

Universidade Federal de Uberlândia  
Faculdade de Computação

# VIII FACOM TECHWEEK E XV WORKSHOP DE TESES E DISSERTAÇÕES EM CIÊNCIA DA COMPUTAÇÃO

Anais

13 a 17 de dezembro de 2021

ISSN: 2447-0406



FACOM  
TECHWEEK

**WTDC**   
XV Workshop de Teses e Dissertações  
em Ciência da Computação

Uberlândia  
2021

# VIII FACOM Techweek

## **Anais do XV WTDC**

Evento virtual

13 a 17 de Dezembro de 2021

### **Comissão Organizadora**

Elaine Ribeiro de Faria Paiva (UFU)

Henrique Coelho Fernandes (UFU)

Murillo Guimarães Carneiro (UFU)

Renan Gonçalves Cattelan (UFU)

Wendel Alexandre Xavier de Melo (UFU)

### **Organização, Execução, Promoção e Realização**

Universidade Federal de Uberlândia (UFU)

Faculdade de Computação (FACOM/UFU)

# **FACOM Techweek e WTDCC**

A TechWeek é um evento realizado pela Faculdade de Computação (FACOM) da Universidade Federal de Uberlândia (UFU), que busca promover atualização técnica e apresentar as últimas tendências tecnológicas, reunindo e proporcionando uma maior interação entre estudantes, professores, profissionais e empresas das áreas de Tecnologia da Informação. Trata-se de um evento tradicional, já em sua oitava edição, composto por palestras, minicursos, mesas redondas, competições técnicas e apresentação de trabalhos e pesquisas científicas.

Realizado em conjunto com a Techweek desde 2016, o Workshop de Teses e Dissertações em Ciência da Computação (WTDCC) é um evento do Programa de Pós-Graduação em Computação (PPGCO) da UFU, na sua décima quinta edição, o qual conta com a participação de estudantes de graduação e pós-graduação de universidades de Uberlândia e região. Dentre os objetivos do WTDCC destacam-se: incentivar a troca de experiência entre pesquisadores, apoiar os trabalhos produzidos pelos discentes, promover a interação entre os diversos grupos de pesquisa e refletir sobre a atual situação deles no cenário nacional. Nesse sentido, o evento consiste em um importante espaço de integração para a troca de experiências acadêmico-científicas, objetivando o desenvolvimento da ciência e da tecnologia nesse domínio do conhecimento.

Considerando as recomendações de isolamento social em virtude da pandemia do COVID-19, ambos os eventos aconteceram de maneira online em 2021.

# Prefácio

O Workshop de Teses e Dissertações em Ciência da Computação foi realizado junto com a VIII FACOM Techweek e teve como objetivo divulgar os projetos de pesquisa científica e tecnológica na área de computação, realizados por estudantes de graduação e pós-graduação da Universidade Federal de Uberlândia (UFU) e região. Também teve por objetivo contribuir para a formação dos participantes despertando o interesse pelas descobertas científicas e pela resolução de problemas complexos.

Na sua décima quinta edição, o WTDCC contou com palestras temáticas, bem como a submissão e apresentação de trabalhos. As palestras compreenderam tópicos relacionados às linhas de pesquisa desenvolvidas pelo Programa de Pós-Graduação em Ciência da Computação (PPGCO/UFU), com representantes da indústria e da academia. O evento também contou com uma mesa redonda sobre os 20 anos de pós-graduação da FACOM-UFU, que discutiu experiências e perspectivas futuras para o programa. As submissões de trabalho, em formato de resumo, contemplaram três trilhas relacionadas à formação dos discentes: trabalhos de graduação, mestrado e doutorado, vários deles apresentados durante o evento em sessões técnicas virtuais.

Além de contar com os trabalhos de dezenas de autores, o WTDCC também contou com boa presença do público, interessado em conhecer mais sobre os trabalhos desenvolvidos em Uberlândia e região. Nesse sentido, agradecemos a todos os alunos e orientadores que submeteram seus trabalhos para o evento e também ao público que prestigiou as palestras e apresentação de trabalhos.

Por fim, agradecemos a comissão de avaliação dos trabalhos e também às empresas e instituições de Uberlândia e região que marcaram presença no evento.

Profa. Elaine Ribeiro de Faria Paiva

Prof. Henrique Coelho Fernandes

**Coordenadores do XV WTDCC**

# Organização

## **Coordenação Geral**

Prof. Wendel Alexandre Xavier de Melo (UFU)

## **Coordenação do XIV WTDC**

Profa. Elaine Ribeiro de Faria Paiva (UFU)

Prof. Henrique Coelho Fernandes (UFU)

## **Comissão de Avaliação dos Trabalhos**

Profa. Elaine Ribeiro de Faria Paiva (UFU)

Prof. Henrique Coelho Fernandes (UFU)

## **Equipe de Trabalho**

Beatriz Ribeiro Borges

Bruno de Paula Nery

Carlos Humberto Martins Junior

Cássio de Souza Campos

Dahlan Pereira Gardim

Gabriel Teodoro Ribeiro

Giullia Rodrigues de Menezes

Higor Raphael Faria e Sousa

João Barbosa de Souza Neto

Lucas Gabriel Teodoro Araújo

Lucas Guimarães Mendes

Marcelo Mendonça Borges

Marcus Vinícius Torres Silva

Mateus Ferreira Silva

Matheus Henrique Onofre de Oliveira

Murielly Oliveira Nascimento

Maycon Douglas Batista dos Santos

Paulo Henrique Alves Santos

Paulo Kiyoshi Oyama Filho

Rodrigo Castro Vieira Gomes

Rodrigo Zamboni Silva

Yan Stivaletti e Souza

## **Editoração dos Anais**

Profa. Elaine Ribeiro de Faria Paiva (UFU)

Prof. Henrique Coelho Fernandes (UFU)

# Contents

<b>I TRABALHOS DE GRADUAÇÃO</b>	<b>1</b>
Estudos de algoritmos e estruturas de dados compactas para grafos temporais . . . .	2
Sistema para auxiliar na certificação de pulverizadores agrícolas . . . . .	3
Aula em vídeo: uma nova forma de levar o conhecimento aos idosos . . . . .	4
Análise de microdados do censo INEP de educação superior utilizando técnicas de aprendizado de máquina . . . . .	5
Criação de um método de ensino de programação baseado no perfil dos melhores programadores jovens do Brasil . . . . .	6
Análise das discussões sobre problemas de segurança no Stack Overflow usando modelagem de tópicos . . . . .	7
Plataforma DebugandoEd auxiliando nas estruturas de dados . . . . .	8
Identificação de ataques não-conhecidos em sistemas de detecção de intrusão baseados em anomalia . . . . .	9
Classificação visual de situação e eventos de trânsito veicular baseado em texturas dinâmicas e redes complexas . . . . .	10
Reforma da previdência no Brasil: uma análise a partir de dados do Twitter . . . . .	11
Uma biblioteca peso-leve para Blockchain . . . . .	12
Comparação de algoritmos de aprendizado de Máquina na classificação de neoplasias mamárias . . . . .	13
Redes Bayesianas para previsão de doadores de sangue . . . . .	14
Avaliação de entradas e funções de recompensa para aprendizado de máquina por reforço para agentes jogadores de videogames . . . . .	15
<b>II TRABALHOS DE MESTRADO</b>	<b>16</b>
Análise do desempenho acadêmico de estudantes universitários utilizando técnicas de visualização de informação . . . . .	17
Programação linear aplicada à otimização de dietas para portadores de diabetes . . . .	18
Avaliação de aspectos de acessibilidade em sistemas integrados de gestão empresarial com desenvolvedores, analistas e usuários finais . . . . .	19
Detecção de linhas de plantio em plantações de cana-de-açúcar utilizando deep learning	20
Entendendo a classificação multiclasse de linfomas através de descritores morfológicos e não-morfológicos . . . . .	21
<b>III TRABALHOS DE DOUTORADO</b>	<b>22</b>
Automatic classification of Acute Lymphoblastic Leukemia with Deep Residual Neural Network improved with Genetic Algorithm . . . . .	23
Arcabouço computacional para o mapeamento das áreas cultivadas em solo brasileiro	24
Integração de dados oriundos de registro de imagens para classificação de imagens por meio de redes neurais profundas . . . . .	25

