Revista de Ciências Agroveterinárias
Universidade do Estado de Santa Catarina (UDESC)

Research Article

Automation in exam grading with Python and web hosting: data science applied to education

Automação na correção de provas com Python e hospedagem web: Ciência de dados aplicada à educação

Paulo Henrique Cerutti *(ORCID 0000-0001-6664-8449), Luan Tiago dos Santos Carbonari (ORCID 0000-0002-7797-5194), Carlos Zacarias Joaquim Junior (ORCID 0000-0001-8681-9946), Mauro Bitencourt de Souza (ORCID 0009-0003-8723-6285), Henrique de Sá Albino (ORCID 0009-0002-9654-7539), Jefferson Luís Meirelles Coimbra (ORCID 0000-0001-9492-6055)

State University of Santa Catarina, Lages, SC, Brazil. *Corresponding author: paulohcerutti@gmail.com

Submission: November 1, 2024 | Acceptance: March 15, 2025

ABSTRACT

Data science employs statistical methods and computational tools to automatically extract knowledge from data or information on a specific subject. In the educational context, automating exam grading enhances efficiency and accuracy, which is especially relevant in social and agrarian sciences, where a high volume of students, questions, and the need for quick grading are common. Thus, the objective of this study was to develop an automation system called *AutoCorrect*, utilizing *Python* and web hosting for the grading of academic exams. The mechanism allows for the generation of customized exams with randomized answer keys and performs automatic grading of submitted responses, significantly optimizing the evaluation process. The methodology included the use of specific *Python* libraries, such as *Pandas*, *Numpy*, and *Matplotlib*, for data processing, statistical analysis, and graphical visualization. A practical application was conducted in Agronomy and Veterinary Medicine classes, in which students took the generated exams. Student responses were directly compared to the answer key provided by instructors, resulting in the automatic generation of scores for each student based on correct and incorrect answers. Additionally, the system enabled a detailed analysis of grades through descriptive measures, such as mean, mean standard deviation, variance, and skewness, providing a clear view of class performance. Supplementary charts were generated to facilitate visualization of the grade distribution and individual performance relative to the overall average. The integration of data science with computational tools like *Python* and web hosting not only optimizes the grading process but also provides greater clarity and agility in analyzing student performance. The *AutoCorrect* system proves to be an efficient and innovative tool for various educational fields, especially in agrarian and social sciences.

KEYWORDS: *AutoCorrect.* Data Analysis. Machine Learning. Computational Tools. Educational Reports.

RESUMO

A ciência de dados utiliza métodos estatísticos e ferramentas computacionais para extrair conhecimento a partir de dados ou informações sobre um determinado assunto de forma automatizada. No contexto educacional a automação da correção de provas proporciona maior eficiência e precisão, sendo especialmente relevante nas ciências sociais e agrárias, onde o volume de alunos, de questões e a necessidade de correção rápida são eventos frequentes. Assim, o objetivo deste trabalho foi desenvolver um sistema de automação chamado *AutoCorrect*, utilizando *Python* e hospedagem Web para a correção de provas acadêmicas. O mecanismo permite a geração de provas personalizadas, com gabaritos aleatórios, e realiza a correção automática das respostas submetidas, otimizando significativamente o processo de avaliação. A metodologia incluiu o uso de bibliotecas específicas do *Python*, como *Pandas*, *Numpy* e *Matplotlib*, para o processamento de dados, análise estatística e visualização gráfica. Para tanto, foi realizada uma aplicação prática em turmas dos cursos de Agronomia e Medicina veterinária, nas quais os alunos responderam às provas geradas. As respostas dos alunos foram comparadas diretamente com o gabarito fornecido pelos professores, o que resultou na geração automática das pontuações de cada estudante, baseadas em seus acertos e erros. Além disso, o sistema permitiu a análise detalhada das notas

por meio de medidas descritivas, como média, desvio padrão da média, variância e assimetria, oferecendo uma visão clara do desempenho da turma. Gráficos complementares foram gerados para facilitar a visualização da distribuição das notas e do desempenho individual em relação à média geral. A integração da ciência de dados com ferramentas computacionais como *Python* e hospedagem web, não apenas otimizam o processo de correção de provas, mas também fornecem maior clareza e agilidade à análise do desempenho dos alunos. O sistema *AutoCorrect* apresenta-se como uma ferramenta eficiente e inovadora para várias áreas da educação, especialmente nas ciências agrárias e sociais.
PALAVRAS-CHAVE: AutoCorrect. Análise de dados. Aprendizagem de Máquinas. Ferramentas Computacionais. Relatórios Educacionais.

INTRODUCTION

Data science has established itself as an essential tool for education, providing solutions that optimize teaching and evaluation processes. In areas such as Agronomy and Veterinary Medicine, the application of data science techniques allows for the automation of repetitive tasks, such as grading exams and making evaluation faster, more accurate, and more efficient.

This process goes beyond simply analyzing large volumes of data ("Big Data"), as it primarily involves creating practical solutions for real challenges faced by both educators and students. By utilizing machine learning algorithms and statistical techniques, data science transforms information into concrete actions, facilitating the work of instructors and improving the educational experience (VAN DER AALST 2016).

The COVID-19 pandemic further highlighted the need for automated solutions in the educational environment. Educators, particularly in the field of agricultural sciences, have recognized the importance of using technologies that facilitate the grading of assessments and the analysis of student performance. In the field of veterinary medicine, authors highlighted the importance of developing digital platforms to disseminate teaching and learning during the pandemic, noting that this model is supplementary and never a substitute for classic forms of teaching (MASSARI et al. 2022).

With automated corrections, data science exemplifies its ability to transform raw data into effective solutions, providing faster and more accurate assessments (ALDRIYE et al. 2019, KASINATHAN et al. 2022, LIGUORI & WINKLER 2020). In automated test grading, for example, data collection (students' answers to administered tests) is followed by a preparation process, which includes organizing and pre-processing the data. Algorithms are applied to identify patterns in students' responses and automate grading, providing consistent and accurate feedback. This approach not only saves time but also ensures a fairer and more standardized assessment for all students (JAMES et al. 2023).

In this regard, several websites and applications are already used for the automated correction of exams, such as Kahoot, Google Forms, ExamSoft, ZipGrade, and Moodle, among others. For example, ExamSoft is designed for secure assessments, blocking internet access during exams, while ZipGrade transforms mobile devices into scanners for the quick grading of multiple-choice assessments. Moodle offers an open-source virtual learning environment, used in some institutions, but the learning curve for new users without knowledge of Learning Management Systems can be difficult. Also, Moodle may require technical knowledge to work with servers, updates, and security; and depending on the number of users and the

functionalities enabled, Moodle may require significant processing power and bandwidth from the server used (MUSTAFA & ALI 2023, RODILLAS et al. 2023).

However, many of these platforms focus on a broader scope of educational activities and were not essentially developed for grading exams, which limits their efficiency and effectiveness. Based on this, the system proposed in this work differs by directly emphasizing the automated correction of exams, and has been tested in the Agronomy and Veterinary Medicine courses, with the flexibility to expand to all areas of science. It uses code written in the *Python* language with specific functions to organize and correct students' answers effectively.

