https://doi.org/10.53656/math2026-3-6-adm

APPLICATION OF DATA MINING TO MONITOR THE LEARNING OUTCOMES OF HIGHER EDUCATION STUDENTS

Olha Korotun, Galyna Marchuk, Vitalii Levkivskyi, Oleksandra Svintsytska
Faculty of Information and Computer Technologies
Zhytomyr Polytechnic State University, Ukraine

Abstract. Analysis of student test results is an important stage in the learning process. It provides valuable information about the effectiveness of the learning process, identifies problem areas and develops measures to address them. The problem of choosing a method for assessing the level of training, the quality of knowledge, skills and abilities is extremely important. In order to improve the efficiency of the educational process in the discipline "Programming", a comprehensive analysis of student test results was carried out. The study included both quantitative data analysis (average score, percentage of correct answers) and qualitative analysis of typical mistakes and difficulties faced by students. Five tests covering the entire spectrum of the C programming discipline were selected for the study. The use of the R programming language allowed for in-depth data analysis and a detailed picture of students' learning achievements. The study proves that data analysis can be a powerful tool for improving the educational process. A detailed analysis of the test results of students allowed us to identify weaknesses in the curriculum and develop a model that can predict student performance. The analysis can be used to create more individualized curricula and improve the overall efficiency of the educational proces.

Keywords: data analysis; test results; descriptive statistics; forecasting; R language; education quality; monitoring

1. Introduction

In the modern world, distance learning is becoming increasingly popular, which creates a need for effective monitoring of the quality of the educational process. Monitoring studies reveal aspects such as the alignment or misalignment of student preparation with the required task level, potential gaps in the study of specific topics, and the assessment of learning progress. Testing can be one of the components of overall academic performance monitoring. To assess the level of knowledge acquisition, it is necessary to analyse test results.

The analysis of such data will contribute to improving the quality of education and personalizing learning at the university, specifically by identifying individual learning trajectories for students and adapting educational materials to their level of knowledge and needs. It should be noted that data analysis is just a tool that must be used in conjunction with traditional teaching methods while taking into account the human factor. This approach is becoming an essential tool in modern education. However, it is crucial to remember that the combination of human intuition and the analytical capabilities of data analysis enables the creation of an optimal learning environment.

Thus, the analysis of student testing is necessary to assess and improve the quality of the educational process in the subject, identify and eliminate knowledge gaps, motivate students, plan curricula, and ensure compliance with educational standards.

2. Problem statement

The aim of the study is to improve the efficiency of the educational process in the "Programming" course by monitoring student learning quality based on the results of testing using data mining methods. To achieve this objective, the following methods must be applied: statistical analysis to identify patterns, trends, and differences in test results; visual analysis to provide a clear representation of data and examine results in detail; cluster analysis to identify student groups with different knowledge levels and test difficulty; and regression analysis to determine the dependencies of various factors.

3. Theoretical background

Ensuring high-quality education is a complex task that requires not only the development of new curricula but also the implementation of effective systems for assessing students' knowledge, as well as engaging and motivating them to learn. The authors of this paper (Marchuk et al., 2023) emphasise the importance of motivation and propose an innovative software system that helps to engage students in learning. This system was integrated into the educational process at Zhytomyr Polytechnic State University and demonstrated its effectiveness in engaging students in both learning and other activities. The research (Pol et al., 2024) showed that teachers consistently misjudge both students' ability for self-assessment and their actual academic achievements. Teachers tend to underestimate students' awareness of their own learning processes and overestimate their actual results. This indicates the need for new approaches in teacher training and the development of tools for more accurate assessment of both students' monitoring skills and their academic performance.

The article (Bilyakovska, 2022) focuses on an important aspect of modern pedagogy – the evaluation of students' knowledge levels through testing. Testing serves as an effective tool for objectively measuring how well students have mastered the learning material and developed the necessary competencies. The authors have developed a detailed plan, including organizational and methodological measures, for conducting high-quality testing. In the paper (Pantsyr and Semenyshena, 2024), a thorough analysis of testing as a tool for assessing students' academic achievements is conducted. It explores how tests can be used to develop key competencies and measure the level of knowledge acquisition. The authors also examine how technology can improve the objectivity and reliability of the testing process. Particular attention is given to the creation of test tasks for technical disciplines. The authors of this paper (Bergbauer et al., 2024) have conducted a study on the impact of different testing modes on student achievement. The study is based on data collected from 59 countries worldwide. The results highlight that in low- and middle-performing countries, more standardized testing is associated with higher student achievement, while the introduction of additional monitoring in tests has a negative impact. The paper (Donets et al., 2023) emphasizes that testing plays a distinct role in the overall quality assurance system of the educational process in higher education institutions. When properly organised, tests help students critically evaluate their achievements, provide information on how well they are assimilating learning material, identify aspects of the educational process that require improvement, and assist in determining which corrective measures should be introduced into the content and formats of students' cognitive activities. An empirical study (Marchuk et al., 2024) aimed at identifying the variance in academic performance among different student groups demonstrated significant differences in the level of training. The findings indicate the necessity of a differentiated approach to the organization of the educational process. The use of data visualisation made it possible to identify groups of students who need additional support. The publication (Hill and LoPalo, 2024) focuses on the research of online and offiine testing of college students. The authors emphasize that the testing method can have significant consequences for student performance and fairness of learning outcomes. As a result, it was found that students perform significantly better on online tests than offiine ones, but the result is significantly reduced when questions that have never been used before are introduced. The article (Barkovska et al., 2024) explores the use of eye-tracking analysis to improve online testing. The proposed model allows for automatic tracking of where students are looking during a test, providing a more objective assessment of their knowledge. This is particularly relevant for distance learning, which has become widespread in Ukraine. The

research focuses on developing tools for the objective assessment of students' knowledge and skills. The study (Rodríguez-Villalobos et al., 2023) aims to compare the effectiveness of different test formats in distance education (unsupervised, traditionally supervised, and software-supervised) and to analyse their impact on student performance. The findings suggest that the average scores of students who took online tests under remote supervision were seven points lower than those of traditional testing. However, this gap does not necessarily indicate an increase in academic integrity during online testing. Possible reasons for lower scores may include test anxiety, technical issues, or other factors that could have affected the reliability of the obtained results.

