Assessment and comparison of meta-features for educational chatbots data and survey data

DIJANA OREŠKI, DIJANA PLANTAK VUKOVAC, GORAN HAJDIN
Faculty of Organization and Informatics
University of Zagreb
Pavlinska 2, Varazdin
CROATIA

Abstract: - Usage of chatbot platforms is acquiring great attention at all levels of education. Human-chatbot interactions generate huge amounts of data which are a valuable source of information, when properly analyzed by means of data and text mining. One of the most challenging tasks in the mining process is the selection of the appropriate algorithm for a data set at hand. This is a complex task and depends on characteristics of the dataset used in the analysis. Those characteristics are formalized through meta-features. In this paper, we identified meta-features of chatbot and survey data. As a case study, we evaluated two data sets and identified their general meta-features along with discussion. This is, to the best of our knowledge, the first examination of meta-features for chatbot interactions data and their comparison with survey data.

Key-Words: - meta-features, data characteristics, chatbot interactions, EDUBOTS, educational chatbot, survey data, meta-learning

Received: June 9, 2021. Revised: January 7, 2022. Accepted: February 17, 2022. Published: March 23, 2022.

1 Introduction

During the past few years, chatbots, which engage users in conversation to find out their opinions, have been adopted for a wide variety of applications [1]. Among various chatbot applications, a promising one is conducting interviews with students. Chatbots serve as a tool for giving feedback to the students [2], and have been also used as a new channel of collecting feedback from students. They are less time and resource-demanding from traditional surveys [3]. So far surveys have been widely used, although their application has several limitations such as low data quality in the open-ended questions ([4]; [5]). To overcome these limitations individual interviews are used to gain deeper insights [6]. Recently, chatbot is receiving attention among practitioners and researchers as a potentially valuable tool comprising advantages from qualitative and quantitative evaluations. Their capability of communication with users through natural language interaction interfaces serve as excellent basis to overcome challenges [7]. There is a small number of papers focusing on the education [8], especially research papers researching chatbot usage on students' evaluation of the course ([9], [10], [3]).

Inspired by these efforts, we are taking a step forward, and providing description and characterization of chatbot data. First step in the characterization process is to identify meta-features for meta-learning. So far, meta-learning has been used in general, on publicly available datasets. Most of the research papers are focused on the analytical system design, experimental methods or survey methods [11]. There are only a small number of papers (e.g. [12]) tackling the educational domain. Among that, we have not found any papers that tackle meta-learning and meta-features of chatbot data within the educational domain. Given the challenges mentioned above, we focus on the chatbot data meta-features and their comparison with traditional survey data. Our investigation has a goal to discover if there are differences in general meta-features of chatbot interactions data and traditional survey data. To achieve that goal, we evaluated both: chatbots` data with 82 participants and survey data with 50 participants, both from University of Zagreb, Faculty of Organization and Informatics. Both evaluations were focused on the students' course evaluation.

The rest of the paper is organized as follows. In Section 2, data is explained and the process of collecting both chatbot and survey data. In Section 3, an overview of meta-learning is given. The meta-features are defined and explained. Section 4

Section 5 summarizes our work, outlines directions for future research and ends in discussing how meta-features

2 Literature review

Our work is related to research in three areas: chatbots in higher education, survey data in higher education and meta-learning in data mining. All are explained in the following three subsections, with the focus of the paper on the first one.

