

LEVANTAMENTO, COLETA, ORGANIZAÇÃO, PROCESSAMENTO E ANÁLISE DE INFORMAÇÕES SOBRE ESTUDANTES EVADIDOS DOS CURSOS DO IFSP PARA FINS DE GESTÃO EDUCACIONAL ¹

Guilherme Dinhani CONCEIÇÃO²

Graduando em Tecnologia em Análise e Desenvolvimento de Sistemas
IFSP/Câmpus Salto

Claudio Haruo YAMAMOTO³

Doutor em Ciências de Computação/USP
Docente de Ciências de Computação
IFSP/Câmpus Salto

RESUMO

Atualmente, o acompanhamento de alunos evadidos é uma ferramenta essencial na construção de indicadores para as instituições de ensino. Os indicadores gerados a partir desse acompanhamento podem ser utilizados para avaliar a qualidade dos cursos oferecidos e aprimorá-los. Entretanto, esta atividade não é fácil, pois exige a construção/manutenção de portais para cadastro de alunos evadidos e a realização de pesquisas por meio de formulários. O uso destes recursos, de modo isolado, tem um efeito limitado porque precisa que os evadidos forneçam dados. Desta forma, este projeto visa criar, com a linguagem de programação R e utilizando o ambiente de desenvolvimento integrado RStudio, um sistema de coleta, organização, processamento e análise de informações públicas da rede social profissional LinkedIn sobre estudantes evadidos do IFSP – Câmpus Salto para fins de gestão educacional. O sistema foi utilizado para coletar dados dos cursos de Análise e Desenvolvimento de Sistemas (ADS) e Gestão da Produção Industrial (GPI) do Câmpus Salto.

Palavras-chave: Evasão escolar; *dashboard*; LinkedIn; linguagem de programação R.

¹ Trabalho resultante de Projeto de Iniciação Científica. Orientador Prof. Dr. Claudio Haruo Yamamoto.

² Endereço eletrônico: guilhermedinhani@hotmail.com

³ Endereço eletrônico: haruo@ifsp.edu.br

Introdução

É de grande importância que as instituições de ensino possuam dados sobre seus estudantes, incluindo os que evadiram de seus cursos. Estudos como o de Grossi, Gomes e Nunes (2013), que identificou o perfil dos estudantes evadidos a partir de dados institucionais, reforçam essa importância. A partir de dados de estudantes, incluindo dos que evadem, é possível obter informações que podem auxiliar na gestão educacional da instituição, por exemplo, formulando hipóteses que podem relacionar motivos pelos quais a evasão pode estar acontecendo. Estas informações podem ser úteis para diminuir os índices de evasão destes cursos.

Entretanto, estes dados, normalmente, não estão organizados, o que dificulta o seu uso pela gestão educacional na tomada de decisões. É comum o uso de portais, questionários e redes sociais para acompanhamento de estudantes egressos. Em portais, em geral, os estudantes cadastram informações, que comumente não são atualizadas posteriormente. Em questionários, faz-se a aplicação de formulários que os estudantes respondem. Em geral, como a aplicação dos questionários não é repetida, os dados também não são atualizados posteriormente. No caso das redes sociais, é possível manter contato entre os estudantes e a instituição, de modo que o dinamismo da relação permite que as informações permaneçam atualizadas. Em redes sociais profissionais, é possível obter informações mais atualizadas que nas demais abordagens descritas.

Dentro deste contexto, este trabalho realizou o levantamento, a coleta, a organização, o processamento e a análise de informações sobre estudantes evadidos de cursos do Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP), obtidas a partir de currículos públicos da rede social profissional LinkedIn (2019), com o apoio de ferramentas de Inteligência de Negócios. O objetivo da Inteligência de Negócios é obter conhecimento adequado para tomar decisões sábias sobre dados e informações (XINGSEN *et al*, 2009). Neste estudo, foram considerados os cursos de Tecnologia em Análise e Desenvolvimento de Sistemas (ADS) e de Tecnologia em Gestão da Produção Industrial (GPI), ambos ofertados no Câmpus Salto.

Para apoio a este estudo, foi criado um sistema, escrito em linguagem R no ambiente RStudio. Este sistema é capaz de extrair informações a partir de currículos do LinkedIn, preprocessar os dados extraídos, e criar uma *dashboard* com informações acadêmicas e outra com informações profissionais. O levantamento e a coleta de currículos de estudantes evadidos foram realizados manualmente por duas razões. Primeiramente, foi preciso obter a informação

da situação dos estudantes, e esta não se encontrava disponível em um sistema integrado. Em segundo lugar, foi preciso identificar os perfis dos estudantes evadidos no LinkedIn, já que não há uma correspondência entre cada estudante e o seu currículo na rede social.

Sistema para coleta, processamento e análise

O sistema de coleta, organização, processamento e análise de dados foi construído na linguagem de programação R 3.5.3 (R CORE TEAM, 2019), utilizando pacotes disponibilizados pela comunidade e o ambiente de desenvolvimento integrado RStudio versão 1.2.1335 (RSTUDIO TEAM, 2019) em um Microcomputador Desktop Core i7-2600, 8 GB RAM, 120 GB SSD com sistema operacional Windows 7 Ultimate 64 bits SP 1 com conexão à Internet.

