The International Journal of E-Learning and Educational Technologies in the Digital Media (IJEETDM) 2(2): 59-72
The Society of Digital Information and Wireless Communications (SDIWC), Apr - 2016 (ISSN: 2410-0439)

Apprentices Identifying Groups with Difficulties in Programming Education Using Data Mining

Valter S. M. Neto, Rodrigo M. Feitosa, Dejailson N. Pinheiro, Milson L. Lima, Sofiane Labidi
Post-graduation Program in Electrical Engineering
Federal University of Maranhão
São Luís, Brazil 65080–805
valternetosnt@gmail.com, feitosamiranda@gmail.com, dejailson.pinheiro@gmail.com, milsonlima@hotmail.com, soflabidi@gmail.com

ABSTRACT

In the literature, one can find various research whose focus are the difficulties faced by students during the teaching and learning of programming. Among the proposals made to improve this situation called for the application of differentiated teaching, personalized since it is considered that the classrooms are formed by heterogeneous students with different ways of learning and who have needs and learning preferences specific. However, the customization of teaching in classroom mode is complicated to be made by the teacher. But the personalized attention to homogeneous groups of learners is a possibility to be considered. From this perspective, this article aims to describe an experience with the use of techniques of data mining along with a taxonomy of educational objectives, Bloom's Taxonomy, to identify similar groups of learners with learning difficulties in programming teaching with data obtained through assessments. With this, we hope to contribute to the construction of appropriate teaching strategies to student groups with the purpose of improving the learning process on the part of these students.

KEYWORDS

Programming Education; Data Mining; Groups of Students; Learning difficulty; Bloom's Taxonomy

1 INTRODUCTION

Learning to program is something essential in the formation of a Computer Science professional, especially for software developers. So know and know how to apply programming concepts is part of the programmer literacy. Evidence of this is the presence of disciplines, especially in the early stages, with the focus on programming, composing the curriculum of the area courses and other related. However, the programming education is something delicate, since it requires, on the part of students, attention and logical reasoning determined, leading those who do not have such facilities or not carrying to acquire them, the impending failure and in extreme cases, evasion.

It is observed that the difficulty is aggravated during the learning process since the content in the introductory lessons are sequential, dependent on each other and increasing difficulty.

In recent years, because of high rates of evasion and failure, the process of teaching and learning programming has generated a growing concern among researchers [1]. Developed studies are motivated above all by the importance of the acquisition of programming skills in shaping the field of computing professional.

Amongst the problems identified in the research highlights the development of own logical reasoning; the idea of programming as an extremely difficult hurdle to be overcome [2]; the traditional way of teaching [3] and the different pace of learning of each student [4]. About this is important to stress that the administration of discipline, in most cases, it is not conducted in the rate of uptake of each student.

However, in the classroom environment, usually consisting of large groups of learners with skills and heterogeneous knowledge, this form of teaching, personalizing teaching, is a difficult and impractical task by the teacher, even with the decrease in the number of students per class. Thus, the same lesson is given to all students, learning becoming liable to failures.

Among the proposals to change this reality in teaching programming is the use of data mining techniques that when applied to a data set, can

generate useful information for making educational decisions in order to maximize student learning. In this sense, this paper proposes to identify homogeneous groups of learners who have similar learning gaps that allows the teacher to customize the classroom teaching.

This work presents the results of applying a clustering technique on data collected from learning assessments by technical school students in a programming course in an attempt to group students with similar learning difficulties. However, it stands out in this work by creating assessments for each subject, with cognitive issues sorted levels of Bloom's taxonomy [5], [6] and contextualized, mostly with everyday matters through problem situations. The paper is organized as follows: Section 2 presents the taxonomy of educational objectives of Bloom, applied in the preparation of assessments. Section 3 presents concepts of data mining, its performance in the educational context and describes the clustering technique used in this study. Section 4 cites relevant works that have applied data mining techniques within the context of the study. Section 5 describes the conducted case study. In Section 6 the results are presented. Finally, in Section 7, are some considerations about this work.

2 BLOOM'S TAXONOMY

The Bloom's Taxonomy was proposed by Bloom et al. [5] in order to assist in the planning, organization and control of the learning goals.

This is presented in three domains: affective, psychomotor and cognitive. However, the cognitive domain, related to learn, mastering knowledge, acquire new information for intellectual development, is the best known and used.

Thus, the cognitive domain, the objectives were grouped in six different categories (Knowledge, Comprehension, Application, Analysis, Synthesis and Evaluation) and are structured in increasing levels of complexity: from the simplest to the most complex.

Thus, the Taxonomy in the educational context becomes relevant, since it aims to create opportunities to use different strategies for the development of students in knowledge acquisition and evaluation process; also provides educators plan in a structured way, so that the students acquire skills from simple skills (such everyday facts) to then ascend to the more complex (the example of scientific knowledge).

2.1 Revised Bloom's Taxonomy

With the great transformations in the educational context, such as: the use of new technologies, new educational methods, new concepts and theories, among others, comes the need for adjustments to the initial research of Bloom and his collaborators.

In 2001, Anderson and Krathwohl [6] form a commission that, in order to maintain the same practicality proposed by Bloom [5] and seek a balance between what already existed with the current positions, published a study on the retrospective use of taxonomy.

According Krathwohl [7], it was observed that generally the objectives state what students are expected to learn, however, is not explicit, consistent, what they are able to accomplish with the knowledge acquired.

