Data Mining Techniques Applied to Recommender Systems for Outdoor Activities: A Systematic Literature Review

Pablo Arévalo[a], John Calle[b], Marcos Orellana[c] and Priscila Cedillo[d]
Laboratorio de Investigación y Desarrollo en Informática - LIDI,
Universidad del Azuay, 24 de mayo y Hernán Malo, Cuenca, Ecuador

Keywords: Recommending Systems, Health, Air Quality, Data Mining, Air Pollutants.

Abstract: Currently, many pollutants are released into the air, representing a risk to the environment and human health. There are significant volumes of data generated by the devices that monitor these pollutants. This information can represent a relevant input that allows the construction of applications, techniques, and methodologies to reach a prediction of the state of the air. On the other hand, recommender systems are present in numerous data processing methods, supporting the decision-making and promoting the improvement of the quality of service of solutions. Although several studies have been presented, no secondary studies have been proposed. Therefore, this paper presents a systematic review of the literature, which aims to identify the knowledge areas, tools, methods, and data mining approaches used in recommender systems for outdoor activities related to atmospheric pollutants. The results obtained contribute to creating new ways of recommendation systems based on the previous topics.

1 INTRODUCTION

Currently, everybody is exposed to high levels of environmental pollution. This situation causes a significant impact on people's health and increases the risk of suffering from different types of diseases directly related to pollution of the environment. Moreover, poor air quality affects a more significant proportion of the most vulnerable population, such as children and the elderly (Singla, 2018) (An et al., 2018). In addition, with the rise of technology use in recent decades, solutions generate large volumes of data; this situation allows the construction of models, techniques, or methodologies (Singla, 2018). For this reason, being aware of the current air quality when carrying out any outdoor activity is of vital importance to prevent possible damage to health.

Recommender systems today have become essential when choosing a product or service. Currently, there are recommender systems in various fields and industries to be found in any system; these solutions are of substantial help in decision-making and a boost to the quality of service (Ricci et al., 2011). However, lately, studies related to secondary studies have not been reported that support researchers in discovering findings related to these topics. Therefore, the evidence is scattered and difficult to find, being necessary to join the most effective approaches summarized in a unique study.

Then, this paper presents a Systematic Literature Review (SLR) delving into the issues related to air quality and its impact on health and the implementation of recommender systems. This SLR follows the methodology presented by Kitchenham & Charters (2007) and answers the following research questions: RQ1. What kind of information is required to develop an outdoor recommender? RQ2. What methodologies are used to address the development of an outdoor recommender? RQ3. How are outdoor recommenders addressed in data science? And RQ4. How has the research on recommender systems for outdoor activities been carried out?

Finally, this paper is structured as follows: Section 2 presents related work, Section 3 develops

[a] https://orcid.org/0000-0002-6085-541X
[b] https://orcid.org/0000-0003-0299-9279
[c] https://orcid.org/0000-0002-3671-9362
[d] https://orcid.org/0000-0002-6787-0655

Arévalo, P., Calle, J., Orellana, M. and Cedillo, P.
Data Mining Techniques Applied to Recommender Systems for Outdoor Activities: A Systematic Literature Review.
DOI: 10.5220/0011045400003188
In *Proceedings of the 8th International Conference on Information and Communication Technologies for Ageing Well and e-Health (ICT4AWE 2022)*, pages 228-235
ISBN: 978-989-758-566-1; ISSN: 2184-4984

the SLR methodology, Section 4 analyzes the results of the SLR execution, Section 5 presents a discussion, Section 6 analyzes the threads of validity, and Section 7 presents the conclusions and the next steps of the research.

2 RELATED WORK

Air pollution places a heavy burden on human health, and understanding the effects of pollutants is a constant challenge for our society. People die due to diseases induced by air pollution, such as ischemic heart disease, lung cancer, among others (Singla, 2018) and (An et al., 2018). There are various sources of environmental pollution. In urban areas, vehicular traffic is the most predominant source of emissions, mainly composed of exhaust emissions of carbon monoxide, nitrogen oxides, and suspended particles from vehicles in megacities (Suresh et al., 2015).