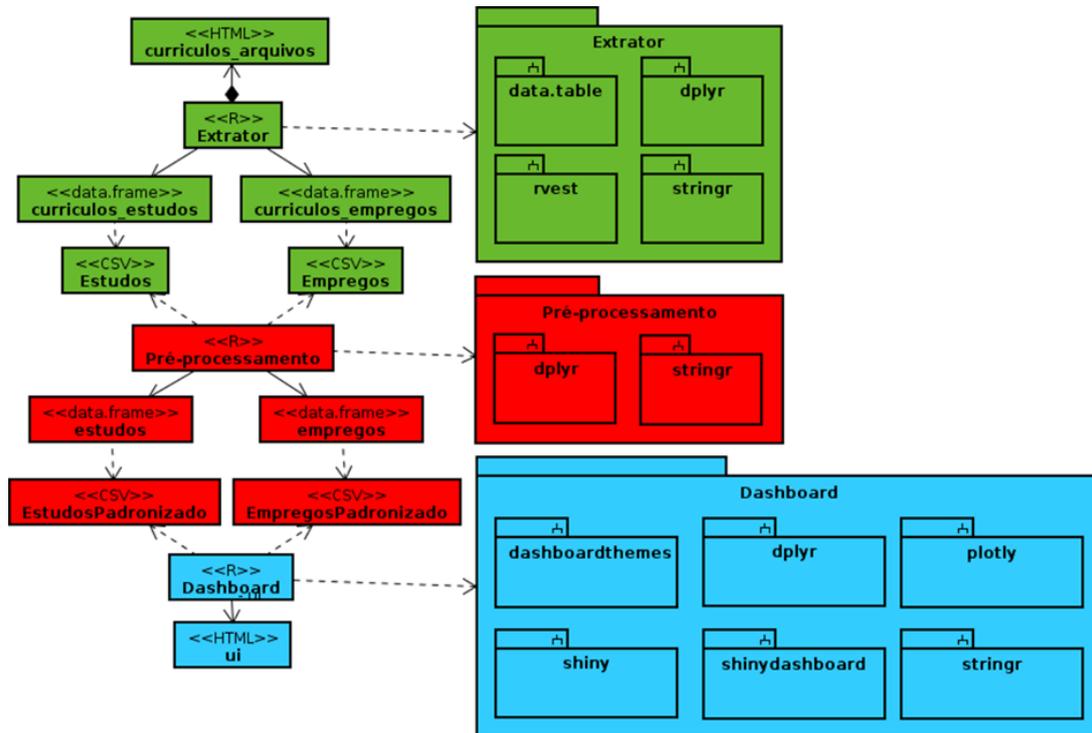
O sistema é composto por três módulos, conforme indicado na arquitetura do sistema (Figura 1): (1) módulo extrator de dados; (2) módulo padronizador de dados; (3) módulo gerador de *dashboards*.

O módulo **extrator de dados** salva em arquivos com a extensão CSV (Comma-Separated Values) os dados selecionados dos currículos baixados manualmente da rede social profissional LinkedIn. São gerados dois arquivos, um para armazenar os dados acadêmicos (Figura 2), e outro para armazenar os dados profissionais (Figura 3).

O módulo **padronizador de dados** lê os dados extraídos no módulo extrator e realiza o pré-processamento dos dados, salvando-os em um segundo arquivo CSV.

No pré-processamento, são realizadas as seguintes alterações: (1) remoção de acentos e caracteres como traço, barra etc.; (2) mudança de palavras para caixa alta; (3) correção ortográfica; (4) agrupamento de palavras semelhantes para instituições, empresas, cargos, áreas de estudo, áreas de emprego. Por exemplo, IFSP e *strings* que iniciam com “Instituto Federal” e terminam com “São Paulo”, são transformadas em IFSP.

Figura 1 – Arquitetura do sistema



Fonte: Autores

Figura 2 – Trecho do arquivo com dados acadêmicos gerado pelo módulo extrator

```

Nome,"Escola","Curso","Area","Data"
[REDACTED],"Instituto Federal de Educação, Ciência e Tecnologia de São Paulo - IFSP","Te
[REDACTED],"Faculdades Atibaia - FAAT",NA,"Engenharia de Produção","2015 – 2018"
[REDACTED],"Instituto Federal de Educação, Ciência e Tecnologia de São Paulo - IFSP",NA,"Gestã
[REDACTED],"Esamc","Master of Business Administration (MBA)","Finanças","2013 – 2014"
[REDACTED],"Instituto Federal de educação ciencia e tecnologia de São paulo","Tecnologia
[REDACTED],"Instituto Federal de Educação, Ciência e Tecnologia de São Paulo - IFSP","Tec
[REDACTED],"Instituto de Ciências Matemáticas e de Computação (ICMC) - USP","Doutorado e
    
```

Fonte: Autores

Figura 3 – Trecho do arquivo com dados profissionais gerado pelo módulo extrator

```
Nome,"Empresa","Cargo","Cidade","Data","Atividade"  
_____, "Diadeis", "Software Engineer Technical Lead", "São Paulo Area, Brazil", "abr de 2018 – o  
_____, "Diadeis", "Java Software Engineer", "São Paulo e Região, Brasil", "jul de 2016 – abr de 20  
_____, "Logus Retail", "Java Software Engineer", "Jundiai, São Paulo", "ago de 2015 – jun de 2016  
_____, "The Gideons International", "Software Developer", "NA", "jun de 2014 – ago de 2015", "NA  
_____, "IBM", "Software Developer Internship", "Hortolândia São Paulo", "jun de 2012 – jan de 2  
_____, "TE Connectivity", "Quality Assurance Analyst", "Bragança Paulista e Região, Brasil", "dez de 201  
_____, "TE Connectivity", "Estagiária em Engenharia de Qualidade Avançada", "Bragança Paulista, São F
```

Fonte: Autores

Por fim, o módulo **gerador de dashboards** utiliza os dados padronizados para gerar as *dashboards*, usando principalmente a biblioteca *shinydashboard* (SHINY, 2019) para gerar dois conjuntos de gráficos, um para dados acadêmicos e outro para dados profissionais.