Thus, with the knowledge dimension (content) and more clearly differentiated cognitive process, the new structure proposed in the Revised Bloom's Taxonomy, created a new model of the Cognitive Domain consists of six categories: Remember, Understand, Apply, Analyze, Evaluate and Create. Later sets up each of the categories of taxonomy.

2.2 Interpretation of the Taxonomy in Programming

In his studies, Whalley et al. [8] and Thompson et al. [9], report their efforts to categorize the issues of programming assessment tools according to Bloom's taxonomy. Table 1 summarizes the main interpretations of each of Bloom's taxonomy of categories included in the programming, the result of this research and that formed the basis for this study.

Table 1. Bloom's taxonomy programming

Category	Interpret in programming
Remember	**retrieving relevant Knowledge from long-term memory** - identifying a particular construct in a piece of code; - recognizing the implementation of a subject area concept.
Understand	**constructing meaning from instructional messages, including oral, written, and graphical communications.** - translating an algorithm from one form of representation to another form; - explaining a concept or an algorithm or design pattern
Apply	**carrying out or using a procedure in a given situation.** - that the process and algorithm or design pattern is known to the learner and both are applied to a problem that is familiar, but that has not been solved previously in the same context or with the same data or with the same tools;
Analyze	**breaking material into its constituent parts and determining how the parts relate to one another and to an overall structure or purpose.** - breaking a programming task into its component parts (classes, components, etc.); - organizing component parts to achieve an overall objective;
Evaluate	**making judgments based on criteria and standards.** - determining whether a piece of code satisfies the requirements through defining an appropriate testing strategy;
Create	**putting elements together to form a coherent or functional whole; reorganizing elements into a new pattern or structure.** - coming up with a new alternative algorithm or hypothesizing that a new combination of algorithms will solve a problem;

3 DATA MINING

The term Data mining defines the automated process of capture and analysis of large data sets to extract a meaning, being used both to describe characteristics of the past as to predict trends for the future [10].
According to Fayyad et al. [11] the process involves the application of specific algorithms that extract patterns from the data. Moreover, it is one step in a larger process known as Knowledge Discovery in Databases, KDD.
Thus, with regard to the processing of knowledge, their role is to apply algorithms on the data and using abstraction, generate knowledge models through the data exploration.

3.1 Tasks and Mining Techniques

According Pimentel and Omar [12], data mining tasks are defined as certain classes of problems according to the type of knowledge to be mined and the desired goals for the solution. Since the choice of mining technique and algorithm to be used depends on the task to be executed.
So, considering the objective of the study, describes the following task grouping or clustering highlighting the partitioning technique and k-means algorithm based on [13], [14].

3.1.1 Clustering

Clustering has proposed to identify and approach the similar records. Cluster is defined as a collection of similar records with each other but different from other records in other clusters.
So, considering that the purpose of this study is to form homogeneous clusters of learners, the clustering task has been chosen.
Among the existing techniques, the partitioning implement this type of task being the *K-means* algorithm the best known. The algorithm divides the data set provided in clusters, requiring initially set the number of clusters to be created for him. This number is set to *K*, the *K-means* behalf reason.

3.2 Educational Data Mining

Data mining techniques can be applied to a variety of decision-making contexts such as telephony, marketing, finance, health and education, the focus of the work.
Siemens and Baker [15] states that given the large amount of data on students, educational institutions show is traditionally inefficient in the use of these data, in most cases, making analysis with serious delays, delaying actions and preventing possible interventions.
In this context, it creates an environment in which new approaches are needed to discover, understand and properly apply the valuable information that exist within these data.

In his research, Baker and Carvalho [16] reported that the mining educational data (EDM) starts more significantly in 2005 in Pittsburgh, USA, with the first *Workshop on Educational Data Mining*, as part of the 20th *National Conference on Artificial Intelligence* (AAAI 2005), with sequels in 2006 and 2007. In 2008 launches in Montreal, Canada, the first conference in EDM: *First International conference on Educational Data Mining*, which turned out to settle down and earn an annual basis. The following year sees the launch of the first volume of the specialized journal, JEDM (*Journal of Educational Data Mining*) and in 2011 constituted a scientific society for EDM (*International Educational Data Mining Society*).

According to Baker and Carvalho [16], the educational data mining is an area of research that has focused primarily on the development of methods to explore data sets collected in educational settings.

Thus, it is possible to obtain information that helped understand more effectively the various aspects of the learning processes as well as improve the environment and methods of this process as development of instructional materials, monitoring and forecasting, among others.

4 RELATED WORK

Between the significant research linking cognitive levels of Bloom's revised taxonomy and programming education, as [8], [17] which discuss how each of the categories of taxonomy can be interpreted and used in evaluation program aiming to bring help in activities related to educational practices.

Another relevant study [18], as he sought to build the Three-phase method for teaching-learning (MTEA) applied in educational programming based on the taxonomy of Bloom [5], according to the cognitive and affective domain and applied by programming technique in doubles.

In [12] and [19] presented a model for the application of data mining techniques, using standard extraction algorithms, in order to discover knowledge about a learner or a group, in data collected through performance evaluations.

Therefore, it is observed that, in this universe, you can still get a lot of information that will enable better decision making with regard to educational programming.