Furthermore, the creation of a web platform with the "Multclick" functionality offers a simple and intuitive interface, facilitating instructors' access to automated grading and providing fast and accurate results. For example, consider a classroom with 40 students enrolled in a particular subject and eligible to take an exam. When administering a test with 10 questions, and each question has 10 items, there are 40 * 10 * 10 = 4000 items to be graded.

Assuming the instructor administers four exams during the semester, there are 16,000 items to grade. If the instructor takes 30 seconds to correct each item, that totals approximately 480,000 seconds dedicated to this task. This amounts to approximately 133 hours of work, or six days.

With *AutoCorrect*, after creating the necessary files, administering the test, and collecting the students' answers, the correction is done in just a few minutes, with the option for the student to fill in their answers themselves and submit them to the system. Besides, the professor can have access to a question bank for their subject, and based on the desired number of questions, the system will randomly select the appropriate number of questions for each assessment. This expands the possibilities for conducting exams and drastically reduces plagiarism among students.

The solution presented demonstrates the importance of integrating data science into education, offering a concrete proposal to address the specific demands of the evaluation process in higher education (ZHANG et al. 2022).

Thus, the objectives of this work were: to develop code in the *Python* programming language and create an internet environment in the form of web hosting, with the Multclick function (click and get the result), intended for the automated correction of exams, in order to facilitate the work of instructors and guarantee a fair, efficient and accurate evaluation for students.

MATERIALS AND METHODS

Automated test grading can be performed in two environments: *Python* and web hosting.

Libraries Used

Several *Python* libraries were used in the code to perform the data analysis and processing. The most relevant libraries are highlighted in Table 1.

Table 1. Description of main libraries used for the creation of the test correction system.

Library	Function
Pandas	Data manipulation and analysis, including reading and writing files in various formats (csv, xlsx, among others).
NumPy	Mathematical operations and manipulation of arrays (data structures that contain elements of the same type, organized contiguously in memory).
Matplotlib and *Seaborn*	Data visualization and chart creation.
Scikit-learn	Application of machine learning algorithms and model validation.
OpenPyXL or *xlrd*	Reading and writing of Excel files.

Source: HELMUS & COLLINS 2016.

These libraries were installed in the Anaconda virtual environment using the command *conda install library_name* or, in some cases, using the command *pip install library_name*, to install packages that are not directly available through Anaconda (ROLON-MÉRETTE D et al. 2016).

Use of Local Files

When considering both the *Python* code developed and the web hosting, both make use of files stored on the personal computer. These files may include datasets in formats such as: .txt; .csv; .tsv; .xls; .xlsx; .json; .jpg; .png; .zip, among others. The files used were accessed and manipulated directly from the *Jupyter Notebook* working directory. The path to these files was specified in the code to ensure that the data was read and processed correctly (from folders on the personal computer), in accordance with the disciplines of plant genetic improvement and basic statistics, from the Agronomy and Veterinary Medicine courses, respectively.

To execute this code, intended for grading exams and for use with web hosting, it was necessary to create several files. These are directly related to each instructor's data and should be saved in a directory folder on the user's personal computer. The files are: *i)* Student names: a document containing the list of students enrolled in the course and eligible to take the exam. This document can be in .xlsx format; *ii*) Document containing the exam questions: in .docx format; *iii*) Instructor´s answer key: document containing the answers to the selected questions from the exam file, in .xlsx format; *iv*) Weight of each question: file that indicates the weight assigned to each question, calculated including the number of items per question, in .xlsx format. Figures 1 through 4 demonstrate how the files are created, using fictitious names and email addresses to protect the students' identities. These files are essential for the start of the process.

A	B	C
Index	Nome	Email
1	Ana Clara	ana.clara@example.com
2	Pedro Henrique	pedro.henrique@example.com
3	Mariana Souza	mariana.souza@example.com
4	João Vitor	joao.vitor@example.com
5	Laura Mendes	laura.mendes@example.com
6	Gabriel Oliveira	gabriel.oliveira@example.com
7	Beatriz Santos	beatriz.santos@example.com
8	Rafael Lima	rafael.lima@example.com
9	Júlia Ferreira	julia.ferreira@example.com
10	Lucas Pereira	lucas.pereira@example.com

1

Q1. Os dados a seguir representam o rendimento de grãos de parcelas de soja, em gramas. 120; 100; 150; 250; 250; 155; 175; 185; 190; 200; 210; 215; 220. Pergunta-se: qual a escala de medida adotada? Qual o rendimento médio de grãos? A variação em g do conjunto? A conversão desta variação em percentual? Uma estimativa de erro na amostragem? Com base nisso, responda assinalando no gabarito apenas a(as) alternativa(s) correta(s): (1) Ordinal; 187; 45,74; 25% e 13. (2) Racional; 187; 45,74; 25% e 13; (3) Intervalar; 186,15; 45,74; 24,57% e 12,68; (4) Racional e Ordinal; 187; 45,90; 23,9% e 15; (5) Racional; 186,15; 45,74; 24,57% e 12,68; (6) Nenhuma das anteriores.

Q2. Sobre medidas para dados quantitativos, escalas de medida e tabela de frequência, assinale a/as alternativas corretas: (1) Mesmo um especialista em uma área de pesquisa pode não ser capaz de obter uma amostra que seja verdadeiramente representativa, então é melhor escolher uma amostra aleatória; (2) o objetivo da estatística é fazer inferências sobre uma população baseadas na informação de amostras; (3) Distribuições de frequência são sempre simétricas; (4) A escala ordinal é algumas vezes usada, mesmo que informação numérica mais precisa seja disponível; (5) Pressão barométrica é usualmente registrada em escala ordinal; (6) Operações aritméticas podem ser efetuadas sobre escalas nominais e de intervalo, mas não podem ser aplicadas a números de escalas de razão e ordinal; (7) A estatística provê estimativas de parâmetros.

2

A	B	C	D	E	F	G	H	I	J	K
QUESTAO	1	2	3	4	5	6	7	8	9	
Q1					**5**	**0.2525**				
Q2	**x**	**x**		**x**			**x**			
Q3	**1**	**2**	**3**		**5**		**7**			

3

A	B	C	D	E
Questão	peso	Subitens		
Q1	3	2		
Q2	3	4		
Q3	4	5		

4

Figures 1 to 4. Creation of files for the complete preparation test correction process, being: 1- file with the name and email of the students; 2- file with the test questions; 3- file with the correct answers per question and 4- weight assigned to each question.

Parts of the code

The *Python* code for automating exam grading was formulated in parts, namely: *i)* Randomization of answer keys according to the names of the students listed; *ii)* Generation of randomized questions, according to the randomized answer keys; *iii)* Compilation of randomized exam files for printing, checking, and distribution to students on the day of the exam; *iv)* Grading of the exams, according to the answers marked on the randomized answer keys. This correction is performed both by marking an "x" on the answer sheet, as well as by choosing numbers corresponding to the correct items (e.g., question 3: correct items 1, 2, 3, 5 and 7) and also by marking values resulting from calculations on the answer sheet (e.g., question 1 of the figures shown previously, where the correct answer for item 6 of that question is 0.2525; v) correction with upload of the question weighting file. In this item on the weight of each question is informed in a file for grading; *vi)* Generation of a heat map. Command used to create a heat map based on student scores; and *vii)* Creation of a graph showing the standard deviation in relation to the overall mean.

