

(I) Brazilian-German Symposium on Learning Analytics Mainz, May 12th 2023

LEARNING ANALYTICS IN **INTRODUCTORY PROGRAMMING**

Elaine Harada Teixeira de Oliveira Research Group on Computing Education Institute of Computing Federal University of Amazonas (UFAM) Manaus Amazonas Rrazil



JOHANNES GUTENBERG UNIVERSITAT MAINZ









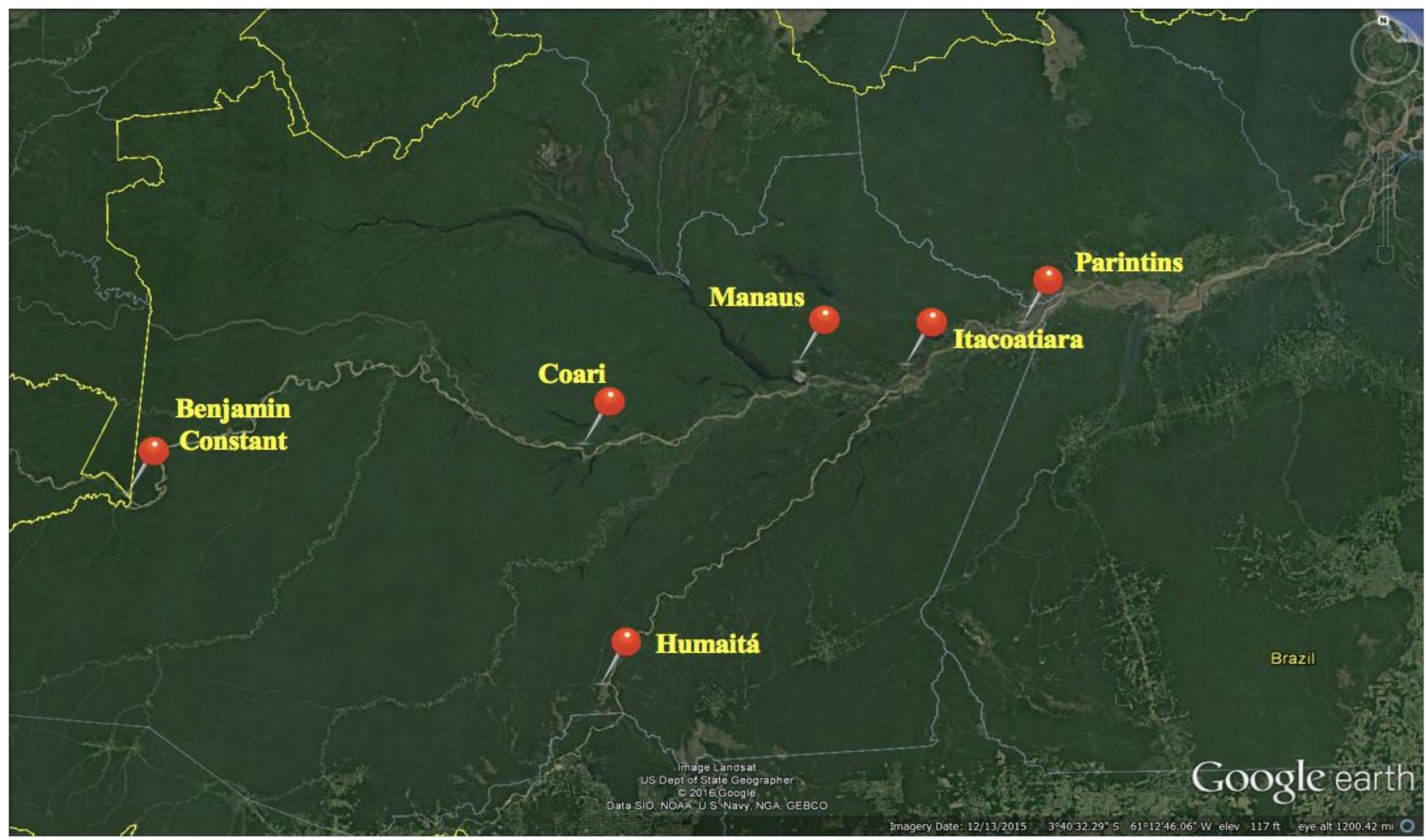


Manaus campus embedded in 600ha of native forest

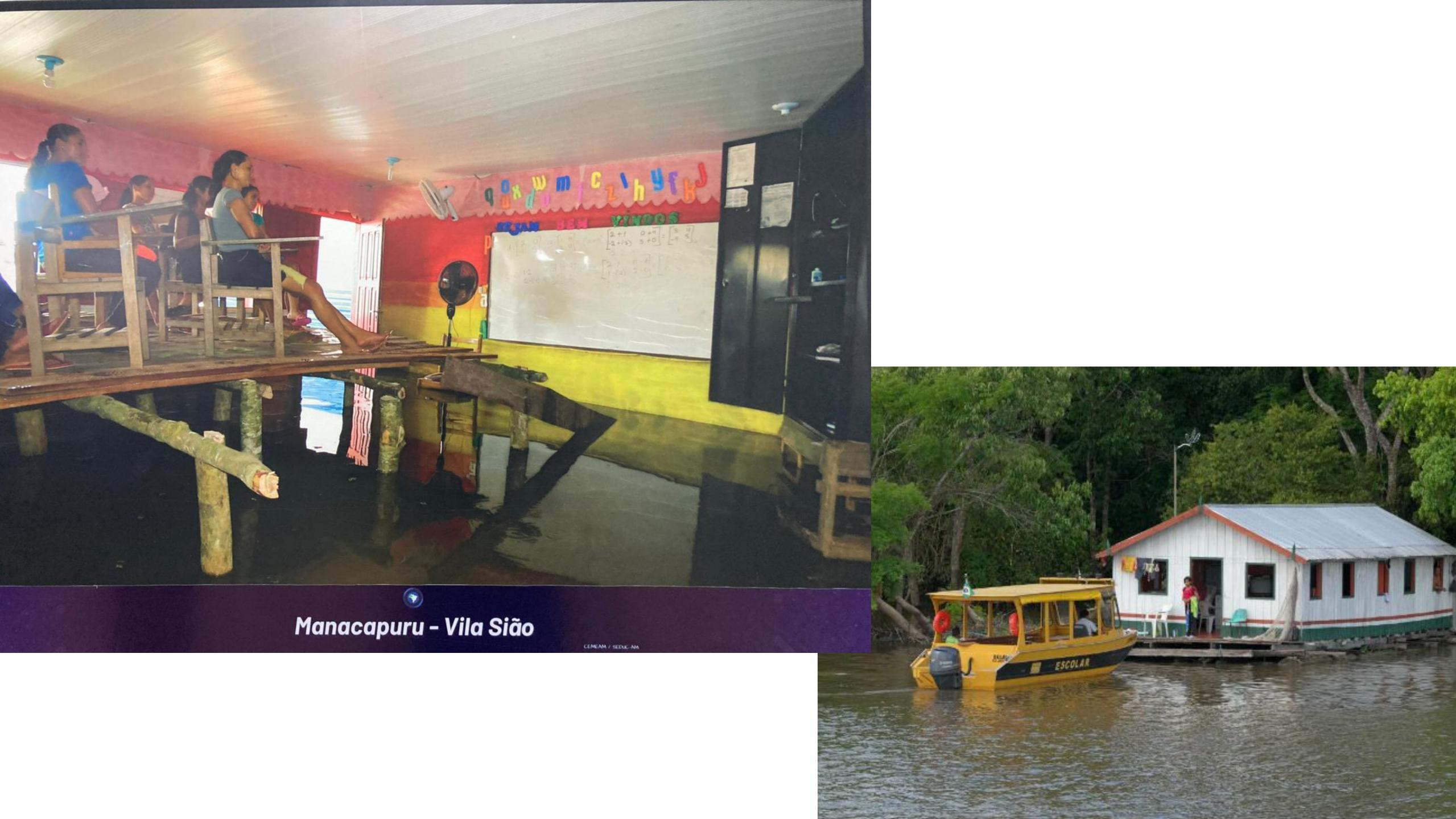












CS1 at UFAM THE FOUR ENGINES

Educational Data

01

Interaction data, user data, tests and submissions, correctness and errors

Online Tools

Adoption of online tools for remote, hybrid and face-to-face teaching

Learning Analytics

03

04

02

Predictions, descriptions, interventions, recommendations, customization, etc.

Gamification

To encourage and engage

LEARNING ANALYTICS THE FOUR ENGINES

01

In computer courses, online judges are essential tools to support teaching and tata. tests bupport teaching and tests data. tests estimation of the sector of the sector of the sector of the learning activites

Online Tools

Adoption of online tools for remote, hybrid and face-to-face teaching

Learning Analytics

03

04

02

Predictions, descriptions, interventions, recommendations, customization, etc.

