Contextual Ontology-based Feature Selection for Teachers

Nader Nashed, Christine Lahoud, Marie-Hélène Abel

Contextual Ontology-based Feature Selection for Teachers*

Nader N. Nashed[1,2][0000−0001−5849−1322], Christine Lahoud[2][0000−0002−4520−634X], and Marie-Hélène Abel[1][0000−0003−1812−6763]

[1] HEUDIASYC, Université de Technologie de Compiègne, Compiègne, France
{nader.nashed, marie-helene.abel}@utc.fr
[2] Université Française d'Égypte, Cairo, Egypt
{nader.nashat,christine.lahoud}@ufe.edu.eg

Abstract. The context of teacher is indescribable without considering the multiple overlapping contextual situations. Teacher Context Ontology (TCO) presents a unified representation of data of these contexts. This ontology provides a relatively high number of features to consider for each context. These features result in a computational overhead during data processing in context-aware recommender systems. Therefore, the most relevant features must be favored over others without losing any potential ones using a feature selection approach. The existing approaches provide struggling results with high number of contextual features. In this paper, a new contextual ontology-based feature selection approach is introduced. This approach finds similar contexts for each insertion of new teacher using the ontology representation. Also, it selects relevant features from multiple contexts of a teacher according to their corresponding importance using a variance-based selection approach. This approach is novel in terms of representation, selection, and deriving implicit relationships for features in the multiple contexts of a teacher.

Keywords: ontology · feature selection · context · teacher · education · machine learning

1 Introduction

Education is one of the most developed fields in the past few years due to mandatory variables such as digital transformation, mindset of learners, etc.[3]. The quality of education is maintained through new technologies to help all actors in the field. The teacher is considered an essential actor in the field who needs the help of these new educational technologies to enhance the quality of this job. Context-aware recommender systems (CARS) is one of these technologies

* This work was funded by the French Research Agency (ANR) and by the company Vivocaz under the project France Relance - preservation of R&D employment (ANR-21-PRRD-0072-01) in collaboration with project Imhotep "Preventing teachers' psychosocial risks through contextual support of educational resources".

that provide the teacher with personalized recommendations [7]. Data representation is an essential step to achieve such objective as a teacher coexists in multiple contexts such as living context and working context. Ontology is the semantic representation of a formal domain by defining categories, properties, and relations between data. Ontology offers an extensive representation of contextual data within same context, as well as different contexts [4]. Therefore, for each teacher, two types of contexts are considered: representational contexts and interactional context. Representational context is described by social or living context and work context, while interactional context is described by sentimental state. These different types of contexts are represented by Teacher Context Ontology (TCO) [16] and its extension with Mood-Detection Ontology (MDO) and MEMORAe Core Collaboration ontology (MCC) [17]. These ontologies introduce high number of features to describe each context. The most relevant features need to be selected to reduce the computational cost of the contextual data processing.

Recently, there has been a grow in the amount of conducted research towards dimensionality reduction and feature selection to cope with the increasing amount of data in addition to favorability towards computational costs reduction. Dimensionality reduction can be achieved by one or more of these approaches: feature selection and/or feature extraction. Feature extraction approaches obtain new features from the input data, and it is widely used with the image processing applications. On the other hand, feature selection approaches get a reduced subset of the full set of existing features which directly affects the computational cost. The feature selection process enhances the computational complexity by reducing number of features and accordingly, size of data [12, 11]. The elimination of irrelevant features enhances level of data personalization for each teacher. On the other hand, this complicated process has various challenges to achieve the optimal result in terms of type of data and selection criteria.

In this paper, a new ontology-based semi-hybrid feature selection approach is introduced that considers multiple contexts of a teacher and numerous features in each context. The main objective of this approach is summarized as representing the multiple contexts of a teacher with minimum number of features without neglecting relevant ones. Throughout this research, a variance-based feature selection approach is found as a suitable answer to this objective. The semantic representation of data, or ontology, provides a precise description of the multiple contexts. The proposed variance-based approach adapts with various features in different contexts. A final clustering step, using Formal Concept Analysis (FCA), is added to easily find intentions between teachers and extensions between features. The rest of this paper is assembled as follows: Section 2 discusses previous research related to this work in terms of feature selection trends and education-related ones. Section 3 introduces the architecture of the proposed feature selection approach and illustrates the function of ontology in it. Section 4 presents a summary of the obtained results with additional discussion. Section 5 concludes the proposed approach and states future perspectives.