Once the environment was configured and the necessary libraries were installed, the code was executed in *Jupyter Notebook*. All parts of the code can be found at a Google Drive address, in the folder called "*Python* Exam Corrections," available for download by any user:

https://drive.google.com/drive/folders/1xUYkB7EsXo-F9CXleLhZ0wrgK-SN3wtH?usp=drive_link.

Web Hosting

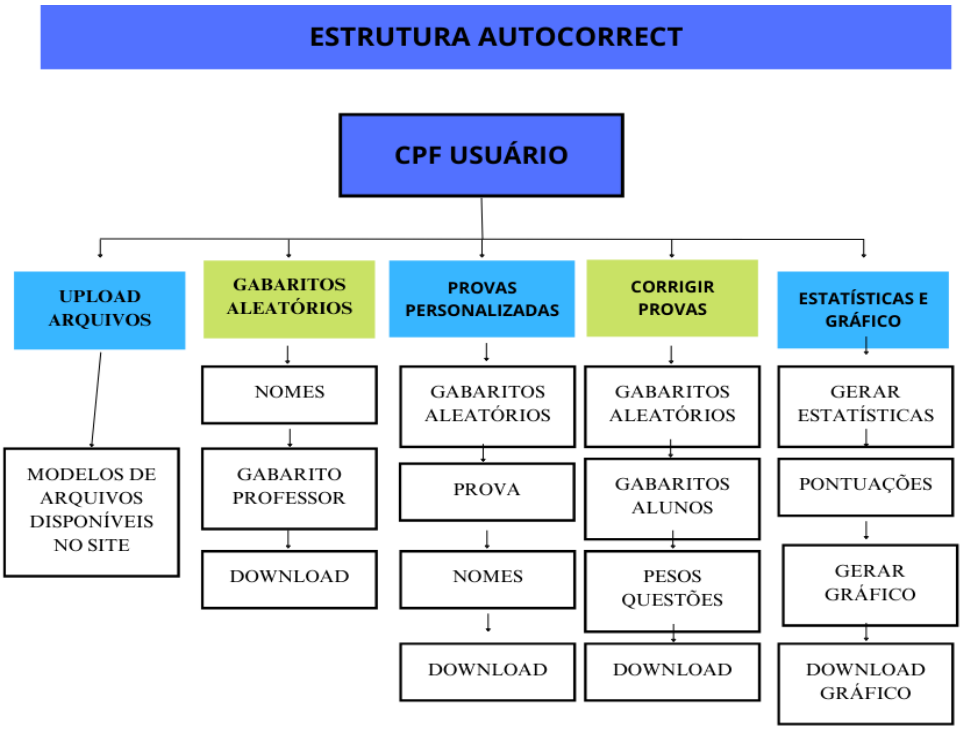
The website called *AutoCorrect* was designed primarily to provide a Multiclick option to users of this automated test correction system. It can be accessed at the following address: http://coimbrajefferson.pythonanywhere.com.

AutoCorrect offers open access, however, the authors hold the patent and trademark registration for the computer programs. To access, the CPF (Brazilian individual taxpayer registry number) must be entered only with numbers, for example: 11111111111. The choice of CPF as a form of entry is due to the fact that each person has a specific identification number.

However, we emphasize that if users define the use of fictitious CPF numbers and happen to use the same number (11111111111), the system will not function correctly, because each instructor has their own files.

The *AutoCorrect* website operates according to the following flowchart:

.



Each of the parts has the following general overview:

Figure 5. Overview of access to the website for automated correction and the website's initial interface for uploading the files required to complete the process.

After logging in, the user is presented with the website interface (Figures 5 to 9), where it is possible to upload files for the automated exam correction process, according to the needs of each instructor.

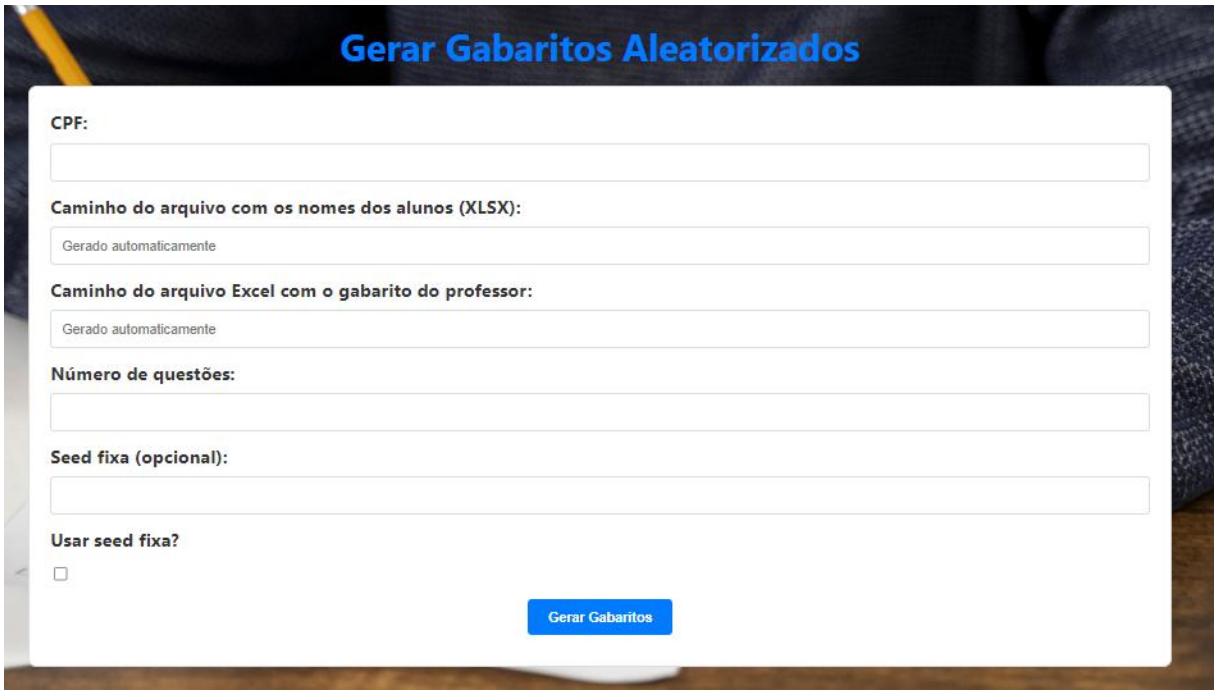


Figure 6. Fraction of the site directly responsible for creating randomized templates for each student.

