IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 5 March 2023, accepted 18 March 2023, date of publication 21 March 2023, date of current version 24 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3260103

RESEARCH ARTICLE

Travel Direction Recommendation Model Based on Photos of User Social Network Profile

PAVEL STEFANOVIČ AND SIMONA RAMANAUSKAITĖ
Faculty of Fundamental Science, Vilnius Gediminas Technical University, 10223 Vilnius, Lithuania
Corresponding author: Pavel Stefanovič (pavel.stefanovic@vilniustech.lt)

This work was supported in part by the European Union Structural Funds, Lithuanian Science, Innovation and Technology Agency under Grant 01.2.1-MITA-T-851-01-0159.

ABSTRACT Travelling is one of the most enjoyable activities for people of all ages. It is constantly looking for innovative solutions on how to tailor travel recommendations to the needs of its customers. The purpose of our proposed recommendation model is to suggest travelling countries based on photos from the user's social network account and metadata associated with the photos. Such recommendation models are highly dependent on the data used in the model preparation steps and on the technologies and methods implemented in the model. The newly collected data from the Instagram users' accounts were used in the model preparation. The recommendation system is based on the combination of four methods: object detection, similarity measures, classification, and data clustering. The novelty of the proposed recommendation model is that it adopts different data (Instagram photos) for travel direction recommendation, defines a new combined method, integrates results of similarity measurement and SOM application results into one final recommendation, and estimates the parameter impact for different components of recommendation model. A proposed evaluation measure has been used to conclude the results of the recommendation model and as a result the names of the travelling countries have been recommended. The results of the proposed recommendation model are promising, and the validation results demonstrate that on average 63% of the users who visited countries match the recommendations provided for the trip directions, while the accuracy of recommendations, matching user visited countries, but not presented in the photos for recommendation estimation, on average was 96%. The accuracy performance is very positive, while the recommendation system is fully automated and machine learning based. With time, the accuracy of the model may even increase by adopting the photo metadata (location).

INDEX TERMS Classification, object detection, recommendation model, self-organizing maps, similarity measure, photo of social networks.

I. INTRODUCTION

According to various statistical studies around the world, many people missed out on travelling during the pandemic situation. Therefore, when the situation of COVID improves and various restrictions have disappeared, people start to travel again. It is one of the activities that people like the most at different ages, so for business it is not only a lucrative field but also a competitive market. Therefore, to increase the success of travel operators in finding potential customers, targeted advertising is very effective. This happens because a potential user is automatically selected for a trip that may be of interest to him.

Predicting the most accurate travel destination for a particular user can be difficult. However, research has shown that there are similarities between user groups that determine what type of travel a user group may like [1]. Nonetheless, to implement a recommendation system qualitatively, the data used for such a system are crucial. Recommendation systems are widely used in different areas [2], [3], [4], but travel recommendation systems face the problem of the need and accuracy of labelled data [5]. Although many travel

The associate editor coordinating the review of this manuscript and approving it for publication was Alberto Cano.

recommendation systems currently rely on data provided by social networks and other platforms for user hobbies and travel [6], [7], [8]. Some of the systems incorporate data from a wide variety of systems and even smart devices from the Internet of Things [9]. There is always the possibility that the consumer was not impressed or even disappointed with his or her travel, but the data do not show it. Therefore, an attempt is made to address this problem by integrating as much data as possible [10] or including interactive user surveys as additional information [11], [12].

Currently, one of the most popular social networks, Instagram, has over a billion users worldwide. The main advantage of Instagram over other social networks is the predominance of posting photos. Photos are related to hobbies, travel, etc. Therefore, the analysis of publicly available consumer photos can provide travel agencies with the necessary information to enable them to offer the appropriate type of travel to the consumer. The fact that photos reflect the user's opinion rather than the responses to questionnaires or even similarities with other users is also observed in the research by Linaza et al. [13].

The main objective of this article is to reduce uncertainty to the extent that modern data analysis methods, such as data classification and clustering, are appropriate for recommending different types of travel to users based on photos published on their social networks. The goal of the experimental investigation would be to examine existing solutions, compare them with each other, and propose the most promising model, adopting different results of the analyzed methods.

The idea of a travel recommendation should be based on the objects identified in the user's Instagram profile photos rather than post metadata only. This potentially will allow travel direction recommendations based on preferred activities, rather than geographical location alone. In practice, such a proposed recommendation model would increase the chances that the travel agency actually chooses at least one of the several trips shown to the consumer. This would meet his or her needs. As a result, the advertising from the travel agency would be used more purposefully, and the consumer would receive a set of travel offers that better meet his expectations. Furthermore, the consumer's satisfaction with the experience would increase accordingly.

In this paper, a recommendation model based on a combination of supervised and unsupervised methods results has been proposed. First, Instagram user data has been collected and pre-processed using Microsoft Azure to identify objects in photos [14]. The final pre-processed data consist of 4683 attributes, where four attributes are metadata and the rest are object detection in the photos results. The data collected will be used in the future to train some components of the recommendation model. These components will be able to identify countries that users have already visited and suggest new countries to visit. To determine which countries users have already travelled to, two methods have been used: the Python geopy library [28] for location identification based on the textual description of the location and the classifier of objects identified in the photos. When the data item fed to the recommendation model does not have any metadata or the visited country list has not been determined, similarity distance and self-organizing maps have been applied to identify possible countries based on object detection results. The results of the proposed model have been concluded by combining the results of similarity distance and clustering into a final recommendation model. The model incorporates different aspects of the similarity distance and clustering results to determine the final travel destination recommendation list based on the user's previous travelling photos. A more detailed description of the proposed recommendation model is presented in Section III.

The novelty of the proposed recommendation model is that it is fully automated and does not require any manual changes. Artificial intelligence methods allow us to retrain the model over time, improving its accuracy. Unlike most other recommended models, the recommendation is performed by extracting data from photos, and if appropriate, the metadata of each photo has been included and analyzed too. Such input data for travel direction, country recommendation was not presented before. The scientific novelty of the manuscript is the combination of a few well-known methods to perform recommendations using not only the well-known similarity distance, but also self-organizing maps. Usually, the self-organizing maps are used to cluster or visualize the data in a general form, but going deeper into the structure of the self-organizing maps the neighbouring rank can be modified and adapted to find out the most related data items in the self-organizing maps. In this manuscript, we modified and adapted the usage of neighbouring rank in the self-organizing map cell by combining it with similarity measures to find the most similar data item to the new data item fed into the model. In this way, the countries' recommendations are provided. In addition, self-organizing maps have difficulties in performance using high-dimensional data, so they have been combined with dimensionality-reduced methods. The proposed evaluation measure allows to summarize the results obtained with all methods and to provide a travel recommendation. The proposed recommendation model can be beneficial to businesses and can be easily implemented on websites and customised.

The structure of the manuscript is as follows. In Section II, related works are reviewed. In Section III, the scheme of the proposed recommendation model is presented and all steps are described. In Section IV, we describe the experimental investigation and validation of the proposed recommendation model. In Section V, the discussion and possible limitations of the proposed model are presented. Section VI concludes the paper.