2 Related Work

During this section, we start with an introduction of feature selection algorithms in Section 2.1. Afterwards, we review the most recent approach related to our problem in general in Section 2.2 and education-related ones in Section 2.3. Finally,the reviewed papers are summarized in Section 2.4.

2.1 Feature Selection Algorithms

Feature selection algorithms are classified into supervised, semi-supervised and unsupervised approaches [12, 11]. Supervised selection algorithms need labeled dataset to select the relevant features and they follow wrapper, filter, or intrinsic methodologies by using target variables. Wrapper algorithms extract highest performing features such as recursive feature elimination (RFE) while filtering algorithms use the relationship between features and target variables such as statistical and importance methods. Moreover, the intrinsic algorithms select features automatically as part of the training model such as decision trees and lasso regression. The most significant difference between supervised and semi-supervised algorithms, is the percentage of labeled data that is used throughout the algorithm. On the contrary, labelled datasets are not required for the unsupervised feature selection algorithms and these algorithms are divided into three approaches: filter, wrapper, and hybrid methods [1, 5]. The filtering-based approach evaluates each feature according to the intrinsic features of the provided dataset while wrapper approach is based exclusively on a machine learning algorithm to evaluate all possible combinations according to the evaluation criterion. Wrapper approach has a high computational overhead because it uses a certain clustering method to evaluate each feature. Hybrid approaches try to compromise between both previous approaches as it tries to combine the unbiased methodology and the data over-fitting avoidance of unsupervised selection with the computational efficiency of supervised selection [22].

2.2 Variance-based Feature Selection

Variance-based feature selection is considered as a filtering-based unsupervised algorithm. It eliminates features with variance less than a certain threshold across all samples. In addition, it considers the high variance as an indication to the existence of useful information. The related research to variance-based feature selection is selected using a search process using the following keywords: "education", "variance-based", "feature", and "selection". It was difficult to obtain relevant results in the educational domain. Most of the found existing research is associated with medical, textual, imaging, networking, or manufacturing applications. Therefore, the most relevant research to our proposed approach, is reviewed regardless of the domain of application. Lakshmi and Vishnuvardhan [13] apply Random Forest (RF) algorithm for feature reduction and random subset feature selection (RSFS) algorithm for feature extraction of various cancer

dataset. This approach considers variance of features without setting a threshold depending on their variance which increases the computational cost. Roberts et al. [18] propose a new feature selection approach for colon and lung cancer classification. Features are selected according to differential variance of cancer subtypes classification which is computed using the ratio of variance between the cancer samples and the samples of non-cancerous tumors. Therefore, they choose not to directly use the variance across features as the selection criteria. Sadeghyan [20] combines the usage of sensitivity analysis (SA) technique with extended Fourier amplitude sensitivity test (EFAST) for a variance-based feature selection for medical and biological datasets. The variance of input features is estimated using the contribution of each feature where the threshold is selected once across each dataset without considering any other adaptable options.

The research study by Kamalov et al. [9] and extended in [8], uses an orthogonal variance-based decomposition for feature selection to identify network traffic features for intrusion detection systems (IDS). They use the uniform distribution using variance of traffic features to distinguish DDoS attacks. This approach considers variance of all features only without providing a clear explanation of the threshold setting. Veisi et al. [23] proposes a keyword extraction approach from Persian and English text documents. Keywords are selected according to token weighting depending on the corresponding variance. This research proposes a hybrid approach of variance-based feature selection with Term Frequency Inverse Document Frequency (TF-IDF). The variance of keywords acts as the selection criteria but no insightful detection of a more precise threshold. Mabkhot et al. [15] propose a variance-based manufacturing process feature selection for a decision support system. The proposed approach represents manufacturing process using ontology and introduces a set of reasoning rules to reduce the computational overhead of the algorithm execution at every new insertion. The model considers 9 features only which can be considered as a drawback to this proposition.