Resultados obtidos e discussão

Na versão atual do sistema, os dados extraídos do LinkedIn são utilizados para gerar duas dashboards por curso: a primeira com informações acadêmicas do estudante evadido (nome da escola, nome do curso, área do curso, data de início e de término); e a outra com informações profissionais (empregador do estudante, cargo do estudante, cidade onde o empregador se localiza, data de início e término). Na Figura 4, é apresentada a dashboard com informações acadêmicas para o curso de ADS.

Figura 4 – Parte superior da *dashboard* acadêmica do curso de ADS



Fonte: Autores

Na parte profissional, um marcador é gerado para verificar se o empregador atual do estudante é do setor público ou privado, e uma outra coluna é gerada com base em um processamento que fornece qual é a área do atual cargo de trabalho de cada estudante.

Neste estudo, foram utilizados dados dos estudantes evadidos de ADS e GPI dos anos de 2009 a 2019. Dos 332 estudantes de ADS (conforme indicado na Figura 5), 132 perfis (39,76%) foram encontrados no LinkedIn (Figura 6). Para GPI, este percentual foi ligeiramente menor, sendo encontrados 109 perfis de 303 estudantes evadidos (35,97%).

Figura 5 – *ValueBox* com a quantidade total de estudantes evadidos de ADS



Fonte: Autores

Figura 6 – *ValueBox* com a quantidade de perfis encontrados (estudantes evadidos de ADS)



Fonte: Autores

Em relação aos evadidos de ADS, dos 132 perfis encontrados no LinkedIn, 111 possuem dados acadêmicos (33,43%), como visto na Figura 7. Para GPI, dos 109 perfis encontrados no LinkedIn, 18 não possuem dados acadêmicos, ou seja, 91 estudantes evadidos possuem estes dados (30,03%).

Figura 7 – *ValueBox* com a quantidade de perfis encontrados (estudantes evadidos de ADS) com dados acadêmicos

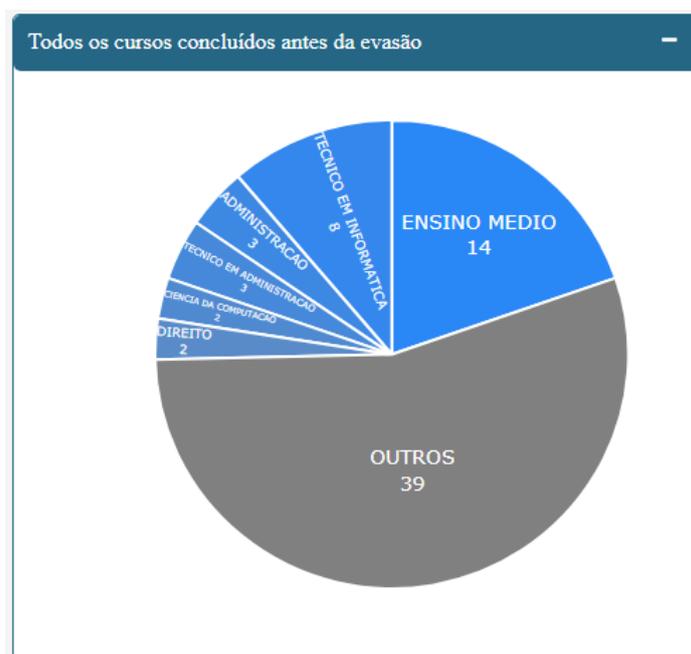


Fonte: Autores

A seguir são apresentados os gráficos das dashboards. O primeiro deles, apresentado na Figura 8, indica os cursos concluídos pelos estudantes de ADS antes de ingressarem no curso,

portanto, antes de evadirem. É possível notar que, para os evadidos de ADS, com exceção do ensino médio, o curso mais comum está relacionado com o eixo tecnológico do curso – Tecnologia da Informação, com 8 estudantes tendo cursado Técnico em Informática previamente. Para GPI, este fato não se aplica, pois apenas na terceira posição aparece um curso do mesmo eixo tecnológico – Processos Industriais – que é o Bacharelado em Engenharia, com apenas três estudantes.

