

**XV SIMPÓSIO DOS PROGRAMAS DE MESTRADO PROFISSIONAL  
UNIDADE DE PÓS-GRADUAÇÃO, EXTENSÃO E PESQUISA**



*Desafios de uma sociedade  
digital nos Sistemas Produtivos  
e na Educação*



**Estudo de Caso Sobre Disciplinas com Maior Reprovação no 1º  
Ano do Curso Integrado de Informática do IFF Campus Quissamã**

Orpheu de Souza AYRES<sup>1</sup>, Aline Estaneck Rangel Pessanha dos Santos<sup>2</sup>,  
Geórgia Regina Rodrigues Gomes<sup>3</sup>, Milton Erthal Júnior<sup>4</sup>

**Resumo:** Este trabalho tem por objetivo elencar as disciplinas que mais reprovam os alunos no primeiro ano do curso técnico de informática do IFF. Os dados foram levantados a partir do sistema acadêmico e compostos no período entre 2013 e 2017, estes, constituem a base de dados com informações sobre a situação pessoal e acadêmica dos alunos. Utilizou-se técnica de mineração KDD sobre os dados e obteve-se como resultados informações significativas quanto às disciplinas que mais reprovam.

**Palavras-chave:** Mineração de Dados, Reprovação no Instituto Federal, Índices de Evasão e Reprovação no 1º ano.

**Abstract:** This work aims to list the subjects that most fail students in the first year of the IFF technical course. The data were collected from the academic system and composed in the period between 2013 and 2017, these constitute the database with information about the students' personal and academic situation. KDD mining technique was used on the data and significant results were obtained as to the disciplines that most disapprove.

**Keywords:** Data Mining, Failure at the Federal Institute, Evasion and Failure Rates in the 1st year.

---

<sup>1</sup>Instituto Federal Fluminense, Campus Quissamã – orpheu.ayres@iff.edu.br

<sup>2</sup>Instituto Federal Fluminense, Campus Quissamã - aestaneck@iff.edu.br

<sup>3</sup>Universidade Federal Fluminense, INFES - georgiargomes@gmail.com

<sup>4</sup>Instituto Federal Fluminense, Campus Guarus - miltonerthal@hotmail.com

## 1 – Introdução

O IFF *campus* Quissamã teve início como núcleo avançado em 2006, com aulas do curso técnico em eletromecânica, ministradas em uma escola municipal, a partir de 2013, já em sede própria, o curso Integrado em Informática teve início, com duração prevista para três anos de formação. Foi observado, entre os anos de 2013 e 2017, que os índices de evasão e reprovação no 1º ano desses cursos foram altos, tendo somados evadidos, cancelados ou reprovados em média 62,91%. Fato nada satisfatório em relação ao Plano de Desenvolvimento Institucional (PDI)<sup>5</sup>.

De acordo com o Sistema Nacional de Informações da Educação Profissional e Tecnológica (SISTEC)<sup>6</sup>, o *campus* Quissamã conta atualmente com um quantitativo de cerca de 700 estudantes entre os cursos Concomitante, Subsequente, Integrado, EAD e FICs. Recebe estudantes oriundos dos municípios de Quissamã, Carapebus, Conceição de Macabu, Campos dos Goytacazes, e Macaé. Eventualmente de outros. São Municípios com IDEB muito parecido e com algumas diferenças de IDH.

A tabela 1 apresenta as taxas de reprovação, evasão e cancelamento somados, para os primeiros anos do curso de informática, entre os anos de 2013 e 2017, com valores muito significativos, de acordo com registros do Sistema Acadêmico (QA).

**Tabela 1-** Taxas de rendimento escolar do 1º ano do curso técnico integrado de informática

Ano	Reprovação + evasão + cancelamentos
2013	54,29%
2014	65,22%
2015	55,56%
2016	65,00%
2017	74,47%

**Fonte:** Registro Acadêmico IFF *campus* Quissamã, 2018.

Nos indicadores do INEP entre os anos de 2014 e 2018 na tabela 2, pode-se perceber uma grande diferença percentual entre os resultados do IFF de Quissamã das demais instituições de ensino federal (Qedu, 2020) <sup>7</sup>.

