Statistical Analyses of Learning Metrics in High School: a Study for Personalized Feedback Purposes

Maely Moraes[1], Gabriel Leitão[2], Áurea Melo[3], Raimundo Barreto[4]

[1]Departamento de Ciência da Computação – Universidade Federal de Roraima (UFRR)

[2]Campus Manacapuru – Instituto Federal do Amazonas (IFAM)

[3]Escola Superior de Tecnologia – Universidade do Estado do Amazonas (UEA)

[4]Instituto de Computação – Universidade Federal do Amazonas (UFAM)

maely.moraes@ufrr.br, gabriel.leitao@ifam.edu.br, asmelo@uea.edu.br, rbarreto@icomp.ufam.edu.br

Abstract. *Improving the teaching-learning process depends on the availability of useful information for decision-making. However, considering the sizeable increase in data generated by educational platforms, performing statistical analyses which can support proper pedagogical choices for teachers and personalized guidance for students has become a real challenge. This paper is in this context, where the main goal is understanding the data and learning metrics and identifying important information about the studied sample to support personalized feedback in a timely manner, in addition to subsequent quantitative analysis. Thus, this article presents an exploratory analysis of interaction data on an educational platform while applying preparatory simulations for the entrance exam in high school classes. The methodology adopted consisted of organizing and tabulating the data, generating information from graphs, calculating statistics on the variables of interest, and interpreting them. The results indicated the most relevant metrics to predict the student's situation at the end of the course besides the Traditional Score (TS). In addition, feedback examples were proposed based on the identified scenarios from the outcomes of the statistical analysis and learning metrics that aim to expand the evaluative elements beyond right and wrong.*

1. Introduction

According to Leitão (2017), the school must integrate technology resources into the teaching processes, proposing new methodologies and pedagogical practices to fulfill its social function more effectively. In this sense, Costa and Mattos (2016) and de Macêdo Santiago et al. (2016) state that technology integration in educational environments has enabled the construction of new tools that collaborate with the teaching process and the evaluation of student performance. An example of tools widely used nowadays are educational content management and distribution platforms. They are commonly used in Distance Education as they enable complete classroom management and the integration of various media and learning resources. Such computational systems are called Virtual Learning Environments (VLE) [Leitão 2017]. However, using VLEs in high school is not widespread yet [Maier and Klotz 2022]. Possibly, for this reason, there are few published studies on applying analytical intelligence to high school students' interaction data in VLEs to understand useful scenarios and patterns to support decision-making [Maier and Klotz 2022; Sousa et al. 2021].

Since the decline in motivation and performance is common among high school students when they move from elementary school to high school, Koenka and Anderman (2019) claim that personalized feedback is essential to improve the performance of these students. However, searching and extracting useful information to bring forth feedback takes time and effort that teachers and students can use to focus on the teaching-learning process. Therefore, the main objective of this article is to study a sample of interaction data of high school students in a Virtual Learning Environment while applying preparatory questionnaires for the entrance exam to identify helpful information for creating personalized feedback and for future predictive and prescriptive analytics. Briefly, the main contributions of this article are: (1) the identification of some variables that are most relevant to predicting students' future performance; (2) the proposition of hypotheses to be investigated about some variables that may or may not influence student success; and, (3) the detection of different usage scenarios to provide timely personalized feedback.

The rest of this article is structured as follows: Section 2 presents the basic concepts necessary for a better understanding of this work; in Section 3, some related works are listed; Section 4 describes the database, materials, and methods adopted in the descriptive modeling performed; Section 5 highlights the main results obtained; in Section 6 shows examples of feedback; and, finally, Section 7 presents the final considerations and future directions.

2. Theoretical Background

2.1. Learning Analytics (LA)

The use of appropriate technology is vital to interpret an extensive set of data that can produce relevant and personalized information for users. When this extensive amount of data comes from Virtual Learning Environments and requires treatment and analysis, it is precisely through learning analysis that this is done [Baldassarre 2016]. Learning Analytics consists of an important application of analytical intelligence (Analytics) in the educational context to guide better decision-making through three types of analysis, namely, descriptive, predictive, and prescriptive. Descriptive analytics is concerned with what happened and why. Prescriptive analytics involves uncovering hidden relationships in the data to predict what will happen. Prescriptive analytics is concerned with the best course of action by combining the results of descriptive analytics and predictive [Daniel 2015]. This article uses descriptive analysis to describe and summarize learning data and metrics to identify the most important information that can support future analysis (predictive and prescriptive) and feedback generation. This article focuses on the descriptive analysis of high school student interaction data in a Virtual Learning Environment for feedback and future analysis (predictive and prescriptive).