2.3 Educational-related Feature Selection

Throughout our search operation, two important reviews are found that demonstrate the major contributions to feature selection techniques in educational data processing. The first review performed by Alyahyan and Düştegör [2] discusses the recent approaches of predicting academic success within the higher education level. The review illustrates the recent methodologies to achieve such goal for the different stages of predication including data collection and selection, data cleaning and preprocessing, feature selection, and data mining for different datasets of students. This research concludes that despite the importance of the usage of either feature selection methods (filtering or wrapper methods), it is better to use the embedded methods with the data mining tools to reduce the computational overhead. The other review introduced by Zaffar et al. [24] presents a state of the art of trends in educational data mining (EDM) but it highlights the classification algorithms and the feature selection algorithms. This research delivers a guide to researchers planning to build an EDM with all

its components. Through the feature selection section, it is observable that the wrapper methods are the most used with a percentage of 78.5% of all reviewed approaches. However, the researchers state that the wrapper methods are not the quickest when used with large datasets. All reviewed approaches target the performance of students and learners and generate useful predictions to help them as well as the educational administration.

2.4 Summary

From the reviewed research, we can state that the importance of using feature selection is summarized into three scenarios: decreasing computational time by selecting the most important features, the combination of feature selection with feature extraction that improves its performance significantly, and finally, the effect of feature selection on the classification accuracy. Multiple approaches follow the two-stage feature selection by applying two feature selection techniques at once. Such approach enhances the efficiency and quality of the selected features. None of these papers discusses the importance of the accurate selection of a threshold or introduces an explicit explanation of the threshold selection and the level of flexibility of this selection with other applications.

3 Methodology

Contexts of a teacher can form a problematic issue when it comes to feature selection methodology as well as high computational time complexity. Additionally, the lack of feature selection approaches for the context of a teacher motivates this research with a semi-hybrid feature selection approach. The proposed contextual ontology-based feature selection approach for teachers, is designed to achieve the following main functions:

i *Map teacher data into ontology representation.* When feature selection process is performed for a new teacher, the teacher data is mapped first into the contextual ontology.
ii *Search for similar teachers based on rules.* Semantic reasoning rules are used to search the ontology for teachers with a similar context as the new teacher.
iii *Select features for a teacher context.* For a new teacher, features are selected based on the context using two filter-based selection methods: information gain importance, and variance-based selection.
iv *Clustering the selected features.* Formal concept analysis is used to cluster the selected features and to construct the corresponding hierarchy for extents and intents which facilitates the future searching process for similar teachers.

The proposed approach consists of ontology representation, information grain importance calculation, variance-based feature selection, and formal concept analysis (FCA) clustering as shown in Fig.1.

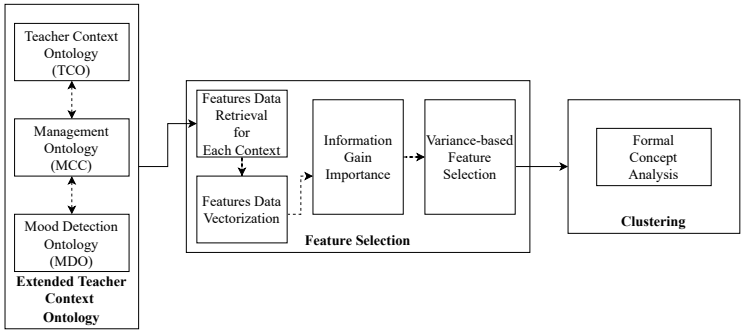

Fig. 1. An overview of the proposed contextual ontology-based feature selection approach for teachers.

3.1 Teacher Context Ontology Representation

The used ontology is based on teacher-context ontology and mood-detection ontology proposed in a previous research regarding the teacher contexts [16, 17]. Protégé[3] is used as the ontology construction tool and DR2Q platform[4] is used as the relational data mapping tool. The teacher contexts representation involves four main concepts: teacher profile, living environment, working environment, and sentimental state, as shown in Fig. 2. Teacher profile involves the main information of a teacher such as name, age, experience, gender, contact details, interests, and competences. Living context is represented by personal address and type of the living area. Working context is accentuated by address and affiliation which forms the connection between the geographical location and educational level of this work environment or the educational institution. Sentimental state of a teacher is described by the mood-detection ontology through three main concepts: mood level, emotional commitment, and flow occurrence. Through this proposition, the mood level is used without considering the other two concepts for the purpose of simplification. The SWRL rule, as shown in Table 1, is applied to any new teacher entry against all saved teachers. This rule checks the similarity of the new teacher with other teachers in terms of experience, teaching style, spoken languages, living environment information, working environment information, and field of science in which a teacher is specialized. The utilization of such rule prevents the computational overhead accompanied by executing the totality of this approach for each new teacher.

