Recommendation and Prediction in a Microservice Web Application for Cyber Ranges

Luis de-Marcos, Daniel Rodríguez, José-María Gutiérrez-Martínez, Adrián Domínguez-Díaz, Sergio Caro-Álvaro

Universidad de Alcalá

Computer Science Department

Alcalá de Henares 28805. Madrid, Spain

{luis.demarcos, daniel.rodriguezg, josem.gutierrez, adrian.dominguez, sergio.caro}@uah.es

Abstract*. The EU research project between industry and academia µDevOps is a collaborative research project formed by an international network of organizations including industry and academia that aims to tackle current challenges of microservice development operations. An important case study considered in this project is the Cyber Ranges application a cyber security training and capability development exercises using microservices for the design, delivery, and management of simulation-based, experiences in cyber security developed by Silensec as one of the partners. This work describes the results of analyzing the scenario usage dataset of the Cyber Ranges training platform. This includes the matrix of starts for scenario/user and the attributes of scenarios. The aims are to produce recommendations of scenarios for users based on previous activity and to predict the success of scenarios as measured by the number of starts.*

Keywords. Recommender system, prediction, classification, microservice, web application, cyberrange.

1 Introduction

In online systems users are often overwhelmed by offer and need help finding what they're looking for, so recommendations are essential, since they lead to happier customers and more sales. Recommender systems are like salesmen who know, based on your history and preferences, what users may like. They reduce transaction costs of finding and selecting items in an online environment. The use of efficient and accurate recommendation techniques is very important for a system that will provide good and useful information to its individual users (Enríquez et al, 2019)

Similarly, online businesses can be overwhelmed by the amount of data produced by users when interacting with their platforms. Making sense of all the data can be difficult in a modern environment, where data is highly distributed, diverse, and dynamic. Predicting sales and classifying users are important to businesses because it helps them make better decisions. Sales forecasting helps in overall business planning, budgeting, and risk management. It allows companies to efficiently allocate resources for future growth and manage its cash flow (Atkins et al, 2023). Classification of users is important because it helps businesses understand their customers better and tailor their products or services to meet their needs. This leads to happier customers and more sales (Mahalingam, 2023).

The aim and objectives of this paper are to produce recommendations for new scenarios that a user can take based on the activity of other users in the *Cyber Ranges* training platform. In this preliminary work, we also try to predict number of starts of a given training as a measure of its success based on its descriptive attributes since it determines the incomes produced by the training.

The rest of the paper is structured as follows. Section 2 covers the background. Section 3 presents the state of the art. Section 4 briefly outlines the approach. Section 5 reports findings. The paper concludes with final remarks and future research.

2 Background

This work is part of the µDevOps project[1], which aims to make software better by encouraging testing throughout the software lifecycle. It is a joint research program with academic and industrial partners to share knowledge. The research program has four goals: learning from context, testing continuously in real settings, assessing and managing risks in µDevOps,

[1] https://microdevops.wordpress.com/

and offering µDevOps development and testing as a service. The program is organized into work packages and tries to solve the problems of a fast-changing environment where development and operations are integrated. And as one of the case studies, we use one partner's application, the Cyber Ranges training platform which employs a microservice architectural design. The aim of this research is to provide recommendations for users and predict the success of trainings.

2.1 Microservices

Microservices - also known as the microservice architecture - is an architectural style that structures an application as a collection of services that are independently deployable, loosely coupled, organized around business capabilities, and owned by a small team. Microservices architectures make applications easier to scale and faster to develop, enabling innovation and accelerating time-to-market for new features (Hasselbring and Steinacker, 2017) [4].

Some examples of web applications that use microservices architecture are:

- Amazon: Initially, Amazon was a monolithic application but when microservice came into existence, Amazon was the first platform to break its application into small components, thereby adapting microservice.
- Netflix uses microservices with APIs.
- Uber: When Uber switched from monolithic nature to a microservice, it allowed to scale its services more efficiently contrasting with the bottleneck produced by the previous monolithic architecture.

