Tourism recommendation system using spatial and demographic features

Uma Perumal [1]

[1] Department of Computer Science, College of Engineering and Computer Science, Jazan University, Jazan 45142, Saudi Arabia, Email: uperumal@jazanu.edu.sa , prmluma@gmail.com

Abstract

Tourism Recommendation System (TRS) becomes a significant guide for tourists as it analyzes the various factors of tourism. In order to make a foolproof recommendation, a variety of factors need to be taken into consideration, including environmental factors, exact geo coordinates, trip destination, preferences of tourists, and so on. Various Artificial Intelligence (AI) techniques have been developed nowadays. However, these techniques failed to concentrate on spatiotemporal characteristics, user privacy and data secrecy, traffic information, and so on while recommending tourism to users. Also, the existing techniques failed to consider the demographic factors for tourism recommendations, thus resulting in inappropriate recommendations. Hence, in this paper, a TRS is proposed based on the Non-central Chi-squared distribution-based Deep Learning Neural Network (NC-DLNN) classification technique by utilizing the Shapefile, Google External Application Programming Interface (API), and Geographic Information System (GIS) map details, which are stored in the Geo database. Moreover, the Direction-based Fire Hawks Optimization (D-FHO) and Alignment-based Bidirectional Encoder Representations from the Transformers (A-BERT) techniques are introduced for filtering and embedding the keywords in Google external API. Further, the average ratings in Google API are considered along with other factors for making decisions and the corresponding tourism recommendation by using the Fuzzy method and NC-DLNN model, respectively. The experimental outcomes depicted that the proposed method achieved an accuracy of 97.91%, precision of 97.9%, and specificity of 97.92%. Furthermore, the proposed embedding algorithm achieved a better Bleu Score value than the prevailing methods.

Keywords:

Rapid Automatic Keyword Extraction(RAKE), Geographic Information System(GIS), Direction-based Fire Hawks Optimization (D-FHO), Alignment-based Bidirectional Encoder Representations from the Transformers (A-BERT), Quintic Interpolation(QI), Non-central Chi-squared distribution-based Deep Learning Neural Network(NC-DLNN), and Application Programming Interface(API).

Highlights:

- This paper focuses on Tourism Recommendation Systems (TRS), which play a crucial role in assisting tourists by providing personalized recommendations.
- The TRS aims to recommend tourist destinations, accommodations, and other relevant items based on various factors like environmental conditions, geocoordinates, and user preferences.
- Non-Central Chi-Squared Distribution-based Deep Learning Neural Network (NC-DLNN): A classification technique used for accurate recommendations.
- Direction-based Fire Hawks Optimization (D-FHO): Enhancing recommendation quality. Filtering techniques.
- The proposed embedding algorithm outperformed existing methods in terms of Bleu Score, indicating better recommendation quality.

Submitted: 22-AUG-24

Accepted: 3-NOV-24

Published:

DOI: 10.5455/jeas.2024021108

1. Introduction

Among the various sectors, Tourism plays a crucial role in the prosperity, development, and well-being of the country. So, it is also considered a pivotal component of international economic activity [1]. Tourism Attractions (TA) and Related Infrastructures (RI) are the two main factors to be considered while visiting a tourist spot as per the Tourism-Spatial System Theory (TSST). The cultural heritage or physical attractiveness of the location (tourist spot) that the individual tourists wish to attain in their leisure time is defined by the TA. Climate, vegetation, culture, scenery, theater performance, museums, and so on are some of the factors considered under TA. RI is the other factor to be considered while visiting the tourist spot. Information regarding accommodation, entertainment, transportation, dining, and so on is provided by RI to tourists [2]. Hence, tourism is also considered as a tool in fighting poverty, with initiatives, such as the World Tourism Organisation´s Sustainable Tourism-Eliminating Poverty (ST-EP) programme [3].

Tourism plays a crucial role in alleviating poverty by means of the Backward Integration Strategy (BIS). It provides employment opportunities and aids in the development of the nation's infrastructure [4]. Foreign exchange earnings earned through trading commodities are one of the major significance of tourism [5]. Tourism is divided into two different categories, namely International Tourism (IT) and Domestic Tourism (DT) by the United Nations World Tourism Organization (UNWTO) [6]. International tourism involves the activities of both resident and non-resident tourists outside their country of residence, whereas domestic tourism constitutes the activities of resident tourists within their country. Domestic tourism emerged in the earlier civilization itself [7].

In recent days, AI has contributed to the enhanced tourism system in several ways. AI has the tendency to analyze the users' preferences, previous experience, and their activities. So, personalized recommendations are provided to the users by using AI, thus enabling a higher level of user satisfaction. Many travel agencies and tourism guidance institutions utilize AI-based chatbots, especially for immediate customer services, including travel bookings and tourist spot details. Furthermore, AI-based systems offer up-to-date suggestions related to the weather conditions and current scenarios around the tourism sites. By using the AI models, the customer reviews, ratings, as well as social media-based comments are deeply analyzed. So, the popularity and quality of the tourist sites are informed to the users to relish the high-end experience. These advantages and capabilities of the AI-based approaches produce better recommendations and elevate the overall tourism experience of the users.

The cultural richness and historical monuments in Saudi Arabia attract tourists from every part. Riyadh, which is the capital of Saudi Arabia with fifteen different municipalities within it, is one of the largest cities in the Arabian Peninsula. Riyadh is surrounded by numerous archeological sites of historical importance. Numerous people visit Riyadh during vacations to see the palaces and monuments and take photographs to memorize their cherished moments. Thus, the Saudi Government hosts the Riyadh Season with a variety of events to promote tourism. Sports, BangTanSonyeondan (BTS) concerts, Fashion shows, and World Wrestling Entertainment (WWE) events are encompassed in some of the events. In this way, the attractiveness of the palace and monuments, the effective exchange of information, and the communication procedures of the Riyadh provincial are revealed to the other nations [8].

To evaluate the structural and topological properties involved in the tourism flow determination process, neural networks are utilized for analyzing multiple destinations and their characteristics [9]. The conventional techniques focused only on static characteristics like degree centrality, closeness centrality, and structural attributes, such as network density and structural holes. But, it failed to concentrate on the spatial and dynamic relationships of multiple destinations. Thus, this paper proposes an efficient tourism Recommendation System (RS) using the NC-DLNN technique via a QI-based spatial interpolation and demographic data-based approach. All the spatial and dynamic information about the tourist spots is provided by the usage of spatial interpolation [10].

1.1. Problem Description

There exist certain downsides despite the various advantages provided by conventional methodologies.

- Existing research methodologies failed to focus on the demographic factors to analyze in and around the circumstances of tourism sites.
- In the existing tourism models, the usage of the unprocessed spatial data resulted in classification errors.

- Rating-based tourism recommendations produced inaccurate results due to the possibility of getting fake review ratings for historic sites and the inability to determine the polarity score.

1.2 Main Contributions

Thus, NC-DLNN-based tourism RS is proposed in this work in order to alleviate the aforementioned issues, and its main contributions are detailed further.

- The usage of the demographic data in the proposed framework aids in the selection of secured tourist areas.

- In the proposed approach, ratings and reviews are considered along with the effective consideration of the polarity score in the tourist spot selection phenomena.

- The utilization of the pre-processing stage after obtaining the data from the geodatabase provides error-free classification results.

