Data Mining to Identify Learning Groups with Difficulties in Programming Education

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

The difficulties encountered by students in the teaching and learning of programming has been the focus of much research. Among the proposals discussed to improve this situation called for the use of differentiated instruction, personalized, since it is considered that the classrooms are composed of heterogeneous students with different ways of learning and who have needs and learning preferences private. However, the customization of teaching in classroom mode is difficult 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

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.

Research indicates that problems with 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

In order to assist in the planning, organization and control of the learning goals, it was proposed by Bloom [5] a taxonomy known as Bloom's Taxonomy.

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.

2.1 Revised Bloom's Taxonomy

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, the new structure proposed in the Revised Bloom's Taxonomy, with the knowledge dimension (content) and more clearly differentiated cognitive process, 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 works, Whaley et al. [8] and Thompson et al. [9] describe 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;

ISBN: 978-1-941968-27-7 ©2016 SDIWC

3 DATA MINING

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, their role, with regard to the processing of knowledge, 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 is the task that proposes 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.

In his research, Baker and Carvalho [15] 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 [15], 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

There are significant research linking cognitive levels of Bloom's revised taxonomy and programming education, as [8], [16] 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 [17], 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 [18] 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], [16] 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 [18]. 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	[$\mathbf{AM_1..AM_n}$] / [$\mathbf{AV_1..AV_n}$]
IdSession	Code identifying the evaluation session number	Nominal	[$\mathbf{S_1..S_n}$]

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	0	1	2
	(368)	(116)	(179)	(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:

- **Cluster 0** - In this cluster, with 116 occurrences, the students (22%) with the concept "D" in the ratings were grouped. In addition, the label presented INCOR on issues, especially in Remember level of Bloom's taxonomy. It was also noted, the options chosen on the issues, computational logic problems and mathematical operations. Create the level, it identified problems in creating solutions that solve the problem situations proposed.
- **Cluster 1** - In this cluster, with 179 occurrences, the students (74%) who achieved the concept of "D" in the ratings were grouped. Moreover, although the label had COR on the issues, the Create level, showed, difficulties in creating consistent solutions that solve the problems posed.
- **Cluster 2** - In this cluster, with 73 occurrences, the students were grouped (4%) that despite having received the concept "A" in the ratings, presented the label INCOR on issues, especially at the level Remember. Furthermore, the Create level or not responded or created wrong solutions.

Table 4. Clustering in class 2 (Introduction)

Attribute	Cluster# 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:

- **Cluster 0 -** In this cluster, with 128 occurrences, the students (70%) who had similar results to the cluster 0 class 1 were grouped, except the Create level, where there were difficulties in creating solutions consistent with the proposed issue and the use of concepts submitted.
- **Cluster 1 -** In this cluster, with 68 occurrences, the students (6%) who achieved the concept of "D" in the ratings were grouped. Moreover, although the label had COR on the issues, the Create level, students had difficulties in creating solutions that solve the problem.
- **Cluster 2 -** In this cluster, with 76 occurrences, the students (24%) who did not present significant difficulties that content were grouped.

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

Table 5. Clustering in class 1 (Conditional)

Attribute	Cluster# 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:

- **Cluster 0**: In this cluster, with 195 occurrences were grouped 44% of the students that even reached the concept of "B" on reviews and presenting the label COR on issues, especially at the level taxonomy Apply, they reported difficulties in variable types and understanding of the structures. Moreover, it was observed that in Create level problems in developing solutions to completely solve the problem situations proposed.
- **Cluster 1 -** In this cluster, with 162 occurrences were grouped 43% of students who achieved the concept of "D" in the ratings. In addition, the label presented INCOR on issues, especially in Remember level, and great difficulties with logical operators (AND, OR), understanding the structures, mathematical operations and understanding of the statement.
- **Cluster 2 -** In this cluster, with 80 occurrences were grouped 13% of students who achieved the concept "A" in the reviews and had the label COR on issues, especially at the level understand. In addition, the students presented no difficulties in content.

Table 6. Clustering in class 2 (Conditional)

Attribute	Cluster# Full Data (373)	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:

- **Cluster 0 -** In this cluster, with 133 occurrences were grouped 29% of students who achieved the concept of "C" in the reviews and had the label INCOR in questions, especially at the level Evaluate. Moreover, it identified difficulties with math operations, use of structures (mostly if / else) and building solutions that solve the proposed problem (level Create).
- **Cluster 1 -** In this cluster, with 130 occurrences were grouped 59% of students who have achieved the "A" concept in the ratings and showed understanding of the subject, especially on the Create of Bloom's taxonomy level.
- **Cluster 2 -** In this cluster, with 60 occurrences, 12% were grouped the students who achieved the concept of "D" in the reviews and had the label INCOR in questions. In addition, it was observed great difficulties regarding the operation of conditional structures (level Understand) and the

differences between the logical operators AND / OR.