An architecture of microservices is a software design approach focusing on application creation, formed by a set of small, independent services that communicate with each other through APIs. In the case of online learning platforms, microservices architectures can be beneficial for the development of these platforms as they allow for greater flexibility and scalability. By using this architecture, developers can create small, independent services that communicate with each other through APIs. This allows each service to be scaled and updated independently. Most online learning platforms use microservice architecture including Coursera, edX, Udacity, Udemy and Skillshare (Dreiko, 2021) [5].

2.2 Recommender Systems

Recommender systems are a type of information filtering system that provide suggestions for items that are most relevant to a user. They can help users to choose from many options, such as products, music, news, or movies. Some examples of recommender systems are present in Amazon, IMDb, Facebook, Netflix, and Spotify.

Recommender systems can use different methods to make suggestions, such as collaborative filtering, content-based filtering, or hybrid methods. Content-based filtering uses the features and attributes of items to find similar items. Hybrid methods combine both approaches to improve the quality and diversity of recommendations (Isinkaye et al., 2015; Aymen and Imène, S, 2022).

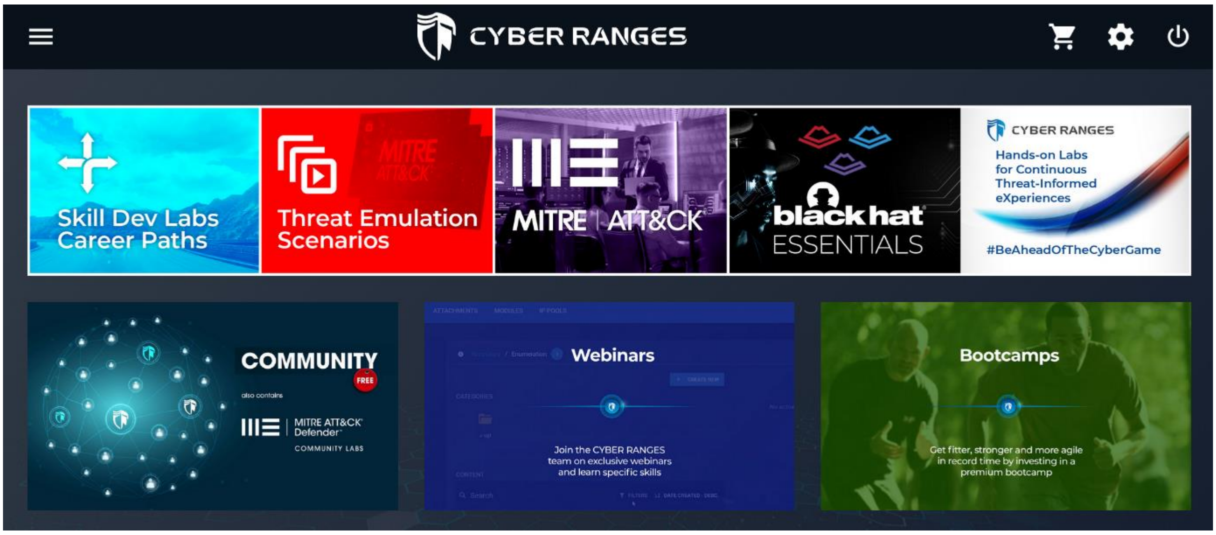
There are several examples of recommender systems used in online learning platforms. For instance, many online education platforms have built diverse recommender systems that utilize traditional data mining methods, such as Collaborative Filtering (CF) (Li and Kim, 2021). Some researchers have also proposed a deep learning-based course recommender system (DECOR) which captures high-level user behaviours and course attribute features (Liu et al, 2022).