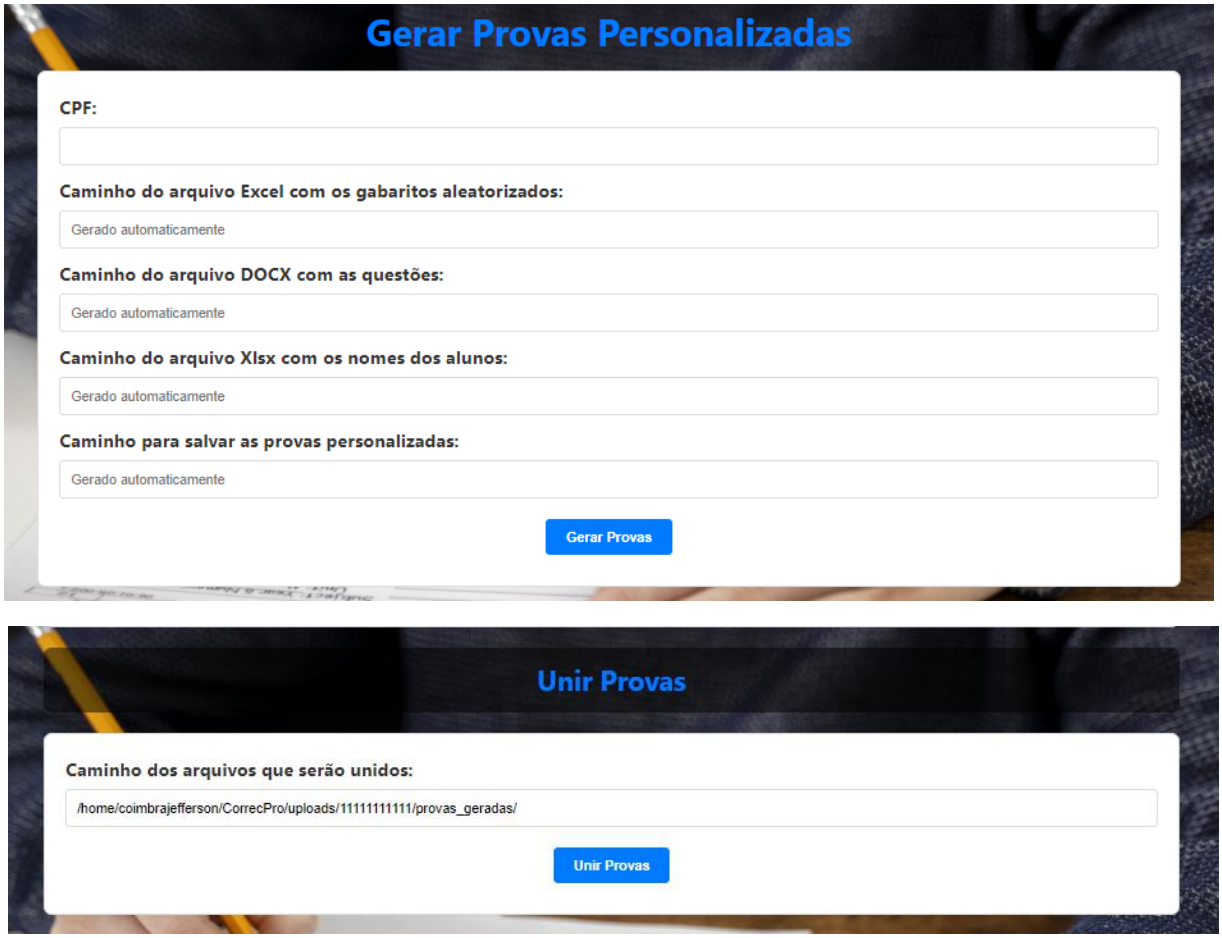


Figure 7. Fractions responsible for generating the tests for each person and combining these into a single Microsoft Word document for printing.

The tests for each student are prepared according to the randomized answer key generated for each student.

Figure 8. Part of the website intended for correcting tests and the possibility of downloading results.

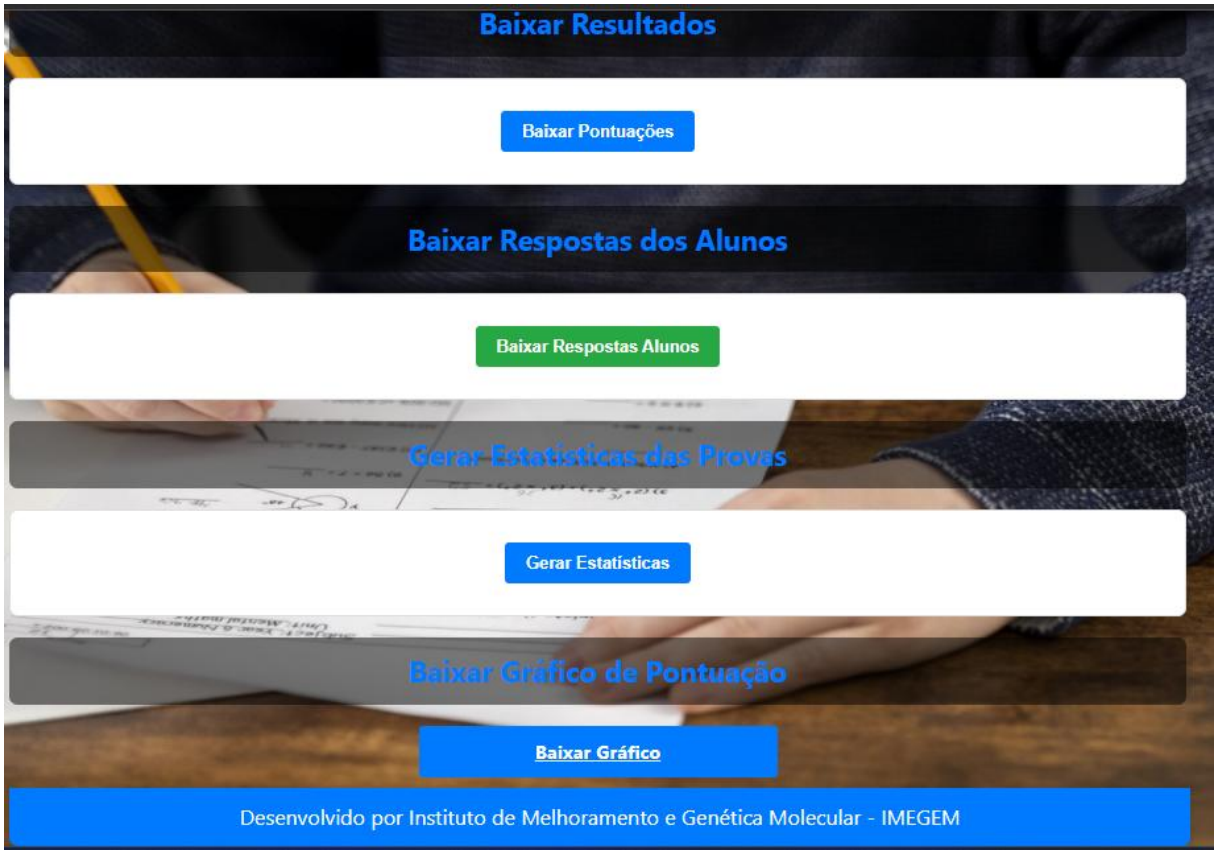


Figure 9. Part of the website intended for downloading students' responses, generating descriptive statistics and a graph with scores.