II. RELATED WORKS

The development of a recommendation system based on users' published photos and recommending a travel destination accordingly requires the interoperability of different technologies and methods. A review of recommendation systems in the field of tourism shows that an increasing proportion of them rely on big data processing and artificial intelligence solutions [15]. Solutions to many existing user-based travel recommendation systems are based on finding a specific location on a photo or a specific user found on a photo [16]. However, in our research, we investigated the extent to which photographs depict common objects, such as animals, notes, food, and more. This is because photographs may reflect the user's profile and interface with the country of interest to the user. Today, the problem of object detection in photographs is also a highly analyzed field. The main key is to provide a list of recognized objects and the probability of their identification [17]. Having a list of objects that have been detected in the photo, the list could be used in classification tasks. This list could be used to determine a country similar to the data item. The object detection results are influenced by various factors, such as the algorithm selected and the way in which it has been trained.

Scientific literature analysis showed that there is no publicly available dataset that could be used to prepare a recommendation model based on objects detected in the images. Usually, all datasets are focused on different aspects and therefore aren't applicable. Compiling a dataset and preparing the data for research is not a trivial task. This is because the accuracy of modern artificial intelligence solutions is highly dependent on the data used for training and their preparation. Sometimes, including too much context in a decision does not increase the accuracy of the decision but reduces it [18]. Therefore, it is necessary to find a balance between the completeness of the data and redundancy. The classification of multiple levels in travel recommendation systems helps to solve the problem of data redundancy [19]. This problem is often due to the very high integration of different data sources [20], but it is also possible in the object detection field, where the objects depicted in the photos may be too specific to be identified (for example, a specific dish is identified rather than simply a category of food). In addition to multilevel classifiers, a knowledge graph base [21] and interfaces between objects the analyzed are also developed.

In many scientific studies, recommendation systems employ well-known methods. One of the old but often used techniques is calculating the similarity distance between vectors. There are a variety of similarity measures. Effectiveness is trusted and can be used when pair-wise similarity has to be determined, but new heuristic measures are emerging. In the manuscript by Ali et al. [22], the performance of three newly formulated similarity measures, namely the difference-based similarity measure, the hybrid difference-based similarity measure, and the triangle-based cosine measure, has been investigated. There is more research in which novel measures are proposed and analyzed [23], [24]. Based on the obtained results, it can be concluded that the efficiency of the newly proposed measures does not differ significantly from traditional similarity measures. These measures are usually used for specific tasks. The various classification and clustering algorithms are usually combined with the similarity distance results to develop the recommendation model. The support vector machine, decision trees, random forests, or even deep learning algorithms, like long short-term memory [26] are examples of traditional classification algorithms [25]. There are also many data clustering algorithms, but the most commonly used are K-means, hierarchical clustering, and K-nearest neighbour [27]. It is obvious that there are many different combinations of how these types of methods can be combined to obtain a high accuracy recommendation.

The literature review has shown that the travel recommendation model developed by other researchers usually uses structured data and traditional machine learning algorithms.

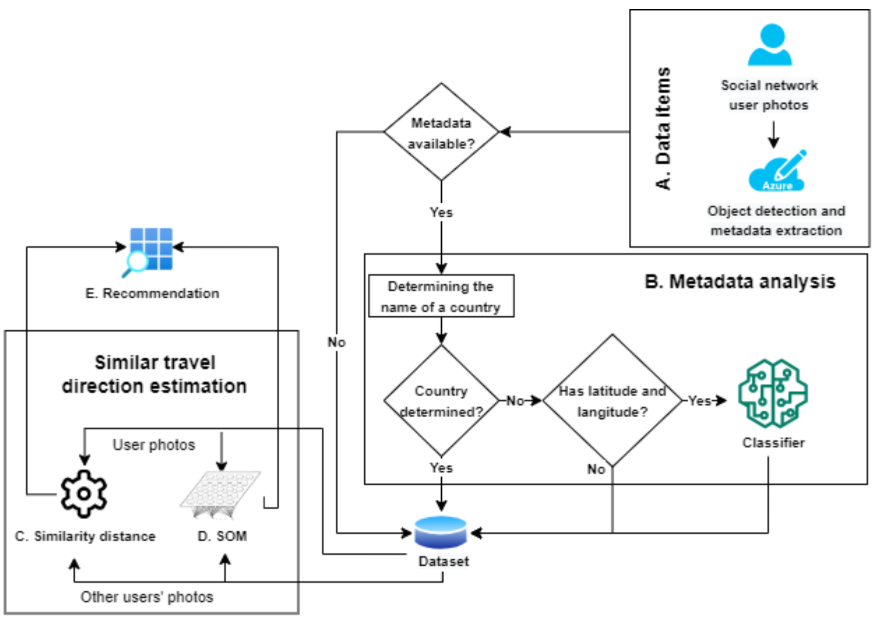

FIGURE 1. The abstract schema of the proposed travel direction recommendation model.

In this manuscript, differently from other researches, the unique features of self-organizing maps have been implemented for the travel recommendation model. Also, the difficulty of this research is to analyze not just the metadata of social networks but also images. This leads to a complex problem and a possible solution in such a situation. The research carried out in the following sections gives promising results.

III. THE PROPOSED RECOMMENDATION MODEL

We propose a recommendation model based on photos from Instagram social profiles and their metadata to recommend user travel destinations. The concept of the proposed recommendation model is presented in Fig. 1. The main components of the model are: data scraping from users' social network accounts; object detection in user photos and obtaining the metadata of each photo; detection of countries already visited; detection of possible countries to visit; recommendations based on all obtained data.

The results of this recommendation model can depend on various factors, such as the similarity distance measure, object detection, and the classification method that has been used. Also, the importance of the chosen parameters in each method, data quality, model evaluation, etc. Therefore, primary research has been performed to find the most appropriate parameters and algorithms for this type of recommendation model. We must agree that any recommendation model has limitations and, using another dataset, the results could vary, but the overall results of our proposed recommendation model are promising. Using photos and metadata to predict travel countries is a complicated task, so the verification and preparation of the model are also complicated and influenced by a variety of factors.

A. DATASET USED IN MODEL PREPARATION

Any recommendation model must be prepared and trained using historical data (in the model, is marked as Historical data). So, first, the Instagram users' profiles that have agreed to share their traveling photos have been web-scraped. The criteria for user selection were: 1) to have at least 10 photos taken during travel; 2) for at least half of the photos to define the country where the photo was taken (done by the user or by estimating photo metadata or text description); 3) user to participate in the research and define how he or she evaluates the recommendation. After that, object detection and metadata retrieval of each user photo were performed using the Microsoft Azure tool. The dataset has been filtered, and only the unique data items that had the full metadata have been included. The collected dataset is a set $X_s = \{x_1, x_2, \ldots, x_n\} \in R^n, s = 1, \ldots, N$, where $N = 12460$, $n = 4683$ [14]. The each data item X_s has 4 metadata records: country name, full location address, latitude, and longitude. The remaining data item attributes (4679) are the probabilities of objects detected in the photos, such as landscape, outdoor, tree, human, sea, water, etc. The distribution of the dataset by country names is presented in Fig. 2.