1.3 Research novelty

Many articles related to the TRS were existent. In most of the existing works like [11],[14], and [18], the spatio-temporal characteristics along with the demographic features were not considered for analyzing the tourism spot scenarios. Further, the existing TRS solely depended on limited data either geographic location, customer reviews, or demographic data. It could not effectively cover the various criteria of the tourism spots and their quality to develop the proper recommendation to the visitors. Conversely, the proposed TRS utilizes three different data related to the tourism site, namely GIS Map, Google external API, and Shapefile of the Geo database.

So, the geographical location, spatial interpolation, customer ratings, and demographic information of the cultural tourist sites are analyzed in this work. Moreover, the inclusion of data preprocessing, keywords filtering with D-FHO, word embedding using A-BERT, and Fuzzy-based decision-making aided in producing the exact recommendation using the proposed NC-DLNN recommendation model. Hence, it is clarified that these innovative or different things are carried out in the proposed work over the other related existing TRS.

The remaining part is arranged as: The related works concerning the proposed method are surveyed in Section 2, the proposed tourism RS using the NC-DLNN is elucidated in Section 3, the results and discussion of the proposed method based on performance metrics are delineated in Section 4, and lastly, the paper is concluded in Section 5 with future enhancements.

2. Literature Review

Torres-Ruiz et al. [11] developed a Hybrid Augmented Reality (HAR) based RS to visit museum itineraries. To generate the recommendation for museums, the data obtained from various sources underwent an Interpolation and Combination (IC) technique, which in turn was visualized using HAR. Hence, the presented methodology provided better recommendations along with visitor and route information through the effective usage of sensors. But, the presented approach failed to analyze the spatiotemporal characteristics.

Ojagh et al., [12] presented a location-based Orientation Aware Recommendation System (OARS). The various steps involved were similarity detection, ranking, and filtering of the events based on the current location and orientation of the user. Thus, the usage of the user orientation details addressed the cold start problem and provided personalized recommendations. However, user privacy and data secrecy were not concentrated.

Abbasi-Moud et al., [13] deployed Semantic Clustering and Semantic Analysis (SC-SA) for the formation of tourism RS. In this, the recommendation was provided based on the reviews given by the user. The different stages involved were pre-processing, semantic clustering, and sentimental analysis. Hence, this technique provided context-aware recommendations with better f-measure values. Nevertheless, the traffic information was not included, which in turn produced invalid results.

X. Zhang et al., [14] recommended a spatial analytical approach to provide recommendations for tourist spots. For determining the temporal and spatial characteristics of the tourist spot, time geographical methodology was used. The two techniques used for determination were space-time path and prism. Thus, this technique avoided the resource unavailability and unfamiliarity problem. Yet, the presented Point Of Interest (POI)-based methodology failed to consider all the tourist spots in the prescribed locations.

Pantano et al., [15] invented an Open Data Analysis (ODA)-based tourism decision-making technique. Various stages involved in this approach were pre-processing, normalization, and optimization.

The system's efficacy was provided by the experimental outcomes. However, the dependence on the initial data for generating recommendations affected the performance.

Cao et al., [16] designed Diversified Personalized Recommendation Optimization (DPRO) technique. For the optimization of non-dominated ranking, the Distributed Parallel Evolutionary Algorithm (DPEA) was utilized. Hence, the diversified recommendation problem was avoided with the effective increase in user satisfaction. Nevertheless, this system didn't consider the matrix decomposition problem.

Zhao et al., [17] presented Sentimental and Spatial Context (SSC) for personalized location recommendation. To generate recommendations, the POI mining technique was fused with SSC. Hence, the system's efficacy was improved by the effective usage of the characteristics and attributes of the geographical locations. Yet, due to the varying user behaviour, the recommendation generation based on sentiment similarity was not effective.

Bahari Sojahrood & Taleai, [18] exhibited Group RS in Location-Based Social Networks (LBSN). In terms of distance, time-hour, and category, the user influence was evaluated. Next, to generate group recommendations, the aggregation of the user influence was done. Hence, the presented approach performed well even for larger groups. However, the technique's transparency was affected by the exclusion of the demographic features.

Nassar et al., [19] developed Multi-criteria Collaborative Filtering-based Recommendation System (MCF-RS). Primarily, the features were extracted from the user reviews and ratings and further fed into the Deep Learning (DL) classifier to provide the prediction results. The system's performance was improved by the inclusion of Collaborative Filtering (CF). But, inaccurate results were offered by the usage of Multi-criteria ratings.

Turkar et al., [20] exaggerated Recommendation System (RS) based on the clustering of the tourist spot images for the effective clustering process. In this, Enhanced Moving K-Means (EMKM) was utilized. Primarily, the pixels were grouped depending on their texture, color, and shape features. Hence, the overlapping of the regions was effectively avoided. However, the dependence of the algorithm on the initialization criterion was a major drawback.

3. Proposed Tourism Recommendation System

This paper proposes an energy-efficient tourism RS through the effective determination of spatial and demographic characteristics of the tourist spot. The Geodatabase considered in the proposed system, including Google external API, Shapefile, and GIS map provides details about the cultural tourism spots and the safety level of the spots among different offenses in Riyadh country. Such a Geo database is gathered from publicly available data sources and is utilized as the input data in the proposed approach for developing the TRS. By employing the Geo database of Riyadh, the tourist site's location, the customer reviews, and the demographic data are processed using improved techniques in this paper to provide better recommendations for tourism sites.

So, the cultural aspects of tourist locations are identified and integrated into the proposed system for efficient recommendation. Moreover, spatial interpolation is employed to know about the status of regions around the tourism sites. For the effective generation of a recommendation list, the framework also utilized fuzzy-based decision-making. For the effective generation of a recommendation list, decision-making is utilized. In Fig. 1, the structural design of the proposed system is modeled.

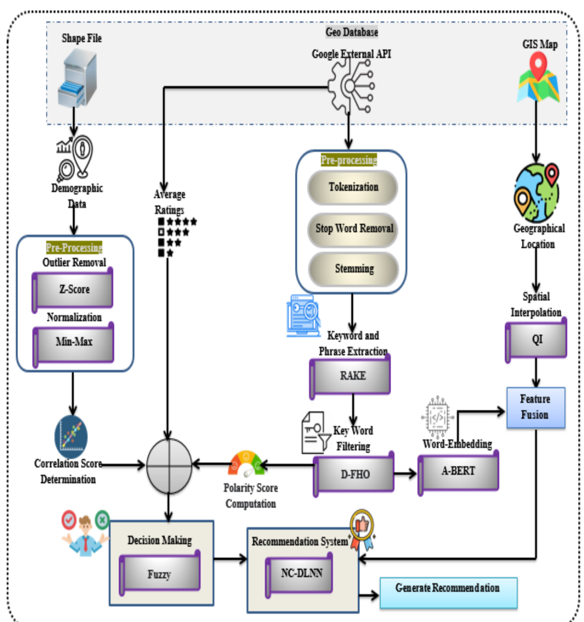

Fig. 1. Structure of the proposed Tourism recommendation system

The process involved in the proposed framework is detailed further.

3.1. Shapefile processing

Initially, the demographic data is obtained from the shapefile of the Geodatabase. Demographic data refers to the details regarding the societal status of the tourist places in Riyadh by considering its socioeconomic factors. The shapefile contains information regarding the violence rate, offenses rate, robbery rate, and certain other notifiable offenses in Riyadh. The demographic data $\left(D^{d}\right)$ obtained from the shapefile is described as,

$$D^{d} = \left\{D^{1}, D^{2}, D^{3}, ..., D^{n}\right\} \tag{1}$$

Where, $d = 1,2,3,...,n$ implies the number of demographic data obtained.