**Tabela 2 -** Taxas de rendimento escolar do 1º ano do ensino médio

	Reprovações	Abandonos	Aprovações
2017	16,60%	2,50%	80,90%
2016	17,10%	2,60%	80,30%
2015	18,50%	3,50%	78,00%
2014	19,70%	2,80%	77,50%

**Fonte:** Censo Escolar 2017, Inep. Classificação não oficial.

<sup>5</sup>Plano de Desenvolvimento Institucional (PDI) - <http://portal1.iff.edu.br/desenvolvimento-institucional/arquivos/pdi-2018-2022-com-resolucao-menor.pdf>

<sup>6</sup>SISTEC - <https://sistec.mec.gov.br/consultapublicaunidadeensino>

<sup>7</sup>Qedu.org.br – Disponível em: <https://www.qedu.org.br/brasil/taxas-rendimento/rede-federal/rural-e-urbana?year=2017> Acessado em: 10/08/2020.

Quando comparadas com as taxas do ensino médio geral entre todas as escolas: particulares, municipais, estaduais e federais, a diferença torna-se ainda maior, conforme tabela 3.

**Tabela 3** - Taxas de rendimento no ensino médio – Brasil 2014 – 2018

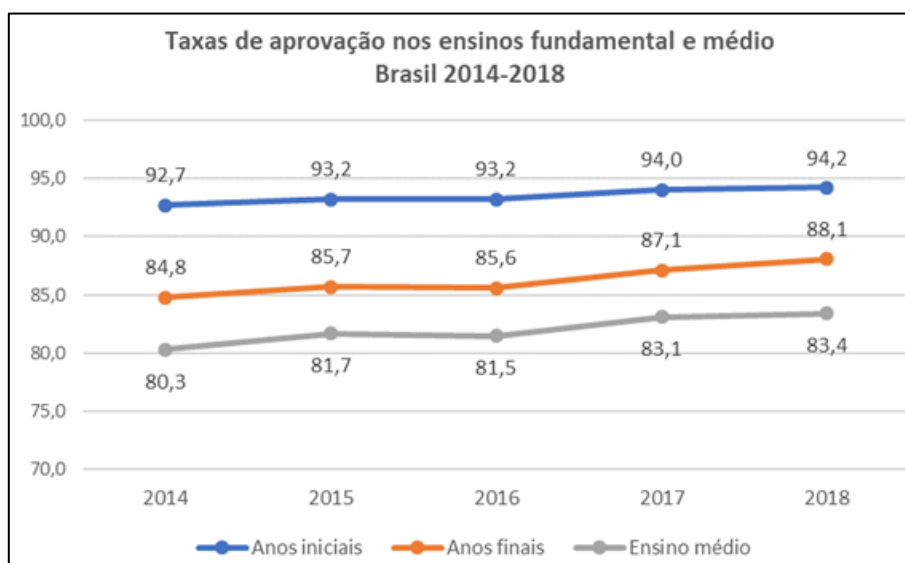
	Reprovações	Abandonos	Aprovações
2017	16,60%	2,50%	80,90%
2016	17,10%	2,60%	80,30%
2015	18,50%	3,50%	78,00%
2014	19,70%	2,80%	77,50%

**Fonte:** MEC/Inep/Deed, 2020

As taxas de rendimento informadas pelas escolas revelam a situação dos alunos entre os anos 2014 e 2017, quanto a aprovação, reprovação ou evasão. Essas taxas de rendimento, que são indicadores no cálculo do Índice de Desenvolvimento da Educação Básica (Ideb), combinados as taxas de rendimento do Censo Escolar com o desempenho dos estudantes no Sistema de Avaliação da Educação Básica (Saeb) contribuem para a formação do censo escolar, considerado o mais importante levantamento estatístico educacional brasileiro. (MEC, 2020).

Pode-se observar uma tendência de melhora nos resultados ao visualizar os dados entre 2014 e 2018, em todos os cenários entre ensino fundamental e ensino médio, conforme figura 1, contrastando ainda mais com os resultados obtidos pela instituição neste período.

