

## ADEGA: Análise de Dados Estatísticos da Grade Acadêmica

Ana Paula Andrade Sodr <sup>1</sup>, Bruno Henrique Meyer<sup>1</sup>, Bruno M ller Junior<sup>1</sup>,  
Henrique Margotte<sup>1</sup>, Matheus Paolini de Miranda<sup>1</sup>, Odair Mario Ditzkun Junior<sup>1</sup>,  
Pedro Pasqualini de Andrade<sup>1</sup>, Tiago Serique Valadares<sup>1</sup>, Vin cius Miotto<sup>1</sup>

<sup>1</sup> Programa de Educa o Tutorial Computa o  
Universidade Federal do Paran  (UFPR)  
Caixa Postal 19.081 – Curitiba – PR – Brasil

{apas19, bhmeyer, bmuller, clfs16, hm19, mpm19,  
omdj17, ppa19, tsv19, vm20}@inf.ufpr.br

**Abstract.** *This paper proposes an academic data processing model, which assists the management of course coordinators. The model was implemented through a system called ADEGA (Academic Grid Statistical Data Analysis). The academics management systems usually don't provide greater functionality than a repository of student grades and attendance. The ADEGA uses this data to provide graphics and information, which are not usually included in other systems. Among the information provided by the system, there are a risk of academic retirement's list, a comparative chart of failures by subjects, among others.*

**Resumo.** *Este artigo prop e um modelo de processamento de dados acad micos que auxilie a gest o de coordena es de curso. O modelo foi implementado por meio de um sistema chamado ADEGA (An lise de Dados Estat sticos da Grade Acad mica). Os sistemas de gest o acad mica normalmente n o fornecem maiores funcionalidades do que um reposit rio de notas e frequ ncias de alunos. O ADEGA utiliza esses dados para fornecer gr ficos e informa es, as quais n o costumam ser abordadas em outros sistemas. Dentre as informa es fornecidas pelo sistema, est o presentes uma lista de alunos em risco de jubilamento, um quadro comparativo entre reprova es por disciplina, entre outros.*

### 1. Introdu o

Coordena es de curso s o respons veis por v rias tarefas em uma institui o de ensino superior, dentre elas, uma das que demandam tempo   o aconselhamento de alunos. Para faz -lo de maneira coerente, o coordenador precisa ter conhecimento do progresso do estudante no curso, por exemplo saber em quais disciplinas o aluno j  foi aprovado, em quais ele consta como reprovado e quais ainda n o foram cursadas. As ferramentas que s o disponibilizadas para esta tarefa, como relat rios em texto do progresso do aluno, em geral n o s o adequadas, por n o fornecerem muitos m todos de visualiza o dos dados, dificultando a an lise realizada pelos coordenadores.

A Universidade Federal do Paran  (UFPR) utiliza um sistema de acompanhamento acad mico chamado Sistema de Informa es para o Ensino (SIE)<sup>1</sup>, o qual oferece

<sup>1</sup><https://avmbasten.com/sie/>

um repositório de dados para as coordenações da universidade, porém tal sistema é incapaz de gerar informações críticas para a gestão de ensino. Além desse sistema, os coordenadores de curso da universidade em questão também utilizavam planilhas eletrônicas, as quais tornavam-se obsoletas e complicadas de manipular com o tempo. Com isso, o trabalho de gestão acadêmica e a tomada de decisões, discutidas em reuniões de colegiados podem ser dificultados ou comprometidos.

No contexto de repositório de dados, caso o coordenador de um curso queira a lista de alunos que tiveram uma acentuada queda de desempenho em dado período de tempo, alguns problemas podem ser encontrados para cada abordagem, dentre os quais destacamos:

1. utilizar os relatórios fornecidos pelos repositórios. Infelizmente tais relatórios contém apenas listagens simples, além de estarem no formato texto e não sintetizarem as informações, o que dificulta a análise;
2. os repositórios fornecem algumas consultas padronizadas (por exemplo, a lista de todos os alunos do curso) cujos resultados são disponibilizados em arquivos para *download* e posterior importação para análise em planilhas eletrônicas. Entretanto, as consultas não são personalizáveis (podendo faltar algumas informações), somado a isso, alguns arquivos podem conter milhares de linhas e colunas, o que torna a planilha lenta;
3. executar consultas *SQL* personalizadas diretamente do repositório de dados, importando os resultados em planilhas para análise. Todavia, estes repositórios de dados nem sempre estão acessíveis diretamente aos coordenadores por questões de privacidade e segurança.