3.2 Feature Selection Approach

The proposed feature selection approach consists of 2 filtering algorithms followed by a clustering approach as shown in Fig. 1. We start our approach by

[3] `https://protege.stanford.edu/`
[4] `http://d2rq.org/`

transforming the retrieved list of features from its feature-value form to easy-to-use vectors. Binary one-hot encoding is selected for this task as it constructs one Boolean-valued feature for each possible string value of a single feature [21]. Afterwards, information gain importance is calculated to be used in the threshold calculation for the next step [10, 19]. Variance-based feature selection uses the computed threshold to eliminate the less importance features [6]. Formal concept analysis (FCA) acts a clustering methodology to evaluate the selected features [14].

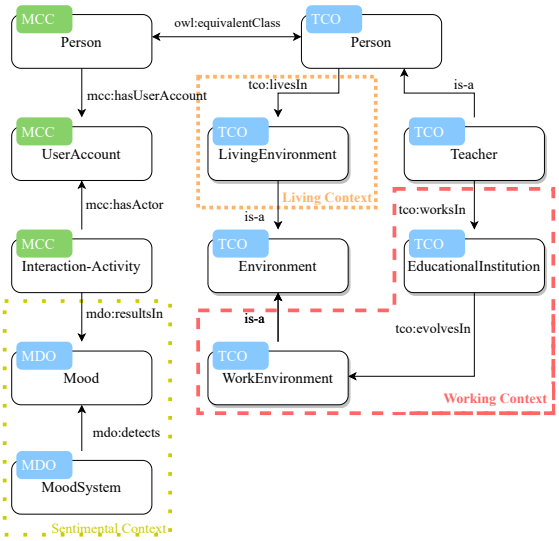

Fig. 2. A partial T-Box representation of the extended ontology.

Table 1. SWRL rule for finding a match of new teacher

Antecedent	Consequence
tco:teacher(?t)^tco:teacher(?ts)^tco:hasYearsOfExperience(?t,?ex)^tco:hasYearsOfExperience(?ts,?exs)^swrlb:greaterThanOrEqual(?exs,?ex)^tco:hasTeachingStyle(?t,?tch)^tco:hasTeachingStyle(?ts,?tchs)^swrlb:stringEqualIgnoreCase(?tch,?tchs)^swrlb:stringEqualIgnoreCase(?tch,"MixedStyle")^tco:hasLanguage(?t,?lan)^tco:hasLanguage(?ts,?lans)^swrlb:contains(?lan,lans)^tco:livesIn(?t,?livenv)^tco:LivingEnvironment(?livenv)^tco:is-a(?livenv,?env)^tco:environment(?env)^tco:hasType(?env,?envtype)^tco:hasCountry(?env,?coun)^owl:country(?coun)^tco:livesIn(?ts,?livenvs)^tco:LivingEnvironment(?livenvs)^tco:is-a(?livenvs,?envs)^tco:environment(?envs)^tco:hasType(?envs,?envtypes)^tco:hasCountry(?envs,?couns)^owl:country(?couns)^swrlb:equal(?envtype,?envtypes)^swrlb:equal(?coun,?couns)^tco:worksIn(?t,?inst)^tco:EducationalInstitution(?inst)^tco:evolvesIn(?inst,?workenv)^tco:WorkingEnvironment(?worknv)^tco:is-a(?workenv,?wenv)^tco:environment(?wenv)^tco:hasType(?wenv,?wenvtype)^tco:hasEducationLevel(?inst,?edulvl)^dcterms:EducationLevel(?edulvl)^tco:worksIn(?ts,?insts)^tco:EducationalInstitution(?insts)^tco:evolvesIn(?insts,?workenvs)^tco:WorkingEnvironment(?worknvs)^tco:is-a(?workenvs,?wenvs)^tco:environment(?wenvs)^tco:hasType(?wenvs,?wenvtypes)^tco:hasEducationLevel(?insts,?edulvls)^dcterms:EducationLevel(?edulvls)^swrlb:equal(?envtype,?envtypes)^swrlb:equal(?edulvl,?edulvls)^tco:hasScience(?t,?sci)^modsci:Science(?sci)^tco:hasScience(?ts,?scis)^modsci:Science(?scis)^swrlb:equal(?sci,?scis)	sameAs(?t,?ts)