2.1 About chatbots and chatbot data

Depending on the main use of the chatbot data can be stored in different ways. Most chatbots have a database which provides the basis for their output. Educational chatbots are still in an early phase of artificial intelligence teaching assistants and thus provide different ways of providing answers to the users, mostly students [13]. Some chatbots use predefined entries, while others employ keyword recognition, closed type decision trees or simple database searching and matching algorithms [14]. While in the public domain some rely on an open data approach, others are focused on privacy and confidentiality, thus using proprietary code [15].
Chatbots are used in different fields, and education is just a small fraction of their general use. When considering all fields, the most common chatbot type in published research papers is CALM-Systems, followed by Mobile Chatbot, FIUTEBOT and NDLtutor [11].
Chatbots rely on their data but can also be trained by different datasets to improve their effectiveness [8]. Some chatbots employ text-mining techniques to increase interpretation quality of user inputs. Additionally, they can use event sequence analysis to further increase the quality of the interpretation [16].
Some results indicate that chatbots which apply humanization techniques provide higher data quality in the context of self-disclosure and social desirability bias [17]. In the higher education context chatbots should have a character, consistent responses, delve into a topic and have some understanding of factors related to a culture [18].
Most of the papers do not describe chatbot data in detail, nor do they provide detailed information about its structure or storing techniques.

2.2 Chatbot and survey data comparisons

Literature review indicated few approaches to examination of chatbot and survey data. Celino and Calegari [19] investigated the effectiveness of a conversational survey in comparison to a traditional questionnaire. Their results showed that users prefer conversational form over a traditional approach and that, from a data collection point of view, the conversational method shows higher response quality with respect to a traditional questionnaire. Yet another perspective is informativeness and clarity of the responses. Xiao et. al. [20] performed research of quality responses measured by Gricean Maxims in terms of their informativeness, relevance, specificity, and clarity. Their conclusions indicated a high level of participant engagement when applying for a chatbot survey. However, they indicated several drawbacks and provided guidelines for creating AI-powered chatbots to conduct effective surveys. Athreya, Ngonga Ngomo and Usbeck [14] introduced the DBpedia Chatbot, a knowledge chatbot developed to optimize interaction. The bot was designed to facilitate the answering of recurrent questions.
Recent paper of Rhim et. al [17] introduced humanization survey chatbot, which is another level of improvement. Authors compared how applying humanization techniques to survey chatbots can affect survey-taking experience in three aspects: respondents' perceptions of chatbots, interaction experience, and data quality. Regarding data quality, authors reported better results in the terms of self-disclosure. Te Pas et al. [21] also compared the user experience of a chatbot questionnaire with a traditional questionnaire.
Literature review showed comparisons of chatbot and traditional survey from different perspectives: response rate, informativeness or relevance. However, we did not find any paper tackling this issue from the perspective of data quality measured and characterized by meta-features.

2.3 Meta-learning and meta-features

Meta-learning is the process of learning from previous experience gained during applying various learning algorithms on different kinds of data and hence reducing the needed time to learn new tasks [22]. Main idea of meta-learning is to exploit the knowledge gained out of previous data analysis experience [23] and to use this experience of previous experiments to learn how to improve automatic learning and recommendation of algorithms. Meta-learning consists of three steps: (i) to establish a meta-learning space using meta-data comprising of meta-features and a performance measure (meta-response) for machine learning mining algorithms particular datasets [23], (ii) developing meta-model out of the meta-dataset constructed in the first phase, (iii) predictive meta-model from second step is used to predict the performance of an algorithm.

Meta-learning success depends on the input, meta-features used to describe the given problem. Finding appropriate meta-features which explain specific tasks well is a basic problem of meta-learning [24]. Vanschoren [25] 40 meta-features grouped into six categories: simple, statistical, information, model-based, and landmarkers. Simple measures are commonly known and easily extracted from data [26]. They are also called general measures [24]. Those measures are [25]: Number of instances, Number of features, Number of classes, Number of missing values, Number of outliers. Statistical meta-features give information about data distribution: average, standard deviation, correlation, and kurtosis. Statistical measures are used only for numerical attributes Those measures are: Correlation, Covariance, Concentration, Skewness, Kurtosis, ANOVA p-value, Coefficient of variation, Sparsity, Gravity, PCA 1, PCA skewness, PCA 95%, Class probability. Information meta-features are from information theory. Information measures are based on entropy, and they are used for categorical attributes. Those measures are Mutual information, Class entropy, Uncertainty coefficient, Equivalent number of features, and Noise-signal ratio. Model based features and landmarkers are specific groups of meta-features which depend on the modeling algorithm used in the data analysis. Since this paper is focused on data characterization, and modeling is not performed, those two groups are not subject of the investigation in this paper.