5 CASE STUDY

In order, to obtain the data necessary to understand more precisely on the assimilation of knowledge and learning difficulties, by grouping students that have similarities in these aspects, using data mining techniques applied lists of exercises questions created and / or adapted according to the levels of the revised Bloom's Taxonomy [6], [8], [17] Technical course in Computer classes an educational institution.

5.1 The Research Environment

The study was conducted in a private educational institution, with 40 students aged between 16 and 17, who joined on 2014 at the Technical course in Computer Science with practical emphasis in programming, analysis and development of systems for the discipline of Logic programming.

By process of selection of the institution, the group of students was divided into two classes in different shifts. Thus, 23 students in the morning shift and 17 in the afternoon shift.

The environmental choice made by the history of the institution with a high rate of low-income students in programming courses and a significant number of failures and evasions in the course of Computer Technician.

Moreover, classes are composed of students who did not have any contact with the computational logic or with a programming language. Thus, in the initial phase of the course, there are the greatest difficulties for students, it is the moment in which these get to know the concepts related to development programs based on algorithms.

In this context, it began the data collection process for the construction of this work as described below.

5.2 Data Collection

Data collection for the development of this work took place through exercise lists applied to students during throughout the program learning process.

For the construction of lists and resolution of questions they put forward, we used the algorithm concepts and C language (Introduction, Conditional, Loops, Arrays and Subroutines). These mostly were contextualized with everyday matters through problem situations, characterized as real or hypothetical situations of theoretical and / or practical and had as reference the cognitive levels of the Taxonomy revised Bloom (Remember, Understand, Apply , Analyze, Evaluate and Create), described in Table 1.

The issues at all levels except at the level Create, were like multiple choice with options A, B, C and D. In item classified as Create, students built programs such as resolution of the proposed problem.

Thus, we used data collected from 12 assessments sessions with 40 students answering, 104 programming problems involving 5 different concepts and a total of 4160 instances.

To compose the datasets of this work, adapted to the attributes proposed by França and Amaral [19]. Table 2 shows these attributes, with its description, type and range of values they can assume within the proposed framework.

Table 2. Collected attributes in the lists of exercises

Attributes	Description	Datatype	Domain
IdStudent	Code that identifies the participant student of evaluation sessions	Nominal	[**AM_1..AM_n**] / [**AV_1..AV_n**]
IdSession	Code identifying the evaluation session number	Nominal	[**S_1..S_n**]
IdQuestion	Code identifying the assessment of the question number	Nominal	[**Q_1..Q_n**]
CognLevel	Cognitive level of the item evaluated according to Bloom's Taxonomy	Nominal	**REM** – Remember **UND** – Understand **APP** – Apply **ANA** – Analyze **EVA** – Evaluate **CRE** – Create
ResultQuest	Label of multiple-choice questions	Nominal	**COR**: Correct **PAR1**: Partially Right **PAR2**: Partially Right **PAR3**: Partially Right **INCOR**: Incorrect
AnswerQuest	Option selected by the learner in multiple-choice questions	Nominal	**A** **B** **C** **D** **X**
ConceptAss	Learner's level of performance in a specific assessment	Nominal	**A**: 8≥note≤10 **B**: 6≥note<8 **C**: 4≥note<6 **D**: 2≥note<4 **E**: 0≥note<2

In the application of mining techniques on the data collected was used WEKA tool - Waikato Environment for Knowledge Analysis, developed by the University of Waikato in New Zealand [14].

6 RESULTS OBTAINED

After collection, the data were preprocessed (removal of any inconsistencies, incompleteness and problems with data types) and transformed to a more appropriate way for mining.

Thus, it created 10 datasets (5 for each group) who underwent WEKA tool to generate groups of students to each programming concept by clustering algorithm *k-means*.

During testing of two groups were made up to 6 clusters. However, it was found that the group with 3 resulted in more consistent cluster centroids for that context.

Tables 3 and 4 show the results provided by WEKA tool from the data obtained in the sections relating to the content Introduction to Programming (variables, variable types) in morning classes (class 1) and evening (class 2), respectively.

Table 3. Clustering in class 1 (Introduction)

	Cluster#			
Attribute	Full Data (368)	0 (116)	1 (179)	2 (73)
IdStudent	AM1	AM23	AM1	AM18
IdSession	S1	S1	S1	S7
IdQuestion	Q1	Q1	Q4	Q2
CognLevel	REM	REM	UND	REM
ResultQuest	D	C	D	A
AnswerQuest	COR	INCOR	COR	INCOR
ConceptAss	C	D	C	A

Looking at Table 3 above and other results provided by the tool was able to do some interpretations:

a) Cluster 0

In this cluster, with 116 occurrences, the students (22%) with the following characteristics were grouped:

- Concept "D" (down) in the assessments, particularly in S1, they worked up the initial subjects using algorithm.
- The label "INCOR" on the issues, especially at the Remember level of Bloom's taxonomy. The question that level required to the students remember the order of priority of the operators NOT, AND and OR in logical expressions. However, despite various exercises as examples, most did not remember.
- Analyzing the alternative chosen, there were computational logic problems, the use of structures and mathematical operations.
- At Create level, if identified problems in building solutions that solve the problem situation proposed. It was observed that the codes possessed a logical sequence, however, incorrect calculations and error contained in the use of the structures.

b) Cluster 1

In this cluster, with 179 occurrences, the students (74%) which showed the following characteristics were grouped:

- Concept "C" (median) in the assessments, particularly in S1, when it worked that the initial topics using algorithms
- The label "COR" on the issues, highlighting the Understand level.
- Despite the label above the Create level, students had difficulties in building coherent solutions that solve the proposed problem.

c) Cluster 2

In this cluster, with 73 occurrences, the students (4%) with the following characteristics were grouped:

- Concept "A" (excellent) in the ratings, especially in S7, they worked up the initial topics using the C programming language
- Despite the above concept, the students presented the label "INCOR" on the issues, especially at the Remember level. The question that level required to students who remember the exercises performed in class involving NOT operators AND and OR in logical expressions. However, few remembered.
- At Create level, it was observed that, despite the problems being contextualized and require the basics and knowledge of mathematical operations, some students did not respond and others, created incorrect solutions.

