grouped under a specific topic name. Thus, for further analysis, we decided to use the three topics generated by the NMF algorithm. On the other hand, the topics generated by K-means and LDA show minimal correlation to NMF.

4.3. Feedback models

This research explores the application of the feedback model for different use cases. This research generates and analyses feedback models for the following five cases:

- All restaurants in the USA
- Restaurants that serve a particular cuisine (Japanese cuisine)
- Japanese restaurants in a particular state (North Carolina state)
- A single Japanese restaurant (Musahi restaurant) in a particular state (North Carolina state)
- Compare between two restaurants in a chain (Pizza Hut chain) in two different states (Pennsylvania and Arizona)

Table 4 shows the cross-tabulation analysis for the four feedback models for the three topics. These values have been graphed out in Figs. 3, 4. The different parameters recorded for each model are total reviews, observed positive reviews, expected positive reviews, salience, valence, the total number of restaurant reviews, the total number of positive reviews and the total number of negative reviews. with the service time. Lastly, 'Staff' is the least salient topic (13.03%) but with good positive valence (13.44%). Musahi Japanese restaurant in North Carolina: On analysing a specific Japanese restaurant ('Musahi') in North Carolina state, we observe that most reviews about the restaurant are about its 'Service Time' (50%). The 'Service Time' at Musahi Restaurant has a negative valence of -11.23%. This implies that though the 'Service Time' is the most reviewed topic, it does not have many positive reviews. Moreover, the 'Staff' is the second salient topic (39.32%) and has a positive valence (11.55%). This indicates that consumers are happy with the 'Staff' and the 'Staff' topic is more likely mentioned in positive reviews than negative ones. 'Food' is the least salient topic (10.67%) but has a positive valence of 10.05%.

4.4. Analysis over the last feedback model

Comparing Pizza Hut across two states, Arizona and Pennsylvania: The feedback model can be used to compare the performance of a chain of restaurants across different states. The cross-tabulation analysis performed for Pizza Hut in Arizona and Pennsylvania is shown in Table 5. The expected and observed positive reviews of Pizza Hut restaurants across Arizona and Pennsylvania is shown in Fig. 5 and the salience and valence of the same is shown in Fig. 6. As observed from Fig. 6 there are noticeable differences in the salience and valence 378 of topics for the same restaurant in different states. We see that 'Service Time' is the most salient topic (62.52%) for Pizza Hut restaurants in Arizona. 'Food' is the second most salient topic (30.65%) followed by 'Staff' (6.82%). The 'Service Time' topic has a negative valence of -10.86% while 'Staff' has the most positive valence (57.89%). 'Food' has a lower positive valence for Pizza Hut restaurants in Arizona (9.25%) in comparison for Pizza Hut restaurants in Pennsylvania (17.96%). Similar observations can be drawn for Pizza Hut restaurants in Pennsylvania. About 70.95%

Table 4
Recorded values for different parameters of feedback models

Hypothesis	Topics	Total	Observed	Expected	Salience	Valence	Total Restaurants	Total Positive Reviews	Total Negative Reviews
Restaurant Reviews in USA	ST	578363	197818	386402	23.18	-32.58	2494488	1666082	828406
	Staff	803801	710494	536975	32.22	21.6			
	Food	1112324	757770	742705	44.59	1.33			
Japanese Restaurant Reviews in USA	ST	82757	44932	58673	46.33	-16.6	178602	1226626	51976
	Staff	20457	15527	14503	11.45	5			
	Food	75388	66167	53448	42.21	16.87			
Japanese Restaurant Reviews in North Carolina	ST	4127	1937	2719	2719	-18.94	9875	6506	3369
	Staff	1287	1021	847	847	13.44			
	Food	4461	3548	2939	2939	13.65			
Musahi Japanese Restaurant Reviews in North Carolina	ST	89	56	66	50	-11.23	178	132	46
	Staff	70	60	52	39.32	11.55			
	Food	19	16	14	10.67	10.05			