(a) Pre-processing

In this subsection, the demographic data $\left(D^{d}\right)$ obtained undergoes pre-processing. Pre-processing is carried out to reduce the training time of the proposed RS since the demographic data $\left(D^{d}\right)$ is noisy and contains certain irrelevant data within it. Pre-processing is the process of converting the raw demographic data into a useful format for effective processing. Outlier removal and normalization are the two main pre-processing steps involved in the proposed framework. It is detailed as follows,

(b) Outlier Removal

Primarily, all the outliers (irrelevant) present in the demographic data are removed. The Z-score normalization is used for the effective detection and removal of outliers. The mathematical formulation for the z-score approach $\left(D^{d}_{out}\right)$ is expressed as,

$$D^{d}_{out} = \frac{D^{d} - \sigma}{\mu} \tag{2}$$

Where, σ implies the mean value of the demographic data, and μ signifies the standard deviation.

(c) Normalization

Then, normalization is performed in order to make the demographic data $\left(D^{d}\right)$ suitable for further processing. Normalization is the process of effective removal of redundant data from the Geodatabase so that only the related data is stored within it. The normalization technique used is Min-Max normalization. Min-Max normalization performs normalization by transforming the minimum value of D^{d}_{out} to 0 and the corresponding maximum value to 1, whereas the rest of the values lie between 0 and 1. Thus, the Min-Max normalized data $\left(\hat{D}^{d}_{out}\right)$ is modelled as,

$$\hat{D}^{d}_{out} = \frac{D^{d}_{out} - \min\left(D^{d}_{out}\right)}{\max\left(D^{d}_{out}\right) - \min\left(D^{d}_{out}\right)} \tag{3}$$

Where, $\min\left(D^{d}_{out}\right)$ signifies the minimum value of D^{d}_{out}, and $\max\left(D^{d}_{out}\right)$ implies the maximum value of D^{d}_{out}. Hence, the preprocessed shape file is simply mentioned as $\aleph$.

3.2 Google External API processing

Afterward, the processing of the information obtained via the reviews and the ratings using API takes place. In this, API signifies the code used for the effective integration of application software. The information obtained using Google External API $\left(G^{e}\right)$ is explained as G^{e}, which includes $E-$ number of data obtained via Google external API.

(a) Pre-processing

To make the tourist recommendation system reliable for further processing, the data obtained using Google external API is pre-processed. The pre-processing steps involved are tokenization, stop word removal, and stemming. Each of the pre-processing steps is discussed further.

(b) Tokenization

It is the process of dividing the incoming Google External API data $\left(G^{e}\right)$ into numerous meaningful subunits named tokens. Tokenization is mainly performed depending on the spacing between the words in the text. For instance, the sentence "Diwali is celebrated all over the country" is tokenized as "Diwali", "is", "celebrated", "all", "over", "the", and "country". In this way, the entire text present is tokenized.

(c) Stop word removal

Then, the removal of stop words takes place. The words that occur repeatedly in the document and provide no specific meaning come under the stop word category. Some of the examples are conjunction ("and", "but", etc.,), Pronouns ("he", "she", etc.,), and so on. Thus, removing the stop words without the loss of crucial information not only reduces the corpus size but also results in a reduction of the time required for training the classifier.

(d) Stemming

Afterward, stemming is performed. The process of removing prefixes and suffixes present in the document by converting the complex word into its root word is termed stemming. For example, the word "eat", "ate", "eaten", and "eating" are converted into "eat". In this way, the words are transformed for further processing. This transformation results in improved classification outcomes. In the end, the data obtained after pre-processing is represented as $\left(G_{pre}^{e}\right)$

3.3 Keywords and phrase extraction by RAKE

Next, keywords and phrases are extracted from the pre-processed output. Keywords and phrases present in the pre-processed API are extracted to provide sufficient information regarding the tourist spot in Riyadh. The RAKE technique has the ability to handle multiple data. So, it is used to effectively extract keywords and phrases from the API. RAKE is an unsupervised approach used to extract the keywords related to review content. The steps involved in the selection of keywords using the RAKE algorithm are detailed further.

(a) Selection of keywords and phrases

Primarily, all the possible keywords and phrases related to tourism are selected from the pre-processed API data, and the selection process is expressed as,

$$G_{sel}^{e} = \frac{\sum_{e=1}^{E} G_{pre}^{e}}{E} \times 100 \qquad (4)$$

Where, G_{sel}^{e} implies the selected keywords and phrases.

(b) Keyword score-matrix determination

Then, the score values of each selected keyword are determined. The two main factors utilized for the determination of the keyword score matrix are keyword frequency (ω) and degree (ϕ). The keyword frequency refers to the number of frequent terms in G_{sel}^{e}, whereas the degree corresponds to the co-occurrence of each selected word. Thus, the score matrix $\left(\kappa^{(s)}\right)$ is formulated as,

$$\kappa^{(s)} = \frac{\omega}{\phi}\left(G_{sel}^{e}\right) \qquad (5)$$

The position ranking method is utilized to evaluate the frequently occurring phrases. The position ranking $\left(P^{r}\right)$ method is detailed as,

$$P^{r} = \aleph \cdot \beta\left(G_{sel}^{e}\right)(1-\aleph) \qquad (6)$$

Where, $\aleph$ implies the damping factor, and β signifies the position ranking constant.

(c) Keyword extraction

Finally, the keywords with the top scores and the phrases in the topmost position are extracted and modeled as A_N, which includes the $n-$ number of extracted keywords.

3.4 Keyword filtering using D-FHO

Thereafter, the optimal keywords are selected from the extracted keywords. To select the most frequently used keyword, the selection of optimal keywords takes place. D-FHO is used for the effective keyword filtering phenomena in the proposed methodology. The most popular Meta-heuristic algorithm inspired by the fire hawk's behavior to protect its prey from attacks is the Fire Hawks Optimization (FHO). FHO can converge faster in an efficient way. Nevertheless, the traditional FHO failed to consider the fire hawk's direction while searching for prey, which made the FHO fall in the local optimum. Thus, to alleviate this issue, the direction calculation is introduced in the conventional FHO. Hence, it is named as D-FHO. The steps involved in the proposed D-FHO are detailed further.

(a) Population initialization

Initially, the $n-$ number of fire hawks population (extracted keywords) $\left(A_N\right)$ is initialized in the D dimensional search space and is modeled as,

$$\mathrm{A}_{N} = A_1, A_2, A_3, ..., A_n \qquad (7)$$

Then, the position of the fire hawk $\left(\rho_{pos}\right)$ is randomly initialized as,

$$\rho_{pos}(0) = \rho_{\min} + rnd(\rho_{\max} - \rho_{\min}) \qquad (8)$$

Where, $\rho_{pos}(0)$ implies the initial prey position, $\rho_{\min}$ and $\rho_{\max}$ signify the minimum and maximum bound limits of the prey, respectively, and rnd elucidates the random integer uniformly distributed between $[0,1]$.