Student performance plays a crucial role, as it is often used as an indicator of an educational institution's effectiveness. Early identification of at-risk student groups, combined with preventive measures, can significantly improve their performance. Recently, machine learning methods have become more widely used for this purpose. In studies (Hooda et al., 2022) and (Alyahyan and Düştegör, 2020), the authors analysed the most commonly used artificial intelligence and machine learning algorithms and emphasised their potential for improving student performance in higher education institutions. While traditional artificial intelligence and data mining technologies have become more frequently applied in education, more advanced methods are still rarely used (Ouyang et al., 2022).

In the paper (Niessen et al., 2016), test results are used to predict students' academic achievements and progress in certain types of courses. All tests demonstrated significant positive correlations with academic criteria. The authors concluded that test results are a good tool for predicting academic performance. The analysis of scientific publications highlights the importance and necessity of continuous monitoring of students' academic performance and the application of advanced technologies for this purpose.

An analysis of scientific literature shows that despite the growing use of artificial intelligence and data mining technologies in education, modern analytical methods are still rarely used, and the subjectivity of teachers' assessments often leads to inaccuracies in determining students' actual achievements. The current study complements existing research on the role of testing in ensuring the quality of education by proposing a transition from simple recording of grades to comprehensive intellectual data analysis for monitoring learning outcomes. The originality of the contribution lies in the use of the R programming language for in-depth analysis in the context of studying the discipline "Programming" (C language), which made it possible to combine methods of descriptive statistics, visualization, and machine learning. In particular, the use of the k-means method for clustering students and assessing the complexity of test tasks, as well as the development of a multiple linear

regression model for predicting final scores based on interim results, ensures high monitoring accuracy and the possibility of personalizing the learning trajectory.

4. Methods

The "Programming" course was chosen as the basis for this study. It is taught to first-year students of the following specialities at the Faculty of Information and Computer Technologies of Zhytomyr Polytechnic State University: 125 "Cybersecurity" (academic groups KB-23-1, KB-23-2), 123 "Computer Engineering" (academic group KI-23-1), 122 "Computer Science" (academic groups KH-23-1, KH-23-2, KH-23-3), and 126 "Information Systems and Technologies" (academic group ICT-23-1). A total of 177 students participated in the testing. The course was conducted in a blended format – both in-person and online – depending on the schedule and circumstances (such as air raid alerts in the region and other factors).

Systematic monitoring of student test results involves the collection, thorough analysis and interpretation of data on students' academic achievements. Five tests from the "Programming" course were selected for the study. The tests were designed to assess students' knowledge of fundamental concepts, such as data types in C, operations and operators (Test 1); branching and loop operators (Test 2); one-dimensional and two-dimensional arrays and their sorting (Test 3); functions and recursion (Test 4); and pointers and strings (Test 5). The results of the final test, which covers the entire course material and is more challenging than the previous tests, were also used for the analysis. The number of tasks in each test is adapted to the complexity of the material and the time allocated. The distribution of points is proportional to the complexity and importance of the topic. The use of different types of questions (closed-ended, open-ended, matching, etc.) contributes to a more objective assessment of students' knowledge and skills. Closed-ended questions require students to choose an answer from a given set of options. Open-ended test questions, on the other hand, require students to formulate their answers independently, without provided choices. Matching test tasks involve pairing elements from two given lists. Each of the tests has a different number of questions and a different point value for each question. The first test consists of 30 questions, each worth 0.56 points for a correct answer; the second test includes 18 questions, each worth 0.33 points; the third and fifth tests have 20 questions, each worth 0.5 points; the fourth test consists of 45 questions, each worth 0.22 points. The final test comprises 20 questions, each worth 0.5 points. Thus, the assessment system is based on the principle of differentiating the complexity and significance of the material.

The study was conducted on a sample of 177 students. The input vari-

ables were scores from five thematic tests, and the target variable was the final test score. Prior to the analysis, the completeness of the data was verified, the rating scale was normalized to the interval [0, 10], distribution analysis was performed (QQ plots were used to assess normality), and outliers were identified using the interquartile range. Mean values, medians, standard deviations, and coefficients of variation were calculated to assess group homogeneity, which made it possible to evaluate the level of central tendency, the homogeneity of the groups, and the presence of "bottlenecks" in specific topics. The k-means method was used to group students by academic performance level (low, medium, high), which made it possible to identify a "risk group." In addition, test tasks were clustered according to their complexity. Clustering was performed based on the average score for each question. In the future, this will allow for a review of the course structure and subsequent adjustments to the teaching methodology of the discipline. A linear regression model was constructed to predict the final result, and a logistic regression model was developed to classify students (successful/unsuccessful). The quality of the models was evaluated using R^2, RMSE, Accuracy, and ROC–AUC. Interpretation of significant linear regression coefficients shows which topics have the greatest impact on the final result. Boxplots, scatter plots, and heatmaps were used to interpret the results as tools to support management decisions.