## **Chapter I**

# **TRABALHOS DE GRADUAÇÃO**

## Estudos de algoritmos e estruturas de dados compactas para grafos temporais

Mateus F Silva (Universidade Federal de Uberlândia)\*; Marcelo Albertini (Federal University of Uberlandia); Bruno A. N. Travençolo (Universidade Federal de Uberlândia)

mateus.silva1@ufu.br\*; albertini@ufu.br; travencolo@gmail.com

**Resumo:** Um grafo é um tipo abstrato de dados que é amplamente utilizado para modelar diversos cenários e uma das suas variações é o grafo temporal que estende as características de um grafo e adiciona a dimensão do tempo, assim possibilitando consultas com base no tempo. Por exemplo, um grafo pode ser utilizado para mapear o contato de uma comunidade a um vírus, onde cada indivíduo seria representado por um vértice  $v$ , e cada relação de contato entre indivíduos seria uma aresta  $e$ . Assim, seria possível realizar consultas para saber se um indivíduo  $X$  está em contato com um indivíduo  $Y$  que está contaminado com o vírus, não é possível saber quando este contato ocorreu, pois o grafo não contempla a dimensão do tempo. Seguindo o exemplo anterior, caso um grafo temporal fosse utilizado haveria a possibilidade de realizar consultas com base em um intervalo de tempo, então poderíamos saber quando o contato entre os dois indivíduos aconteceu e, além disso, também poderíamos saber se houve múltiplos contatos ao longo do tempo. Grafos podem ser representados de diversas formas e cada representação possui uma complexidade de espaço distinta. Estruturas de dados mais comuns para se armazenar grafos são: listas de adjacências, matrizes de adjacência e lista de arestas. Em listas de adjacência, o custo de espaço é  $(-V- + -E-)$ . Matrizes de adjacência custam  $(-V-^2)$  de espaço. Listas de arestas custam  $(-E-)$  de espaço. Para estimar o custo de espaço de um grafo temporal, podemos multiplicar o uso de cada estrutura de dados por  $T$ , sendo  $T$  a quantidade de instantes de tempo considerada no grafo. A quantidade de informação produzida hoje é cada vez maior e para que consigamos armazenar e consultar essas informações precisamos de estruturas de dados mais eficientes no uso de memória, do que as que são utilizadas majoritariamente atualmente. Uma ferramenta que pode nos ajudar com isso são as estruturas de dados compactas, que são estruturas que gastam menos espaço que as estruturas de dados convencionais e sem aumento significativo no tempo de execução original. Fica evidente que o estudo e implementação de algoritmos e estruturas de dados compactas para grafos temporais deve ser levado em consideração para que possamos tornar mais eficiente nossas computações e viabilizar processamento em conjuntos de dados cada vez maiores. Diante dos fatos supracitados, este trabalho tem como objetivo de implementar estruturas de dados compactas para grafos temporais, sem perdas significativas no desempenho por meio de um estudo aprofundado no trade-off entre espaço e tempo.

**Trilha:** Trabalho de Graduação



## **Sistema para auxiliar na certificação de pulverizadores agrícolas**

Ronaldo Alves Pereira Filho (Universidade Federal de Uberlândia)\*; Thiago Pirola Ribeiro (Federal University of Uberlândia - UFU); Cleyton Alvarenga (Universidade Federal de Uberlândia)

ronaldoalves@ufu.br\*; thiagopirola@gmail.com; cleytonalvarenga@ufu.br

**Resumo:** A agricultura, que antes era tratada como algo realizado por pessoas com pouca instrução, modificou-se ao longo do tempo. Atualmente, a agricultura está cada vez mais tecnológica com a utilização de tratores com GPS para a plantação ocorrer com um erro máximo de 2 cm ou até a utilização de tratores guiados remotamente e auxiliados por GPS para a locomoção e utilização sem operadores. Ao mesmo tempo que os produtores estão se especializando para obter maiores e melhores produções, os compradores, principalmente estrangeiros, impõem exigências cada vez maiores. Dentre essas exigências estão as certificações que a unidade produtora deve ter: certificações sanitárias, certificação ISO 9001, Programa 5S de qualidade, além de exigirem rastreabilidade dos produtos. Para auxiliar na obtenção dessas certificações, as máquinas agrícolas utilizadas na propriedade devem ser inspecionadas e suas condições de uso e conservação devem ser certificadas. Atualmente, o levantamento dos dados é realizado de forma manual, em formulários de papel, o que leva um longo e demorado processo para realizar a inspeção de várias máquinas. O formulário atual conta com 216 itens a serem avaliados, sendo que alguns deles possuem até 50 subitens. O presente projeto desenvolveu um sistema para armazenar e gerenciar as informações coletadas por dispositivos móveis (tablets), gerando listagens dessas informações, visando inicialmente auxiliar na avaliação das condições de uso e conservação de pulverizadores agrícolas. Após o levantamento de requisitos, percebeu-se a necessidade da criação e modificação das questões do formulário pelo usuário aplicador e, a internacionalização do sistema, uma vez que na Europa, os governos obrigam a prática de inspeção periódica de pulverizadores e, com isso, o sistema foi implementado para 3 idiomas nativos: português, inglês e espanhol. Durante a avaliação, pode ocorrer de não ter acesso à internet e com isso, o sistema móvel armazena a avaliação e em momento posterior, envia ao sistema no servidor para o processamento. Após a fase de testes, o sistema será disponibilizado para utilização pública e gratuita, porém o usuário deverá conceder permissão para que se possa utilizar os dados coletados (em conformidade com a LGPD) para o diagnóstico da utilização do sistema nas diversas propriedades. Os testes iniciais, demonstraram que o sistema tem grande potencial, pois conseguiu-se informatizar e acelerar todo o processo. As avaliações agora podem ser conduzidas com o auxílio de aplicativos para dispositivos móveis e compiladas no sistema desenvolvido e, que por sua vez tornam a inspeção mais ágil e com menos erros para as equipes avaliadoras.

**Trilha:** Trabalho de Graduação

## **Aula em vídeo: uma nova forma de levar o conhecimento aos idosos**

Giullia R Menezes (Universidade Federal de Uberlândia)\*; Renan G Cattelan (Federal University of Uberlândia)

giulliar@gmail.com\*; renan@ufu.br

**Resumo:** A exclusão digital é um problema que afeta parte significativa da população, com incidência maior entre o público idoso, justamente um dos segmentos que mais cresce considerando-se o aumento da expectativa de vida. Visando auxiliar o público idoso a se adaptar às novas tecnologias e poder delas usufruir, o CompPET desenvolve diversas iniciativas para promover sua inclusão digital. Em geral, a metodologia consiste na oferta de oficinas voltadas ao público idoso, com as seguintes características: flexibilidade parcial de conteúdo; material didático próprio, gratuito e livremente disponível, complementado por conteúdo criado por especialistas do NIC.br e voltado ao público idoso; aulas dinâmicas, com colaboração em pares; aproveitamento de infraestrutura dos laboratórios da universidade; emprego de recursos de acessibilidade; e formação de parcerias. Este projeto tem procurado renovar tal metodologia, buscando melhorar o aprendizado e permitindo que o conteúdo repassado seja realmente absorvido e que os idosos se capacitem para utilizar a tecnologia em seu cotidiano. O modelo de ensino passou por mudanças com a inclusão de jogos, mas com a chegada da pandemia, essa atividade não pode ser realizada de forma presencial, e com a tentativa de renovar esse modelo de ensino, considerando as ferramentas disponíveis e a situação presente, notou-se a necessidade de disponibilizar o conteúdo de forma diferente. A ideia da iniciativa é a produção de materiais em vídeo ensinando a utilização de aplicativos presentes em smartphones. O piloto foi a gravação de vários vídeos de curta duração (de 2 a 4 minutos) ensinando a criar e manusear uma conta em um aplicativo utilizado para comunicação, o WhatsApp. Foi feito um passo a passo ensinando atividades como criar uma conta, adicionar uma foto ao seu perfil, salvar contatos para futuras conversas envolvendo o aplicativo, enviar mensagens de texto, enviar mensagens de áudio, enviar fotos e escutar áudios recebidos. Acredita-se que apenas com o conteúdo apresentado nos vídeos, o idoso já consiga se comunicar de forma natural utilizando o aplicativo. Um novo desafio para continuidade e aperfeiçoamento do projeto é a disponibilização desses vídeos em massa, além de análise da qualidade do aprendizado de quem acessa esse conteúdo. Vale ressaltar também, que se os resultados apontarem números positivos, a criação de novos vídeos com novos conteúdos pode ser uma forma de levar mais conteúdo a mais pessoas de maneira mais rápida.