With this problem, systematic reviews have been carried out that have been interested in air quality in different areas. For example, seeking to determine the best statistical model based on machine learning techniques to capture the non-linear relationship between the concentration of air pollutants and their emission and dispersion sources (Rybarczyk & Zalakeviciute, 2018). In another case, the scientific evidence linking air pollution to physical activity in China was systematically reviewed, showing that it affects behaviours related to daily physical activity in residents (An et al., 2019). In another example, the objective is to investigate the applications of deep learning in the forecast of air quality in time series, demonstrating that the accuracy of the results is greater than that of individual models (Zaini et al., 2021). In the same way, another study provides a valuable synthesis of the relevant literature on smart cities by analyzing and discussing the key findings in creating sustainable cities and communities considering environmental control and air quality (Ismagiloiva et al., 2019).

Additionally, considering the particulate matter, related studies have been evaluated in order to assess the impact on health in terms of possible reductions in premature deaths due to the reduction of this pollutant (Jahn et al., 2011).

On the other hand, recommender systems are techniques or software tools that provide suggestions for elements, which can be useful to a user. These suggestions or recommendations are related to decision-making processes normally focused on a specific topic (Ricci et al., 2011). Therefore, these recommenders represent a substantial aid in decision-making and an impetus to improve the quality of services. For example, recommender systems have been developed focused on educational data mining in order to predict student performance (Thai-Nghe et al., 2010) or focused on text mining (Betancourt & Ilarri, 2020). Therefore, having adequate information at the right time, especially when the individual is exposed to a lesser amount of pollution, is a substantial issue to protect their health and achieve a better quality of life.

3 RESEARCH METHOD

A systematic review is a research method for obtaining, evaluating, and interpreting information related to a specific research question or area of interest. Its objective is to facilitate an objective evaluation of a research topic in a reliable, rigorous, and methodological way. For the process, the study carried out by Kitchenham, who provides a methodology to carry out systematic reviews, it was considered as reference (Kitchenham & Charters, 2007) due to it is mainly focused on three stages:

1) Planning the review: The review needs to be identified, the research questions are specified, and the review protocol is defined.

2) Conducting the review: the primary studies are selected, the quality assessment used to include the studies is defined, the data is extracted and monitored, and the data are synthesized.

3) Reporting the review: the dissemination mechanisms are specified, and the review report is presented.

According to the research carried out on recommender systems focused on the domain of the human health, in physical activity specifically, there is evidence of intervention of these systems in 2012 (Knoch et al., 2012). Therefore, this study considers articles retrieved from digital databases between 2012 and 2021.

3.1 Research Questions

Research questions are part of the support that manage the entire research process as it allows the relevant data to be determined and transformed into a research contribution. They should be formulated in four sections: population, intervention, comparison, and result (Kitchenham & Charters, 2007); this to carry out a complete examination of the variation in the study factors and between populations, for which it was necessary to relate three variables, the use of recommender systems, development of the physical

activity, and the effect that it causes of the quality of air when carrying out any type of activity. The research questions that were asked are:

- RQ1. What kind of information is required to develop an outdoor recommender?
- RQ2. What methodologies are used to address the development of an outdoor recommender?
- RQ3. How are outdoor recommenders addressed in data science?
- RQ4. How has the investigation of recommender systems for outdoor activities developed?

3.2 Search Process

The digital libraries and indexers selected for this study were chosen because they cover a large number of articles related to recommender systems, which are detailed below: Digital library ACM, ScienceDirect, SpringerLink, Scopus, IEEEXplore, Taylor and Francis, EBSCO, Web of Science and Hinari (OARE)

The search string was then developed using concise words, relevant terms, and alternative terms that emerged from the research questions. Moreover, the relevant terms defined for this study were: “Recommender”, “Data mining”, “Activities”, “Weather”, “Pollution”, and “Air”. Also, a set of alternative terms was developed consisting of: “Big data”, “Data Science”, “Physical activities”, “Sport”, “Fitness”, “healthy”, “Air Pollution” and “Air quality” to complement the relevant terms.

By using connectors “AND” and “OR” various attempts were made with combinations of these terms, adapting the search string to each library or index, the better results were obtained with the following combination:

Abstract: recommend* AND (data mining OR big data OR data science) AND All Metadata: (activities OR physical activi* OR sport OR fitness OR health*) OR (weather OR pollu* OR air*).

3.3 Exclusion and Inclusion Criteria

The protocol, inclusion, and exclusion criteria must be developed. This strategy reduces the number of selected primary studies. In this study, the exclusion criteria help eliminate studies that matched at least one of the following:

- Duplicate publications that have reference to the same study in several digital libraries.
- Short publications of less than five pages.
- Publications that are not in English.
- Gray literature (They do not have a digital object identifier, also called DOI for its acronym in English).
- Publications made before 2012.

