

## Análise da relação entre perfil e desempenho acadêmico dos alunos matriculados na disciplina de Introdução à Programação utilizando algoritmos de classificação

*Roberto Silva de Oliveira Júnior<sup>a</sup>, Douglas Valentim de Almeida Cardins<sup>b</sup>, Wendell Juler Pereira de Lima<sup>c</sup>,  
Thereza Patrícia Pereira Padilha<sup>d</sup> e Vanessa Farias Dantas<sup>e</sup>*

**Resumo:** Ao ingressar na Universidade, a maioria dos alunos enfrenta dificuldades no aprendizado dos conteúdos da disciplina de Introdução à Programação, primeiro componente curricular de programação nos cursos de Computação e Informática. Muitos alunos terminam evadindo ou reprovando vários semestres nesta disciplina e, assim, comprometem a conclusão de seus cursos. Sabe-se, com base em estudo da literatura, que isso se deve principalmente a aspectos como falta de motivação dos alunos, deficiências do ensino médio e dificuldades com raciocínio lógico e conceitos abstratos. Entretanto, outros fatores podem interferir nos resultados dos discentes. Diante disso, o presente artigo tem como proposta utilizar algoritmos de classificação, disponíveis na ferramenta Weka, para investigar se existe relação entre o perfil (social, acadêmico e econômico) dos alunos e o seu desempenho acadêmico nessa disciplina, com o objetivo de guiar as ações da coordenação do curso e do corpo docente no combate à evasão e à reprovação. Para isso, foram testados dados coletados de um curso de Licenciatura em Ciência da Computação da Universidade Federal da Paraíba. Foram encontradas regras que classificam os alunos em aprovado ou reprovado. Assim, foi constatado que alunos com gasto elevado com transporte, sem conhecimento prévio de computação e idade maior ou igual a 19 anos são propensos a reprovação, já os que não trabalham e têm idade menor ou igual a 18 anos e têm conhecimento prévio de computação são mais propensos a aprovação na disciplina.

**Palavras-chave:** Programação. Evasão. Reprovação. Mineração de dados.

## Analysis of the relationship between profile and academic performance of students enrolled in the Programming Introduction course using classification algorithms

**Abstract:** Upon entering the University, most of the students show difficulties to learn contents of the Introduction to Programming course, first programming course in the Computer Science Degree. Many students dropped off or failed several semesters in this course and, thus delayed their degrees. In the literature, several studies show that the main factors are: lack of motivation of the students, deficiencies of high school and difficulties with logical reasoning and abstract concepts. However, other factors may interfere with student outcomes. The present paper shows the use of classification algorithms, available in the Weka tool, to investigate if there is a relationship between the students (social, academic and economic) profile and their academic performance in this course, in order to guide actions of the course coordination and faculty to minimize evasion and retention. So, data were collected of a Computer

- 
- a Universidade Federal da Paraíba (UFPB). E-mail: [roberto.silva@dce.ufpb.br](mailto:roberto.silva@dce.ufpb.br). Currículo: <http://lattes.cnpq.br/5095958238014357>
- b Universidade Federal da Paraíba (UFPB). E-mail: [douglas.valentim@dce.ufpb.br](mailto:douglas.valentim@dce.ufpb.br). Currículo: <http://lattes.cnpq.br/0509156062041263>
- c Universidade Federal da Paraíba (UFPB). E-mail: [wendell.juler@dce.ufpb.br](mailto:wendell.juler@dce.ufpb.br). Currículo: <http://lattes.cnpq.br/2195755521034980>
- d Universidade Federal da Paraíba (UFPB). E-mail: [thereza@dce.ufpb.br](mailto:thereza@dce.ufpb.br). Currículo: <http://lattes.cnpq.br/2084409307725504>
- e Universidade Federal da Paraíba (UFPB). E-mail: [vanessa@dce.ufpb.br](mailto:vanessa@dce.ufpb.br). Currículo: <http://lattes.cnpq.br/5196069557142312>

Science course from Federal University of Paraiba. Rules were found that classify students as success or fail. So, students with high transportation costs, no prior knowledge of computing and age greater than or equal to 19 years possibly will be to fail, and students that do not work and are less than or equal to 18 years old and have previous knowledge of computation are predisposed for success in the course.