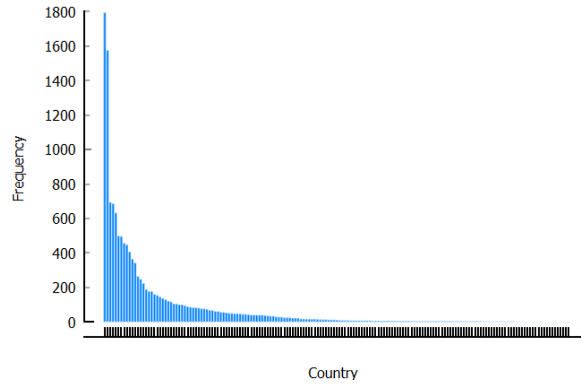

FIGURE 2. The countries names distribution in the dataset.

The dataset is unbalanced and it has totally 169 unique countries, where the most photos (of all data) belong to the United States (13.16%), Italy (11.55%), Spain (5.08%), India (5.03%), Brazil (4.64%), France (3.65%), Greece (3.64%), United Kingdom (3.35%), Germany (3.28%), Turkey (2.98%), Portugal (2.68%), Mexico (2.51%), and the other countries have less than 2% of the data items, starting from various countries on the European and other continent to small islands.

The large number of objects detected in the overall dataset of user photos results in a large number of 0 values. This indicates that the subject was not captured in the photo. In this way, the so-called thinned vectors are obtained when the dominant value in each vector is 0, and only the probabilities where the object has been detected are not equal to 0. It is obvious that in one photo all 4679 objects cannot be seen at the same time. Some of the objects are unique and can only be found in certain countries. Therefore, it can be used to find similarities between photos.

B. METADATA ANALYSIS

The described dataset has been used in three parts of the proposed recommendation model: 1) the metadata used to train the classifier; 2) to calculate similarity distance between new data items fed to the model and historical data items, to determine the most similar TOP 10 countries names in the dataset; 3) to train the self-organizing maps.

To determine which similarity distance measure and classifier are the most suitable, fits in the proposed recommendation model, the dataset has been split into 5 parts for the cross-validation method. The same parts of the data have been used in two stages: classification and similarity detection; training and evaluation. To train the self-organizing maps, the entire dataset has been analyzed. Metadata have been omitted from the model, since the main purpose is to recommend travel destinations based solely on photos. Metadata is used to determine which country the consumer has already visited.

In the perfect case, the analyzed data items fed to the recommendation model consist of full data: 4 metadata attributes and 4679 object detection results. In this case, we have the list of countries where consumers have already been. Sometimes, the data cannot be retrieved fully; for example, the metadata

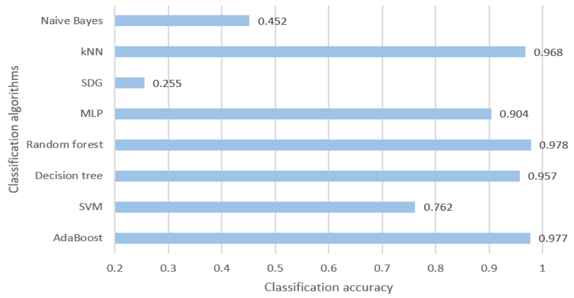

FIGURE 3. The efficiency of classification algorithms to identify the country, based on its latitude and longitude data.

of photos may not be full or may be partially full. First, if the data have at least one metadata attribute, the Python library geopy [28] is used to determine previously visited countries and adjust the data record. The classification algorithm is used to determine the list of countries that have been visited when the Python library geopy fails, but the updated data item has latitude and longitude metadata. In case, if the data items have full and correct metadata, the data items still have been fed to the model to pre-process the data and save it to the historical data database. Over time, the accuracy of the model should increase and provide higher quality recommendations.

When the visited countries of the consumer have not been identified using the Python geopy library [28], but the photo data contain latitude and longitude information, the classifier is used to estimate the country by using two data attributes: latitude and longitude. There are many different classification algorithms that have their own advantages and disadvantages in different fields. Efficiency analyses of various classification algorithms have been performed in the past decade [29], [30], [31]. Usually, traditional classification algorithms have been used for different solutions, because they are faster than deep learning algorithms, so such algorithms are more suitable for real-time tasks as recommendation systems. In this paper, the efficiency of eight classification algorithms has been investigated: AdaBoost, Decision trees, Random forest, a support vector machine (SVM), naive Bayes, multilayer perception (MLP), stochastic gradient descent (SGD), and k nearest neighbour (kNN). The related work analysis shows that these algorithms are suitable for fast upcoming data prediction and fast model retraining after historical data are updated.

The experimental investigation has been performed using the same five folds of the dataset as used in similarity distance calculations described in Section III-A. During the training of each classification algorithm, hyperparameter optimization has been used to optimize the accuracy obtained. The overall results are presented in Fig. 3.

As we can see, the lowest accuracy was obtained using the stochastic gradient descent algorithm (22.5%). Naive Bayes and support vector machines are, respectively, 45.2% and 76.2% accurate. The highest accuracy is obtained using the random forest algorithm (97.8%), but the accuracy of the AdaBoost algorithm is slightly lower (97.7%). There is no big difference between them that should be considered in the proposed recommendation model, so the random forest algorithm has been chosen.

TABLE 1. The similarity distance has been analyzed.

Distance name	Equation	Comments
Euclidean	$d(X,Y) = \sqrt{\sum_{k=1}^{n}(x_k - y_k)^2}$	The Euclidean distance is a special case of the Minkowski distance, where $p = 2$.
City block	$d(X,Y) = \sum_{k=1}^{n}\lvert x_k - y_k\rvert$	The city block distance is a special case of the Minkowski distance, where $p = 1$.
Chebychev	$d(X,Y) = max_k\{\lvert x_k - y_k\rvert\}$	The Chebychev distance is a special case of the Minkowski distance, where $p = \infty$.
Minkowski	$d(X,Y) = \sqrt[p]{\sum_{k=1}^{n}\lvert x_k - y_k\rvert^p}$	-
Cosine	$d(X,Y) = \frac{\sum_{k=1}^{n} x_k y_k}{\sqrt{x_k^2} \times \sqrt{y_k^2}}$	-

C. SIMILARITY DISTANCES

The related work showed that there are many different similarity distance measures that could be used to find out how one data item is similar to the rest of the dataset items. In this paper, we have been analyzing the effectiveness of several commonly used similarity distances using our newly collected data. Most of the chosen similarity distances have been used and analyzed in overviewed related work recommendation systems, so it is important to investigate the performance of them in our proposed model, too. Suppose that we have two data items of the same dimensionality n: $X = (x_1, x_2, \ldots, x_n)$ and $Y = (y_1, y_2, \ldots, y_n)$. The calculation of similarity distances between these data items $d(X, Y)$ is presented in Table 1.