Gamification

To encourage and engage



👨 Editar Exercício × @ _

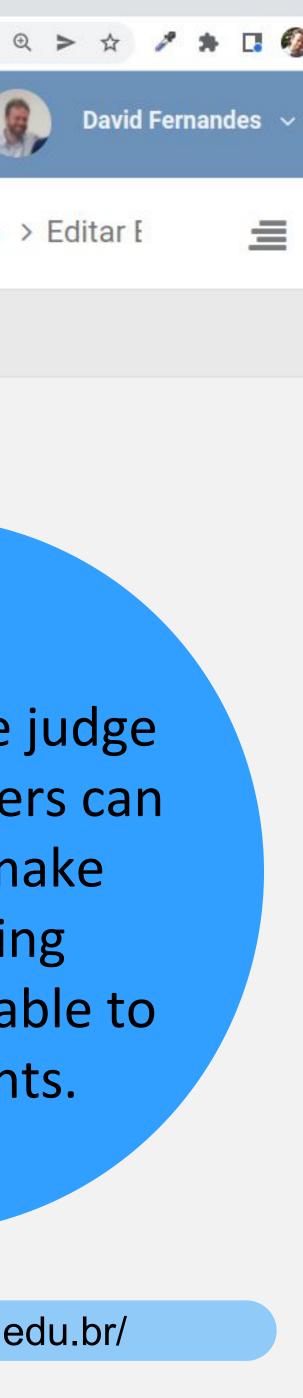
← → C
admin.codebench.icomp.ufam.edu.br/index.php?r=exercicio%2Fupdate&id=672&trabalho=3175

× +

Ð	CODEBENCH ≡	8	Busca	•	Contexto de Bus	sca ~						David
	PROFESSOR	A	Todas as Turm	as > Introdução	o à Program	nação de Co	omputad	lores – ⁻	Furma EN	/IEC1/C2	> Traba	Ihos > Lab 5 - Vetores e Strings > Editar I
P	Turmas	>										
8	Disciplinas	>	 Exercício 	Editar Exercício	Estatística	as do Exercício						
	Mensagens	>	Enunciado •									
	ADMINISTRAÇÃO		Editar - Formata	r ▼ Inserir ▼ Vis	ualizar 🔻 🛛 Fer	rramentas •						
4	Usuários	>	B I Format	os • ≣ ≣	≡ := - 1	∃ - ⊒	- B		- Σ	o 🖶		
盦	Laboratórios	>	Vetor de uns								*	
	Docker Containers	>	Escreva um programa que leia um número inteiro N, positivo, este número N será relativo a quantidade de posições de um vetor que será criado. Em seguida, crie e imprima um vetor de N posições, todas preenchidas com o valor 1 (um), do tipo inteiro.									
Þ	Minha Conta	-	р									
۲	Logout		*								lin.	design and make
Front-end Casos de teste para correção do exercício Adicionar							programming exercises available to					
	USO DO SERVIDOR		Entrada do primeiro	caso de teste •								their students.
	Memória 37%	% ■	34									
	CPU 0.1%	6	Saída esperada do p	rim <mark>eiro c</mark> aso de teste	•							
			[11111111111	11111111111	111111111	1111]					https:/	//codebench.icomp.ufam.edu.br/
				annes come								



÷.





Introduction to Computer Programming Programming Exercises

Start

,

Start	Descript	ion	File	Edit Searc	h Run Tools			
01	Write a pro	graph that chocks	Python 3 🛧 main.py					
02		gram that checks string entered by the	1	1 def is_palindrome				
03	user is pali	ndrome or not.	3	<pre>2 left_pos = 0 3 right_pos = ler</pre>				
04	Tips		4					
05	TIPS		6					
Grades		ne is a word, phrase, quence of characters	7					
		the same backward as ich as level or radar.	9 10 11		ght_pos -= 1 n True			
	Input/Output Example				put("Please ent _palindrome(str			
	Input	kayak	•	Console	Shell			
	Output True			<pre>\$ python3 main.py Please enter a string: madam True</pre>				



Home > ES01 > Homeworks

meworks 2

Materials Messages

ing):

ing) - 1

left_pos: eft_pos] == string[right_pos]:

iter a string: ")
ir))

Students code solutions for the proposed exercises in our system

https://codebench.icomp.ufam.edu.br/







CODEBENCH

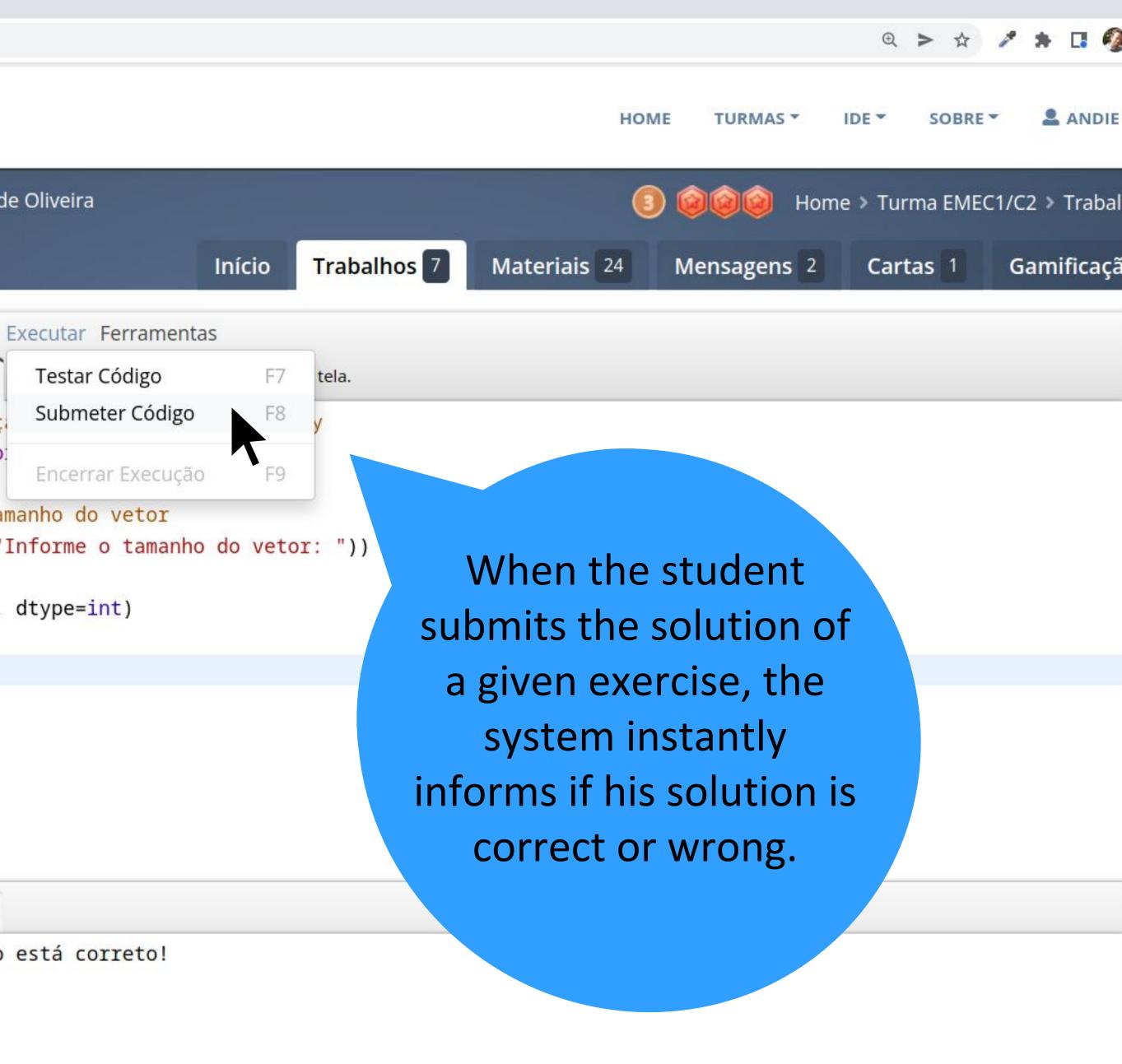
←

Introdução à Programação de Computadores — Turma EMEC1/C2 — professor(a) Elaine Harada Teixeira de Oliveira

Lab 5 - Vetores e Strings

Início	"uns"	não podem estar seguidos de ponto;	Arquivo Editar	Buscar
01	2. Use a	função ones(), do módulo numpy, bem como sua	Python 3	main.py
02 03 04 05	Exemplos	dtype=int . s de Entrada e Saída	1 # Nāo se 2 from num 3 4 # Leitur 5 N = int(e esqueç npy impo ra do ta
□ 06 □ 07 □ 08	Entrada Saída	6 [1 1 1 1 1]	6 7 vetor = 8	
□ 09 □ 10	Resultad	0	9 print(ve	etor)
 11 12 13 14 15 16 		Parabéns! Seu esforço e desempenho enchem nosso mundo de	Console	Shell
 16 17 Notas 		riquezas! Por isso, um baú com 5 moedas de ouro surgiu nesta dimensão.	Parabéns, seu	ı código

https://codebench.icomp.ufam.edu.br/index.php?r=trabalho%2Fview&id=3175#



https://codebench.icomp.ufam.edu.br/

LEARNING ANALYTICS IN ONLINE JUDGE SYSTEMS

DATA COLLECTION IN ONLINE JUDGES



 \checkmark

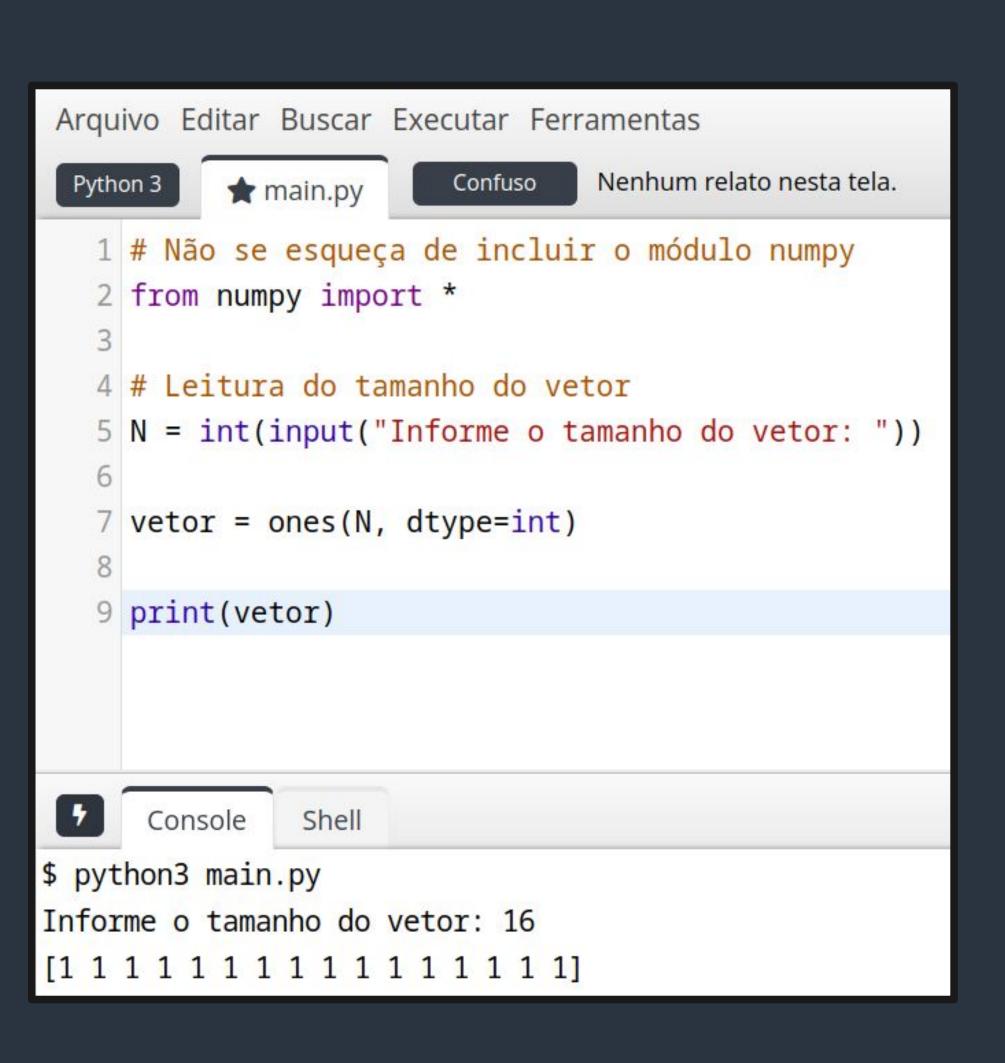
 \checkmark

Data generated by online judges differs from data generated by other types of web applications