**Figura 1** - Taxas de aprovação nos ensinos fundamental e médio Brasil 2014 - 2018



**Fonte:** MEC/Inep/Deed, 2020

A reprovação pode gerar sujeitos com baixa autoestima, com bloqueios e com suas capacidades e habilidades inibidas pela aceitação do fracasso. Impactando assim seu futuro como profissional e cidadão. Então, compreende-se que a reprovação escolar, não é um assunto pontual, simples, e de ação individual, o campus Quissamã deve ter um olhar atento quanto a comunidade em que está inserido pois as consequências do fracasso escolar na vida de um indivíduo atingem a todos (SILVA et al. 2016).

A reprovação escolar na educação brasileira é um problema recorrente. Tanto discentes, quanto docentes atribuem aos alunos a maior parte da responsabilidade pela reprovação escolar na instituição federal, sem descartar a necessidade de melhorar o processo de acompanhamento pedagógico e de fazê-lo mais efetivo (SILVA et al. 2016).

A evasão escolar é um problema de nível mundial, que precisa ser combatido. Desperta interesse de pesquisadores acadêmicos do mundo e dos órgãos governamentais, uma vez que, de acordo com a teoria do capital humano, a educação é essencial ao progresso econômico, para o aumento de competitividade de um país e para a melhoria de bem-estar social (SOUZA et al., 2015).

Os alunos entram nos Institutos Federais por diversas razões, como: a busca de ensino de qualidade, ensino gratuito, status, potencializar os resultados acadêmicos e mais uma variedade de motivos. Os cursos são técnicos e atuam em alguma área como Eletromecânica, Informática, Química, Farmácia e outras. Suas escolhas dentro deste espectro limitado ofertado, frequentemente se torna um fator de desestímulo, pois, tem afinidade distante ou não correspondente às suas características individuais.

Com o objetivo de reduzir as reprovações e evasões, bem como compreender melhor o que leva estes estudantes à reprovação, é que se propõe um estudo sobre os primeiros anos do curso Integrado em Informática, sobre os índices de reprovação de cada disciplina do primeiro ano. Este estudo pretende, utilizando a mineração de dados, através do software Weka, apresentar tais índices de reprovação. Ranquear as disciplinas que mais reprovam, criando indicadores para que esta discussão fomente, dentro do Instituto a propositura de estratégias para minimizar essas reprovações.

## 2 – Metodologia

Para realizar tal intento, faz-se necessário conhecer melhor, tanto o perfil do estudante do primeiro ano do curso de informática como as características das disciplinas cursadas neste primeiro ano. Entretanto, esbarra-se em informações incompletas no próprio Registro Acadêmico. Para que esses dados se transformem em conhecimento foi utilizada a Descoberta de Conhecimento em Bancos de Dados, *Knowledge Discovery in Databases* (KDD).

O processo passou por um fluxo operacional compreendido por: seleção, pré-processamento e limpeza, transformação, mineração de dados (*data mining*) e interpretação. Utilizando-se por referencial teórico USAMA FAYYAD. Definição: “o processo, não trivial, de extração de informações implícitas, previamente desconhecidas e potencialmente úteis, a partir dos dados armazenados em um banco de dados”. (FAYYAD et.al, 1996).

**Figura 2** - Processo de KDD (FAYYAD, apud GONÇALVES, 2001)



Figura 1 Processo de KDD (FAYYAD, apud GONÇALVES, 2001)

Fonte: <http://www.ufrgs.br/gianti/files/orientacao/mestrado/defesa/pdf/>

## 2.1 – Seleção dos dados

Nesta primeira fase do processo, escolhemos um conjunto de dados para trabalhar, com todas as possíveis variáveis, as quais também chamamos de características ou atributos e as observações ou registros que farão parte da análise.

O processo de seleção utilizado neste estudo foi viabilizado por recebimento dos dados via planilhas eletrônicas. Foram dispensados campos para extração de dados como: nome do aluno, número de telefone, email, número de matrícula, número de turma, por serem de natureza individual, não geral, embora utilizados para conferência e validação dos dados na etapa de pré-processamento e limpeza.

Este trabalho foi realizado mais de uma vez, visto que as tentativas iniciais não geraram padrões significativos sobre os dados, nem consistência percentual, que pudessem correlacionar os campos com evasão e ou retenção. Nas inserções finais, os dados foram restringidos aos campos: Turma, Matrícula, Nome do Aluno, código de cada disciplina com respectivos campos de Nota, Situação, Frequência e Carga Horária, Média do Aluno, Cidade, Escola de Origem, Per. Letivo Inicial, Percentual Frequência, Renda Familiar, Renda Familiar Per Capita, Renda Familiar Per Capita INEP, Renda Familiar Per Capita SIG, Sexo, Situação Matrícula, Situação Período e Tipo Forma Ingresso no Período.