The inclusion criteria to select works to meet the following items:

- Publications that address issues related to air quality.
- Publications that address issues related to recommender systems.
- Publications that relate health, air quality, and outdoor activities.
- Publications that implement methodologies to develop recommender systems.
- Scientific articles, conferences, books.
- Recommenders that integrate data mining techniques.

3.4 Quality Assessment

It is necessary to provide individual quality control to each study. This information should be included when answering the research questions posed. Table 1 shows a list of quality criteria. Additionally, each question was answered, dividing them into sub-questions identified by the prefix EC ##. The studies that passed the inclusion and exclusion criteria were identified by a code that has the following format: [A + sequential number Author Title].

The information collected per study had the following characteristics: name of the library, the title of the article, author(s), DOI, year of publication, number of pages, and the number of times it was cited, the latter using the search tool of articles provided by Google Scholar

The most important characteristics were obtained through the support of data mining experts and an analysis of the literature, which will allow to answer the research sub-questions. In the full reading stage of the articles, each of them will be scored with a zero or one, based on the presence or absence of that characteristic within the article. This strategy ensures that the same data extraction criteria are applied to each article.

To carry out this process, a matrix was designed in which both the score of each article – based on the sub-questions – and the evaluation of its quality were recorded. Subsequently, the bubble diagrams used in the reporting stage were obtained based on this same matrix. Each research question with its respective extraction criteria, the options available to each, and a summary of the data collection process are in the following url: https://bit.ly/3HRujm5.

4 RESULTS REPORT

It is necessary to externalize the results obtained when conducting this study. At this stage, primary studies were identified, selected, and evaluated using the exclusion, inclusion, and quality criteria defined above.

The list of selected articles is in the following url: https://bit.ly/3HRujm5. The procedure was divided into the following stages:

- Systematic search: The search string was adapted according to each library or index in this activity. Subsequently, the searches and downloads of the resulting articles were carried out, obtaining 3,417 studies.
- First selection: Since certain libraries and indexers have a high number of resulting works, a reading of the title of each of the articles was previously carried out to discard works that do not have a relationship to the topic and scope of this research. In this step, 2,337 articles were discarded.
- Second selection: At this stage, of each work obtained, the title, abstract, and keywords were read, and later it was graded with "Accepted", "Undecided", and “Rejected” based on the inclusion and exclusion criteria. Only articles that in their entirety had a rating of “Accepted” by all investigators were considered for the next stage, resulting in 108 studies.
- Third selection: In this step, the disagreements and doubts in the selection of certain articles collected in the previous step were resolved, by means of a consensus among all the researchers after the total reading of each article; the same ones that were identified using coding, mentioned before. At this stage, the primary studies were reduced to 56.
- Quality evaluation: As a final step, from the resulting articles, those that met the quality criteria were selected, resulting in 44 primary studies.

The quality of the studies was evaluated by applying the criteria presented above. The results are in the following url: https://bit.ly/3HRujm5, where the percentage of studies that answered each research question is shown. Articles EC03, EC08, EC10, and EC15 have the highest scores.

5 DISCUSSION

The intention of this study is to show the areas of interest, tools, and data mining methods that have been used in recommender systems for outdoor activities based on atmospheric pollutants. After the systematic review, and based on the articles analyzed, it can be observed that the main area of study of the recommender systems lies in health, with the main goal of improving the quality of life of people regardless of the input data. These include, for example, IoT events [A02], patient data [A04] or various groups of contaminants [A33].

For this, they have different methods in data mining to issue recommendations, focusing mainly on classification techniques [A08] and clustering [A16]. Also, they consider the users’ knowledge, whether this is the content that the user develops when using the system (i.e., source of content knowledge) or generalized information from all users (i.e., source of social knowledge).

5.1 Information Required to Develop an Outdoor Activities Recommender

The results obtained after the review indicate that the required information is based on physical activities and human health [A04, A05, A07, A08, A14, A21, A28, A29, A30, A31, A32, A35, A39] [A01, A02, A04, A05, A06, A07, A08, A09, A10, A11, A12, A13, A14, A15, A16, A18, A19, A23, A24, A25, A26, A28, A29, A30, A31, A32, A33, A35, A36, A37, A38, A39, A41].