2.2. Learning Metrics

The learning metrics aim to assess the learning experience and the conditions of the teaching environment. The idea is to assimilate through them whether the strategies, resources, activities, and learning objects available in the Virtual Learning Environment are contributing to the achievement of the results expected by teachers and students. Several works use learning analytics [Paiva et al. 2019; Martin and Ndoye 2016; Govindarajan et al. 2015] based on traditional performance metrics. However, Leitão et al. (2020) present several metrics that expand the evaluative elements beyond right and wrong. Through the statistical analysis of such learning metrics, among other variables, this article intends to support the generation of feedback to encourage the learner to become independent and responsible for their own learning, as recommended by Cutumisu and Schwartz (2021).

3. Related Works

An exploratory search was carried out in the Scopus (Elsevier) database in search of works that used statistical analysis of metrics that evaluated the learning experience and the conditions of the teaching environment for students in Virtual Learning Environments for feedback purposes. In this way, it was possible to notice, in the literature, several works that use statistical analysis of interaction and performance data in Virtual Learning Environments, aiming to understand the data to assist in making pedagogical decisions that lead to improvements in the learning process. However, in the first instance, no studies were identified that carried out statistical analyses of learning metrics in high school courses from Virtual Learning Environments and used such analyses to provide feedback to students and teachers.

Table 1 presents a comparative summary of related works under the following aspects: type of course, statistical technique used, use of educational data from interactions and grades in Virtual Learning Environments, assessment of student performance, and use of analysis for feedback purposes. In all these works, the application of statistical techniques to educational data and their importance in the study and understanding of educational processes stood out.

Table 1. Comparison of related works.

Author	Type of Course	Technique Used	ED[1]	SPE[2]	Fb[3]
[Purwoningsih et al. 2020]	Open and Distance Learning	Descriptive statistics	yes	yes	no
[Gonçalves et al. 2018]	University education	Descriptive statistics	yes	yes	no
[Reino et al. 2015]	University education	Descriptive statistics	yes	no	no
[Detoni et al. 2014]	University education	Machine learning	yes	yes	no
[Ramos et al. 2014]	University education	Descriptive statistics	yes	yes	no
[Souza et al. 2013]	University education	Descriptive and multivariate statistics	yes	yes	no
This paper	High school	Descriptive and multivariate statistics	yes	yes	yes

[1] Educational Data. [2] Student Performance Evaluation. [3] Feedback.

Thus, the central differential of this article is the use of descriptive analysis, applied to a set of data from high school students in a Virtual Learning Environment, to understand the behavior of the variables and identify those most relevant for future statistical analyses. Based on this, an attempt was made to determine their relationships, aiming to use such information to generate feedback that could enhance student learning and support teachers in decision-making.

4. Methodology

This section presents the procedure adopted in the descriptive analysis of the data of this study to achieve the proposed objective.

4.1. Database

The data analyzed in this article were collected by Leitão (2023) from the interaction of 339 high school students in a virtual learning environment. The collection was carried out during the application of preparatory simulations for the entrance exam at the institution where the researcher teaches. The students were divided into 11 classes, where each class answered a simulation of 40 multiple-choice questions with five alternatives, grouped into questionnaires by subject. Each student obtained a grade for each simulated subject. After excluding 365 missing data records, the analyzed data resulted in 2,251 observations from multiple-choice questionnaires and student interaction data with the learning environment, used to calculate the evaluation metrics that enabled analysis of difficulties and students' needs. This article analyzes the metrics presented by Leitão et al. (2020) that are related to the questionnaires as a whole, that is, by discipline. Since the metrics related only to the questions are part of the calculation of the metrics related to the questionnaires, namely: Traditional Score (TS), Weighted Score (WS), Set Deviation (SD), Priority (P), Response Time, Assertiveness Degree (AD) and Questionnaire Comprehension Level (QuCL).