**Keyword:** Programming. Evasion. Retention. Data mining.

## 1 Introdução

Uma preocupação nas Instituições de Ensino Superior (IES) no Brasil e no mundo é a evasão universitária, definida como a saída do aluno de seu curso, sem concluí-lo. Sabe-se, ainda, que os índices de evasão são maiores nos primeiros períodos dos cursos (Bardagi e Hutz, 2009). No caso dos cursos da área de Computação e Engenharia, outra preocupação é o alto índice de retenção em disciplinas de programação. Nos últimos anos, pesquisadores têm concentrado estudos na busca de encontrar respostas para estes problemas analisando dados acadêmicos, sociais e econômicos (Silva, 2013; Araújo et al., 2011; Campello e Lins, 2008).

Segundo Fagundes et al. (2014), o desempenho acadêmico constitui também um potente indicador de qualidade dos cursos de uma instituição, ou seja, sinaliza se os objetivos estão sendo alcançados. Portanto, entende-se que há uma grande pertinência na realização de estudos sobre desempenho acadêmico no curso de Licenciatura em Ciência da Computação (LCC) da Universidade Federal da Paraíba, a fim de estabelecer ações para aumentar/manter a sua qualidade que, atualmente, possui conceito máximo (5) atribuído pelo Ministério da Educação.

O aumento da disponibilidade dos dados dos alunos e o interesse de pesquisadores das áreas de Mineração de Dados e Informática na Educação, principalmente, fizeram surgir uma nova área de pesquisa conhecida como Mineração de Dados Educacionais (MDE). MDE visa explorar dados do ambiente educacional para descobrir padrões (isto é, conhecimento oculto/escondido nos dados) para contribuir no entendimento do processo de aprendizagem e, assim, auxiliar as tomadas de decisão sobre novas práticas pedagógicas (Baker, 2010). Embora MDE ainda seja pouco explorada no Brasil, esta é vista como promissora na compreensão de problemas na área da Educação (Gottardo et al., 2013).

## 2 Objetivos

Este artigo tem como objetivo geral investigar uma relação entre o perfil (social, cultural, acadêmico e econômico) dos alunos e o seu desempenho acadêmico na disciplina de Introdução à Programação, do curso de Licenciatura em Ciência da Computação, a fim de minimizar os altos índices de evasão e de reprovação.

Em termos de objetivos específicos, tem-se:

- Categorizar, em um nível maior de detalhes, o perfil dos alunos ingressantes no curso de Licenciatura em Ciência da Computação da Universidade Federal da Paraíba;
- Realizar experimentos com o conjunto de dados construído (perfil, econômico e acadêmico) utilizando três algoritmos de classificação;
- Identificar variáveis que influenciam os alunos no sucesso e insucesso da disciplina de Introdução à Programação (retenção e evasão);
- Apresentar aos gestores acadêmicos e ao professor da disciplina os resultados da pesquisa.

## 3 Procedimentos metodológicos

No meio do semestre letivo 2017.1, foi aplicado um formulário web para 42 alunos matriculados na disciplina de IP com 19 perguntas objetivas envolvendo aspectos pessoais (Q1 a

Q6), econômicos (Q10 a Q12) e acadêmicos (Q13 a Q19). Uma identificação sobre o conteúdo de cada pergunta é apresentada a seguir:

- Q1 – gênero;
- Q2 – idade;
- Q3 – estado civil;
- Q4 – possui filhos;
- Q5 – possui computador;
- Q6 – possui internet em casa;
- Q7 – cidade de origem;
- Q8 – moram na cidade em que estudam;
- Q9 – gastos mensais com transportes;
- Q10 – trabalha;
- Q11 – recebe ajuda de custos;
- Q12 – horas semanais de trabalho;
- Q13 – tipo de escola frequentada no ensino médio;
- Q14 – cursou a disciplina de IP antes de entrar no curso;
- Q15 – motivo pelo qual desistiu ou reprovou anteriormente na disciplina;
- Q16 – conhecimento da língua inglesa;
- Q17 – conhecimento de computação antes de entrar no curso;
- Q18 – conhecimento do perfil do curso de LCC;
- Q19 – LCC foi a primeira opção no Enem.

Todas as perguntas do formulário fizeram parte do conjunto de dados obtido como atributos. Ao fim do semestre, foi disponibilizada a situação final dos alunos (aprovado ou reprovado), que foi adicionado ao conjunto de dados como atributo-classe (rótulo). Dos 42 registros, 34 eram da classe aprovado e 8 reprovado. Em seguida, os dados foram transformados e disponibilizados no formatado .arff, formato padrão de entrada para a ferramenta Weka.