## 2.2 – Pré-processamento e limpeza dos dados

Nesta etapa eliminou-se dados singulares como nome e matrícula, dados repetidos e os inconsistentes, corrigiu-se dados incompletos e avaliou-se divergências entre os dados.

Após ordenação dos dados e identificação dos campos sem informação, ou com maioria das informações sem preenchimento, eliminou-se os campos: Matrícula, Nome do Aluno, Frequência, Renda Familiar, Renda Familiar Per Capita, Renda Familiar Per Capita INEP (contém as informações necessárias para a construção de faixas de renda Familiar Per Capita) e O campo Renda Familiar Per Capita SIG.

A informação relativa à carga horária foi retirada uma vez que se tratava de um dado único e específico de cada disciplina.

## 2.3 – Transformação dos dados

Utilizou-se o software Weka, aplicativo que utiliza uma coleção de algoritmos de aprendizado de máquina (*machine learning*), bem como, ferramentas de preparação de dados, classificação, agrupamento e outras diversas ferramentas para a mineração de dados.

Para tanto os dados precisaram sofrer algumas alterações essenciais. O Weka não lê adequadamente os dados se estes tiverem acentos gráficos, bem como “ç” ou caracteres especiais. Os dados numéricos foram agrupados em faixas, para evitar ter uma quantidade de informações tão grande que não pudesse ser identificada como um padrão. As notas exigiram um campo específico que as associasse à disciplina e foram divididas da seguinte forma:

Valor para classificação da faixa de nota da disciplina e para a faixa de Média do Aluno.

**Quadro 1** - Classificação da faixa de nota da disciplina e para a faixa de Média do Aluno

Conteúdo da Célula Anterior	Valor a ser exibido no novo campo
“Campo Vazio”	Sem Preenchimento
Menor que 4	Muito Baixa
Menor que 6	Insuficiente
Menor que 8	Média
Maior ou igual a 8	Alta

**Fonte:** próprio autor

A classificação neste intervalo foi determinada de acordo com estrutura avaliativa do IFF. No IFF quaisquer notas bimestrais que ficarem abaixo de 4, fazem com que os alunos tenham que fazer provas de recuperação semestral, mesmo que tenham obtido média maior que 6 (seis) que é a média de aprovação exigida.

Quaisquer dados numéricos não podem apresentar a vírgula no número, uma vez que o formato CSV, necessário ao aplicativo Weka, trata todas as vírgulas como separador de campo, o que geraria quantidades de colunas diferentes por linha de registro.

Após salvar a planilha no formato CSV, o arquivo fora aberto no aplicativo bloco de notas, converteu-se primeiro todas as vírgulas (“,”) em pontos (“.”), e em seguida todos os pontos e vírgulas (“;”) em vírgulas(“,”) preparando o arquivo para ser adequadamente utilizado pelo Weka.

O Percentual de Frequência fora removido uma vez que seus resultados abaixo de 25%, valor que significaria a reprovação, representam menos de 10% dos registros, foram visivelmente associados à evasão ou cancelamento, com apenas 4 registros matriculados. Dados estes obtidos com uma simples ordenação do campo Percentual de Frequência.

## 2.4 Mineração de dados (*Data Mining*)

A etapa de Mineração de Dados (*Data Mining*) é entendida por teóricos como FAYYAD como sendo a etapa mais importante do processo de busca de

conhecimento em base de dados. Para BERRY e LINOFF (1997) apud PRASS “*data mining* é a exploração e análise, de forma automática ou semiautomática, de grandes bases de dados com objetivo de descobrir padrões e regras”.

O conhecimento produzido pelo trabalho de *data mining* se dá por meio das tarefas: análise de regras de associação, classificação e predição, análise de padrões sequenciais, análise de *clusters*, análise de *outliers* (Elmasri e Navathe, 2002).