4.2. Materials and Methods

The R and Python programming languages were used to generate the descriptive analyses, summarize the data and provide graphs for viewing and understanding the metrics and the dependence relationship between the variables studied in this article. For data adjustment and analysis, attributes engineering was performed on the data set, and three new variables were created, as described in Table 2.

Table 2. New Variables.

Variable	Description
TIME	Which stores the total time, in hours, spent by the student to answer a questionnaire, based on the sum of the Response Time for each question in the questionnaire, originally computed in seconds.
SITUATION	Which assumed the value "Passed," if the student's Traditional Score (TS) in a given questionnaire is equal to or greater than 5, or "Failed," otherwise.
CONCEPT	Also categorized according to the student's Traditional Score (TS), assuming the following levels: "A" ($8.75 \leq TS \leq 10$ - superior passed); "B" ($6.25 \leq TS < 8.75$ - average passed); "C" ($5 \leq TS < 6.25$ - passed); "D" ($3.75 \leq TS < 5$ - needs intervention); "E" ($0 \leq TS < 3.75$ - failed).

Before proceeding with the descriptive analysis of the data, missing data was detected in the Assertiveness Degree (AD) variable, and it was decided to exclude such records. Then, adjustments were made to the types of variables, their basic statistics were calculated, and their balancing was verified using the Skewness asymmetry measure and visual analysis of histograms and boxplots. In addition, we performed a correlation analysis of variables using Pearson's correlation coefficient and scatter plots, a dependency analysis of variables using comparative boxplots, and a chi-square test for each crossing some categorical variables. Principal components analysis and graph analysis resulting from applying the Random Forest model to the database were used to find the most relevant characteristics (predictive variables) to predict students' final performance.

5. Results

The basic statistics on the numerical variables of the dataset examined in this article are visually distributed in the diagrams in Figures 1 and 2. Regarding these statistics, the following stand out:

- Among the variables analyzed, only the TIME variable presents a measure of asymmetry (Skewness = 2.125) outside the range [-1,+1], indicating a strong asymmetry in the distribution of this variable [Fávero and Belfiore 2017]. Therefore, if the TIME variable needs to undergo parametric analysis in future work, it will be necessary to adapt it to the assumptions of the normal distribution through a transformation to verify whether the asymmetry persists, as the lack of data symmetry dramatically limits the application of the main statistical techniques or makes the results inconclusive [Mandrekar and Mandrekar 2003].
- The second variable with the highest asymmetry (0.693) is the Degree of Assertiveness (AD), but it is still close to normal, that is, in the interval [-1,+1].
- Most of the variables studied (TS, WS, SD, P, QuCL, TIME) present mean and median with approximate values, indicating a certain symmetry of the distributions of these variables.

- The Traditional Score (TS), Priority (P), and TIME variables have outliers.

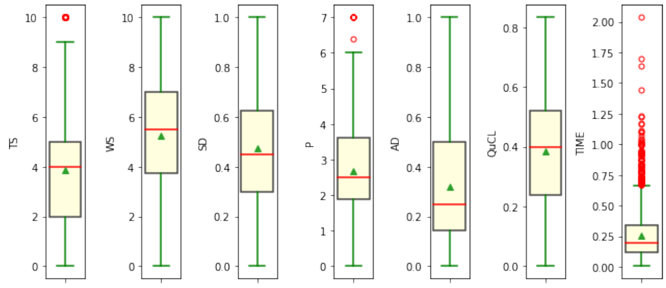

Figure 1. Boxplot representing the dataset.