Por fim, os dados foram aplicados aos algoritmos de classificação OneR, PART e JRip, disponíveis na ferramenta Weka, e que têm como saída um conjunto de regras.

## 4 Resultados

Nesta seção serão apresentados os resultados obtidos utilizando os algoritmos OneR, PART e JRip.

### 4.1 OneR

A Figura 1 apresenta as quatro regras geradas pelo algoritmo OneR. Neste caso, somente o atributo gastoComTransporte foi usado para separar as classes “aprovado” e “reprovado”. Onde classificou trinta e cinco instâncias corretamente (83.3%) e sete incorretamente (16.6%).

Observa-se que alunos com gasto elevado com transporte têm sido classificado como reprovado. Analisando o conjunto de dados, verifica-se que esses alunos moram em outras cidades e, até mesmo, em outros estados, e, assim, precisam se deslocar para as suas cidades diariamente. Sabe-se que este deslocamento no cotidiano acarreta em cansaço e falta de tempo para estudos e, conseqüentemente, influencia no desempenho das disciplinas cursadas.

Figura 1: Resultado Obtido pelo Algoritmo OneR.

```
=== Classifier model (full training set) ===

gastoComTransporte:
    Alto    -> Reprovado
    Baixo   -> Aprovado
    Medio   -> Aprovado
    Zero    -> Aprovado
(35/42 instances correct)
```

Fonte: Autores.

## 4.2 PART

A Figura 2 apresenta as quatro regras geradas pelo algoritmo PART. Neste caso, o algoritmo classificou 36 instâncias corretamente (85.7%) e 6 incorretamente (14.3%).

Figura 2: Resultado Obtido pelo Algoritmo PART.

```
PART decision list
-----

trabalha = Nao AND
idade <= 18: Aprovado (20.0)

conhecimentoPrevioDeComputacao = Sim: Aprovado (13.0/2.0)

trabalha = Nao AND
conhecimentoPrevioDoCurso = Sim: Aprovado (5.0/2.0)

: Reprovado (4.0)

Number of Rules :      4
```

Fonte: Autores.

Foram determinadas três regras para classificar os alunos como aprovados. De modo geral, o perfil dos alunos aprovados são: 1) não trabalham e possuem idade menor ou igual a 18; 2) tem conhecimento prévio sobre computação; e 3) não trabalham e tem conhecimento prévio do curso que ingressaram. Caso não esteja dentro de um desses perfis, o algoritmo determina a classe como reprovado.

## 4.3 JRip

Na Figura 3 podem ser observadas as duas regras extraídas pelo algoritmo JRip. Este algoritmo classificou 35 instâncias corretamente (83.3%) e 7 incorretamente (16.7%).

Figura 3: Resultado Obtido pelo Algoritmo Jrip.

```
JRIP rules:
=====
(conhecimentoPrevioDeComputacao = Nao) and
(idade >= 19) => Class=Reprovado (9.0/3.0)
=> Class=Aprovado (33.0/2.0)
```

Fonte: Autores.

Neste caso, basicamente, alunos são classificados como reprovados, aqueles que não possuem conhecimento prévio de computação e com idade maior ou igual a 19. A partir dos resultados obtidos com a ferramenta de mineração de dados Weka, pôde-se encontrar fatores que influenciam a reprovação e a aprovação de alunos matriculados em IP. De modo geral, gasto elevado com transporte, não conhecimento prévio de computação e idade maior ou igual a 19 anos podem levar a uma reprovação. Por outro lado, não trabalhar e ter idade menor ou igual a 18 anos, ter conhecimento prévio de computação são fatores que contribuem para uma aprovação na disciplina.

## 5 Considerações Finais

O crescente número de reprovações e evasões na disciplina de Introdução à Programação é alarmante e cabível de medidas que visem cessar tal crescimento. Dessa forma, foi primordial a captação de dados, a averiguação e o estudo deles para se obter respostas sobre os casos. Assim, a aplicação de Mineração de Dados apresentou-se como alternativa para obtenção de informações pertinentes.