Here, the final objective of the recommendations seeks to reduce the possible damage to health and an improvement in the quality of life. The data sources of those solutions are databases on diabetes to predict this disease [A10]. Information collected on diseases and their respective symptoms to implement a personalized system of recommendation of prevention techniques [A11]; data regarding lifestyle habits and patterns [A41] or sensory data from IoT devices to recommend interventions to promote an active lifestyle [A29].

Similarly, the articles related to health are directly related to environmental pollution [A03, A05, A16, A17, A28, A33, A37, A40]. Although the input data comprises various pollution particles (e.g., pollutants per particle, ozone, carbon monoxide, sulphur oxides, nitrogen oxides), many studies share a central goal: to prevent complications to human well-being. For example, they are the vehicular mitigation pollution

[A16] or recommended pedestrian routes that minimize the time of exposure to allergens [A28].

Consequently, before developing a recommendation system based on pollutants, it is necessary to consider the benefit or positive impact that can be provided to human well-being. A recommender to determine the best times for outdoor activities fully covers this objective to improve the quality of life.

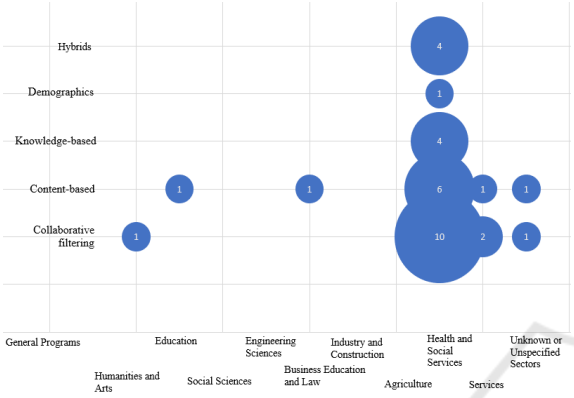

Figure 1: Study areas with the various approaches

It should be emphasized that it is necessary to know the main approaches (algorithms) involved in health issues when trying building a recommender system. In Figure. 1, the influence of the health area on collaborative and content-based filtering approaches can be seen as they are the most required. [A01, A06, A08, A14, A15, A18, A29, A35, A36, A38] [A01, A15, A18, A24, A35, A38].

A recommender with collaborative filtering looks for the system to issue recommendations taking into account the users' interactions in the past. For example, consider that if users shared similar preferences long ago, they might have equivalent tastes (Yu, 2018). Then, through various methods, citing some: frequency, weighting, or similarity of the cosine, it is possible to obtain a measure that indicates this similarity among users, generally based on the available ratings (Yu, 2018).

On the other hand, in a content-based approach, the system learns to recommend elements similar to those that interested the user long ago. The similarity of these elements is calculated based on the characteristics associated when comparing them. In other words, each element has a table with its main attributes. When issuing a recommendation, a search is done for similar attributes between the profile of the user and the items they liked in the past (Ricci et al., 2011) (Yu, 2018).

In collaborative filtering, other additional factors such as demographics can also be considered by taking additional attributes (e.g., gender, age, location) to issue recommendations demonstrating its usefulness in health by extracting more information from patients.

5.2 Methodologies Used in Addressing the Development of an Outdoor Activities' Recommender

The extraction criteria ranging from EC05 to EC08 are taken into account to answer the RQ2 research question. As mentioned above, collaborative filtering and content filtering approaches are the most frequently considered algorithms that are most frequently considered in the development of recommenders. However, it is also necessary to consider the source and type of knowledge used in the system.

The results indicate that a large part of the primary studies uses different sources of knowledge when addressing the construction of recommender systems. [A01, A02, A03, A04, A05, A06, A07, A08, A09, A10, A11, A12, A13, A15, A16, A17, A18, A19, A20, A21, A22, A23, A24, A25, A26, A28, A29, A30, A31, A33, A34, A35, A36, A37, A39, A40, A41, A42, A43, A44].

The sources of knowledge include the understanding that one has about the target user to issue personalized recommendations. The knowledge can be divided into three sections: the knowledge that is available from the target user (Individual), the knowledge about the characteristics of the article to recommend (Content), and knowledge about the broader community, including the target user (Social) (Ricci et al., 2011).