5. Results

5.1. Descriptive statistics and preliminary data analysis

For the analysis, we use the R programming language and the RStudio environment. To do this, the test results were downloaded from the Moodle learning management system for the corresponding course in the form of Excel files.

The internal structure of the dataset containing the results for Test 2 (Figure 1) is presented in the form of numeric and character data types. The datasets for other tests have a similar structure.

To obtain detailed information about the test results and identify patterns that affect student performance, we proceed to the second stage of data analysis. We begin by reviewing the descriptive statistics of the obtained datasets, which reflect the key statistical indicators of our empirical datasets, presented in the form of a listing:

```
> summary(test1$‘Rating/10,00‘)
          Min    1st Qu  Median  Mean   3rd Qu  Max
          3.670  9.000   9.670   9.275  10.000  10.000
> summary(test2$‘Rating/10,00‘)
```

```
> str(test1)
tibble [136 × 34] (S3: tbl_df/tbl/data.frame)
 $ Lastname     : chr [1:136] "Bailo" "Balenko" "Bezshkura" "Bilan" ...
 $ Name         : chr [1:136] "Oleksandr" "Hlieb" "Ielyzaveta" "Maksym" ...
 $ Group        : chr [1:136] "KH-23-2" "KB-23-2" "KH-23-2" "KH-23-3" ...
 $ Raiting/10,00: num [1:136] 10 9.67 10 7.33 7 8.33 9.67 9.33 10 9.33 ...
 $ Q.1 /0,33    : num [1:136] 0.33 0.33 0.33 0.33 0 0.33 0.33 0.33 0.33 0.33 ...
 $ Q.2 /0,33    : num [1:136] 0.33 0.33 0.33 0.33 0.33 0.33 0.33 0.33 0.33 0.33
 $ Q.3 /0,33    : num [1:136] 0.33 0.33 0.33 0.33 0.33 0.33 0.33 0 0.33 0.33 ...
 $ Q.4 /0,33    : num [1:136] 0.33 0.33 0.33 0 0 0 0.33 0.33 0.33 0.33 ...
 $ Q.5 /0,33    : num [1:136] 0.33 0.33 0.33 0.33 0.33 0.33 0.33 0.33 0.33 0.33
 $ Q.6 /0,33    : num [1:136] 0.33 0.33 0.33 0.33 0.33 0.33 0.33 0.33 0.33 0.33
 $ Q.7 /0,33    : num [1:136] 0.33 0.33 0.33 0 0 0.33 0.33 0.33 0.33 0.33 ...
 $ Q.8 /0,33    : num [1:136] 0.33 0.33 0.33 0.33 0 0 0.33 0.33 0.33 0.33 ...
 $ Q.9 /0,33    : num [1:136] 0.33 0.33 0.33 0 0 0.33 0.33 0.33 0.33 0.33 ...
 $ Q.10 /0,33   : num [1:136] 0.33 0.33 0.33 0 0 0.33 0 0.33 0.33 0 ...
```

Figure 1. Internal structure of the dataset containing the results for Test 2

```
          Min    1st Qu  Median  Mean   3rd Qu  Max
          2.220  5.560   7.780   7.189  8.330   10.000
> summary(test3$‘Rating/10,00‘)
          Min    1st Qu  Median  Mean   3rd Qu  Max
          1.000  6.500   7.500   7.312  8.000  10.000
> summary(test4$‘Rating/10,00‘)
          Min    1st Qu  Median  Mean   3rd Qu  Max
          0.220  6.890   7.560   7.312  8.220  9.780
 > summary(test5$‘Rating/10,00‘)
          Min    1st Qu  Median  Mean   3rd Qu  Max
          5.000  8.000   8.500   8.606  9.500  10.000
```

The first aspect to examine is the minimum and maximum scores achieved in the tests. In almost all tests, there were students who successfully completed the tasks and received the highest score of 10 points. We can conclude that the range of minimum scores varies significantly, from 0.2 points, indicating that only one task was completed, to 5 points, which corresponds to completing half of the tasks. Another metric is the calculation of the average score for each test, which ranges from 7.2 to 9.3 points, showing a difference of approximately 2 points. The data spread is small, indicating that most students received relatively high test scores and successfully completed more than 75% of the tasks, meaning their level of preparation was above average. Additionally, the listing includes the values of the 1st and 3rd quartiles, which represent the scores below 25% and 75%, respectively. Thus, for Test 2, students received the lowest scores, as 25% of them managed to complete only 50% of the test. This is further confirmed by the 3rd quartile value, indicating that the topic “Branching and Loop Operators” was particularly

challenging for them. There is a distinct group of students who have not mastered the basic logic of algorithms. This indicates that this topic is a "bottleneck" that requires additional practical hours, or a change in teaching methods to overcome the cognitive gap. In contrast, the topics "Data Types in C, Operations and Operators" and "Pointers and Strings" were relatively easy, as evidenced by the obtained values of the 1st and 3rd quartiles.

For speciality 122, the dataset includes three groups (KH-23-1, KH-23-2, KH-23-3). The knowledge levels of students in groups KH-23-1 and KH-23-2 are almost at the same level, while the knowledge level of students in group KH-23-3 is slightly lower. Regarding speciality 123, group KI-23-1, the test scores indicate that students in this group have a lower level of knowledge compared to students in other groups. The situation is slightly better for students of speciality 125, as both groups KB-23-1 and KB-23-2 performed almost equally well across all tests. Students of the group ICT-23-1, speciality 126, have the same level of preparation as students of speciality 123.