Dos problemas citados acima, destaca-se a dificuldade em obter informações relevantes a partir dos dados disponibilizados aos profissionais da gestão de ensino. Uma melhoria significativa seria exibir as informações de maneira gráfica. Por exemplo, para visualizar o progresso acadêmico de um estudante, é possível apresentar a grade curricular com cores diferentes para as disciplinas em que esse aluno está aprovado, reprovado, matriculado, etc.

A partir dessa ideia, desenvolveu-se um sistema que importa arquivos do repositório de dados contendo notas e frequências de todos os alunos, cria um banco de dados próprio e esquematiza automaticamente a grade curricular na tela do computador, a partir do registro acadêmico de um dado aluno. Essa implementação foi o que deu início ao ADEGA, que atualmente contempla dezenas de funcionalidades, as quais foram testadas por professores com experiência em gestão de ensino.

O ADEGA foi implementado por meio de um sistema web que permite que múltiplos usuários enviem os repositórios de dados exportados do SIE. As informações necessárias para os repositórios consistem em informações da matrícula e o histórico escolar de cada aluno de um curso de graduação. Esses repositórios são processados utilizando tecnologias de processamento e visualização de dados, de forma que o usuário tenha acesso a resumos desses repositórios. Esses resumos podem representar métricas como a nota média de alunos e gráficos como a evolução da taxa de evasão. O código-fonte do projeto foi disponibilizado de forma que a comunidade possa contribuir com a definição de novos gráficos e métricas calculadas no processamento dos repositórios.

Além disso, um conversor foi implementado de forma que seja facilitada a exportação de dados provenientes de repositórios diferentes do SIE.

O texto está organizado da seguinte forma: A seção 2 apresenta trabalhos relacionados. A seção 3 descreve a arquitetura do sistema ADEGA, a seção 4 aborda a metodologia de desenvolvimento e a seção 5 discorre sobre a avaliação e validação do sistema. A seção 6 apresenta alguns dos resultados obtidos. Por fim, a seção 7 contém a conclusão e alguns trabalhos futuros.

## 2. Revisão Bibliográfica

A maior parte dos trabalhos publicados na área de mineração de dados educacionais analisa o desempenho de alunos em disciplinas que utilizam ambientes virtuais de aprendizagem (AVA) [Maschio et al. 2018], como alunos em risco de reprovação [Rigo et al. 2014, Santos et al. 2015] análise de desempenho [Ferreira e Vasconcelos 2017] e de comportamento [Dionísio et al. 2017] dos estudantes, entre outros.

Em geral, esses trabalhos têm como principal objetivo a solução de problemas específicos, como a predição de evasão de curso ou a identificação de perfis de desempenho. Dessa forma, nenhum dos trabalhos mencionados consideram a perspectiva de gestores e coordenadores de cursos, que podem utilizar as informações extraídas de dados para tomar decisões práticas.

O trabalho apresentado aqui se diferencia daqueles pois o “cliente” não é o professor, mas sim a coordenação de um curso de uma instituição de ensino superior. Foram encontrados poucos trabalhos que abordam este contexto, dos quais destacamos, [Manhães et al. 2011] que identifica precocemente os alunos em risco de evasão, utilizando o software de mineração de dados *weka*. [Išljamović e Lalić 2014] importa as informações do banco de dados acadêmicos da Universidade de Belgrado para o software proprietário QlikView<sup>2</sup> e utiliza as ferramentas lá disponibilizadas para fazer diversas comparações entre o desempenho de grupos de estudantes.

O ADEGA foi construído de forma que possa incorporar no futuro os métodos e tecnologias propostas pelos trabalhos mencionados nesta seção, como a predição de evasão nos cursos. Porém, este artigo tem como principal objetivo a proposta de um sistema que permita coordenadores de cursos acessarem informações extraídas de um repositório de dados. Além disso, o ADEGA também se diferencia por permitir exportar as informações computadas no formato de planilhas, que podem ser facilmente utilizadas em algoritmos de aprendizado de máquina [Manhães et al. 2011].