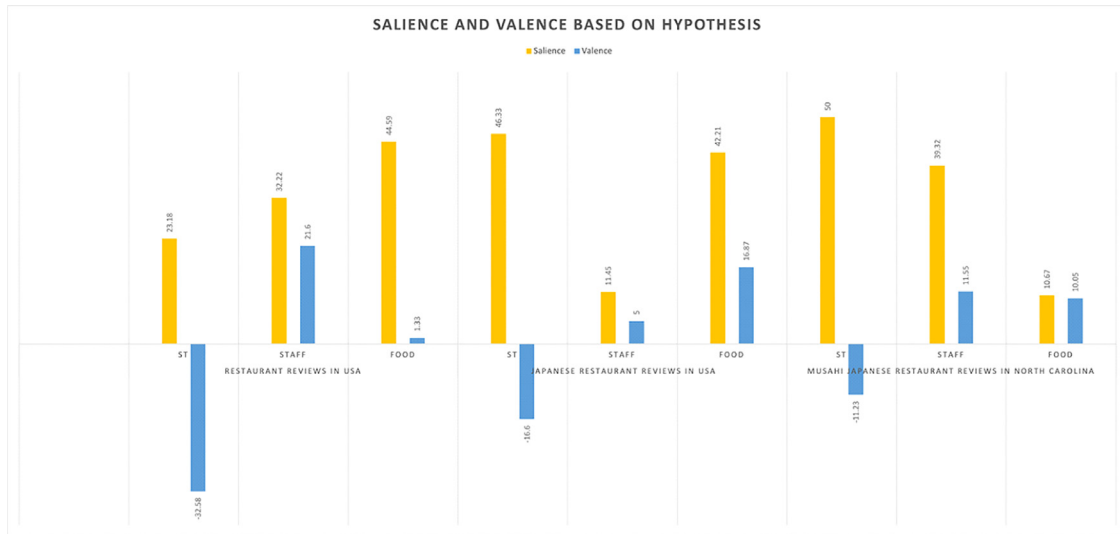


Fig. 3. Expected and observed positive reviews of all the four feedback models.

Table 5
Recorded values for Pizza Hut across Arizona and Pennsylvania

Hypothesis	Topic	Total Reviews	Observed Positive Reviews	Expected Positive Reviews	Salience	Valence	Total Restaurant reviews	Total Positive Reviews
Pizza hut reviews in Arizona	ST	724	107	185	62.52	-10.86	1158	297
	Staff	79	66	20	6.82	57.89		
	Food	355	124	91	30.65	9.28		
Pizza hut reviews in Pennsylvania	ST	149	17	27.67	70.95	-7.16	210	39
	Staff	9	3	1.67	4.28	14.76		
	Food	52	19	9.65	24.76	17.96		

of the reviews are about 'Service Time' topic and the topic has negative valence of -7.16%. The 'Food' is the second most salient topic (24.76%) followed by 'Staff' (4.28%). Thus, this research shows how the scope of the feedback model can be extended (i.e.) the model can be made as specific as to a restaurant in a city or generalised to restaurants in a country. In fact, the model can be made specific to cuisine-oriented restaurants in a country. Moreover, the feedback model can also be used to analyse restaurant chains.

4.5. Predicting star ratings from reviews

The performances of different classifiers for predicting the star ratings from the reviews are analysed by splitting the dataset into a test and a train set. The train-test split ratio is 80:20. The accuracy of the various classifiers (Random Forest, Decision Tree, Logistic Regression, Regression, K-Nearest Neighbours and Support Vector Machines) are given in Table 6. The accuracy of a classifier is the ratio of the num-

Table 6
Performance of different classifiers

Model	Accuracy
Random Forest	87.59%
Decision Tree	84.93%
Logistic Regression	74.19%
K-Nearest Neighbours	67.03%
Support Vector Machines	43.66%

ber of correct predictions to the total number of predictions made. It can be seen from Table 6 that Random Forest and Decision Tree classifiers give the same and the highest accuracy (99.98%) when predicting star ratings from the reviews, followed by Logistic Regression (74.19%), K-Nearest Neighbours (67.03%) and Support Vector Machines (43.66%) classifiers.