Information gain :

Information gain utilizes the mutual information (MI) I between two random features to compute the non-negative value of dependency between these two features. The calculations are based on entropy estimation used in k-nearest neighbor distance estimation as in Eq. 1. After computing the importance of all features, its mean $mean_I$ is calculated to be used as the threshold value for the variance-based feature selection.

$$I(f_1, f_2, \cdots, f_m) = \psi(k) - \frac{(m-1)}{k} + (m-1)\psi(N) - < \psi(n_{f_1}) + \psi(n_{f_2}) + \cdots + \psi(n_{f_m}) > \tag{1}$$

where $f_1, f_2, \cdots, f_m$ are the m features
ψ is digamma function
k is k-nearest neighbours
N is set of features
m is total numner of features
n_{f_i} is boundry limit inwhich all features fall

Variance-based threshold selection :

Variance-based threshold algorithm is baseline approach to selects features with high variances regardless of the desired output with an unsupervised approach. It eliminates all features with a variance that is less than a predetermined threshold thr. In our case, this threshold is calculated using the average importance $mean_I$ as shown in Eq.2. The resulting set of features are used by the formal concept analysis to cluster the provided set of teachers.

$$thr = mean_I(1 - mean_I) \tag{2}$$

Formal Concept Analysis :

Formal concept analysis (FCA) delivers a model to represent a set of objects (teachers) with another set of properties (features) where each object is associated with a subset of these properties. The formal context matrix represents such relationship between teachers and features. Using this matrix, a formal ontology is constructed to describe the hierarchical representation of the relation between each sub-concept and subset of teachers using the theory of lattices.

Each formal context is represented by a set of teachers, a set of features, and incidence, or binary relation, between each teacher and feature. The pairing between a subset of teacher T and a subset of features F is said to be a formal concept if every teacher in T has all features in F, and T and F contain only

a subset of their main sets of teachers and features respectively. This formal concept is visualized into extensions and intensions using lattices as a directed acyclic graph-like representation of concepts.

4 Scenarios and Discussion

When we apply the proposed approach to a vectorized set of context features of a teacher as shown in Fig. 3, the correlation coefficient of all features is computed in the same manner as shown in Fig.4. The illustrated heatmap shows that there are dependencies between features which should be eliminated by the feature selection algorithms. Afterwards, the information gain importance is calculated for all features. For example, the feature number 12 "lang1=English", represents English language as the mother tongue of a teacher, is dependent on several other features, in addition to the obtained low importance using the information gain method. Therefore, this feature can be eliminated by all feature selection methods except for the variance-based threshold method according to the used scoring algorithm. After the above data analysis, the proposed approach can be applied to any dataset of choice. For example, if a mean importance of 0.1, is computed from the information gain algorithm, a threshold of 0.09 is used for the variance-based selection. As shown from the two numbers, the threshold is slightly lower than the mean importance across all features. This threshold selection results in the avoidance of neglecting low-variance but relatively important features such as mood. This approach targets to select at least 78% of the input features without neglecting any relevant ones.

In order to validate the variance selection, we use the formal concept analysis which finds the extensions and intentions for all teachers and selected features. All teacher-feature combinations are tested to create a closed connection which

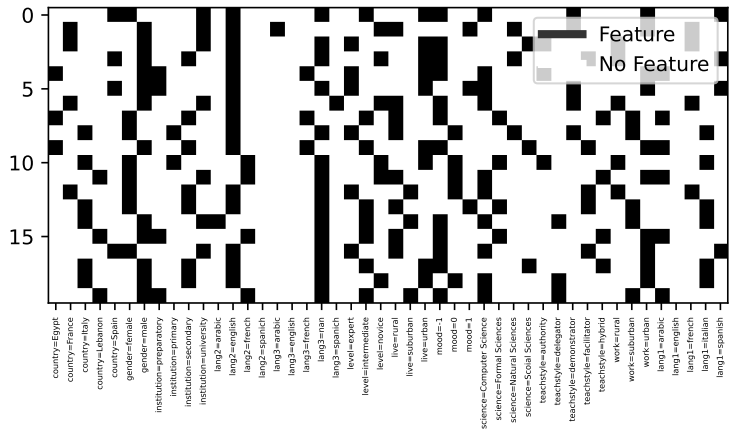

Fig. 3. Vectorized list of features for 20 teachers.