(b) Fitness evaluation

Then, each fire hawk's fitness in the population is determined. The fitness corresponds to the maximum frequency of repeated keywords $(\max(Af))$. Therefore, the fitness evaluation criterion is equated as,

$$f(A_N) = \sum_{N=1}^{n} \max(Af) \qquad (9)$$

Where, $f(A_N)$ specifies the fitness function. Hence, the one with the better fitness value is considered the strongest fire hawk (ξ_i), while the other members of the fire hawk population are considered the prey (ϑ_m) and are detailed below,

$$\xi_i = \xi_1, \xi_2, \xi_3, ..., \xi_I \qquad (10)$$
$$\vartheta_m = \vartheta_1, \vartheta_2, \vartheta_3, ..., \vartheta_M \qquad (11)$$

Where, $m = 1,2,3,...,M$ implies the $M-$ number of prey in the search space, and $i = 1,2,3,...,I$ signifies the $I-$ number of fire hawks.

(c) Prey searching process

Next is the searching phase. In this, the fire hawks move around for searching prey. Once the fire hawk identifies the prey, it spreads the fire around the prey to make the hunting process easier. Hence, the distance between the prey and the fire hawk is determined using the following expression,

$$d_x^y = \left(\sqrt{(u_2 - u_1)^2 + (v_2 - v_1)^2}\right) \qquad (12)$$

Where, d_x^y signifies the distance between x^{th} prey and y^{th} fire hawk, and (u_1, u_2) and (v_1, v_2) denote the coordinates in the search space. Next, the direction of movement of the fire hawk is evaluated using eqn (13). Therefore, the local optimum problem is minimalized.

$$\varphi = \arctan \frac{x}{y} \qquad (13)$$

Where, φ_x^y signifies the direction of movement of the y^{th} fire hawk towards the x^{th} prey.

(d) Collection of burning sticks

After the successful identification of the prey, the fire hawk begins to collect the burning sticks to set fire to the area, where the prey is identified. During the burning stick collection process, the fire hawks change their position. Hence, the updated position $(\hat{\rho}_{pos})$ is,

$$\hat{\rho}_{pos} = \rho_{pos} + (\wp^1 \times \xi_i - \wp^2 \times \rho_{near}), i = 1,2,3,...,I \qquad (14)$$

Where, $\wp^1$ and $\wp^2$ signify the random variables, and ρ_{near} implies the nearest fire hawk. Now, the fire hawk drops the burning stick around the prey zone.

(e) Movement of the prey

The prey decides to protect itself upon the reception of the burning stick. Thus, it tries to run away or hide, but unfortunately, it moves much closer to the second fire hawk while trying to escape from the first one. Hence, the updated position of the (ϑ_m^{pos}) becomes,

$$\vartheta_m^{pos} = \overline{\vartheta}_m + \wp^3 \times \xi_i - \wp^4 \times \hbar \qquad (15)$$

Where, $\overline{\vartheta}_m$ signifies the initial prey position, $\wp^3, \wp^4$ imply the random variables, and $\hbar$ denotes the safe place. Hence, the prey inside the burning stick is considered to be attacked, whereas the others outside are considered safe. So, the process is repeated until an optimal solution is obtained, and the $M-$ number of optimal features filtered is represented as A_m^{opt}.

Pseudocode for D-FHO

Input: Extracted keywords $A_1, A_2, A_3, ..., A_n$

Output: Optimal keywords

Begin

Initialize Hawks population and position (ρ_{pos}), $\rho_{\min}, \rho_{\max}$, maximum iteration $(It_{\max})$

Set $It = 1$

While $(It \leq It_{\max})$ **do**

Estimate Fire hawks (ξ_i) and prey (ϑ_m) within $\rho_{\min}, \rho_{\max}$

Determine the total distance d_x^y and direction $\left(\arctan\frac{x}{y}\right)$ between ξ_i and ϑ_m

Determine Fire Hawk territory

For hawks ξ_i **do**

Update hawk position with respect to $\rho_{pos} + (\wp^1 \times \xi_i - \wp^2 \times \rho_{near})$

For prey $\vartheta_m, m = 1,2,...,M$ **do**

Estimate safe place

Update prey position ϑ_m^{pos}

End For

End For

Estimate fitness of $\hat{\rho}_{pos}, \vartheta_m^{pos}$

If $(\hat{\rho}_{pos} \geq \rho_{pos})$ {

Obtain theGlobal best solution

} **Else** {

$It = It + 1$

Repeat While

}

End If

End While

Return optimal solutions A_m^{opt}

End

3.5 *Word embedding using A-BERT*

Now, for the effective generation of the recommendation list, the optimal keywords filtered are converted into numerals. Such conversion is carried out by means of A-BERT in the proposed work due to its ability to create contextual embedding. Usually, Bidirectional Encoder Representations from Transformers (BERT) is a pre-trained technique provided with an encoder component that accepts the input text data and provides the recommendation using the decoder. However, the inability to incorporate the target keywords in the classification output is the major downside. Hence, to overcome this downside, the target keywords are aligned and fed into the fully connected layer of BERT. This inclusion of alignment in the conventional BERT is the so-called A-BERT and is detailed further.

(a) Prey searching process

Token embedding layer: Initially, the filtered optimal keyword (A_m^{opt}) is fed into the token embedding layer of the A-BERT model. The input-filtered optimal keyword (A_m^{opt}) is mapped into its respective embedding vector through the following steps,

- Primarily, a [CLS] token is inserted at the beginning of the sentence and a [SEP] token is added at its end. CLS indicates the starting point, whereas SEP defines the sentence separator.
- Then, a sentence embedding indicating sentence A or sentence B is supplemented to each word in the sentence.
- Later, position embedding is executed to indicate the position of a word in the sequence.

(b) Transformer encoder layer

Afterward, the token embeddings obtained are further converted into numerical vectors in the encoder layer. To obtain long-range dependencies, the encoder layer is provided with a large number of self-attention layers. The output from the encoder layer is aligned before feeding it into the fully connected layer. The alignment process $(\varsigma(\upsilon, \zeta, T))$ is modeled as,

$$\varsigma(\upsilon, \zeta, T) = \sum_{m=1}^{M} \left(A_m^{opt} \zeta\right)^2 \qquad (16)$$

Where, υ implies the representation, ζ is the label vector, and T signifies the threshold value.

(c) Output layer

The aligned encoder output is then converted into vectors using the Softswish activation functio $\left(\chi\left(\varsigma\left(\upsilon,\zeta,T\right)\right)\right)$ and is represented as,

$$\chi(\varsigma(\cdot))=\frac{\varsigma(\cdot)*\gamma}{\sum e^{Q^t}} \tag{17}$$

Where, γ signifies the activation parameter. The output obtained at the A-BERT is O^{bert}.

3.6 Polarity score computation

At the same time, the polarity score is determined for the filtered optimal keywords of reviews. The polarity score corresponds to the positive, negative, and neutral comments. Therefore, the polarity score is determined as,

$$\overline{P}_{pos}\left(A_m^{opt}\right)=\frac{\varpi_{pos}\left(A_m^{opt}\right)}{\varpi(m)} \tag{18}$$

$$\overline{P}_{neg}\left(A_m^{opt}\right)=\frac{\varpi_{neg}\left(A_m^{opt}\right)}{\varpi(m)} \tag{19}$$

$$\overline{P}_{neu}\left(A_m^{opt}\right)=\frac{\varpi_{neu}\left(A_m^{opt}\right)}{\varpi(m)} \tag{20}$$

Where, $\overline{P}_{pos}\left(A_m^{opt}\right)$, $\overline{P}_{neg}\left(A_m^{opt}\right)$, and $\overline{P}_{neu}\left(A_m^{opt}\right)$ imply the polarity score of the positive, negative, and neutral keywords filtered out, respectively, and $\varpi_{pos}\left(A_m^{opt}\right)$, $\varpi_{pos}\left(A_m^{opt}\right)$, and $\varpi_{pos}\left(A_m^{opt}\right)$ refers to the frequency of the respective positive, negative, and neutral components, correspondingly. Hence, the polarity score for the filtered optimal keywords is determined as $\left(Pol^s\right)$.