- Compiler/interpreter error messages
- Code development process

Research opportunities on the use of \checkmark Learning Analytics in online judge systems are wide and diverse



LEARNING ANALYTICS IN ONLINE JUDGE SYSTEMS

- The code development process can be collected at different levels of granularity
- Each level is associated with a data size and collection frequency



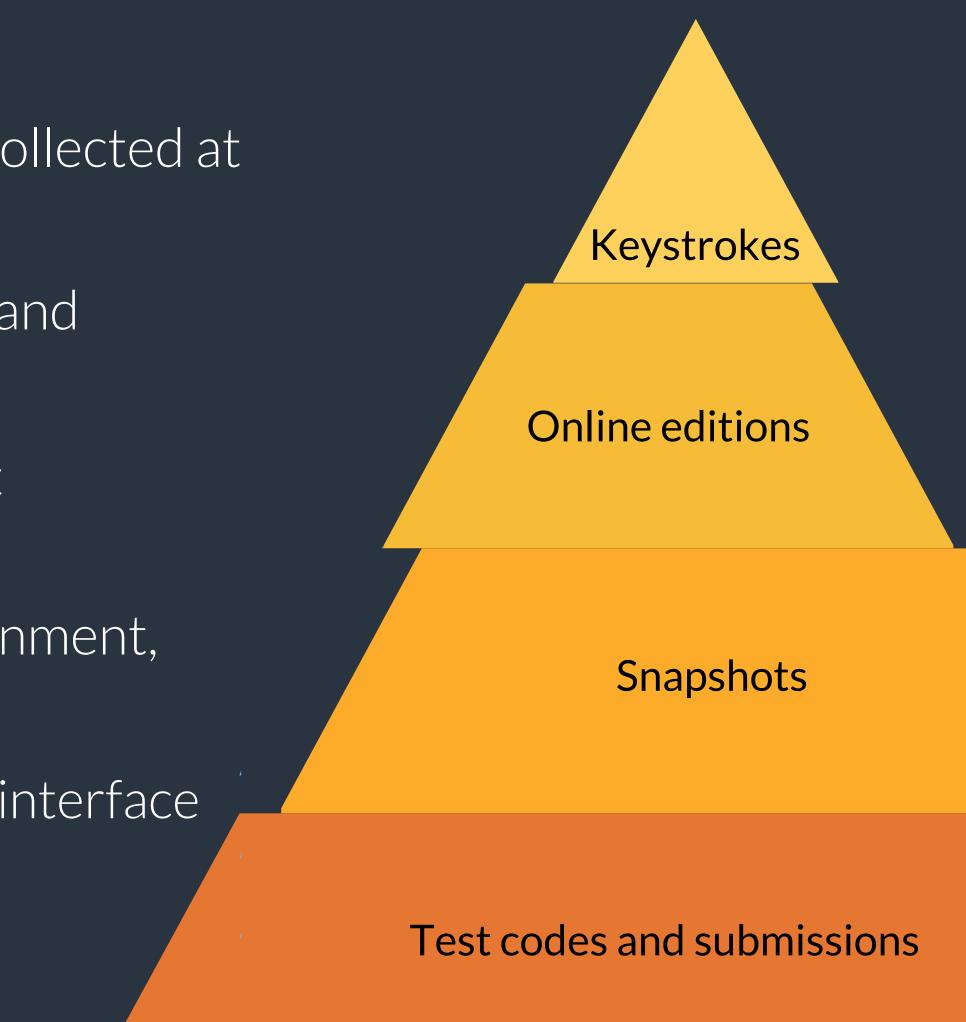
 \checkmark

 \checkmark

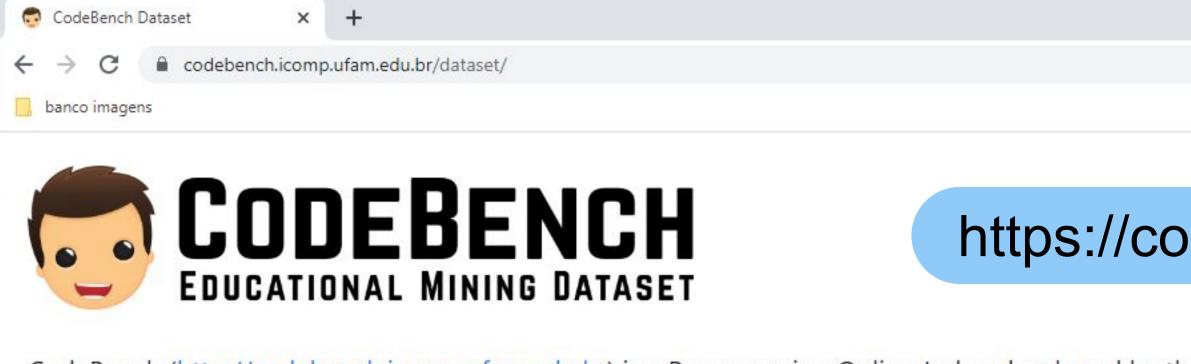
In **CodeBench**, this collection is done at keystroke granularity



- In addition to the activities in the environment, the following are collected:
- Mouse movements and clicks in the interface
 - Learning resources accessed
 - Logins and logouts
 - Activities on the gamification platform







CodeBench (http://codebench.icomp.ufam.edu.br) is a Programming Online Judge developed by the Institute of Computing (IComp) of the Federal University of Amazonas, Brazil. Through Codebench, teachers can provide lists of programming exercises to their students, who in turn must develop solutions for each exercise through an embedded IDE. Once a student submits a source code for a given exercise, the system instantly notifies the student whether him/her solution is correct or not.

The CodeBench automatically logs all actions performed by students on embedded IDE during their attempts to solve the proposed exercises. This dataset contains all logs collected from CS1 students during 2016 to 2021.

Releases

CodeBench Dataset 1.6

- This release of the dataset contains all logs collected from CS1 students during the years 2016 to 2021. Each academic year is divided into two semesters.
- Obs: Because of the COVID-19 pandemic that started in 2020, the semesters 1 and 2 of such year occurred throughout 2021. In 2020, there was a single semesters Emergency Remote Education), where only 3 classes of IPC were taught.
- Release date: 2022-10-21
- Download options:

• A file per semester: 2016-1, 2016-2, 2017-1, 2017-2, 2018-1, 2018-2, 2019-1, 2019-2, 2020-ERE, 2020-1, 2020-2, 2021-1 and 2021-2.

Release statistics:

	2016- 1	2016-	2017- 1	2017- 2		2018- 2		2019- 2	2020-ERE	2020- 1	2020- 2	2021- 1	2021- 2	Total
		2												
CS1 classes	9	5	10	5	9	5	9	7	3	7	6	7	8	90
Total number of students	471	172	463	177	465	180	489	297	492	228	200	224	219	4077
Homework exercises	681	447	1278	556	1550	893	1559	1116	420	1008	882	962	1180	12532
Exam exercises	124	110	163	103	182	107	176	138	8	12	8	154	195	1480

https://codebench.icomp.ufam.edu.br/dataset

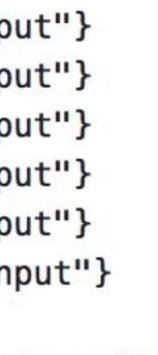
Since 2013, **CodeBench** has been used by about 6700 students and has more than 5600 exercises.

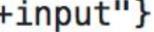


https://codebench.icomp.ufam.edu.br/dataset

2017-3-13 17:28:44.965#viewportChange#0 2017-3-13 17:28:46.733#change#{"from":{"line":2,"ch":0},"to":{"line":2,"ch":0},"text":["p"],"removed":[""],"origin":"+input"} 2017-3-13 17:28:46.885#change#{"from":{"line":2,"ch":1},"to":{"line":2,"ch":1},"text":["r"],"removed":[""],"origin":"+input"} 2017-3-13 17:28:47.109#change#{"from":{"line":2,"ch":2},"to":{"line":2,"ch":2},"text":["i"],"removed":[""],"origin":"+input"} 2017-3-13 17:28:47.373#change#{"from":{"line":2,"ch":3},"to":{"line":2,"ch":3},"text":["n"],"removed":[""],"origin":"+input"} 2017-3-13 17:28:47.605#change#{"from":{"line":2,"ch":4},"to":{"line":2,"ch":4},"text":["t"],"removed":[""],"origin":"+input"} 2017-3-13 17:28:50.369#change#{"from":{"line":2,"ch":5},"to":{"line":2,"ch":5},"text":["()"],"removed":[""],"origin":"+input"} 2017-3-13 17:28:50.369#keyHandled#"'('" 2017-3-13 17:28:52.083#change#{"from":{"line":2,"ch":6},"to":{"line":2,"ch":6},"text":["\"\""],"removed":[""],"origin":"+input"} 2017-3-13 17:28:52.083#keyHandled#"'\"'" 2017-3-13 17:30:31.907#viewportChange#0 2017-3-13 17:30:56.332#mousedown#{"isTrusted":true} 2017-3-13 17:30:56.349#focus#







LEARNING ANALYTICS IN ONLINE JUDGE SYSTEMS DASHBOARDS IN ONLINE JUDGE SYSTEMS



High rate of failures and dropouts in programming disciplines, notably for STEM courses, is a serious and recurring problem.



