Co-occurrence patterns of issues and guidelines related to ethics and privacy of learning analytics in higher education—literature review

Barbara Šlibar*
barbara.slibar@foi.unizg.hr
University of Zagreb
Faculty of Organization and Informatics
Varaždin, Croatia

Jelena Gusić Munđar
jelena.gusic@foi.unizg.hr
University of Zagreb
Faculty of Organization and Informatics
Varaždin, Croatia

Sabina Rako
sabina.rako@srce.hr
University of Zagreb
University Computing Centre
Zagreb, Croatia
University of Zagreb
Faculty of Organization and Informatics
Varaždin, Croatia

Diana Šimić
diana.simic@foi.unizg.hr
University of Zagreb
Faculty of Organization and Informatics
Varaždin, Croatia

ABSTRACT

Ethics and privacy issues have been recognized as important factors for acceptance and trustworthy implementation of learning analytics. A large number of different issues has been recognized in the literature. Guidelines related to these issues are continuously being developed and discussed in research literature. The aim of this research was to identify patterns of co-occurrence of issues and guidelines in research papers discussing ethics and privacy issues, to gain better understanding of relationships between different ethics and privacy issues arising during implementation of learning analytics in higher education. A total of 93 papers published between 2010 and 2021 were qualitatively analyzed, and nine categories of issues and respective guidelines related to ethics and privacy in learning analytics were identified. Association rules mining Apriori algorithm was applied, where 93 papers represented transactions, and 18 categories of issues or guidelines (nine each) represented items. Two clusters of issues co-occurring in papers were identified, corresponding to deontology ethics (related to rules and duties), and consequentialism ethics (related to consequences of unethical behavior).

*All authors contributed equally to this research.

LAK22, March 21–25, 2022, Online, USA

ACM ISBN 978-1-4503-9573-1/22/03.
https://doi.org/10.1145/3506860.3506974

CCS CONCEPTS

• **Social and professional topics → Privacy policies**; **Privacy policies**; • **Security and privacy → Privacy protections**; **Privacy protections**; • **Information systems → Association rules**; • **Applied computing → Education**.

KEYWORDS

Learning analytics, Ethics, Higher education, Apriori algorithm

ACM Reference Format:
Barbara Šlibar, Jelena Gusić Munđar, Sabina Rako, and Diana Šimić. 2022. Co-occurrence patterns of issues and guidelines related to ethics and privacy of learning analytics in higher education—literature review. In *LAK22: 12th International Learning Analytics and Knowledge Conference (LAK22), March 21–25, 2022, Online, USA*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3506860.3506974

1 INTRODUCTION

With rapid intrusion of digital technologies into every aspect of our lives, issues of discrimination, privacy, security, surveillance, and trust have emerged from the closed professional communities into the limelight of general public interest. Learning analytics (LA) is not an exception. Thus, Griffits [10] argues about the ethics of LA in a historical context, concluding that it is necessary, beside practice reflection, to view LA as a manifestation in the education sector of wider trends which are transforming society, and trying to recognize future trends. Generally speaking, three major ethical theories could be considered when designing ethical approaches to LA: consequentialism, deontology, and virtue ethics [17]. Consequentialism emphasizes the consequences of actions often defining a set of rights and duties whose purpose is to minimize harms and to maximize utility [27]. In contrast, deontological approach relies on normative theories, based on the belief that some choices can not be justified by their effects, and should be morally required, forbidden, or permitted [2]. Virtue ethics emphasizes the development of virtues or moral character with the focus on reasoned tendency to act [14].

On a more pragmatic level, Pardo and Siemens [24] have identified four principles to categorize numerous issues derived from privacy in LA: transparency, student control over the data, security, and accountability and assessment. Even more practical approach, with strong emphasis on privacy and ethical aspects of LA implementation at higher education institutions (HEIs), was creation of the DELICATE checklist by Drachsler and Greller [7] containing eight action points that managers and decision makers should consider when planning an implementation of LA.

Focusing on empirical research on ethical issues in the LA literature, Cerratto Pargman and McGrat [4] have concluded that the top three ethical areas most often addressed in analyzed papers were transparency, privacy, and informed consent. They also noticed that survey was a dominant research strategy with most respondents representing institutional views rather than student's perspective.