3 Case Study: The CyberRanges Web Application

Cyber Ranges consist of virtual environments that provide secure and legal environment for cybersecurity education, practice, and training. They are designed to isolate trainees from threats by allowing them to recognize and respond to real-world challenges in a controlled environment. This approach ensures that client infrastructure and data are never at risk due to cybersecurity training (Aries Security, 2020). The main commercial cyber range solutions available in the market for are:

- IBM X-Force Command Cyber Range: This is a cloud-based platform that offers immersive and gamified cyberattack simulations, as well as customized scenarios and training modules. It also provides access to IBM's security experts and tools, such as IBM Security QRadar and IBM Resilient.
- Palo Alto Networks Cyber Range: This is a physical facility that hosts hands-on workshops and exercises for cybersecurity professionals, students, and enthusiasts. It also leverages Palo Alto Networks' products and services, such as Cortex XDR and Prisma Cloud, to demonstrate best practices and techniques for threat prevention and detection.
- Fortinet NSE Cybersecurity Training Institute: This is an online platform that offers self-paced courses and certifications for various levels of cybersecurity skills and roles. It also includes access to Fortinet's cyber range labs, where learners can practice their skills in realistic scenarios using Fortinet's security solutions, such as FortiGate and FortiSandbox.

CyberRanges (see Figure 1) is a microservice web application developed by Silensec that provides cybersecurity training and capability development exercises using technology and services for the design, delivery, and management of simulation-based, deep-dive experiences. CyberRanges can be used to learn, train, test, measure, and improve the digital dexterity and cyber resilience of professionals, teams, and organizations by using a platform and technology that resembles a real-world scenario. A scenario is a training session.



Join the CYBER RANGES team on exclusive webinars and learn specific skills

Get fitter, stronger and more agile in record time by investing in a premium bootcamp

Figure 1. CyberRanges Web Application

As there are several competing platforms offering cyber range solutions. To be competitive it is not only required to provide the better offering in terms of platform functionality, library of scenarios and user base, but it is also necessary include recommender systems that suggests trainings based on to the capabilities of the users. Furthermore, as the incomes of the platform are based on the number of starts of trainings, providers of cyber ranges are usually interested in knowing the features that determine training success from users' perspective. Data analytics can also provide further insights, like the categories that are underrepresented in the library but are particularly successful or appealing to users. All this information can help providers to identify the most meaningful ways to expand their library of scenarios. The present research deals with all these issues by analyzing the data of user activity of the CyberRanges application.

The data used for the recommendation system is composed of two datasets. The first dataset (Dataset 1) is a matrix of IDs of users and IDs of scenarios. Each cell contains a 1 if user started the scenario, 0 it not. There are 11,093 users and 1,534 scenarios.

The second dataset (Dataset 2) is described in Table 1. It comprises multiple attributes and the target attribute (Sum) as the number of times a user starts a scenario (computed for this analysis as an aggregation from Dataset 1)

Table 1. Dataset with scenario information

ID	String
Name	String
Owner	String
Author	String
Status	Categorical. PUBLISHED, DRAFT, PENDING_APPROVAL
Categories	Categorical (multiple values as CSV. E.g. "Web, System Exploitation".
Tags	Categorical (multiple values as CSV describing the scenario). E.g., "SQLi, File Upload, PHP" Arbitrary.
Visible:	Boolean
Type:	Categorical. CCL, CPL, CTF, MEL, TTX
Mode	Categorical. SINGLEPLAYER, MULTIPLAYE
Difficulty	Categorical. EASY, INTERMEDIATE, HARD, IMPOSSIBLE
Technical requirements	Numerical (integer) attributes describing the technical requirements of the scenario: Power, Duration, CPU, RAM (MB), Disk (GB), #Servers, #Containers, #Networks, #Subnets, #Ports, #CloudConfig, #Routers, #Windows, #Linux, #Unknown, #ZoneNova, #ZoneWindows, #Zone BSD, #Zone Passthrough
Sum.	Numerical: Integer. Number of times a user starts a scenario (Computed for this analysis as an aggregation from Dataset 1)

4 Experimental Work

4.1 Data Analysis Methods

In this work, we tackle data analysis through correlations and machine learning. First, we check the correlations of each tuple of scenarios based on the users that started both. From the machine learning point of view, apply: (i) Regression of the number of starts of scenarios (Sum), and (ii) Classification of scenarios (based on the discretization of Sum attribute). Before data analysis, we use descriptive statistics to identify possible anomalies and to describe the main attributes.