Figura 8 – Cursos concluídos pelos estudantes de ADS antes da evasão

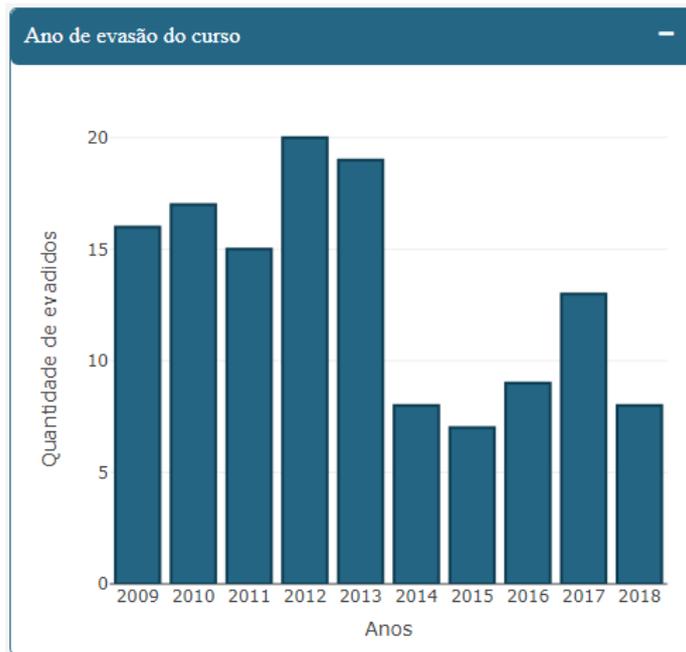


Fonte: Autores

Na Figura 9, é apresentado o ano de ingresso dos estudantes evadidos. É possível perceber que a turma de ADS que ingressou em 2012 foi aquela que teve a maior quantidade de estudantes evadidos (20 estudantes). A partir de 2014, a quantidade de estudantes evadidos é menor. É importante destacar que os estudantes que ingressaram a partir de 2014 podem estar ainda matriculados. Para GPI, a turma com a maior quantidade de evadidos é a de 2013, com 18 estudantes evadidos.

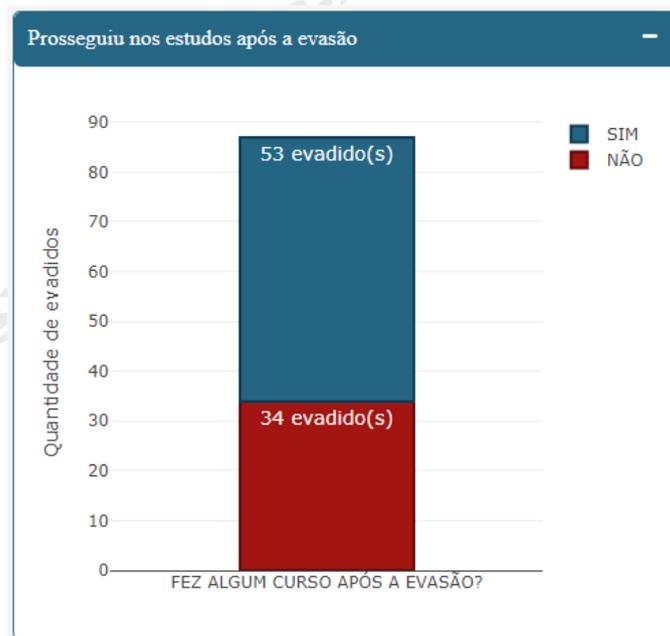
Na Figura 10 é indicado se o estudante evadido prosseguiu ou não nos estudos após a evasão de ADS. Como pode ser visto, 53 estudantes (47,75%) fizeram outro curso, 34 não fizeram outro curso e de 24 não há dados sobre os cursos, apenas sobre a instituição de ensino. Entre os evadidos de GPI, 47 estudantes (51,65%) prosseguiram com os estudos, 25 não prosseguiram e de 19 não há dados sobre os cursos, apenas sobre a instituição de ensino.

Figura 9 – Quantidade de estudantes evadidos de ADS por ano



Fonte: Autores

Figura 10 – Quantidade de estudantes evadidos de ADS que prosseguiram nos estudos após a evasão

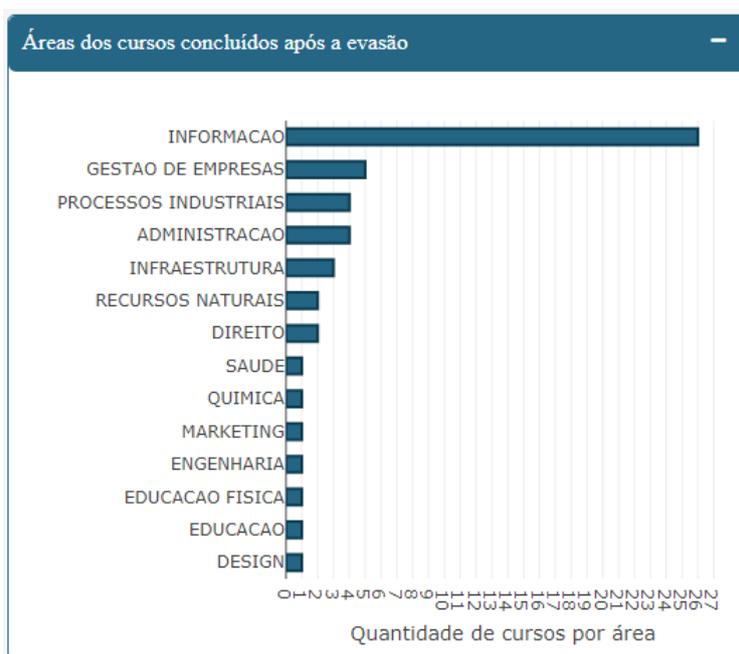


Fonte: Autores

Nas Figuras 11 e 12, são apresentados, respectivamente, as áreas dos cursos e os tipos de cursos concluídos, após a evasão, pelos estudantes evadidos de ADS. A área de Informação

foi a mais comum para os estudantes evadidos de ADS (Figura 11), sendo o próprio curso – ADS – o mais comum, com 8 dos 53 estudantes evadidos (Figura 12). Para GPI, a área mais comum após a evasão foi de Processos Industriais e o curso mais comum foi Engenharia de Produção, com 6 dos 47 estudantes evadidos.