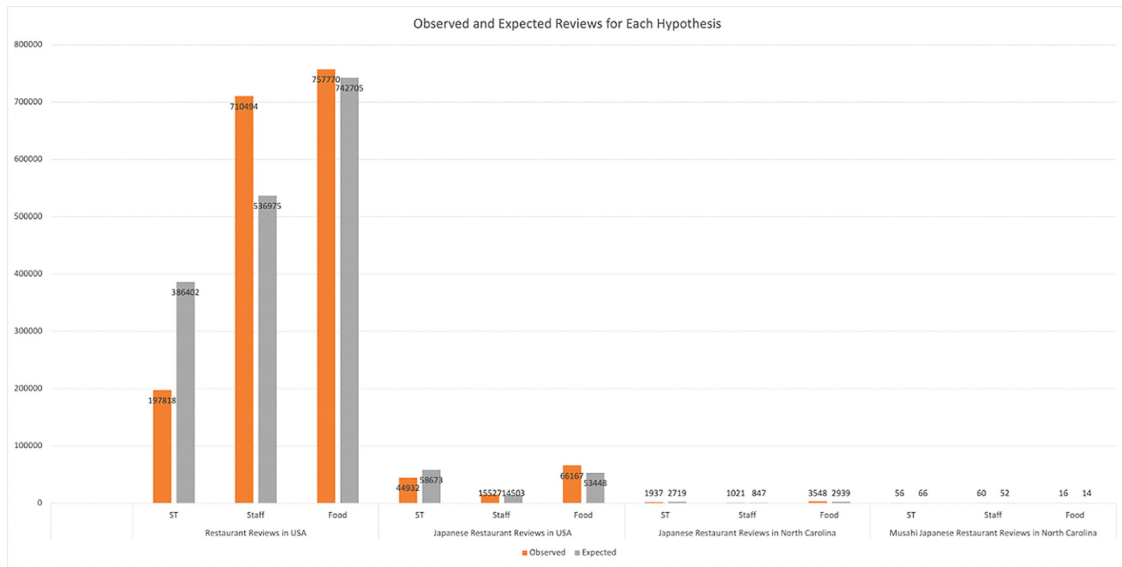


Fig. 4. The salience and valence of all four feedback models.

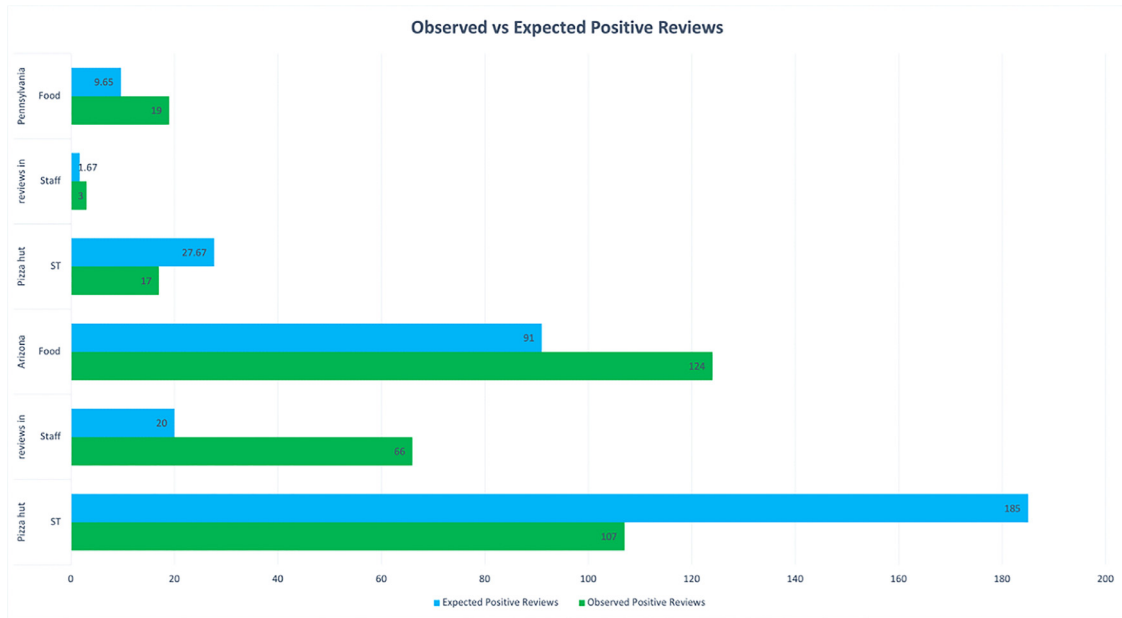


Fig. 5. Expected and observed positive reviews of Pizza Hut across Arizona and Pennsylvania.

4.6. Knowledge based recommendation result

The restaurant needs to use a set of relevant features to create input data for the recommender system. The user reviews are essential features of building knowledge-based recommendation systems. The collection of concatenated reviews has been converted into a matrix of token occurrences where the columns represent the features of a given restaurant, and the rows represent the individual restaurants. This is achieved using a hash vector with parameters set to 5 features per restaurant using English stop words. After this, data use normalisation using the term frequency helps to improve the features.

The normalise features pass into a recommender system. The goal is to use these input features to evaluate a restaurant based on feature vectors. As the matrix is already prepared in the preprocessing data step, the group of restaurant features are used to create the model. Similar locations have been defined based on the 5 features in the matrix. This is done using the K-Nearest Neighbors(KNN) algorithm with cosine dis-

Table 7

Similar recommended hotel

S.No.	Similar recommended restaurants
1	U-Pack Moving
2	Starbucks
3	Fuji San
4	Crazy Sushi
5	King Tut Restaurant and Hookah Bar
6	Larmon Photo
7	Empire Today
8	Kendra's Floral Kreations
9	The Franklin Residences
10	RE/MAX Premier Properties

tance on the CSR matrix. For inference, Dimitri is used as a restaurant liked by our consumers. The model then predicted 10 restaurants shown in Table 7 as ones that this same customer would like.