Thus, it is observed that the class 1, with respect to the introductory threads programming had a median income. However, it presented some difficulties in remembering concepts and examples seen in the classroom (Remember level), and it presented problems to create solutions to the problems posed, actually understandable because they are at the beginning of the course.

Table 4. Clustering in class 2 (Introduction)

	Cluster#			
Attribute	Full Data (272)	0 (128)	1 (68)	2 (76)
IdStudent	AV1	AV14	AV7	AV16
IdSession	S1	S1	S1	S7
IdQuestion	Q1	Q2	Q6	Q4
CognLevel	REM	REM	ANA	UND
ResultQuest	D	C	D	A
AnswerQuest	INCOR	INCOR	COR	COR
ConceptAss	C	C	D	A

Looking at Table 4 above and other results provided by the tool was able to do some interpretations:

a) Cluster 0

In this cluster, with 128 occurrences, the students (70%) which showed the following features were grouped:

- Concept "C" (median) in the assessments, particularly in S1, they worked up the initial subjects using algorithm.
- The label "INCOR" on the issues, especially at the Remember level of Bloom's taxonomy. The question that level required to students who remember the particulars related to the declaration of variables. However, despite several examples carried out in the room, most did not remember.
- Analyzing the alternative chosen, there were computational logic problems, the use of structures and mathematical operations.
- At Create level, there were difficulties in creating solutions consistent with the proposed issue and the use of the concepts presented.

b) Cluster 1

In this cluster, with 68 occurrences, the students (6%) which showed the following features were grouped:

- They reached the concept "D" (bad) in the ratings, particularly in S1, they worked up the initial subjects using algorithm.
- The label "COR" on the issues, highlighting the level Analyze.
- At Create level, students had difficulties in creating solutions that solve the problem even before a contextualized issue and required only the basics of programming and mathematical operations (percentage).

c) Cluster 2

In this cluster, with 76 occurrences, the students (24%) who did not present significant difficulties that content were grouped. Thus, the afternoon class had a median income. However, presented some difficulties in construction programs (Create level), reflecting the problems presented in the Remember level.

Thus, the class 2 has a median income. However, it presents some difficulties in construction programs (level Create), reflecting the problems presented in the level Remember.

Importantly, by comparing the two groups, the morning group showed the worst results showing great difficulty in understanding the initial concepts.

Tables 5 and 6 show the results provided from the data obtained in the sections relating to the content Conditional Structures (if/else, switch case) in morning classes and afternoon, respectively

Table 5. Clustering in class 1 (Conditional)

	Cluster#			
Attribute	Full Data (437)	0 (195)	1 (162)	2 (80)
IdStudent	AM1	AM9	AM7	AM19
IdSession	S2	S2	S8	S2
IdQuestion	Q1	Q5	Q1	Q4
CognLevel	REM	APP	REM	UND
ResultQuest	A	C	A	A
AnswerQuest	COR	COR	INCOR	COR
ConceptAss	B	B	D	A

Checking Table 5 above and other results provided by WEKA it was found that:

a) Cluster 0

With 195 events were grouped in this cluster, 44% of students who showed the following features:

- In the assessments, students got the concept "B" (good), especially in S2, with written questions using the algorithm.

- In matters presented the label COR, particularly the Apply level taxonomy.
- Despite the above results, the students reported difficulties in variable types and understanding of structures, especially as the functioning of the command IF/ELSE.
- At Create level, there were problems in developing solutions that fully solve the problem situations proposed.

b) Cluster 1

With 162 events were grouped in this cluster, 43% of students who had the following characteristics:

- In the evaluations, students achieved the concept of "D" (bad), most significantly in the S8 built using the C language
- In matters reached labeled "ERR", especially at the Remember level. It (Q1), demanded to the students who remember, conceptually, the characteristics and functioning of conditional structures.
- There was the chosen responses, most option "A" great difficulties with logical operators AND and OR, understanding of structures, mathematical operations and understanding of the statement.

c) Cluster 2

80 patients were grouped in this cluster, 13% of students who achieved the concept "A" (excellent) in the assessments and presented the label COR on issues, especially at the Understand level. In addition, students reported no difficulties in the displayed content.

Thus, it appears that the class 1 had difficulty understanding the functioning of taught commands, in particular, the IF/ELSE command. In addition, there was the continuing difficulty with the logical operators (AND, OR). However, in general, the class got a good result.