3.7 Correlation score determination

Further, for the effective identification of secured tourist spots in Riyadh city, the correlation between the normalized demographic data $\left(\hat{D}_{out}^d\right)$ of various locations present in the Geodatabase is determined. In the proposed methodology, the correlation score determination phase provides details regarding the linear relationship between the various locations present in Riyadh depending on their offense rate. The correlation score determination is detailed as,

$$\hat{D}_{corr}^d=\frac{\delta\left(\hat{D}_{out}^d\right)}{\overline{\mu}\left(\hat{D}_{out}^d\right)} \tag{21}$$

Where, $\hat{D}_{corr}^d$ signifies the correlation score, $\delta\left(\hat{D}_{out}^d\right)$ implies the mean value of the normalized demographic data, and $\overline{\mu}\left(\hat{D}_{out}^d\right)$ signifies the standard deviation of $\hat{D}_{out}^d$.

3.8 Decision-making through fuzzy

Meanwhile, the decision is made depending upon the polarity score and correlation score obtained using the fuzzy method along with the effective usage of the average ratings provided. The Fuzzy approach is used for the decision-making process because of its ability to obtain the best data from the vague ones. The fuzzy decision-making process is detailed below,

- **Fuzzification interface**: Primarily, the polarity score and the correlation score obtained are transferred into a range of unique values, which in turn are transformed into low, very low, and high.
- **Knowledge base:** The low, very low, and high transformation is based on the membership function determined. Hence, the fuzzification of the variables is achieved.
- **Decision-making:** Decision-making based on the membership function is generally centered on the minimum value of the positive score values $\left(Po\bar{l}^s\right)$ obtained. Thus, the membership function becomes,
$$B_{dec}=\min_{1\le d\le D,1\le s\le l}\left(\hat{D}_{corr}^d,Po\bar{l}^s\right) \tag{22}$$
- **Defuzzification interface:** The obtained fuzzy output is converted into a crisp value $\left(\overline{R}\right)$, which is computed as follows,
$$\overline{R}=\frac{\left(B_{dec}\right)^q\cdot r}{B_{dec}} \tag{23}$$

Where, q implies the fuzzy rating, and r signifies the numerical rating.

3.9 GIS map details processing

Then, the location details obtained using the GIS map are effectively processed. The tourist location in Riyadh city obtained using a GIS map $\left(L^z\right)$ is depicted as L^z.

(a) Spatial Interpolation using QI

The spatial details of the corresponding location are identified from the location details obtained using the spatial interpolation phenomena. Spatial interpolation is used for the effective estimation of unknown locations using the details of the known area surrounding the tourist spot. Quintic interpolation is used for the effective determination of spatial characteristics because of its ability to operate on both smooth and continuous surfaces. Hence, the spatial characteristics obtained using QI (QI) is expressed as,

$$QI = \frac{ab(L^z) - c^2 \times L^z}{a^2 \times L^z} \tag{24}$$

Where, a, b, c illustrate the spatial characteristics.

(b) Feature fusion

For further classification, the feature values obtained using QI and the A-BERT are then fused. The process of combining the most suitable features obtained from corresponding historical sites is termed feature fusion. Hence, the output obtained after feature fusion is articulated as H^f.

3.10 Classification by NC-DLNN

Finally, the generation of tourist spot recommendations takes place using NC-DLNN based on the decision made using Fuzzy and the fused features obtained. A Deep Learning Neural Network (DLNN) made of the input layer, hidden layer, and output layer is used for recommendation list generation. However, backpropagation of the neural network takes place due to the random initialization of weight values, which in turn increases the training time of the neural network. In order to overcome this downside, the weight values are optimized using a Non-central Chi-squared distribution (NC). Hence, it is named NC-DLNN. In Figure 2, the structure of NC-DLNN is demonstrated.

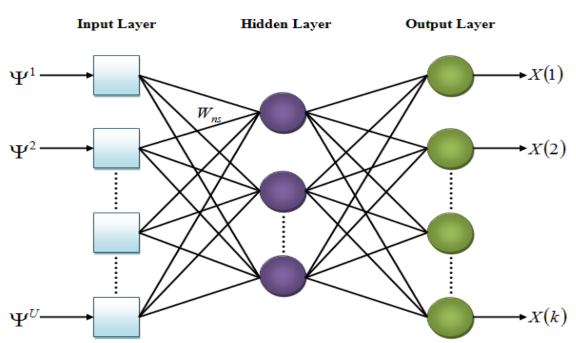

Fig. 2. Architecture of the proposed NC-DLNN model

Step 1: Initially, the input is fed to the input layer of NC-DLNN, and the input is multiplied with the input-hidden weight at each node. Hence, the hidden layer $(\hbar^{out})$ output is given as,

$$\hbar^{out} = \sum_{u=1}^{U} \Psi^u W_{ns} + \tilde{B}_{hid} \tag{25}$$

$$\Psi^u = \{O^{bert}, H^f\} \tag{26}$$

Where, U signifies the number of hidden layer nodes, W_{ns} denotes the weight value initialized using Non-central Chi-Squared Distribution (NC), which is defined in eqn (34), and $\tilde{B}_{hid}$ entitles the bias value of the hidden layer.

$$W_{ns} = \sum_{u=1}^{U} (\overline{F} + \eta^u)^2 \tag{27}$$

Where, $\overline{F}$ implies the standard variable, and η^u signifies the non-centrality parameter.

Step 2: The sigmoid activation function $(\gamma_{\hbar^{out}})$ is calculated for each hidden layer node as,

$$\gamma_{\hbar^{out}} = \frac{1}{1 + e^{\Psi^u}} \tag{28}$$

Step 3: The final output at each node of the output layer $(X(k))$ is calculated using equation (35).

$$X(k) = \left(\gamma_{\hbar^{out}} \left(\sum_{u=1}^{U} \Psi^u W_{ns} + \tilde{B}_{out} \right) \right), k = 1,2,3,\ldots,K \tag{29}$$

Where, $\tilde{B}_{out}$ signifies the bias value of the output layer. The output of this NC-DLNN network generates the low-recommended, more-recommended, and most-recommended tourist spots.

Pseudocode of the proposed NC-DLNN

Input: Fuzzy output $\left(\overline{R}\right)$ and H^f
Output: Recommendation output

Begin
Initialize input neurons Ψ^u, bias values $\left(\widetilde{B}_{hid}, \widetilde{B}_{out}\right)$
  For input Ψ^u **do**
    Initialize weights W_{ns} with the NC technique
  Perform hidden layer operation $\sum_{u=1}^{U} \Psi^u W_{ns} + \widetilde{B}_{hid}$
  Activate hidden neurons with sigmoid activation $\left(\gamma_{\hbar^{out}}\right)$
  Estimate output layer results $X(k)$
  End For
  Return recommended result
End

4. Result and Discussion

In this section, the detailed exploration of the final outcome of the proposed technique is explained. For the performance analysis, the GIS of Riyadh is gathered from the Saudi Arabia points of Interest, an open street map export data source. This data includes various features, including amenities, opening hours, tourism, source, street address, city address, and so on. Further, the demographic data of Riyadh is collected from Saudi Arabia's population in Riyadh data of Statista web source.