With numerous ethics and privacy issues arising in implementation of LA, it is difficult to find a paper addressing all the issues. Many factors can influence the selection of issues discussed in individual papers. Although there are a few review papers on ethical aspects of LA in higher education, most of them are very recent, published during 2021. Therefore, some of them were refereed in online databases after the time of conducting the literature review that was used as the primary source of data for this research [31]. The aims of these reviews were different from the aim of this paper (e.g., to characterize the type of empirical research that has been conducted on ethics in LA, to identify the main ethical areas and knowledge gaps, to identify the possibility of alignment between the LA and General Data Protection Regulation, etc.) [3, 5]. Also, some of them are focused on all education stages, from early childhood stage to higher education [13, 30], while the focus of this paper is on higher education. The goal of this research was to identify combinations of issues that are usually, or more frequently, discussed together. One of the approaches to solving this problem is through association rules mining.

Association rules mining is a data mining technique for finding rules or associations in large numbers of unordered lists of items. Its beginnings are connected to market basket analysis, but it is now used in all areas of human activity, including the education and learning analytics. Thus, Sutch [28] used association rules mining to detect combinations of elective subjects taken by high school students. Liu and Li [20] searched for reasons of MOOC dropouts. Moubayed et al. [22] looked for associations between engagement and performance in e-learning. Khaled et al. [16] applied association rules mining together with item response theory to analyze curricula. Wu and Zeng[35] and Wangand Chung [33] identified associations among students' success in different courses, while Kong et al. [18] combined association rules mining with contrast set mining to deduce psychological features from academic performance. Ma et al. [21] aimed to identify students at financial risk, and Yu et al. [36] identified patterns of co-occurrence of typical student errors in English as foreign language. In a comprehensive review of EDM and LA in higher education by Aldowah et al. [1], association rules mining was identified as a relevant technique, mainly used to identify relationships between students' behaviors, learning materials, and characteristics of performance discrepancy.

This study aims to identify patterns of co-occurrence of ethics and privacy issues, and respective guidelines appearing in research literature on applications of learning analytics in higher education. The patterns can help identify gaps in research and inform decisions on mitigation of these issues in practice. Therefore, the proposed research questions are: 1) Which ethical and privacy issues and guidelines are frequently addressed together, i.e. what are patterns of their co-occurrence in research literature? and 2) What was the dynamics of appearance of ethical and privacy issues and guidelines in research literature?

The paper is structured as follows. The importance of the topic on ethical and privacy issues of LA in HEIs is emphasized in Section 1. The research design is described in detail in Section 2. Results of the research are presented in Section 3 and discussed within Section 4. Summarized findings are presented in Section 5.

2 RESEARCH METHODOLOGY

Our research design follows Knowledge Discovery in the Databases process (KDD) [8]. The KDD is an iterative process for discovering useful knowledge from data, applying a data mining method. It consists of five steps: selection, preprocessing, transformation, data mining, and interpretation/evaluation. The first four steps are described in the following subsections, and the last step is covered by sections: Results, Discussion, and Conclusions.

2.1 Selection, Preprocessing and Transformation

The data used for analysis were extracted from a literature review dataset published in open research data portal Harvard Dataverse [31, 32]. The dataset contains information about 93 research papers mentioning ethical and privacy issues, or guidelines for ethically correct application of learning analytics in higher education. For each paper, three sets of variables were recorded. The first set of 12 variables comprises bibliographic information. Ethical and privacy issues, identified by qualitative analysis of the papers, were classified into nine categories described in Table 1. The second set of variables contains information on the presence/absence of these nine categories of ethics and privacy issues. The final variable contains a comma-separated list of guidelines for dealing with these issues that were proposed or discussed in each study. The subset selected for data mining contains 11 variables (paper id, one variable per issue category, and one variable with guidelines).

Input variables were preprocessed to create a list of vectors containing issues and guidelines present in the respective papers. The preprocessing included standard descriptive statistics and visualizations. Finally, the list was transformed into a transactions object of the R package arules [12].

2.2 Data Mining

Association rules mining, also known as association analysis, or frequent itemset analysis, is a widely accepted data mining method. The association rules mining was conceived for the analysis of consumer buying behavior. Hence, association rules mining terminology reflects customer transactions (i.e., baskets) that consist of a specific set of items (i.e., products). The association rules mining

Table 1: Abbreviations and descriptions of the categories of issues.