Table 6. Clustering in class 2 (Conditional)

	Cluster#			
Attribute	Full Data (323)	0 (133)	1 (130)	2 (60)
IdStudent	AV1	AV12	AV10	AV4
IdSession	S2	S8	S2	S8
IdQuestion	Q1	Q7	Q9	Q2
CognLevel	UND	EVA	CRE	UND
ResultQuest	C	A	X	D
AnswerQuest	COR	INCOR	COR	INCOR
ConceptAss	A	C	A	D

Checking Table 6 above and other results provided by WEKA it was found that:

a) Cluster 0

With 133 events were grouped in this cluster, 29% of students who exhibited the following relevant points:

- Students reached the concept "C" (median) in the assessments, highlighting the S8 session, with the focus on the structures using the C language teaching.
- Presented the INCOR label on the issues, especially at the Evaluate level. In the problem (Q7), required is knowledge of the functioning of the command IF/ELSE with a condition involving the logical operators AND and OR.
- It was identified in the chosen responses, most option "A", difficulties with math operations, use of structures (especially IF/ELSE) and building solutions that solve the proposed problem (Create level).

b) Cluster 1

With 130 events were grouped in this cluster, 59% of students who have achieved the "A" concept in the evaluations and showed understanding of the subject, especially in the Create level of Bloom's taxonomy.

c) Cluster 2

In this cluster, with 60 occurrences, 12% of students which have the following relevant points were grouped:

- Students reached the concept "D" (bad) in the assessments, highlighting the S8 session, with the focus on the structures using the C language teaching

- Presented the INCOR label on the issues. Highlighted the question of Understand level (Q2), which presented a program with an empty section (IF condition) and asked that among the options offered, the student chose the one that best completed the code.
- Observed great difficulties regarding the operation of conditional structures (Understand level) and the differences between the logical operators AND / OR.

Thus, it analyzing the information, it is observed that the class 2 continued with difficulty with the use of logical operators. Moreover, the level Understanding was one in which the students found most difficult.

Comparing the two results, the morning class continued to show poor performance. In addition, the two had problems in IF/ELSE structure, the Evaluate level (class 2) and another in Remember (class 1).

Tables 7 and 8 show the results obtained from the data obtained in the sections relating to the content Loop (while, do/while, for) in the analyzed classes.

Table 7. Clustering in class 1 (Loop)

	Cluster#			
Attribute	Full Data (506)	0 (179)	1 (205)	2 (122)
IdStudent	AM1	AM21	AM17	AM5
IdSession	S3	S3	S3	S9
IdQuestion	Q1	Q2	Q4	Q5
CognLevel	APP	REM	UND	APP
ResultQuest	C	D	B	C
AnswerQuest	INCOR	COR	INCOR	COR
ConceptAss	D	D	D	B

Analyzing the data from Table 7 above and other results provided by WEKA tool was noted that:

a) Cluster 0

With 179 events were grouped in this cluster the students (30%) with the following:

- The "D" concept (bad) was obtained by the students in applied evaluations. Highlighting the S3 session, with the focus on learning the structures using the algorithm.
- Even presenting the label COR on issues, especially in Remember level of Bloom's taxonomy, they were observed by the options chosen on the issues, difficulties in differentiating the functioning of the structures used. Mainly between WHILE and DO/WHILE commands.
- At level Create, it was realized that the solutions created not completely solved the problem situations proposed. In some cases, the codes presented consistent but not bring the solution.

b) Cluster 1

With 205 events were grouped in this cluster, the students (52%) with the following:

- The "D" concept (bad) was obtained by the students in applied evaluations. Highlighting the S3 session, with the focus on learning the structures using the algorithm.
- Presented the INCOR label on the issues, especially at the level Understanding. It (Q4) was asked the student's understanding of a piece of code using the WHILE command. In addition, it identified little understanding about the functioning of loop structures (input and output conditions), confusion increments and decrement, and failures in the Create level.

c) Cluster 2

With 122 events were grouped in this cluster, the students (17%) that have reached the concept "B" (good) in the assessments and obtained the label COR on issues, especially the Apply level. In addition, students did not have difficulties in large proportions on the subject studied.

Checking the data, you can see that in general, the group presented a poor performance, especially at the Understand level of Bloom's Taxonomy that required to understand the functioning of repetitions structures. Also, if realized, problems related to increments and decrements in the present codes and problems in construction solutions using commands.

Table 8. Clustering in class 2 (Loop)

	Cluster#			
Attribute	Full Data (374)	0 (165)	1 (107)	2 (102)
IdStudent	AV1	AV6	AV7	AV12
IdSession	S3	S3	S9	S3
IdQuestion	Q1	Q7	Q4	Q9
CognLevel	APP	APP	UND	CRE
ResultQuest	C	C	B	X
AnswerQuest	INCOR	INCOR	COR	INCOR
ConceptAss	D	B	C	D

Analyzing data of Table 8 above and other results provided by WEKA tool was noted that:

a) Cluster 0

165 events were grouped in this cluster, the students (70%) with the following characteristics:

- Although the students have reached the concept "B" (good) in the assessments (S3 session), presented the INCOR label on the issues, particularly the Apply level. At issue highlighted Q7, requested that the student, after reading the statement, analyze among the options provided, the most appropriate to solve the problem using the FOR structure.
- Little understanding of the conditions of the loops and increment / decrement.

b) Cluster 1

With 107 events were grouped in this cluster, the students (12%) with the following characteristics:

- The "C" concept (median) in the assessments performed by the students, most in the S9 session using the C programming language
- After obtaining the label COR on issues, especially at the Understand level of Bloom's taxonomy, students showed difficulties concerning the operation of the loops.

c) Cluster 2

With 102 events were grouped in this cluster, the students (12%) with the following characteristics:

- Students reached the "D" concept in the evaluations, most in S3 session.
- Presented the INCOR label on the issues, particularly the Create level. Featured in session, the categorized question at this level (Q9), required to solve a problem contextualized the solution using one of repeating structures seen in the classroom.
- Little understanding of the operation and use (Create level) of repeating structures.