## 3. O Sistema ADEGA

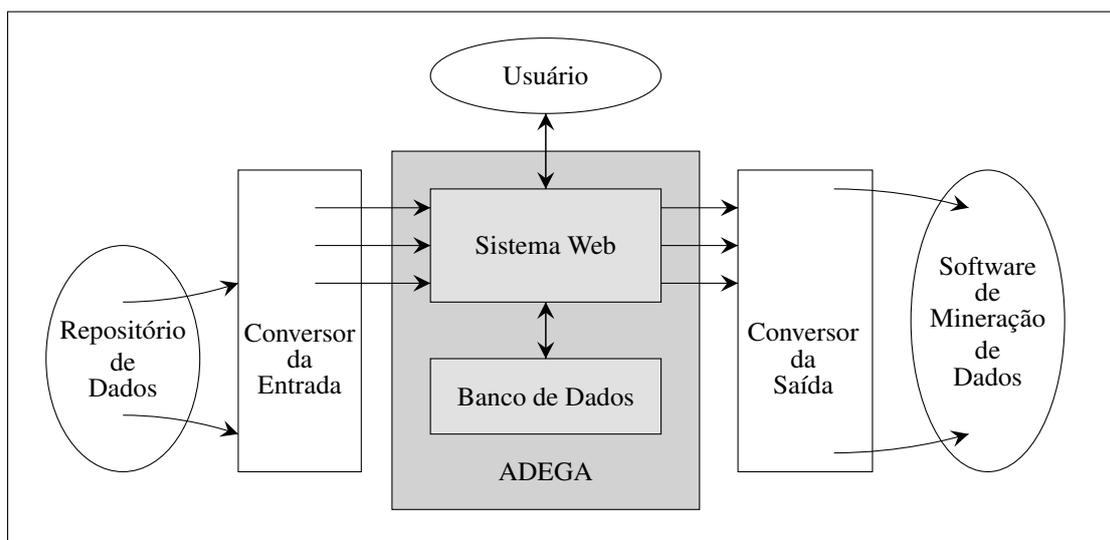
O ADEGA é um sistema *web* que visa auxiliar os coordenadores na análise de dados sobre o curso de Bacharelado em Ciência da Computação da Universidade Federal do Paraná, portanto as ferramentas de síntese, como os gráficos, tabelas e métricas foram definidas conforme a demanda desses profissionais. Outras funcionalidades serão desenvolvidas à medida que novas necessidades forem detectadas. Para que seja possível utilizar tal sistema em outros contextos, considera-se necessário a realização de levantamentos das

---

<sup>2</sup><https://www.qlik.com/us/>

necessidades específicas de outros cursos, sejam estes da UFPR ou de outra instituição, para serem agregados ao ADEGA.

A Figura 1 apresenta o modelo esquemático do ADEGA. O sistema funciona no modelo ETL (*Extraction, Transformation, Load*) [Christen 2012]. Do repositório de dados indicado no lado esquerdo da Figura 1 são extraídos os dados do sistema, por exemplo em arquivos `csv`.



**Figura 1. Modelo esquemático do ADEGA**

O conversor de entrada é responsável por converter os dados de um repositório para arquivos no formato aceito pelo *software*, esse padrão está documentado no sítio do ADEGA<sup>3</sup>, onde utiliza-se o formato de entrada do SIE, por exemplo. Dessa forma, para o usuário utilizar o sistema com o objetivo de analisar dados provenientes de outros repositórios, é necessário adaptar a implementação do conversor de entrada, o que possibilita o uso da ferramenta em outros contextos.

Após convertidos, os arquivos são repassados para o banco de dados do ADEGA via *upload* com o uso de um dos módulos do sistema *web* desenvolvido. Esse mesmo sistema também é o responsável pela interação com o usuário, apresentando consultas que sintetizam questões referentes a alunos e disciplinas.

Essas consultas têm como resultados diferentes gráficos e métricas específicas, produzidos durante o processamento do repositório enviado pelo usuário. Um exemplo de consulta no sistema *web* consiste no fluxo de atividades onde o usuário acessa o sistema, envia o repositório de dados, acessa a página relacionada a um aluno e obtém um gráfico que informa a evolução da aprovação média de um aluno em diferentes semestres. As métricas e gráficos implementadas no Sistema Web estão disponíveis no repositório público do projeto ADEGA, apresentado na Seção 4. Devido à alta variedade e complexidade técnica dessas implementações, este artigo não irá detalhá-las.