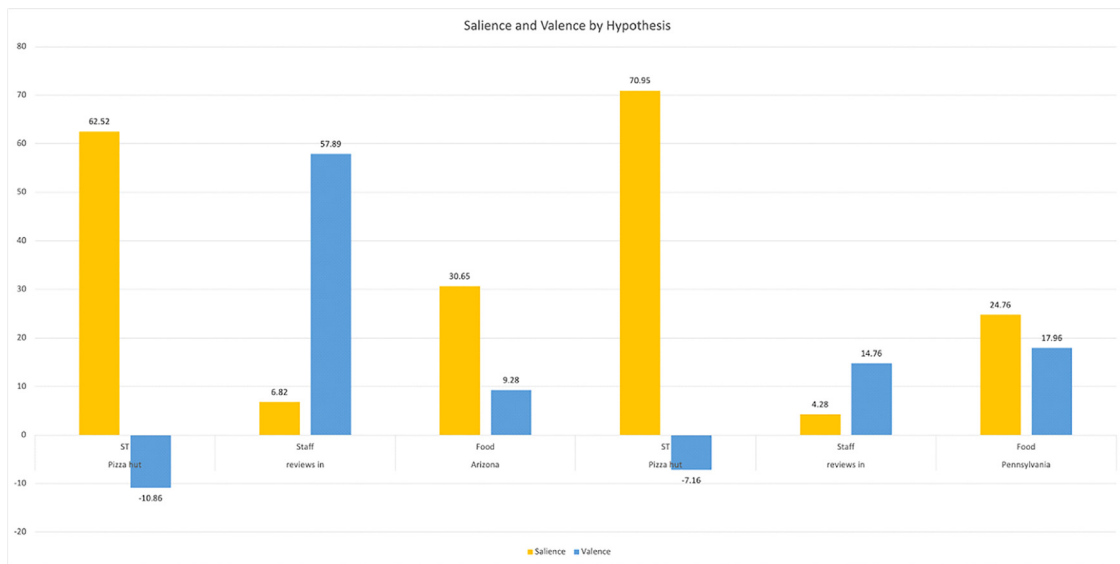


Fig. 6. The saliency and valence of Pizza Hut across Arizona and Pennsylvania.

5. Discussions

5.1. Review of literature

Many researchers in different domains carry out the study on recommendation systems. The primary techniques used by authors are content-based (Pazzani & Billsus, 2007), (Gulati & Bhagat, 2021), collaborative filtering (Mustafa et al., 2017), (Dubey et al., 2018), and hybrid approach (Li et al., 2020). The existing approaches have limited approaches to capturing and extracting the features to the full extent. While these approaches have made larger advancements, they do not fully utilise these features to provide in depth models or analysis of such features. The proposed work uses the feedback model to summarise the reviews for tourism authorities by providing the topics, their importance, and the sentiment of those topics.

5.2. Contributions to literature

This paper explores a new area by creating a recommender system drawn not from star ratings or handcrafted features but rather from automatically generated features that are drawn from the text-based reviews of the restaurants. Here we use a variety of unsupervised methods (Kmeans, LDA and NMF) to create features based on the parts of the reviews that showed patterns across the restaurants. These features proved highly useful as they could then be passed into either a supervised learning model (Section 4.5) to predict star ratings or used in an unsupervised knowledge-based recommender system to filter and recommend restaurants based on their reviews. This personalized recommendation will accelerate the decision-making process for tourists. Though the recommender system developed in this research has the feature of single topic selection, it can be easily extended to multiple topic selection. The feedback model summarises the reviews for tourism authorities by providing the topics, their importance, and the sentiment of the topics. The tourism authorities can take the necessary action to improve the F&B sector and consequently boost tourism. By calculating the topic's saliency, it can be understood which restaurant topic is the most addressed in the online reviews. This provides insight not only into areas that covered by the location but also reveals which areas may be lacking and in need of attention which will allow for a shift in focus towards areas which may be lacking. This will in turn help tourism authorities to take the necessary action to improve the F&B sector and consequently boost tourism.

5.3. Implications for practice

The restaurants and locations that a person likes can often be subjective and biased due to personal experience and preferences. Because of this, simple star ratings (Reddy et al., 2017) cannot always help recommender systems select restaurants, even with added features such as cuisine and location. This new knowledge-based recommender system is built using features learned from text reviews, which allow the model to find a restaurant with reviews similar to ones the user has previously enjoyed. This type of model allows for all types of preferences and biases since the model draws from a corpus with those same preferences and biases. Moreover, the valence of the same topic can talk about the restaurant's performance on the topic. The developed system in research is a knowledge-based recommender system that tourists can use to obtain recommendations based on his/her preferences and in-depth knowledge about the restaurants. Moreover, it allows personal recommendation by selecting the topic based on the tourist's interest. After tabulation of results, inferences regarding the restaurant and its management can be drawn. For instance, restaurant management would aim to have 40% of staff reviews as positive. This could be because the management does not consider staff as an essential topic that would help them achieve their long-term goals. In such a case, the expected number of positive reviews for the 'Staff' topic could be set to 40% of the number of reviews in the 'Staff' topic; hence, the valence will change accordingly. Moreover, the valence of the same topic can tell us about the restaurant's performance in the topic Hypothesis - Variability of the expected number of positive reviews.