Considerations regarding the operation of AutoCorrect: In the same Google Drive folder, a PDF file with instructions for using AutoCorrect (AutoCorrect Step-by-Step) has been made available, accessible at: https://drive.google.com/drive/folders/1xUYkB7EsXo-F9CXleLhZ0wrgK-SN3wtH?usp

=drive_link. All parts of the website require the user to provide the path to the file to be uploaded. After uploading, the files are stored in the cloud for a maximum of 24 hours. It is important that the file names are correct before submission (that they are identical to the templates available for download and editing). During the generation of randomized answer keys, the user must indicate whether they want to use a fixed seed. In Python, a fixed seed, especially in operations involving random numbers, ensures that the results are reproducible. In general, functions that depend on randomness, such as number generation or data sampling, produce different results with each execution. However, by defining a seed, the starting point of the random number generator is "locked," ensuring that the same results are obtained in subsequent runs (DOE 2024).

In the part of the code responsible for generating the randomized answer sheets, the system uses the file containing the students' names and the answer key provided by the instructor. AutoCorrect creates this file, called "randomized answer keys," which is automatically saved in the user's downloads folder. To personalize the tests for each student, the user must open the file in .xlsx format, select all the worksheets, and remove the first row (which contains the students' names). After making these edits, the file must be sent back to AutoCorrect to generate the personalized exams.

RESULTS AND DISCUSSION

After the code is executed or through the use of AutoCorrect, the essential files are generated to achieve the final objective: correcting the questions and obtaining scores for each student. The algorithm used is based on controlled randomization (seed), which allows for the shuffling of questions and answer keys using Python's Random library, making the teaching and learning process more dynamic and efficient (SANDVE et al. 2013). Figure 10 exemplifies the generation of two randomized answer keys for students "Ana Clara" and "Pedro Henrique".

A	B	C	D	E	F	G	H	I	J	K
Ana Clara										
Questão	Questão C	Subitem 1	Subitem 2	Subitem 3	Subitem 4	Subitem 5	Subitem 6	Subitem 7	Subitem 8	Subitem 9
Q1	Q2	x	x		x			x		
Q2	Q1					5	0.2525			
Q3	Q3	1	2	3		5		7		

A	B	C	D	E	F	G	H	I	J	K
Pedro Henrique										
Questão	Questão C	Subitem 1	Subitem 2	Subitem 3	Subitem 4	Subitem 5	Subitem 6	Subitem 7	Subitem 8	Subitem 9
Q1	Q1					5	0.2525			
Q2	Q3	1	2	3		5		7		
Q3	Q2	x	x		x			x		

Figure 10. Randomized answer sheets for specific students in the subject in question.

One of the main benefits of randomizing answer keys and questions is the ability to create unique versions of exams for each student, which significantly reduces the chance of plagiarism or improper collaboration during exams. Besides, using a fixed

seed ensures that the generated versions can be reproduced, allowing the instructor to have complete control over the process while maintaining the integrity of the assessment (DOE 2024).

Another benefit is the ability for the instructor to use an extensive question bank, allowing the system to randomly select different questions for each student. This personalizes the assessments and offers a unique evaluation experience for each student, ensuring that everyone is assessed fairly and equitably. This allows the instructor to create a diversified assessment without the need to manually generate multiple versions of the test. This approach also allows for the continuous use of a question bank over several semesters, without students receiving the same tests repeatedly (KURDI et al. 2020).

In addition to optimizing the instructor's time, the use of randomized questions encourages learning, since each student receives different challenges, promoting a more comprehensive understanding of the content (LIGUORI & WINKLER 2020). The combination of answer keys and randomized questions makes the assessment environment more secure and effective, standing out as a practice of excellence in higher education, especially in courses such as Agronomy and Veterinary Medicine, which require robust and differentiated assessments.

Several algorithms were used in generating personalized tests for each student based on a randomized answer key. Among the main features are: the use of regular expressions to identify patterns and locate the questions in the document; algorithms for reading and processing .csv and .xlsx files; shuffling and mapping algorithms that associate the randomized order of the questions with each student's answer key; conversion of tables present in the questions into images using the matplotlib library; and the generation of personalized Word documents, inserting text and images according to each student's specific answer key. Figure 11 illustrates an example of a test paper, given to the student with randomized questions, where the first question corresponds to Q1, followed by Q3 and Q2, exactly according to the randomized answer key generated for that student.

The correct correspondence between the generated test and the answer key is fundamental to ensure that the assessment takes place in a fair and accurate manner, guaranteeing that each student is evaluated based on the specific set of questions assigned to them. This type of automated and highly customizable classification method has been used by a number of programmers around the world, due to its ability to apply a standard for classifying a task and managing libraries in order to assist in conducting tests in schools (AKAHANE et al. 2015).

After the randomization and generation of the test papers, the test-taking phase begins. Students can fill out the randomized answer sheet by marking the correct item with an "x" or using the corresponding numerical character, such as 1, 2, or 3 for each question. This automation is essential to avoid errors in reading the answers. In questions involving calculations, the student must enter the calculated value directly into the question, such as a probability (0.2525) associated with an event. An important aspect is the anti-fraud option, which penalizes random answers. If the student incorrectly marks an alternative, the system eliminates a correct answer, ensuring that the assessment reflects actual knowledge and not "luck" (KASINATHAN et al. 2022).

.

Nome do Aluno: Pedro Henrique

Questão Q1:Q1. Os dados a seguir representam o rendimento de grãos de parcelas de soja, em gramas. 120; 100; 150; 250; 250; 155; 175; 185; 190; 200; 210; 215; 220. Pergunta-se: qual a escala de medida adotada? Qual o rendimento médio de grãos? A variação em g do conjunto? A conversão desta variação em percentual? Uma estimativa de erro na amostragem? Com base nisso, responda assinalando no gabarito apenas a(as) alternativa(s) correta(s): (1) Ordinal; 187; 45,74; 25% e 13. (2) Racional; 187; 45,74; 25% e 13; (3) Intervalar; 186,15; 45,74; 24,57% e 12,68; (4) Racional e Ordinal; 187; 45,90; 23,9% e 15; (5) Racional; 186,15; 45,74; 24,57% e 12,68; (6) Nenhuma das anteriores.

Questão Q2:Q3. Os dados a seguir representam a frequência de plantas de cada cultivar de soja: Zeus=10; Fibra= 9; Vênus=16; Valente= 18; Raio= 20 e Trovão =5. Com isso, assinale a/as alternativas corretas: (1) As cultivares fibra, raio e trovão apresentam a predominância de plantas; (2) Os extremos dessa distribuição são representados por Trovão e Raio; (3) O maior percentual de plantas ocorre em Raio; (4) O menor percentual de plantas é observado em Zeus; (5) Valente, Zeus e Fibra representam 47,42%; (6) Valente, Zeus e Fibra representam 47,44%; (7) A variável foi medida em escala nominal.

Figure 11. Example of a test document generated for each student eligible to take the assessment.