The choice of data analysis methods depends on the research questions, the data characteristics, and the available resources. In this project, we decided to use correlations and machine learning as our main tools for exploring the relationships between variables and predicting outcomes. Correlations are useful for measuring the strength and direction of linear associations between two or more variables and can help us identify potential causal factors or confounders. Machine learning is a branch of artificial intelligence that uses algorithms to learn from data and make predictions or classifications. Machine learning can handle complex and nonlinear patterns, deal with missing or noisy data, and discover hidden features or structures in the data. We did not use other methods, such as clustering, or hypothesis testing, because they either did not fit our research goals, required more assumptions or data preprocessing, or were less efficient or accurate than machine learning.

4.2 Exploratory Analysis

Descriptive statistics do not show any significant anomaly. The target attribute Sum (number of starts) seems to follow a power-law, long tail distribution (N=1534, Mean: 25.9. SD: 68.87, SE Mean) as shown in Figure 2, where we can see a few instances with many starts and many instances with a limited number of starts. This makes the analysis more challenging since several methods rely on specific distributions.

It is worth noting that there are 187 instances are missing the values for the following attributes: Power, Duration, CPU, RAM (MB), Disk (GB). Also 215 instances are missing all other numerical attributes (the ones that start with # in Table 1).

Kruskal-Wallis non-parametric tests return statistical differences for all categorical attributes when response is Sum: Status (H=193.4), Visible (H=195.5), Type (H=312.6), Mode (H=37.7) & Difficulty (H=69.8), with p<0.01 in all cases. Although in several cases the factors have a reduced number of instances. This is shown in Figure 3. This means that for all this attributes there is a statistical difference for at least one of the groups which can be visually examined in the figure. For instance, when it comes to Difficulty, scenarios rated as 'Impossible' have significantly higher number of starts. Exploratory analysis also shows very high values for several instances of the numerical attributes (particularly Power, Duration, CPU, RAM (MB), Disk (GB)) which may be outliers or bias several results.

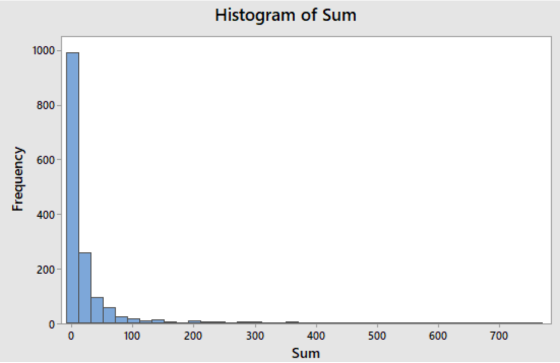


Figure 2. Histogram of the Sum attribute

5 Experimental Results and Discussion

This section provides the results of the experimental work to produce recommendations based on user activities and to predict the number of starts of courses. It also discusses results briefly discussions.

5.1 Correlation-based Recommendations

We computed the correlation between each pair of scenarios based on users' starting them. The matrix was inputted for the correlation as provided (Dataset 1 as described in Section 3). Only valid pairs of correlations were included. These results in a set of 583740 correlations. It takes 3-5 minutes with the dataset provided. Correlations are then filtered by value selecting only those that are larger than 0.6 (moderate correlation) and which have a minimum number of instances in which the correlation is based (initially 10). This rules out highly correlated scenarios which are only based in a very low number of users taking both. Parameters of both filters can later be tuned if needed.

The correlation value can be roughly interpreted as the percentage of number of users who started one of scenarios and the other based on a minimum sample on the dataset. For instance, for a given tuple of courses that were started 238 times we found a correlation value of 0.94 indicating that a significant amount of the users who started the first scenario also started the second. Correlation value is a measure of the strength of the recommendation. Higher means stronger. The minimum number of instances is a measure of number of instances on which the recommendation is based. Higher means better.