Abbreviation	Description of the category
Access	Storing, using and accessing learning analytics data issues
ConfPriv	Learning analytics data confidentiality, privacy and transparency issues
Discr	Prejudice, discrimination and other psychological issues
Educ	Awareness and educational issues of learning analytics usage
GovPol	Lack of learning analytics governance models and policies
Misuse	Data misuse, surveillance and data-profiting issues
OwnSha	Learning analytics data ownership and sharing issues
Power	Unequal power relation and manipulation issues
Trust	Mutual trust issues

is based on the market-basket model of data that depicts relationships between two types of objects (items, transactions) with the following assumptions: each transaction contains a set of items (i.e., itemset); the number of items in a transaction is usually assumed to be small (smaller than the number of observed items); the number of transactions is usually assumed to be large; the data is assumed to be depicted by a sequence of transactions in a file [19]. The idea is to extract itemsets that occur in many transactions, represented as a corpus of association rules. An association rule $\{X \rightarrow Y\}$ consists of a left hand side (LHS) itemset $\{X\}$ and a right hand side (RHS) item $\{Y\}$ [12]. It implies that if all items in $\{X\}$ occur in a particular transaction, then conditional probability of $\{Y\}$ appearing within the same transaction is high. The total number of items in both LHS and RHS is the length of a rule. In our research each paper represents a transaction, and categories of issues/guidelines represent items. Thus, there are 93 transactions, and 18 items (nine issues, and respective guidelines).

Usually, the association rules mining produces a large corpus of rules. In order to find interesting or relevant rules, numerous measures of quality or interestingness of rules were proposed [29]. The most often used measures are support, confidence, coverage, and lift [19].

Support is an indication of how frequently a set of items (issues/guidelines) appears together in the set of transactions (papers).

$$Support\{X \rightarrow Y\} = \frac{frequency(X, Y)}{N},$$

where N is the total number of transactions (papers). Coverage is defined as the support of the left hand side of a rule [12]. It is used together with support to assess the relevance of a rule, or here, to how many papers the rule can be applied.

Confidence is a conditional frequency of the RHS, conditioned on the presence of items from the LHS in a transaction [12].

$$Confidence\{X \rightarrow Y\} = \frac{frequency(X, Y)}{frequency(X)} = \frac{Support\{X \rightarrow Y\}}{Support(X)}$$

Lift is a measure of association between the $\{X\}$ and $\{Y\}$, computed as a ratio of the observed frequency of both $\{X\}$ and $\{Y\}$ appearing within the same transaction, divided by the expected frequency under the assumption of their independence [12].

$$Lift\{X \rightarrow Y\} = \frac{Support\{X \rightarrow Y\}}{Support(X) \cdot Support(Y)} = \frac{Confidence\{X \rightarrow Y\}}{Support(Y)}$$

Another important concept for interpretation of rules is redundancy. For two rules $\{X1 \rightarrow Y\}$ and $\{X2 \rightarrow Y\}$ where $X1 \subset X2$ the rule $\{X1 \rightarrow Y\}$ is considered more general. A rule $\{X2 \rightarrow Y\}$ is redundant if there is a more general rule $\{X1 \rightarrow Y\}$ with equal or higher confidence.

The association rules mining and visualization were carried out in R and RStudio using packages arules, arulesViz, igraph, and ggplot2 [6, 11, 12, 25, 26, 34]. The rules were extracted using the Apriori algorithm. The rules were restricted to minimal support of 0.15, and minimal confidence of 0.8, without restrictions on the rule length. Rules that had a RHS item with support higher than the chosen confidence (i.e., higher than 0.8) were also removed. Such items appear very frequently in transactions, and therefore such rules are not very relevant. Redundant rules were also identified and removed. The resulting association rules were visualized as a directed graph. Vertices represent items and rules (items as rectangles, rules as circles). Arcs connect items of the LHS to rules, and rules to their respective RHS items. Circle size is proportional to the rule confidence, and higher color intensity represents higher rule lift. Color palette in the area graph is the palette "Paired" from R package RColorBrewer, specifically designed to be color blind friendly [23].