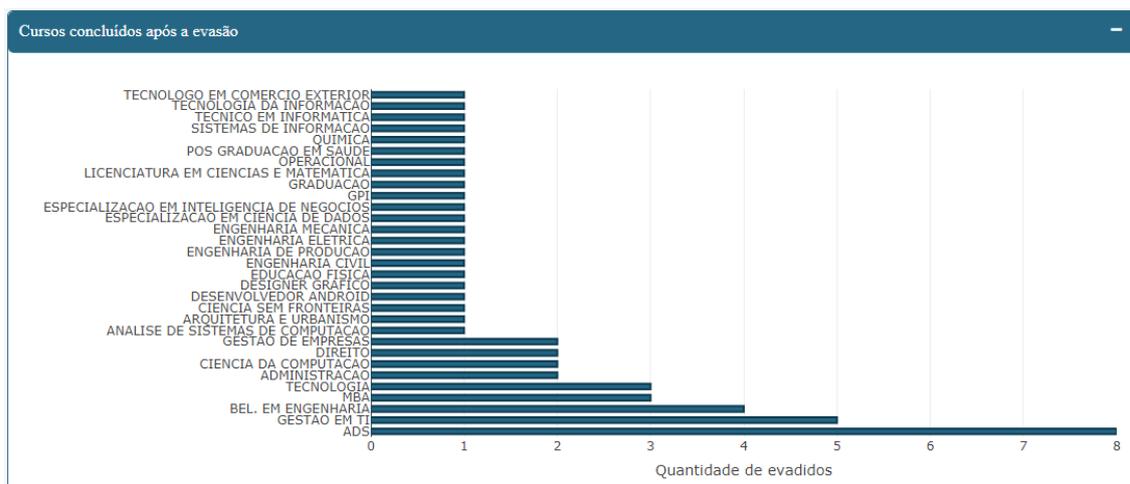
Figura 11 – Área dos cursos concluídos por estudantes de ADS após a evasão



Fonte: Autores

Em relação aos gráficos do grupo de dados profissionais, é possível observar que: dos 332 estudantes evadidos de ADS, somente 110 perfis (33,13%) possuem dados profissionais no LinkedIn (Figura 13), ou seja, dos 132 perfis totais encontrados, 22 não possuem dados profissionais. Para GPI, dos 109 estudantes encontrados, 101 possuem dados profissionais em seu perfil (33,33% do total de 303 estudantes evadidos). Estes estudantes evadidos de ADS estão empregados atualmente por 95 empregadores diferentes (Figura 14), enquanto que para GPI são 91 empregadores distintos.

Figura 12 – Cursos concluídos por estudantes de ADS, após a evasão



Fonte: Autores

Figura 13 – ValueBox com a quantidade de perfis de estudantes evadidos de ADS encontrados no LinkedIn com dados profissionais



Fonte: Autores

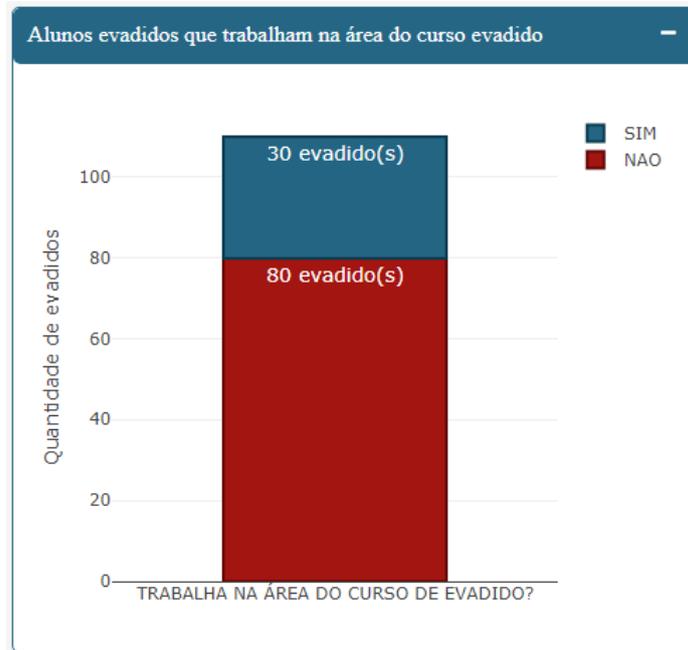
Figura 14 – ValueBox com a quantidade de empresas que empregam atualmente os estudantes evadidos de ADS



Fonte: Autores

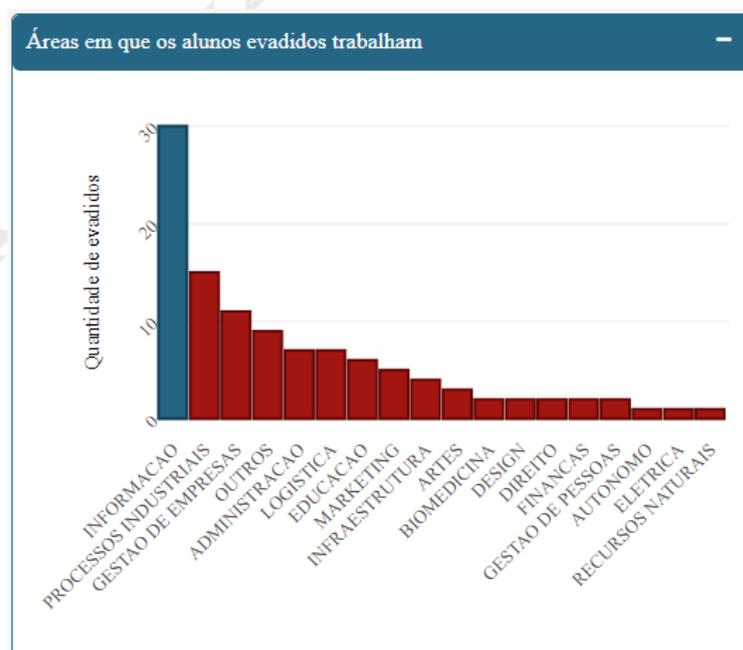
Nas Figuras 15 e 16, é possível notar que a maior parte dos estudantes evadidos de ADS não trabalham na área do curso. Dos 110 estudantes evadidos de ADS, 80 não trabalham na área (Tecnologia da Informação), e sim em Processos Industriais (15 evadidos), Gestão de Empresas (onze evadidos), entre outros. Em GPI essa quantidade é de 68 estudantes evadidos dos 101, variando-se entre as áreas de Gestão de Empresas (14 evadidos), Administração (nove evadidos), Logística (nove evadidos), entre outros.

