

Minerando dados de um juiz on-line para prever a evasão de estudantes em disciplinas introdutórias de programação

Dayvson Silva¹ Sergio Cleger^{1,3} Marcela Pessoa^{1,2}
Fernanda Pires^{1,2} David B. F. Oliveira² Elaine H. T. Oliveira²
Leandro S. G. Carvalho²

¹Escola Superior de Tecnologia – Universidade do Estado do Amazonas
Manaus – AM – Brasil

²Programa de Pós-graduação em Informática
Instituto de Computação – Universidade Federal do Amazonas
Manaus – AM – Brasil

³Sidia - Instituto de Ciência e Tecnologia
Manaus – AM – Brasil

- **Evasão:** alunos que não completam **cursos** ou **programas** de estudo (se **matriculam** e desistem até antes iniciar).¹
- As Instituições de **Ensino Superior** há anos **enfrentam** o problema de evasão escolar
 - Especialmente nas áreas de **Ciência Exatas**
 - Disciplinas **Introdutórias de Programação**

¹Maia, M. d. C. and Meirelles, F. d. S. (2005)

Pergunta Norteadora



É possível prever **prematuramente** se estudantes de introdução à programação vão **desistir**, utilizando algoritmos de **Aprendizagem de Máquina**, sobre dados de *logs* de um juiz on-line?

Objetivos

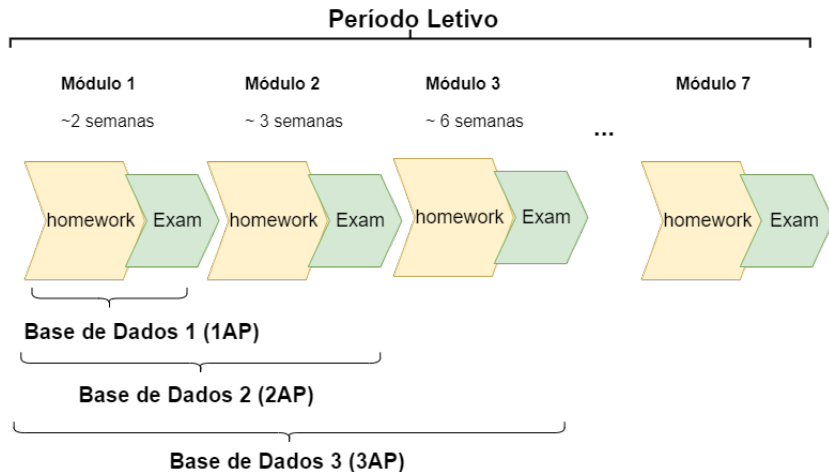
- **Geral:** Aplicar **técnicas de mineração de dados** nos arquivos de *logs* do juiz on-line
- **Específicos:**
 - **Explorar** os logs do juiz on-line **CodeBench**
 - **Criar** bases dados com os dados minerados
 - **Treinar** e **testar** diferentes modelos de Aprendizagem de Máquina

Metodologia aplicada

Descrição dos dados

- Base de Dados:
 - 2010 estudantes
 - 37 turmas com aproximadamente 40 alunos cada
 - Período: de 2016 a 2019
- Tipos de dados (38 atributos):
 - Logs de avaliações:
 - homework
 - exam
 - Logs de utilização do sistema:
 - número de logins
 - tempo médio

Metodologia aplicada



Algoritmos de Aprendizagem de Máquina

Algoritmos de Aprendizagem de Máquina utilizados:

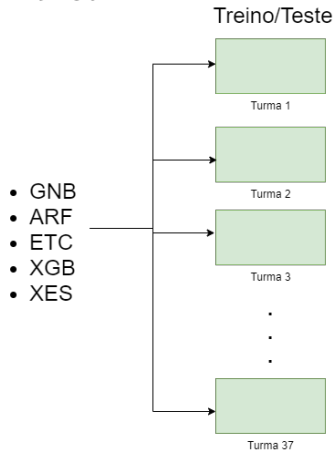
- *Gaussian naive bayes* (GNB)
- *Random Forest* (ARF)
- *Extra tree classifier* (ETC)
- *Xgboost* (XGB)
- *Xgboost with early stopping* (XES)