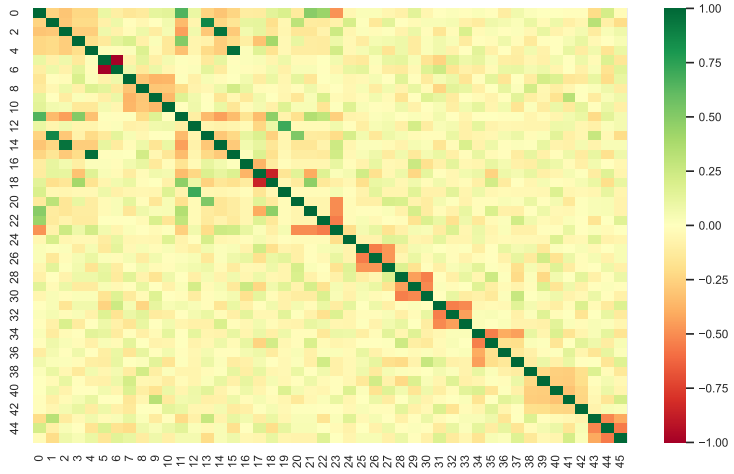

Fig. 4. Correlation coefficient of the selected features.

indicates the correctness of the proposed algorithm. Also, it validates the completeness of connections between concepts for each subset of teachers along with the associated subset of features. The graph illustration highlights the effectiveness of the usage of the information gain importance as the threshold of the variance-based feature selection along with the ontology-based representation. Additionally, this algorithm shows a significant enhancement in the computational time when compared to other techniques due to the usage of mathematical models with low complexity.

The main importance of the used ontology can be summarized in facilitating the reasoning of the SWRL rules that are introduced in Table 1, positioning the expertise level of each teacher, and reduction of the computational overhead corresponding to repetition of the approach at each new teacher insertion into the dataset. Using the SWRL rules reasoning, the insertion of new teachers does not require re-execution of the proposed approach, as at least 10% of teachers share the same context in the used dataset. This observation directs this research to reduce the computational cost of the initial processed data. Some remarks are worthy to be highlighted concerning the selected features by this approach. Mood features are essential during this difficult period and due to the increasing importance and impact of sentimental state on the performance of teachers. The proposed approach considers mood features in the final selection regardless of its relatively moderate importance. The proposed feature selection approach keeps the main formation of all features, while on the contrary, the other methods may have dropped many important features. The usage of formal concept analysis is essential for clustering of teachers, and it can be considered as a more accurate evaluation metric during the application of this approach in a recommender system. FCA clustering is not used for evaluation due to visualization complexity and lack of wide experimentation of this approach with a

recommender system. From the limited performed tests, it is observed that FCA obtains faster computational time and precise intention of teachers with the proposed feature selection approach. On the other hand, it results in low-accuracy intentions while applied to data from other feature selection methods.

5 Conclusion and Perspectives

This research provides an overview of a new contextual ontology-based feature selection approach for teachers. The proposed approach highlights the effect of the combination of ontology knowledge representation of features and their selection in the educational domain, or more specifically, the teacher-context representation. Also, this work follows an indirect merging of two feature selection algorithms into one mathematical model. Although many of the reviewed methods use a hybrid approach, this paper introduces a semi-hybrid approach between two filter-based feature selection without using inter-categorical hybridization. Accordingly, this paper delivers a novel work regarding the educational domain, in general, and context of teachers, specifically.

Despite the level of novelty of this work, more validation is needed to test this approach in a real-time recommender system which may enhance the resulting accuracy and shows the role of FCA clustering. We plan to test it with a dataset that represents the contexts of 100 teachers from different field of sciences, and work at different institution levels: primary education (elementary), preparatory education (middle school), secondary education (high school), and higher education (university, etc.). In this dataset, the contexts of each teacher are represented by 46 features distributed as follows: 28 features for teacher profile, 8 features for living environment context, 7 features for working environment context, and 3 for the sentimental state context. Moreover, it can be used to serve an educational resources recommender system (ERRS) to provide teachers with personalized recommendations using the full representation of their contexts, in addition to the new feature selection approach. Overall, this study is a new step towards the research in diversity of contexts in which a teacher coexists and the utilization of these contexts to enhance the information systems provided to teachers.