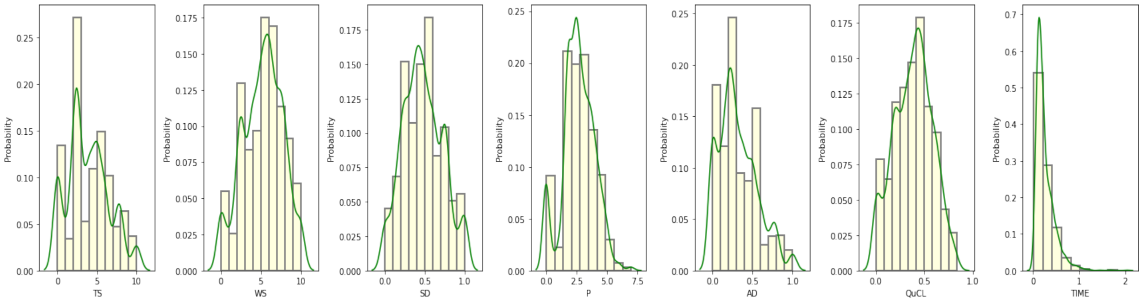

Figure 2. Histograms of the original numeric variables.

From the comparative boxplots in Figure 3, it is possible to see, for example, that the time spent by students who passed the questionnaires presents practically the same variability as that of students who failed. This suggests that no defined relationship exists between the time spent answering the questionnaires and whether or not the student passes the subject. Figure 3 shows that students who passed have a higher degree of assertiveness and understanding than students who failed, as expected. It is also verified that the outliers occurred more in the following cases: (1) approved and failed students who spent much time answering the questionnaires; (2) approved students with a Questionnaire Comprehension Level (QuCL) below that presented by failed students; and, (3) failed students with a degree of assertiveness similar to that achieved by approved students. Figure 3 also shows that as the CONCEPT variable oscillates from A to E, Assertiveness Degree (AD) decreases. There appear outliers that did not occur when considering the SITUATION variable. The same happened with the Questionnaire Comprehension Level (QuCL) variable in the face of the same CONCEPT variation. However, such behavior is not verified when considering the variable TIME. Therefore, it is necessary to investigate the outliers detected when comparing the variables in question (QuCL, AD, and TIME) with the levels of the CONCEPT variable.

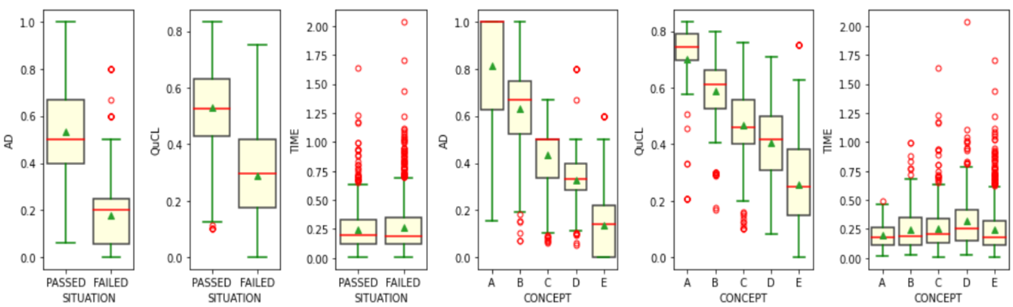

Figure 3. Comparative boxplots.

When applying the chi-square test on the SITUATION x CLASS and

SITUATION x SUBJECT crosses, it was found that, in both cases, the *p-value* assumed a value lower than the significance level of 0.05, indicating that the fact that the student is part of a particular class or is studying a specific subject seems to interfere with his/her performance.

The alleged lack of relationship between the time spent answering the questionnaire and the student's passing or not in the subject is confirmed, at least linearly, by observing the correlation matrix in Figure 4, given that Pearson's correlation coefficient between Traditional Score (TS) and TIME is very close to 0. Figure 4 shows a moderate positive correlation between the Assertiveness Degree (AD) and Questionnaire Comprehension Level (QuCL) variables. On the other hand, the pairs Traditional Score (TS) and Assertiveness Degree (AD), Traditional Score (TS) and Questionnaire Comprehension Level (QuCL), Weighted Score (WS) and Assertiveness Degree (AD), and Weighted Score (WS) and Questionnaire Comprehension Level (QuCL) reveal a high positive correlation.