The links for these data sources are given below in the reference list. From each data source, the data is collected in the ratio of 80:20 for training and testing the proposed system, respectively. The proposed methodology is employed in the working platform of PYTHON by deploying these publically available datasets.

4.1 Performance analysis of keyword filtering

Initially, the proposed D-FHO's performance for keyword filtering is analyzed by comparing it with the prevailing FHO, Dwarf Mongoose Optimization (DMO), Crow Search Optimization (CSO), and Particle Swarm Optimization (PSO).

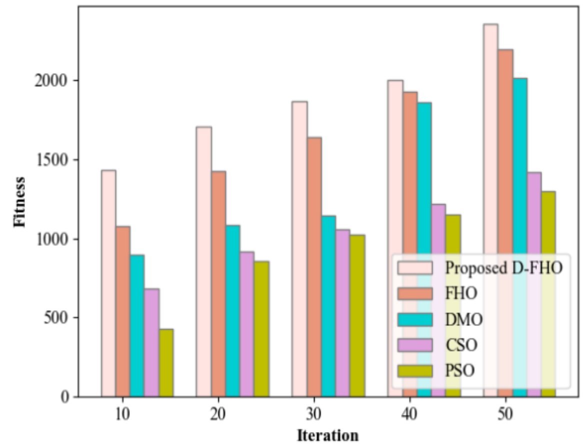

Fig. 3. Fitness analysis of proposed D-FHO for keyword filtering

The performance of the proposed approach based on fitness vs iteration is evaluated in Figure 3. The proposed D-FHO obtained a fitness of 1432 within a minimum of 10 iterations. However, the existing techniques like PSO delivered a lower fitness value of 432 for the same 10 iterations. Similarly, the fitness values vary (lower) for other methods also. Thus, when analogized with the proposed techniques, the proposed method obtained an increased fitness level. This is because the exploration capability is enhanced by analyzing the direction of Fire Hawks. It leads to better searching for prey and attaining better fitness than the existing methods. Hence, the consideration of the direction of movement in the proposed method tends to achieve the best fitness values, and also ensure classification accuracy.

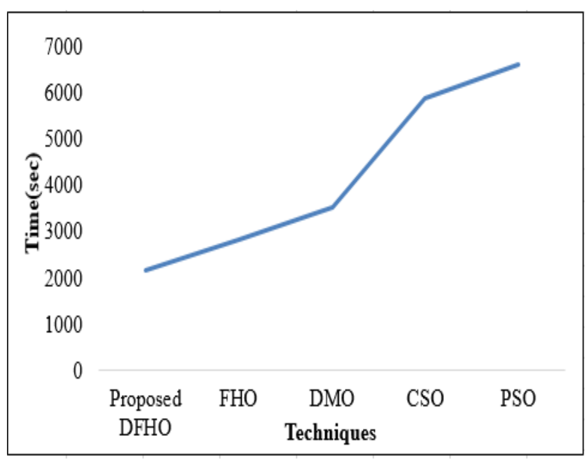

Fig. 4. Time complexity analysis of proposed D-FHO over the existing algorithms

The performance of the proposed technique based on its feature selection time is illustrated in Figure 4. For selecting features, the proposed D-FHO attained 2175 sec, whereas the existing techniques like FHO, DMO, CSO, and PSO showed higher feature selection times of 2837 sec, 3527 sec, 5896 sec, and 6632 sec, respectively. As the Fire Hawk's exploration for prey is improved regarding its direction, the optimal solution is achieved in a reduced time duration. Thus, the proposed system's performance is better than the conventional techniques.

4.2 Performance analysis of word embedding using proposed A-BERT

In this subsection, regarding the Bleu score, the proposed technique's performance for word embedding of filtered keywords is analogized with the prevailing BERT, GloVe, Fast text, and Skip Gram approaches.

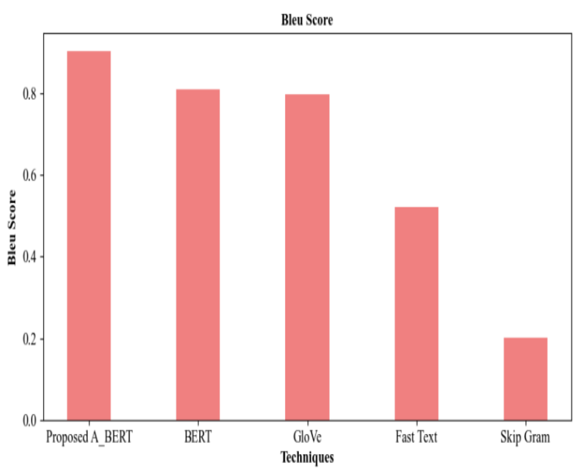

Fig. 5. Performance of proposed A-BERT based on Bleu score

The Bleu (Bilingual evaluation understudy) score is a performance metric that expresses the text quality produced by machine learning-based translators. It shows the similarity between the generated text and the original reference text. Generally, the Bleu score ranges from 0 to 1. The higher score closer to 1 indicates the better performance and vice versa. So, it is used to analyze the word or text embedding of Google external API, which is done using the proposed A-BERT model. The Bleu score obtained by the proposed technique is elucidated in Figure 5. A higher Bleu score indicates better performance. This improved performance of the proposed method is due to the alignment of target keywords within the model. So, the keywords are properly vectorized, thus resulting in efficient word embedding. Thus, the proposed methodology attains a better Bleu score of about 0.902, whereas the Bleu scores obtained by the existing BERT, GloVe, Fast Text, and Skip Gram are lower of the order of 0.809, 0.796, 0.521, and 0.201, respectively. So, the proposed technique produced better word embedding than the existing techniques.

4.3 Performance evaluation of the proposed NC-DLNN model

The proposed NC-DLNN's performance is analogized with the prevailing DLNN, Deep Belief Network (DBN), Artificial Neural Network (ANN), and Convolutional Neural Network (CNN).

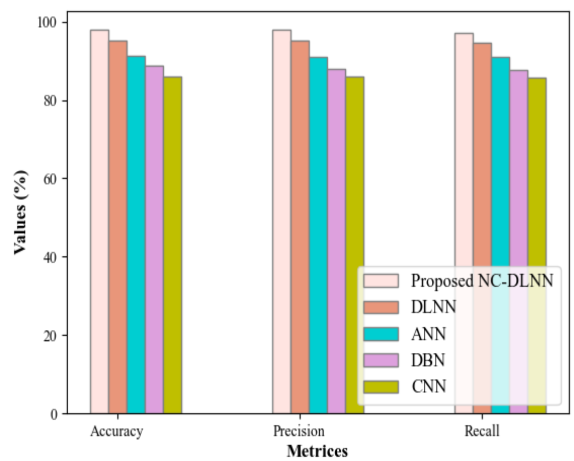

Fig. 6. Performance analysis of proposed NC-DLNN

Regarding accuracy, precision, and recall, the performance of the proposed NC-DLNN is explicated in Figure 6. For accuracy, precision, and recall, the NC-DLNN attains 97.91%, 97.90%, and 97.12%, respectively. But, the performance of the existing DBN (88.76% accuracy, 88.02% precision, and 87.70% recall) is low. Similarly, the rest of the existing classifiers also obtained lower accuracy, precision, and recall compared to the proposed method. Thus, due to the effective usage of Non-central Chi-Squared Distribution-based weight initialization, the gradient flow among the neuron layers gets regulated, resulting in improved data learning. So, the proposed NC-DLNN had better performance than the other existing techniques.