Figura 15 – Quantidade de estudantes evadidos de ADS que trabalham na área do curso



Fonte: Autores

Figura 16 – Áreas em que os estudantes evadidos de ADS trabalham atualmente

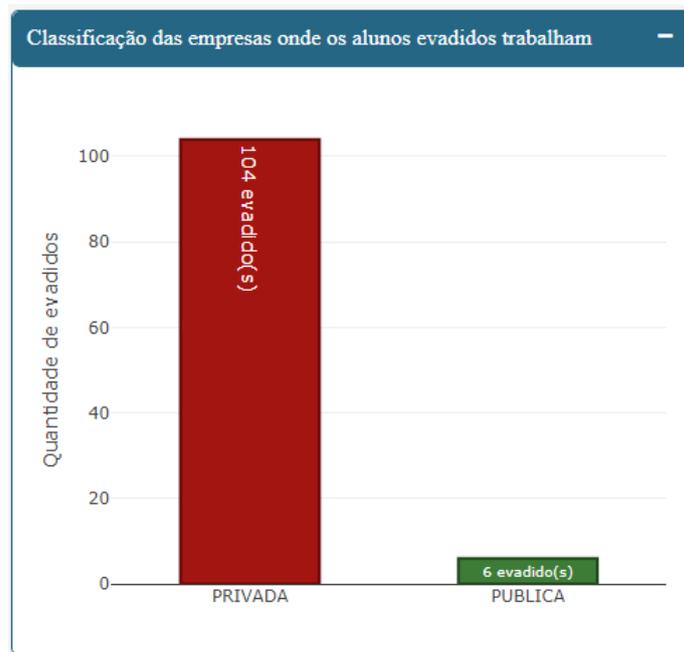


Fonte: Autores

Nas Figuras 17, 18 e 19, é possível notar que, dos 110 estudantes evadidos de ADS, 104 estão empregados no setor privado, e seis atuam em cinco órgãos distintos dentro do setor público. Dentre as empresas, a IBM é que mais emprega (quatro evadidos). No curso de GPI, apenas um estudante evadido trabalha no setor público e outros cem em empresas privadas, sendo a Eucatex, Pepsico e Toyota os maiores empregadores (três estudantes cada uma).

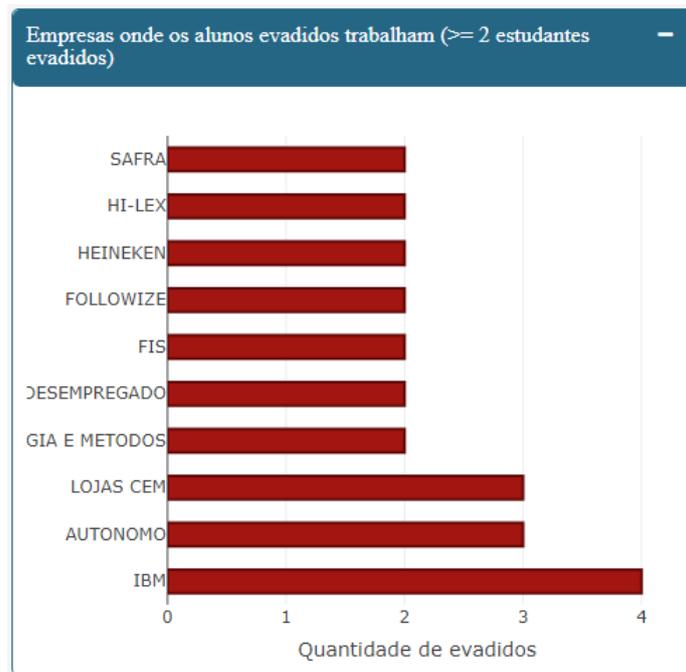
Na *dashboard*, é possível obter mais detalhes do empregador a partir do gráfico da Figura 17. Selecionando Privada (setor privado), é apresentado o gráfico da Figura 18. Caso seja selecionado Pública (setor público), é apresentado o gráfico da Figura 19.