3 This research

3.1 Research design: research questions and methods

Aim of this paper is to research meta-features of chatbot and survey data. Research design is constructed around the first three steps of the CRISP DM process for data mining. CRISP DM consists of six steps: (i) domain understanding, (ii) data understanding, (iii) data preparation, (iv) modeling, (v) evaluation, (vi) deployment. Domain understanding is described through literature review in section 2. Data understanding and data preparation are presented through research results in section 3. Modeling, evaluation and deployment, refers to model development and usage of the results in real-world scenarios, and those steps are not part of this research.

By using above described methodology and data, following research questions are addressed:

RQ1: What are general and information-based meta-features of chatbot interactions data?

RQ2: What are the differences in general and information-based meta-features between chatbot interactions data and traditional survey data?

3.2 Research data: chatbot data from the EDUBOTS project and the faculty's survey

In order to answer to the aforementioned research questions, we compared data from two different sources: i) a chatbot Hubert which was investigated within the project EDUBOTS [27], and ii) a faculty's survey which is conducted each semester at the Faculty of Organization and Informatics from the University of Zagreb [28].

One of the goals of the project EDUBOTS (Best practices of pedagogical chatbots in higher education) was to document best practices for the use of chatbots in higher education, by introducing two chatbots into university courses: a chatbot BO within the chat application Differ [29] and a chatbot Hubert [30], a web application with integrated AI algorithms developed to automate recruiting process but also for gathering opinions in the form of the survey with open-ended questions.

In our research Hubert was used to provide students with more responsive and entertaining means for collecting their opinions about the university course. Educators at three courses created an evaluation for their course by using a customized educational template in Hubert. The template consisted of several main questions that were asking the following: I) What is working well on (course_name) and should it continue in the same way? II) What could teachers start doing, that would improve it? III) What could teachers stop doing, that would improve (course_name)? IV) What is your overall experience of (course_name)? Please write a sentence or two. V) Do you want to add something more [3].

Students were provided the link to the course evaluation in Hubert and were asked to respond in the form of the sentence in English language. Subquestions were asked from the Hubert side if it did not understand the student's answer (Figure 1).

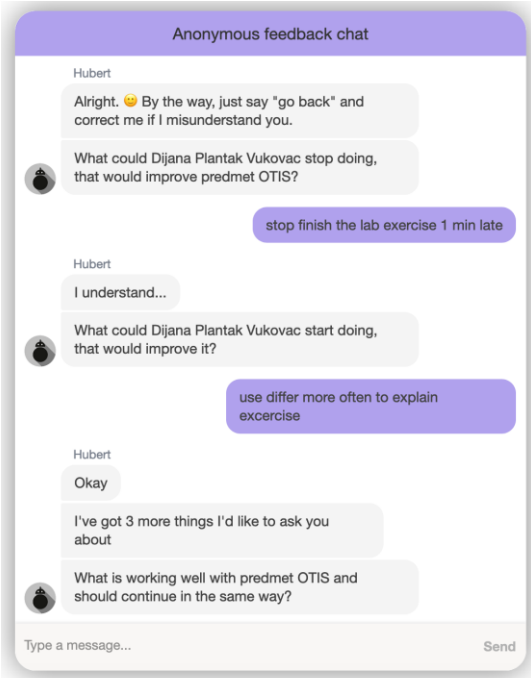

Figure 1. Conversation in Hubert for course evaluation

All conversations related to one course could be downloaded in the form of a transcript in a .txt file. In addition, collected answers are visualized in a Hubert evaluation dashboard, in the form of numerical data and graphs. Data are also classified according to the positive or negative tone of the answer (see Figure 2).