Table 1: Comparative analysis based on F-Measure, and specificity

Techniques	Performance Metrics (%)	
	F-Measure	Specificity
Proposed NC-DLNN	98.25	97.92
DLNN	94.93	95.34
ANN	90.91	91.11
DBN	87.25	88.54
CNN	85.36	86.17

The value of the F-measure and the specificity of the proposed technique are given in Table 2. For F-measure and specificity, the proposed method attains 98.25% and 97.92%, while the prevailing CNN attained 85.36% and 86.17%. The proposed technique achieved literally high performance based on F-measure and specificity metrics. As various criteria like fused features of word-embedded keywords, spatial interpolation, and fuzzy-based decision are assumed for training, the proposed approach generates the recommendation list in a more precise way.

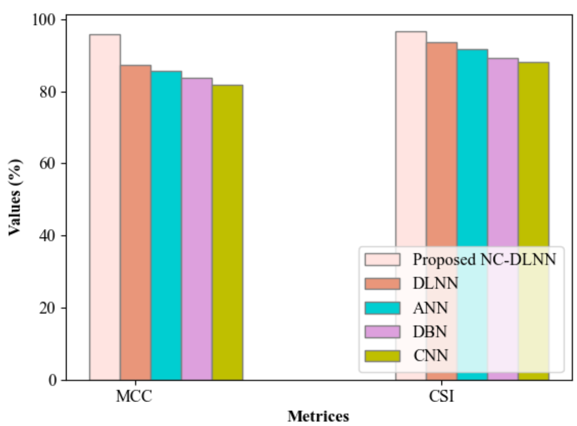

Fig. 7. Performance of the proposed NC-DLNN regarding MCC and CSI

The performance of the proposed NC-DLNN for tourism recommendation based on the Mathews Correlation Coefficient (MCC) and Critical Success Indicator (CSI)is analyzed in Figure 7. The higher MCC and CSI value denotes the robustness of the model. The proposed NC-DLNN attained 95.66% MCC and 96.55% CSI. Meanwhile, the existing DLNN attained an MCC of 87.39% and CSI of 93.71%, ANN attained an MCC of 95.74% and CSI of 91.75%, DBN attained an MCC of 83.69% and CSI of 89.2%, and CNN attained an MCC of 81.87% and CSI of 88.03%, which are lower than the proposed method. As multiple data factors, including GIS map, shape file, and Google external API are analyzed, the proposed model achieved increased performance than the other methods.

Table 2: Comparative measure of proposed NC-DLNN

Techniques	Performance metrics (%)		
	FNR	FDR	FPR
Proposed NC-DLNN	0.021	0.0141	0.0208
DLNN	0.0407	0.0276	0.0673
ANN	0.0574	0.0457	0.0875
DBN	0.0745	0.8521	1.3201
CNN	0.9897	1.6324	2.5410

The performance evaluation of the proposed NC-DLNN regarding FPR, False Discovery Rate (FDR), and FNR is evaluated in Table 2. The lower value of FNR, FDR, and FPR interprets the reduced misprediction. The proposed NC-DLNN achieved FNR of 0.021%, FDR of 0.0141%, and FPR of 0.0208%. The proposed model effectively learned the data by proper weight initialization using NC. So, the complex relations among the data were exactly learned without being trapped in an overfitting issue. So, the proposed model generated recommendations with reduced error. But, the FNR, FDR, and FPR values of the existing DLNN are 0.0407%, 0.0276%, and 0.0673%, respectively. Similarly, these metrics rates are also increased for other techniques, indicating their inferior performance. Thus, the proposed work is more reliable and outperforms the existing approaches.

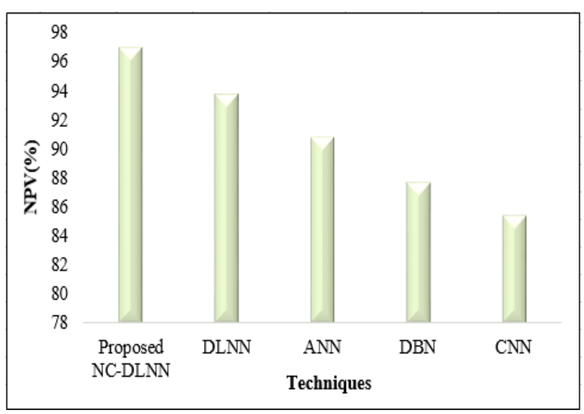

Fig. 8. Graphical illustration of proposed NC-DLNN's performance based on NPV

The superiority of the proposed methodology regarding Negative Predictive Value (NPV) is exhibited in Figure 8. The NPV is a measure that represents the true negative result among the overall negative test outcomes. It also represents the accurate evaluation of the recommendation outcomes. The NPV achieved by the proposed methodology is higher at the rate of 96.91%, whereas the existing DLNN, ANN, DBN, and CNN achieve the NPV at the order of 93.75%, 90.78%, 87.63%, and 85.39%, respectively. Thus, the proposed method outperforms the existing methodologies.

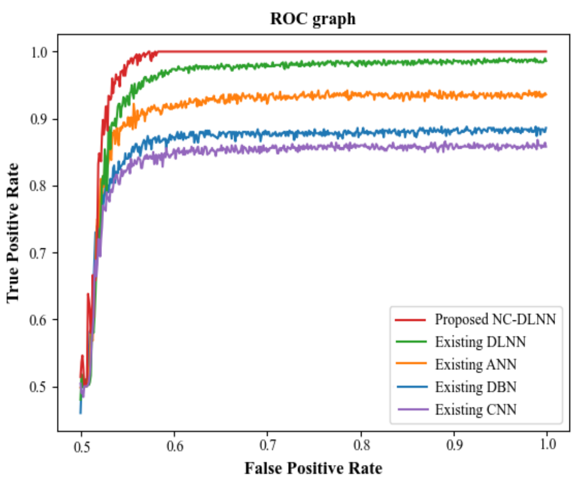

Fig. 9. ROC plot-based performance evaluation of proposed NC-DLNN

The Receiver Operating Characteristics (ROC) of the proposed NC-DLNN are exhibited in Figure 9. The ROC curve of the proposed technique converged at a higher TPR, which indicated better performance, whereas the performance of the other existing methods showed inferior performance. The inclusion of NC-based weight initialization along with the analysis of various factors of tourist sites contributed to the better performance of the model. Thus, the proposed approach performs better than the traditional techniques.

4.4 Comparative analysis with literature papers

Regarding efficiency, the proposed technique's performance is analogized to the prevailing OARS [12], DPRS[16], LSBN [18], and MCF-RS[19].

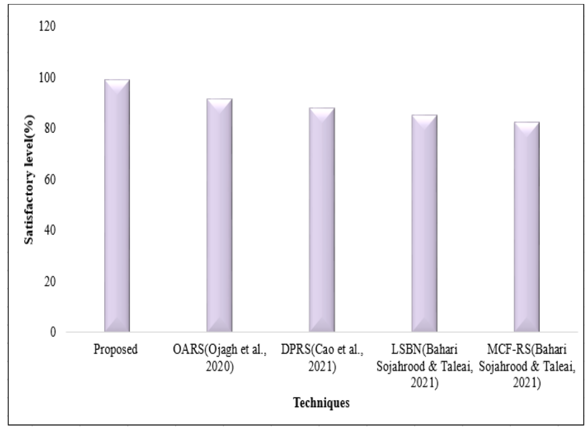

Fig. 10. Comparative analysis of proposed methodology in terms of Satisfactory level

The superiority of the proposed approach depending on the satisfactory level is revealed in Figure 10. The proposed tourism RS attains a satisfactory level of 98.95%, whereas the existing DPRS achieves only 88%. Likewise, the satisfactory level is also lower for other existing techniques. Thus, the usage of the spatial and demographic features along with the decision-making criterion improved the satisfactory level. Hence, the proposed approach withstands the conventional methodologies.

