FEDERATED LEARNING SETTING FOR E-LEARNING COURSE RECOMMENDATIONS

Miloš Jolović*,
[0009-0003-6580-2039]
Dušan Kostić,
[0009-0000-3432-6074]
Aleksandar Joksimović,
[0009-0008-5711-7636]
Talib Tahirović,
[0009-0004-9515-8237]
Petar Lukovac
[0000-0003-4561-8886]

University of Belgrade,
Faculty of Organizational Sciences,
Belgrade, Srebia

Abstract:

The main research problems addressed in this article refer to the complexity of maintaining Learning Management Systems, ensuring data privacy throughout any analysis of that data, and personalizing learning, which can be a task requiring significant resources. The research aims to provide an answer that can address these problems through a Federated Learning setting, enabling cross-institutional cooperation and retaining the data in its place of origin. The research includes a simulation of such a Federated learning setting, which proved to be very interesting for identifying future challenges and directions for a tangible, real-world application. The simulation was built with a dataset comprised of students' grades and interests in a first-year mandatory subject, E-business, taught at the University of Belgrade, Faculty of Organizational Sciences. This dataset was suitable for building a recommender system that can produce an intelligent suggestion for an elective course for each student individually based on their interests and academic achievements.

Keywords:

Federated Learning, E-learning, Personalized Learning.

INTRODUCTION

Recent advances in the generation of immense volumes of data have raised concerns about leveraging the generated data in a privacy-preserving manner. Consequently, a Federated Learning (hereinafter: FL) setting has emerged. The FL setting is a distributed machine learning technique where models are brought to local data on each node of the network, instead of centralizing vast volumes of data [1]. Several authors have proposed the use of FL in education for data analysis [2] [3] or for detecting dropout rates [4]. One of the potential applications of the FL setting is training models across academic institutions to develop and sustain personalized learning. Personalized learning can be defined as a complex activity that considers individual needs and goals in the process of learning [5]. This research paper aims to develop a FL simulation for making intelligent recommendations for students regarding choosing an appropriate elective course depending on their previous knowledge and interests.

Correspondence:
Miloš Jolović

e-mail:
milos.jolovic@elab.rs

DOI: 10.15308/Sinteza-2025-35-40

2. RELATED WORK

The term FL was first introduced in [6] as a way to distribute the process of training a machine learning algorithm by a federation of clients, such as mobile devices or several different companies and institutions. The term FL has since evolved into a machine learning setting in which client nodes collaborate in training a machine learning model on their local data, with a central server orchestrating the training process [7].

The training process of a machine learning algorithm via FL can be described in four steps. Firstly, the server initializes a global machine learning model. Next, client nodes download the global model and train it on their local data. Then, client nodes send back the model parameters to the server node. Finally, the server aggregates the parameters using the FL algorithm and updates the global model. These steps are repeated until the model converges [8].

The FL paradigm has three different categories, depending on how data is partitioned in both feature and sample space: Vertical FL, Horizontal FL, and Federated Transfer Learning. Horizontal FL refers to a situation where clients hold data with the same features and different samples. Vertical FL refers to a situation where clients hold data with different features and share samples. Federated Transfer Learning refers to a situation where datasets differ in feature and sample spaces with limited overlaps [9] [10].

FL can also be categorized based on the amount of decentralization between the nodes as centralized (CFL), decentralized (DFL) and semi-decentralized FL (SDFL). In DFL, participants perform all four steps of the FL process independently, SDFL participants perform the first three steps independently, the aggregator node handles the aggregation process and then passes the aggregator role to a new node in the next iteration of training. Centralized FL functions between a server node and several client nodes, as described in Figure 1 [11].

The justification for usage of the FL setting can be summarized as follows. First, machine learning algorithms can be trained on separate data silos (e.g., several medical institutions) in a privacy-preserving manner, without the need for any silo to share its local data. Second, some data sources provide a large amount of real-time data, making it more efficient to move the model to the data rather than vice-versa. Third, many legal regulations can make sensitive data hard to move from the place of origin [12] [13] [14] [15].