Apesar do ADEGA disponibilizar algumas consultas, seria inviável elaborar um conjunto que atendesse às necessidades de todos os cursos. Por esta razão, foi criado um

<sup>3</sup><https://gitlab.c3sl.ufpr.br/adega/conversor-de-dados-adeqa>

módulo de saída, ilustrado do lado direito da Figura 1. Este módulo contém um conjunto de funcionalidades que permitem o usuário exportar informações contidas no banco de dados para uso em programas de mineração de dados como, por exemplo o *Weka*<sup>4</sup>. No momento, a exportação se dá no formato CSV.

Nesse contexto, é importante destacar que os componentes ilustrados no lado esquerdo da Figura 1 tornam o sistema ADEGA independente de entrada, enquanto os ilustrados no lado direito facilitam a exportação para múltiplos sistemas de mineração de dados.

### 3.1. Entidades e Métricas

Os estudos que precederam o desenvolvimento do ADEGA apontaram que os sistemas de gestão acadêmica baseiam-se no relacionamento de quatro entidades. Para cada entidade foi desenvolvido um conjunto de métricas que sintetizam as informações lá contidas.

As métricas de entidades são informações numéricas, em sua maioria porcentagens, que resumem dados relacionados às aprovações, notas e outros indicadores individuais dos diferentes estudantes. Com a submissão de novos dados pelo usuário, a base de dados é atualizada e as métricas e entidades são recalculadas, gerando novas informações que permitem as coordenações de curso identificar quais aspectos do curso de graduação requerem maior atenção e observar o efeito de mudanças pedagógicas do curso.

Essas métricas podem ser aplicadas a conjuntos de um ou mais estudantes selecionados, isso é possível com o uso de algum critério que pode incluir todos os alunos ingressantes em um determinado ano ou os que foram aprovados em alguma disciplina.

A entidade Aluno contém informações de um estudante e suas relações com as outras entidades. Cada aluno está associado a uma grade curricular, a qual pode ser modificada durante sua jornada no curso. Ao considerar essa entidade, podemos elencar métricas como o tipo de evasão do curso (ou sem evasão no caso de ser um aluno vinculado ao curso de graduação), o Índice de Rendimento Acadêmico (IRA) individual e a taxa de aprovação semestral. A Equação 1 representa o cálculo necessário para obter o IRA de um aluno, onde  $N_i$  e  $C_i$  representam respectivamente a nota e o total de créditos de uma disciplina presente no histórico do aluno.

$$IRA = \frac{\sum_i(N_i \times C_i)}{100 \times \sum_i(C_i)} \quad (1)$$

Na entidade Disciplina é possível encontrar informações do desempenho de todos os alunos que se matricularam em uma determinada disciplina. Além disso, o sistema permite comparar informações de duas disciplinas distintas. As métricas dessa entidade resumem a importância da matéria na grade acadêmica do curso em questão. Nota média (tanto geral quanto do último ano), taxas de reprovação e de trancamento, assim como a média da quantidade de vezes que os alunos cursaram cada disciplina até obterem aprovação, são algumas das métricas desta entidade.

As informações relacionadas ao desempenho de todos os estudantes de um curso, ativos e inativos, são sumarizadas pela entidade Curso. Essa entidade possui um conjunto de grades curriculares (que incluem pré-requisitos) e, para cada uma delas, a lista das

<sup>4</sup><https://www.cs.waikato.ac.nz/ml/weka/>

disciplinas cujas aprovações são necessárias para concluir o curso. As métricas de um Curso informam o seu estado atual em relação ao histórico. Dessa forma, o IRA médio, a quantidade de alunos ativos e as taxas de reprovação são métricas importantes para essa entidade.

A entidade Turma de Ingresso sumariza informações de um conjunto de alunos que ingressaram no curso em um mesmo período. Para tal entidade, deve-se levar em consideração os índices de reprovação geral e de evasão por tipo, assim como as médias de IRA e o tempo de formatura em anos da turma em questão.