3 RESULTS

The most frequent issue mentioned in the analyzed papers was ConfPriv. It appeared in 82 (88.2%) papers. Other issues appearing in more than 50% papers were Access (53 or 57.0%), OwnSha (48 or 51.6%), and Misuse (47 or 50.5%). The most frequent guidelines mentioned in the analyzed papers were guidelines on ConfPriv (G.ConfPriv) which appeared in 37 (24.5%) papers, followed by G.Educ (36 or 23.8%). The third most frequently mentioned guidelines were G.Access (26 or 17.2%). Other guidelines appeared in less than 10% of papers.

Apriori algorithm resulted in 95 rules. After removing rules with ConfPriv on the right hand side there were 20 rules left. Finally, after removing seven redundant rules, there were 13 rules to interpret. There was one rule with length two, nine rules with length three, and three rules with length four.

Table 2 shows the extracted rules. Minimal support and confidence were chosen in advance, and the resulting coverage was also satisfactory, i.e., above 0.16 for all rules. The value of lift for all rules was relatively high, ranging between 1.42 and 2.42. For rules with issues on the RHS, values of lift were below two, while for rules with guidelines on the RHS, lift was above two. This means that for all rules, the conditional frequency of the RHS conditioned on the LHS was much higher than the unconditional frequency of the RHS. Thus, the strength of association was high for all rules, and stronger for guidelines than for issues. Only five items appeared on the RHS, three of them were issue categories, and two were categories of guidelines. Five rules described associations for Access. Issues of Access were more frequently discussed in papers that discussed Trust, or Discr and Educ, or Educ and GovPol, or Educ and Misuse and OwnSha, or Misuse and guidance on ConfPriv (G.ConfPriv). The rule $\{Trust \rightarrow Access\}$ had the confidence of 1, meaning that every paper that discussed Trust, also discussed Access. Guidelines on ConfPriv were the only guidelines appearing on the LHS in these rules.

Misuse appeared on the RHS of three rules. Issues of Misuse appeared more frequently in papers that discussed GovPol and OwnSha, or Access and Power, or ConfPriv and GovPol and OwnSha. OwnSha appeared on the RHS of only one rule, with ConfPriv and GovPol and Misuse on the LHS.

Two guidelines appeared on the RHS: G.ConfPriv and G.Educ. G.ConfPriv appeared in three rules. It appeared more frequently in papers that also discussed G.Access and G.Educ, or G.Educ and Misuse, or G.Educ and Power. G.Educ appeared on the RHS of only one rule, with Discr and G.ConfPriv on the LHS.

Graph representation of rules presented in Figure 1 enables easy identification of issues and guidelines that only appear on the LHS or RHS of rules, comparing rules by confidence and lift, and detecting groups of issues or guidelines that are connected by common rules. Lift is usually interpreted as reflecting the strength of association. Thus, associations among guidelines may be seen as stronger than associations among issues. However, one should take into account that lift is bounded above by the inverse of the frequency of the RHS. Since frequencies of issues appearing in the rules are around 50%, their values of lift are bounded above by two. On the other hand, values of lift for the guidelines are bounded above by four. Therefore, it is meaningful to compare values of lift among rules that have similar frequency of the RHS. In our case that means among the issues or among the guidelines. Within both these groups of RHSs the values of lift are similar.

After identifying patterns of co-occurrence, it was interesting to visualize the timing of appearance of issues and guidelines. Figure 2 shows area plots of absolute and relative frequencies of issues (in the left) and guidelines (in the right) by year of publication. For both, issues and guidelines, the number of papers grows exponentially, with an extra peak in 2016. The number of papers published in 2021 was smaller, because the year had not ended yet, and papers that were already published may not have been referenced in databases. The shape of the growth curve is the same for issues and guidelines, however the guidelines were discussed three times less often than the issues. Among the most frequent categories of issues, confidence and privacy, ownership and sharing, misuse, and education demonstrated growth in the number of papers, and a constant share of papers. The issue category access, on the other hand, was constantly present with some papers, but the share of papers was slowly decreasing. Out of the three most frequently discussed guidelines categories, guidelines on confidence and privacy showed growth of papers and a constant share. Guidelines on education were the first to appear, but were constantly present only after 2016. And category guidelines on access showed the same pattern as the category access issues—constant presence, but slow diminishing of the share of papers.