From a practical perspective FL setting can be used in healthcare informatics [16] [17], and the banking industry in several directions, such as assessing credit risks [18], open banking [19] or credit fraud detection [20]. Practical applications can be found in IoT systems [21] [22], wireless communications [23], the automotive industry [24] Etc.

More specifically, practical usage of FL can be found in an e-learning setting. The concept of e-learning can be described as an educational process that leverages digital platforms and resources to facilitate the learning process and educational resources [25]. Research conducted by [26] states that students often use the same edge devices (e.g., PCs, laptops, mobile phones) for both entertainment and study assignments, which can cause problems with maintaining focus on study tasks. The authors propose an FL architecture that collects data from students' edge devices in a privacy-preserving way to train a classification model for students' on-screen

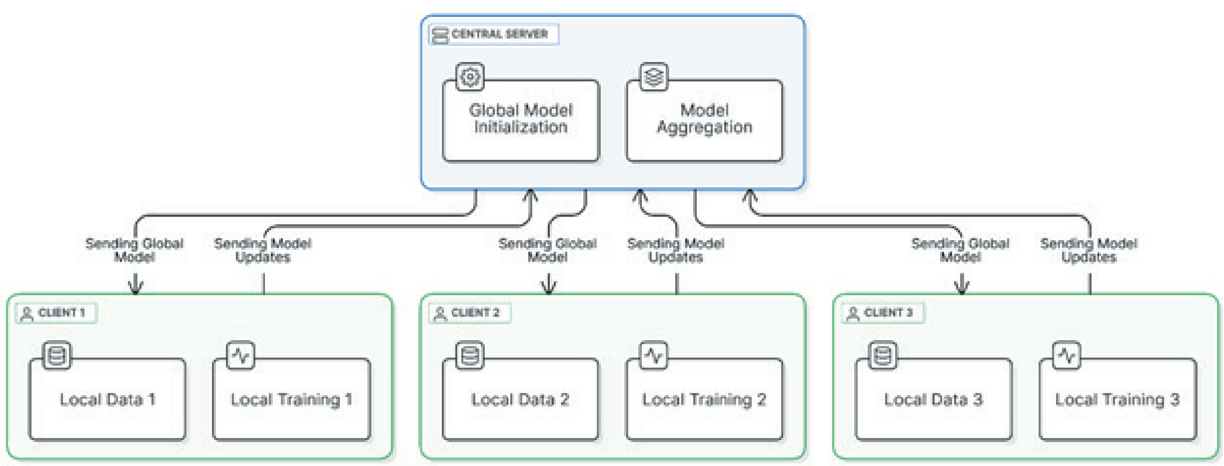

Figure 1. Federated learning process

time that detects situations in which study tasks lose students' attention due to entertainment. Research [27] recognizes the significance of personalized learning and the possibilities of implementing it in an e-learning setting, where security and data privacy may arise. Recommendation systems trained via an FL architecture are proposed as a solution for implementing personalized learning in a secure and privacy-preserving way.

The simulation of an FL setting for this paper was implemented through the Flower Framework. Flower is an FL framework that provides a unified approach to FL, analytics, and evaluation that can be applied to federating any workload or machine learning framework [28]. This research paper proposes a simulation of an FL setting in e-learning, specifically building a recommendation model to suggest elective courses to students depending on their previous interests and academic accomplishments. The simulation environment represents an opportunity to identify the benefits of FL setting in e-learning, such as cross-institutional cooperation between academic institutions without the need to share raw data, learning personalization and possible integration with Learning Management Systems.

3. METHODOLOGY

Data used for this research paper was collected from students enrolled in the Information Systems and Technologies undergraduate study program at the Faculty of Organizational Sciences, University of Belgrade. This study program focuses on applying computer science to construct solutions for business-oriented problems. More precisely, the dataset comprised students' grades and topics of project-based activities from the first-year E-business course. E-business provides students with both theoretical and practical, introductory knowledge of building web-based applications. Course assessment consists of mandatory closed-format tests once a week, four project-based homework assignments and one exploratory essay designed to encourage students to explore state-of-the-art technologies regarding various topics of computer science such as blockchain, artificial intelligence, big data, cloud technologies, IoT solutions, etc. The procedure is described below.