Além de tudo é muito importante saber diferenciar tarefa de técnica de mineração de dados. Técnica de mineração é o estabelecimento de métodos para como encontrar os padrões que serão válidos nessa pesquisa. A tarefa reside em descrever aquilo que se deseja descobrir nos dados (AMO, 2003).

O Esquema (*Scheme*) utilizado em *CLUSTER* com 7 (sete) *clusters*:  
weka.clusterers.SimpleKMeans -init 0 -max-candidates 1000 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 7 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10, gerou o seguinte resultado:

- Disciplinas que demonstraram notas predominantes abaixo de 4,0 : CINFO13nota; CINFO14nota; CINFO11nota; CINFO6nota
- (CINFO.13 - Algoritmo e Estrutura de Dados; CINFO.6 - Matemática I; CINFO.14 - Sistemas Digitais; CINFO.11 - Sociologia I)
- Em 111 registros, aproximadamente a metade dos alunos pesquisados, a renda familiar per capita predominante foi de 0,5 e entre 1 e 1,5 SM;
- São predominantes: a Renda Familiar Per Capita inferior a 0,5 SM, a escola de origem: escola pública municipal, o Sexo Feminino.
- Utilizando 10 ou mais *clusters* encontramos os resultados abaixo:
- É observado que as mesmas disciplinas que aparecem com médias predominantemente muito baixas em *clusters*, aparecem com médias altas em escolas de origem particular.
- 48 registros com Média Baixa e várias reprovações, para o sexo Feminino com renda familiar per capita inferior a 0,5 SM se destaca no *cluster* 1 de 10 *clusters*

Quanto à Classificação, utilizou-se tanto o algoritmo J48, quanto o *NavieBayes*, em ambos os casos todos os campos apresentaram percentual de erro maior que o de acertos, com exceção do campo Média. O *NavieBayes* em todos os casos, apresentou um percentual de assertividade maior que o J48 e por isso foi eleito para os resultados obtidos sobre a pesquisa.

O Esquema (*Scheme*) utilizado foi: weka.classifiers.bayes.NaiveBayes

*Relation*: planilhas weka janeiro-weka.filters.unsupervised.attribute.Remove -R1- weka.filters.unsupervised.attribute.Remove R3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45 - weka.filters.unsupervised.attribute.Remove-R33, resultando em: O único cabeçalho de classificação que retornou um percentual elevado de assertividade foi a Média.

Pode-se observar que: a renovação por reprovação é de aproximadamente 20%, que as notas insuficientes somadas às notas muito baixas são 2,25 vezes maiores que as notas altas, que as notas altas são apenas 8% das notas e que as notas médias representam aproximadamente 55% das notas.

### 3 – Resultados e Discussões

A identificação da disciplina de algoritmos como uma das maiores responsáveis por retenção e evasão, um problema clássico a nível mundial, foi ratificada a partir da aplicação de mineração de dados fornecidos pelo Registro Acadêmico no período entre 2013 e 2018 do Instituto Federal Fluminense de Quissamã para o curso técnico em informática.

Como já era esperado, a melhor condição socioeconômica e a escola de origem particular demonstraram favorecer os resultados dos alunos, entretanto, uma vez passado o primeiro ano, os alunos, independentemente da escola de origem ou de condição socioeconômica, tiveram índice de reprovação semelhante.

Sobre as disciplinas responsáveis pela maioria das reprovações, duas são de disciplinas técnicas do curso de informática e duas são da área propedêutica, a saber: Algoritmo e Estrutura de Dados, Sistemas Digitais, Matemática I e Sociologia I.

De acordo com a matriz curricular do projeto Pedagógico do curso em questão, a disciplina de Algoritmo e Estrutura de Dados possui o dobro de carga horária da maioria das disciplinas, 160h e é responsável pela base de conhecimentos necessários às disciplinas técnicas de programação nos anos posteriores e razoável que seja cobrada com mais rigor uma vez que dela resulta a formação do egresso quanto ao perfil técnico. Da mesma forma a disciplina de Sistemas Digitais, base de conhecimento para diversas disciplinas técnicas do mesmo ano e dos anos seguintes.