References

1. Alelyani, S., Tang, J., Liu, H.: Feature selection for clustering: a review. Data clustering: algorithms and applications **29**(1) (2013)
2. Alyahyan, E., Düştegör, D.: Predicting academic success in higher education: literature review and best practices. International Journal of Educational Technology in Higher Education **17**(1), 1–21 (2020)
3. Cabaleiro-Cerviño, G., Vera, C.: The impact of educational technologies in higher education. GIST Education and Learning Research Journal **20**, 155–169 (2020)
4. Chen, H., Finin, T., Joshi, A.: An ontology for context-aware pervasive computing environments. The knowledge engineering review **18**(3), 197–207 (2003)

5. Dong, G., Liu, H.: Feature engineering for machine learning and data analytics. CRC Press (2018)
6. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. Journal of machine learning research **3**(Mar), 1157–1182 (2003)
7. Haruna, K., Akmar Ismail, M., Suhendroyono, S., Damiasih, D., Pierewan, A.C., Chiroma, H., Herawan, T.: Context-aware recommender system: A review of recent developmental process and future research direction. Applied Sciences **7**(12), 1211 (2017)
8. Kamalov, F.: Orthogonal variance decomposition based feature selection. Expert Systems with Applications **182**, 115191 (2021)
9. Kamalov, F., Moussa, S., El Khatib, Z., Mnaouer, A.B.: Orthogonal variance-based feature selection for intrusion detection systems. In: 2021 International Symposium on Networks, Computers and Communications (ISNCC). pp. 1–5. IEEE (2021)
10. Kraskov, A., Stögbauer, H., Grassberger, P.: Estimating mutual information. Physical review E **69**(6), 066138 (2004)
11. Kuhn, M., Johnson, K.: Feature engineering and selection: A practical approach for predictive models. CRC Press (2019)
12. Kuhn, M., Johnson, K., et al.: Applied predictive modeling, vol. 26. Springer (2013)
13. Lakshmi Padmaja, D., Vishnuvardhan, B.: Variance-based feature selection for enhanced classification performance. In: Information Systems Design and Intelligent Applications, pp. 543–550. Springer (2019)
14. Lindig, C.: Fast concept analysis. Working with Conceptual Structures-Contributions to ICCS **2000**, 152–161 (2000)
15. Mabkhot, M.M., Al-Samhan, A.M., Hidri, L.: An ontology-enabled case-based reasoning decision support system for manufacturing process selection. Advances in Materials Science and Engineering **2019** (2019)
16. Nashed, N.N., Lahoud, C., Abel, M.H.: Tco: a teacher context ontology. In: 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD). pp. 757–762. IEEE (2021)
17. Nashed, N.N., Lahoud, C., Abel, M.H., Andrès, F., Blancan, B.: Mood detection ontology integration with teacher context. In: 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA). pp. 1710–1715. IEEE (2021)
18. Roberts, A.G., Catchpoole, D.R., Kennedy, P.J.: Variance-based feature selection for classification of cancer subtypes using gene expression data. In: 2018 International Joint Conference on Neural Networks (IJCNN). pp. 1–8. IEEE (2018)
19. Ross, B.C.: Mutual information between discrete and continuous data sets. PloS one **9**(2), e87357 (2014)
20. Sadeghyan, S.: A new robust feature selection method using variance-based sensitivity analysis. arXiv preprint arXiv:1804.05092 (2018)
21. Seger, C.: An investigation of categorical variable encoding techniques in machine learning: binary versus one-hot and feature hashing (2018)
22. Solorio-Fernández, S., Carrasco-Ochoa, J.A., Martínez-Trinidad, J.F.: A review of unsupervised feature selection methods. Artificial Intelligence Review **53**(2), 907–948 (2020)
23. Veisi, H., Aflaki, N., Parsafard, P.: Variance-based features for keyword extraction in persian and english text documents. Scientia Iranica **27**(3), 1301–1315 (2020)
24. Zaffar, M., Hashmani, M.A., Savita, K., Khan, S.A.: A review on feature selection methods for improving the performance of classification in educational data mining. International Journal of Information Technology and Management **20**(1-2), 110–131 (2021)