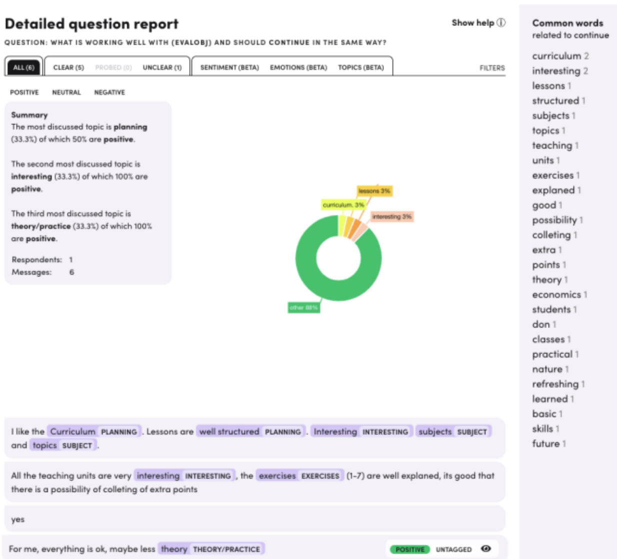

Figure 2. Data visualisation in Hubert

The second set of data is from the faculty's survey about perceptions of the students about the course quality and its delivery in the online environment.

This survey was introduced in summer semester 2019/2020 after all teaching was shifted to the full online in both asynchronous mode (Moodle, video lectures etc.) and synchronous mode (e.g. videoconferences and chats with the students) due to COVID-19 pandemic. The survey is now used regularly in each semester to evaluate every faculty course and its teaching quality, to identify trends and room for improvements.

The survey consists of 27 questions grouped into the following categories: I) course organization and communication, II) teaching materials on LMS, III) knowledge and skills assessment, and IV) delivery of the course. Students rated their opinions about the course on the Likert scale from 1 (totally disagree) to 5 (totally agree) or selected a single or multiple response from a predefined list, but also had the opportunity to provide the answers to four open-ended questions. Figure 3 shows the example of the data visualization of the answers to two questions from the category Course organization and communication.

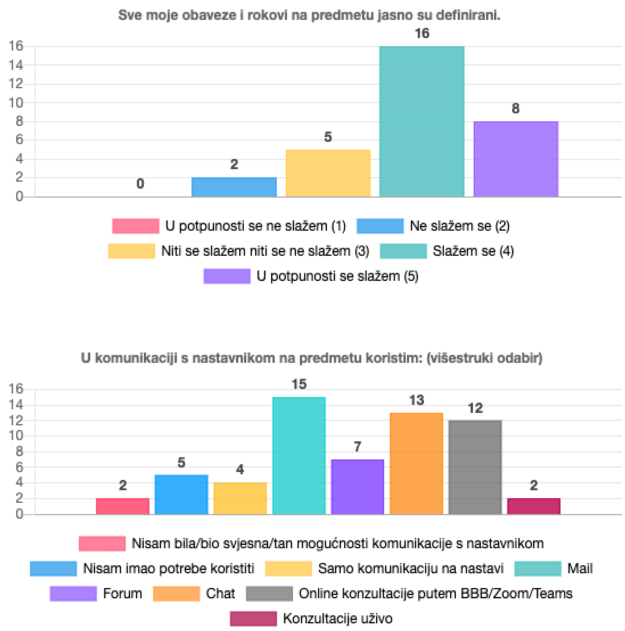

Figure 3. An example of data visualisation in FOI Students surveys

In the section 4, research results are presented with an aim to answer research questions.

4 Results

Meta-learning process consists of the: (i) data collection, (ii) data understanding activities which are focused at detecting data quality. Two datasets are presented in this paper: one collected with the chatbot Hubert within the EDUBOTS project and the second one collected with the classical survey. First dataset consists of chatbots` data with 82 participants, students which were asked to evaluate the course. Second dataset consists of survey data with 50 participants, students which were asked to evaluate the course through a survey questionnaire. Research was conducted at the winter semester of 2020/2021 academic year (survey data) and summer semester (chatbot data) of 2019/2020 academic year. Chatbot data was collected within the following three undergraduate courses at University of Zagreb, Faculty of Organization and Informatics: Software Engineering, Text and Image Editing and Business Informatics. Via forum messages on Moodle, students were asked to give feedback and evaluate the course using chatbot application Hubert.