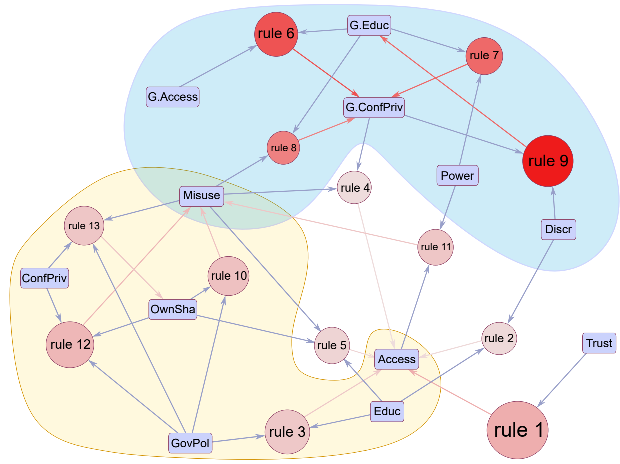

Figure 1: Directed graph representation of extracted association rules. Size of rule vertices is proportional to confidence, and color intensity is proportional to lift.

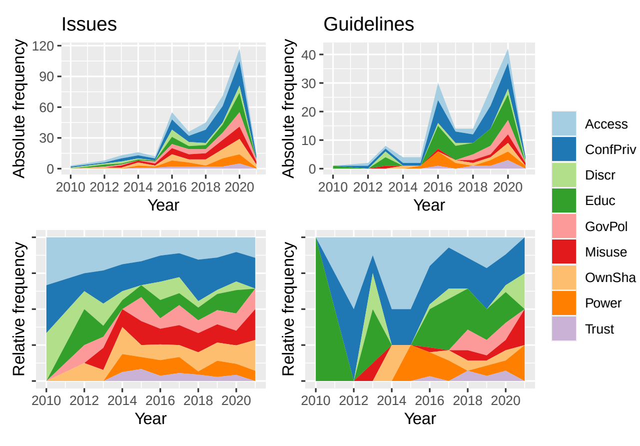

Figure 2: Absolute and relative frequencies of issues and guidelines by year of publication.

4 DISCUSSION

Cerrato Pargman and McGrat [4] identified transparency, privacy, and informed consent as the top three ethical areas addressed in papers they review. This is consistent with our finding that issues

Table 2: Association rules for issues and guidelines. Abbreviations starting with G. indicate guidelines for the category of issues indicated by the subsequent text.

LHS	RHS	Support	Confidence	Coverage	Lift	N
{Trust}	{Access}	0.161	1.000	0.161	1.755	15
{Discr,Educ}	{Access}	0.151	0.824	0.183	1.445	14
{Educ,GovPol}	{Access}	0.183	0.895	0.204	1.570	17
{G.ConfPriv,Misuse}	{Access}	0.183	0.810	0.226	1.420	17
{Educ,Misuse,OwnSha}	{Access}	0.172	0.842	0.204	1.478	16
{G.Access,G.Educ}	{G.ConfPriv}	0.151	0.875	0.172	2.199	14
{G.Educ,Power}	{G.ConfPriv}	0.172	0.842	0.204	2.117	16
{G.Educ,Misuse}	{G.ConfPriv}	0.172	0.800	0.215	2.011	16
{Discr,G.ConfPriv}	{G.Educ}	0.161	0.938	0.172	2.422	15
{GovPol,OwnSha}	{Misuse}	0.194	0.818	0.237	1.619	18
{Access,Power}	{Misuse}	0.172	0.800	0.215	1.583	16
{ConfPriv,GovPol,OwnSha}	{Misuse}	0.194	0.857	0.226	1.696	18
{ConfPriv,GovPol,Misuse}	{OwnSha}	0.194	0.818	0.237	1.585	18

of confidence and privacy, the category containing transparency and privacy, were the most often mentioned issues.

Two clusters of issues / guidelines were identified in Figure 1. The first cluster was formed by the four rules in the upper right (rules 6–9) and the associated issues and guidelines. Issues of misuse, discrimination, and power appear in the left hand sides of these rules. Papers discussing issues of power or misuse in combination with guidelines on education are more likely to also include guidelines on confidence and privacy. Papers discussing the issues of discrimination and guidelines on confidence and privacy are more likely to also discuss guidelines on education. These are the only association rules involving guidelines, i.e. normative texts, thus they can be seen as representing papers taking predominantly a position of deontology branch of ethics.