Several algorithms have been implemented for this process. The pd.read_excel function is used to read the students' answers and the answer key, while the correct_exam function compares the answers with the answer key and calculates the final score. The grades are then organized and saved in an Excel file using the pd.DataFrame.to_excel function from the Pandas library, which structures the data into rows and columns to facilitate the recording and analysis of the results.

The use of digital technologies to automate the evaluation of semester results in a course becomes important because it reduces or eliminates subjectivity in the assessment of students. The arduous task of grading a large number of exams in a single semester, within a limited timeframe, can lead instructors to make grading errors, negatively impacting their students' performance (SANDVE et al. 2013, SMETANA & BELL 2012). On the other hand, this system requires training, since, in general, students from various educational institutions may not be familiar with the evaluation method, and errors such as incorrect filling of the answer sheet by students may occur (KASINATHAN et al. 2022, BLATTLER et al. 2023). Figure 12 illustrates how the student's grade is assigned, in comparison to the teacher's answer key.

A	B	C	D	E	F	G	H	I	J	K
QUESTAO	1	2	3	4	5	6	7	8	9	
Q1					**5**	**0.2525**				
Q2	**x**	**x**		**x**			**x**			
Q3	**1**	**2**	**3**		**5**		**7**			

Gabarito do professor

A	B	C	D	E	F	G	H	I	J	K
Questão	Questão Original	Subitem 1	Subitem 2	Subitem 3	Subitem 4	Subitem 5	Subitem 6	Subitem 7	Subitem 8	Subitem 9
Q1	Q3	1	2	3		5		7		
Q2	Q2	x	x		x			x		
Q3	Q1						0.2121	7		

Aluna: Beatriz Santos

A	B	C	D	E	F	G	H	I	J	K
Questão	Questão Original	Subitem 1	Subitem 2	Subitem 3	Subitem 4	Subitem 5	Subitem 6	Subitem 7	Subitem 8	Subitem 9
Q1	Q2	x	x		x			x		
Q2	Q3					5		7		
Q3	Q1				4		0.1111			

Aluna: Mariana Souza

Figure 12. Example of student responses when completing the assessment.

The Figures demonstrate how the grades were assigned to students Beatriz and Mariana by comparing their answers with the instructor's answer key. The student Beatriz answered question Q1 (which corresponds to question 3 in the instructor's original answer key) completely correctly, marking sub-items 1, 2, 3, 5 and 7 correctly. In question Q2 (corresponding to Q2 in the original answer key), she also answered correctly, marking "x" in columns 1, 2, 4, and 7. However, in question Q3 (which corresponds to Q1 in the instructor's answer key), Beatriz made a mistake, marking 0.2121 in column 6 and the number 7, resulting in zero points for that question. As a result, Beatriz achieved a final grade of 6.0. The student Mariana, on the other hand, answered question Q1 (corresponding to Q2 in the original answer key) completely correctly, marking "x" in columns 1, 2, 4 and 7. In question Q2 (which corresponds to Q3 in the instructor's answer key), she answered partially correctly, marking sub-items 5 and 7 correctly, but failing to mark numbers 1, 2, and 3, which resulted in a partial score. In question Q3 (corresponding to Q1 in the instructor's answer key), Mariana answered completely incorrectly, marking the numbers 4 and 0.1111, resulting in zero points for that question. As a result, Mariana obtained a final grade of 4.6.

The automated system performs these instant comparisons, ensuring that incorrect answers cancel out correct answers, guaranteeing a fair assessment. The

students' answers are entered into the "student_answer_sheets" file, and the grades are assigned according to the established rules. This grading method not only prevents possible fraud or attempts to "guess" answers, but also promotes a more rigorous assessment, where the student's real knowledge is effectively tested. Invalidating correct answers in case of incorrect marking forces students to reflect on their answers, reducing the likelihood of random markings and encouraging a more careful analysis of the alternatives (BASHITIALSHAAER et al. 2021, ELTAHIR et al. 2022).

Besides, this automated grading system increases the transparency and objectivity of the evaluation process, as it eliminates potential human errors when manually checking the exams. The standardization provided by this tool is especially beneficial in large-scale assessments, such as those applied in subjects with many students, ensuring that everyone is evaluated fairly and consistently. With the ability to generate different versions of exams, the system also becomes more robust against possible collaboration between students, since each student can receive a different combination of questions and answer keys. This mechanism also allows the instructor to focus on more strategic aspects of assessment, such as the qualitative analysis of student performance, instead of worrying about the actual grading process itself. By reducing the time and effort required for grading, the instructor can dedicate more attention to reviewing the concepts that proved most difficult, promoting a more efficient feedback cycle aligned with the students' learning needs (AYOUB-AL-SALIM & ALADWAN 2021)

In this system, the instructor types in the students' answers or allows the students themselves to enter their answers. To do this, the instructor simply logs into the AutoCorrect system and clicks on the "access answer key" option.

Upon clicking "access answer key," the user is presented with the following screen, where the password is the student's first name (Figure 13).

For this operation to work, the instructor must upload the file called "answer_keys". This file comes from the "randomized_answer_keys" file, and from there, using an available computer, the student can enter their answers according to the questions present in their assessment.

From that point on, the student themselves is responsible for entering the chosen answers for each question. In addition to creating randomized test templates, customized tests with ordered questions, and automatic answer correction, AutoCorrect also allows for the analysis of student performance through calculations of descriptive measures and their graphical representation. These descriptive measures are essential for understanding the distribution of scores and the overall performance of students in relation to the test.

Questão	Subitem 1	Subitem 2	Subitem 3
Q1	None	None	None
Q2	None	None	None
Q3	None	None	None

Salvar Respostas

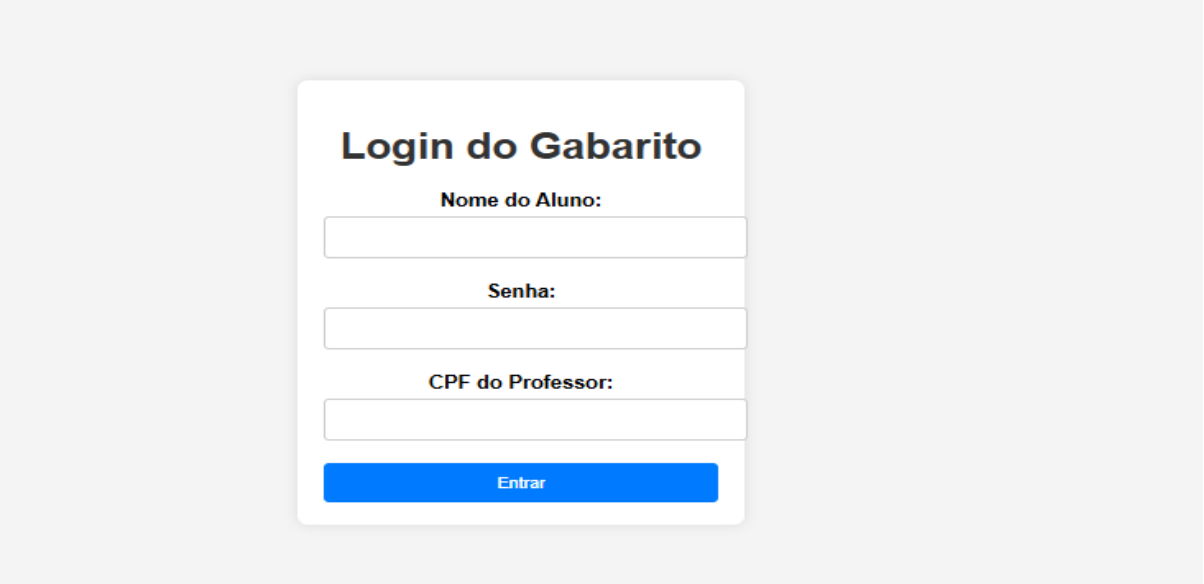


Figure 13. Access for students to fill in their test answers.