**Trilha:** Trabalho de Graduação

## **Análise de microdados do censo INEP de educação superior utilizando técnicas de aprendizado de máquina**

Dahlan P Gardim (Universidade Federal de Uberlândia)\*

dahlanpg@ufu.br\*

**Resumo:** A evasão escolar é um dos fatores mais preocupantes e desafiadores no contexto do sistema educacional como um todo, pois representa a total frustração dos objetivos educacionais. Contudo, para além de uma análise puramente quantitativa, deve-se perceber que, em se tratando de um retrato da realidade, é necessário também relacionar qualitativamente os números aos sujeitos, ou seja, explorar essa questão de forma a delimitar cuidadosamente suas causas, eliminando ambiguidades e tornando o resultado mais compreensível. Nota-se que, apesar de uma noção inicial apreendida dos dados brutos, pelo censo do INEP, ela comumente apresenta lacunas para um entendimento mais fidedigno, isto é, a forma padrão dos estudos sobre evasão escolar, consideram em geral a porcentagem bruta que não se vale situações específicas, como aqueles que são matriculados em mais de um curso de Ensino Superior e que optam por um ou outro no decorrer de ambos, o que seria contabilizado de forma equivocada como evasão escolar. Entende-se que é preciso estabelecer um cruzamento entre os dados dos censos ao longo do tempo, isto é, para saber as influências para um determinado grupo em análise sobre suas decisões, considerando evasões. Dos objetivos: Selecionar e organizar os dados, relativos à UFU, disponibilizados no site do INEP, compreendidos entre 2009 e 2015; Analisar estatisticamente a base de dados e seu comportamento; Utilizar ferramentas de aprendizado de máquina para identificar padrões e correlações entre os atributos; Buscar correlações desta pesquisa com outros trabalhos para ampliar suas aplicações. A metodologia utilizada consiste no processo de Descoberta de Conhecimento em Base de Dados (DCBD), descrito por Fayyad, Piatetsky-Shapiro e Smyth. A DCBD explora diversos ramos da Inteligência Artificial, como o Aprendizado de Máquina, Reconhecimento de Padrões, Sistemas Inteligentes, além de disciplinas como Estatística e Matemática. De acordo com os 694886 dados de entrada colhidos, cerca de 37139 alunos evadiram e 7187 trocaram de curso. Isto define as seguintes condições: 1) Porcentagem dos que trocaram curso (não-evasão) pelo Total: 1.03%; 2) Porcentagem dos que perderam vínculo (evasão) pelo Total: 5.34%; 3) Porcentagem dos que trocaram em relação à evasão de fato: 19.35%. Percebe-se que estes resultados já indicam uma oportunidade para se explorar demais contextos nesse sentido. Por exemplo, na 8ª edição do Mapa do Ensino Superior foi apontado 18,5% de evasão na rede pública no país. Alarmante, mas ainda possui métodos de análise pouco criteriosos. Assim, conforme os resultados, a proporção de 1 que não perdeu vínculo para 5 evadidos de fato, este estudo pode contribuir para se obter resultados mais condizentes com a realidade, bem como enseja uma oportunidade de se identificar as causas de uma alta taxa de trocas, ou seja, de forma mapear características que levam às trocas e evasões, contribuindo, assim, com melhor entendimento sobre os dados na educação brasileira.

**Trilha:** Trabalho de Graduação

## **Criação de um método de ensino de programação baseado no perfil dos melhores programadores jovens do Brasil**

Giullia R Menezes (Universidade Federal de Uberlândia)\*; João Henrique de Souza Pereira (Universidade Federal de Uberlândia); Luiz Theodoro (Universidade Federal de Uberlândia)

giulliam@gmail.com\*; joaohs@ufu.br; luiz.theodoro@ufu.br

**Resumo:** O Trabalho de Conclusão de Curso visa o desenvolvimento de um método de ensino de programação baseado nos resultados de uma coleta de dados sobre a rotina de jovens talentos que foram medalhistas na Olimpíada Brasileira de Informática (OBI) e sua aplicabilidade em um Projeto de Ensino de Programação para Alunos do Ensino Médio e Fundamental (PEPAEMF). Em 2019 na Semana Olímpica que ocorreu na UNICAMP, foi feita uma coleta de várias informações dos premiados na OBI 2019 de modo a conhecer um pouco mais sobre seu perfil e sua rotina de estudos. Após a coleta desses dados, foi feita uma análise e discussão dos pontos observáveis e relevantes. Um próximo passo para esse estudo, foi utilizar o que foi coletado para avaliar a aplicabilidade de alguns pontos relevantes na rotina de alunos de um projeto local (PEPAEMF), que utilizava o mesmo método sem alterações a algum tempo. Os resultados obtidos na coleta de dados na Semana Olímpica como faixa etária, forma e plataforma de estudos, quantidade de horas, dias de estudo e definição de cronograma de estudos foram de suma importância para identificar pontos que necessitavam de mudanças no método de ensino tradicional do projeto citado. Foram feitas 3 aplicações para comparação de resultados entre elas, a primeira envolvia o método de ensino tradicional do projeto (nomeado como modelo PEPAEMF), a segunda com algumas modificações considerando resultados obtidos na Semana Olímpica (nomeado como PEPAEMF 1.1) e a terceira foram adaptações de pontos críticos observados na segunda aplicação (nomeado como PEPAEMF 1.2). Como benefício visivelmente observado em um primeiro momento, com a implantação das mudanças sugeridas no método PEPAEMF 2.1 comparado ao método tradicional PEPAEMF a participação de alunos na maratona no final do último ciclo teve um crescimento de aproximadamente 400% e um crescimento de aproximadamente 150% na média de exercícios resolvidos por pessoa. Como resultado foi possível concluir que o acompanhamento e apoio ao aluno durante todo o seu processo de aprendizado é essencial e indispensável. Além disso, o projeto obteve um recorde de números de medalhistas na edição da OBI 2021, totalizando 5 alunos, entre eles alunos do projeto ou que já fizeram parte dele. Um novo desafio para a continuidade dessa pesquisa é aprimorar o método e aplicar em um maior número de alunos do país como um todo para análise de novos resultados.

**Trilha:** Trabalho de Graduação

## **Análise das discussões sobre problemas de segurança no Stack Overflow usando modelagem de tópicos**

Lucas G Mendes (Universidade Federal de Uberlândia)\*; Marcelo Maia (UFU); Rodrigo Miani (Federal University of Uberlandia)

lucas.guimaraes@ufu.br\*; marcelo.maia@ufu.br; miani@ufu.br

**Resumo:** Nos últimos anos, vivenciou-se um período de diversos ataques cibernéticos de grandes proporções, tanto em sistemas governamentais como em sistemas de grandes empresas em todo o mundo. Este ocorrido evidencia a necessidade urgente do debate, pesquisa e difusão de conhecimentos e boas práticas associadas à segurança da informação. Uma plataforma útil para tais necessidades são os fóruns de discussão como o Stack Overflow, cujo público alvo é voltado para programadores, e que também podem ser utilizados para discutir problemas sobre segurança da informação. Sendo assim, o presente trabalho tem como objetivo analisar e compreender a dinâmica das discussões sobre problemas de segurança no Stack Overflow e estudar o comportamento dos usuários em relação à assuntos de cibersegurança, analisando a reação da comunidade acerca das perguntas na plataforma. Trabalhos anteriores analisaram questões sobre segurança em tal plataforma, estudando como a compreensão de segurança e práticas seguras são desenvolvidas e disseminadas entre os profissionais neste ambiente on-line. Como diferencial, este estudo terá escopo mais amplo, de forma a abranger perguntas de diferentes assuntos e compará-las com perguntas relacionadas à segurança da informação. Para tanto, modelou-se um banco de dados contendo todas as postagens do Stack Overflow e posteriormente levantou-se estatísticas básicas sobre a base de dados. O próximo passo será a aplicação de técnicas e métodos de modelagem de tópicos como o Latent Dirichlet Allocation (LDA) e Open Coding. Essas técnicas permitirão criar uma classificação dos posts de acordo com o seu conteúdo. Isso fornecerá subsídios para entender, entre outras coisas, se perguntas que não são respondidas estão relacionadas a determinados conteúdos específicos. Apesar do trabalho ainda estar em andamento, alguns resultados preliminares indicam que as perguntas relacionadas à segurança da informação demoram mais para serem respondidas em comparação com as perguntas sobre outros assuntos. Além disso, essas perguntas também possuem menos respostas.

**Trilha:** Trabalho de Graduação

## Plataforma DebugandoEd auxiliando nas estruturas de dados

Lucas F Borges (UFU)\*; Ana Cláudia Martinez (UFU); Thiago Pirola Ribeiro (Federal University of Uberlândia - UFU)

lucasfernado84@gmail.com\*; aclaudia.martinez@gmail.com; thiagopirola@gmail.com

**Resumo:** O mundo atual está cada vez mais tecnológico e as tecnologias estão presentes cada vez mais no dia a dia. Na sala de aula, tentar conciliar o uso das tecnologias acessáveis pelos estudantes para que se consiga estimular o aprendizado, tornar os conteúdos motivadores, atrativos e mais compreensíveis, é um grande desafio. Com o foco na área de computação em seus diversos cursos, analisou que a literatura apresenta diversos trabalhos em que se optou pelo uso de simuladores para que os alunos consigam entender as abstrações presentes nos conteúdos de forma mais tangível e visual. Com isso, na área de estruturas de dados existem diversos simuladores citados na literatura, porém em nenhum dos analisados, contém de fato uma interação totalmente livre para a manipulação das estruturas pelos usuários. Partindo dessa necessidade desenvolveu-se uma plataforma didática na web focada em estruturas de dados que permite a interação do usuário, tendo sido desenvolvida utilizando as tecnologias: HTML 5, CSS 4, Javascript, PHP 7.4 e PostgreSQL 12.3. Na plataforma desenvolvida é possível manipular estruturas como vetores, matrizes, pilhas, filas, ponteiros e listas simples e duplamente encadeadas. Escolhendo uma dessas estruturas, seleciona-se o tamanho e logo em seguida é permitido a inserção ou remoção de valores a depender da estrutura. Todas essas interações são visualizadas de forma gráfica, assim como o código em linguagem C demonstrando a implementação do que foi executado, além de uma memória simulada na qual os dados são armazenadas de forma que se assemelha a um “debug” sem a necessidade da utilização de um Ambiente de Desenvolvimento Integrado (IDE) ou exigir a implementação do código. Foi implementada uma área para que o usuário (restrita por login) possa salvar e recuperar as estruturas simuladas por ele anteriormente, permitindo que ele consiga continuar seus estudos posteriormente. Para o salvamento e recuperação das informações, é necessário a realização de um pequeno cadastro (baseado na LGPD) que permitirá o login na plataforma. Estando logado, é permitido ao usuário a mudança entre estruturas sem perda das informações inseridas anteriormente, e ao voltar a mesma estrutura, a plataforma garante a integridade das informações oportunizando a ele a continuidade na interação, mantendo a linha de raciocínio para os estudos que estão sendo realizados. Percebe-se com os testes iniciais que a plataforma atende às necessidades da utilização baseada nos conteúdos abordados em sala de aula. Para aumentar a motivação e engajamento dos usuários, está sendo estudada a inserção da gamificação na plataforma por meio da adoção de desafios e missões, pontuações, fases, níveis, medalhas e conquistas.