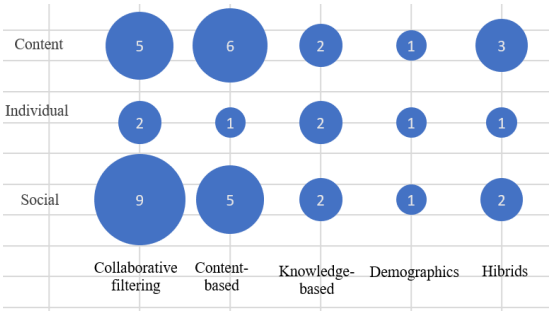

Figure 2: Sources of knowledge used in the various approaches.

As shown in Figure 2, a collaborative filtering approach uses social knowledge represented by the opinions or behaviour of a user community. In other

cases, it is possible that there is only knowledge about the characteristics, uses, or domains of the elements that are recommended, which allows us to infer what particularities attract users. In this case, a focus on content is a better option.

From a perspective focused on knowledge sources, providing a hybrid recommendation is mixing a social source with a content source by adapting the necessary algorithms to accept a knowledge source more typically associated with another type (Ricci et al., 2011). Being clear about the base content of the system helps to determine the algorithms that best adapt to the sources, considering the domain in which the recommendations will be issued.

5.3 Data Mining Techniques Used in Developing an Outdoor Recommender

The use of data mining techniques, to aid recommendation systems in their goal to learn the correct user profiles, is an active area of research (Alabdulrahman et al., 2018). The primary studies analyzed focus on using neural networks as described in Figure 3, where particulate pollutants such as carbon monoxide are targeted. These pollutants are the most common in most studies as the main component for generating predictions or recommendations on air quality.

The data found in the literature analysis are primarily concentrated on neural network techniques since it is identified as one of the most efficient processing of the collected data [A09]. Thus, many primary studies address particular pollutants, like carbon monoxide, due to their more significant impact on people's health and greater environmental concentration.

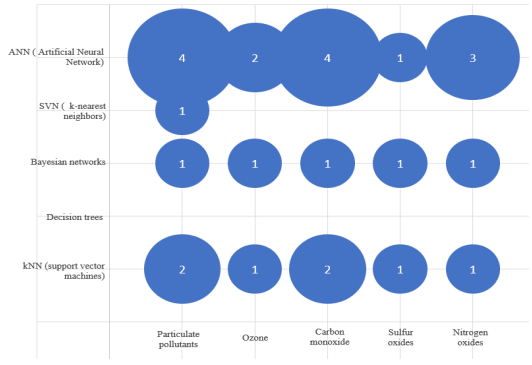

Figure 3: Classification variables in air quality.

The studies dealing with Artificial Neural Network (ANN) and the criteria regarding air quality as pollutants per particle and carbon monoxide are: [A03, A05, A06, A08, A09, A11, A12, A16, A17, A21, A27, A28, A33, A34, A37, A38, A40, A43].

Next, Figure 4 shows a high concentration of primary studies that focus on Bayesian networks as the main method to use in collaborative and content filtering approaches. Bayesian networks are helpful in domains with high stability in user preferences, meaning user preference changes slowly concerning the time required to build the model. The articles that refer to the Bayesian network classification method are: [A01, A08, A10, A13, A23, A24, A27, A28, A29, A34, A42, A43].

The trend of using data mining techniques occurs because it allows exploring the relationship between elements based on how users have rated them (Ricci et al., 2011). However, these techniques need to compare each user with others, which is not practical with huge data sets as well as requiring a class tag and, in many applications, such tags are not available, leading to domain wide expert tagging. They also require a class label, and in many applications such labels are not available, leading to domain-wide expert labelling (Alabdulrahman et al., 2018). So, it can be beneficial to perform a dimensionality reduction that, although they require an extensive offline calculation, the results scale much better (Ricci et al., 2011).

It is essential to consider that clustering techniques can identify groups with similar characteristics. Once these groups have been discovered, it is possible to make predictions taking these characteristics into account. Although these grouping methods have less precision than the ranking methods, they can be applied as a preliminary step to reduce the number of candidates or distribute them to different recommendation engines. A pre-grouping can be a valuable trade-off between accuracy and performance (Ricci et al., 2011).

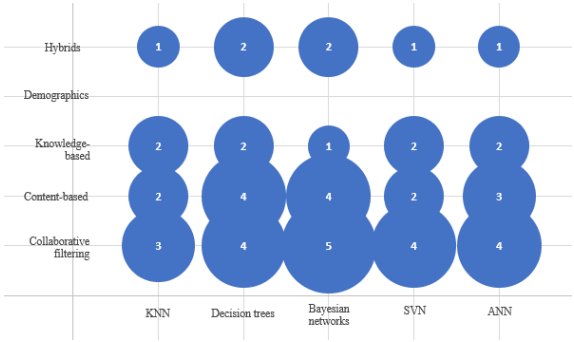

Figure 4: Most used classification methods in the various approaches.