The recommender system developed in this work is a knowledge-based recommender system which tourists can use to obtain recommendations based on his/her preferences and knowledge about the restaurants. Moreover, it allows personal recommendation by selecting the topic suitable to the tourist. This can make the decision-making process faster for tourists. Though the recommender system developed in this research has the feature of single topic selection, it can be easily extended to multiple topic selection. The recommender system can also help a restaurant understand its position in the recommendations list. The restaurant can learn about its ranking in the recommendations list when a user filters restaurants according to his or her preferences. One can also consider how the saliency, valence and even the final model might be affected after the industry has changed. The model described in this paper shows a flexible approach that allows for change in the industry, after which the information gained and insights found will change.

6. Conclusion

This research develops a recommender system designed to work based on text analysis and star ratings from user reviews while highlighting the relation between machine learning and tourism. The proposed machine learning-based system can achieve three subgoals: predicting star ratings from reviews, a feedback model, and a knowledge-based recommender system. The automated system employed several machine learning classifiers to predict star ratings from reviews. The highest accuracy achieved by the random forest classifier is 99.98%. In addition, there was an exploration of text-based reviews using machine learning, linking these reviews to customer preferences, and grouping restaurants based on these reviews. The research also addressed the flexibility of the feedback model. The final part of this process was building a knowledge-based recommender system to generate a restaurant recommendation list based on the reviews of previous restaurants they enjoyed. The current research has considered only the restaurants in the F&B sector, and similar feedback models can be generated for other tourism sectors by simply expanding the CSR matrices and feature columns to improve the recommendation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Adomavicius, G., Sankaranarayanan, R., Sen, S., & Tuzhilin, A. (2005). Incorporating contextual information in recommender systems using a multidimensional approach. *ACM Transactions on Information Systems (TOIS)*, 23(1), 103–145 Jan.. [10.1145/1055709.1055714](https://doi.org/10.1145/1055709.1055714).
- Aggarwal, C. C. (2016). An introduction to recommender systems. *Recommender Systems*, 1–28. [10.1007/978-3-319-29659-3_1](https://doi.org/10.1007/978-3-319-29659-3_1).
- al Amrani, Y., Lazaar, M., & el Kadiri, K. E. (2018). Random forest and support vector machine based hybrid approach to sentiment analysis. *Procedia Computer Science*, 127, 511–520. [10.1016/j.procs.2018.01.150](https://doi.org/10.1016/j.procs.2018.01.150).
- Alnogaithan, O., Algazlan, S., Aljuraiban, A., & Shargabi, A. A. (2019). Tourism recommendation system based on user reviews. In *Proceedings of the international conference on innovation and intelligence for informatics, computing, and technologies 3ICT 2019*, Sep.. [10.1109/3ICT.2019.8910312](https://doi.org/10.1109/3ICT.2019.8910312).
- Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *American Statistician*, 46(3), 175–185. [10.1080/00031305.1992.10475879](https://doi.org/10.1080/00031305.1992.10475879).
- Amzad, H., Vijayalakshmi, K., “Tourism recommendation system: A systematic review cite this paper tourism recommendation system: A systematic review”, Accessed: Aug. 29, 2022. [Online]. Available: www.ijert.org
- Asrani, C., Kar, A.K., “Diffusion and adoption of digital communications services in India,” 2022, [10.1080/02681102.2022.2046536](https://doi.org/10.1080/02681102.2022.2046536).
- Bawack, R. E., Wamba, S. F., Carrillo, K. D. A., & Akter, S. (2022). Artificial intelligence in E-Commerce: A bibliometric study and literature review. *Electronic Markets Mar..* [10.1007/S12525-022-00537-Z](https://doi.org/10.1007/S12525-022-00537-Z).
- Blei, D. M., Ng, A. Y., & Edu, J. B. (2003). Latent dirichlet allocation michael I. Jordan. *Journal of Machine Learning Research*, 3, 993–1022.
- Chang, C.-Y., et al., “The corpus of emotional valences for 33,669 Chinese words based on big data,” pp. 141–152, 2022, [10.1007/978-3-031-05544-7_11](https://doi.org/10.1007/978-3-031-05544-7_11).