Cada entidade também pode ser apresentada por meio de uma tabela em que cada linha representa uma instância da mesma entidade e as colunas representam dados relativos à instância. Com a utilização de filtros e com o auxílio de uma barra de busca, a tabela permite ao usuário visualizar esses dados e instâncias de modo dinâmico e sintetizado.

## 4. Metodologia do ADEGA

Utilizando o paradigma de programação orientada a objetos, cada uma das entidades descritas acima foi mapeada como uma classe. Assim, uma instância da classe Curso é composta por instâncias das classes Aluno e Disciplina. Cada classe contém uma série de métodos, os quais fazem análise dos dados lá contidos, como o desempenho histórico de aprovações em uma disciplina. Esta generalização permite que o ADEGA possa ser utilizado em outros cursos ou em ambientes além do universitário.

### 4.1. Implementação e Ambiente de Desenvolvimento

O sistema ADEGA foi implementado combinando várias tecnologias, dentre as quais destacamos o *framework web* Django, com a linguagem de programação Python 3, o qual é a base para a construção do sistema que aceita o repositório de dados dos cursos, analisa os dados em um servidor e gera páginas *web* com diferentes relatórios gerenciais.

As bibliotecas Pandas e Numpy [McKinney 2012] foram utilizadas para processar as informações dos repositórios, enquanto o PlotlyJS [Sievert et al. 2016] foi utilizado para a criação de gráficos que sintetizam diferentes informações (*dashboards*) [Schwendimann et al. 2017] das entidades do sistema.

Por fim, o Docker [Merkel 2014] foi a ferramenta empregada para gerenciar a modularização do banco de dados, da aplicação web e do servidor *Hypertext Transfer Protocol* (HTTP). Essa tecnologia permite simplicidade na reprodutibilidade do trabalho e nos fatores de segurança envolvidos nas tecnologias utilizadas.

O ADEGA é um *software* livre, logo é possível encontrar seu código fonte, além de um manual com instruções de uso e configuração em um repositório público disponível no GitLab do grupo PET Computação da UFPR<sup>5</sup>.

## 5. Validação de Tecnologia

Para entender o potencial do modelo proposto, o seguinte experimento foi desenvolvido: em um primeiro momento o participante iria interagir com o sistema, executando uma série de tarefas predefinidas que foram disponibilizadas por meio de um roteiro.

---

<sup>5</sup><https://gitlab.c3sl.ufpr.br/adega/adega>

Após realizar todas as tarefas, o participante deveria responder um questionário elaborado com base no Modelo de Aceitação de Tecnologia (TAM) [Davis 1989], relacionado à percepção de utilidade e facilidade de uso do sistema. As questões foram formuladas como afirmações e, seguindo a escala Likert [Likert 1932], o participante deveria indicar o quanto concordava ou discordava de uma determinada afirmação.

As afirmações estavam divididas em 3 seções: (i) experiência com gestão acadêmica, (ii) percepção de utilidade do sistema e (iii) percepção de facilidade de uso do sistema. O grau de concordância deveria ser expresso de acordo com a seguinte classificação: Discordo Plenamente, Discordo, Opinião Neutra, Concordo e Concordo Plenamente. Além disso, espaços para opiniões gerais e sugestões de melhoria também foram incluídos.

## 6. Resultados Obtidos

O ADEGA é composto por dezenas de funcionalidades as quais permitem o usuário extrair informações relevantes de um curso de graduação. Dentre as principais funcionalidades do sistema, destaca-se a visualização da grade curricular (seção 6.1) e visualização gráficos interativos (seção 6.2).

### 6.1. Visualização da Grade Curricular

A página de visualização da entidade Aluno possui a grade curricular, que consiste em uma tabela onde cada coluna representa um período do curso e cada célula representa uma disciplina daquele período, como exemplificado na Figura 2.