We will now analyse the standard deviation values to understand the extent to which individual data points for students in different groups deviate from the mean. This will help us assess the homogeneity of the data within the established datasets. For Test 1, the highest standard deviation is observed among students in groups KB-23-1 and KH-23-3, indicating greater variability in their scores. In contrast, the results in other groups are more homogeneous. For Test 2, the most heterogeneous results are observed in groups KH-23-3, KB-23-1 and KB-23-2, while the most homogeneous results are found in group ICT-23-1. Regarding Tests 3 and 5, results across all groups are heterogeneous. For Test 4, the results for group KI-23-1 are the most heterogeneous compared to students in other groups, with noticeable variability also present in groups KH-23-2 and KB-23-2. The heterogeneity observed in the groups indicates a gap in the students' basic training. In terms of course outcomes, this requires a differentiated approach, where stronger students can be involved in mentoring, and weaker students can be given adapted tasks.

5.2. Visual analysis and assessment of the distribution of results

To gain a better understanding of the calculated measures of central tendency, as well as variability (the range of values within the dataset) and data skewness (the deviation of data around the mean), we will create visualisations, specifically box plots, and analyse them (Figure 2). For Test 1, a high level of knowledge is observed among students in group KH-23-1, while the lowest knowledge level is found in group KH-23-3. Outliers are present in groups ICT-23-1, KB-23-2, and KH-23-3, where the outliers correspond to low test scores. For groups KH-23-3 and KB-23-1, these values are abnormally low, indicating data variability. Data asymmetry is present in groups

ICT-23-1, KB-23-2, and KH-23-3. Moving on to the box plot for Test 2, outliers are also observed in groups KB-23-2 and KH-23-2, with data asymmetry evident in groups KI-23-1, KH-23-1, and KH-23-2. For Test 3, a single outlier is found in groups KB-23-2, KH-23-1, and KH-23-3, while data asymmetry is observed in groups KB-23-1, KI-23-1, and KH-23-3. In Test 4, two outliers appear in groups KB-23-2, KI-23-1, and KH-23-3, with data asymmetry present in groups ICT-23-1, KB-23-1, and KB-23-2.

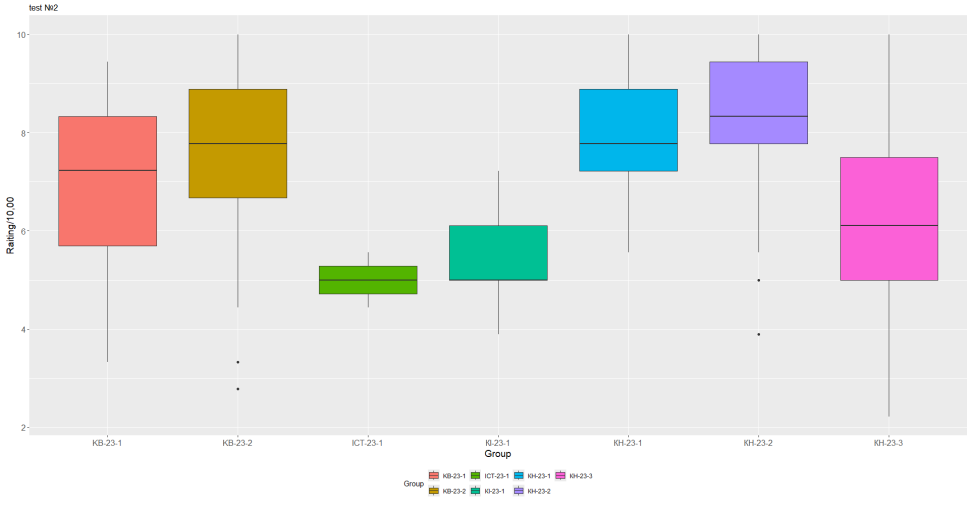


Figure 2. Box plot of student Test 2 scores by group

The results of the final Test 5 show one outlier in groups KB-23-2, KI-23-1, KH-23-2, and KH-23-3, as well as data asymmetry in groups KI-23-1, KH-23-2, and KH-23-3. Looking at the overall knowledge levels across all tests, nearly every group includes students with either a high level of preparation or, conversely, a very low level. This trend is particularly evident in Tests 2 and 5.

To assess the normality of the distribution of student test scores, Q-Q plots were constructed. The majority of student test scores across all tests demonstrate a tendency towards normal distribution, as the points in all graphs lie almost along a straight line. This trend is most evident in Tests 2 and 3. A different pattern is observed in Tests 1, 4, and 5, where the points deviate from the straight line, indicating the presence of either excessively high or low test scores in the empirical data.

Next, we will review the results of the final student assessment and present the descriptive statistics, as shown in the listing:

```
> summary(test_pids$‘Rating/10,00‘)
          Min    1st Qu  Median  Mean   3rd Qu  Max
          1.100  6.900   7.900   7.571  8.400  9.400
```

The minimum score obtained by students is 1, while the maximum score is 9.4, with an average score of 7.5. Based on the values of the 1st and 3rd quartiles, it can be concluded that students have mastered the theoretical and practical material of the course studied throughout the semester at an average level. Next, it is necessary to determine which groups have best mastered the entire course material. For each academic group, the maximum and minimum scores, average scores, and standard deviation have been calculated and are presented in the following listing:

```
   > test_pids %>% group_by(Group)%>%
+ summarise(mean_grupa  =  mean(‘Rating/10,00‘),
+ median_grupa  =  median(‘Rating/10,00‘),
+ max_grupa  =  max(‘Rating/10,00‘),
+ min_grupa  =  min(‘Rating/10,00‘),
+ sd_grupa = sd(‘Rating/10,00‘))
# A tibble: 7 × 6
        Group mean_grupa median_grupa max_grupa min_grupa sd_grupa
        KB-23-1  8.05    8.0          9.4       6.4       0.882
        KB-23-2  7.49    7.6          8.9       5.6       1.020
        ICT-23-1 5.76    6.0          9.2       1.1       2.940
        KI-23-1  6.73    7.3          8.3       3.5       1.530
        KH-23-1  7.80    8.4          9.4       3.4       1.550
        KH-23-2  8.09    8.3          9.3       6.1       0.862
        KH-23-3  7.33    7.6          9.1       1.9       1.460
```