To find the efficiency of similarity distances, an experimental investigation was performed. As mentioned above, the dataset has been split into five folds. 80% of the dataset has been used as historical data, and the rest 20% to test whether the similarity distance can determine which country the test data item really belongs to. First, the distance between each test data item and all historical data has been calculated using all similarity distances presented in Table 1. In the next step, a calculation was performed to determine at which position the country name of the test data item is detected. We have calculated three variants: 1) The real country of the test data item has been recognized in the first place; 2) The real country of the test data item falls into TOP 5; 3) The real country of the test data item falls into TOP 10. Experiments have been carried out with each fold separately. The results of all the folds were averaged. The concluded results are presented in Fig. 4.

As we can see, the results are similar and there is no significant difference in how the similarity distance is calculated. This is not a very high result when we talk about

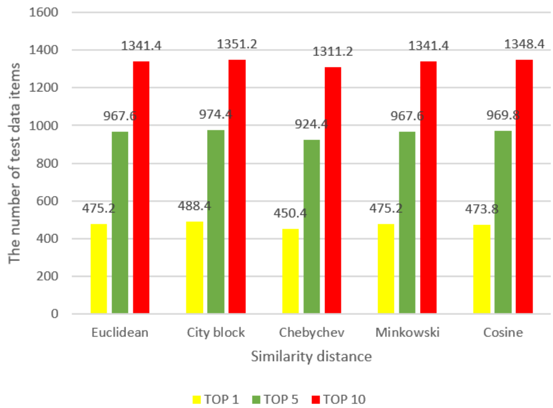

FIGURE 4. The efficiency of similarity distance, when 2492 items were tested with five different similarity distance methods to select the TOP most similar countries and evaluate is the country in the photo is present in the TOP similar photos.

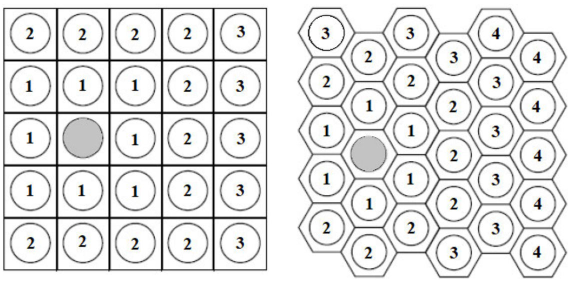

FIGURE 5. Two-dimensional SOM, rectangular, and hexagonal topology.

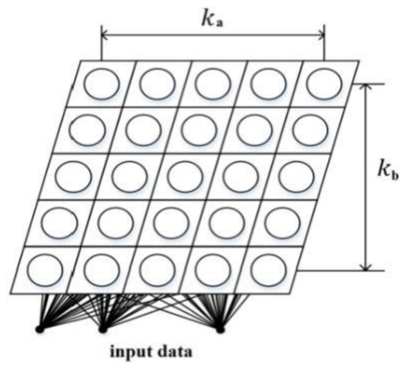

FIGURE 6. Two-dimensional SOM, rectangular topology.

country identification based on detected objects in one photo. However, it is natural as some countries are very similar, as well as some countries have very different regions or tourist activities. Therefore, the similarity distance is not enough to recommend a travel direction based on one photo but plays an important role in the final recommendation model.

Talking about the differences between similarity distance methods, the Chebychev distance provides the poorest results, because the total number of correct test data countries determination is equal to 1311.2 (TOP 10), so it is about 30-40 cases less than the results of the other distances. The highest number (1351.2) of correct determinations in TOP 10 is obtained using the city block distance. Also, the highest number of TOP 1 (448.4) and TOP 5 (974.4) determinations. Therefore, the city block distance has been incorporated into the development of the proposed recommendation model.

D. SELF-ORGANIZING MAPS

Classification and clustering algorithms are among the most commonly used methods in modern solutions. In this research, clustering is important to indicate clusters of similar photos (showing similar objects in it) photos. It was mentioned that, based on country only, the recommendation of trip direction might not be accurate, and we should be concentrating on what is common in user photos, what he or she likes. Therefore, by using photography clustering, we can estimate to which cluster the photo belongs, what other photos are similar to it, and in which countries (maybe even specific locations) it was taken.

Various clustering methods can be applied [32], such as density-based clustering, hierarchical clustering, K-means, etc. In our proposed recommendation model, we used self-organizing maps (SOMs) to draw conclusions about all the results in each part of the list of models and to recommend countries to visit. SOM is one of the artificial neural network models proposed by Kohonen [33]. The main advantage of this method is that it not only clusters the data but also shows the results in a visual form. This can be much easier to interpret by a researcher. Visual forms can be presented in various ways [34], [35], but the main objective of SOM is to preserve the topology of multidimensional data when they are transformed into a lower-dimensional space (usually two-dimensional). SOM can be applied in various fields, such as data mining [36], text mining [37], finding outlier points in data [38], multi-label text data class adjustment [39], and even in image analysis tasks [40], [41]. SOM can be used to cluster, classify, and visualize data. SOM is a set of nodes connected via a rectangular or hexagonal topology (Fig. 5).

The main difference in the SOM topology is that the neighbouring rank compared to the grey node in the SOM is determined differently. The neighbouring rank is used in the SOM training process. The set of weights forms a vector M_{ij}, $i= 1, \ldots, k_a, j= 1, \ldots, k_b$ that is usually called a neuron or codebook vector, where k_a is the number of rows, and k_b is the number of columns of SOM (Fig. 6).

The learning process of the SOM algorithm starts from the initialization of the components of the vectors M_{ij}, where they can be initialized at random, linear, or by the principal components. At each learning step, an input vector X_s is passed to the SOM. The vector X_s is compared to all neurons M_{ij}. Usually, the Euclidean distance between this input vector X_s and each neuron M_{ij} are calculated. The vector M_w with the minimal Euclidean distance to X_s is designated as a neuron winner. All neuron components are adapted according to the learning rule:

$$M_{ij}(t+1) = M_{ij}(t) + h_{ij}^{w}(X_s - M_{ij}(t)) \tag{1}$$

there t is the number of learning step, h_{ij}^{w} is a neighboring function, w is a pair of indices of the neuron winner of vector X_s, $s = 1, \ldots, N$.

The learning is repeated until the maximum number of iterations is reached. Many SOM visualization methods use coloring techniques to show the distance on the map. It shows how close the vectors of the neighbouring cells are in the dimensionality space of the analysed data. The most popular visualization technique is based on the so-called unified distance matrix (u-matrix) [42]. The SOM is colored by the values of u-matrix elements. If grayscale is used, a dark color between neurons corresponds to a large distance. A light color between the neurons signifies that the codebook vectors are close to each other in the input space. Light areas can be thought of as clusters, and dark areas can be thought of as cluster separators.

To recommend the list of travel countries, first of all, the self-organizing map has been trained using the X_s dataset. The dimensionality of dataset X_s has been reduced to 10 using the PCA [43] to improve the performance in the SOM training process. In our previous research [44], [45], the influence of SOM parameters has been investigated, so the parameters have been chosen according to the results of previous research.