Figura 17 – Tipo de empregador atual dos estudantes evadidos de ADS



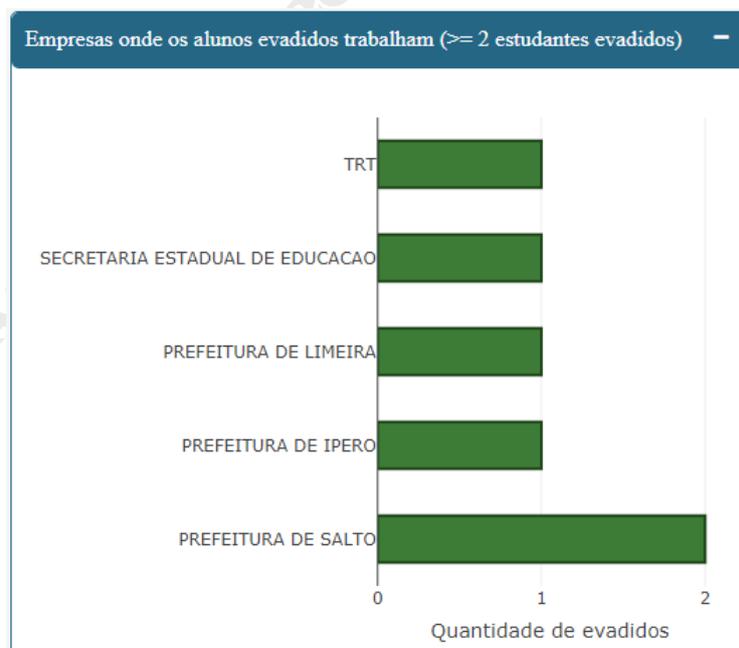
Fonte: Autores

Figura 18 – Empregador (setor privado) dos estudantes evadidos de ADS



Fonte: Autores

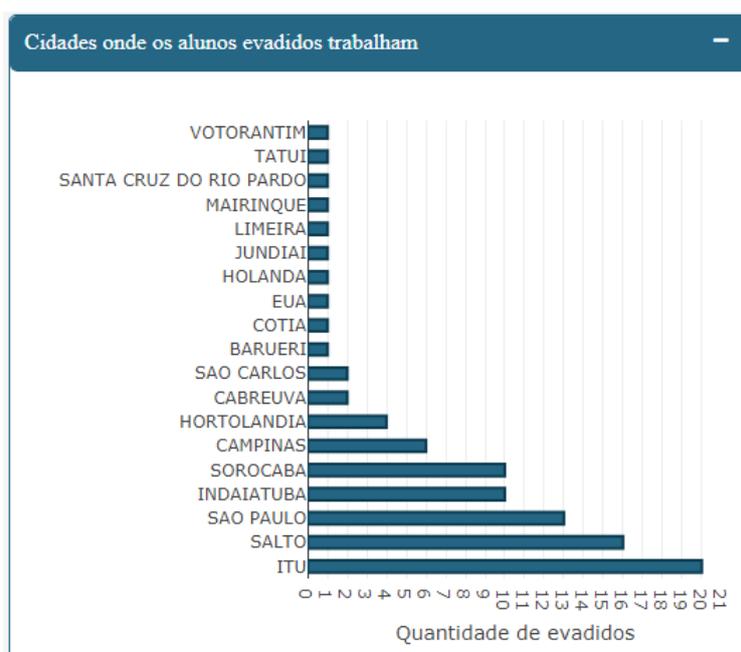
Figura 19 – Empregador atual (setor público) dos estudantes evadidos de ADS



Fonte: Autores

Na Figura 20, pode-se aferir que grande parte dos estudantes evadidos de ADS está atuando em cidades com distância menor ou igual a 100 km de Salto. As cidades com maior quantidade de estudantes evadidos são: Itu (20), Salto (16), Indaiatuba (10) e Sorocaba (10), entre outras. De GPI, 12 estudantes evadidos trabalham em São Paulo, onze em Salto, dez em Itu, oito em Campinas, entre outras.

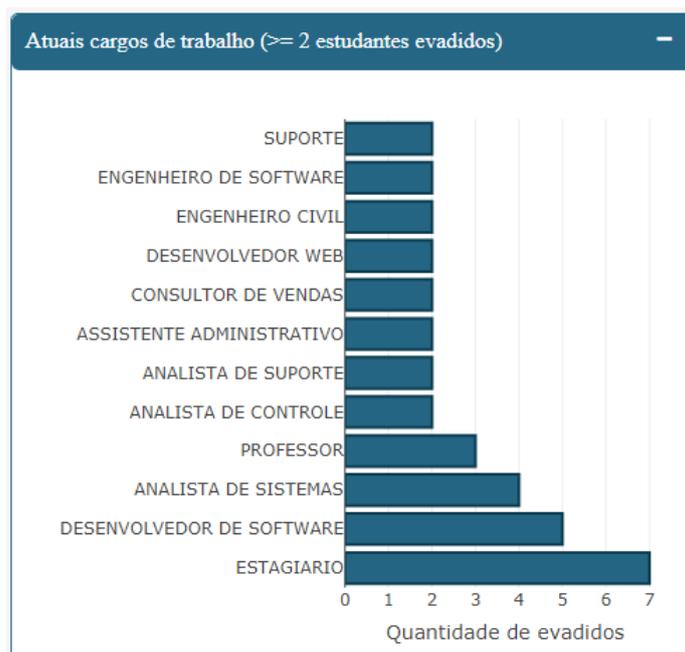
Figura 20 – Localidade onde os estudantes evadidos de ADS trabalham atualmente



Fonte: Autores

Na Figura 21, são apresentados os cargos mais frequentes entre os estudantes evadidos de ADS: estagiário (sete estudantes evadidos), desenvolvedor de software (5), analista de sistemas (4), professor (3), entre outros. Para o curso de GPI, o cargo mais frequente é assistente administrativo (5 estudantes evadidos).