Let us now analyse the existing standard deviations. The group ICT-23-1 has the highest standard deviation, indicating a large deviation from the average score and a significant spread of scores. This group includes both high- and low-performing students, as confirmed by the maximum and minimum scores obtained within the group. Continuing the analysis for students in groups KI-23-1, KH-23-1, KB-23-2, and KH-23-3, we see that they are predominantly made up of students with average levels of knowledge. Examining groups KB-23-1 and KH-23-2, we can conclude that the majority of students in these groups have mastered the course material to an almost high level. As a result, they are likely to find it easier to succeed in future courses that build upon the knowledge and skills gained in the “Programming” course.

5.3. Clustering students and tasks

For a more detailed study of the results of the final testing, we will conduct clustering based on the students' levels of knowledge within the groups and determine the number of students with high, average, and low levels of knowledge. The number of clusters was chosen based on a combination of pedagogical expediency and statistical analysis. From a methodological point of view, the division into three groups allowed for a clear differentiation of students by level of preparation (low, medium, high) and test tasks by degree of complexity. To mathematically confirm the optimality of this choice, the elbow method was used, which showed a characteristic "break" (elbow) at point k=3, indicating that a further increase in the number of clusters does not lead to a significant improvement in the compactness of the groups, but only complicates the interpretation of the results. To achieve this, we will use the k-means method. Its distinguishing feature is that objects within the same cluster are as similar as possible to each other, while objects in different clusters are as distinct as possible. In our case, the clustering will be conducted into three clusters, and the following results will be obtained (Figure 3). The number of students in the 1st cluster (low level of knowledge) is 6 students, in the 2nd cluster (average level of knowledge) there are 54 students, and in the 3rd cluster (high level of knowledge), the largest group consists of 69 students. Identifying the first cluster allows teachers to identify the "risk group" at an early stage. This makes it possible to apply preventive measures and individual support to them even before the final testing, which directly affects the overall success rate of the course.

To better understand the clustering results, we will visualise the distribution of students across clusters by academic group (Figure 4), showing the number of students in each cluster within each group. This allows us to determine which groups have successfully mastered the proposed learning material in the course.

We will conduct clustering based on the difficulty level of the final test questions. This will allow us to identify easy, medium, and difficult questions. Accordingly, the clustering will be performed into three clusters, and the results will be described and presented in the form of a listing.

```
> kmeans_results_test_1 <- kmeans(mean_all_quer, centers = k)
> cluster1
[1] 39
> cluster2
[1] 44
> cluster3
[1] 17
```

```
> k <- 3
>  kmeans_results_test_pids <- kmeans(test_pids$`Raiting/10,00`, centers = k)
>  test_pids$cluster <- kmeans_results_test_pids$cluster
>  print(test_pids)
# A tibble: 129 × 105
   Lastname Name  Group `Raiting/10,00` `Q.1 /0,10` `Q.2 /0,10` `Q.3 /0,10` `
   <chr>    <chr> <chr>           <dbl>       <dbl>       <dbl>       <dbl>
 1 Zahilsk… Niki… ICT-…             1.1         0.1         0.1         0.1
 2 Kravchuk Olek… ICT-…             6.7         0.1         0.1         0.1
 3 Lysiuk   Alisa ICT-…             9.2         0.1         0.1         0.1
 4 Poliekh… Bohd… ICT-…             6           0.1         0.1         0.1
 5 Stakhov  Andr… ICT-…             5.8         0.1         0.1         0
 6 Bryzhen… Dmyt… KB-2…             7.6         0.1         0.1         0.1
 7 Hrabins… Davyd KB-2…             8.3         0.1         0.1         0.1
 8 Danchuk  Illia KB-2…             6.4         0.1         0.1         0.1
 9 Derzhan… Anas… KB-2…             9           0.1         0.1         0.1
10 Zahlada  Vlad… KB-2…             7.6         0.1         0.1         0
# ℹ 119 more rows
# ℹ 93 more variables: `Q.9 /0,10` <dbl>, `Q.10 /0,10` <dbl>, `Q.11 /0,10` <db
#   `Q.14 /0,10` <dbl>, `Q.15 /0,10` <dbl>, `Q.16 /0,10` <dbl>, `Q.17 /0,10`
#   `Q.20 /0,10` <dbl>, `Q.21 /0,10` <dbl>, `Q.22 /0,10` <dbl>, `Q.23 /0,10`
#   `Q.26 /0,10` <dbl>, `Q.27 /0,10` <dbl>, `Q.28 /0,10` <dbl>, `Q.29 /0,10`
#   `Q.32 /0,10` <dbl>, `Q.33 /0,10` <dbl>, `Q.34 /0,10` <dbl>, `Q.35 /0,10`
#   `Q.38 /0,10` <dbl>, `Q.39 /0,10` <dbl>, `Q.40 /0,10` <dbl>, `Q.41 /0,10`
# ℹ Use `print(n = ...)` to see more rows
>  kmeans_results_test_pids %>%
+    augment(test_pids) %>%
+    count(cluster)
# A tibble: 3 × 2
  cluster     n
    <int> <int>
1       1    54
2       2     6
3       3    69
```

Figure 3. Results of the dataset clustering based on the final test

Thus, the third cluster, which includes questions of higher difficulty, contains 17 questions. This allows the lecturer to review the content of the teaching materials. If most students are unable to cope with certain issues, this does not indicate a weakness on their part, but rather a need to adjust the structure of the course or provide additional teaching materials to explain complex topics. The second cluster (questions of medium difficulty) includes 44 questions, while the first cluster (questions of low difficulty) consists of 39 questions. In subsequent work, the lecturer can identify the topics that need to be emphasised in teaching the course as they present difficulties for students in terms of understanding, and those topics that require less focus, as they are clearer and easier for students to grasp.