One way to deal with this problem is to provide information and recommendations that help teachers to mitigate student difficulties.



Which students are having difficulty in the subject?



What are the best exercises for each student?



What aspects does the student need to improve to do well in the discipline?



Which students are 'cheating' on assignment lists?



LEARNING ANALYTICS IN ONLINE JUDGE SYSTEMS DASHBOARDS IN ONLINE JUDGE SYSTEMS



High rate of failures and dropouts in program disciplines, notably for STEM courses, is a se recurring problem.



One way to deal with this problem is to provide information and recommendations that help te to mitigate student difficulties.



Which students are having difficulty in the subject?



What are the best exercises for each student?



What aspects does the student need to improve to do well in the discipline?

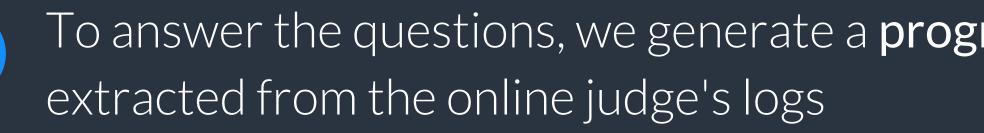


Which students are 'cheating' on assignment lists?

Would it be possible to answer all these questions through the usage of logs of online judges?



LEARNING ANALYTICS IN ONLINE JUDGE SYSTEMS **PROGRAMMING PROFILES**



Ratio between pasted characters and typed characters Code changes between submissions

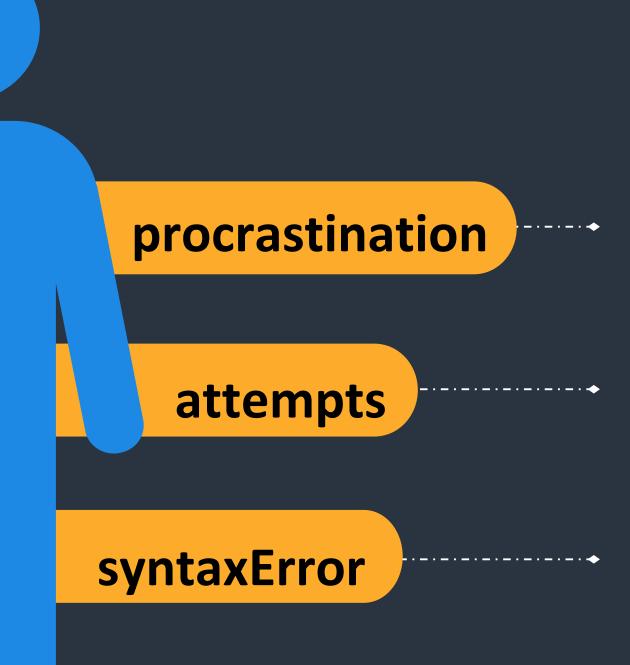
 \checkmark

Number of events during development

propPaste changes

qtyEvents

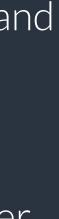
To answer the questions, we generate a **programming profile** for each student. from attributes

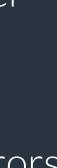


Time between the start of coding and the deadline

Average number of submissions per exercise

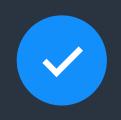
Proportion of submissions with errors





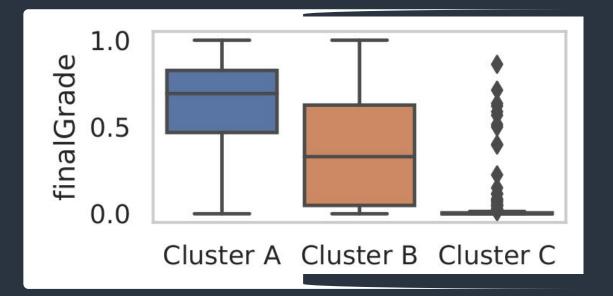
WHO ARE THE STUDENTS WITH DIFFICULTY?

EFFECTIVE AND INEFFECTIVE ATTITUDES



Using clustering techniques (K-means) on programming profiles, students are grouped into 3 large groups:

Effective, average and non-effective students





The graph on the side shows that the student's grade is directly related to the student's behavior in the online judge

Pereira, Filipe D.; OLIVEIRA, ELAINE HT; OLIVEIRA, DAVID BF; CRISTEA, ALEXANDRA I.; CARVALHO, LEANDRO SG; FONSECA, SAMUEL C.; TODA, ARMANDO; ISOTANI, SEIJI. Using learning analytics in the Amazon: understanding students? behavior in introductory programming. British Journal of Educational Technology, vol. 51, p. 955-972, 2020.



Cluster C Cluster A Cluster B errorQuotient keystrokeLatency syntaxError watWinScore ideUsage copyPaste correctness procrastination events firstExamGrade systemAccess lloc attempts countVar event_activity amountOfChange



PERFORMANCE PREDICTION



To deal with the high failure and dropout rates, it is essential to identify struggling students in advance.



Using Deep Learning around programming profiles, it was possible to predict student performance with an accuracy of 82.5% in the first two weeks of class.



- With advance performance prediction, teachers can:
 - Provide extra assignments for students who do well in the class



Provide personalized support for struggling students

PEREIRA, FILIPE DWAN; FONSECA, SAMUEL C.; OLIVEIRA, ELAINE HT; OLIVEIRA, DAVID BF; CRISTEA, ALEXANDRA I.; CARVALHO, LEANDRO SG. Deep learning for early performance prediction of introductory programming students: a comparative and explanatory study. BRAZILIAN JOURNAL OF COMPUTING IN EDUCATION, v. 28, p. 723-748, 2020.







EXPLAINING EFFECTIVE AND INEFFECTIVE BEHAVIORS



In addition to predicting performance, prediction methods such as SHAP, based on games theory, are able to explain the results of their predictions.

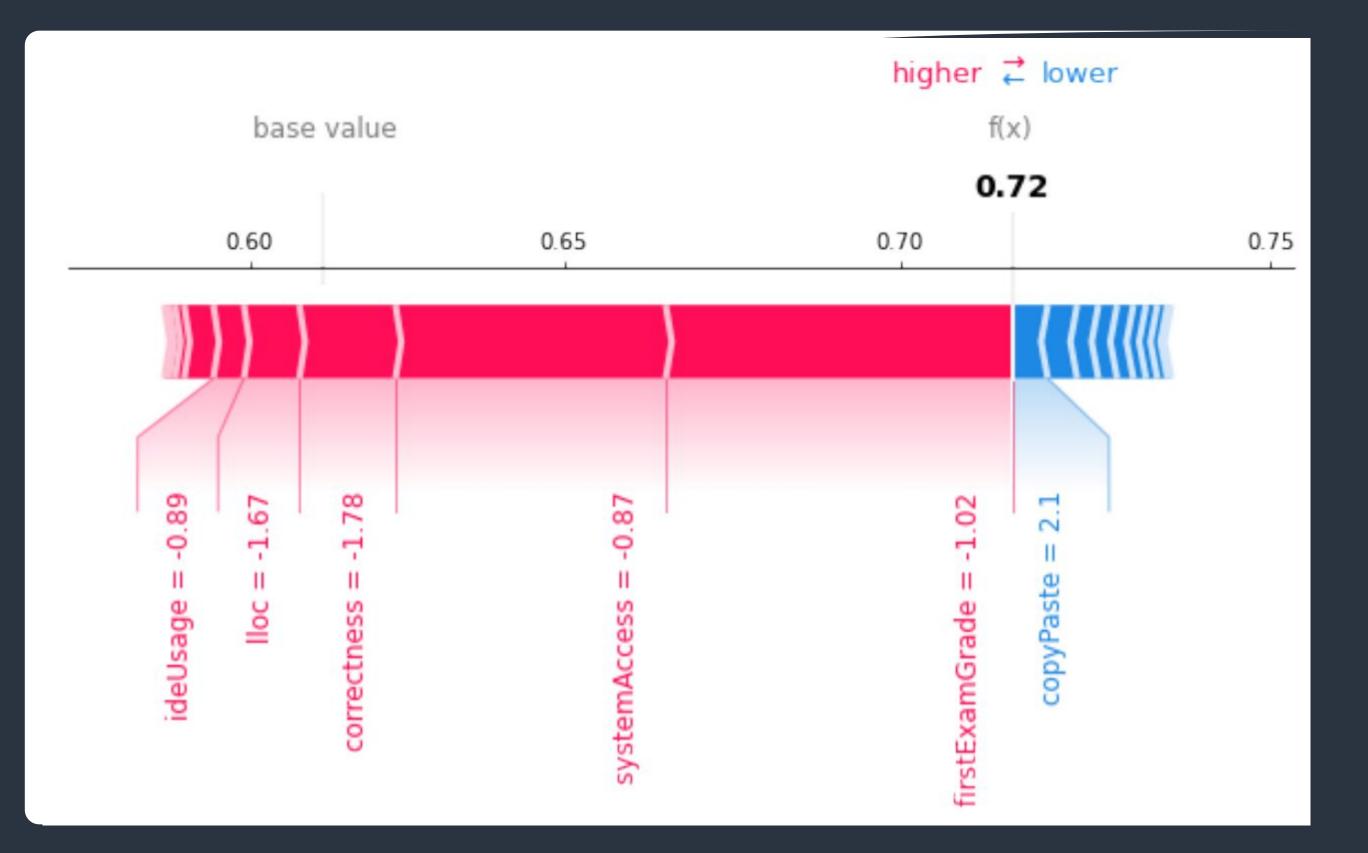


Early prediction empowered by its explanation may lead to better interventions.

PEREIRA, FILIPE DWAN; FONSECA, SAMUEL C.; OLIVEIRA, ELAINE HT; CRISTEA, ALEXANDRA I.; BELLHAUSER, HENRIK; RODRIGUES, LUIZ; OLIVEIRA, DAVID BF; ISOTANI, SEIJI; CARVALHO, LEANDRO SG Explaining Individual and Collective Programming Students? Behavior by Interpreting a Black-Box Predictive Model. IEEE Access v. 9, p. 117097-117119, 2021.