- Chatterjee, S., Kar, A.K., Dwivedi, Y.K., “Intention to use IoT by aged indian consumers,” vol. 62, no. 4, pp. 655–666, 2021, [10.1080/08874417.2021.1873080](https://doi.org/10.1080/08874417.2021.1873080).
- Chen, F., & Huang, Y. (2019). Knowledge-enhanced neural networks for sentiment analysis of Chinese reviews. *Neurocomputing*, 368, 51–58 Nov.. [10.1016/j.neucom.2019.08.054](https://doi.org/10.1016/j.neucom.2019.08.054).
- Chica, M., Hernández, J. M., & Perc, M. (2022). Sustainability in tourism determined by an asymmetric game with mobility. *Journal Cleaner Production*, 355, Article 131662 Jun.. [10.1016/j.jclepro.2022.131662](https://doi.org/10.1016/j.jclepro.2022.131662).
- Dubey, A., Gupta, A., Raturi, N., Saxena, P., “Item-based collaborative filtering using sentiment analysis of user reviews,” 2018, pp. 77–87.
- Ezin, E., Palomares, I., & Neve, J. (2019). Group decision making with collaborative-filtering ‘in the loop’: Interaction-based preference and trust elicitation. In *Proceedings of the IEEE international conference on systems, man and cybernetics: 2019-October* (pp. 4044–4049). Oct.. [10.1109/SMC.2019.8914224](https://doi.org/10.1109/SMC.2019.8914224).
- “Ecommerce 101: The History and Future of Online Shopping.” <https://www.bigcommerce.com/articles/ecommerce/#growth-of-ecommerce> (accessed Jun. 03, 2022).
- “Electronics, Cars, Fashion, Collectibles & More | eBay.” <https://www.ebay.com/> (accessed Aug. 29, 2022).
- “Full article: Knowledge-based recommendation system using semantic web rules based on Learning styles for MOOCs.” <https://www.tandfonline.com/doi/full/10.1080/23311916.2021.2022568> (accessed Jun. 03, 2022).
- Gandhi, M., & Kar, A. K. (2022). How do Fortune firms build a social presence on social media platforms? Insights from multi-modal analytics. *Technological Forecasting and Social Change*, 182, Article 121829 Sep.. [10.1016/j.techfore.2022.121829](https://doi.org/10.1016/j.techfore.2022.121829).
- Goel, K., Sindhgatta, R., ILAVARASAN, V., & Kar, A. (2021). Building a knowledge society: Learnings from a digital literacy program. *ECIS 2021 Research Papers*. Jun. Accessed: Aug. 29, 2022. [Online]. Available: https://aisel.aisnet.org/ecis2021_rp/121.
- Gretzel, U., & Yoo, K. H. (2008). Use and impact of online travel reviews. *Faculty of Commerce - Papers (Archive)*, 35 Jan.. [10.1007/978-3-211-77280-5_4](https://doi.org/10.1007/978-3-211-77280-5_4).
- Grover, P., Kar, A. K., & Dwivedi, Y. (2022). The evolution of social media influence - A literature review and research agenda. *International Journal of Information Management Data Insights*, 2(2), Article 100116 Nov.. [10.1016/j.ijime.2022.100116](https://doi.org/10.1016/j.ijime.2022.100116).
- Gulati, R., & Bhagat, M. (2021). Content and knowledge based Food Recommending Intelligence Enabled System (FRIES). In *Proceedings of the 5th international conference on computing methodologies and communication, ICCMC* (pp. 1743–1749). Apr.. [10.1109/ICCMC51019.2021.9418030](https://doi.org/10.1109/ICCMC51019.2021.9418030).
- Gupta, S., Tuunanen, T., Kar, A. K., & Modgil, S. (2022). Managing digital knowledge for ensuring business efficiency and continuity. *Journal of Knowledge Management, ahead-of-print*(ahead-of-print). [10.1108/JKM-09-2021-0703/FULL/XML](https://doi.org/10.1108/JKM-09-2021-0703/FULL/XML).
- Herath, H. M. K. M. B., & Mittal, M. (2022). Adoption of artificial intelligence in smart cities: A comprehensive review. *International Journal of Information Management Data Insights*, 2(1), Article 100076 Apr.. [10.1016/j.ijime.2022.100076](https://doi.org/10.1016/j.ijime.2022.100076).
- Jain, R., Batra, J., Kar, A. K., Agrawal, H., & Tikkiwal, V. A. (2021). A hybrid bio-inspired computing approach for buzz detection in social media. *Evolutionary Intelligence* 2021 15:1, 15(1), 349–367 Apr.. [10.1007/S12065-020-00512-7](https://doi.org/10.1007/S12065-020-00512-7).
- Kumar, P., Gupta, G. P., Tripathi, R., Garg, S., & Hassan, M. M. (2021). DLTF: Deep learning-driven cyber threat intelligence modeling and identification framework in IoT-enabled maritime transportation systems. *IEEE Transactions on Intelligent Transportation Systems*, 1–10 Nov.. [10.1109/ITITS.2021.3122368](https://doi.org/10.1109/ITITS.2021.3122368).
- Kumar, P., Hollebeek, L. D., Kar, A. K., & Kukuk, J. (2022). Charting the intellectual structure of customer experience research. *Marketing Intelligence and Planning, ahead-of-print*(ahead-of-print). [10.1108/MIP-05-2022-0185/FULL/XML](https://doi.org/10.1108/MIP-05-2022-0185/FULL/XML).