The SOM size is usually chosen experimentally and depends on the size of the analyzed dataset. The primary experiments have been carried out by changing the size of the SOM from 10 to 40, by step 5 when $k_a = k_b$ in rectangular topology. An example of 10 × 10 SOM using the u-matrix is presented in Fig. 7. An Orange data mining tool has been used for the visual presentation of SOM [46]. The circles' size in the SOM indicates how much data falls in the same cell of the SOM, usually, it means that the dataset items are similar. Using the u-matrix visualization method, additional cells are inserted to show the distance between the items of the dataset. In this way, the size of the SOM is presented larger, but in the learning process the size of the SOM is 10 × 10. In addition, the neighbouring rank indicated the similarity between the datasets. The smaller the neighboring rank, the more similar the dataset items are in the SOM. Example of countries' names that fall into the one marked cell are given close to the SOM.

Primary experimental investigation finds that the optimal size of the SOM is equal to 30 × 30. This is because the data items are distributed over the number of SOM and the data items are not high in one cell of the number of SOM. Using the small size of the SOM, in the same cell of the SOM fall a few hundred data items, which makes the recommendation process not accurate.

E. COMBINED MODEL FOR TRAVEL DIRECTION RECOMMENDATION

Photo data (tag values for objects in photos) are compared to an existing photo dataset to recommend a travel direction based on a user's photo. The photos that are the most similar are selected as potential travel directions. The list is limited to a defined number of countries ($L = 10$), excluding the country where the photo was taken. By eliminating the same

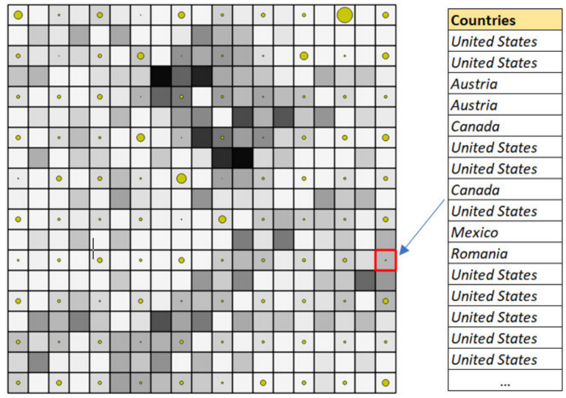

FIGURE 7. Dataset X_s presented in 10 × 10 SOM using u-matrix.

country, a variety of destinations will be presented, otherwise the same direction could be proposed as the one that the user has already travelled. With the SOM application, the list of potential countries is selected from the cluster, taking the first L countries. In addition to the list of countries, each item was accompanied by a metric, how often the country appeared in the top similar photos until there were more than 10 different countries, excluding the visited country. Additionally, this metric includes information about the country that was visited and shown in the most similar photos. In addition, it adds information on how often the country appears in the most popular photos.

Recommending a travel direction based on one photo is not effective. Users usually have multiple photos, therefore different solutions for combined recommendation were generated to estimate the most suitable solution for travel direction recommendation:

- SIM_combination. Estimating the most similar items for each photo, and then selecting the top 10 associated counties. These countries are combined into one list to identify the most popular as recommendations.
- SOM_combination. Items from the relevant SOM cluster are selected to get a list of 10 countries for each photo. These countries are combined into one list to identify the most popular as recommendations.
- SIM_common. A combined photo data set created by combining multiple photos' tags and the one item used to find the most similar photos and countries associated with them.
- SOM_common. A combined photo data set created by combining multiple photos' tags and the one item used to identify the most similar clusters and countries in it.
- Combined. A combination of the above models has been developed to reflect their different benefits.

For the models, on the basis of the combination of separate photo results, a variety of combination schemas were tested: ordering the recommended countries according to the cumulative number of repetitions for each country, sum of the repetitions for those countries which were mentioned more than r times in one photo, sum of photos in which the country was mentioned in more than two photos of the user, etc. The most effective method of combining multiple photo recommendation results was to sum the country-mentioning facts in each photo without further filtering. Therefore the recommendation is based on set of photos $p = \{pl_1, pl_2, \ldots, pl_R$, where pl_i defines a list of recommended countries for photo i, and i is each photo out of R photos, used for the recommendation. The list of potential recommended countries RC is created by sorting the list of mentioned countries and selecting TOP L countries, based on the number of mentions m_C of each country C (2).

$$m_C = \sum_{k=1}^{R} C \in pl_k \quad (2)$$

Meanwhile common photo model $Y = \{y_5, y_6, \ldots, y_n\}$ was build to represent one record with all objects, detected in all R photos, used for the travel direction recommendation. The value of each attribute y_i in the model Y is calculated by averaging the corresponding attributes in each of the analysed R photos:

$$y_i = \frac{\sum_{k=1}^{R} x_{k,i}}{R} \quad (3)$$

there i is the i-th attribute, representing the probability of the detected object in photo k. This model also had a modification, dedicated to average only those attributes that are not equal to 0.

The combined model is dedicated to recalculate the score S' of each recommended country C by multiplying each model t score $S_{t,C}$ with parameter w_t and add those scores from each model t (4). M is the number of combined models, while t is the index of the model, which is integrated into the combined model. The weighted scores should be aggregated by summing to one score, used for country ordering and TOP recommendation selection.

$$S'_C = \sum_{t=1}^{M} w_t S_{t,C} \quad (4)$$

Such a combination of multiple models into one weighted model allows for the integration of different solutions. As well the score $S_{t,C}$ can be extended variously, taking into account the rank of the recommendation in the list, the sum of country mentions, relative frequency of the country, etc.

IV. VALIDATION OF THE PROPOSED RECOMMENDATION MODELS

There was a dataset of 12460 photos [14] collected and available to compare user photos to it. However, the data were not fully associated with the user profiles. Therefore, the data could not be used for recommendation testing. The validation of the proposed model was carried out in two steps: 1) 9 people were selected and asked to share their Instagram photos, which will be used for model modeling; 2) 10 photos of each user were selected for model input, and a list of visited by the user was gathered from the photos and their answers to compare model output to model input. Each part of the validation steps and their results are presented in the following subsections.

TABLE 2. The summary of user photo data for validation.

User no.	Number of countries: In photos for input	Number of countries: Not in input photos	Average of tags in photo
1	4	2	13.0
2	3	0	13.8
3	3	0	17.5
4	1	0	14.0
5	7	0	15.0
6	5	5	14.6
7	5	2	15.4
8	6	2	16.2
9	9	1	13.5

A. DATASET AND METRICS FOR TRAVEL DIRECTION RECOMMENDATION MODEL VALIDATION

To validate the proposed travel direction recommendation models, a dataset of user photos was collected. Nine respondents shared 10 to 20 photos of travel, but not from more than 10 countries. All photos were labelled with the country in which they were taken (using the methods described in Sections III-A and III-B or manually by the user). 10 random photos from each user were selected as input data for the recommendation models. The list of all countries visited by users was compiled for comparison with the model output. The number of countries shown in the ten analyzed photos and the additional country list. In addition, the average number of detected objects in each of the 10 photos is shown in Table 2.