Research began with collecting data from Moodle Learning Management System (LMS). The Department of E-business utilizes Moodle throughout the teaching process for posting lecture resources, communication with students, assessments and grading. As mentioned before, the dataset consists of students' grades on vari-

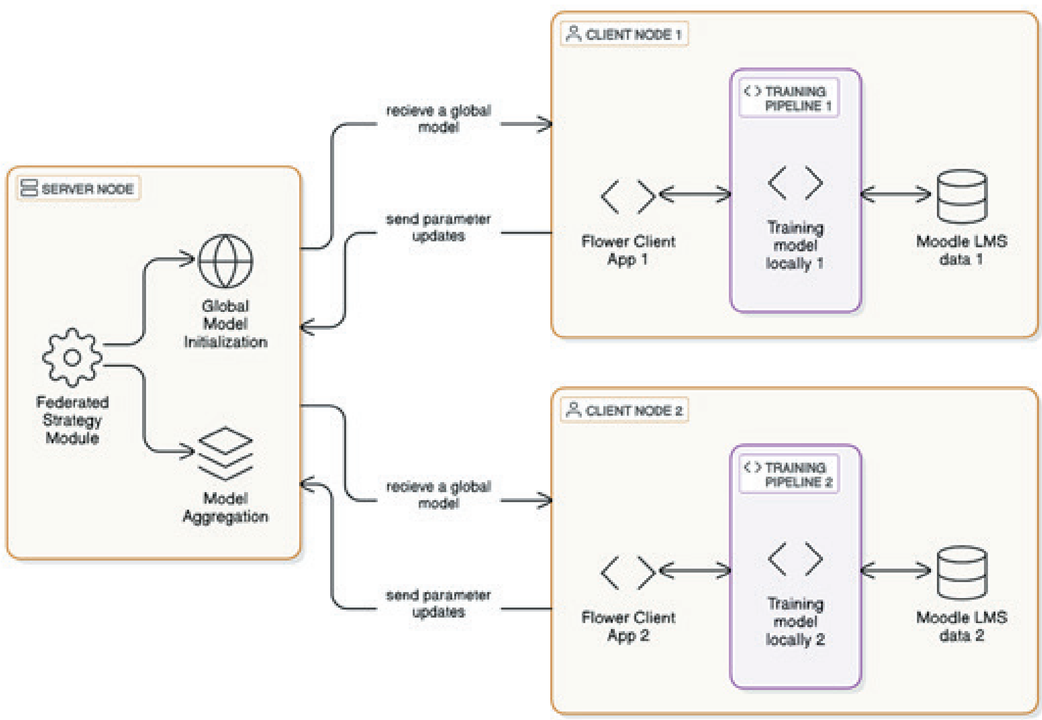

Figure 2. Federated learning project architecture

ous homework assignments and tests, with topics of their exploratory essays. The collected data was used to build an intelligent recommendation system that suggests elective courses for students based on their grades and previous interests expressed in essays done on the E-business course. The Dataset contains 609 student records.

In the process of building a recommender model data preparation involved replacing missing values with zeroes – assigning zero points to students who did not complete a particular assignment. Students who did not successfully complete the course were removed from the dataset. Numerical columns were normalized, and students were grouped into three clusters based on their performance in the E-business course. These clusters served as recommendations for the level of elective courses that students should take. The clustering was performed using the K-Means algorithm, implemented with the scikit-learn library. Data containing elective courses topics and descriptions was processed using the Stanza library and transformed into TF-IDF matrices. The same procedure was applied to students' essay topics and their respective categories. Cosine similarity was computed between the matrices representing course descriptions and students' essays. The three highest similarity scores were used to generate personalized elective course recommendations for each student.

The architecture of FL was built with Flower Framework using two client nodes. Each client node received a global model initialized on the server node and returned updates to the global model. The server node aggregated the model updates using the Federated Averaging strategy.

4. RESULTS

As a result of training on students' data, three clusters emerged:

- Cluster 0 – students with good performance and a minor lack of theoretical knowledge.
- Cluster 1 – students that lack both theoretical and practical knowledge and should cover the basics again.
- Cluster 2 - students that excel in the E-business course and have a great understanding of both theoretical and practical knowledge.