1°	2°	3°	4°	5°	6°	7°	8°
CI068 Matrícula: 2 Última Nota: 70,00	CI210	CI212	CI215	CI162	CI163	CI221	OPT
CI055 Matrícula: 1 Última Nota: 70,00	CI056	CI057	CI062	CI065	CI165	CI211	OPT
CI046	CI067	CI064	CE003	CI059	CI009	OPT	OPT
CI043 Matrícula: 2 Última Nota: 58,00	CI009	CI237	CI058	CI061	CI218	OPT	OPT
CI001 Matrícula: 1 Última Nota: 76,00	CI002	CI166	CI164	SA214	CI220	TG I	TG II

1°	2°	3°	4°	5°	6°	7°	8°
CI068 Taxa: 0,60 Nota: 48,54	CI210 Taxa: 0,48 Nota: 39,15	CI212 Taxa: 0,50 Nota: 40,52	CI215 Taxa: 0,65 Nota: 51,62	CI162 Taxa: 0,81 Nota: 64,67	CI163 Taxa: 0,72 Nota: 51,31	CI221 Taxa: 0,82 Nota: 61,26	OPT
CI055 Taxa: 0,48 Nota: 39,33	CI056 Taxa: 0,62 Nota: 50,57	CI057 Taxa: 0,58 Nota: 47,51	CI062 Taxa: 0,82 Nota: 63,62	CI065 Taxa: 0,70 Nota: 50,76	CI165 Taxa: 0,66 Nota: 43,13	CI211 Taxa: 0,75 Nota: 59,52	OPT
CI046 Taxa: 0,45 Nota: 39,47	CI067 Taxa: 0,53 Nota: 45,36	CI064 Taxa: 0,48 Nota: 50,11	CE003 Taxa: 0,59 Nota: 43,95	CI059 Taxa: 0,58 Nota: 46,12	CI009 Taxa: 0,69 Nota: 51,88	OPT	OPT
CI043 Taxa: 0,53 Nota: 42,21	CI009 Taxa: 0,50 Nota: 39,55	CI237 Taxa: 0,36 Nota: 34,64	CI058 Taxa: 0,59 Nota: 44,88	CI061 Taxa: 0,92 Nota: 66,81	CI218 Taxa: 0,71 Nota: 54,56	OPT	OPT
CI001 Taxa: 0,53 Nota: 43,06	CI002 Taxa: 0,54 Nota: 42,89	CI166 Taxa: 0,73 Nota: 53,65	CI164 Taxa: 0,78 Nota: 45,87	SA214 Taxa: 0,88 Nota: 66,00	CI220 Taxa: 0,90 Nota: 68,34	TG I	TG II

Figura 2. Visualização da Grade Curricular do Aluno

Essas células sintetizam informações sobre o desempenho de um determinado aluno naquela disciplina, como o número de vezes que já se matriculou e a nota da última vez que a cursou. Também utiliza cores para indicar se o estudante consta naquela disciplina como aprovado (cor verde), reprovado (cor vermelha) e matriculado (cor azul). Células com um ícone de cadeado indicam uma disciplina que exige pré-requisito não cumprido, e células na cor branca indicam que a disciplina não foi cursada.

A mesma estrutura de grade curricular existe para a entidade Curso, com a diferença que os dados exibidos em cada célula são nota média e taxa de aprovação de cada disciplina. A grade do Curso possui a opção de ser visualizada como mapa de calor, auxiliando o usuário a identificar disciplinas críticas: quanto menor a taxa de aprovação da disciplina mais vermelha é a sua célula e quanto maior a taxa de aprovação da disciplina mais verde é a sua célula.

A grade do Curso mostra quais são as dependências e pré-requisitos de cada disciplina de forma interativa: ao passar o *mouse* sobre uma célula, as outras células das disciplinas que são pré-requisito daquela são destacadas em azul e as células da disciplinas que a tem como pré-requisito são destacadas em roxo.

## 6.2. Gráficos

As tabelas de métricas e os gráficos interativos (ver Figura 3) são gerados a partir dos dados das entidades do sistema (ver seção 3.1), além de se adequarem aos filtros aplicados pelo usuário. Ambos são utilizados para representar diversas relações entre determinados parâmetros, como IRA, quantidade de alunos, semestres e anos, etc. Com isso, podem auxiliar na análise detalhada dos dados do curso.