5.4 Research Developed in Recommender Systems for Outdoor Activities

The criteria EC13, EC14, and EC15 are considered to answer this question. A large majority of studies have been developed in the field of academic research [A01, A02, A03, A04, A05, A06, A07, A08, A09, A10, A11, A12, A13, A14, A16, A18, A19, A20, A21, A22, A23, A24, A25, A26, A27, A28, A29, A30, A31, A33, A34, A36, A37, A38, A39, A40, A41 A43, A44]. The results also indicate that evaluations are developed through experiments and case studies [A01, A02, A04, A06, A07, A08, A09, A11, A12, A13, A15, A16, A17, A18, A19, A21, A24, A26, A28, A32, A34, A35, A41, A42, A44] [A03, A05, A29, A31, A33, A37, A40].

In [A44], a group recommendation system has been made to explore network document resources using the knowledge graph and Long short-term memory (LSTM), carrying out experimentation of the system in the field of Big Data applications in the packaging industry.

A case study based on a national level project focused on the pre-processing and analysis of the data collection of a city was developed in order to provide the public with a guide by establishing a Big Data platform for fine dust and provide administrative guidance to public institutions and local governments and inform the integrated indoor and outdoor air management service [A37].

6 THREATS TO VALIDITY

6.1 Lack of Important Primary Studies

The primary studies were obtained from 9 sources between libraries and indexes. Although the selected databases are compliant for this type of study, the possibility of a small group of primary studies being eliminated is not ruled out. In the same way, when building the search chain, certain words can be chosen incorrectly, obtaining wrong results.

A pre-evaluation was considered when determining the search chain to observe the results obtained to mitigate this threat. In the beginning, the chain lacked words like "Sport" or "Health", omitting a considerable number of results. Later, these words were included for a second search. Then all the connectors used in the chain were of type "AND". These actions caused a shortage of results, obtaining one or in specific libraries no results.

Finally, changing the AND connectors that joined the secondary terms for connectors of the OR type was considered, solving these problems. Although many studies were obtained, one more stage was included when choosing the primary articles.

6.2 Reliability of the Selection

In each of the previously defined selection stages, primary articles related to the proposed topic may also be excluded. Each researcher scored Yes, No, and Review each article to mitigate this threat. The objective is to reach a consensus among all researchers on specific issues that are not clear. Then, there is no doubt about the selected options. This procedure was carried out at each selection stage to prevent the elimination of articles relevant to the proposed investigation.

6.3 Data Extraction

A threat during data extraction can arise due to misunderstanding or disagreement between the reviewers. So then, five papers were randomly selected, and each of the researchers involved issued their respective interpretations of each one of them. Subsequently, the Fleiss 'kappa calculation was performed with a resulting value of 0.65.

According to Landis and Koch (Landis & Koch, 1977), values between 0.61 and 0.80 are interpreted as substantial agreement. Many studies did not provide clear answers to the extraction as mentioned above criteria, so this obtained result provides a good coefficient of agreement.

7 CONCLUSIONS

The systematic review process has satisfactorily complied with the proposed validations and evaluations. The agreement for the selection of studies and the understanding of the extraction criteria applied later in the reading of said studies was verified and improved. In addition, a reasonable assessment of the quality of the articles was obtained using a quantified score. Likewise, there is an acceptable validity in all stages of the systematic review, so it can be deduced that the planning of said review was adequate

As the main conclusion, it has been determined that currently, the recommender systems use collaborative filtering to make recommendations, whatever the case. However, there is very little

research and applications that use a knowledge-based approach. Therefore, it would be optimal for this type of application; that is, it does not need or depend so much on the information provided by the system's users, but rather that the system can provide recommendations based on the history that has been saved in the system.

ACKNOWLEDGEMENTS

The authors wish to thank the Vice-Rector for Research of the University of Azuay for the financial and academic support and all the staff of the School of Computer Science Engineering and the Laboratory for Research and Development in Informatics (LIDI).