Survey data was collected within the following two courses at the undergraduate level: Knowledge based systems and Knowledge discovery in data, and one course at the graduate level: Intelligent systems.

On the previously explained data sets, meta-features are extracted. General meta-features include general information related to the dataset at hand. To a certain extent they are conceived to measure the complexity of the underlying problem. Some of them are the number of instances, the number of attributes, dataset dimensionality, the ratio of missing values. Table 1. depicts main meta-features for both datasets.

Table 1. General meta-features for chatbot and survey data

Meta-feature	Chatbot data	Survey data
Number of instances	82	50
Number of attributes	7	17
Number of categorical attributes	7	17
Number of numerical attributes	0	0
Ratio categorical to numerical attributes	0/7	0/17
Number of missing values	15 %	22 %

Statistical meta-features are defined for numerical attributes since those meta-features describe attribute statistics and class distributions of a dataset sample. They include various summary statistics per attribute like mean, standard deviation, class entropy, etc. Hereinafter, information based meta-features are calculated since those features are intended for categorical attributes, and two datasets involved in the research consists of categorical attributes

Table 2. Information based meta-features for chatbot and survey data

Meta-feature	Chatbot data	Survey data
Class entropy	0.43	0.66
Mutual information	0.57	0.34
Uncertainty coefficient	0.26	0.33
Equivalent number of features	1	1
Noise-signal ratio	0.17	0.33

According to information based meta-features, chatbot data gives more relevant information (measured by mutual information) and gives less noise (measured by noise-signal ratio).

5 Discussion

In this section, we provide answers on research questions and discuss our results. General meta-features of chatbot data are: low number of instances and low number of attributes, higher number of categorical attributes than numerical, and low number of missing values (see Table 1, RQ1). In our sample data, both sets had low number of instances and low number of attributes. Information-based meta-features of chatbot data are low level of mutual information, low level of uncertainty coefficient and noise-signal ratio (see Table 1, RQ1). Comparison of chatbot interactions data and traditional survey data yielded differences in number of attributes, number of instances, number of missing values, class entropy, mutual information and noise-signal ratio (see Table 2, RQ2)

In this research, we performed data characterization through meta-features for both human-chatbot interaction and survey data and provided their comparison. Our approach focuses on general and information meta-features with an aim to detect data quality. General and information meta-features show higher data quality for chatbot data. In order to give a broader conclusion, this paper investigated previous comparisons of chatbot and survey data, from various perspectives. Some of the previous research papers on the topic are presented in the second section of the paper. Our results are in line with previous research papers investigating a given topic, but from different perspectives than ours. Celino and Calegari [19] reported higher response quality of chatbot evaluations with respect to a traditional questionnaire. Furthermore, Rhim et. al [17] showed better data quality of chatbot data measured through self-disclosure. Our investigation showed that chatbot data gives more relevant information (measured by mutual information) and gives less noise (measured by noise-signal ratio). Implications of such results provide valuable contributions. Results indicate that chatbots can serve as valuable and quality tool for data collection.

6 Conclusion

Despite recent advances in machine learning, it is still challenging to find appropriate algorithms for data analysis. Especially, with a growing body of data sources emerging every day. Selection of appropriate algorithms is dependent on the meta-features of employed data. Thus, meta-features should be explored and elaborated to characterize a specific domain. In this research, we focused on chatbot data and compared it with traditional sources in the educational domain in order to understand characteristics of datasets. One group of meta-features, general meta-features, were investigated to understand data properties. Our research resulted with the following scientific contributions: (i) identification of meta-features in chatbot data, (ii) comparison of chatbot data meta-features with survey data meta-features, (iii) Since this is only the first part of the research, it has several limitations. We have employed only general and information meta-features into account. Secondly, only two datasets were included in the research. In the future research, there will be a broader number of chatbot datasets to increase the possibility of results generalization. Furthermore, other groups of meta-features will be employed in order to make a reliable base for meta-model development.