Figure 14 presents the descriptive statistics of the tests, where it is possible to observe the overall mean grade (7.33), which indicates a satisfactory performance in the class. The median, which is the central value of the distribution, was slightly below the mean, at 6.95, which suggests a slight skew towards lower grades. The standard deviation of the mean of 2.32 and the variance of 5.39 reveal that there was considerable dispersion in the students' grades, which may indicate significant differences in their level of knowledge.

Furthermore, the minimum value recorded was 4.6, which shows that no student received an extremely low grade, while the maximum value was 10.0, suggesting that some students managed to achieve maximum performance. The asymmetry of 0.09 indicates that the grades are evenly distributed around the class mean (0.09 ≈ 0). The negative kurtosis (-1.72) indicates that the distribution of grades is platykurtic, meaning it is flatter than a normal distribution, indicating a greater concentration of grades around the mean. These measures, when applied in conjunction with the automatic grading system, allow for an accurate analysis of student performance and provide the instructor with valuable insights into the strengths and weaknesses of the class (WEISSGERBER et al. 2016) The combination of randomized answer keys, automated grading, and statistical analysis makes the evaluation process more robust and data-driven, facilitating the planning of more effective pedagogical interventions.

.

.

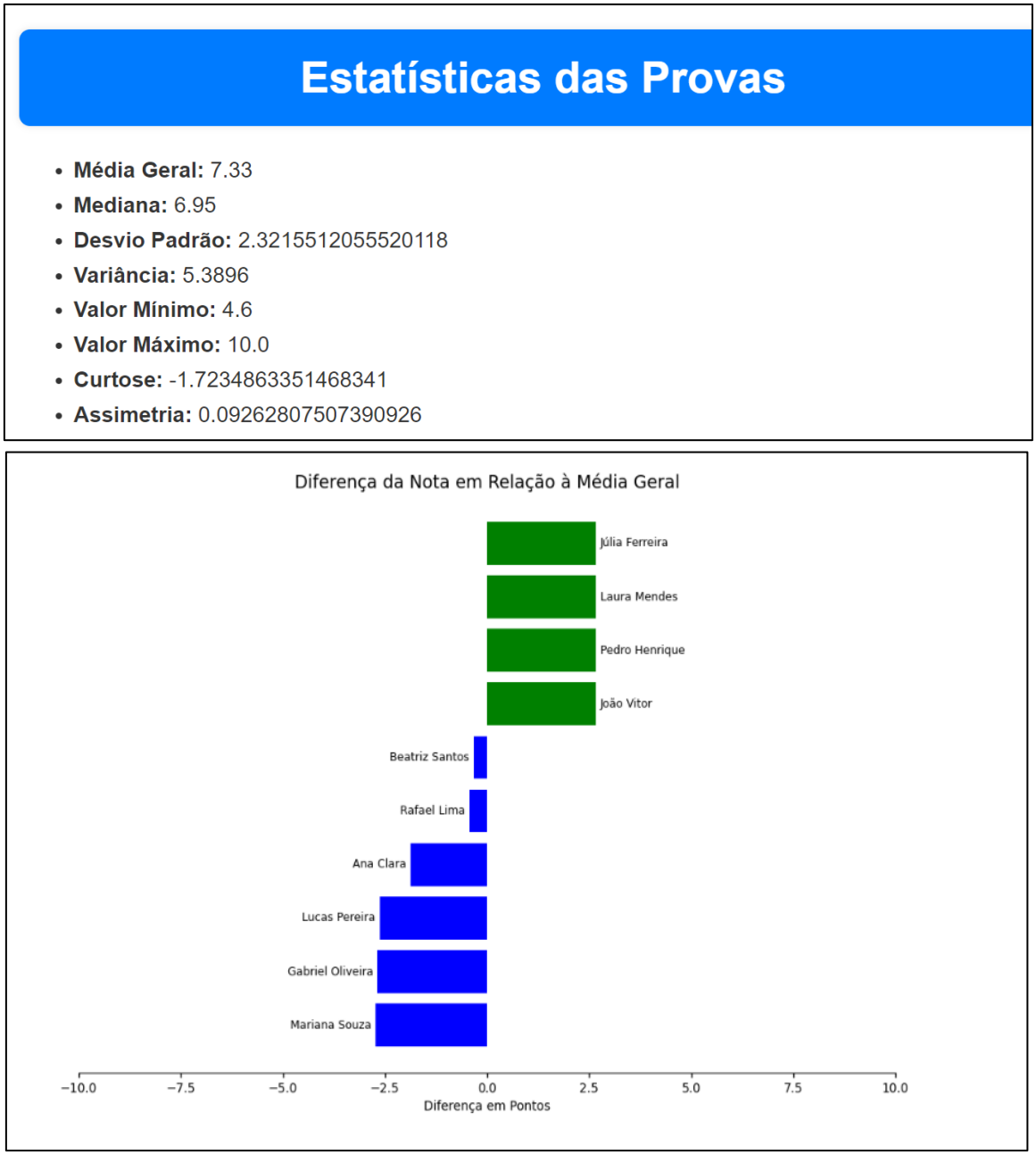


Figure 14. Calculation of descriptive measures and graphical representation of student performance.

In addition to descriptive statistics, the graph illustrates the difference between each student's grade and the overall class mean. This visualization allows for quick identification of which students performed above or below the mean of 7.33. Based on the analysis of the figure, we observe that students Júlia Ferreira, Laura Mendes, Pedro Henrique, and João Vitor obtained grades above the mean, with Júlia Ferreira standing out as having the best performance, with a positive difference of approximately 2.5 points compared to the mean. On the other hand, students like Mariana Souza, Gabriel Oliveira, Lucas Pereira, and Ana Clara performed below the mean, with Mariana Souza showing the largest negative difference, about 3 points below the overall mean. Beatriz Santos and Rafael Lima achieved scores close to the mean, indicating less variation compared to the other students.

This graphical representation is useful for the instructor to clearly visualize the relative performance of the students, allowing them to identify those who may need additional support or content review. At the same time, students with above-mean performance can be recognized for their good results. The use of charts like this helps in making targeted pedagogical decisions, facilitating the monitoring of student progress in a visual and objective way (SIAL 2021).

Through the development of this system, its usefulness in the field of education is observed, promoting precise gains in student assessment. Table 2 shows a brief comparison between the main systems used in the educational environment in the country and AutoCorrect.