5.2 Performance of Regression

To analyze the performance of regression, we firstly performed the following preprocessing:

1. Transpose the matrix (Dataset 1) to compute the Sum of the number of starts for all users.
2. Join Sum to Dataset 2 to use it as the target attribute.
3. Filter out the scenarios with Sum = 0 (never started by any user).
4. Remove all instances with any missing value in any of the numeric attributes.
5. Partition of the dataset for training and testing: 70/30 (draw randomly)

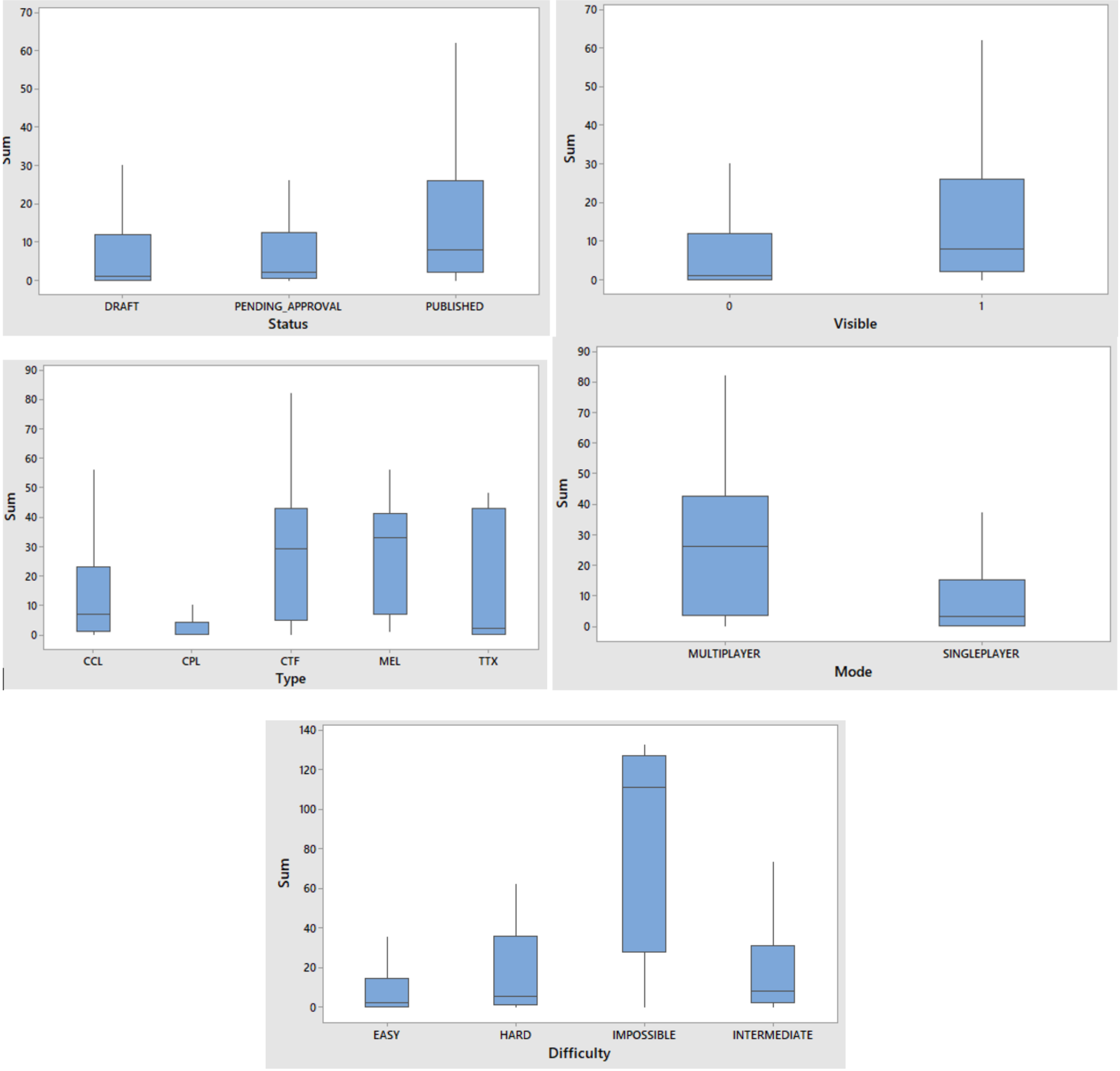


Figure 3. Box plots of the categorical attributes (Status, Visible, Type, Mode & Difficulty)