The second cluster was formed by the four rules at the bottom left (rules 3, 10, 12, and 13). All these rules had issues of governance and policy on the left hand side. Papers discussing issues of governance and policy were more likely to also discuss issues of confidence and privacy, misuse, ownership and sharing, and access. These papers are more focused on consequences of unethical behavior, and can be interpreted as taking the position of the consequentialist branch of ethics.

All papers discussing the issue of trust also discussed the issue of access (rule 1 with confidence of 1). Trust issues can be view through the lens of virtue ethics.

We have mapped categories of issues and guidelines to the three ethical theories recognized by Kitto and Knight [17] as relevant for learning analytics. However, there are other approaches to ethics that might also be applied in this context. For instance, Fukuda-Parr and Gibbons [9] suggest that human rights provide a robust framework for assessment of emerging guidelines on ethical artificial intelligence. Categories of issues in learning analytics related to confidence and privacy, access, ownership and sharing may be viewed through the lens of the rights ethics. On the other hand, the rights-based approach to morality was criticized by Kapoor [15] as promoting selfishness, and going "against the ethics of care, compassion, benevolence and solidarity" by promoting individual interests at the expense of social solidarity. These opposing views provide support for Kitto and Knight's [17] conclusion that combining different views and ethics theories would be useful for practical application of ethical learning analytics.

Regarding the dynamics of appearance of ethical and privacy issues and guidelines in research literature, there were 11 papers addressing ethical issues published in 2016, the same as the total number of papers on ethical issues published between 2010 and 2015. The categories of issues appearing in the earliest papers were those related to access, confidence and privacy, and discrimination, issues already discussed within business analytics, and applied to the education sector. Issues appearing later were those more specific to learning analytics, reflecting maturing of the field. All nine categories of issues were addressed already during the early period until 2015. It is interesting to note the peak in the number of papers published in 2016. It is due to the eight papers published in the special issue of the journal Educational Technology Research and Development with the topic Exploring the Relationship of Ethics and Privacy in Learning Analytics and Design: Implications for the Field of Educational Technology.

The limitation of this research is that it is based on data generated for a literature review with different objectives and research questions. Thus, mapping of issues and guidelines into categories did not involve identification of ethical theories underlying the reviewed papers. The methodology used in the paper is exploratory in its nature—aiming to detect patterns in data. An independent study should validate interpretation of the detected clusters of issues and guidelines.

5 CONCLUSION

The use of association rules mining enabled identification of patterns of co-occurrence of issues and guidelines that were interpreted as reflecting deontological (patterns including guidelines) and consequentialism (patterns including issues) ethical theories. When deliberating ethical issues people tend to base their arguments on one of the ethical theories, e.g. deontology, consequentialism, virtue or rights. We propose that approaching the issues of ethics and privacy in learning analytics combining the points of view of different ethical theories would be beneficial. Structuring guidelines and communication about ethical issues in learning analytics based on combination of different ethical theories may broaden our understanding of different ethical issues, and increase buy-in from a wider range of stakeholders.

ACKNOWLEDGMENTS

This work has been fully supported by the Croatian Science Foundation under the project IP-2020-02-5071.

REFERENCES

[1] Hanan Aldowah, Hosam Al-Samarraie, and Wan Mohamad Fauzy. 2019. Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics* 37 (2019), 13–49.

[2] Larry Alexander and Michael Moore. 2021. Deontological Ethics. In *The Stanford Encyclopedia of Philosophy* (Winter 2021 ed.), Edward N. Zalta (Ed.). Metaphysics Research Lab, Stanford University, Stanford, CA.