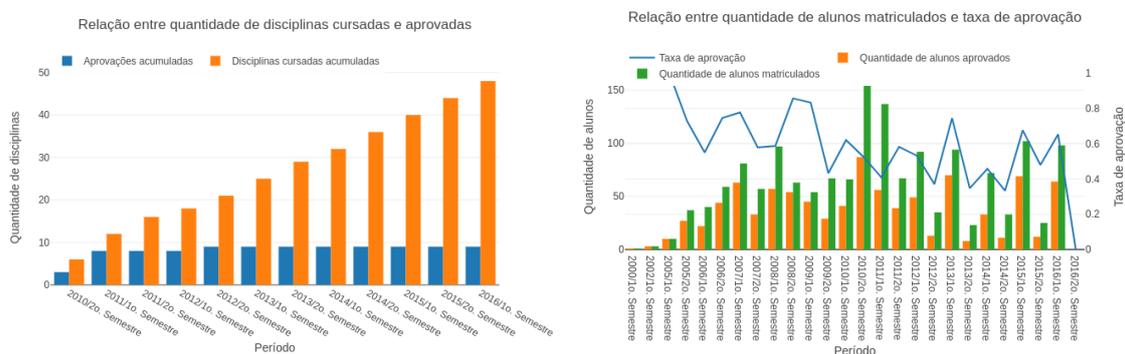


Figura 3. Gráficos Ilustrando Matrículas e Aprovações.

Na entidade Curso, foram utilizados gráficos de barras juntamente com gráfico de linhas para exibir o histórico de evasão do curso. Para essa variável existem dois gráficos, o primeiro mostra as evasões para os semestres de todos os anos contidos nos dados, enquanto o segundo considera quantas evasões cada período do curso obteve. Esse formato também foi utilizado para exibir a distribuição de alunos por faixas de IRA.

Para a entidade Disciplina, relaciona-se em um histograma à quantidade de alunos matriculados e aprovados em uma disciplina, combinado a um gráfico de linhas que representa a taxa de aprovação. Enquanto que na entidade Aluno, emprega-se um gráfico de linhas que relaciona o IRA semestral e, juntamente com um histograma, a quantidade de disciplinas cursadas naquele semestre. O mesmo ocorre na entidade Turma de Ingresso, para exibir as aprovações acumuladas e os índices de aprovações por semestre.

## 6.3. Resultados da Avaliação Qualitativa

Dentre os que participaram do experimento de validação da tecnologia (ver seção 5), todos declararam ter experiência prévia com gestão acadêmica e ter participado de coordenação de curso, além de que os três participantes relataram contato com alguma tecnologia de auxílio na gestão acadêmica, sendo que dois dos respondentes afirmaram ter tido algum tipo de experiência com o ADEGA anteriormente. Todos os participantes conseguiram executar as tarefas propostas, havendo apenas um participante que relatou dificuldades para encontrar um campo na interface gráfica durante uma das atividades.

Todos participantes concordaram que o sistema pode auxiliar na gestão da informação e tomada de decisão pelos coordenadores de curso e nenhum deles relatou di-

ficuldades para utilizar o sistema de modo geral ou para interpretar as métricas e gráficos gerados pelos diferentes módulos do ADEGA.

Diversos aspectos da avaliação podem comprometer a validade dos experimentos, como o baixo número de participantes e o fato de que parte dos participantes já tiveram algum conhecimento prévio do sistema. Contudo, essas limitações podem ser superadas em avaliações e trabalhos futuros, considerando que os resultados gerais dos experimentos neste artigo indicam que o ADEGA poderia ser utilizado para auxiliar gestores de cursos do ensino superior.

## 7. Conclusão e Trabalhos Futuros

Este trabalho analisa os desafios relacionados à interpretação de dados de alunos de graduação a partir de seus históricos acadêmicos. Além disso, também é proposto um sistema chamado ADEGA, capaz de auxiliar gestores e coordenadores de cursos para a análise desses dados. O sistema proposto foi implementado considerando quatro principais entidades: Curso, Aluno, Disciplina e Turma Ingresso.

O sistema foi implementado utilizando ferramentas de desenvolvimento *web* como Django e Plotly, além de ferramentas de processamento de dados como Numpy e Pandas, da linguagem de programação Python. O principal fluxo de funcionamento do sistema consiste em um usuário realizar *upload* de uma planilha para o sistema, que irá processá-la e criar uma página *web* com informações detalhadas de cada entidade, incluindo métricas e gráficos interativos.