Dentre os padrões apresentados com diferentes configurações na aplicação da mineração de dados sobre os dados do sistema acadêmico, destacam-se:

**Tabela 4** - Disciplinas que demonstraram notas predominantes abaixo de 4,0

Ordem de maior incidência	Código utilizado no Weka	Nome da Disciplina
1º	CINFO13nota	Algoritmo e Estrutura de Dados
2º	CINFO14nota	Sistemas Digitais
3º	CINFO11nota	Sociologia I
4º	CINFO.6	Matemática I

**Fonte:** do próprio autor

Características predominantes observadas

- Aproximadamente a metade dos alunos pesquisados, a renda familiar per capita predominante foi de 0,5 e entre 1 e 1,5 SM;
- Renda Familiar Per Capita inferior a 0,5 SM,
- Escola de origem: escola pública municipal,
- Sexo Feminino.

Percebe-se que o enfrentamento a evasão do aluno depende de múltiplos fatores, do suporte pedagógico, social e cultural tanto para o indivíduo quanto para as instituições de ensino. Há a necessidade de uma nova postura ativa tanto de gestores, quanto de pesquisadores, educadores e da comunidade em geral no combate ao problema de âmbito geral.



#### 4 – Conclusão

O ranqueamento foi realizado e foram encontradas as disciplinas responsáveis pela maioria das reprovações. Os dados minerados revelaram além de duas disciplinas técnicas que necessitam de maior orientação pedagógica, também duas disciplinas da propedêutica a serem tratadas com maior atenção. Os padrões encontrados foram bastante significativos em termos comparativos tanto a outras disciplinas, quanto a outras instituições. Os resultados encontrados foram apresentados à direção de ensino e à coordenação do curso de informática para a devida discussão e tratamento, atendendo ao objetivo da pesquisa para nortear os trabalhos de redução de evasão e retenção no IFF de Quissamã.

#### Referências

BRASIL, Sistema Nacional de Informações da Educação profissional e Tecnológica. SISTEC. Disponível em: <<http://sistec.mec.gov.br/login/login>> Acesso em: 18 jan 2019

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. *From data mining to knowledge discovery: An overview*. In: *Advances in Knowledge Discovery and Data Mining*, AAAI Press/The MIT Press, England, 1996, p.1-34

GONÇALVES, Loren P.F. Avaliação de Ferramentas de Mineração de dados como fontes de dados relevantes para a tomada de decisão: aplicação na Rede União de supermercados São Leopoldo. Dissertação de Mestrado, UFRGS, 2001. Disponível em: <[http://www.ufrgs.br/gianti/files/orientacao/mestrado/defesa/pdf/24\\_dissertacao\\_goncalves.pdf](http://www.ufrgs.br/gianti/files/orientacao/mestrado/defesa/pdf/24_dissertacao_goncalves.pdf)> Acesso em: 18 jan

MEC - Ministério da Educação - Disponível em: [http://portal.inep.gov.br/artigo/-/asset\\_publisher/B4AQV9zFY7Bv/content/inep-divulga-taxas-de-rendimento-escolar-numeros-mostram-tendencia-historica-de-melhora/21206](http://portal.inep.gov.br/artigo/-/asset_publisher/B4AQV9zFY7Bv/content/inep-divulga-taxas-de-rendimento-escolar-numeros-mostram-tendencia-historica-de-melhora/21206). Acessado em: 10/08/2020.

PRASS, Fernando Sarturi. KDD – UMA VISAL GERAL DO PROCESSO. Disponível em: <<https://docplayer.com.br/7949839-Kdd-uma-visal-geral-do-processo.html>> Acesso em: 17 Janeiro 2019.

Qedu – Qedu.org.br – 2020

SILVA, TOC, SANTOS, A, CUNHA, PC, MAGRONE, E – Reprovação Escolar No Ensino Médio Integrado à Educação Do Instituto Federal Do Paraná (IFPR) – Campus Ivaiporã. 2016.

SOUSA, JNM, TABOSA, FJ, SIMONASSI, A, CASTELAR, PUC – Principais Fatores Que Impactam Na Reprovação E Evasão Dos Alunos Dos Cursos Tecnológicos No Instituto Federal De Ciência E Tecnologia Do Ceará – 2015

SOUZA, CT, SILVA, C, GESSINGER, RM - Congressos CLABES, 2012 – revistas.utp.ac.pa - Um estudo sobre evasão no ensino superior do Brasil nos últimos dez anos - 2012