Thus, despite the class 2 income, it was noticed that the students had difficulty understanding and, above all, to apply the concepts of the repeat commands, especially increments and decrements associated with the operation of the loops.

Noting the results of the two groups, it is clear that the morning class had more difficulties as learning structures, especially in understanding the structures.

In Tables 9 and 10 are observed the results obtained from the data obtained in the sections relating to Arrays content in groups under study.

Table 9. Clustering in class 1 (Arrays)

	Cluster#			
Attribute	Full Data (759)	0 (357)	1 (258)	2 (144)
IdStudent	AM1	AM22	AM1	AM6
IdSession	S4	S5	S10	S11
IdQuestion	Q1	Q5	Q2	Q3
CognLevel	UND	APP	UND	UND
ResultQuest	C	C	D	A
AnswerQuest	INCOR	COR	INCOR	INCOR
ConceptAss	D	D	D	D

Exploring the data in Table 9 above and other results provided by WEKA tool was noted that:

a) Cluster 0

Were grouped in this cluster, 61% of students, a total of 357 events, with the following criteria:

- The "D" concept on ratings, especially in S5 session, in which we used algorithm to build the issues related to the subject 1D Arrays.
- Even presenting the label COR on issues, especially the Apply level, it was identified in the 2D Arrays subject, difficulty levels Analyze and Evaluate. In these questions, we used analysis of codes by a statement describing the problem, and evaluating proposals for a specific problem solutions.

- At Create level, it was realized that some of the solutions created not completely solved the problem situations proposed.

b) Cluster 1

Were grouped in this cluster, 35% of students, a total of 258 events, with the following criteria:

- Students reached the concept "D" (bad) on ratings, especially in the S10. In it approached the vector content using the C programming language.
- Most of the issues presented INCOR label, especially at the Understand level. The question (Q2) had a code in which values were assigned and the reverse of them was printed by the code that was requested. Analyzing the options chosen, it was observed that there was no understanding of the code.
- In addition, it identified little understanding for the declaration and operation of 1D Arrays.
- At Create level, it was identified that largely did not propose solutions to the problems posed.

c) Cluster 2

Were grouped in this cluster, 4% of the students, a total of 144 events, with the following criteria:

- Students reached the "D" concept (bad) on ratings, especially in the S11. In it he approached the 2D Arrays content using the C programming language
- Most of the issues presented INCOR label, especially at the Understand level. The question (Q3) had a code with a missing piece that provided as output a 3x3 identity 2D Arrays hus, the student should choose among the options that best fill the space.
- Difficulties in understanding the concepts and 2D Arrays operation, especially regarding the use of loops (rows and columns).

Analyzing the points, it is clear that the group had great difficulties in the subject 2D Arrays. Especially in issues that needed to analyze and evaluate codes that were provided. In addition, they identified problems with the use of nested loop (lines and columns).

Table 10. Clustering in class 2 (Arrays)

	Cluster#			
Attribute	Full Data (561)	0 (291)	1 (194)	2 (76)
IdStudent	AV1	AV4	AV15	AV4
IdSession	S4	S10	S11	S5
IdQuestion	Q1	Q6	Q1	Q9
CognLevel	UND	UND	UND	CRE
ResultQuest	C	D	C	X
AnswerQuest	INCOR	INCOR	COR	INCOR
ConceptAss	D	D	C	D

Exploring the data of Table 10 above and other results provided by WEKA tool was noted that:

a) Cluster 0

Were grouped in this cluster, 59% of students, a total of 291 events, with the following criteria:

- The concept of "D" (bad) in the assessments has been achieved by the students, particularly in S10.
- The INCOR label was obtained on issues, especially at the level of Bloom's Taxonomy. The questions that Understand level had codes in which values were assigned and the output was requested. But checking the options chosen, it was observed that there was no understanding of the code.
- There was little understanding of the concept, the declaration and the operation about the 1D Arrays.

b) Cluster 1

Were grouped in this cluster, 35% of students, a total of 194 events, with the following topics described:

- Although the students have achieved the label COR in questions, especially at the level of Understand taxonomy, they reached the concept "C" (median) in the assessments (S11 session).
- Difficulties in Analyze level. In it, the student was asked to analyze the proposed solutions to the problem.
- Problems in building codes consistent with problem situations proposals regarding 2D Arrays.

c) Cluster 2

Were grouped in this cluster, 35% of students, a total of 194 events, with the following topics described:

- The concept of "D" (bad) was achieved by students in the evaluations, especially in S5 session, in which they tested the vectors of knowledge using the algorithm.
- The label "INCOR" was awarded the issues. Highlighted, has Q9, the Create level, which requested the creation of a solution to a problem similar to one previously seen in the classroom.
- Major difficulties regarding the operation and use (Create level) of 2D Arrays.

The above points show that in the group that works with 1D and 2D Arrays, the group demonstrated great difficulty in understanding the functioning of the structures.