STIMULATING **EFFECTIVE BEHAVIORS**



We already know what behaviors lead to good performance in programming disciplines



In this way, would it be possible to use customized gamification to encourage each student to behave in a more effective way?



Gamification in CodeBench is based on RPG games, in which students need to learn programming to break free from a virtual world where they are trapped.

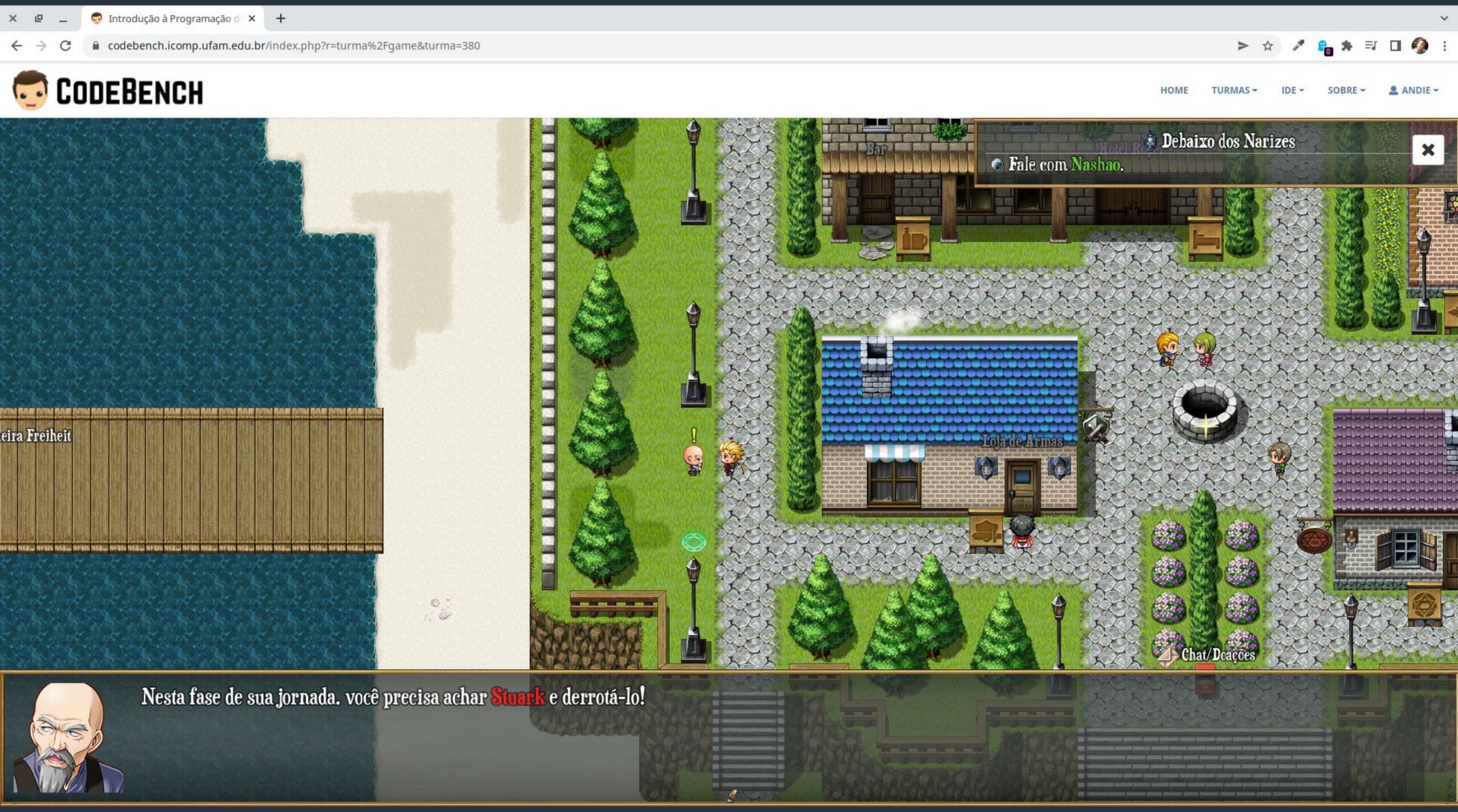
PESSOA, MARCELA; MELO, RAFAELA; HAYDAR, GABRIEL; OLIVEIRA, DAVID BF DE; CARVALHO, LEANDRO SG; OLIVEIRA, ELAINE HT DE; CONTE, TAYANA; PEREIRA, FILIPE DWAN; RODRIGUES, LUIZ; ISOTANI, SEIJI. An analysis of player types on a gamification platform incorporated into an online judge system. In: Brazilian Symposium on Informatics in Education, 2021, Brazil.