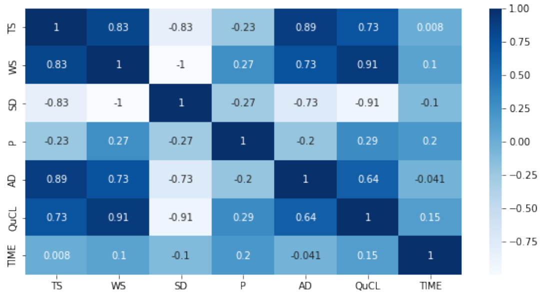

Figure 4. Pearson's Correlation Coefficient.

Regarding the scatter plots in Figure 5, it is worth to note that in graphs (a), (b), (c), and (d), the TIME variable appears to have no association with the other variables analyzed because there is no clear trend of the points. The trend of the points in graphs (e), (f), (g), and (h) is clearly perceived as positive. Also, in the graph (i), the points have a positive trend, although their scattering indicates a moderate correlation.

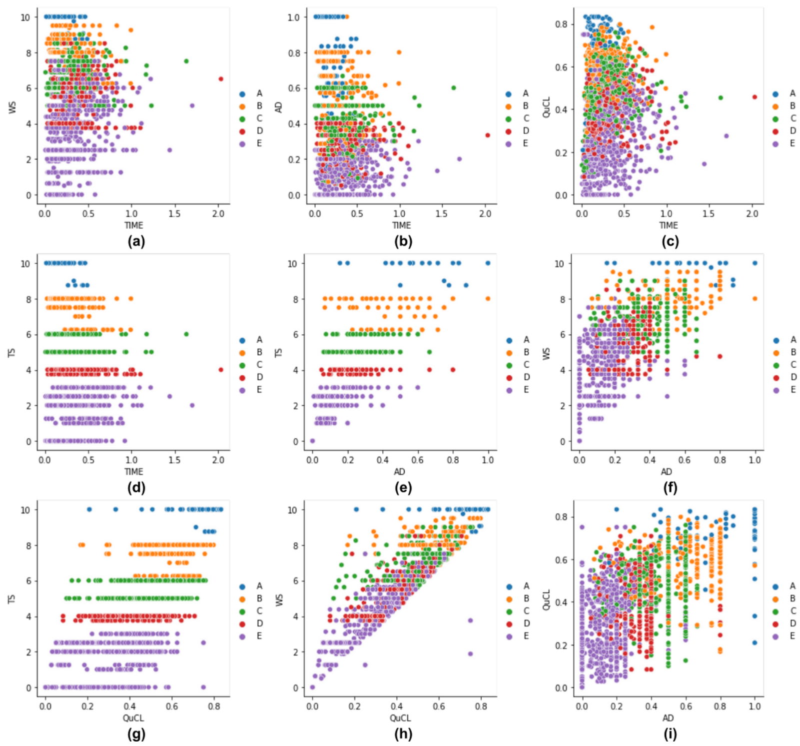

Figure 5. Scatter diagrams.

The Random Forest model was applied to the data to predict student performance and indicate the most relevant characteristics to predict final student performance. In this model, a set of decision trees establishes a score for each variable, and those with the highest score should be selected for model construction [Jones and Linder 2015]. Thus, Figure 6 shows that the model detected that the Traditional Score (TS) is the most relevant variable to predict the student's situation at the end of the course, as expected since it was used to create the variable SITUATION, followed by Assertiveness Degree (AD), Weighted Score (WS), Set Deviation (SD), Priority (P) and Questionnaire Comprehension Level (QuCL).

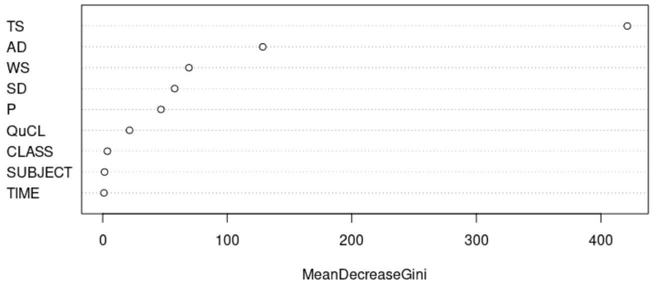

Figure 6. Relevance of variables for building the forecast model.