- Kumar, S., Kar, A.K., Ilavarasan, P.V., “Using big data analytics on social media to analyze tourism service encounters,” pp. 121–130, 2022, doi: 10.1007/978-981-16-8542-2_10.
- Kushwaha, A. K., Kar, A. K., & Ilavarasan, P. V. (2021). Predicting retweet class using deep learning. *Trends in Deep Learning Methodologies: Algorithms, Applications, and Systems*, 89–112 Jan.. [10.1016/B978-0-12-822226-3.00004-0](https://doi.org/10.1016/B978-0-12-822226-3.00004-0).
- Kushwaha, A. K., Kar, A. K., Roy, S. K., & Ilavarasan, P. V. (2022). Capricious opinions: A study of polarization of social media groups. *Government Information Quarterly*, 39(3), Article 101709 Jul.. [10.1016/j.giq.2022.101709](https://doi.org/10.1016/j.giq.2022.101709).
- Kushwaha, A. K., Pharswan, R., Kumar, P., & Kar, A. K. (2022). How do users feel when they use artificial intelligence for decision making? A framework for assessing users’ perception. *Information Systems Frontiers*, 1–20 May. [10.1007/S10796-022-10293-2/TABLES/3](https://doi.org/10.1007/S10796-022-10293-2/TABLES/3).
- Kyaw, N. E. E., & Wai, T. T. (2019). Inferring user preferences using reviews for rating prediction. In *Proceedings of the international conference on advanced information technologies, ICAIT* (pp. 194–199). Nov.. [10.1109/AITC.2019.8921179](https://doi.org/10.1109/AITC.2019.8921179).
- Li, M., Li, Y., Lou, W., & Chen, L. (2020). A hybrid recommendation system for Q&A documents. *Expert Systems with Applications*, 144, Article 113088 Apr.. [10.1016/j.eswa.2019.113088](https://doi.org/10.1016/j.eswa.2019.113088).
- Malhotra, S., Chaudhary, K., & Alam, M. (2022). Modeling the use of voice based assistant devices (VBADs): A machine learning base an exploratory study using cluster analysis and correspondence analysis. *International Journal of Information Management Data Insights*, 2(1), Article 100069 Apr.. [10.1016/j.ijime.2022.100069](https://doi.org/10.1016/j.ijime.2022.100069).
- Masripah, S., et al., (2020). Comparison of naïve bayes algorithm with genetic algorithm and particle swarm optimization as feature selection for sentiment analysis review of digital learning application. *Journal of Physics: Conference Series*, 1641(1), Article 012040 Nov.. [10.1088/1742-6596/1641/1/012040](https://doi.org/10.1088/1742-6596/1641/1/012040).
- Mishra, R. K., Raj, H., Urolagin, S., Angel, J., Jothi, A., & Nawaz, N. (2022). Cluster-based knowledge graph and entity-relation representation on tourism economical sentiments. *Applied Sciences*, 12(16), 8105 2022, Vol 12, Page 8105Aug.. [10.3390/AP12168105](https://doi.org/10.3390/AP12168105).
- Mishra, R. K., Urolagin, S., & Angel Arul Jothi, J. (2020). Sentiment analysis for POI recommender systems. In *Proceedings of the 7th international conference on information technology trends, ITT* (pp. 174–179). Nov.. [10.1109/ITT51279.2020.9320885](https://doi.org/10.1109/ITT51279.2020.9320885).
- Mishra, R. K., Urolagin, S., & Jothi, A. A. J. (2019). A Sentiment analysis-based hotel recommendation using TF-IDF Approach. In *Proceedings of the international conference on computational intelligence and knowledge economy, ICCIKE* (pp. 811–815). Dec.. [10.1109/ICCIKE47802.2019.9004385](https://doi.org/10.1109/ICCIKE47802.2019.9004385).
- Mustafa, N., Ibrahim, A. O., Ahmed, A., & Abdullah, A. (2017). Collaborative filtering: Techniques and applications. In *Proceedings of the International Conference on Communication, Control, Computing and Electronics Engineering, ICCCEE Feb.* [10.1109/ICCC-CEE.2017.7867668](https://doi.org/10.1109/ICCC-CEE.2017.7867668).
- Neidhardt, J., Rümmele, N., & Werthner, H. (2017). Predicting happiness: User interactions and sentiment analysis in an online travel forum. *Information Technology and Tourism*, 17(1), 101–119 Mar.. [10.1007/S40558-017-0079-2/TABLES/6](https://doi.org/10.1007/S40558-017-0079-2/TABLES/6).
- Neogi, A. S., Garg, K. A., Mishra, R. K., & Dwivedi, Y. K. (2021). Sentiment analysis and classification of Indian farmers’ protest using twitter data. *International Journal of Information Management Data Insights*, 1(2), Article 100019 Nov.. [10.1016/j.ijime.2021.100019](https://doi.org/10.1016/j.ijime.2021.100019).
- Nilashi, M., Ibrahim, O., Yadegaridehkordi, E., Samad, S., Akbari, E., & Alizadeh, A. (2018). Travelers decision making using online review in social net-