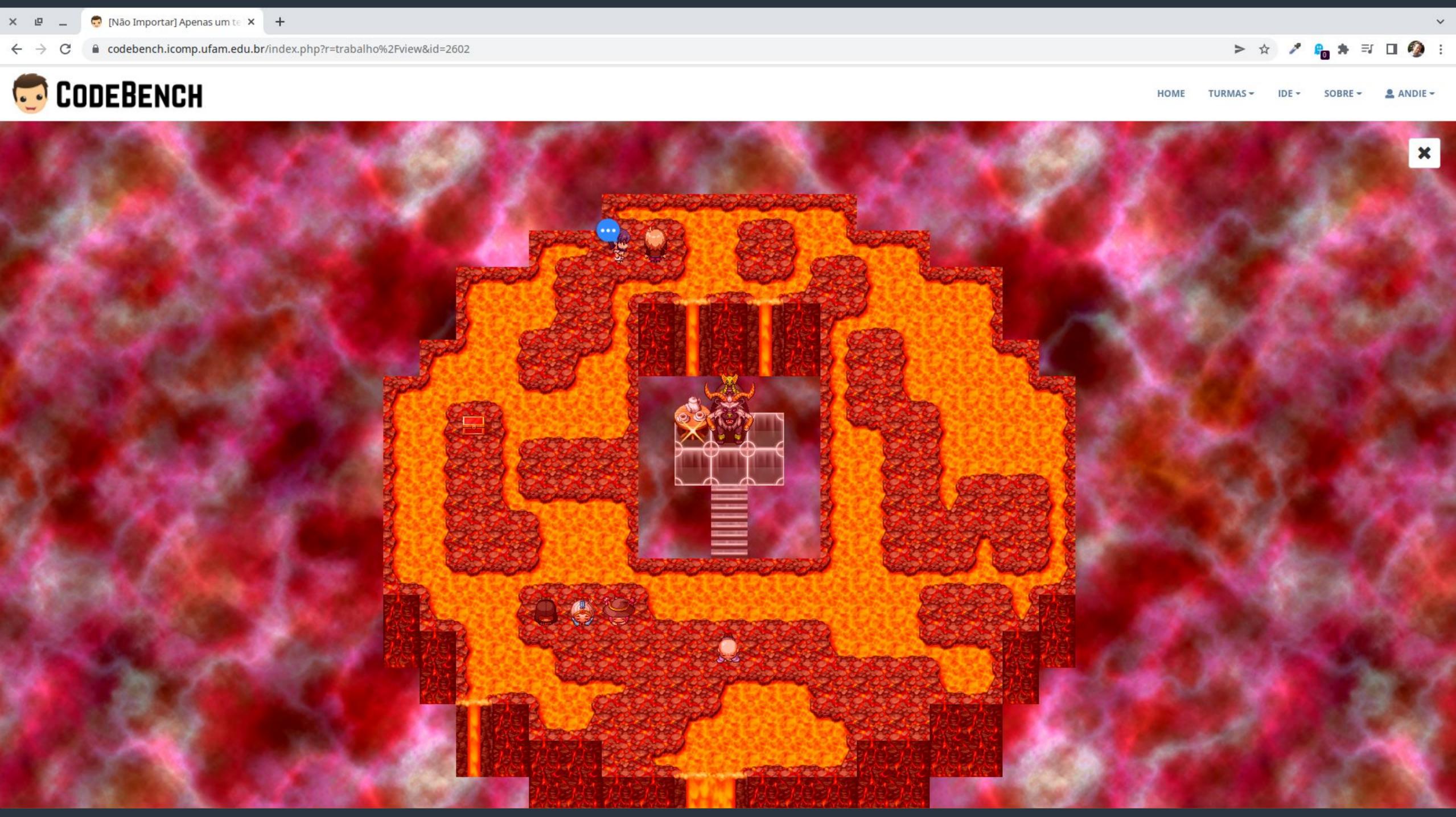




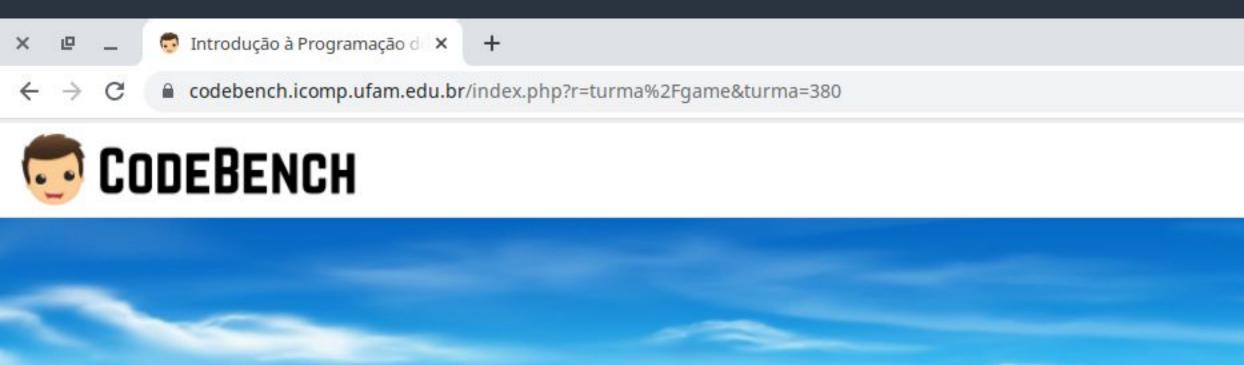


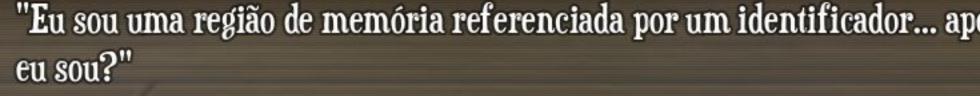


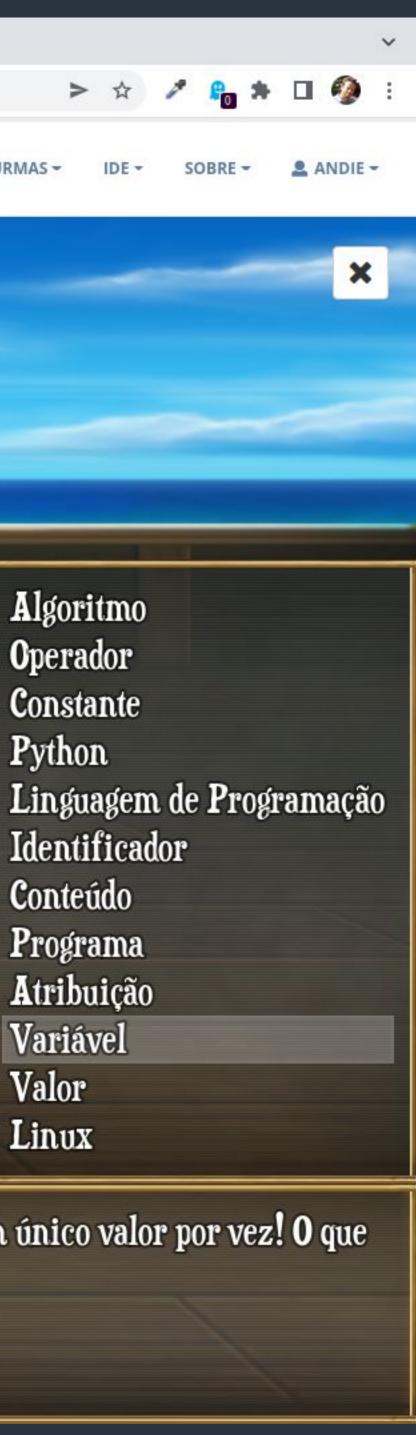












HOME TURMAS



Algoritmo **O**perador Constante Python Identificador Conteúdo Programa Atribuição Variável Valor Linux

"Eu sou uma região de memória referenciada por um identificador... apesar de poder receber vários valores distintos. eu só posso assumir um único valor por vez! O que

codebench.icomp.ufam.edu.br/index.php?r=turma%2Fgame&turma=380

🗔 CODEBENCH

× e

Elaine Harada Teixeira de Oliveira - Turma EMEC1/C2 Introdução à Programação de Computadores

Sob o Domínio de Marduk

Enredos da gamificação:

Geral Ca

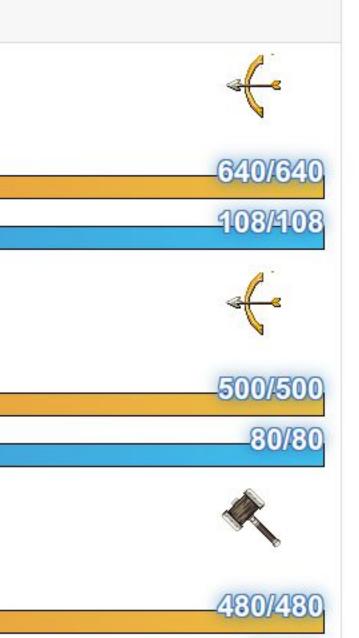
Capítulo 1 Capítulo 2

Capítulo 3

Você e um grupo de amigos foram presos em um mundo virtual controlado por um ser extremamente poderoso chamado Marduk. Esse reino virtual, uma espécie de Matrix, é chamado de Midgard e foi desenvolvido através de uma tecnologia computacional muito sofisticada. A única forma de se libertar desse reino virtual é através de um portal mágico protegido por um monstro, a terrível Quimera, uma besta virtual muito impiedosa e poderosa. Como o mundo virtual foi desenvolvido através de tecnologia computacional e de linguagens de programação disponíveis no mundo real, a única forma de vencer a Quimera e se proteger dos demais perigos do mundo virtual é através do conhecimento de programação. Estudar programação faz você aumentar seu poder, possibilitando que você se transforme em um ser realmente poderoso, conhecido no mundo virtual como **CodeMeistre**.

 Guerreiros e Guerreiras — Grupo Ouro
 Image: Comparison of the state of the s





Definições de termos

CARTAS DE RECOMPENSA (CARTAS DE THORIEL)



São as cartas sorteadas para um aluno quando um exercício é resolvido corretamente. As recompensas são: moedas, pontos de experiência, abertura de novos locais exploráveis e progressão

nos capítulos. Também conhecidas como **cartas de** Thoriel.

EXPERIÊNCIA

Pontos de experiência (EXP) podem ser sorteados ao fim da resolução de exercícios, encontrados explorando os mapas, e realizando missões.

NÍVEIS

Indica o nível de evolução do aluno. Quanto maior o nível, mais poderoso ele se torna. O nível é definido pela

DETECTING PLAGIARISM USING LEARNING ANALYTICS



In the initial programming classes, the codes developed by the students tend to be simple and small, making it difficult to use conventional plagiarism detection techniques.



In this way, instead of using traditional techniques, we chose to analyze the students' behavior during the creation of codes in the online judge's IDE



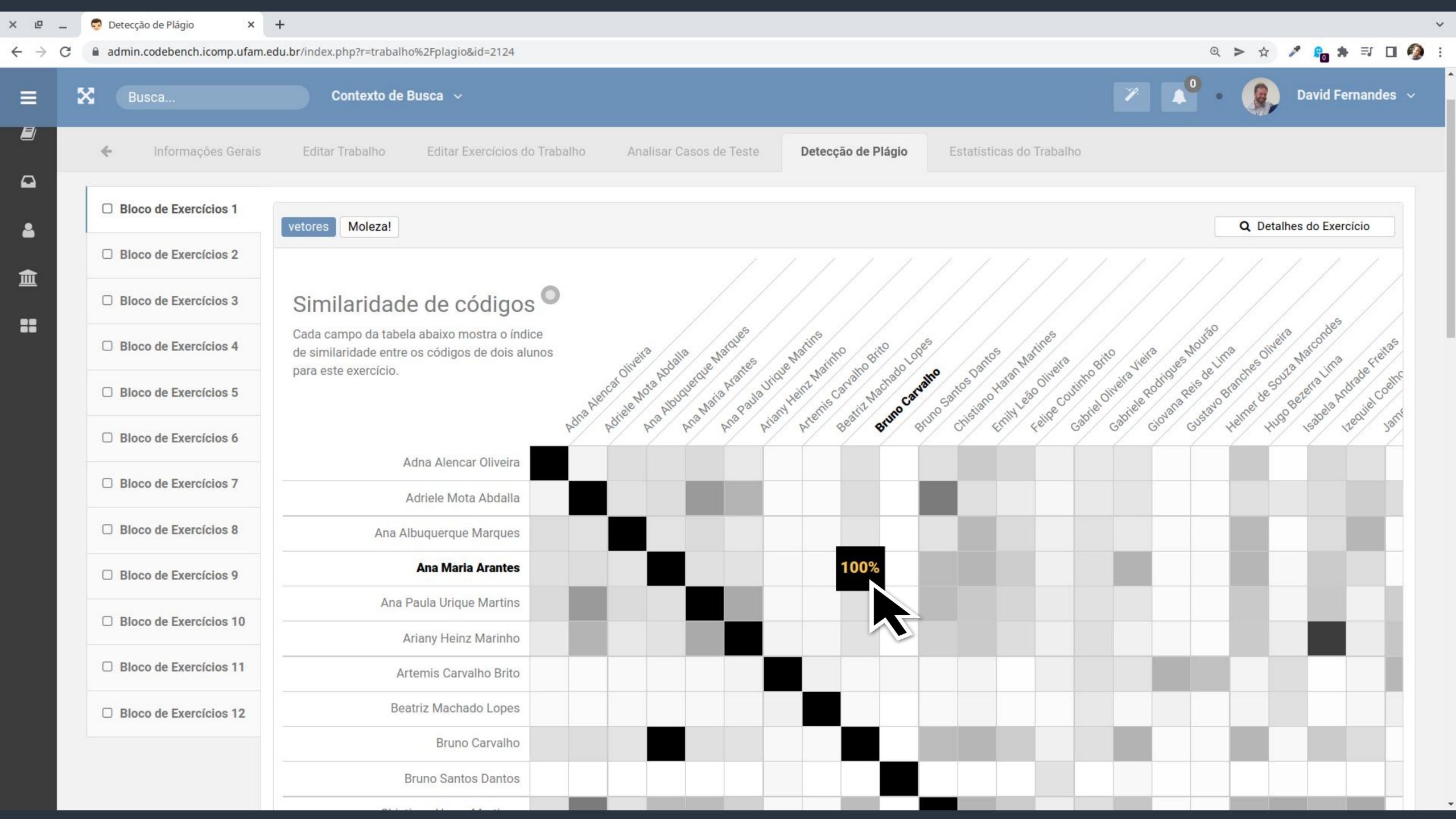
Starting from the premise that the behavior of those who are pasting is different from those who are actually solving the exercises, we arrived at the following results:

	Precisão	Revocação	Medida-F
Plágio	0.806	0.844	0.824
Não-plágio	0.846	0.809	0.827
Média	0.827	0.826	0.826

OLIVEIRA, DAVID BF DE; LAVAREDA FILHO, RONEM M.; OLIVEIRA, ELAINE HT; CARVALHO, LEANDRO SG; PEREIRA, FILIPE DWAN; COLONNA, JUAN G.; MENEZES, ADRIA. A Plagiarism Detection Method for Online Judge Systems Based on Student Behavior. In: Brazilian Symposium on Informatics in Education, 2021, Brazil. P. 836.