REFERENCES

Alabdulrahman, R., Viktor, H., & Paquet, E. (2018). Beyond k-NN: Combining cluster analysis and classification for recommender systems. *IC3K 2018 - Proceedings of the 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, *1*(Ic3k), 82–91. https://doi.org/10.5220/0006931200820091

An, R., Shen, J., Ying, B., Tainio, M., Andersen, Z. J., & de Nazelle, A. (2019). Impact of ambient air pollution on physical activity and sedentary behavior in China: A systematic review. *Environmental Research*, *176*, 108545. https://doi.org/10.1016/j.envres.2019.108545

An, R., Zhang, S., Ji, M., & Guan, C. (2018). Impact of ambient air pollution on physical activity among adults: a systematic review and meta-analysis. *Perspectives in Public Health*, *138*(2), 111–121. https://doi.org/10.1177/1757913917726567

Betancourt, Y., & Ilarri, S. (2020). Use of text mining techniques for recommender systems. *ICEIS 2020 - Proceedings of the 22nd International Conference on Enterprise Information Systems*, *1*(Iceis), 780–787. https://doi.org/10.5220/0009576507800787

Ismagiloiva, E., Hughes, L., Rana, N., & Dwivedi, Y. (2019). Role of Smart Cities in Creating Sustainable Cities and Communities: A Systematic Literature Review. In Y. Dwivedi, E. Ayaburi, R. Boateng, & J. Effah (Eds.), *ICT Unbounded, Social Impact of Bright ICT Adoption* (pp. 311–324). Springer International Publishing.

Jahn, H. J., Schneider, A., Breitner, S., Eißner, R., Wendisch, M., & Krämer, A. (2011). Particulate matter pollution in the megacities of the Pearl River Delta, China – A systematic literature review and health risk assessment. *International Journal of Hygiene and Environmental Health*, *214*(4), 281–295. https://doi.org/10.1016/j.ijheh.2011.05.008

Kitchenham, B. A., & Charters, S. (2007). *Guidelines for performing Systematic Literature Reviews in Software Engineering. EBSE Technical Report EBSE-2007-01. School of Computer Science and Mathematics, Keele University*. *January*, 2007.

Knoch, S., Chapko, A., Emrich, A., Werth, D., & Loos, P. (2012). A context-aware running route recommender learning from user histories using artificial neural networks. *Proceedings - International Workshop on Database and Expert Systems Applications, DEXA*, 106–110. https://doi.org/10.1109/DEXA.2012.49

Landis, J. R., & Koch, G. G. (1977). Landis amd Koch1977_agreement of categorical data. *Biometrics*, *33*(1), 159–174.

Ricci, F., Rokach, L., Shapira, B., & Kantor, P. B. (2011). Recommender Systems Handbook. In *Journal of Physics A: Mathematical and Theoretical* (Vol. 44, Issue 8). https://doi.org/10.1088/1751-8113/44/8/085201

Rybarczyk, Y., & Zalakeviciute, R. (2018). Machine Learning Approaches for Outdoor Air Quality Modelling: A Systematic Review. *Applied Sciences*, *8*(12). https://doi.org/10.3390/app8122570

Singla, S. (2018). Air ality friendly route recommendation system. *PhD Forum 2018 - Proceedings of the 2018 Workshop on MobiSys 2018 Ph.D. Forum, Part of MobiSys 2018*, 9–10. https://doi.org/10.1145/3212711.3212717

Suresh, L. P., Dash, S. S., & Panigrahi, B. K. (2015). A16 - A Bi-level clustering analysis for studying about the sources of vehicular pollution in Chennai. In *Advances in Intelligent Systems and Computing* (Vol. 325). https://doi.org/10.1007/978-81-322-2135-7

Thai-Nghe, N., Drumond, L., Krohn-Grimberghe, A., & Schmidt-Thieme, L. (2010). Recommender system for predicting student performance. *Procedia Computer Science*, *1*(2), 2811–2819. https://doi.org/10.1016/j.procs.2010.08.006

Yu, L. (2018). A35 - Cloud storage–based personalized sports activity management in Internet plus O2O sports community. *Concurrency Computation*, *30*(24), 1–10. https://doi.org/10.1002/cpe.4932

Zaini, N., Ean, L. W., Ahmed, A. N., & Malek, M. A. (2021). A systematic literature review of deep learning neural network for time series air quality forecasting. *Environmental Science and Pollution Research*, *29*(4), 4958–4990. https://doi.org/10.1007/s11356-021-17442-1