Figure 7 shows the Principal Component Analysis graphs, whose objective is to condense the information contained in the original variables into a smaller set of statistical variables (components). With minimal loss of information, hidden data patterns are identified, dimensionality is reduced, and correlated variables are identified [Kassambara 2017]. Figure 7 (a) shows that the first two explain more than 80% of the data variation of the seven main components. As the variability of these two components is greater than 50%, graph (b) of Figure 7 can be considered adequate to evaluate the relationships between the variables and identify how this data sample is ordered and its relationship between the variables for cluster analysis. Furthermore, graph (b) in Figure 7 shows the variables that most contributed to the differentiation of points and the grouping between them. In this way, the Principal Component Analysis (PCA) helped define the main variables directly related to the student's situation in the discipline, whether passed or failed. PC1 and PC2 explained 82.3% of the data, with the first factor explaining 61.5% and the second factor 20.8%.

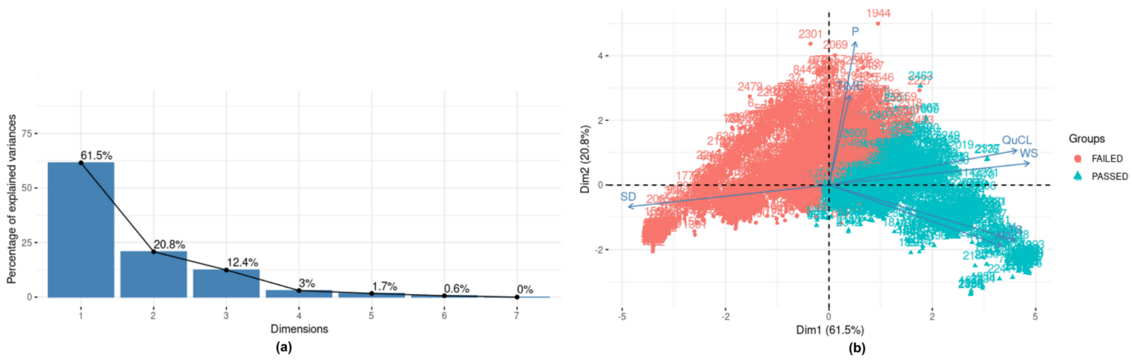

Figure 7. Principal Component Analysis Graphs.

6. Examples of Feedback

Based on the experience of the researchers involved in this study, on the suggestion of high school teachers, on the categorization adopted in the CONCEPT variable (Table 2), and on the analysis of the evaluation metrics, some scenarios were identified that may be useful for providing personalized feedback to students and teachers. Even more so if one considers the difficulty that the teacher has to individualize the feedback for each student in a large class. Therefore, for each scenario, an action was proposed to be adopted by the interested party in the feedback. Some examples of feedback are listed in Table 3. The recommended actions for each scenario aim to help the teacher make pedagogical decisions and minimize the chances of student failure through early interference by interested parties in the teaching-learning process.

Table 3. Examples of Feedback.