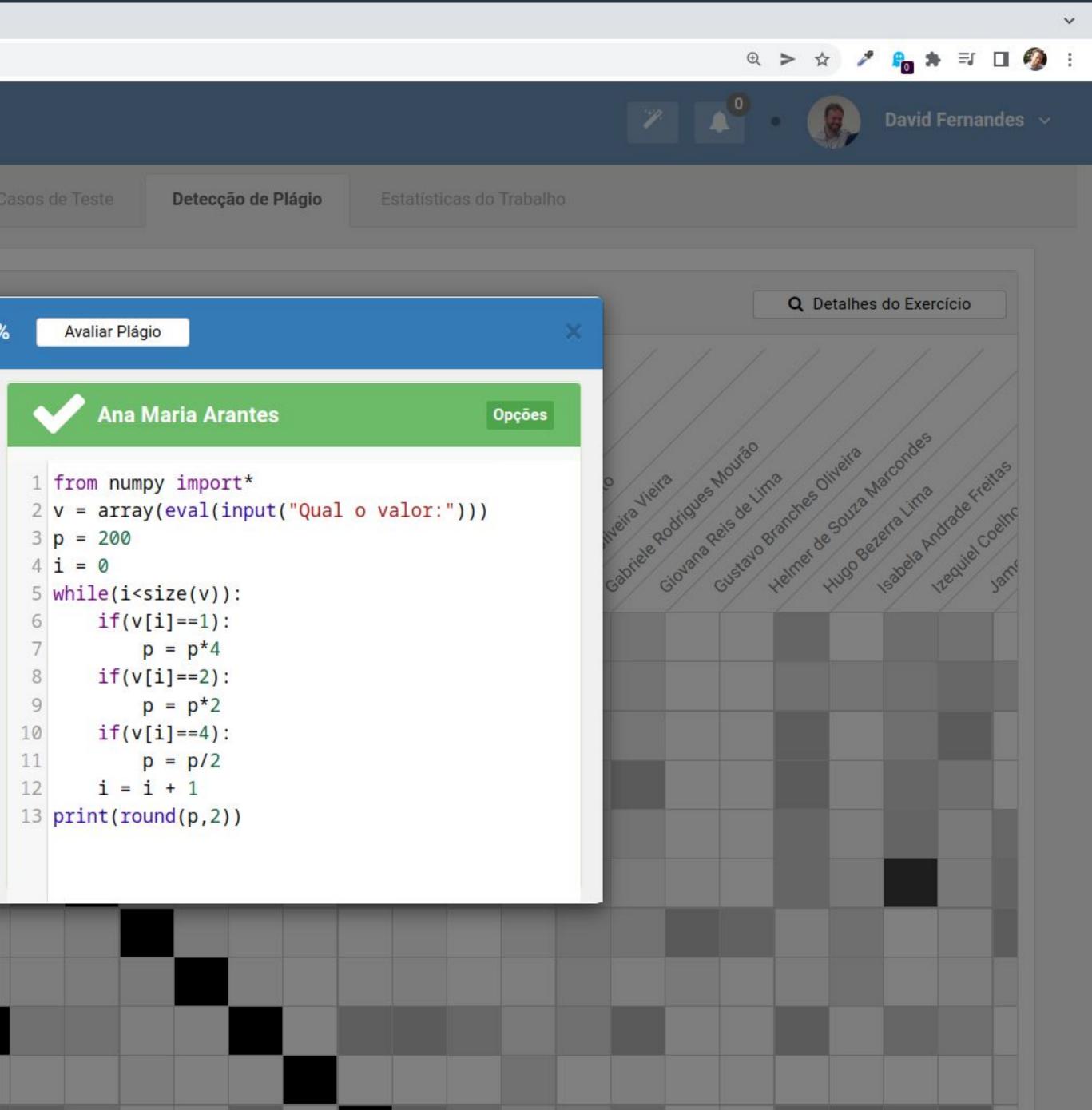






× + 👨 Detecção de Plágio × @ _ admin.codebench.icomp.ufam.edu.br/index.php?r=trabalho%2Fplagio&id=2124 \leftrightarrow C X Contexto de Busca 🗸 \equiv 4 Bloco de Exercícios 1 * Bloco de Exercícios 1 – Similaridade entre os códigos: 100% Bloco de Exercícios 2 盦 Bloco de Exercícios 3 **Bruno Carvalho** Opções Bloco de Exercícios 4 1 from numpy import* 2 v = array(eval(input("Qual o valor:"))) □ Bloco de Exercícios 5 3 p = 2004 i = 0 5 while(i<size(v)):</pre> Bloco de Exercícios 6 if(v[i]==1): 6 7 p = p*4Bloco de Exercícios 7 8 if(v[i]==2): 9 p = p*2□ Bloco de Exercícios 8 10 if(v[i]==4): 11 p = p/2i = i + 112 □ Bloco de Exercícios 9 13 print(round(p,2)) □ Bloco de Exercícios 10 □ Bloco de Exercícios 11 Artemis Carvalho Brito Beatriz Machado Lopes □ Bloco de Exercícios 12 Bruno Carvalho

Bruno Santos Dantos



OTHER RESEARCH INVOLVING LEARNING ANALYTICS



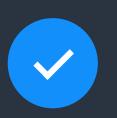
Classification of difficulty of programming exercises: by analyzing the behavior of students when proposed to an exercise, it is possible to infer the difficulty of it.



Detecting plagiarism: by analyzing the behavior of students in the system, we're trying to identify the characteristics of who is 'cheating'.



Automatic detection of students feeling confused: by analyzing the behavior of students in the system, it is possible to identify students who are feeling confused in a given exercise.



Group formation: by analyzing the students' learning paths, we'll analyse the best collaborative techniques and group arrangements.



LEARNING ANALYTICS IN ONLINE JUDGE SYSTEMS NEW CODEBENCH



In 2023, version 3.0 of the CodeBench system will be launched, which is more secure, scalable and modern



The new system will be released to institutions that want to conduct research involving online judges and learning analytics



- Some features of the new system:
 - Support for multiple languages

 \checkmark

Support for multiple institutions



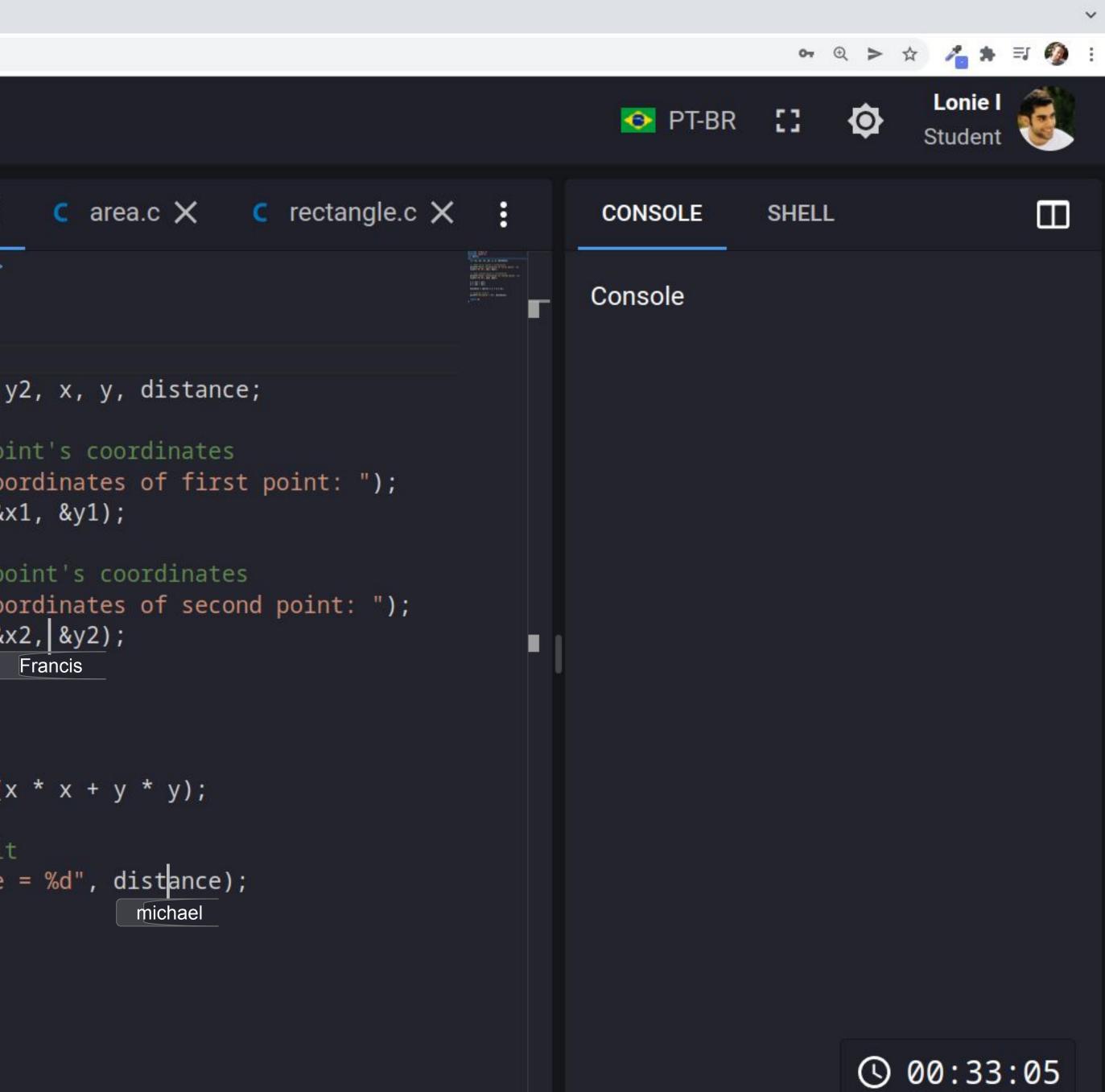
 \checkmark

- Support for creating study materials and interactive learning objects
- Classroom (similar to Google Classroom)