Embora o conjunto de dados aplicado tenha sido restrito, principalmente, no que diz respeito à quantidade de registros de alunos reprovados, foi possível ter um modelo de perfil de alunos que perdem a disciplina de IP. Neste experimento não houve a presença de nenhum registro de aluno como evadido. Apesar de ter tido casos de alunos nesta situação durante o período avaliado, não foi possível obter seus dados pessoais, econômicos e acadêmicos destes alunos porque a coleta aconteceu no meio do semestre e os alunos já tinham desistido da disciplina. Portanto, para evitar estas perdas de dados, recomenda-se coletar os dados dos alunos até a segunda semana após o início das aulas.

## Referências

ARAÚJO, E. A. T; CAMARGOS, M. A.; CAMARGOS, M. C. S. Desempenho Acadêmico de Discentes do Curso de Ciências Contábeis: uma Análise dos seus Fatores Determinantes em Uma IES Privada. **XXXV Encontro da Associação Nacional de Pós-graduação e Pesquisa em Administração**. Rio de Janeiro – RJ, Brasil, 2011.

BAKER, R. S. J. Data mining for education. **International Encyclopedia of Education**, v. 7, n. 3, p. 112 – 118, 2010.

BARDAGI, M. P.; HUTZ, C. S. **Não havia outra saída": percepções de alunos evadidos sobre o abandono do curso superior.** Psico-USF (Impr.) vol.14 no.1 Itatiba Jan./ Apr. 2009.

BARROSO, M. F; FALCÃO, E. B. M. **Evasão Universitária: O Caso do Instituto de Física da UFRJ,** IX Encontro Nacional de Pesquisa em Ensino de Física. Rio de Janeiro, 2004.

BRANCO NETO, W. C.; SCHUVARTZ, A. A. Ferramenta Computacional de Apoio ao Processo de Ensino-Aprendizagem dos Fundamentos de Programação de Computadores. **18º Simpósio Brasileiro de Informática na Educação.** São Paulo, Brasil. 2007.

CAMPELLO, A. V. C.; LINS, L. N. Metodologia de Análise e Tratamento da Evasão e Retenção em Cursos de Graduação de Instituições Federais de Ensino Superior. **XXVIII Encontro Nacional de Engenharia de Produção.** Rio de Janeiro, RJ, Brasil, 2008.

DIAS, A. F. M.; CERQUEIRA, G. S.; LINS, L. N. Fatores Determinantes da Retenção Estudantil em um Curso de Graduação em Engenharia de Produção. **COBENGE 2009 – XXXVII Congresso Brasileiro de Ensino de Engenharia.** Recife – PE, 2009.

FAGUNDES, C. V.; LUCE, M. B.; ESPINAR, S. R. O desempenho acadêmico como indicador de qualidade da transição Ensino Médio-Educação Superior. **Ensaio: aval. pol. públ. Educ.,** Rio de Janeiro, v.22, n. 84, p. 635-670, jul./set. 2014.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From Data Mining to Knowledge Discovery: an overview. **Advances in Knowledge Discovery and Data Mining,** AAAI Press / The MIT Press, MIT, Cambridge, Massachusetts, and London, England, p.1-34, 1996.

GOTTARDO, E.; KAESTNER, C.; NORONHA, R. V. **Aplicação de Técnicas de Mineração de Dados para Estimativa de Desempenho Acadêmico de Estudantes em um AVA Utilizando Dados com Classes Desbalanceadas.** ICBL2013 – International Conference on Interactive Computer aided Blended Learning, p. 24-28, Florianópolis-SC, Brasil, 2013.

HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H. **The weka data mining software: an update.** SIGKDD Explor. Newsl., ACM, New York, NY, USA, v. 11, n. 1, p. 10–18, nov 2009. ISSN 1931-0145.

LODER, L. L., NAKAO, O. S. **Evasão e Retenção em Cursos de Engenharia.** Sessão Dirigida. XXXIX Congresso Brasileiro de Educação em Engenharia (COBENGE 2011). Blumenau – SC, 2011.

MARQUES, A. F. **Aplicação de clusterização de dados na base de dados do saneamento ecológico e econômico de Minas Gerais. 88p. Monografia (Bacharelado em Sistemas da Informação) – Universidade Federal de Lavras, Lavras/MG, 2014.**

SCHULTZ, M. R. O., **Metodologias para Ensino de Lógica de Programação de Computadores**. Monografia de Especialização (Ciência da Computação). Universidade Federal de Santa Catarina (UFSC). 69p. Florianópolis, SC, Brasil. 2003.