With the resulting dataset, we trained several machine learning regressors to predict Sum (number of starts of a given scenario) as a function of the rest of the attributes of the scenario. Regressors used and results are shown in Table 2 .

Table 2. Results of the regression techniques

Regressor	**Adjusted R^2**	**Mean absolute error**
Linear (only with numeric attributes)	0.024	72.10
Simple regression tree	-0.193	41.03
Tree ensemble	0.047	43.66
Gradient boosted trees	-0.087	41.32
Random forest	0.056	42.94

As shown on the table, regression results show that the regression models do not fit the data. In the best case, only 5% of the variability is explained by the best model. Negative R^2 in several cases means that the model fits worse than a linear model. Results then suggest that Sum (number of starts) of a scenario cannot be explained as a measure of the other attributes of the scenario given (in the dataset)

5.3 Classification

To analyse the performance of classification, we firstly pre-processed the dataset using the same procedure described in the previous subsection. We also discretized based on the mean of Sum (number of starts) which is around 25. So, all instances with less than 25 starts where classified as ‘Low’, and all others (Sum >=25) as ‘High’. Partition used stratified sampling. We then trained several classifiers of different types to predict the discretized Sum as a function of the rest of the attributes of the scenario. Results are summarized in table 3.

Table 3. Results of the classifiers

Classifier	Accuracy	Precision (High)	Precision (Low)	AUC
KNN	.77	.46	.83	.63
Decision Tree (C4.5)	.81	.57	.86	.70
Gradient Boosted Tree	.80	.59	.83	.76
Random Forest	*.81*	*.61*	*.83*	*.79*
Tree Ensemble	.80	.59	.83	.78
Naïve Bayes	.76	.39	.81	.68
Fuzzy Rule	.76	.36	.84	.61

Analysis from table 3 shows that Random Forest outperforms other classifiers, although only by a small margin when compared to other ensemble methods (Gradient Boosted Tree, Tree Ensemble), and simple classification tree (C4.5). Differences are significant when compared with instance-based methods (KNN), probabilistic (Bayes) and fuzzy rules. The ROC curve of the Random Forest classifier is presented in Figure 4. It plots two parameters: True Positive Rate (TPR) and False Positive Rate (FPR) at different classification thresholds. Classifiers that give curves closer to the top-left corner indicate a better performance. The red line represents a random classifier. The area under the ROC curve (AUC) is a measure of how well a model can distinguish between two classes.

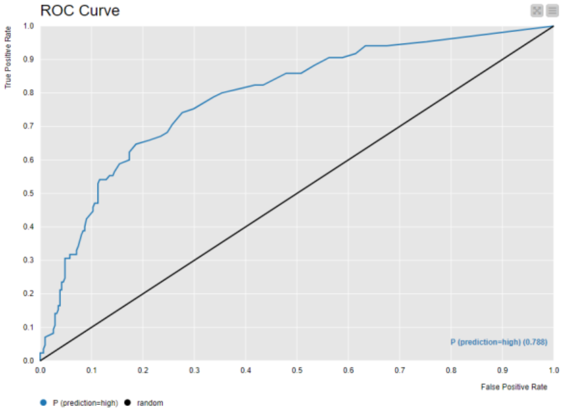


Figure 4. ROC Curve of the best classifier (Random Forest)

Results show that ensemble methods, with Random Forest in particular. produce a moderate successful method to predict the number of starts of a given scenario. In all cases, we can observe a low precision rate for ‘High’ instances as compares to ‘Low’. This is probably due to the imbalance of the resulting discretized dataset.