When a photo is analyzed for recommendation, the country where the photo was taken is removed from the recommendation list. While when common or combined models are used, no countries are eliminated. Therefore, to estimate the accuracy of the recommendation, we estimate how many countries of the 10 recommendations match the countries the user has already visited (5).

$$acc_t = \sum_{l=1}^{U} \frac{c_l \in RC_t}{U} \quad (5)$$

there acc_t is accuracy of model t, calculated by matching U user visited countries c_l (l is the index of the country) with a list of model recommendations RC_t.

This metric was selected as some of the users visited only one country, which was presented in the photos. In addition, about half of the users had not been to other countries than those shown in the photos. Additionally, the idea of asking users how they evaluate the recommendations was dropped. This is because most people would be willing to visit another country if they had time and money. In such a survey, only people who are currently looking for directions for their next trip should be included.

B. ACCURACY OF PROPOSED RECOMMENDATION MODELS

To estimate the accuracy of the proposed models acc, different parameters (data aggregation functions, weights, model combinations, etc.) were modelled, and the highest average results for the nine users are presented in Table 3. The standard deviation values indicate that accuracy is highly variable. There was a very different number of countries previously

TABLE 3. Summary of all five model accuracy results.

Model	Accuracy, %	Standard deviation, %
SIM_combination	50	29
SOM_combination	53	24
SIM_common	37	30
SOM_common	39	28
Combined	63	25

TABLE 4. The summary combined model case accuracy.

Model	Count	Rank	Frequency
SIM (City block similarity)	48%	50%	16%
SOM clustering	53%	53%	17%

TABLE 5. The summary common model case accuracy.

Model	General average	Adapted average
SIM (City block similarity)	29%	37%
SOM clustering	39%	31%

visited in the tested cases. From a research perspective, accuracy variation is not very promising. However, the test cases reflect scenarios where clients have very different photos and travel experiences, which adds additional value to the research.

In the experiments, the approach where recommendations for separate photos are generated and then combined is more accurate in comparison to the method where a common photo is generated from all photos with one common photo indicating travel directions - at least 10% better accuracy was achieved. The combined model developed is 10% more accurate than the later models. In addition, it has one of the smallest standard deviation values, which indicates that the recommendation accuracy is the most stable among all users. This high standard deviation means that there are no statistically significant differences between these models. However, by calculating the P-value as the area of the t distribution with $n-1$ degrees of freedom, that falls outside $\pm$ t and with a 95% confidence interval, we can say that the combined model can recommend travel directions to all users, which are 44% to 82% accurate.

When analyzing how many recommendations matched the countries provided as additional photos and not included in the inputs, only one country was absent from the ten recommendations for the user number 6, which included 5 additional countries. All other countries presented as additionally visited by the user were on the recommendation list for this user. This would lead to recommendation accuracy from 85% to 100% (with a 95% confidentiality interval), where the average accuracy is 96%.

The aggregation functions for combining separate photo recommendations into one were analyzed in several cases. The most interesting results are presented in Table 5. The "Count" case reflects the sum of each country included in each photo recommendation. The "Rank" case sums up the ranks of each of the most similar photos. The case 'Frequency' sums up the relative frequency of photos in each recommendation. This means that the number of photos on the top list is divided by the number of records in the dataset for the country. Comparison of the model cases indicates that "frequency" is not suitable for recommending travel directions. The count and rank aggregation functions show very similar results.

We analyzed how common photo construction affects recommendation accuracy. In the case of a general average, the sum of tag probabilities is added and divided by the number of photos. One or a few photos show the impact of a final object. Another option is an adapted average, which is calculated only among photos that contained an appropriate tag in the picture, whose object detection probability was not 0. This table presents the comparisons between accuracy of those two common cases when City block similarity and SOM based common models are provided. In both SIM and SOM models, the results show that the average calculation has the opposite effect - for SIM, the adapted average is better, while for SOM the general average is better.

In terms of the combined model, it achieved 63% accuracy for all the countries visited by the users and 96% accuracy for the countries listed in addition to the input photos. The results were obtained by combining the SIM_common and SOM_combination models. The weights were equal to 1 for both models. The SIM_common model used the adapted average, while the SOM_combination model used a "Count" metric for the country score.

V. DISCUSSION

Validation of the proposed travel recommendation model has been performed using a small amount of data collected from real people and their travel experience data to simulate a real situation in the model. To validate the model, the artificially made dataset could be used instead of the real people's data, with a larger amount, but in that case it will distort the real situation. All parts of the model have been tested in the model development process, so there is no reason to test the final travel recommendation model proposed on the large dataset. This proposed model will be used in the real information system, and the collected user experience data will be taken into account in the future for a deeper analysis.

Comparison to other travel recommendation models is very limited. None of them use the same test cases or input data, Instagram photos. For example, Xiang Huang proposes a place of interest text description and user comment comparison-based travel recommendation system. The model is able to achieve up to 88% recommendation accuracy and outperform existing analogue solutions [47]. Considering text descriptions are more representative of user opinions than photos on Instagram, our results are in the same range - we achieved 66% accuracy for all user visited locations, and 96% accuracy for recommending countries not shown in the input photos.

Meanwhile, research on travel recommendations based on purely image data has just gained popularity. One of the first attempts to use images for travel recommendations was made by M.T. Linaza et al. [13]. Images were used to classify users into four profiles, while travel directions were entered manually. Four travel profiles were estimated by object detection in photos by M. Figueredo et al. work as well. These authors estimated 10 recommended travel directions and compared them with the user's top 10 preferences. The accuracy of the recommendations was 50%-60% depending on the profile. This is similar to our achieved results; however, this model has only 4 possible profiles, not a recommended country, therefore, it cannot be directly compared with our results.

In other research papers and recommendation systems, user images are considered additional data. The recommendations are based on additional data, such as user reviews [50], user location, and moving trajectory [51], etc.

As a result of existing work in user photo adoption for travel direction recommendation, our proposed method provides a distinct approach, a country recommendation rather than a POI recommendation, and it is based entirely on user photos alone (metadata is used to eliminate countries the user has already visited from the recommended list). To compare its results with the random recommendation model, we would bet less than 1% accuracy to guess the country the user visited from 169 possible countries in the dataset. As 10 countries are provided as a recommendation and on average, each user visited 6 countries, the random model would achieve about 10% accuracy, while our model indicated 63% accuracy. Evaluating how accurate the model would be to recommend additional countries visited by the user, but not presented in the photos, the random model would achieve about 25% accuracy, as there were from 2 to 3 additional countries visited by the user. Meanwhile, our model on average achieved 96% accuracy. In both cases, our proposed model demonstrates a significant difference in comparison to the random model.

VI. CONCLUSION

In existing research articles, the recommendation of travel destinations is still relevant, but there is a lack of datasets for visited countries, locations, travel photos and objects in them. We designed a fully automated solution that gathered user Instagram photos, detected objects in the photos, and analyzed the item data to improve location accuracy. Automating this process enables us to create a data set of needed data and adjust or discretize its metadata.