The next stage of the study will involve creating a model to predict the final

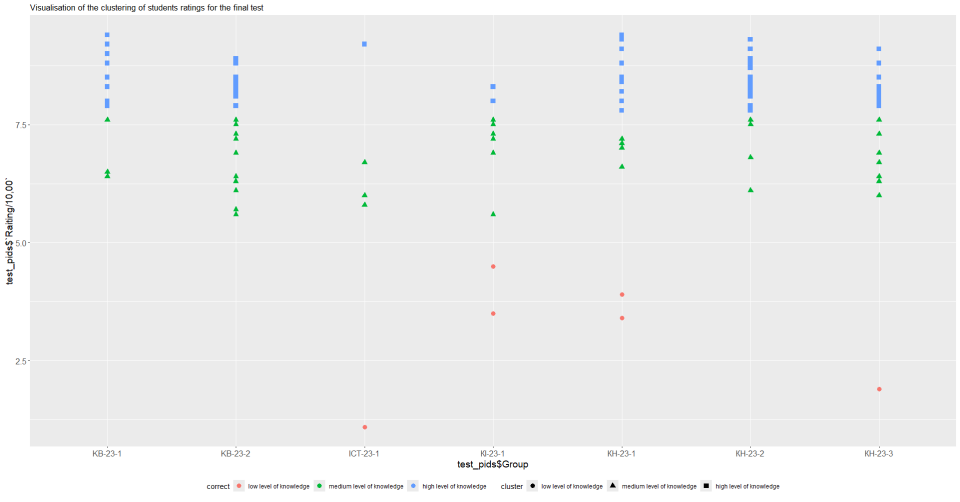


Figure 4. Visualisation of the dataset clustering with the final test results

test results of any student. To do this, we will model the relationship between the final test score of the students and the scores they previously received in interim assessments (Test 1, Test 2, Test 3, Test 4, Test 5). As an example, we will take students from the KH-23-1 group. We will apply the statistical method of multiple linear regression, which will allow us to assess how well the students have mastered the learning material throughout the semester while studying the specified course. This will help the lecturer predict their academic performance at the end of the semester. We will create a new dataset test_reg, containing students' results for all tests taken. To continue working with the new dataset, we need to examine its internal structure, paying special attention to the column types (Figure 5).

Statistical analysis and clustering allow teachers to move away from a one-size-fits-all approach and towards personalized learning. Based on identified individual trajectories (e.g., through quartile and outlier analysis), lecturers can adapt teaching materials to the knowledge level of specific students and create individualized plans. This allows students with low levels of preparation to receive additional support, while strong students receive more challenging tasks, which improves the overall quality of education.

Clustering test tasks by level is directly related to the course content. Identifying specific "problem areas," such as the topic "Branching Operators and Loops" (Test 2), where students performed the worst, indicates the need to review the teaching methodology for this particular topic. The lecturer can

```
> str(test_reg)
tibble [22 × 9] (S3: tbl_df/tbl/data.frame)
 $ Lastname  : chr [1:22] "Vorona" "Hrushnyk" "Dmytrevych" "Ivanytskyi" ...
 $ Name      : chr [1:22] "Viktoriia" "Mykhailo" "Nazarii" "Dmytro" ...
 $ Group     : chr [1:22] "KH-23-1" "KH-23-1" "KH-23-1" "KH-23-1" ...
 $ Final test: num [1:22] 8.2 8.4 8 8.5 8.5 8.8 6.6 7 7.1 9.3 ...
 $ test 1    : num [1:22] 9.33 9.67 9.33 9.67 9 9.67 8.67 9.67 9 10 ...
 $ test 2    : num [1:22] 10 7.78 7.22 9.44 6.11 7.78 8.33 9.44 7.22 9.44 ...
 $ test 3    : num [1:22] 7.83 8.83 7.5 8.5 5.17 6.5 7.5 7.83 5.17 8 ...
 $ test 4    : num [1:22] 7.11 8.81 8.37 7.56 8 7.33 7.56 7.33 7.11 8 ...
 $ test 5    : num [1:22] 8.5 9.5 10 9.5 8.5 8 10 0 0 10 ...
```

Figure 5. Internal structure of the test_reg dataset

reallocate teaching time, paying more attention to topics that cause difficulties and reducing the emphasis on material that is easily understood (for example, "Data Types").