SILVA, R. F. **Fatores que Influenciam o Desempenho Acadêmico**. Dissertação de Mestrado (Economia). Instituto de Ensino e Pesquisa (Insper). 42p. São Paulo, SP, Brasil, 2013.

SILVA FILHO, R. L. L.; MONTEJUNAS, P. R., HIPÓLITO, O.; LOBO, M. B. C. M. **A evasão no ensino superior brasileiro**. Caderno de Pesquisa, v. 37, n.132, p. 641-659. São Paulo/SP. 2007.

SOARES, I. S. **UFRJ – Escola Politécnica – Vestibular 1993-2009 – Revisão Histórica – Vagas, Evasão e Retenção**. In: COBENGE 2009 – XXXVII Congresso Brasileiro de Educação em Engenharia. Recife, PE, 2009.



### Vídeo da apresentação

Título: Análise da relação entre perfil e desempenho acadêmico dos alunos matriculados na disciplina de Introdução à Programação utilizando algoritmos de classificação.



Disponível em: [http://dadosabertos.info/enhanced\\_publications/idt/video.php?id=20](http://dadosabertos.info/enhanced_publications/idt/video.php?id=20)

## **Transcrição da apresentação**

Este ano será realizado entre os dias 27 e 29 de novembro na cidade de João Pessoa na Paraíba no auditório de superintendência de tecnologia da informação da Universidade Federal da Paraíba.

O título do trabalho submetido é “Análise da relação entre perfil e desempenho acadêmico dos alunos matriculados na Disciplina de Introdução à Programação utilizando algoritmos de classificação desempenho acadêmico dos alunos matriculados”.

Os atores envolvidos são eu Roberto Silva Oliveira, Douglas Valentim de Almeida Cardins, Wendell Juler Pereira de Lima, Thereza Patrícia Pereira Padilha, Vanessa Farias Dantase.

Uma preocupação do ensino superior do mundo é o índice de evasão são maiores nos primeiros períodos dos cursos. Nos casos dos cursos de computação e engenharia uma preocupação é a falta de atenção nas disciplinas de programação.

Nos últimos anos pesquisadores têm concentrado os estudos na bolsa de encontrar respostas para esses problemas analisando dados acadêmicos sociais e econômicos. O aumento da disponibilidade dos dados dos alunos no interesse dos pesquisadores na área de mineração de dados informática na educação faz surgir uma nova área de pesquisa conhecida como mineração de dados educacionais, visa explorar dados do ambiente educacional para descobrir padrões, isto é conhecimento oculto escolhido nos dá para contribuir no entendimento do processo de aprendizagem e assim acelerar a tomada de decisão sobre novas práticas pedagógicas.

Este artigo tem como objetivo geral investigar uma relação que o perfil social cultural acadêmico e econômico dos alunos em seu desempenho acadêmico na disciplina de introdução a programação a fim de minimizar os altos índices de evasão de reprovação na disciplina.

No meio do semestre letivo de 2017.1 foi aplicado formulário web para 42 alunos matriculados na disciplina de IP com 19 perguntas objetivas envolvendo aspectos pessoais econômicos e acadêmicos todas as perguntas do formulário fizeram parte do conjunto de dados obtidos como atributo ao fim do semestre foi disponibilizado a situação final dos alunos aprovados e reprovado que foi adicionado ao conjunto de dados como atributo clássico hotton.

Dos 42 registros, 34 eram da classe aprovado, 08 reprovado, em seguida os dados foram transformados e disponibilizados em formato .ARFF formato padrão da entrada para ferramenta WEKA.

Por fim os dados foram aplicados aos algoritmos de classificação OneR, PART e JRip, disponíveis na ferramenta WEKA e tem como sair do conjunto de regras resultados no OneR neste caso somente um atributo gasto no transporte foi usado para separar a classe aprovados e reprovados, onde classificou 33 instâncias aproximadamente, 78 pontos 6% e 9 incorretamente 21.4%. O alto gasto com transporte é uma possibilidade de reprovar na disciplina.

Resultados com algoritmos PART classificou 36 instâncias corretamente aproximadamente 85.7% e seis incorretamente 14,3%.