Students from Cluster 0 were recommended intermediate-level elective courses, Cluster 1 students were advised to choose elective courses which will cover the basics again, and students from Cluster 2 were encouraged to enroll in advanced elective courses.

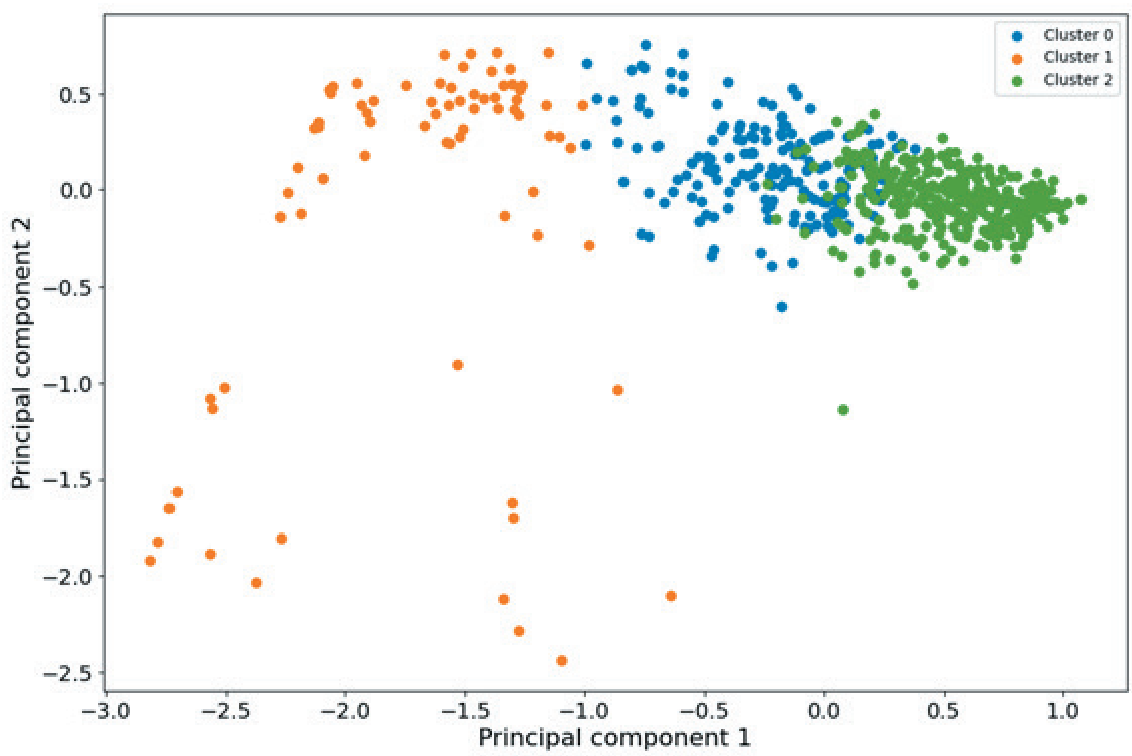

Figure 3. Student clusters emerged from the K-Means clustering algorithm

The top three elective course recommendations were generated for each student based on the similarity between their exploratory essay topics and elective course descriptions. For instance, students who wrote about IoT in their essays were recommended IoT-based elective courses.

5. CONCLUSION

This pilot research demonstrated a simulation of an FL setting implemented with Flower Framework. The study aimed to highlight the potential of applying federated learning in possible cross-institutional collaboration and personalization of e-learning. The research also proves to be valuable in identifying potential challenges that may arise during the development of a true-to-life federated architecture.

Future research could focus on expanding cross-institutional cooperation by building and deploying a fully functional FL system across multiple universities or faculties. This system could integrate real-time student progress tracking and incorporate more complex metrics. However, several challenges must be addressed, including security risks in federated architectures, data heterogeneity across institutions (client nodes), and the need for a custom aggregation strategy tailored to varying data quality and volume at each institution.