Durante a pesquisa e desenvolvimento do sistema, foi identificado que o modelo ETL (*Extraction, Transformation, Load*) é adequado para processar dados acadêmicos e auxiliar tarefas relacionadas a gestões de cursos de graduação. Isso foi demonstrado por meio de uma avaliação qualitativa realizada neste estudo. Além de proporcionar ao usuário uma interação *web* para acessar informações, extraídas do repositório de dados, o sistema *web* implementado também possibilita a exportação dessas informações para planilhas, de forma que elas sejam utilizadas por outros sistemas de mineração de dados ou de aprendizado de máquina. As análises que extraem as informações dos repositórios foram definidas em conjunto de coordenadores de cursos de graduação, e estão disponíveis em um repositório público com código-aberto, de forma que novas contribuições possam ser propostas pela comunidade. Além disso, o desenvolvimento do sistema foi orientado por um formato genérico de repositório de dados, de forma que seja simplificada o uso do ADEGA em diferentes Instituições de Ensino Superior.

O sistema desenvolvido foi avaliado por meio de um experimento onde diferentes docentes que possuem experiências com coordenações de cursos utilizaram o ADEGA e responderam um formulário com perguntas relacionadas ao seu uso. Apenas um participante não conseguiu encontrar a informação desejada no sistema para uma das perguntas. A avaliação geral do sistema permitiu concluir que o sistema tem potencial para auxiliar os coordenadores, considerando que todos os participantes já tiveram algum contato com gestões de cursos de graduação. Entretanto, algumas dificuldades foram identificadas em relação à experiência do usuário ou *User Experience (UX)*. Por esse motivo, diversos trabalhos futuros podem ser realizados, de forma que novas funcionalidades possam ser adicionadas ao sistema como melhorias na interface do usuário ou a integração direta com outros sistemas como o SIE.

## Referências

- Christen, P. (2012). *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection*. Springer.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*.
- Dionísio, M., Ferreira, R., Cavalcanti, A., Carvalho, R., e Neto, S. (2017). Mineração de Texto Aplicada à Identificação de Colaboração em Fóruns Educacionais. *Anais do XXVIII Simpósio Brasileiro de Informática na Educação (SBIE 2017)*.
- Ferreira, V. A. S. e Vasconcelos, G. C. (2017). Recomendação de recursos educacionais baseada em aprendizagem de máquina para autorregulação da aprendizagem. *Anais do XXVIII Simpósio Brasileiro de Informática na Educação (SBIE 2017)*.
- Išljamović, S. e Lalić, S. (2014). Academic dashboard for tracking students' efficiency. In *Proceedings Of XIV International Symposium New Business Models And Sustainable Competitiveness*.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology*.
- Manhães, L. M. B., da Cruz, S. M. S., Costa, R. J. M., Zavaleta, J., e Zimbrão, G. (2011). Previsão de estudantes com risco de evasão utilizando técnicas de mineração de dados. *Anais do XXII Simpósio Brasileiro de Informática na Educação*.
- Maschio, P. d. T. M., Vieira, M. A., Sara, N. T. d. C., de Melo, L., e Junior, C. X. P. (2018). Um panorama acerca da mineração de dados educacionais no brasil. *Anais do XXIX Simpósio Brasileiro de Informática na Educação*.
- McKinney, W. (2012). *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*. "O'Reilly Media, Inc."
- Merkel, D. (2014). Docker: lightweight linux containers for consistent development and deployment. *Linux journal*.
- Rigo, S. J., Cambuzzi, W., Barbosa, J. L. V., e Cazella, S. C. (2014). Aplicações de mineração de dados educacionais e learning analytics com foco na evasão escolar: oportunidades e desafios. *Revista Brasileira de Informática na Educação*, 22(01).
- Santos, F. D. S., Bercht, M., e Wives, L. (2015). Classificação de alunos desanimados em um asea: uma proposta a partir da mineração de dados educacionais. *Anais do XXVI Simpósio Brasileiro de Informática na Educação*.
- Schwendimann, B., Rodriguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., e Holzer, A. (2017). Perceiving learning at a glance: A systematic literature review of learning dashboard research. *IEEE Transactions on Learning Technologies*.
- Sievert, C., Parmer, C., Hocking, T., Chamberlain, S., Ram, K., Corvellec, M., e Despouy, P. (2016). plotly: Create interactive web graphics via plotly's javascript graphing library [software].