Experimentos

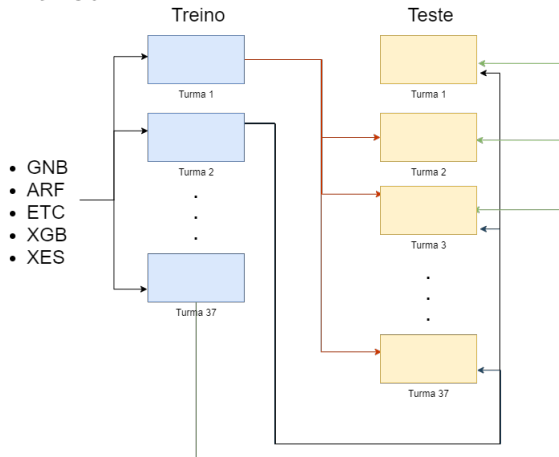
- Duas análises:
 - Avaliação da performance por turma
 - Avaliação do desempenho geral (com as 37 turmas)
- Métrica para avaliação dos resultados
 - Acurácia
- Os cinco algoritmos de **Aprendizagem de Máquina** foram utilizados em todos os experimentos
- Os **resultados** foram **validados** com o teste estatístico de **Friedman**

Descrição dos experimentos

Análise 1



Análise 2



- **Análise 1:** performance por turma.
 - Teste de Friedman: ($\rho < 0.05$) nos resultados de 1AP e 3AP

Tabela: Médias do desempenho de todos os experimentos

	ARF	ETC	XGB	XES	GNB
1AP	0,8136	0,8561	0,8035	0,8035	0,7691
2AP	0,8698	0,8959	0,8627	0,8627	0,8351
3AP	0,9187	0,9194	0,9057	0,9057	0,9022

- **Análise 2:** generalizando os modelos.
 - Teste de Friedman:
 - $\rho < 0.01$ nos resultados de 1AP e 2AP
 - $\rho < 0.05$ para 1AP e 3AP

Tabela: Desempenho médio com os melhores e os piores resultados.

	ARF	ETC	XGB	XES	GNB	Melhor	Pior
1AP	0,8136	0,8543	0,7845	0,8035	0,7691	0,8871	0,4565
2AP	0,8678	0,8913	0,8360	0,8627	0,8351	0,9166	0,4770
3AP	0,9147	0,9196	0,9016	0,9057	0,9022	0,9473	0,4876

Resultados

Tabela: Acurácias médias das melhores e piores turmas como generalizadoras.

	1AP	2AP	3AP
1 º Melhor	0,8742	0,9034	0,9351
2 º Melhor	0,8666	0,8998	0,9339
3 º Melhor	0,8637	0,8985	0,9334
4 º Melhor	0,8617	0,8977	0,9329
5 º Melhor	0,8597	0,8960	0,9323
5 º Pior	0,7451	0,8223	0,8913
4 º Pior	0,7392	0,8164	0,8902
3 º Pior	0,6722	0,7715	0,8563
2 º Pior	0,6112	0,6738	0,8003
1 º Pior	0,5787	0,6591	0,7431

Considerações finais

- Dos cinco algoritmos utilizados em cada experimento, destaque para o ETC
 - **Análise 1:** Acurácia média de 88,6% e 91,6%
 - **Análise 2:** Acurácia média 85,6% e 91,9%
 - **Precisão em classificar desistentes:** 69,7%, 80,1%
- Adição de dados para realização de experimentos adicionais

Agradecimentos

- Projeto Samsung-UFAM de Ensino e Pesquisa (SUPER)
- Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES)
- Programa de Bolsas de Iniciação Científica e Tecnológica da UEA (PBICT/UEA).

Minerando dados de um juiz on-line para prever a evasão de estudantes em disciplinas introdutórias de programação

Dayvson Silva¹ Sergio Cleger^{1,3} Marcela Pessoa^{1,2}
Fernanda Pires^{1,2} David B. F. Oliveira² Elaine H. T. Oliveira²
Leandro S. G. Carvalho²

¹Escola Superior de Tecnologia – Universidade do Estado do Amazonas
Manaus – AM – Brasil

²Programa de Pós-graduação em Informática
Instituto de Computação – Universidade Federal do Amazonas
Manaus – AM – Brasil

³Sidia - Instituto de Ciência e Tecnologia
Manaus – AM – Brasil