REFERENCES

[1] K. M. Jawadur Rahman et al., “Challenges, Applications and Design Aspects of Federated Learning: A Survey,” *IEEE Access,* vol. 9, pp. 124682–124700, 2021, doi: 10.1109/ACCESS.2021.3111118.

[2] C. Fachola, A. Tornaría, P. Bermolen, G. Capdehourat, L. Etcheverry, and M. I. Fariello, “Federated Learning for Data Analytics in Education,” *Data 2023*, Vol. 8, Page 43, vol. 8, no. 2, p. 43, Feb. 2023, doi: 10.3390/DATA8020043.

[3] S. Guo, D. Zeng, and S. Dong, “Pedagogical data analysis via federated learning toward education 4.0,” *American journal of education and information technology*, vol. 4, no. 2, pp. 56–65, 2020, doi: 10.11648/J.AJEIT.20200402.13.

[4] T. Zhang et al., “Enhancing Dropout Prediction in Distributed Educational Data Using Learning Pattern Awareness: A Federated Learning Approach,” *Mathematics 2023*, Vol. 11, Page 4977, vol. 11, no. 24, p. 4977, Dec. 2023, doi: 10.3390/MATH11244977.

[5] A. Shemshack and J. M. Spector, “A systematic literature review of personalized learning terms,” *Smart Learning Environments,* vol. 7, no. 1, pp. 1–20, Dec. 2020, doi: 10.1186/S40561-020-00140-9/TABLES/4.

[6] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, “Communication-Efficient Learning of Deep Networks from Decentralized Data,” Apr. 10, 2017, *PMLR*. Accessed: Feb. 24, 2025. [Online]. Available: https://proceedings.mlr.press/v54/mcmahan17a.html

[7] P. Kairouz et al., “Advances and Open Problems in Federated Learning,” Foundations and Trends® in *Machine Learning*, vol. 14, no. 1–2, pp. 1–210, Jun. 2021, doi: 10.1561/2200000083.

[8] J. Wen, Z. Zhang, Y. Lan, Z. Cui, J. Cai, and W. Zhang, “A survey on federated learning: challenges and applications,” *International Journal of Machine Learning and Cybernetics 2022* 14:2, vol. 14, no. 2, pp. 513–535, Nov. 2022, doi: 10.1007/S13042-022-01647-Y.

[9] Y. Liu et al., “Vertical Federated Learning: Concepts, Advances, and Challenges,” *IEEE Transactions on Knowledge and Data Engineering,* vol. 36, no. 7, pp. 3615–3634, Jul. 2024, doi: 10.1109/TKDE.2024.3352628.

[10] A. Z. Tan, H. Yu, L. Cui, and Q. Yang, “Towards Personalized Federated Learning,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 12, pp. 9587–9603, Dec. 2023, doi: 10.1109/TNNLS.2022.3160699.

[11] E. T. Martinez Beltran et al., “Decentralized Federated Learning: Fundamentals, State of the Art, Frameworks, Trends, and Challenges,” *IEEE Communications Surveys and Tutorials*, vol. 25, no. 4, pp. 2983–3013, 2023, doi: 10.1109/COMST.2023.3315746.

[12] Q. Li et al., “A Survey on Federated Learning Systems: Vision, Hype and Reality for Data Privacy and Protection,” *IEEE Transactions on Knowledge and Data Engineering,* vol. 35, no. 4, pp. 3347–3366, Apr. 2023, doi: 10.1109/TKDE.2021.3124599.

[13] T. R. Gadekallu, Q.-V. Pham, T. Huynh-The, S. Bhattacharya, P. K. R. Maddikunta, and M. Liyanage, “Federated Learning for Big Data: A Survey on Opportunities, Applications, and Future Directions,” Oct. 2021, Accessed: Feb. 26, 2025. [Online]. Available: https://arxiv.org/abs/2110.04160v2

[14] P. M. Mammen, “Federated Learning: Opportunities and Challenges,” Jan. 2021, doi: 10.1145/nnnnnnn.nnnnnnn.