Figura 21 – Cargo atual dos estudantes evadidos de ADS



Fonte: Autores

Conclusão

O acompanhamento de estudantes evadidos é uma ferramenta essencial para as instituições de ensino, pois permite a elas realizar uma autoavaliação dos seus próprios cursos, podendo realizar possíveis mudanças nos mesmos, visando a diminuição do índice de evasão na instituição com o conhecimento obtido. No entanto, este acompanhamento apresenta desafios, como a obtenção de dados educacionais e profissionais dos estudantes evadidos. Dentro deste contexto, este projeto concebeu um sistema semiautomático para coleta de informações de estudantes evadidos a partir da rede social profissional LinkedIn. Ainda assim, o processo apresenta impasses, como o fato de a maioria dos estudantes evadidos não possuir seus dados em redes sociais. Como visto nos resultados apresentados, pouco mais de um terço do total de evadidos continha dados em seu perfil LinkedIn, e muitos destes não os mantêm atualizados.

Neste trabalho, com o auxílio do sistema extrator criado no projeto, foram extraídos diversos tipos de dados sobre os estudantes evadidos do IFSP – Câmpus Salto a partir de perfis públicos da rede social profissional LinkedIn. Após organização e modelagem dos dados coletados, foi possível visualizá-los em gráficos agrupados em *dashboards* para estudo. Após uma análise mais cuidadosa, os estudos indicaram um perfil médio do estudante evadido, considerando os dois principais tipos de dados extraídos (acadêmicos e profissionais). A seguir,

são apresentadas algumas observações acerca do perfil comum dos dois cursos analisados neste estudo (Análise e Desenvolvimento de Sistemas e Gestão da Produção Industrial):

- Características gerais: nota-se que, entre estudantes evadidos, mesmo saindo de um curso superior e indo para outros, não é comum a criação de um perfil na rede social profissional LinkedIn, esta que é muito popular neste âmbito, pois visa exibir a público os principais atributos destes ex-alunos. Entretanto, observa-se uma pequena particularidade: é mais frequente a criação deste perfil entre os estudantes evadidos de Análise e Desenvolvimento de Sistemas (39,76%) do que entre os de Gestão da Produção Industrial (30,03%). Uma hipótese para este comportamento reside no fato de o LinkedIn ser uma ferramenta de Tecnologia da Informação, área da qual o curso ADS faz parte.
- Características acadêmicas: (1) a grande maioria dos estudantes evadidos de ambos os cursos já havia, antes da evasão, feito algum curso no mesmo eixo tecnológico do curso; (2) é comum que os estudantes evadidos continuem a fazer outros cursos no mesmo eixo tecnológico após a evasão, entendendo-se que a causa da evasão pode estar relacionada com: (1) localidade da instituição, de acesso dificultado por longos trajetos até a instituição, impedindo o estudante de continuar no curso; (2) turno escolar adverso, coincidindo com horário da jornada de trabalho, o que é comum aos estudantes de ensino superior (este fato que ocorre na maioria das vezes com os cursos de turno matutino).
- Características profissionais: (1) é comum que a maioria dos estudantes evadidos de ambos os cursos não trabalhe na área do curso de evasão (área de Informática para o curso de ADS e Processos Industriais para o curso de GPI), assumindo cargos relacionados a outras áreas de formação; (2) para ambos os cursos, a maioria dos estudantes evadidos tem como empregador uma empresa privada da região.

Referências

GROSSI, M. G. R.; GOMES, R. M.; NUNES, R. C. Perfil sócio-demográfico dos alunos evadidos do curso técnico a distância em Segurança do Trabalho do Polo Cabo Frio – Instituto Federal Fluminense: Um estudo de caso. *In*: Congresso Internacional ABED de Educação à Distância, 19, 2013. Salvador. **Anais** [...]. Salvador: ABED, 2013. p. 1-10.

LINKEDIN CORPORATION. **LinkedIn**. Disponível em: <https://www.linkedin.com>. Acesso em: 28 mar. 2019.

R CORE TEAM. R: **What is R?**. [S. l.]: R Foundation for Statistical Computing, 2019. Disponível em: <https://www.r-project.org/about.html> . Acesso em: 26 jun. 2019.

RSTUDIO TEAM. **RStudio**: Integrated Development for R. Boston: RStudio, 2019. Disponível em: <https://www.rstudio.com/>. Acesso em: 26 jun. 2019.

SHINY. Shiny. Disponível em: <http://shiny.rstudio.com/>. [S. l.]: RStudio, 2019. Acesso em: 9 maio 2019.

XINGSEN, L. *et al.* A Systematic Information Collection Method for Business Intelligence. In: INTERNATIONAL CONFERENCE ON ELECTRONIC COMMERCE AND BUSINESS INTELLIGENCE, 2009, Pequim. **Anais [...]**. Pequim: IEEE Computer Society, 2009. p. 116-119.

***SURVEY, GATHERING, ORGANIZATION, PROCESSING AND ANALYSIS OF
INFORMATION ABOUT DROPPED OUT STUDENTS OF IFSP COURSES FOR
EDUCATIONAL MANAGEMENT PURPOSES***

ABSTRACT

Currently, the follow-up of evaded students is an essential tool in building indicators for educational institutions. The indicators generated by this follow-up can be used to evaluate the quality of the courses offered and improve them. However, this activity is not easy because it requires the construction / maintenance of portals to register evaded students and conducting surveys using forms. However, the use of these resources in isolation has a limited effect, as it requires evaders to provide data. Therefore, this project aims to create, using the R programming language and the RStudio integrated development, a system for collecting, organizing, processing and analyzing public information from the LinkedIn professional social network about IFSP - Câmpus Salto students for educational management purposes. The system was used to collect data from the Systems Analysis and Development (SAD) and Industrial Production Management (IPM) courses at Câmpus Salto.

Keywords: *Evaded students; Dashboard; LinkedIn; R.*

Envio: novembro/2019

Aceito para publicação: dezembro/2019