5. Conclusion

In conclusion, the tourism RS is proposed using NC-DLNN techniques via the effective usage of spatial and demographic characteristics. Several operations like pre-processing, keyword extraction, correlation and polarity score determination, and classification were included in the proposed framework. Then, to validate the effectiveness of the proposed algorithm, the experimentation analysis was performed. Various uncertainties could be handled by the developed approach. Also, it could render more promising results. For accuracy, precision, and specificity, the proposed method achieved 97.91%, 97.9%, and 97.92%, respectively. Moreover, a better Bleu score value of 0 was attained by the proposed embedding algorithm. Generally, the proposed recommendation framework outperformed the existing state-of-the-art methods and generated a reliable recommendation list through the usage of both spatial and demographic characteristics. Cultural richness was not considered despite the usage of shape features and the polarity of reviews in the determination of tourist areas in RIYADH. Hence, the work will be extended in the future using a sustainable cultural tourism model by conducting a pictographic survey among locals.

Conflict of Interest

None

References

[1] Banerjee, S., Kabir, M. M., Khadem, N. K., & Chavis, C.; Optimal locations for bikeshare stations: A new GIS based spatial approach. Transportation Research Interdisciplinary Perspectives, 4, 10010, 2020. https://doi.org/10.1016/j.trip.2020.100101.

[2] Zhang, J., & Dong, L; Image monitoring and management of hot tourism destination based on data mining technology in big data environment. Microprocessors and Microsystems, 80, 103515, 2021. https://doi.org/10.1016/j.micpro.2020.103515

[3] Zhang, A., Yang, Y., Chen, T., Liu, J., & Hu, Y. (2021). Exploration of spatial differentiation patterns and related influencing factors for national key villages for rural tourism in China in the context of a rural revitalization strategy, using GIS-based overlay analysis. Arabian Journal of Geosciences, 14(2), 1-14. https://doi.org/10.1007/s12517-020-06381-9

[4] del Carmen Rodríguez-Hernández, M., & Ilarri, S. (2021). AI-based mobile context-aware recommender systems from an information management perspective: Progress and directions. Knowledge-Based Systems, 215. https://doi.org/10.1016/j.knosys.2021.106740

[5] Zheng, W., Liao, Z., & Lin, Z.; Navigating through the complex transport system: A heuristic approach for city tourism recommendation. Tourism Management, 81, 1–33, 2020. http://dx.doi.org/10.1037/xge0000076

[6] Sahani, N., & Ghosh, T.; GIS-based spatial prediction of recreational trail susceptibility in protected area of Sikkim Himalaya using logistic regression, decision tree and random forest model. Ecological Informatics, 64, 101352, 2021. https://doi.org/10.1016/j.ecoinf.2021.101352

[7] Law, R., Li, G., Fong, D. K. C., & Han, X.; Tourism demand forecasting: A deep learning approach. Annals of Tourism Research, 75, 410–423, 2021. https://doi.org/10.1016/j.annals.2019.01.014

[8] Sun, S., Wei, Y., Tsui, K. L., & Wang, S.; Forecasting tourist arrivals with machine learning and internet search index. Tourism Management, 70, 1–10, 2019. https://doi.org/10.1016/j.tourman.2018.07.010

[9] Esmaeili, L., Mardani, S., Golpayegani, S. A. H., & Madar, Z. Z.; A novel tourism recommender system in the context of social commerce. Expert Systems with Applications, 149, 113301, 2019. https://doi.org/10.1016/j.eswa.2020.113301

[10] Zhang, Q., Lu, J., & Jin, Y.; Artificial intelligence in recommender systems. Complex and Intelligent Systems, 7(1), 439–457, 2021. https://doi.org/10.1007/s40747-020-00212-w

[11] Torres-Ruiz, M., Mata, F., Zagal, R., Guzman, G., Quintero, R., & Moreno-Ibarra, M.; A recommender system to generate museum itineraries applying augmented reality and social-sensor mining techniques. Virtual Reality, 24(1), 175–189, 2020. https://doi.org/10.1007/s10055-018-0366-z

[12] Ojagh, S., Malek, M. R., Saeedi, S., & Liang, S.; A location-based orientation-aware recommender system using IoT smart devices and social networks. Future Generation Computer Systems, 108, 97–118, 2020. https://doi.org/10.1016/j.future.2020.02.041

[13] Abbasi-Moud, Z., Vahdat-Nejad, H., & Sadri, J.; Tourism recommendation system based on semantic clustering and sentiment analysis. Expert Systems with Applications, 167(2-3), 1-17, 2021. https://doi.org/10.1016/j.eswa.2020.114324.

[14] Zhang, X., Yang, Y., Zhang, Y., & Zhang, Z.; Designing tourist experiences amidst air pollution: A spatial analytical approach using social media. Annals of Tourism Research, 84, 2020. https://doi.org/10.1016/j.annals.2020.102999

[15] Pantano, E., Priporas, C. V., Stylos, N., & Dennis, C.; Facilitating tourists' decision making through open data analyses: A novel recommender system. Tourism Management Perspectives, 31, 323–331, 2019. https://doi.org/10.1016/j.tmp.2019.06.003

[16] Cao, B., Zhao, J., Lv, Z., & Yang, P.; Diversified personalized recommendation optimization based on mobile data. IEEE Transactions on Intelligent Transportation Systems, 22(4), 2133–2139, 2021. https://doi.org/10.1109/TITS.2020.3040909

[17] Zhao, G., Lou, P., Qian, X., & Hou, X.; Personalized location recommendation by fusing sentimental and spatial context. Knowledge-Based Systems, 196, 105849, 2020. https://doi.org/10.1016/j.knosys.2020.105849

[18] Bahari Sojahrood, Z., & Taleai, M.; A POI group recommendation method in location-based social networks based on user influence. Expert Systems with Applications, 171, 114593, 2021. https://doi.org/10.1016/j.eswa.2021.114593

[19] Nassar, N., Jafar, A., & Rahhal, Y.; A novel deep multi-criteria collaborative filtering model for recommendation system. Knowledge-Based Systems, 187, 1-15, 2020. https://doi.org/10.1016/j.knosys.2019.06.019

[20] Turkar, H. R., & Thakur, N. V.; Performance comparison of clustering algorithms based image segmentation on mobile devices. In Springer, Singapore. 2018 https://doi.org/10.1007/978-981-13-0617-4_45

Dataset links:

https://datasetsearch.research.google.com/search?ref=TDJjdk1URjBjekZqTjJKc2VnPT0sTDJjdk1URnNibkV5WkRaaWF3PT0%3D&query=versions:Saudi+Arabia+Points+of+Interest+%28OpenStreetMap+Export%29&docid=L2cvMTFsbnEyZDZiaw%3D%3D&vers=MTAzODg1ODQyNzc5Mjg5NTMyNDM%3D

https://datasetsearch.research.google.com/search?src=0&query=riyadh%20demographic%20data&docid=L2cvMTFyOHpsaHlnbQ%3D%3D