[15] N. Truong, K. Sun, S. Wang, F. Guitton, and Y. K. Guo, “Privacy preservation in federated learning: An insightful survey from the GDPR perspective,” *Computers & Security,* vol. 110, p. 102402, Nov. 2021, doi: 10.1016/J.COSE.2021.102402.

[16] N. Rieke et al., "The future of digital health with federated learning," n*pj Digital Medicine 2020 3:1,* vol. 3, no. 1, pp. 1–7, Sep. 2020, doi: 10.1038/s41746-020-00323-1.

[17] J. Xu, B. S. Glicksberg, C. Su, P. Walker, J. Bian, and F. Wang, "Federated Learning for Healthcare Informatics," *Journal of Healthcare Informatics Research*, vol. 5, no. 1, pp. 1–19, Mar. 2021, doi: 10.1007/S41666-020-00082-4/TABLES/2.

[18] D. Kawa, S. Punyani, P. Nayak, A. Karkera, and V. Jyotinagar, "Credit Risk Assessment from Combined Bank Records using Federated Learning," *International Research Journal of Engineering and Technology*, p. 1355, 2008, Accessed: Feb. 26, 2025. [Online]. Available: https://www.rbi.org.in/Scripts/NotificationUser.aspx

[19] G. Long, Y. Tan, J. Jiang, and C. Zhang, "Federated Learning for Open Banking," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12500 LNCS, pp. 240–254, 2020, doi: 10.1007/978-3-030-63076-8_17.

[20] W. Yang, Y. Zhang, K. Ye, L. Li, and C. Z. Xu, "FFD: A federated learning based method for credit card fraud detection," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11514 LNCS, pp. 18–32, 2019, doi: 10.1007/978-3-030-23551-2_2/TABLES/4.

[21] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, and H. Vincent Poor, "Federated Learning for Internet of Things: A Comprehensive Survey," *IEEE Communications Surveys and Tutorials*, vol. 23, no. 3, pp. 1622–1658, Jul. 2021, doi: 10.1109/COMST.2021.3075439.

[22] V. Rey, P. M. Sánchez Sánchez, A. Huertas Celdrán, and G. Bovet, "Federated learning for malware detection in IoT devices," *Computer Networks,* vol. 204, p. 108693, Feb. 2022, doi: 10.1016/J.COMNET.2021.108693.

[23] S. Niknam, H. S. Dhillon, and J. H. Reed, "Federated Learning for Wireless Communications: Motivation, Opportunities, and Challenges," *IEEE Communications Magazine,* vol. 58, no. 6, pp. 46–51, Jun. 2020, doi: 10.1109/MCOM.001.1900461.

[24] H. Zhang, J. Bosch, and H. H. Olsson, "Real-time end-to-end federated learning: An automotive case study," *Proceedings - 2021 IEEE 45th Annual Computers, Software, and Applications Conference, COMPSAC 2021*, pp. 459–468, Jul. 2021, doi: 10.1109/COMPSAC51774.2021.00070/VIDEO.

[25] S. Bashir and A. L. Lapshun, "E-learning future trends in higher education in the 2020s and beyond," *Cogent Education,* vol. 12, no. 1, Dec. 2025, doi: 10.1080/2331186X.2024.2445331.

[26] D. Mistry, M. F. Mridha, M. Safran, S. Alfarhood, A. K. Saha, and D. Che, "Privacy-Preserving On-Screen Activity Tracking and Classification in E-Learning Using Federated Learning," *IEEE Access,* vol. 11, pp. 79315–79329, 2023, doi: 10.1109/ACCESS.2023.3299331.

[27] D. Javeed, M. S. Saeed, P. Kumar, A. Jolfaei, S. Islam, and A. K. M. N. Islam, "Federated Learning-Based Personalized Recommendation Systems: An Overview on Security and Privacy Challenges," *IEEE Transactions on Consumer Electronics*, vol. 70, no. 1, pp. 2618–2627, Feb. 2024, doi: 10.1109/TCE.2023.3318754.

[28] D. J. Beutel et al., "Flower: A Friendly Federated Learning Research Framework," Jul. 2020, Accessed: Feb. 20, 2025. [Online]. Available: https://arxiv.org/abs/2007.14390v5