5.4. Predicting success based on regression analysis

We will build a multiple linear regression model for the dependent variable (students' final test scores) based on the independent variables (students' scores on Tests 1–5), with a code fragment presented in the listing:

```
> model_reg<-lm(‘Final
+test‘~‘test1‘+‘test2‘+‘test3‘+‘test4‘+‘test5‘,
+ data=test_reg)
```

The coefficients of the obtained multiple regression model can be found in the "estimate" column; their standard errors (std. error), which are small values, indicate the accuracy of the estimate of the influence of independent variables on the dependent variable. The p-value shows the significance of each model coefficient. In our case, all p-values are below 0.05, meaning we consider all the model coefficients to be statistically significant, as shown in the listing:

```
> tidy(model_reg)
term estimate std.error statistic p.value
1 (Intercept) -20.6 4.37 -4.71 0.000238
2 test1 1.98 0.441 4.48 0.000376
3 test2 0.726 0.212 3.42 0.00351
4 test3 -1.00 0.203 -4.92 0.000153
5 test4 1.32 0.312 4.24 0.000630
6 test5 0.125 0.0564 2.21 0.0419
```

Since the coefficient of determination (R-Squared) is 0.8, which is very close to the maximum value of 1, we can conclude that the model describes the presented data very well. It should be emphasized that success in learning language C is linearly dependent on the sequential mastery of each topic. This justifies the importance of continuous monitoring, since gaps in some tests mathematically determine a low score on the final test.

To confirm the statistical significance and validity of the obtained model, its assumptions were diagnosed. VIF analysis showed no multicollinearity between independent variables (test1–test5), which allows for a correct interpretation of the impact of each topic on the final result.

```
> vif_values <- vif(model_reg)
> print(vif_values)
‘test 1‘ ‘test 2‘ ‘test 3‘ ‘test 4‘ ‘test 5‘
1.110318 1.950102 2.460175 1.733234 1.221699
```

The Breusch-Pagan test was used to check the homoscedasticity condition of the residuals of the constructed model. The test results (BP = 2.5884, df = 5, p-value = 0.7631) indicate the absence of statistically significant heteroscedasticity. Thus, the assumption of constant variance of the residuals is satisfied, which allows us to consider the obtained estimates of the model parameters to be effective.

```
> bptest(model_reg)
 studentized Breusch-Pagan test
data:  model_reg
BP = 2.5884, df = 5, p-value = 0.7631
```

Thus, the results of the model diagnostics confirmed its statistical validity, since the absence of multicollinearity of factors and the proven homoscedasticity of residuals guarantee the effectiveness of parameter estimates and the correct interpretation of the impact of each variable. For students, these results guarantee that the model works fairly and transparently: the absence of statistical anomalies allows them to accurately determine the real contribution of each topic to the final score and make a reliable prediction of success without the risk of random errors or distortions.

To validate the multiple linear regression model, we will predict the final test score for the first student based on their previous test results. The prediction resulted in a score of 7.8, as shown in the listing. However, the student achieved a score of 8.2 on the final test. A small discrepancy is observed between the actual and predicted values, indicating that the model demonstrates a fairly high level of accuracy.

Listing of the prediction for the final test score of one student:

```
> t<-data.frame('test1' = 9.33,'test2' = 10,'test3' = 7.83,
'test4' = 7.11,'test5' = 8.5)
> t
test1 test2 test3 test4 test5
1 9.33 10 7.83 7.11 8.5
> result<-predict(model_reg, t)
> print(result)
1
7.76597
```

We will predict the final test scores for all students in the KH-23-1 group, and the results are presented in the listing:

```
> predicted_values <- model_reg$fitted.values
> predicted_values

1      2      3      4      5      6      7      8      9
7.7659 8.1958 7.9288 8.0801 8.1256 8.3845 6.3607 7.262 6.6964
10     11     12     13     14     15     16     17     18
9.8758 8.3521 9.3874 7.1266 6.1087 8.9479 5.3673 7.5023 8.026350
19     20     21     22
4.4326 9.2084 7.2917 9.5727
```

One of the key stages in building any machine learning model, including multiple linear regression, is evaluating the accuracy of the obtained predictions. Therefore, in addition to R-Squared, which provides this capability, we will also calculate the MSE (Mean Squared Error). This value shows the average deviation between the actual and predicted values. The MSE is 0.52, indicating a fairly high accuracy of the constructed multiple regression model, which can be used for future predictions. The results are presented in the listing:

```
> MSE <- mean((test_reg$‘ ‘ - predicted_values)^2)
> print(MSE)
[1]  0.5242242
```

Based on the MSE indicator, we can calculate the RMSE, which is 0.724. In the context of the study, this means that the average error of the predicted final score of a student is about 0.72 points on a ten-point scale, which

confirms the high accuracy and reliability of the developed predictive model in the context of the educational process. The small error value indicates that the use of mathematical methods minimizes the influence of subjective factors in assessing knowledge. Thus, the obtained indicators confirm that the proposed approach not only meets statistical standards but also has real pedagogical value for improving student performance.

The use of a multiple linear regression model will allow predicting the results of the final test of students based on interim scores. For the purpose of early detection of "risk groups" (students whose predicted score is unsatisfactory), it will enable the teacher to take preventive measures before the student fails the final test. Such monitoring will help students critically evaluate their own achievements and adjust their cognitive activity in a timely manner.

Thus, the mathematical methods used in the study minimize the influence of the teacher's subjective factor. The use of data on the distribution of scores (QQ charts, standard deviation) allows for the validation of the quality of the tests themselves and the fairness of the learning outcomes. This creates a more transparent educational environment where assessment is based on real competencies rather than intuitive judgments, which are often inaccurate. Statistical indicators are the foundation for making informed pedagogical decisions. They allow the learning process to be transformed from passive observation to active management of student performance, ensuring compliance with modern standards of higher education.

6. Conclusions

The analysis of test results can serve as a valuable tool for scientific research in the field of pedagogy, as it contributes to the development of new teaching and assessment methods that can be integrated into educational practice. This study aimed to develop an effective system for assessing the knowledge and skills of first-year students in the "Programming" course based on data analysis obtained from test results.