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	0	1	2
	(506)	(179)	(205)	(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:

- **Cluster 0**: With 179 events were grouped in this cluster the students (30%) with the concept of "D" in the ratings. Moreover, even with the label color in questions, especially in Remember level of Bloom's Taxonomy, it was observed by the options chosen in questions, difficulties in differentiating the functioning of the structures used. Create the level, it was noted that solutions created not completely solved the problem situations proposed.
- **Cluster 1 -** With 205 events were grouped in this cluster the students (52%) who achieved the concept of "D" in the reviews and had the label INCOR in questions, especially at the level understand. Moreover, it identified little understanding about the functioning of looping structures (input and output conditions), confusion increments and decrements and failures in Create level.
- **Cluster 2 -** With 122 events were grouped in this cluster the students (17%) who achieved the concept of "B" in the ratings and obtained the label COR in questions, especially the Apply level. Moreover, students did not present difficulties in large proportions on the subject studied.

Table 8. Clustering in class 2 (Loop)

	Cluster#			
Attribute	Full Data	0	1	2
	(374)	(165)	(107)	(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:

- **Cluster 0**: With 165 events were grouped in this cluster learners (70%) which, although they reach the term "B" in the evaluations presented in the label INCOR- questions, especially the Apply level. Moreover, little is identified understanding of the conditions of loops and increment / decrement.
- **Cluster 1 -** With 107 events were grouped in this cluster the students (12%) who achieved the concept of "C" in the ratings. Moreover, despite obtaining the label COR in questions, especially at the level of Bloom's taxonomy understand, students showed difficulties concerning the operation of Loops.
- **Cluster 2 -** With 102 events were grouped in this cluster the students (12%) who achieved the concept of "D" in the reviews and had the label INCOR in questions. Moreover, it identified little understanding of the operation and use (Create level) of the loop 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	0	1	2
	(759)	(357)	(258)	(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:

- **Cluster 0**: With 357 events were grouped in this cluster, the students (61%) with the concept of "D" in the ratings. Moreover, even with the label COR in questions, especially at the level of the taxonomy Apply, it was identified in the 2D Arrays subject, difficulty levels Analyze and Evaluate. Create the level, it was noted that some of the solutions created not completely solved the problem situations proposed.
- **Cluster 1 -** With 258 events were grouped in this cluster the students (35%) who achieved the concept of "D" in the reviews and had the label INCOR in questions, especially at the level Understand. Moreover, it identified little understanding as the declaration and operation of 1D Arrays. Create the level, it was identified that largely did not propose solutions to the problems posed.
- **Cluster 2 -** With 144 events were grouped in this cluster the students (4%) who achieved the concept of "D" in the ratings and obtained the label INCOR in questions, especially at the level Understand. Moreover, there was difficulty in understanding concepts and functioning of 2D Arrays, especially regarding the use of nested loops (rows and columns).

Table 10. Clustering in class 2 (Arrays)

	Cluster#			
Attribute	Full Data	0	1	2
	(561)	(291)	(194)	(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:

- **Cluster 0**: With 291 events were grouped in this cluster 59% of students who have attained the concept "D" in the ratings and obtained the label INCOR in questions, especially at the level of Understand taxonomy. Moreover, there was little understanding of the concept, the declaration and the operation about the 1D arrays.
- **Cluster 1 -** With 194 events were grouped in this cluster 35% of students who, although have reached the label COR in questions, especially at the level of understand taxonomy, they reached the concept "C" in the ratings. Moreover, it identified difficulties in the level Analyze and building codes consistent with problem situations proposals regarding 2D arrays.
- **Cluster 2 -** With 76 occurrences were grouped in this cluster 6% of students who have reached the "D" concept in the reviews and had the label INCOR in questions. In addition, it was noticed great difficulties regarding the operation and use (level Create) of 2D Arrays.

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:

- **Cluster 0**: In this cluster, with 186 occurrences, the students (74%) were grouped that have obtained the concept of "D" in the ratings and hit the label INCOR in questions, especially at the level Evaluate the taxonomy. Moreover, it identified difficulties in all taxonomic levels mainly on the role and action of the subroutines in a program.
- **Cluster 1 -** In this cluster, with 79 occurrences, the students (13%) who achieved the concept "E" in the reviews and had the label INCOR in questions were grouped. Moreover, there was difficulty in understanding the concept and function of subroutines within a program and creating solutions using subroutines (level Create).
- **Cluster 2 -** In this cluster, with 57 occurrences, the students (13%) were grouped that despite reaching the concept of "B" in the ratings and hit the label COR in questions, especially the Apply level, had some difficulties at the level Understand of Bloom's taxonomy.

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:

- **Cluster 0**: In this cluster, with 116 occurrences, the students (59%) were grouped that have obtained the concept of "D" in the ratings and hit the label INCOR in questions, especially at the level Evaluate the taxonomy. Moreover, it identified difficulties in understanding and application of concepts learned about subroutines. Create the level, most students did not propose solutions to the problems posed.
- **Cluster 1 -** In this cluster, with 76 occurrences, the students were grouped (35%) that despite reaching the label COR in questions, presented the concept "C" in the ratings. Moreover, there was difficulty in Remember level of Bloom's taxonomy and creating solutions using subroutines (level Create).
- **Cluster 2 -** In this cluster, with 46 occurrences, the students (6%) who achieved the concept of "D" in the ratings and hit the label INCOR in questions were grouped. Moreover, there was great difficulty in level Understand and Remember Bloom's taxonomy.

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.

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] 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.

[16] 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.

[17] 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.

[18] 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.