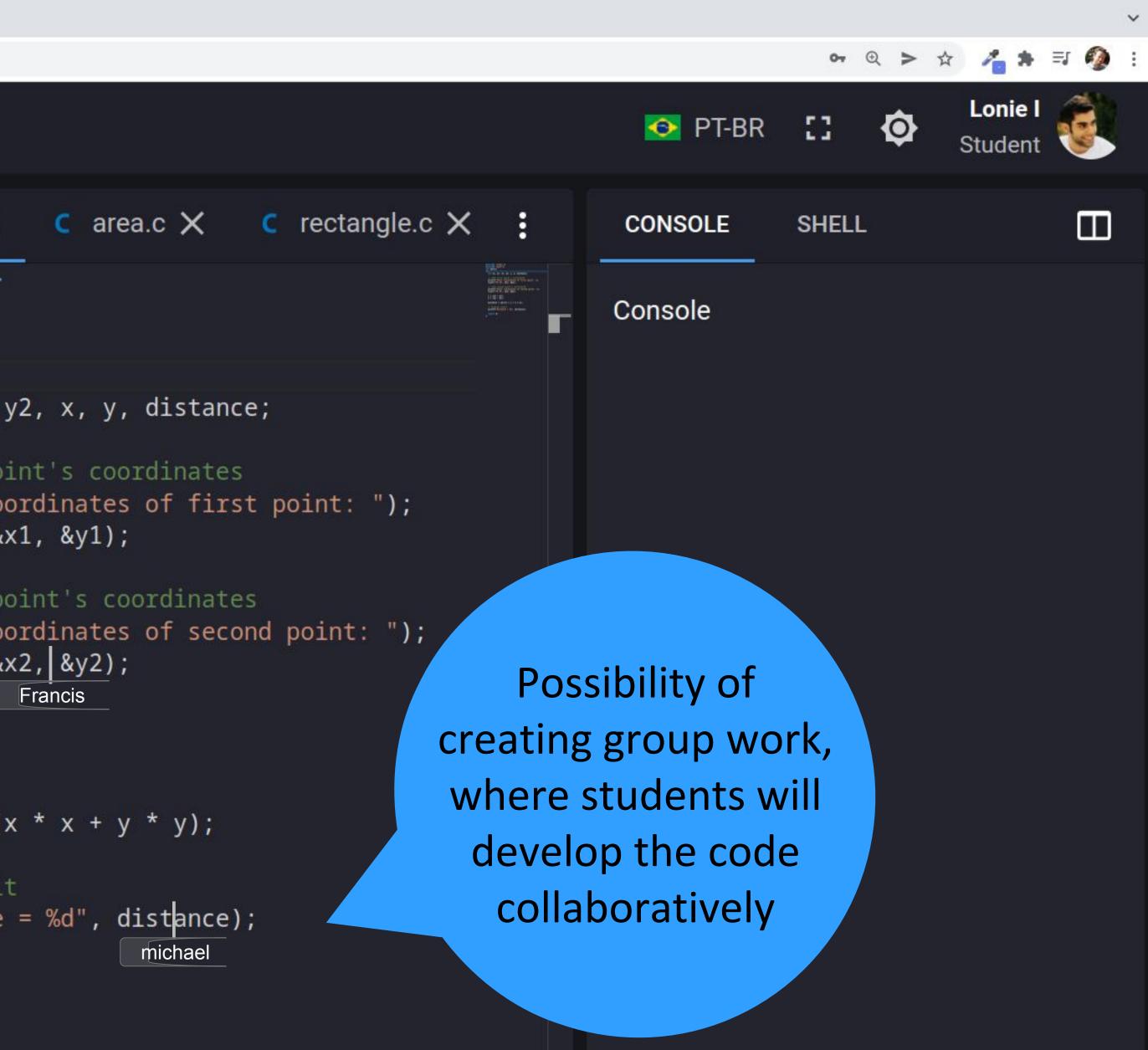


× @ _	😨 CodeMeistre 🗙 🕂	
$\leftrightarrow \ \ $	localhost:8081/ide	
\diamond	Universidade Federal do Amaz	zonas
i		C main.c X
01 02 03 04 05 06	 ENUNCIADO Escreva um programa que imprima o fatorial de um valor informado na entrada. Em matemática, o fatorial de um número natural n, representado por n!, é o produto de todos os inteiros positivos menores ou iguais a n. 	<pre>1 #include <stdio.h> 2 #include <math.h> 3 int main() 4 { 5 int x1, y1, x2, y 6 7 // take first poi 8 printf("Enter coor 9 scanf("%d %d", &x 10 11 // take second point 12 printf("Enter coor 12 printf("Enter coor 12 printf("Enter coor 12 printf("Enter coor 12 printf("Enter coor 12 printf("Enter coor 13 printf("Enter coor 14 printf("Enter coor 15 printf("Enter coor 16 printf("Enter coor 17 printf("Enter coor 10 printf("Enter coor 10</math.h></stdio.h></pre>
07	✓ DICAS	13 scanf("%d %d", &x 14
08 09	É importante não se esquecer dos casos bases, isto é, dos fatoriais de 0 e de 1.	<pre>15 x = (x2 - x1); 16 y = (y2 - y1); 17 18 distance = sqrt(x 19</pre>
10 11	 CASOS DE TESTE DE EXEMPLO 	20// display result21printf("Distance22
12	Caso de Teste 1 EXECUTAR >_ Entrada 16	23 return 0; 24 } 25
~	Saída 0 1 1 2 3 5 8 13 21 34	

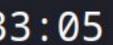


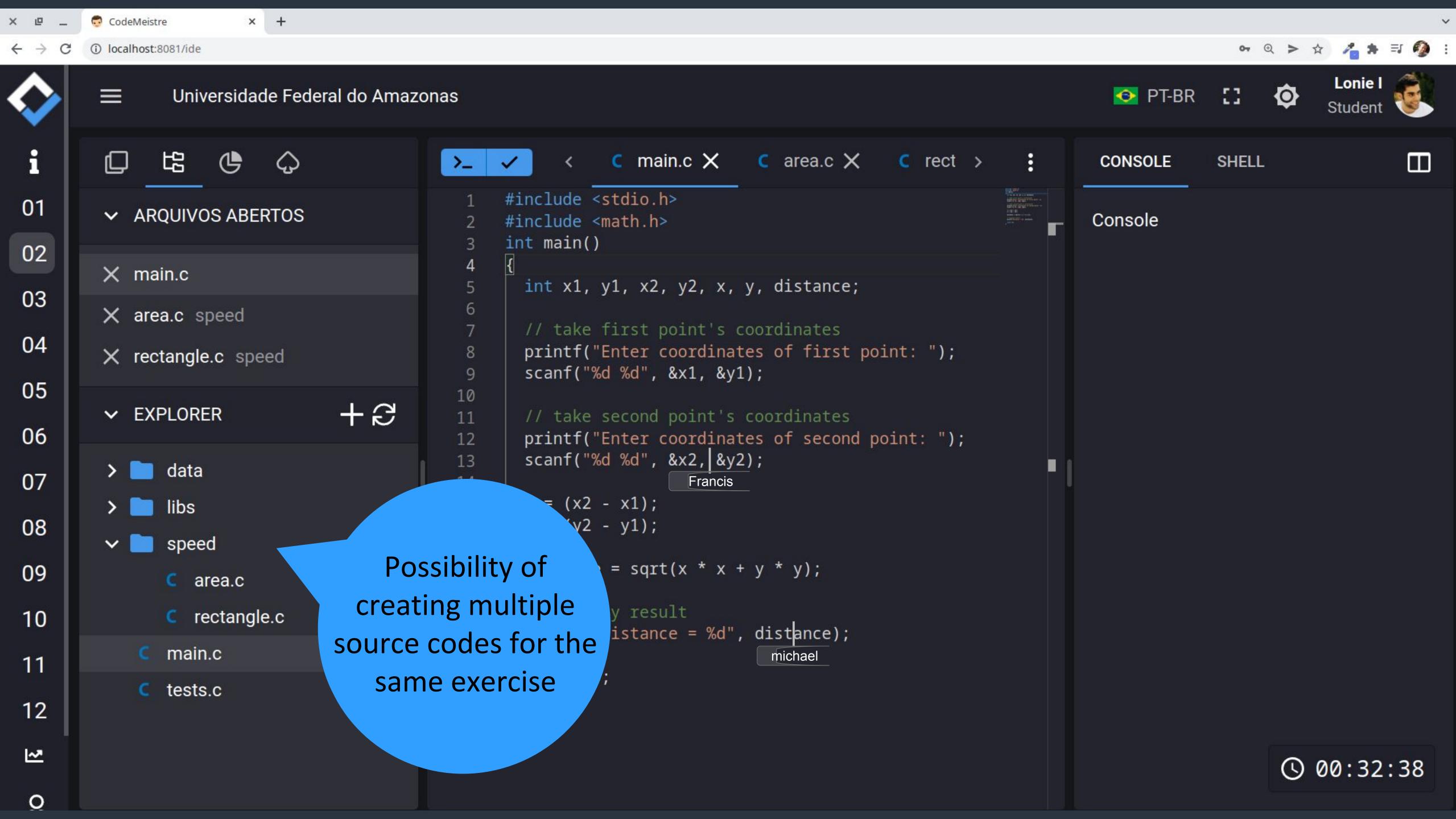
× @ _	CodeMeistre × +	
$\leftrightarrow \ \ \rightarrow \ \ G$	localhost:8081/ide	
\diamond	Universidade Federal do Amaz	onas
i		>_ ✓ C main.c X
01 02 03 04 05 06	 ENUNCIADO Escreva um programa que imprima o fatorial de um valor informado na entrada. Em matemática, o fatorial de um número natural n, representado por n!, é o produto de todos os inteiros positivos menores ou iguais a n. 	<pre>1 #include <stdio.h> 2 #include <math.h> 3 int main() 4 { 5 int x1, y1, x2, y 6 7 // take first poi 8 printf("Enter coo 9 scanf("%d %d", &x 10 11 // take second po 12 printf("Enter coo</math.h></stdio.h></pre>
07	✓ DICAS	13 scanf("%d %d", &x 14
08 09	É importante não se esquecer dos casos bases, isto é, dos fatoriais de 0 e de 1.	<pre>15 x = (x2 - x1); 16 y = (y2 - y1); 17 18 distance = sqrt(x 19</pre>
10 11	 CASOS DE TESTE DE EXEMPLO 	20// display result21printf("Distance22
12	Caso de Teste 1 EXECUTAR >_ Entrada 16	23 return 0; 24 } 25
~	Saída 0 1 1 2 3 5 8 13 21 34	

0



00:33:05







(I) Brazilian-German Symposium on Learning Analytics Mainz, May 12th 2023

Would you like to collaborate? :)

Elaine Harada Teixeira de Oliveira Research Group on Computing Education Institute of Computing Federal University of Amazonas (UFAM)



JOHANNES GUTENBERG UNIVERSITÄT MAINZ



Alexander von **HUMBOLDT** STIFTUNG