- work sites: A case on TripAdvisor. *Journal Computing Science*, 28, 168–179 Sep.. [10.1016/J.JOCS.2018.09.006](https://doi.org/10.1016/J.JOCS.2018.09.006).
- “Online Shopping Site for Mobiles, Electronics, Furniture, Grocery, Lifestyle, Books & More. Best Offers!” <https://www.flipkart.com/> (accessed Aug. 29, 2022).
- Pazzani, M. J., & Billsus, D. (2007). Content-based recommendation systems. In *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*: 4321 (pp. 325–341). LNCS. [10.1007/978-3-540-72079-9_10/COVER](https://doi.org/10.1007/978-3-540-72079-9_10/COVER).
- Qin, Y., Wang, X., & Xu, Z. (2022). Ranking tourist attractions through online reviews: A novel method with intuitionistic and hesitant fuzzy information based on sentiment analysis. *International Journal of Fuzzy Systems*, 24(2), 755–777 Mar.. [10.1007/S40815-021-01131-9/TABLES/18](https://doi.org/10.1007/S40815-021-01131-9/TABLES/18).
- Reddy, C. S. C., Kumar, K. U., Keshav, J. D., Prasad, B. R., & Agarwal, S. (2017). Prediction of star ratings from online reviews. In *Proceedings of the IEEE region 10 annual international conference, proceedings/TENCON: 2017-December* (pp. 1857–1861). Dec.. [10.1109/TENCON.2017.8228161](https://doi.org/10.1109/TENCON.2017.8228161).
- Sahu, T. P., & Ahuja, S. (2016). Sentiment analysis of movie reviews: A study on feature selection and classification algorithms. In *Proceedings of the international conference on microelectronics, computing and communication, MicroCom Jul.* [10.1109/MICROCOM.2016.7522583](https://doi.org/10.1109/MICROCOM.2016.7522583).
- Sharma, S., Kar, A.K., Gupta, M.P., Dwivedi, Y. K., Janssen, M., “Digital citizen empowerment: A systematic literature review of theories and development models,” <https://www.doi.org/10.1080/0268110220222046533>, 2022, [10.1080/02681102.2022.2046533](https://doi.org/10.1080/02681102.2022.2046533).
- Silveira, T., Zhang, M., Lin, X., Liu, Y., & Ma, S. (2019). How good your recommender system is? A survey on evaluations in recommendation. *International Journal of Machine Learning and Cybernetics*, 10(5), 813–831 May. [10.1007/S13042-017-0762-9](https://doi.org/10.1007/S13042-017-0762-9).
- “Statistics.” <https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx> (accessed Jun. 03, 2022).
- Sufi, F. K. (2022). Identifying the drivers of negative news with sentiment, entity and regression analysis. *International Journal of Information Management Data Insights*, 2(1), Article 100074 Apr.. [10.1016/J.JJIMEI.2022.100074](https://doi.org/10.1016/J.JJIMEI.2022.100074).
- Theses, G., Gunden, N., Berezina, E., Faizan Ali, C., “Scholar commons how online reviews influence consumer restaurant selection,” 2017, Accessed: Jun. 03, 2022. [Online]. Available: <http://scholarcommons.usf.edu/etd/6707>
- Tohmo, T. (2018). The economic impact of tourism in Central Finland: A regional input-output study. *Tourism Review*, 73(4), 521–547 Oct.. [10.1108/TR-04-2017-0080](https://doi.org/10.1108/TR-04-2017-0080).
- Tripathi, A., Goswami, T., Trivedi, S. K., & Sharma, R. D. (2021). A multi class random forest (MCRF) model for classification of small plant peptides. *International Journal of Information Management Data Insights*, 1(2), Article 100029 Nov.. [10.1016/J.JJIMEI.2021.100029](https://doi.org/10.1016/J.JJIMEI.2021.100029).
- “Yelp Dataset.” <https://www.yelp.com/dataset>, 2022.