Thus, it is clear that both groups have difficulties in content addressed, especially in arrays. Moreover, the problem has become more latent Understand level reflecting directly on other levels, particularly in the Create.

Finally, Tables 11 and 12 show the results obtained from data collected in the sections relating to Subroutines content in the classes studied.

Table 11. Clustering in class 1 (Subroutines)

	Cluster#			
Attribute	Full Data	0	1	2
	(322)	(186)	(79)	(57)
IdStudent	AM1	AM11	AM8	AM14
IdSession	S6	S12	S12	S6
IdQuestion	Q1	Q5	Q7	Q3
CognLevel	EVA	EVA	CRE	APP
ResultQuest	A	A	X	D
AnswerQuest	INCOR	INCOR	INCOR	COR
ConceptAss	D	D	E	B

Ascertaining Table 11 above and other results provided by WEKA tool was identified that:

a) Cluster 0

In this cluster, with 186 occurrences, the students (74%) with the following criteria were grouped:

- Students received the concept "D" (bad) in the assessments (S12 session).
- The INCOR label was attributed to issues, particularly in Evaluate level the Taxonomy.
- Difficulties in all taxonomic levels, especially on the function and action of the subroutines in a program.

b) Cluster 1

With 79 events were grouped in this cluster, the students (13%) who expressed the following:

- Evaluations (S12 session) with the concept of "E" (bad).
- The INCOR label on the issues.
- Difficulties in understanding the concept and function of subroutines within a program, and creating solutions using subroutines (Create level).

c) Cluster 2

In this cluster, they were grouped the 13% of students, a total of 57 events, with the following characteristics:

- Despite reaching the concept "B" (good) in the ratings and reach the label COR on issues, especially the Apply level, they presented some difficulties at the Understand level of Bloom's Taxonomy.

With the above data, it is clear that the group did not obtain a significant learning about subroutines. Most students got a bad concept in the evaluations and had difficulties in understanding the function of subroutines within a program, as well as in building solutions using the concepts.

Table 12. Clustering in class 2 (Subroutines)

	Cluster#			
Attribute	Full Data	0	1	2
	(238)	(116)	(76)	(46)
IdStudent	AV1	AV4	AV2	AV13
IdSession	S6	S6	S12	S6
IdQuestion	Q1	Q3	Q1	Q2
CognLevel	EVA	EVA	EVA	UND
ResultQuest	A	A	A	B
AnswerQuest	INCOR	INCOR	COR	INCOR
ConceptAss	B	D	C	D

Looking at Table 12 above and other results supplied by WEKA tool was identified that:

a) Cluster 0

In this cluster, with 116 occurrences, the students (59%) were grouped that presented the following points:

- The concept of "D" (bad) in the assessments (S6 session).
- The issues reached labeled "INCOR" on, especially at the Evaluate level the Taxonomy.
- Difficulties in understanding and application of concepts learned about subroutines.
- At Create level, most students did not propose solutions to the problems posed.

b) Cluster 1

76 events were grouped in this cluster, the students (35%) showing the following:

- Despite reaching the label COR on issues, presented the concept "C" (median) in the assessments (S12 session).
- Difficulties in Remember level Bloom's Taxonomy and creating solutions using subroutines (Create level).

c) Cluster 2

46 events were grouped in this cluster, the students (6%) showing the following features:

- Students hit the "D" concept (bad) in the assessments (S6 session) and reached the INCOR label on the issues.
- Great difficulty levels Remember and Understand of Bloom's Taxonomy. In the first, the question required the student to remember the sub concepts and the difference between procedures and function. In the second, an incomplete portion of which called the option that correctly completed the code.

Thus, it is observed that the class 2 had some significant learning about the topic under study (stanzas). Mainly evaluate solutions using the concepts seen in the classroom.

Analyzing the information obtained from classes, it is clear that in both there was a real learning stanzas. Regardless of the context, or example used, students could not understand the usefulness of the concept of subroutines.

7 CONCLUSIONS

The results show that learning assessments can generate important data about the process of teaching and learning, especially when directed by a taxonomy of educational objectives, this work, Bloom's taxonomy.

Also, confirm that the application of clustering techniques are quite useful for the formation of homogeneous clusters of learners. Once identified, these groups allow the teacher to formulate most effective teaching strategies that it will act according to the real needs of students, especially those with learning disabilities.

In the study it was possible to identify, for example, in general, the students presented major problems in the Create level. I.e. difficulties in building solutions with the concepts presented, satisfying the problem situations proposed.

As for the learning of Conditional structures if identified problems, especially in the IF/ELSE structure. This situation is aggravated when the problems required the knowledge of logical operators AND and OR.

In addition, it was noted that in some content, especially Loops and Arrays, students had little understanding (level Understand) in the declaration and functioning of the structures. Loops on the topic, there is confusion with the conditions for entry and exit and the action of increment / decrement in the structures. In the topical arrays, 2D arrays stands out with the use of loops (rows and columns).

A fact to be noted is that, since the student finds it difficult in the early levels of the taxonomy it reflects the other levels. Loops and Arrays observed this fact.

Another fact to note is the poor performance of students in Arrays and Subroutines topics, reflecting the difficulties do not identified and it do not addressed in a timely manner. In some clusters, it was perceived difficulties at all levels of the taxonomy and on topic Subroutines, much was not able to build solutions using the concepts.