[3] Rafael Vescovi Bassani and Silvio Cesar Cazella. 2021. The Alignment between Learning Analytics and the General Data Protection Regulation: A Systematic Literature Review. *ETD Educacao Tematica Digital* 23, 4 (2021), 1022–1040. https://doi.org/10.20396/etd.v23i4.8658829

[4] Teresa Cerratto Pargman and Cormac McGrath. 2021. Mapping the Ethics of Learning Analytics in Higher Education: A Systematic Literature Review of Empirical Research. *Journal of Learning Analytics* 8 (2021), 123–139. Issue 2. https://doi.org/10.18608/jla.2021.1

[5] Teresa Cerratto Pargman and Cormac McGrath. 2021. Mapping the Ethics of Learning Analytics in Higher Education: A Systematic Literature Review of Empirical Research. *Journal of Learning Analytics* 8, 2, SI (2021), 123–139.

[6] Gabor Csardi and Tamas Nepusz. 2006. The igraph software package for complex network research. *InterJournal* Complex Systems, Article 1695 (2006), 9 pages. https://igraph.org

[7] Hendrik Drachsler and Wolfgang Greller. 2016. Privacy and Analytics: It's a DELICATE Issue a Checklist for Trusted Learning Analytics. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (LAK '16)*. Association for Computing Machinery, New York, NY, USA, 89–98. https://doi.org/10.1145/2883851.2883893

[8] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. 1996. From Data Mining to Knowledge Discovery in Databases. *AI Magazine* 17 (1996), 37–54. Issue 3. https://doi.org/10.1609/aimag.v17i3.1230

[9] Sakiko Fukuda-Parr and Elizabeth Gibbons. 2021. Emerging Consensus on 'Ethical AI': Human Rights Critique of Stakeholder Guidelines. *Global Policy* 12, S6 (2021), 32—-44. https://doi.org/10.1111/1758-5899.12965

[10] Dai Griffiths. 2020. The Ethical Issues of Learning Analytics in Their Historical Context. In *Radical Solutions and Open Science: An Open Approach to Boost Higher Education*, Daniel Burgos (Ed.). Springer, Singapore, 39–55. https://doi.org/10.1007/978-981-15-4276-3_3

[11] Michael Hahsler. 2017. arulesViz: Interactive Visualization of Association Rules with R. *The R Journal* 9, 2 (2017), 163–175. https://doi.org/10.32614/RJ-2017-047

[12] Michael Hahsler, Bettina Grün, and Kurt Hornik. 2005. arules—A Computational Environment for Mining Association Rules and Frequent Item Sets. *Journal of Statistical Software* 14, 15 (2005), 1–25. https://doi.org/10.18637/jss.v014.i15

[13] Laura Hakimi, Rebecca Eynon, and Victoria A. Murphy. 2021. The Ethics of Using Digital Trace Data in Education: A Thematic Review of the Research Landscape. *Review of educational research* 91, 5 (2021), 671–717. https://doi.org/10.3102/00346543211020116

[14] Rosalind Hursthouse and Glen Pettigrove. 2018. Virtue Ethics. In *The Stanford Encyclopedia of Philosophy* (Winter 2018 ed.), Edward N. Zalta (Ed.). Metaphysics Research Lab, Stanford University, Stanford, CA.

[15] Rajesh Kapoor. 2019. What Is Wrong with a Rights-based Approach to Morality? *Journal of National Law University Delhi* 6, 1 (2019), 1—-11. https://doi.org/10.1177/2277401719870004

[16] Hayette Khaled, Pablo Gregori, and Raphaël Couturier. 2019. Analysis of bachelor's degree curricula through Item Response Theory and Association Rules. *KEDI Journal of Educational Policy* 16, 1 (2019), 3–20.

[17] Kirsty Kitto and Simon Knight. 2019. Practical ethics for building learning analytics. *British Journal of Educational Technology* 50, 6 (2019), 2855–2870. https://doi.org/10.1111/bjet.12868 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/bjet.12868.

[18] Jie Kong, Jiaxin Han, Junping Ding, Haiyang Xia, and Xin Han. 2020. Analysis of students' learning and psychological features by contrast frequent patterns mining on academic performance. *Neural Computing and Applications* 32, 1 (2020), 205–211. https://doi.org/10.1007/s00521-018-3802-9

[19] Jure Leskovec, Anand Rajaraman, and Jeffrey David Ullman. 2020. *Mining of Massive Datasets* (3rd edition ed.). Cambridge University Press, New York, NY.