As regras geradas pelo algoritmo PART falavam que os alunos que não trabalha que tem idade menor de 18 anos tem grande possibilidade de ser aprovado, os que já tinha conhecimento prévio do curso tem maior possibilidade de ser aprovado, os que não trabalham e têm conhecimento prévio do curso tem possibilidade de ser aprovado.

Resultado dos algoritmos JRip classificou 35 instâncias corretamente aproximadamente 83,3% e 7 incorretamente 16,7%.

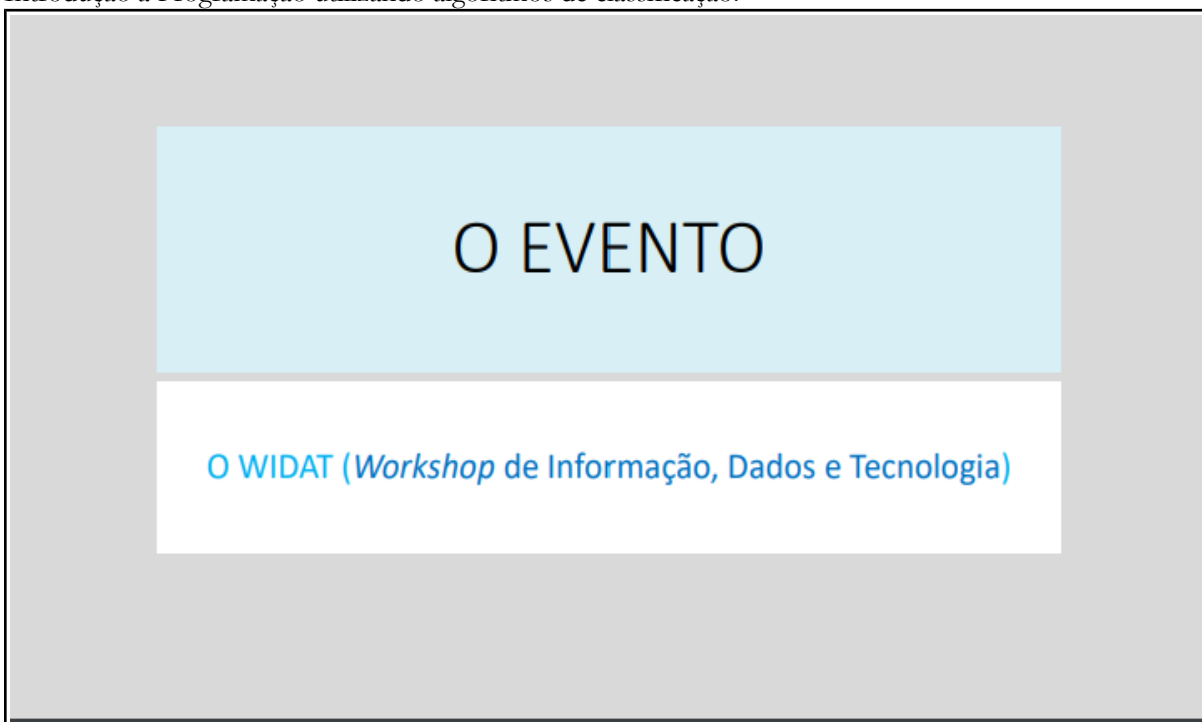
A regra gerada pelo algoritmo JRip falava que os alunos que não têm conhecimento prévio de computação e tem idade maior que 19 tem maior possibilidade de ser reprovado na disciplina de introdução à programação.

Considerações finais o crescente índice de reprovações na educação na disciplina de introdução a programação é alarmante e cabíveis de medidas que visem o seu crescimento.

Logo foi primordial a captação de dados e averiguação e o estudo com a aplicação de mineração de dados apresentou-se como alternativa para a obtenção de informações pertinentes embora conjunto de dados aplicado tenha sido restrito principalmente no que diz respeito à quantidade registro dos alunos reprovados, foi possível ter um modelo do perfil de alunos que perdem a disciplina de IP mesmo com esses empecilhos na descrição do vídeo vocês podem entrar em contato conosco pois deixaremos nossos e-mails.

## Slides da apresentação

Título: Análise da relação entre perfil e desempenho acadêmico dos alunos matriculados na disciplina de Introdução à Programação utilizando algoritmos de classificação.



Disponível em: [http://dadosabertos.info/enhanced\\_publications/idt/presentation.php?id=20](http://dadosabertos.info/enhanced_publications/idt/presentation.php?id=20)