6 Conclusions and Future Work

Results show that a simple correlation of scenarios from a matrix representing the starts of scenarios and users can be used to produce recommendations of new scenarios to users by applying a set of filters with a minimum threshold for the correlation and a minimum number of instances in which the correlation is based. Its computation, though, can scale poorly as the number of scenarios increases.

Regression results are quite poor. The number of starts of a scenario cannot be explained by other attributes. We argue that labeling of scenarios (attributes Categories and Tags) and the quality of the numerical attributes need to be checked. Further, since the provided attributes are not descriptive for the target metric (number of starts) other attributes, like learning performance, should be considered for future analysis.

Classification models return moderate good results possibly as measured by the metrics used in this study (accuracy, precision, and AUC) because of the unbalance of the dataset (instances with a high number of starts vs low).

As for future work, we suggest the following lines for each analysis of this paper. Concerning correlation, it is necessary to assess whether results are meaningful for the developers and users of the Cyber Ranges application. Next steps also include the integration of correlation results as a recommender system in the learning platform. The correlation process should be integrated in the architecture considering its scalability limitations (limited scalability as a function of the number of scenarios). Concerning regression, we suggest checking the integrity of numeric data and providing new attributes before any further development. As for classification, we suggest increasing the class underrepresented, which means having more scenarios with a relevant number of starts. Additionally, it is also possible to try other discretization methods.

We also want to mention that although our method and results focus on a microservice web application, it is also possible to apply it to monolithic systems which are scalable vertically and horizontally.

Comparing results with other methods/state of the art is needed to assess the success of our approach. Given the limited number of existing platforms providing cyber ranges training, we did not find any previous studies (research papers or reports).

Acknowledgments

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 871342

References

Aries Security. (2020). What is a Cyber Range? A Definitive Guide and Definition. https://www.ariessecurity.com/what-is-a-cyber-range-a-definitive-guide-and-definition/ [Accessed on 16th May 2023]

Atkins, C., Uster, M.V.D., Mahdavian, M., and Yee, L. Unlocking the power of data in sales. 2016 [cited 16th May 2023; Available from: https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/unlocking-the-power-of-data-in-sales

Aymen, A.T.M., and Imène, S.: ‘Scientific Paper Recommender Systems: A Review’. Springer International Publishing, 2022, pp. 896-906

Dreiko, J. How to Plan Software Architecture For eLearning Platforms. 2021 [Last accessed 7th Sept 2023]; Available from: https://polcode.com/resources/blog/how-to-plan-software-architecture-for-e-learning-platforms/

Enríquez, J.G., Morales-Trujillo, L., Calle-Alonso, F., Domínguez-Mayo, F.J., and Lucas-Rodríguez, J.M., ‘Recommendation and Classification Systems: A Systematic Mapping Study’, Scientific Programming, 2019, pp. 1-18

Hasselbring, W. and Steinacker, G.. Microservice architectures for scalability, agility and reliability in E-commerce. Software Architecture Workshops (ICSAW), 2017 IEEE International Conference on. IEEE, 2017.

Isinkaye, F.O., Folajimi, Y.O., and Ojokoh, B.A.: ‘Recommendation systems: Principles, methods and evaluation’, Egyptian Informatics Journal, 2015, 16, (3), pp. 261-273

Li, Q., and Kim, J.: ‘A Deep Learning-Based Course Recommender System for Sustainable Development in Education’, Applied Sciences, 2021, 11, (19), pp. 8993

Liu, T., Wu, Q., Chang, L., and Gu, T.: ‘A review of deep learning-based recommender system in e-learning environments’, Artificial Intelligence Review, 2022, 55, (8), pp. 5953-5980

Mahalingam, K. The Importance of Sales Forecasting. 2022 15th May 2023]; Available from: https://www.chargebee.com/blog/importance-of-sales-forecasting/.