In the context of the collected dataset, user photos can be analyzed and compared with the records in the dataset, calculating their similarity and membership of the cluster. By combining county data with similar photos, we can build a travel destination recommendation system. As a result of the experiments conducted, it can be concluded that different variation in the model is capable of improving accuracy. However, the most effective result is achieved by using a combination model, which is able to recommend ten countries, which corresponds to 63% of the countries the user visited. We found that accuracy was even higher, on average 96%, when we analyzed how many countries users travelled to in addition to those provided as input for the model. Based on only user photos, this is a promising result to predict travel directions more accurately.

Our proposed travel recommendation model is based on the comparison of the detected object vector with other records in our dataset. This eliminates the need for repetitive analysis of each photo. But at the same time, it becomes very sensitive to changes in the object detection model. Because of this, our dataset was not able to be used for analyzing photos processed with different object detection solutions. These solutions provide different lists of detected objects or different distributions of the probability of detecting an object. In case of object detection model changes, the whole dataset of compared photos should be revised to get updated scores of detected objects.

The current solution is designed to recommend countries. Since some countries have a wide variety of regions, the model could be updated or expanded to recommend a more specific location. This could be a region or even a specific area. The data is available in the dataset and could be used for more detailed travel planning.

REFERENCES

[1] U. Gretzel, N. Mitsche, Y.-H. Hwang, and D. R. Fesenmaier, "Tell me who you are and I will tell you where to go: Use of travel personalities in destination recommendation systems," *Inf. Technol. Tourism*, vol. 7, no. 1, pp. 3–12, Jan. 2004, doi: 10.3727/1098305042781129.

[2] P. Phorasim and L. Yu, "Movies recommendation system using collaborative filtering and k-means," *Int. J. Adv. Comput. Res.*, vol. 7, no. 29, pp. 52–59, Feb. 2017.

[3] Y. H. Hu, P. J. Lee, K. Chen, J. M. Tarn, and D. V. Dang, "Hotel recommendation system based on review and context information: A collaborative filtering appro," Presented at the Pacific Asia Conf. Inf. Syst. (PACIS), 2016. [Online]. Available: https://core.ac.uk/download/pdf/301369618.pdf

[4] L. Yu, F. Han, S. Huang, and Y. Luo, "A content-based goods image recommendation system," *Multimed. Tools Appl.*, vol. 77, no. 4, pp. 4155–4169, 2018.

[5] A. Felfernig, S. Gordea, D. Jannach, E. Teppan, and M. Zanker, "A short survey of recommendation technologies in travel and tourism," *OEGAI J.*, vol. 25, no. 7, pp. 17–22, 2007.

[6] Z. Yu, H. Xu, Z. Yang, and B. Guo, "Personalized travel package with multi-point-of-interest recommendation based on crowdsourced user footprints," *IEEE Trans. Human-Mach. Syst.*, vol. 46, no. 1, pp. 151–158, Feb. 2016.

[7] L. Tang, D. Cai, Z. Duan, J. Ma, M. Han, and H. Wang, "Discovering travel community for POI recommendation on location-based social networks," *Complexity*, vol. 2019, Feb. 2019, Art. no. 8503962.

[8] V. Subramaniyaswamy, V. Vijayakumar, R. Logesh, and V. Indragandhi, "Intelligent travel recommendation system by mining attributes from community contributed photos," *Proc. Comput. Sci.*, vol. 50, pp. 447–455, Jan. 2015.

[9] C. Bin, T. Gu, Y. Sun, L. Chang, and L. Sun, "A travel route recommendation system based on smart phones and IoT environment," *Wireless Commun. Mobile Comput.*, vol. 2019, Jul. 2019, Art. no. 7038259.

[10] C. Lu, P. Laublet, and M. Stankovic, "Travel attractions recommendation with knowledge graphs," Presented at the Eur. Knowl. Acquisition Workshop, Nov. 2016, doi: 10.1007/978-3-319-49004-5_27.

[11] I. Memon, L. Chen, A. Majid, M. Lv, I. Hussain, and G. Chen, "Travel recommendation using geo-tagged photos in social media for tourist," *Wireless Pers. Commun.*, vol. 80, no. 4, pp. 1347–1362, 2015.

[12] T. Mahmood, F. Ricci, and A. Venturini, "Improving recommendation effectiveness: Adapting a dialogue strategy in online travel planning," *Inf. Technol. Tourism*, vol. 11, no. 4, pp. 285–302, Dec. 2009.

[13] M. T. Linaza, A. Agirregoikoa, A. Garcia, J. Ignacio Torres, and K. Aranburu, "Image-based travel recommender system for small tourist destinations," Presented at the Tourism, Jan. 2011, doi: 10.1007/978-3-7091-0503-0_1.

[14] Kaggle. *Travel Dataset*. Accessed: Dec. 29, 2022. [Online]. Available: https://www.kaggle.com/datasets/pavelstefanovi/travel-dataset