References:

[1] Grech, M. 2017. The Current State of Chatbots in 2017. (Apr 2017). https://getvoip.com/blog/2017/04/21/the-current- state- of- chatbots- in- 2017/

[2] Lundqvist, K. O., Pursey, G., & Williams, S. (2013). Design and implementation of conversational agents for harvesting feedback in eLearning systems. In *European Conference on Technology Enhanced Learning* (pp. 617-618). Springer, Berlin, Heidelberg.

[3] Čižmešija, A., Hajdin, G., Oreški, D., Using chatbot for course evaluation in higher education, INTED2021 Proceedings, 1494-1501.

[4] Erikson, M., Erikson, M. G., and Punzi, E. (2016). *Student Responses to a Reflexive Course Evaluation*. Reflective Practice (17:6), Routledge, pp. 663–675. (https://doi.org/10.1080/14623943.2016.1206877).

[5] Tucker, B., Jones, S., and Straker, L. (2008). Online Student Evaluation Improves Course Experience Questionnaire Results in a Physiotherapy Program. *Higher Education Research and Development* (27:3), pp. 281–296. (https://doi.org/10.1080/07294360802259067).

[6] Steyn, C., Davies, C., and Sambo, A. (2019). Eliciting Student Feedback for Course Development: The Application of a Qualitative Course Evaluation Tool among Business Research Students. *Assessment and Evaluation in Higher Education* (44:1), Routledge, pp. 11–24. (https://doi.org/10.1080/02602938.2018.1466266).

[7] Rubin, V. L., Chen, Y., and Thorimbert, L. M. (2010). Artificially Intelligent Conversational Agents in Libraries. *Library Hi Tech* (28:4), pp. 496–522. (https://doi.org/10.1108/07378831011096196

[8] Baha, T. A. I. T., Hajji, M. E. L., Es-Saady, Y., & Fadili, H. (2022). Towards highly adaptive Edu-Chatbot. *Procedia Computer Science*, 198, 397–403.

[9] Wambsganss, T., Winkler, R., Schmid, P., & Söllner, M. (2020a). Unleashing the

Potential of Conversational Agents for Course Evaluations: Empirical Insights from a Comparison with Web Surveys. In *ECIS*.

[10] Wambsganss, T., Winkler, R., Söllner, M., & Leimeister, J. M. (2020b). A Conversational Agent to Improve Response Quality in Course Evaluations. In Extended Abstracts of the *2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-9).

[11] Wang, J., Hwang, G. H., & Chang, C. Y. (2021). Directions of the 100 most cited chatbot-related human behavior research: A review of academic publications. *Computers and Education: Artificial Intelligence*, 100023.

[12] Molina, M.M., Romero, C. Luna, J.M. Ventura. S. Metalearning approach for automatic parameter tuning: A case study with educational datasets. *5th International Conference on Educational Data Mining*, Chania, Greece,180-183, 2012.

[13] Smutny, P., & Schreiberova, P. (2020). Chatbots for learning: A review of educational chatbots for the Facebook Messenger. *Computers & Education*, 151, 103862.

[14] Athreya, R. G., Ngonga Ngomo, A. C., & Usbeck, R. (2018, April). Enhancing Community Interactions with Data-Driven Chatbots--The DBpedia Chatbot. In *Companion Proceedings of the The Web Conference 2018* (pp. 143-146).

[15] Keyner, S., Savenkov, V., & Vakulenko, S. (2019, June). Open data chatbot. In *European Semantic Web Conference* (pp. 111-115). Springer, Cham.