The results also brought questions such as the fact that students of class 2, which runs in the afternoon shift, with the same teacher and the same classroom, deliver better results compared

to the class 1, which works in the evening shift. Perhaps the shift factor can be analyzed within the context of this work.

The exercise lists drawn up this work, proved a valid assessment tool that will enable other teachers, suffering adjustments in some cases, can better visualize the learning of students in classes in which they operate.

However, it is expected that with the results shown, teaching strategies are built to enhance the learning of programming students. As for the problem with the mathematical operations presented at the beginning of the programming discipline, start classes with a math review with problems involving the subjects that it will be needed later.

As future work, we intend to conduct deeper analysis on the data found by analyzing other points of view and work on a system that provides a faster, more specific feedback for both teachers, and for the students

REFERENCES

[1] SILVA, I. F. A.; SILVA, I. M. M. S; SANTOS, M. S. Análise de problemas e soluções aplicadas ao ensino de disciplinas introdutórias de programação. Universidade Federal Rural de Pernambuco, Recife – PE. 03 p. 2009.

[2] JUNIOR, M. C. R. Como Ensinar Programação? Informática – Boletim Informativo, Canoas, RS, v.1, n.1, 2002.

[3] BORGES, M. Avaliação de uma metodologia alternativa para a aprendizagem de programação. In: WORKSHOP DE EDUCAÇÃO EM COMPUTAÇÃO – WEI, 8., 2000. Curitiba. Anais... Curitiba, [s.n.], 2000.

[4] RAABE, A. L. A. and SILVA, J. M. C. Um Ambiente para Atendimento as Dificuldades de Aprendizagem de Algoritmos. In Anais do XXV Congresso da Sociedade Brasileira de Computação: XIII Workshop sobre Educação em Computação. São Leopoldo: Brasil, pp. 2326-2337, 2005.

[5] BLOOM, B. S. (Ed.).Taxonomy of educational objectives, Handbook 1: Cognitive domain. New York: David McKay, 1956.

[6] ANDERSON, L. W. and KRATHWOHL, D. R. (Ed.).A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives. New York: Addison Wesley Longman, 2001.

[7] KRATHWOHL, D R. A revision of bloom's taxonomy: an overview. Theory into Practice, [s.l.], n. 41, v. 4, pp. 212-218, 2002.

[8] WHALLEY, J. L; LISTER, R.; THOMPSON, E.; CLEAR, T.; ROBBINS, P.; KUMAR, P. K. A; PRASAD, C."An Australasian Study of Reading and Comprehension Skills in Novice Programmers, using the Bloom and SOLO Taxonomies", In: VIII Australasian Computing Education Conference (ACE2006), Computer Society, pp. 243-252, 2006.

[9] THOMPSON, E.; REILLY, A. L.; WHALLEY, J.; HU, M.; ROBBINS, P."Bloom's taxonomy for CS assessment", In: X Australasian Computing Education Conference - ACE, Australian Computer Society,pp. 155-161, 2008.

[10] SFERRA, H. H. and CORREA, Â. M. C. J. Conceitos e aplicações de Data Mining .Revista Ciência & Tecnologia, pp. 19 -34, jul/dez, 2003.

[11] FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From Data Mining to Knowlegde discovery. American Association for Artificial Intelligence.1996.

[12] PIMENTEL. E. P. and OMAR, N. Descobrindo Conhecimentos em Dados de Avaliação da Aprendizagem com Técnicas de Mineração de Dados. In: XII Workshop de Informática na Escola, 2006, Campo Grande, MS. Anais do XXVI CSBC, 2006.

[13] CAMILO, C. O. and SILVA, J. C.; Mineração de Dados: conceitos, tarefas, métodos e ferramentas. Goiânia: Instituto de Informática/UFG, 2009.

[14] WITTEN, I. H. and FRANK, E. . Data Mining: Practical Machine Learning Tools and Techniques. 2nd edition,Morgan Kaufmann Publishers, San Francisco, CA, 2005.

[15] SIEMENS, G. and BAKER, R. S. J. D. Learning analytics and educational data mining: towards communication and collaboration. In: INTERNATIONAL CONFERENCE ON LEARNING ANALYTICS AND KNOWLEDGE, 2., 2012, New York, NY, USA. Proceedings... ACM, 2012. p. 252–254.

[16] BAKER, R. S. J and CARVALHO, A. M. J. B, A. Mineração de Dados Educacionais: Oportunidades para o Brasil. Revista Brasileira de Informática na Educação, [s.l.] v. 19, n.2, pp. 3-13, 2011.

[17] JESUS, E. A. and RAABE A. L. A. Interpretações da Taxonomia de Bloom no Contexto da Programação Introdutória. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 20., 2009. Ford. Anais...2009.

[18] FARIA, E.S.J. Método trifásico de ensino-aprendizagem baseado na taxionomia de objetivos educacionais de Bloom: uma aplicação no ensino de programação de computadores. 2010. 453 f. Tese (Doutorado em Engenharia Elétrica) - Faculdade de Engenharia Elétrica, Universidade Federal de Uberlândia, 2010.

[19] FRANÇA, R. S. and AMARAL, H. J. C. Mineração de Dados na Identificação de Grupos de Estudantes com Dificuldades de Aprendizagem no Ensino de Programação. RENOTE-Revista Novas Tecnologias na Educação, Porto Alegre v. 11, n.1, CINTED/UFRGS., 2013.