**Trilha:** Trabalho de Graduação

## Identificação de ataques não-conhecidos em sistemas de detecção de intrusão baseados em anomalia

Miguel S Rocha (UFU)\*; Rodrigo Miani (Federal University of Uberlandia); Gustavo Bernardo (UFU); Luan Mundim (UFU)

miguel.sanches98b@gmail.com\*; rodrigomiani@gmail.com; gustavo.bernardo@gmail.com; luanmundim@gmail.com

**Resumo:** Este projeto de pesquisa é motivado pelo crescente número de ataques não-conhecidos que se disseminam pela Internet visando infectar sistemas computacionais. Nesse sentido, o projeto se consolida como uma fonte de conhecimento científico capaz de prover informações relevantes acerca de como os Sistemas de Detecção de Intrusão (IDSs) baseados em anomalia se comportam perante a estes ataques, a fim de demonstrar a eficácia de tais sistemas ao serem utilizados na estrutura de segurança de um ambiente computacional. O objetivo do trabalho é analisar e comparar o desempenho de IDSs baseados em anomalias e desenvolvidos com técnicas de aprendizado de máquina, para que seja possível ter clareza sobre seus comportamentos perante ataques não-conhecidos. Por fim, espera-se que, ao final do projeto, seja possível identificar os tipos de ataques não-conhecidos mais propensos a serem detectados assim como os algoritmos de aprendizado de máquina com os melhores desempenhos em tal tarefa. Até o momento, algumas atividades de vital importância foram realizadas de modo a garantir a confiabilidade nos resultados do projeto. Primeiramente, foi feito um estudo detalhado sobre o conjunto de dados usado (CIC-IDS2017) com o objetivo de entender como os dados estavam estruturados e analisar os atributos escolhidos para descrever os fluxos de pacotes. Deste modo, foi possível fazer alguns ajustes no conjunto de dados como retirar dados pouco relevantes para que eles não afetassem o desempenho dos algoritmos de Machine Learning na tarefa de identificação dos ataques. Após isso, diversos algoritmos como “Random Forest”, “Decision Tree”, “Logistic Regression”, entre outros, foram submetidos a treinos e testes e tiveram seus resultados comparados, para que fosse possível destacar os melhores resultados em termos de precisão, revocação e taxa de falsos negativos. Apesar de estar em fase inicial, o projeto já apresenta alguns dados interessantes com relação aos algoritmos usados. Tem-se- que é difícil obter bons resultados quando se realiza testes com ataques que não foram usados para o treinamento dos modelos. Contudo, algumas análises mostram que o algoritmo “Random Forest” se sobressai e possui resultados melhores quando comparado com modelos até mais complexos como redes neurais. O trabalho agora será voltado à análise minuciosa do “Random Forest”, de modo a alterar parâmetros do algoritmo como “n\_estimators” (número de árvores na floresta) e submetê-lo a diferentes bases de treino e teste. Esse processo será importante para realizar uma verificação profunda da eficiência deste classificador ao lidar com ataques não-conhecidos. Por fim, os modelos podem ser submetidos a outros conjuntos de dados com o objetivo de se obter mais uma forma de validação e análise de seus desempenhos ao fazer parte da segurança de um ambiente computacional.

**Trilha:** Trabalho de Graduação

## **Classificação visual de situação e eventos de trânsito veicular baseado em texturas dinâmicas e redes complexas**

Arthur Henrique C Rosa (Universidade Federal de Uberlândia)\*; Leandro Couto (Universidade Federal de Uberlândia)

arthur.rosa@ufu.br\*; leandronc@ufu.br

**Resumo:** A utilização de visão computacional para realizar classificação de situação de trânsito e a detecção de eventos está recebendo forte atenção em pesquisas ao redor do mundo. Entre os motivos está o desenvolvimento de uma mobilidade moderna e sustentável, tornando necessário o aprimoramento da logística e do trânsito nas cidades. Nesse contexto, a automação da classificação de situações e eventos de trânsito terrestre possibilita aos órgãos reguladores, entender o comportamento dos motoristas nas vias e os motivos que causam os acidentes e os congestionamentos. Assim, possibilitando a tomada de decisões regulatórias e urbanísticas que visem sanar os problemas identificados de forma racional. Este projeto tem como objetivo principal realizar a classificação de níveis de tráfego de veículos terrestres com a utilização de métodos de visão computacional e processamento de imagens. Nossa proposta utiliza o método de Fluxo Óptico para estimar o deslocamento de veículos entre dois quadros da imagem através do somatório da magnitude de todos vetores de fluxo óptico. Também utilizamos o YOLOv3, um modelo de aprendizado de máquina que realiza a detecção de objetos e sua localização em uma imagem, computamos a quantidade de objetos identificados na imagem por quadro do vídeo. Os dois métodos em conjunto foram aplicados a uma base de dados bem conhecida da literatura para treinamento de um classificador de tráfego de veículos. Testes foram aplicados a base de vídeos de tráfego Trafficdb (254 vídeos, divididos em 3 classes representando tráfego leve, médio e pesado). O classificador SVM obteve o melhor desempenho com acurácia média de 90,4% pelo método da magnitude do fluxo óptico, e acurácia média de 84,5% com a contagem de carros do método YOLOv3, resultado que inclusive se manteve proporcional a outros projetos encontrados na literatura. Analisando a literatura foi possível encontrar projetos que utilizaram o mesmo banco de imagens utilizado neste trabalho, os resultados foram expressivos para a aplicação em Deep Learning que alcançou a acurácia de 98%, enquanto a aplicação que utiliza o detector de objetos Haar Cascade obteve a acurácia de 84%, dessa forma é notável a superioridade do método baseado em Fluxo Óptico em relação ao Haar Cascade, no entanto é expressiva a preeminência do Deep Learning aos métodos que utilizam detectores de objetos. Dois métodos de visão computacional foram implementados para realizar a detecção de eventos de trânsito em um banco de imagens e ambos obtiveram bons resultados, contudo o Fluxo Óptico é o destaque por sua acurácia média de 90,4% quando testado no classificador SVM. Além da acurácia superior, o Fluxo Óptico também se sobressai em tempo de processamento, chegando a ser mais de 35 vezes menor que o tempo de processamento do YOLOv3. Como trabalho futuro é uma linha potencial buscar outros descritores baseados em fluxo óptico além da magnitude total dos fluxos.