[16] Akhtar, M., Neidhardt, J., & Werthner, H. (2019, July). The potential of chatbots: analysis of chatbot conversations. In 2019 *IEEE 21st Conference on Business Informatics* (CBI) (Vol. 1, pp. 397-404). IEEE.

[17] Rhim, J., Kwak, M., Gong, Y., & Gweon, G. (2022). Application of humanization to survey chatbots: Change in chatbot perception, interaction experience, and survey data quality. *Computers in Human Behavior*, *126*, 107034.

[18] Tsivitanidou, O., & Ioannou, A. (2020). Users' Needs Assessment for Chatbots' Use in Higher Education. In *Central European Conference on Information and Intelligent Systems* (pp. 55-62). Faculty of Organization and Informatics Varazdin.

[19] Celino, I., & Calegari, G. R. (2020). Submitting surveys via a conversational interface: an evaluation of user acceptance and approach effectiveness. *International Journal of Human-Computer Studies*, 139, 102410.

[20] Xiao, Z., Zhou, M. X., Liao, Q. V., Mark, G., Chi, C., Chen, W., & Yang, H. (2020). Tell me about yourself: Using an AI-powered chatbot to conduct conversational surveys with open-ended questions. *ACM Transactions on Computer-Human Interaction* (TOCHI), 27(3), 1-37.

[21] Te Pas, M. E., Rutten, W. G., Bouwman, R. A., & Buise, M. P. (2020). User experience of a chatbot questionnaire versus a regular computer questionnaire: prospective comparative study. *JMIR Medical Informatics*, 8(12), e21982.

[22] Brazdil, P., Giraud Carrier, C., Soares, C., Vilalta, R. (2009). Metalearning: Concepts and Systems (pp. 1–10). https://doi.org/10.1007/978-3-540-73263-1_1

[23] Bilalli, B., Abelló, A., & Aluja-Banet, T. (2017). On the predictive power of meta-features in OpenML. *International Journal of Applied Mathematics and Computer Science*, 27(4), 697–712. https://doi.org/10.1515/amcs-2017-0048

[24] Castiello, C., Castellano, G., & Fanelli, A. M. (2005). Meta-data: Characterization of input features for meta-learning. In *Lecture Notes in Computer Science* (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 3558 LNAI (pp. 457–468). https://doi.org/10.1007/11526018_45

[25] Vanschoren, J. (2018). Meta-Learning: A Survey. http://arxiv.org/abs/1810.03548

[26] Rivolli, A., Garcia, L. P. F., Soares, C., Vanschoren, J., & de Carvalho, A. C. P. L. F. (2018). Characterizing classification datasets: a study of meta-features for meta-learning. http://arxiv.org/abs/1808.10406

[27] EDUBOTS. (2021, January 8) Realizing Chatbots in Higher Education. Retrieved from: www.edubots.eu

[28] FOI Students surveys. (2021) FOI studentske ankete. Retrieved from: https://ankete.foi.hr/

[29] Differ (2021, January 8) For Educators. Retrieved from: www.differ.chat/for-educators

[30] Hubert (2021, October) Hubert.ai, Retrieved from: https://www.hubert.ai/

Contribution of individual authors to the creation of a scientific article (ghostwriting policy)

Dijana Oreški, Dijana Plantak Vukovac and Goran Hajdin collected data, prepared data and analysed data.
Dijana Oreški carried out interpretation of results.
Dijana Plantak Vukovac and Goran Hajdin organized literature review.

Sources of funding for research presented in a scientific article or scientific article itself

This work has been supported by the project "Best practices of pedagogical chatbots in higher education / EDUBOTS" which is funded under the Erasmus+ KA2: Cooperation for innovation and the exchange of good practices – Knowledge Alliances (grant agreement no: 612446, project ref.: 612446-EPP-1–2019-1-NO-EPPKA2-KA).

Creative Commons Attribution License 4.0 (Attribution 4.0 International , CC BY 4.0)