Scenario	Recommended action	Interested party	Feedback sample
High performance of students in a given subject.	Propose challenging questions to encourage them and prevent them from getting discouraged.	Students with grade ''A'' $(8.75 \le TS \le 10)$	Very good! How about more challenging questions?
Medium performance of students in a given subject	Suggests that they study the subjects in which they obtained a grade close to the passing average.	Students with grade ''C'' $(5 \le TS < 6.25)$	Very well, but your performance can improve! We suggest a greater dedication to the following subjects, to obtain better results.
Low performance of students in a given subject.	Suggest that they study the subjects in which they obtained a grade below the passing average. In addition, based on the "priority" metric, indicate an order for studying the subjects.	Students with grades ''D'' $(3.75 \le TS < 5)$ and ''E'' $(0 \le TS < 3.75)$	Your performance must improve! In this order of priority, we suggest a greater dedication to the following subjects to obtain better results without too much effort. `<MATTER LIST IN DESCENDING ORDER OF PRIORITY>`
Very large discrepancy between the average time estimated by the teacher and the response time of all students.	Suggest checking that the time estimated by the teacher is not underestimated or overestimated, as this information influences the calculation of the Questionnaire Comprehension Level.	Professor	Teacher, check that the estimated time for the following questions is not underestimated or overestimated, as there was a very large discrepancy when compared to the response time of most students.
Time spent answering a questionnaire is much higher than the time estimated by the teacher.	Encourage the student to practice more exercises for greater agility next time.	Any student	You need to practice more exercises on the following contents to improve your response time. `<CONTENT LIST>`
Approved student with response time much lower than expected to answer the questionnaires.	Suggest checking whether the student really masters the subject.	Professor	Teacher, student `<STUDENT NAME>` took a long time to complete the questionnaire. Wouldn't it be nice to see if he really mastered the subject, or if it was just luck?
Approved student but with a Questionnaire Comprehension Level below that presented by failed students.	Suggest that it be checked if the student is having any difficulties.	Professor	Teacher, student `<STUDENT NAME>` obtained a Questionnaire Comprehension Level below that presented by students who failed. Wouldn't it be nice to see if he is having any trouble?
Failed student but with a degree of assertiveness similar to that achieved by approved students.	Suggest that the student be helped to master the rest of the content.	Professor	Teacher, student `<STUDENT NAME>` obtained a degree of assertiveness similar to that achieved by approved students. Wouldn't it be nice to help master the rest of the content?
Most students prioritize a particular subject.	Suggest that some corrective action be taken in this situation.	Course coordinator and teacher	Most students (`<PERCENTAGE OF STUDENT>`) have the subject `<NAME OF SUBJECT>` as their study priority. Wouldn't it be good to take some corrective action regarding this situation?
Most students got the same question wrong.	Suggest that the content of the question be reviewed in the classroom.	Professor	Teacher, most students got the following question wrong `<QUESTION-ID>`. Wouldn't it be nice to review the content of this question?

7. Conclusion and Future Works

This study aimed to present the description and summary of data and learning metrics to identify the most important information about the studied sample, which is useful to subsidize the feedback with future predictive and prescriptive analyses. The results of the exploratory data analysis indicated that, in addition to the Traditional Score (TS), the most important variables in this study are Assertiveness Degree (AD), Set Deviation (SD), Weighted Score (WS), Priority (P) and Questionnaire Comprehension Level (QuCL). Another result that this exploratory analysis showed was that, for the analyzed data set, the time taken to take the exam does not seem to be related to the student's approval of the questionnaire. However, it may be useful to investigate further the relationship between access time to the Virtual Learning Environment and final performance in the discipline or course. From the cross-analysis CLASS x SITUATION, which suggested the influence of the class on student achievement, another hypothesis to be investigated is that this is due to some specific characteristic of the class to which the student belongs. This makes it necessary to investigate possible variables that characterize the class. Furthermore, the detected outliers – although they need to be treated because they can distort and affect the precision of the estimates and adjustment forecasts – may also highlight the anomalies that need to be investigated in the next analysis phase. Thus, a more detailed investigation of the outliers is required to handle them properly before any modeling.

In future work, we intend to investigate the detected anomalies (outliers) and whether the class and subject influence the student's final performance in the questionnaire. If this last assumption proves to be true in future research, it may also be useful to identify the losses and possible solutions for this situation. It is also planned to carry out predictive modeling based on the variables identified as important in this article to predict the students' final situation and provide actionable recommendations for both students and teachers in time to reverse the risk of failure in subjects.

8. Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES-PROEX) - Finance Code 001. This work was partially supported by Amazonas State Research Support Foundation - FAPEAM - through the POSGRAD project.

9. References

Baldassarre, M. (2016). Think big: learning contexts, algorithms and data science. *Research on Education and Media*, 8 (2), 69–83.

Costa, C. S. and Mattos, F. R. P. (2016). *Tecnologia na sala de aula em relatos de professores*. Curitiba: CRV, 202.

Cutumisu, M. and Schwartz, D. L. (2021). Feedback choices and their relations to learning are ageinvariant starting in middle school: A secondary data analysis. *Computers & Education*, 171, 104215.

Daniel, B. (2015). Big data and analytics in higher education: Opportunities and challenges. *British journal of educational technology*, 46(5), 904–920.

de Macêdo Santiago, L. B., Vasconcelos, K. C., and Santana, J. R. (2016). O uso dos artefatos tecnológicos virtuais e digitais na escola. *ARTEFACTUM-Revista de estudos em Linguagens e Tecnologia*, 13(2).