**Trilha:** Trabalho de Graduação



## Reforma da previdência no Brasil: uma análise a partir de dados do Twitter

Matheus H. F. Protásio (Federal University of Uberlândia)\*; Paulo H R Gabriel (Federal University of Uberlândia)

matheushfp@ufu.br\*; phrg@ufu.br

**Resumo:** A reforma da previdência votada e aprovada no ano de 2019 foi uma das pautas mais relevantes na política brasileira nos últimos anos. Dada a relevância do tema, é importante analisar a opinião dos cidadãos sobre o mesmo. Isso pode ser feito utilizando dados provenientes de redes sociais que, com o advento da inclusão digital, cada vez são mais utilizadas pelos brasileiros com o passar dos anos. A rede social mais conhecida pelo compartilhamento rápido de notícias e opiniões na forma de curtos textos é o Twitter, onde esses pequenos textos são chamados tuítes. Neste trabalho foi feita uma análise qualitativa dos dados de uma base já montada com tuítes sobre a reforma da previdência no Brasil no ano de 2019. Foi realizada a classificação dos usuários da base, distinguindo usuários comuns de contas automatizadas (conhecidas como bots), esta análise foi feita utilizando a ferramenta Botometer que classificou 8.180 usuários como bots dentre os 237.894 usuários únicos da base. Com a classificação de usuários realizada, foi analisada a possibilidade de uma “rede” de propagação de mensagens ter sido utilizada para propagar uma determinada opinião a respeito do tema, tudo levou a crer que não houve utilização deste tipo de “rede”, pois os bots produziram 40.151 tuítes dos 980.577 tuítes da base. Também foram realizadas classificações dos tuítes, por meio da análise de sentimentos, onde as classificações podiam ser: “positivo”, “negativo” ou “neutro”. Para a realização destas classificações foram utilizados quatro modelos: Naïve Bayes, Regressão Logística, Floresta Aleatória e Máquinas de Vetores de Suporte (SVM). Estes modelos possibilitaram uma comparação de resultados com os resultados obtidos pelo criador da base, pois ele os utilizou em seu trabalho para classificar os tuítes. Com relação à classificação dos tuítes produzidos pelos bots, para todos os modelos utilizados, a maioria dos tuítes recebeu classificação neutra. Algo que pode ter levado a este resultado é o fato de que muitos veículos de comunicação utilizam contas automatizadas para replicação de notícias que foram publicadas em seus sites, assim muitas contas de veículos de comunicação foram consideradas bots nesta análise. Na etapa de pré-processamento da base também houve o tratamento de hashtags, algo que não foi realizado na classificação feita pelo criador da base em seu trabalho. Em seguida foram comparados os novos resultados obtidos após o tratamento de hashtags com os resultados originais e foi possível observar que tanto para a classificação original, como para a nova classificação, Naïve Bayes classificou a maioria dos tuítes como positivos, já os demais modelos classificaram a maioria dos tuítes como neutros. Este comportamento se repetiu em outras análises realizadas na base, como por exemplo na classificação dos mil tuítes mais favoritos.

**Trilha:** Trabalho de Graduação

## Uma biblioteca peso-leve para Blockchain

Guilherme Henrique d de Araújo Santos (Universidade Federal de Uberlândia)\*; Ivan da Silva Sendin (Universidade Federal de Uberlândia)

guilhermeher10@ufu.br\*; sendin@ufu.br

**Resumo:** A notável ascensão das criptomoedas levou pesquisadores do mundo todo a querer descobrir o porquê de sua popularidade rápida e da fama de confiabilidade e segurança em suas transações. O funcionamento de seus algoritmos são baseados no armazenamento das informações em blockchains, que torna a informação "imutável" e descentralizada. Para a geração de novas moedas virtuais são utilizados mineradores, que são usuários na rede que buscam resolver problemas computacionais para organizar as informações contidas em cada novo bloco que será inserido na blockchain e como recompensa pela resposta encontrada, o minerador recebe um "pagamento" em criptomoedas pelos recursos gastos (energia elétrica, CPU, GPU, etc.). O objetivo deste trabalho é produzir um biblioteca peso-leve, simples, de fácil instalação, que consuma poucos recursos computacionais para simular uma blockchain de PoW, com aplicações educacionais e científicas. O experimento do conceito prova de trabalho (proof of work) que será apresentado é uma simulação de PoW porém sem a recompensa pelo trabalho, ele tem o intuito de realizar a criação de uma blockchain composta por uma base de mineradores que possuem como principal característica o poder computacional. A soma do poder computacional de todos os mineradores é definida como poder mundial, ao descobrir o poder mundial é possível realizar uma "loteria", um sorteio para a escolha do minerador que irá efetuar a mineração de um bloco que será inserido na blockchain e para a escolha do minerador por meio da loteria tem-se a influência do poder computacional como um diferencial de prioridade, ou seja, quanto maior o poder computacional de um minerador, maior será sua chance de minerar um bloco. Observa-se que o poder computacional do minerador possui razão proporcional com a quantidade de blocos que ele vai minerar. Essa proporcionalidade não é perceptível ao minerar-se poucos blocos, porém ao executar milhares de minerações fica evidente o quanto o poder computacional influencia blockchains que usam como algoritmo o PoW.

**Trilha:** Trabalho de Graduação

## Comparação de algoritmos de aprendizado de Máquina na classificação de neoplasias mamárias

Tulio Araujo Araujo Santos De Oliveira (Universidade Federal de Uberlândia)\*; Fernanda Santos (UFU)

tulio.oliveira@ufu.br\*; fmcsantos@ufu.br

**Resumo:** O câncer de mama é um problema relevante de saúde pública no Brasil e à nível mundial, o que justifica as intensas pesquisas e esforços para aumentar a expectativa de vida dos pacientes. O fator crucial para uma maior expectativa de vida dos pacientes é a detecção precoce da doença. A mamografia é uma das melhores técnicas para o rastreamento do câncer de mama disponível atualmente, capaz de registrar imagens da mama com a finalidade de diagnosticar a presença ou ausência de estruturas que possam indicar a doença. Aliado à tais metodologias tradicionais da Medicina, sistemas computacionais contribuem substancialmente para o diagnóstico precoce do câncer de mama, ainda mais se tais sistemas fizerem uso de algoritmos de Aprendizado de Máquina para melhorar as taxas de detecção. Portanto, o objetivo deste trabalho é criar um classificador de padrões, constituído por algoritmos de Aprendizado de Máquina (AM) e pela Análise Componentes Principais (PCA), com o propósito de analisar e avaliar a base de dados Wisconsin e, por fim, classificar as neoplasias mamárias. A base de dados utilizada foi Wisconsin Diagnostic Breast Cancer, uma base de dados pública disponibilizada no site UC Irvine Machine Learning Repository e que contém atributos quantitativos extraídos de mamografias, parâmetros estatísticos calculados a partir dos próprios atributos e a classificação da citologia entre maligno ou benigno. Os motivos pela escolha dela deve-se a autenticidade dos dados e sua ampla citação em trabalhos acadêmicos. Os dados são compostos por 569 instâncias divididos entre a classe maligno, compreendendo 357 amostras, e a classe benigno, 212 amostras. A técnica PCA atuou no pré-processamento dos dados, pois agrupou os indivíduos da base de dados de acordo com sua variação auxiliando na identificação dos padrões. Os métodos de AM escolhidos foram Redes Neurais Artificiais, Support Vector Machine, Naive Bayes e Árvore de Decisão. A eficiência dos algoritmos de AM foram analisadas e calculadas pelas métricas quantitativas acurácia, precisão, revocação e medida F. Ademais, utilizou-se da curva ROC (Receiver Operating Characteristic) para fazer um comparativo dos resultados gerados pelos algoritmos de AM. Os resultados apresentaram um desempenho satisfatório para os classificadores constituídos pelas Redes Neurais Artificiais e pelo Support Vector Machine, apresentando acima de 90% de acertamentos para as métricas de avaliação acurácia, precisão, revocação, medida-F1 e área ROC.

**Trilha:** Trabalho de Graduação

## Redes Bayesianas para previsão de doadores de sangue

Cristina Zayra de Nobrega Romani (Universidade Federal de Uberlândia)\*; Fernanda Santos (UFU)

cristina.romani@ufu.br\*; fmcsantos@ufu.br

**Resumo:** Os hemocentros são entidades responsáveis por administrar bancos de sangue para transfusões. Em diversas situações, o hemocentro necessita prever a quantidade de doadores de sangue regulares para garantir o estoque e a rotatividade exigida pelos hospitais. A busca por doadores de sangue regulares é realizada através de vários parâmetros e variáveis, as quais poderiam ser analisadas a partir de uma ferramenta computacional que auxiliaria e aprimoraria a pesquisa. Diante desse cenário, este trabalho apresenta o modelo probabilístico das Redes Bayesianas como um classificador dentre as técnicas de aprendizado de máquinas para a previsão de doadores regulares. Os classificadores empregados foram o Naive Bayes e o Tree Augmented Naive Bayes (TAN). A base de dados explorada foi o "Conjunto de dados do Centro de Serviços de Transfusão de Sangue" disponível no site UCI Machine Learning Repository. O conjunto de dados possui 748 doadores escolhidos de forma aleatória da base de dados do Centro de Serviços de Transfusão de Sangue da cidade de Hsin-Chu em Taiwan. A base de dados possui classes desbalanceadas sendo necessário aplicar as técnicas SMOTE e k-fold estratificado para neutralizar o problema. A validação dos atributos da base de dados foram examinados pelas medidas de associação. As medidas de associação são utilizadas em estudos epidemiológicos e são capazes de quantificar o grau de interferência de uma variável em um determinado ambiente. Logo, em busca de encontrar os atributos da base de dados mais preditivos para definir o perfil de um possível doador de sangue, este trabalho baseou-se nas medidas risco relativo e odds ratio. Os classificadores Bayesianos foram implementados pela linguagem de programação Python, os quais utilizaram recursos da linguagem para criar os algoritmos clássicos definidos na literatura do Naive Bayes e do TAM. A melhor divisão do conjunto de dados para o treinamento e para o teste dos classificadores foi definido pela validação cruzada estratificada, definindo em 3, 5 e 10 o número de folds. Para avaliar os resultados obtidos pelos classificadores Bayesianos, utilizou-se as métricas acurácia, precisão e sensibilidade. Os resultados encontrados demonstraram que o TAN possui maiores taxas de acerto para encontrar um doador de sangue regular. Porém, não foi possível obter melhores resultados devido ao desbalanceamento da base de dados.