[20] Tian-yi Liu and Xiu Li. 2017. Finding out Reasons for Low Completion in MOOC Environment: An Explicable Approach Using Hybrid Data Mining Methods. In *DEStech Transactions on Social Science, Education and Human Science (meit)*. DEStech Publications, Inc., Chongqing, China, 376–384. https://doi.org/10.12783/dtssehs/meit2017/12893

[21] Yong Ma, Xingxu Zhang, Xiaoqiang Di, Tao Ren, Huamin Yang, and Binbin Cai. 2020. Analysis and Identification of Students with Financial Difficulties: A Behavioural Feature Perspective. *Discrete Dynamics in Nature and Society* 2020, Article 2071025 (2020), 18 pages. https://doi.org/10.1155/2020/2071025

[22] Abdallah Moubayed, MohammadNoor Injadat, Abdallah Shami, and Hanan Lutfiyya. 2018. Relationship Between Student Engagement and Performance in E-Learning Environment Using Association Rules. In *2018 IEEE World Engineering Education Conference (EDUNINE)*. IEEE, Buenos Aires, 1–6. https://doi.org/10.1109/EDUNINE.2018.8451005

[23] Erich Neuwirth. 2014. *RColorBrewer: ColorBrewer Palettes.* https://CRAN.R-project.org/package=RColorBrewer R package version 1.1-2.

[24] Abelardo Pardo and George Siemens. 2014. Ethical and privacy principles for learning analytics. *British Journal of Educational Technology* 45, 3 (2014), 438–450. https://doi.org/10.1111/BJET.12152

[25] R Core Team. 2021. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

[26] RStudio Team. 2021. *RStudio: Integrated Development Environment for R.* RStudio, PBC, Boston, MA. http://www.rstudio.com/

[27] Walter Sinnott-Armstrong. 2021. Consequentialism. In *The Stanford Encyclopedia of Philosophy* (Fall 2021 ed.), Edward N. Zalta (Ed.). Metaphysics Research Lab, Stanford University, Stanford, CA.

[28] Tom Sutch. 2015. *Using association rules to understand subject choice at AS/A level.* Cambridge Assessment Research Report. Cambridge Assessment, Cambridge, UK.

[29] Pang-Ning Tan, Vipin Kumar, and Jaideep Srivastava. 2002. Selecting the right interestingness measure for association patterns. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '02*. ACM Press, Edmonton, Alberta, Canada, 32–41. https://doi.org/10.1145/775047.775053

[30] Dimitrios Tzimas and Stavros Demetriadis. 2021. Ethical issues in learning analytics: a review of the field. *ETR&D—Educational Technology Research and Development* 69, 2 (2021), 1101–1133.

[31] Barbara Šlibar, Luka Zlatić, and Nina Begičević Ređep. 2021. Ethical and Privacy Issues of Learning Analytics in Higher Education. In *ICERI2021 Proceedings (14th annual International Conference of Education, Research and Innovation)*. IATED, Online Conference, 3064–3074. https://doi.org/10.21125/iceri.2021.0761

[32] Barbara Šlibar, Luka Zlatić, and Nina Begičević Ređep. 2021. *Ethical issues, privacy issues, and guidelines for ethically correct application of learning analytics in higher education.* Harvard Dataverse. https://doi.org/10.7910/DVN/WYENKU

[33] Liwen Wang and Soo-Jin Chung. 2021. Sustainable Development of College and University Education by use of Data Mining Methods. *International Journal of Emerging Technologies in Learning (iJET)* 16, 5 (2021), 102–115. https://doi.org/10.3991/ijet.v16i05.20303

[34] Hadley Wickham. 2016. *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York. https://ggplot2.tidyverse.org

[35] Xiaodong Wu and Yuzhu Zeng. 2019. Using Apriori Algorithm on Students' Performance Data for Association Rules Mining. In *Proceedings of the 2nd International Seminar on Education Research and Social Science (ISERSS 2019)*. Atlantis Press, Kuala Lumpu, Malaysia, 403–406. https://www.atlantis-press.com/article/125910647

[36] Zhengwang Yu, Rongheng Lin, Ke Song, and Fangchun Yang. 2021. A Framework for Detecting and Summarizing Students' Typical Errors in English Teaching. In *2021 9th International Conference on Information and Education Technology (ICIET)*. IEEE, Okayama, Japan, 268–273. https://doi.org/10.1109/ICIET51873.2021.9419616