Detoni, D., Araujo, R. M., and Cechinel, C. (2014). Predição de reprovação de alunos de educação a distância utilizando contagem de interações. In *Anais do Simpósio Brasileiro de Informática na Educação* (Vol. 25, N. 1, p. 896).

Fávero, L. P. and Belfiore, P. (2017). *Manual de análise de dados: estatística e modelagem multivariada com Excel®, SPSS® e Stata®*. Elsevier Brasil.

Gonçalves, E. J. T., Vilela, J. F. F., and Bezerra, J. D. M. (2018). Análise estatística de notas e interações em cursos a distância. In *Anais do Simpósio Brasileiro de Informática na Educação* (Vol. 29, p. 71).

Govindarajan, K., Kumar, V. S., and Boulanger, D. (2015). Learning analytics solution for reducing learners' course failure rate. In *2015 IEEE seventh international conference on technology for education (T4E)* (pp. 83–90). IEEE.

Jones, Z. and Linder, F. (2015). Exploratory data analysis using random forests. In *Prepared for the 73rd annual MPSA conference*, 1–31.

Kassambara, A. (2017). *Practical guide to principal component methods in R: PCA, M (CA), FAMD, MFA, HCPC, factoextra* (Vol. 2). Sthda.

Koenka, A. C. and Anderman, E. M. (2019). Personalized feedback as a strategy for improving motivation and performance among middle school students. *Middle School Journal*, 50(5), 15–22.

Leitão, G. D. S. (2017). Uma plataforma de suporte ao docente no contexto da educação digital. M.S. thesis, Universidade Federal do Amazonas.

Leitão, G. D. S. (2023). Um método baseado em objetos tangíveis para verificação da aprendizagem. PhD thesis, Universidade Federal do Amazonas.

Leitão, G., Colonna, J., Monteiro, E., Oliveira, E., and Barreto, R. (2020). New metrics for learning evaluation in digital education platforms. *arXiv preprint arXiv:2006.14711*.

Maier, U. and Klotz, C. (2022). Personalized feedback in digital learning environments: Classification framework and literature review. *Computers and Education: Artificial Intelligence*, 3, 100080.

Mandrekar, S. J. and Mandrekar, J. N. (2003). Are our data symmetric?. *Statistical methods in medical research*, 12(6), 505-513.

Martin, F. and Ndoye, A. (2016). Using learning analytics to assess student learning in online courses. *Journal of University Teaching & Learning Practice*, 13(3), 7.

Paiva, R., Bittencourt, I. I., and Lemos, W. (2019). Helping teachers visualize students' performance. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)* (Vol. 30, N. 1, p. 1731).

Purwoningsih, T., Santoso, H. B., and Hasibuan, Z. A. (2020). Data analytics of students' profiles and activities in a full online learning context. In *2020 Fifth International Conference on Informatics and Computing (ICIC)* (pp. 1–8). IEEE.

Ramos, J. L. C., Rodrigues, R. L., Silva, J. C. S., and Gomes, A. S. (2014). Analisando fatores que afetam o desempenho de estudantes iniciantes em um curso a distância. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)* (Vol. 25, N. 1, p. 99).

Reino, L. R. A. C., Hernández-Domínguez, A., Júnior, O. D. G. F., Carvalho, V. D. H., Barros, P. A. M., and de Melo Braga, M. (2015). Análise das causas da evasão na educação a distância em uma instituição federal de ensino superior. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)* (Vol. 26, N. 1, p. 91).

Sousa, E. B. D., Alexandre, B., Ferreira Mello, R., Pontual Falcão, T., Vesin, B. and Gašević, D. (2021). Applications of learning analytics in high schools: A systematic literature review. *Frontiers in Artificial Intelligence*, 4, 737891.

Souza, T. I., Franco, A. O., Silva, T. E., and Vasconcelos, H. L. (2013). Avaliando o desempenho discente em um AVA: Um estudo de caso utilizando estatística multivariada. In *Anais dos Workshops do Congresso Brasileiro de Informática na Educação* (Vol. 2, N. 1, pp. 422–431).