**Trilha:** Trabalho de Graduação

## **Avaliação de entradas e funções de recompensa para aprendizado de máquina por reforço para agentes jogadores de videogames**

Mateus de F Rosa (UFU)\*; Leandro N Couto (Universidade Federal de Uberlândia); Murillo Guimarães Carneiro (Universidade Federal de Uberlândia)

mateusfreitasrosa@hotmail.com\*; landovers@gmail.com; mgcarneiro@ufu.br

**Resumo:** A topologia de uma rede neural é um tema complexo e obscuro, pelo fato da rede neural fornecer pouca informação sobre a influência de variáveis, por este motivo ela pertence aos modelos de caixa preta. Recompensas esparsas, também evidenciam um problema na área de pesquisa do aprendizado por reforço, pois os sinais de recompensas precisam ser retro propagados a ações que à resultaram, e a pouca ocorrência destes sinais torna-se difícil a retro propagação a toda cadeia de ações anteriores. Agentes inteligentes capazes de resolver problemas variados, sem a necessidade de adaptação de parâmetros ou estruturas topológicas do modelo e também sem a necessidade de um especialista evidenciam uma área importante na inteligência artificial. O aprendizado por reforço se assemelha ao aprendizado humano, sendo capaz, através de sensores harmoniosos, a se adaptar a mudanças no ambiente e a funções de recompensas. Videogames oferecem ambientes artificiais com complexidades próximas às do mundo real, sendo ótimos simuladores para agentes inteligentes. Devido à capacidade de exploração, planejamento e aprendizado, o agente baseado em aprendizado por reforço é capaz de superar mínimos locais e resolver problemas oferecidos por jogos. A utilização de informações visuais como entrada para a rede neural aproxima ainda mais o aprendizado do agente ao aprendizado de um ser humano ao jogar um mesmo jogo, uma vez que um ser humano teria apenas a informação visual para a tomada de ações. À vista disso, esta pesquisa propõe o desenvolvimento de um agente inteligente baseado no aprendizado por reforço, que seja capaz de jogar mais de um jogo sem alterações na estrutura de rede, e utilizando apenas dados visuais como entrada da rede neural. O agente inteligente foi desenvolvido baseando-se no aprendizado por reforço, utilizando redes neurais convolucionais, algoritmo Q-Learning e informações visuais. Utilizamos a estrutura com três camadas convolucionais e três camadas ocultas de neurônios de 512x256x64. A entrada da rede neural consiste em um pacote com 4 capturas de imagem do jogo com o objetivo de o agente conseguir entender e reagir ao movimento dos objetos na tela. Os resultados parciais obtiveram bom desempenho no jogo Pong, levando o agente a conseguir vencer seu oponente. Em poucas iterações do algoritmo já é possível notar o aprendizado do agente. Nesse treinamento o sinal de recompensa obedece um intervalo de -20 a 20, sendo a pontuação positiva pontos à frente do oponente e a negativa o contrário. Com 200 episódios o agente obteve -15 pontos, com 600 episódios 0 pontos, e finalmente no episódio 800 o agente ultrapassa o oponente, chegando a 5 pontos. Todavia, jogos com recompensas muito esparsas e/ou ambientes que se modificam muito, necessitam de um alto poder computacional para o treinamento do agente. Dessa forma, por limitações computacionais, estamos buscando topologias de rede e metodologias para incorporar jogos mais complexos à pesquisa.

**Trilha:** Trabalho de Graduação

## **Chapter II**

# **TRABALHOS DE MESTRADO**

## **Análise do desempenho acadêmico de estudantes universitários utilizando técnicas de visualização de informação**

Laura Ferreira Marquez (UFU)\*; Jose Gustavo S. Paiva (Universidade Federal de Uberlândia)

lauraecomarquez@gmail.com\*; gustavo@ufu.br

**Resumo:** O estudo do desempenho acadêmico dos alunos no ensino superior é desafiador na área da Educação. Educadores, coordenadores de programas e professores estão interessados em compreender o aprendizado de tópicos específicos, a influência da frequência, nota, bem como de determinados assuntos no aprendizado de outros tópicos, dentre outras análises. As instituições acadêmicas armazenam uma grande quantidade de informação a respeito do desempenho dos alunos nos cursos oferecidos, coletadas ao longo dos períodos letivos. A análise dessas informações pode auxiliar na compreensão da relação entre componentes curriculares e o desempenho acadêmico dos alunos, propiciando a descoberta de possíveis causas das deficiências dos alunos nesses componentes curriculares. Estratégias que abordam a visualização de desempenho acadêmico de estudantes apresentam resultados satisfatórios quanto a análise de notas, faltas anteriores e predição de notas, características dos alunos, interações dos alunos de um curso e padrões de progressão dos alunos. No entanto, essas abordagens possuem algumas limitações, tais como a falta de correlações das informações para identificar as reais causas do fracasso ou sucesso do aluno no curso matriculado. Diante do cenário apresentado, este trabalho propõe o desenvolvimento de uma estratégia de análise visual dos dados de desempenho acadêmico de estudantes universitários, utilizando técnicas de visualização de informação interativas. Acredita-se que o uso de representações visuais que evidenciam ao usuário padrões e tendências a respeito dos fenômenos relacionados ao processo educacional, permitem uma melhor compreensão desses fenômenos e auxiliam na tomada de decisão por parte dos profissionais da educação e, conseqüentemente, na evolução do processo de aprendizado dos alunos. Os layouts utilizarão técnicas para visualização de evoluções ao longo do tempo, correlações entre informações medidas, relacionamento entre alunos, entre outros. Estes layouts tem como objetivo possibilitar análises mais assertivas, bem como a compreensão de deficiências nos componentes curriculares, criando representações visuais que evidenciam ao usuário padrões e tendências que explicam diversos fenômenos relacionados ao processo educacional, permitindo uma melhor compreensão desses fenômenos e conseqüentemente auxiliando na identificação dos principais problemas associados e melhorando a tomada de decisão por parte dos profissionais da área de Educação. Atualmente o projeto se encontra na fase de coleta de dados e preparação para entrevista com profissionais da educação, para colher as principais necessidades, requisitos e tarefas. Em paralelo estão sendo definidos, por meio de pesquisas a trabalhos relacionados, os layouts para a análise visual, juntamente com ferramentas interativas que permitam a exploração dos dados pelo usuário.

**Trilha:** Trabalho de Mestrado

## **Programação linear aplicada à otimização de dietas para portadores de diabetes**

Thiago F Soares (UFU)\*; Mauricio Escarpinati (Facom); Paulo H R Gabriel (Federal University of Uberlândia)

thiagofs1211@gmail.com\*; mauricio@ufu.br; phrg@ufu.br

**Resumo:** Diabetes mellitus (DM) é um distúrbio no metabolismo de carboidratos, gorduras e proteínas resultantes de falhas na secreção e/ou na ação do hormônio insulina no organismo. Como parte do tratamento, indivíduos portadores de diabetes devem seguir uma dieta rigorosa, com quantidades controladas desses nutrientes. Nesse contexto, um dos maiores desafios enfrentados por médicos e nutricionistas é a gestão alimentar de seus pacientes, visto que é necessário respeitar suas necessidades nutricionais ao mesmo tempo em que se deve reduzir o consumo de alimentos que possam provocar um aumento excessivo da glicemia capilar. Além disso, existe a questão da “satisfação alimentar pessoal”, uma vez que tais pacientes devem, apesar de todas as limitações impostas, buscarem uma dieta agradável a seus paladares. Dado esse cenário, é de grande utilidade o desenvolvimento de uma ferramenta computacional que auxilie na tomada de decisão sobre a elaboração de uma dieta completa que se encaixe no perfil de uma pessoa diagnosticada com DM. Tal ferramenta pode colaborar com profissionais de saúde na definição de quais alimentos devem ser selecionados a partir do perfil alimentar e os gostos pessoais de seus pacientes. Este trabalho tem como objetivo o desenvolvimento de um modelo de programação linear matemático (PLM) que permita encontrar uma dieta adequada para portadores de diabetes, respeitando as demandas nutricionais mínimas e máximas exigidas. No desenvolvimento da pesquisa, serão abordados tópicos relacionados a modelagem matemática e otimização. Nesse sentido, serão exploradas diferentes funções objetivo e seu impacto na obtenção de um conjunto apropriado de alimentos. Como entrada, esse modelo deve receber alimentos comumente encontrados na região e cuja composição será coletada a partir de tabelas nutricionais. A resolução do modelo envolverá o uso de algoritmos exatos (por meio de solvers) presentes em ferramentas como o GNU Linear Programming Kit (GLPK) ou a OR-Tools. É importante destacar que diversos trabalhos têm sido realizados em outros países. Neste trabalho, busca explorar alimentos presentes na dieta da população brasileira. Uma boa fonte de informação nutricional é a Tabela Brasileira de Composição de Alimentos (TACO) de 2011, que contempla uma grande variedade de alimentos, de todas as regiões do Brasil e contém as informações nutricionais relevantes para o modelo proposto. Resultados preliminares mostraram a eficiência dessa abordagem para um conjunto de 172 alimentos presentes no Guia Alimentar da População Brasileira.