The study was comprehensive and based on the application of statistical, visual, cluster, and regression analysis, which provided a detailed picture of test performance, identified student groups with varying levels of knowledge, and revealed factors influencing the results. The use of the R programming language ensured efficiency and flexibility in the analysis. Key indicators were determined for each test, for a specific group, and for an individual student. Results across groups show heterogeneity. Students demonstrated different levels of understanding of the material. This was particularly evident in Test 2 ("Branching and Loop Operators"), highlighting the need for greater focus on this topic. The complexity of the material and the accuracy of test prepa-

ration significantly impact performance. A multiple linear regression model was also developed to predict learning outcomes, providing a reasonably accurate forecast. However, it does not account for all possible factors that may influence results.

The findings confirm the effectiveness of the proposed approach to monitoring students' academic performance. The data obtained can be used to develop recommendations for optimizing curricula and teaching methods, allowing for better consideration of individual student needs, thus making learning more personalized and effective.

REFERENCES

Alyahyan, E. & Düştegör, D. (2020). Predicting academic success in higher education: literature review and best practices. *International Journal of Educational Technology in Higher Education*, *17*(3), doi: 10.1186/s41239-020-0177-7.

Barkovska, O., Liapin, Y., Muzyka, T., Ryndyk, I. & Botnar, P. (2024). Gaze Direction Monitoring Model in Computer System for Academic Performance Assessment. *Information Technologies and Learning Tools*, *99*(1), 63 -– 75, doi: 10.33407/itlt.v99i1.5503.

Bergbauer, A. B., Hanushek, E. A. & Woessmann, L. (2024). Testing. *Journal of Human Resources*, 59(2), 349 – 388. doi: 10.3368/jhr.0520-10886R1.

Bilyakovska, O. (2022). Test as an Effective Means of Assessing the Quality of Students' Knowledge. *Academic Notes Series Pedagogical Science*, 204, 16 – 20. doi: 10.36550/2415-7988-2022-1-204-16-20.

Donets, I., Yeroshenko, H., Vatsenko, A., Shevchenko, K. & Riabushko, O. (2023). Testing as a Means of Control of Knowledge of Medical Students. *Pedagogical Sciences*, *81*, 54 – 59. doi: 0.33989/2524-2474.2023.81.289371.

Hill, A. J. & LoPalo, M. (2024) The effects of online vs in-class testing in moderate-stakes college environments. *Economics of Education Review*, 98, doi: 10.1016/j.econedurev.2023.102505.

Hooda, M., Rana, C., Dahiya, O., Rizwan A. & Md, S. (2022). Hossain Artificial Intelligence for Assessment and Feedback to Enhance Student Success in Higher Education. *Mathematical Problems in Engineering*, 5215722, 1 – 19, doi: 10.1155/2022/5215722.

Marchuk, G. V., Levkivskyi, V. L., Marchuk, D. K. & Liubchenko, D. V. (2023). A System of Rewards and Motivation for Students using Virtual Currency. *Information Technologies and Learning Tools*, *96*(4), 169 – 184. doi: 10.33407/itlt.v96i4.5285.

Marchuk, G., Levkivskyi, V., Suhoniak, V. & Panarina, I. (2024). Application of descriptive statistics for the analysis of student test results. *Technical Engineering*, *1*(93), 185 – 193. doi: 10.26642/ten-2024-1(93)-185-193.

Niessen, A. S. M., Meijer, R. R. & Tendeiro, J. N. (2016). Predicting Performance in Higher Education Using Proximal Predictors. *PLoS ONE 11*(4): e0153663, doi: 10.1371/journal.pone.0153663.

Ouyang, F., Zheng, L. & Jiao, P. (2022). Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020. *Education and Information Technologies* 27, 7893 – 7925, doi: 10.1007/s10639-022-10925-9.

Pantsyr, Y. I. & Semenyshena, R. V. (2024). Testing and Assessing Students' Learning Achievements in Teaching General Technical Disciplines. *Professional and Applied Didactics*, (1), 13 – 18. doi: 10.37406/2521-6449/2024-1-2.

Pol, J. & Oudman, S. (2024). Teachers' judgment accuracy of students' monitoring skills: a conceptual and methodological framework and explorative study. *Metacognition Learning*, (19), 65 – 101. doi: 10.1007/s11409-023-09349-8.

Rodríguez-Villalobos, M., Fernandez-Garza, J. & Heredia-Escorza, Y. (2023). Monitoring methods and student performance in distance education exams. *International Journal of Information and Learning Technology*, *40*(2), 164 – 176, doi: 10.1108/IJILT-04-2022-0085.

✉ **Olha V. Korotun**
ORCID iD: 0000-0003-2240-7891
Faculty of Information and Computer Technologies
Zhytomyr Polytechnic State University
103, Chudnivska St.
10005 Zhytomyr, Ukraine
E-mail: korotun-o@ztu.edu.ua

✉ **Galyna V. Marchuk**
ORCID iD: 0000-0003-2954-1057
Faculty of Information and Computer Technologies
Zhytomyr Polytechnic State University
103, Chudnivska St.
10005 Zhytomyr, Ukraine
E-mail: pzs_mgv@ztu.edu.ua

✉ **Vitalii L. Levkivskyi**
ORCID iD: 0000-0002-1643-0895
Faculty of Information and Computer Technologies
Zhytomyr Polytechnic State University
103, Chudnivska St.
10005 Zhytomyr, Ukraine
E-mail: `levkivskyy@ztu.edu.ua`

✉ **Oleksandra M. Svintsytska**
ORCID iD: 0000-0002-2613-2437
Faculty of Information and Computer Technologies
Zhytomyr Polytechnic State University
103, Chudnivska St.
10005 Zhytomyr, Ukraine
E-mail: `up_som@ukr.net`