Table 2. Comparison between main educational systems used in the Brazilian educational environment.

System	Advantages	Disadvantages
Google Classroom	Integration with Google tools, free for educational institutions, access via browser.	Limited customization, dependence on a good quality internet connection.
Microsoft Teams	Integration with Microsoft tools, synchronous and asynchronous classes, creating and grading tests via Forms.	Complex for new users, the free version has limitations, and it consumes resources from the user's machine.
Moodle	Open source, tools for creating tests, does not depend on servers.	Requires technical knowledge for configuration, less intuitive interface, slow system.
AutoCorrect	Open-source system, flexible to the instructor's needs, efficiency and accuracy in grading, use of simple files, and plagiarism control among students.	Dependence on strict answer sheets, need to type student answers onto the answer sheets.

Source: Adapted from MUSTAFA & ALI 2023, RODILLAS et al. 2023.

Based on these analyses and graphical representations, it is evident that data science has a transformative impact on the automation of exam grading. The application of advanced algorithms for the randomization of answer keys, automatic correction, and generation of descriptive statistics resulted in greater efficiency and accuracy in the evaluation process. This automated system not only minimizes human error, but also promotes a fair, standardized, and transparent evaluation.

Visualizing the results in graphs and tables facilitates the identification of student performance patterns, allowing instructors to implement effective pedagogical interventions (HAND 2019). By adopting these tools, educational institutions can optimize instructors' time and ensure a more balanced and data-driven assessment of classes each semester.

CONCLUSION

Automating the grading of exams using tools such as Python and online AutoCorrect has proven to be a robust solution for managing large-scale assessments. The integration of algorithms and data analysis libraries allowed not only for efficient and accurate correction, but also provided immediate insights into student performance. Furthermore, the personalization of feedback and the elimination of human errors made the process more efficient and reliable. This modernization of the evaluation process provides significant benefits for both instructors, who can focus on more strategic aspects of teaching, and students, who receive a fair and accurate assessment in less time.

AUTHOR'S CONTRIBUTIONS

Conceptualization, methodology, and formal analysis, Paulo Henrique Cerutti; software and validation, Jefferson Luís Meirelles Coimbra; investigation, Mauro Bitencourt de Souza, Luan Tiago dos Santos Carbonari; resources and data curation, Jefferson Luís Meirelles Coimbra; writing - original draft preparation, Paulo Henrique Cerutti; writing - review and editing, Paulo Henrique Cerutti, Carlos Joquim Zacarias Júnior, Henrique de Sá Albino; visualization, Paulo Henrique Cerutti; supervision and project administration, Jefferson Luís Meirelles Coimbra; funding acquisition, Jefferson Luís Meirelles Coimbra. All authors have read and agreed to the published version of the manuscript.

FINANCING

This work was not supported by any funding agency.

STATEMENT OF THE INSTITUTIONAL REVIEW BOARD

Not applicable to studies that do not involve humans or animals.

INFORMED CONSENT STATEMENT

Not applicable because this study did not involve humans.

DATA AVAILABILITY STATEMENT

The data can be made available upon request.

ACKNOWLEDGEMENTS

We thank the State University of Santa Catarina (UDESC) and the Center for Agroveterinary Sciences (CAV) for their intellectual support.

CONFLICTS OF INTEREST

The present manuscript has no conflicts of interest.

REFERENCES

ALDRIYE H et al. 2019. Automated grading systems for programming assignments: A literature review. International Journal of Advanced Computer Science and Applications 10: 250-259.

AKAHANE Y et al. 2015. Design and evaluation of automated scoring: Java programming assignments. International Journal of Software Innovation 3: 18–32.

AYOUB-AL-SALIM MI & ALADWAN K. 2021. The relationship between academic integrity of online university students and its effects on academic performance and learning quality. Journal of Ethics in Entrepreneurship and Technology 1: 43-60.

BASHITIALSHAAER R et al. 2021. Obstacle comparisons to achieving distance learning and applying electronic exams during COVID-19 pandemic. Symmetry 13: 99.

BLATTLER A et al. 2023. One-Shot Grading: Design and Development of an Automatic Answer Sheet Checker. In: 2023 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM). IEEE 562-566.

DOE JO. 2024. Uso de Seed Fixa para Garantir Reprodutibilidade em Processos Computacionais. Journal of Computational Reproducibility 1: 1-5.

ELTAHIR E et al. 2022. Implementation of E-exams during the COVID-19 pandemic: A quantitative study in higher education. Plos One 17: e0266940.

HAND D. 2019. What is the purpose of statistical modeling? Harvard Data Science Review 1: 1-6. https://doi.org/10.1162/99608f92.4a85af74.

HELMUS JJ & COLLIS SM. 2016. The Python arm radar toolkit (py-art), a library for working with weather radar data in the Python programming language. Journal of Open Research Software 4: 1-12.

JAMES G et al. 2023. Introduction. In: An Introduction to Statistical Learning. Springer Texts in Statistics. Stanford: Department of Statistics. https://doi.org/10.1007/978-3-031-38747-0_1

KASINATHAN V et al. 2022. ProctorEx: An Automated Online Exam Proctoring System. Mathematical Statistician and Engineering Applications 71: 876–889.

KURDI G et al. 2020. A systematic review of automatic question generation for educational purposes. International Journal of Artificial Intelligence in Education 30: 121-204.

LIGUORI E & WINKLER C. 2020. From offline to online: Challenges and opportunities for entrepreneurship education following the COVID-19 pandemic. Entrepreneurship Education and Pedagogy 3: 346-351.

MASSARI CHAL et al. 2022. Veterinary anatomy during the COVID-19 pandemic in Brazil: Research focused on pedagogical practice. International Journal of Morphology 40:79-83.

MUSTAFA AS & ALI N. 2023. The adoption and use of Moodle in online learning: A systematic review. Information Sciences Letters 12: 341-351.

RODILLAS MJ et al. 2023. Mediating Effects of Teacher's Assessment Competencies on the Relationship Between the Use of the Zipgrade Application and Student's Achievement. Psychology and Education: A Multidisciplinary Journal 11: 862-876.

ROLON-MÉRETTE D et al. 2016. Introduction to Anaconda and Python: Installation

and setup. Quantitative Methods for Psychology 16: 3-11.
SANDVE GK et al. 2013. Ten simple rules for reproducible computational research. PLoS Computational Biology 9: e1003285.
SIAL M. 2021. A Brief Introduction to Regression Analysis and Its Types. Asian Journal of Probability and Statistics 13: 58-63.
SMETANA LK & BELL RL. 2012. Computer simulations to support science instruction and learning: A critical review of the literature. International Journal of Science Education 34: 1337-1370.
VAN DER AALST W. 2016. Data Science in Action. Berlin: Springer. https://doi.org/10.1007/978-3-662-49851-4_1.
WEISSGERBER TL et al. 2016. From static to interactive: transforming data visualization to improve transparency. PLoS Biology 14: e1002484.
ZHANG L et al. 2022. An automatic short-answer grading model for semi-open-ended questions. Interactive Learning Environments 30: 177-190.