**Trilha:** Trabalho de Mestrado



## **Avaliação de aspectos de acessibilidade em sistemas integrados de gestão empresarial com desenvolvedores, analistas e usuários finais**

Ellen Diana Guissoni (UFU)\*; André Pimenta Freire (UFLA); Rafael Dias Araújo (UFU)

ellendiana45@gmail.com\*

**Resumo:** O desenvolvimento da tecnologia está avançando a cada dia e os softwares de gestão empresarial se tornam cada vez mais importantes para que as empresas automatizem seus processos. Além disso, existe uma grande população no Brasil e no mundo que tem algum tipo de deficiência e que não podem ser ignoradas, pelo contrário, devem ser incluídas em qualquer ambiente e, por isso, é importante considerar aspectos de acessibilidade ao desenvolver qualquer tipo de software, incluindo nos sistemas de gestão, que são fundamentais para que elas tenham garantidos seus direitos de acesso a sistemas imprescindíveis ao exercício de atividades de trabalho. Assim, este trabalho visa avaliar questões de acessibilidade em sistemas integrados de gestão empresarial, analisar os diferentes níveis de conhecimento sobre o tema entre pessoas que fazem o desenvolvimento desses softwares e, ainda, avaliar a usabilidade de tais sistemas junto a usuários com e sem deficiência, a fim de identificar possíveis fragilidades e sugerir melhorias. Existem poucos artigos publicados sobre acessibilidade em softwares de gestão empresarial, e devido a grande importância da acessibilidade em si, este trabalho tem grande relevância para a comunidade. Esta pesquisa utiliza uma abordagem exploratória para análise quantitativa e qualitativa. Foi realizado um workshop de Design Thinking (duração de cinco dias, sendo uma hora por dia) com os colaboradores de uma empresa da área de tecnologia participante para identificar o nível de acessibilidade do ERP com base nas Diretrizes de Acessibilidade para a Web (WCAG) e elencar as principais implementações que poderiam ser feitas pelos desenvolvedores do sistema para melhorar essa navegabilidade e acessibilidade do sistema priorizadas por impacto versus esforço. Como todos os participantes não tinham nenhum tipo de deficiência, esse workshop causou um grande impacto na experiência de navegar pelo sistema por um leitor de tela. Foram levantadas muitas questões sobre a acessibilidade e como a falta dela pode afetar o bem-estar e até a independência de uma pessoa com deficiência. Após vários impasses, foi concluído ao final do workshop, que a maioria das atividades levantadas são de baixo esforço e alto impacto. Também foram coletados dados sobre percepções sobre a acessibilidade de sistemas ERP baseado na Teoria do Comportamento Planejado que ainda serão analisados. Por fim, também serão realizados testes de usabilidade com pessoas com deficiência utilizando o protocolo think-aloud. Esta pesquisa foi submetida e aprovada pelo Comitê de Ética em Pesquisa com Seres Humanos (CAAE 51611821.2.0000.5148).

**Trilha:** Trabalho de Mestrado

## Detecção de linhas de plantio em plantações de cana-de-açúcar utilizando deep learning

João Batista Ribeiro (Universidade Federal de Uberlândia)\*; André R Backes (Universidade Federal de Uberlândia); Mauricio Escarpinati (Facom)

joao42lbatista@gmail.com\*; arbackes@yahoo.com.br; mauricio@ufu.br

**Resumo:** O rápido crescimento populacional, principalmente no último século, tem impulsionado a demanda por alimentos e a utilização inteligente/sustentável dos recursos naturais. Nesse contexto, a agricultura aliada à tecnologia, denominada de Agricultura de Precisão (AP), busca suprir essa demanda utilizando os recursos sob medida com base nas informações coletadas. As imagens utilizadas na AP têm variadas fontes (e.g., câmeras acopladas em Veículos Aéreos Não Tripulados (VANTs) e satélites) dependendo da aplicação e do Ground Sample Distance requisitado. Uma das principais aplicações da AP é a detecção das linhas de plantio, principalmente porque esta é utilizada como uma etapa importante para outras aplicações da AP (e.g., detecção de ervas daninhas, mapeamento e previsão de produção de safra, detecção de falhas). Um dos cenários de grande utilização da Agricultura de Precisão no Brasil está no cultivo de cana-de-açúcar (*Saccharum officinarum*), motivando pesquisadores e empresas a desenvolverem soluções na área. Apesar de na literatura já existirem muitos trabalhos para detectar linhas de plantio, a maioria deles é para outras culturas (e.g., milho e beterraba) ou focados em linhas retas. Além de muitos utilizarem imagens de baixíssima altitude ( $\approx 2$  m), tiradas a partir do solo (manualmente ou acopladas no maquinário agrícola) ou de alta altitude (imagens de satélites). Considerando o cenário de grande utilização dos VANTs para obtenção de imagens para a Agricultura de Precisão, a grande importância da detecção das linhas de plantio, este projeto tem como objetivo principal estudar e analisar os diferentes métodos de segmentação automática em imagens de VANTs tendo, inicialmente, foco em imagens de plantações de cana-de-açúcar e propor melhorias na detecção de linhas de plantio. Dentre as técnicas mais utilizadas para a detecção de linhas de plantio algumas foram escolhidas (e.g., segmentação, operações morfológicas, transformada de Hough e transformada de Radon) para serem avaliadas. A principal hipótese a ser avaliada é de que o uso do deep learning (redes neurais profundas para segmentação semântica) em conjunto com índices de vegetação podem gerar bons resultados na detecção de linhas de plantio. Ao fim do projeto, espera-se obter um método automático capaz de realizar de forma mais eficiente e com maior precisão o processo de segmentação de imagens de plantações de cana-de-açúcar.

**Trilha:** Trabalho de Mestrado

## Entendendo a classificação multiclasse de linfomas através de descritores morfológicos e não-morfológicos

Tiago P de Faria (Universidade Federal de Uberlândia)\*; Luiz G A Martins (UFU); Marcelo do Nascimento (ufu)

tiagofaria@ufu.br\*; lgamartins@ufu.br; marcelo.nascimento@ufu.br

**Resumo:** O câncer que acomete células do sistema linfático é chamado de linfoma. O correto diagnóstico do tipo de linfoma é crucial para o seu tratamento, mas esta tarefa não é trivial, e especialistas podem divergir na classificação de um mesmo caso. Sistemas de diagnóstico apoiado por computador (CAD - computer aided diagnosis) podem ser usados para diminuir esta variabilidade. Técnicas de visão computacional e aprendizado de máquina podem ser empregadas para classificar imagens histológicas de forma automática. Apesar de precisos na tarefa de classificação, muitos destes métodos são construídos como caixas pretas, em que o conhecimento utilizado para a decisão é inacessível para o usuário, ou é representado de forma complexa demais para que possa ser interpretado. Esta característica diminui a confiabilidade do sistema e dificulta sua depuração. Este trabalho estuda técnicas para classificação multiclasse de imagens e métodos de inteligência artificial explicável, assim como suas aplicações na tarefa de classificação de imagens histológicas de linfomas do tipo não Hodkin. Propomos uma metodologia que integra diferentes técnicas de explicação de modelos de aprendizado de máquina, e assim, consegue extrair e mostrar ao usuário informações interpretáveis sobre o funcionamento do sistema preditivo e de suas predições. Em nossa metodologia, segmentamos as imagens, separando os núcleos celulares. Então, extraímos deles, descritores morfológicos e não-morfológicos, que são baseados, respectivamente, em suas formas e cores. Estes dados são submetidos a uma etapa de engenharia de atributos, na qual os atributos de pouca variância ou de alta correlação entre si são removidos. Nesta etapa, também fazemos a explicação dos dados utilizados na criação do modelo, mostrando ao usuário correlações existentes entre os diferentes tipos de descritores e suas relações com as classes de linfoma. Também realizamos uma normalização dos vetores de atributos. Então, treinamos uma rede neural para realizar a tarefa de classificação multiclasse, que obteve acurácia de 0.947 em nossos testes. Utilizamos SHAP e Anchors para interpretar, tanto o funcionamento geral deste modelo, quanto suas predições individuais. Um gráfico também é utilizado para melhor representar a explicação gerada pelo Anchors, onde mostramos o posicionamento espacial da amostra relativa a seus vizinhos da base de treinamento. Em casos em que haja dúvida sobre a classificação de uma amostra, o sistema também disponibiliza explicações sobre os valores dos atributos mais relevantes da amostra e sua relação com outras amostras presentes na base de dados. Nossa metodologia oferece uma classificação com acurácia similar a trabalhos do estado da arte, porém utiliza de descritores mais simples, e utiliza de técnicas que possibilitam a interpretabilidade de todo o sistema, evidenciando relações entre atributos na base de dados, explicando o funcionamento do modelo treinado e de suas predições.