[15] P. Thiengburanathum, S. Cang, and H. Yu, "Overview of personalized travel recommendation systems," Presented at the 22nd Int. Conf. Automat. Comput. (ICAC), Sep. 2016. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7604955
[16] Y. Sun, H. Fan, M. Bakillah, and A. Zipf, "Road-based travel recommendation using geo-tagged images," *Comput. Environ. Urban Syst.*, vol. 53, pp. 110–122, Sep. 2015.
[17] K. Saleh, S. Szénási, and Z. Vámossy, "Occlusion handling in generic object detection: A review," Presented at the IEEE 19th World Symp. Appl. Mach. Intell. Inform. (SAMI), Jan. 2021. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9378657
[18] Y. Zheng, R. Burke, and B. Mobasher, "Differential context relaxation for context-aware travel recommendation," Presented at the Int. Conf. Electron. Commerce Web Technol., Sep. 2012, doi: 10.1007/978-3-642-32273-0_8.
[19] J. An, S. Zhao, X. Lu, and N. Liu, "A two-stage multiple-factor aware method for travel product recommendation," *Multimedia Tools Appl.*, vol. 77, no. 21, pp. 28991–29012, 2018.
[20] L. Ravi and S. Vairavasundaram, "A collaborative location based travel recommendation system through enhanced rating prediction for the group of users," *Comput. Intell. Neurosci.*, vol. 2016, Mar. 2016, Art. no. 1291358.
[21] Y. Fang, K. Kuan, J. Lin, C. Tan, and V. Chandrasekhar, "Object detection meets knowledge graphs," Presented in the 26th Int. Joint Conf. Artif. Intell., Aug. 2017. [Online]. Available: https://core.ac.uk/download/pdf/200253749.pdf
[22] A. A. Amer, H. I. Abdalla, and L. Nguyen, "Enhancing recommendation systems performance using highly-effective similarity measures," *Knowl.-Based Syst.*, vol. 217, Apr. 2021, Art. no. 106842.
[23] A. Gazdar and L. Hidri, "A new similarity measure for collaborative filtering based recommender systems," *Knowl.-Based Syst.*, vol. 188, Jan. 2020, Art. no. 105058.
[24] B. Hawashin, M. Lafi, T. Kanan, and A. Mansour, "An efficient hybrid similarity measure based on user interests for recommender systems," *Expert Syst.*, vol. 37, no. 5, pp. 1–18, Oct. 2020.
[25] P. Thiengburanathum, S. Cang, and H. Yu, "A decision tree based recommendation system for tourists," Presented in the 21st Int. Conf. Automat. Comput. (ICAC), Sep. 2015. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7313958
[26] Y. Zuo, J. Zeng, M. Gong, and L. Jiao, "Tag-aware recommender systems based on deep neural networks," *Neurocomputing*, vol. 204, pp. 51–60, Sep. 2016.
[27] R. Ahuja, A. Solanki, and A. Nayyar, "Movie recommender system using k-means clustering and k-nearest neighbor," Presented in the 9th Int. Conf. Cloud Comput., Data Sci. Eng. (Confluence), Jan. 2019. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8776969
[28] GeoPy's. *Geocoding Web Services*. Accessed: Feb. 8, 2023. [Online]. Available: https://geopy.readthedocs.io/en/stable/
[29] A. Singh, N. Thakur, and A. Sharma, "A review of supervised machine learning algorithms," in *Proc. 3rd Int. Conf. Comput. Sustain. Global Develop. (INDIACom)*, Mar. 2016, pp. 1310–1315. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7724478
[30] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, "Supervised machine learning: A review of classification techniques," *Emerg. Artif. Intell. Appl. Comput. Eng.*, vol. 160, no. 1, pp. 3–24, 2007.
[31] P. C. Sen, M. Hajra, and M. Ghosh, "Supervised classification algorithms in machine learning: A survey and review," Presented in the IEM Graph, Sep. 2018, doi: 10.1007/978-981-13-7403-6_11.
[32] C. C. Aggarwal and C. Zhai, "A survey of text clustering algorithms," in *Mining Text Data*. New York, NY, USA: Springer, 2012, pp. 77–128.
[33] T. Kohonen, *Self-Organizing Maps*, vol. 30. Berlin, Germany: Springer, 2012.
[34] P. Stefanovič and O. Kurasova, "Visual analysis of self-organizing maps," *Nonlinear Anal., Model. Control*, vol. 16, no. 4, pp. 488–504, Dec. 2011.
[35] G. Dzemyda and O. Kurasova, "Comparative analysis of the graphical result presentation in the SOM software," *Informatica*, vol. 13, no. 3, pp. 275–286, 2002.
[36] A. Urueña López, F. Mateo, J. Navío-Marco, J. M. Martínez-Martínez, J. Gómez-Sanchís, J. Vila-Francés, and A. José Serrano-López, "Analysis of computer user behavior, security incidents and fraud using self-organizing maps," *Comput. Secur.*, vol. 83, pp. 38–51, Jun. 2019.
[37] K. Yoshioka and H. Dozono, "The classification of the documents based on Word2 Vec and 2-layer self organizing maps," *Int. J. Mach. Learn. Comput.*, vol. 8, no. 3, pp. 252–255, Jun. 2018.
[38] P. Stefanovič and O. Kurasova, "Outlier detection in self-organizing maps and their quality estimation," *Neural Netw. World*, vol. 28, no. 2, pp. 105–117, 2018.
[39] P. Stefanovič and O. Kurasova, "Approach for multi-label text data class verification and adjustment based on self-organizing map and latent semantic analysis," *Informatica*, vol. 33, no. 1, pp. 109–130, 2022.
[40] S. Licen, A. Di Gilio, J. Palmisani, and S. Petraccone, "Pattern recognition and anomaly detection by self-organizing maps in a multi month e-nose survey at an industrial site," *Sensors*, vol. 20, no. 7, p. 1887, 2020.
[41] S. Aly and S. Almotairi, "Deep convolutional self-organizing map network for robust handwritten digit recognition," *IEEE Access*, vol. 8, pp. 107035–107045, 2020.
[42] A. Ultsch and H. P. Siemon, "Exploratory data analysis: Using Kohonen networks on transputers," Univ. Dortmund, Dortmund, Germany, Tech. Rep. 329, 1989.
[43] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 2, nos. 1–3, pp. 37–52, Aug. 1987.
[44] P. Stefanovič and O. Kurasova, "Influence of learning rates and neighboring functions on self-organizing maps," Presented at the Int. Workshop Self-Organizing Maps. Berlin, Germany: Springer, Jun. 2011.
[45] P. Stefanovič and O. Kurasova, "Investigation on learning parameters of self-organizing maps," *Baltic J. Mod. Comput.*, vol. 2, no. 2, pp. 45–55, 2014.
[46] J. Demšar, T. Curk, A. Erjavec, C. Gorup, T. Hočevar, M. Milutinovič, M. Možina, M. Polajnar, M. Toplak, A. Starič, M. Štajdohar, L. Umek, L. Žagar, J. Žbontar, M. Žitnik, and B. Župan, "Orange: Data mining toolbox in Python," *J. Mach. Learn. Res.*, vol. 14, no. 1, pp. 2349–2353, 2013.
[47] X. Huang, "Personalized travel route recommendation model of intelligent service robot using deep learning in big data environment," *J. Robot.*, vol. 2022, pp. 1–8, Jan. 2022.
[48] M. Figueredo, J. Ribeiro, and N. Cacho, "From photos to travel itinerary: A tourism recommender system for smart tourism destination," Presented at the IEEE 4th Int. Conf. Big Data Comput. Service Appl. (BigDataService), Mar. 2018.
[49] J. Díez, P. Pérez-Núnez, O. Luaces, and B. Remeseiro, "Towards explainable personalized recommendations by learning from users' photos," *Inf. Sci.*, vol. 520, pp. 416–430, May 2020.
[50] Y. Zheng, X. Xu, and L. Qi, "Deep CNN-assisted personalized recommendation over big data for mobile wireless networks," *Wireless Commun. Mobile Comput.*, vol. 2019, pp. 1–16, Apr. 2019.

PAVEL STEFANOVIČ received the Ph.D. degree in computer science from the Institute of Mathematics and Informatics, Vilnius University, Lithuania, in 2015. He is currently a Researcher and an Associate Professor with the Faculty of Fundamental Sciences, Vilnius Gediminas Technical University. He is the author of 16 publications. His research interests include data mining methods, natural language pre-processing, machine learning methods, the visualization of multidimensional data, and data clustering methods.

SIMONA RAMANAUSKAITĖ was born in December 1983. She received the M.Sc. degree in informatics, in 2008, and the Ph.D. degree in informatics engineering from Vilnius Gediminas Technical University, in 2012. She is currently a Professor and a Senior Researcher with the Department of Information Technologies, Vilnius Gediminas Technical University. She has published more than 40 papers in scientific journals and conference materials. Her current research interests include different aspects of information security, e-learning, and artificial intelligence.