**Trilha:** Trabalho de Mestrado

## **Chapter III**

# **TRABALHOS DE DOUTORADO**

# Automatic classification of Acute Lymphoblastic Leukemia with Deep Residual Neural Network improved with Genetic Algorithm

Larissa F. Rodrigues Moreira (Universidade Federal de Uberlândia)\*; Bruno A. N. Travencolo (Universidade Federal de Uberlândia); André R Backes (Universidade Federal de Uberlândia)

lfrodriguesmoreira@gmail.com\*; travencolo@gmail.com; arbackes@yahoo.com.br

**Resumo:** Acute Lymphoblastic Leukemia (ALL) is the most common childhood cancer worldwide, and it is characterized by the production of immature malignant cells in the bone marrow. Microscopy image analysis is the most economical for the initial screening of patients with ALL, but this task is subjective and time-consuming. Computer vision techniques provide automated analysis that can help specialists diagnose this disease. In this study, we propose a hybrid model using a genetic algorithm (GA) and Convolutional Neural Network (CNN) named ResNet-50V2, specifically a Residual Neural Network, to predict ALL using microscopy images available in ALL-IDB dataset. It contains 260 images, each with a single, centered cell in evidence, categorized into two classes: healthy (130 images) and immature (130 images). Given the reduced number of images in the dataset, we used data augmentation based on horizontal and vertical flips and transfer learning to improve the training of the CNN. We evaluated hyperparameter optimization using GA since it is not easy to determine a suitable setup for hyperparameter values of deep learning models. Also, we compare the performance of GA hyperparameter optimization with Bayesian and Random Search methods. The experiments were validated using a five-fold cross-validation procedure over the training and test set. The dataset was divided into 80% for training (further split into 80% for training and 20% for internal validation during 5-fold cross-validation) and 20% for testing (external validation). Our approach excels those found in the state-of-the-art by offering a hybrid model that allows obtaining 98.46% accuracy and 98.55% precision. In this sense, this study improves the classification performance significantly. It demonstrates that Bayesian and Random Search approaches are limited to search space distribution. At the same time, GA can incorporate domain-specific knowledge in all optimization phases, essential for hyperparameter combination in deep learning. Moreover, as far as we know, our result is the best obtained for ALL classification in microscopy images using the dataset evaluated in this study. Our proposal approach provides new perspectives to identify leukemia based on computer vision strategies, which could be an alternative for applications in a real-world scenario contributing to future research to help healthcare workers identify leukemia and manage patients' conditions.

**Trilha:** Trabalho de Doutorado

## **Arcabouço computacional para o mapeamento das áreas cultivadas em solo brasileiro**

Leandro Silva (Universidade Federal de Uberlândia)\*; João F Mari (Universidade Federal de Viçosa); Mauricio Escarpinati (Facom); André R Backes (Universidade Federal de Uberlândia)

leandrofurtado.ufv@gmail.com\*;  
arbackes@yahoo.com.br

joaof.mari@ufv.br;

mauricio@ufu.br;

**Resumo:** A humanidade passou por um rápido crescimento populacional a partir do século XIX. Esse crescimento levou a previsões pessimistas sobre a capacidade do planeta de produzir alimentos suficientes para uma população crescente. No entanto, as previsões pessimistas não se concretizaram principalmente devido às revoluções da agricultura moderna. Esses avanços são atualmente representados pela Agricultura de Precisão (AP). Em geral, o objetivo da AP é definir um Sistema de Apoio à Decisão para toda a propriedade agrícola, a fim de otimizar o uso dos insumos, preservar os recursos e potencializar as safras futuras. Neste sentido, a AP tornou-se dependente da tecnologia de imagem, mapeamento e sensoriamento para fomentar a capacidade produtiva da plantação, além de identificar outras características agronômicas importantes (por exemplo, estresse de nitrogênio). No Brasil, o agronegócio consiste em um setor importante para economia, onde representa cerca de 22% do PIB. Outro ponto de destaque do setor na realidade nacional é que em meio a pandemia de SARS-CoV-2, declarada pela Organização das Nações Unidas (ONU) em 11 de março de 2020, o agronegócio foi o único a apresentar crescimento (cerca de 0,6%), segundo o Instituto Brasileiro de Geografia e Estatística (IBGE). Deste modo, esta pesquisa tem como objetivo construir um arcabouço computacional, o qual seja capaz de, a partir de imagens de satélite, encontrar campos específicos de plantação no território brasileiro, a fim de auxiliar os agricultores em suas diversas atividades. Essa ferramenta também precisará levar em consideração a diversidade de biomas existentes no Brasil (Cerrado, Mata Atlântica, Pantanal e Amazônia, por exemplo) e as características inerentes de cada um deles. Neste sentido, esta pesquisa irá se concentrar em 3 atividades gerais: (i) detecção de nuvens e sombras, visto que apenas imagens nítidas podem permitir estimar adequadamente as características do campo; (ii) detecção dos limites de campo, a fim de prover visualização para que os agricultores possam obter informações sobre os estados de seus campos; e (iii) detecção de tipos de safra, data de semeadura e estágios de crescimento das plantas, com objetivo de determinar os estágios de crescimento da planta, através de imagens multiespectrais. Assim, espera-se com este trabalho construir e disponibilizar uma ferramenta computacional, a qual seja aderente à realidade do agronegócio brasileiro que fomente a produção agrícola. Espera-se também, com essa ferramenta, auxiliar o produtor a obter informações específicas relativas à sua área plantada.

**Trilha:** Trabalho de Doutorado

## Integração de dados oriundos de registro de imagens para classificação de imagens por meio de redes neurais profundas

Cícero L Costa (UFU)\*; Celia AZ Barcelos (Federal University of Uberlândia); Bruno Travencolo (Universidade Federal de Uberlândia)

ciceroclc@hotmail.com\*; celiazb@ufu.br; travencolo@ufu.br

**Resumo:** No século XXI, é crescente o uso de vídeo para transmitir informações; vídeos podem ser gravados e transmitidos instantaneamente por um *smartphone*. O foco deste trabalho é a classificação de imagens sequenciais oriundas de vídeos, por meio da fusão de imagens com dados de registro não rígido. Este trabalho foi desenvolvido sobre a hipótese de que o registro não rígido aplicado a imagens sequenciais pode produzir características (*features*) relevante para as *Convolutional Neural Networks* (CNNs), melhorando o desempenho dessas redes na tarefa de classificação. Para atingir o objetivo proposto no trabalho foram realizadas as seguintes etapas: carregar uma sequência de imagens, registrar a sequência de imagens, adotar as estratégias de fusão inicial e tardia; e, por fim, apresentar os resultados do classificador. As etapas do trabalho foram aplicadas a uma base que contém 600 vídeos de partes de filmes da indústria cinematográfica. Os vídeos são divididos por categorias que representam emoções a partir das expressões faciais. As emoções as quais os vídeos estão associados são: raiva, nojo, medo, feliz, neutro, triste, surpreso. A partir de uma situação que há uma sequência de imagens para análise  $\{S_0, S_1, S_2, \dots, S_n\}$ , a sequência de imagens é registrada, encontra-se o campo de deformação  $u^k$  que registra  $S_{n-1}$  a imagem  $S_n$ , ou seja,  $S_{n-1}(x + u^k(x)) \cong S_n(x)$ . Após encontrar  $u^k$ , tal que  $u^k = (u_1, u_2)$ , haverá a fusão dos campos de deformação  $u^k$  com as imagens. A fusão pode ser inicial, antes das imagens passarem por um classificador. Neste caso, os campos de deformação  $u_1$  e  $u_2$  são fundidos a uma imagem em nível de cinza. Os valores de  $u_1$  e  $u_2$  são acrescentados à imagem em nível de cinza, a imagem passa a ter 3 camadas. Após o processo de fusão, as imagens são submetidas a um classificador. O *F1 Score* obtido pelo classificador, *DenseNet201*, para fusão inicial foi de 32,74, na literatura, para mesma base de dados, alguns autores conseguiram *F1 Score* em torno de 60. A fusão tardia é um outro método explorado. Neste caso, a imagem em nível de cinza e os campos de deformação  $u_1$  e  $u_2$  foram submetidos, sem passar por fusão (separadamente) à rede neural *DenseNet201*. As respostas geradas pela rede que passarão por fusão. Os resultados parciais obtidos foram: O *F1 Score* de 40,70 para a classificação das imagens em nível de cinza e o *F1 Score* de 32,74 para as imagens registradas. A fusão tardia apresenta boas chances de melhoria dos resultados, visto que no caso 1, na classificação de imagens em cinza, ocorreram 46 acertos e no caso 2, na classificação das imagens registradas, foram 37 acertos; juntando os dois casos, para um conjunto de teste de 113 itens, foram acertados 81 itens diferentes da amostra e 2 iguais.

**Trilha:** Trabalho de Doutorado