

REVISTA CIENTÍFICA CAMPUS PARTY

BRASIL
campus
party

realização **GOUVEA**
experience

NOV. 22



**REVISTA
CIENTÍFICA
CAMPUS PARTY**



Sergio Risola – Revista Campus Party e a carta de abertura

No fantástico SHOW dessa 14ª edição da Campus Party em São Paulo nesse novembro.22, é lançada essa Revista para registrar muito além das centenas de palestras nos palcos, muito além das tecnologias apresentadas nos stands e ainda das mentorias e das milhares de conversas nos corredores e salas preparadasSIM, além disso tudo chegam essas páginas físicas e digitais para carimbar / destacar projetos / tecnologias / teses universitárias ... ou a caminho das bancadas de P&D daquelas empresas q apostam em inovação, muito além dos seus laboratórios, ou das 4 paredes como dizemos, mas buscam no mercado a agregação de inovação em seus produtos e serviços.

E vale lembrar os inúmeros investidores, de todos os tamanhos, verdadeiros olheiros q buscam ou farejam sacadas inovadoras q pintam pelos metros quadrados desse rico e incrível ambiente de jovens, também em busca de oportunidades ou de firmarem parcerias q complementem suas iniciativas.

Essa 14ª edição da Campus decreta um novo rumo para Seus Campuseiros presenciais ou virtuais. Sim, uma nova era para os negócios apresentados e discutidos pelos empreendedores participantes e assim provocando os estudantes universitários a trazerem para o palco da Campus sua tese de graduação ou mesmo da pós a caminho, enfim, trazer para os corredores do Anhembi e para as salas de mentorias, super bem preparadas pelos profissionais, destaque aqui para o TIME do SEBRAE, e buscar luzes para as mais diferentes linhas estratégicas à sua tese acadêmica ou ao seu canvas ou até ao seu plano de negócios esboçado. Sempre reforçando q esses DOCS estejam bastante alinhados com as tecnologias da informação e aqui um horizonte sem fim com o digital, com a inteligência artificial, robótica, 3 ou 4 D ou realidade mista ou tb. com as ciências da vida ou as novas energias e as incontáveis outras linhas dos negócios inovadores.

Atenção para as tendências q vimos acompanhando em nosso País, e pelo mundo, e sabendo q elas serão impulsionadas aqui no ambiente da Campus com todo esse forte viés empreendedor. Ainda, o momento Brasil e mundo, quanto às tendências q todas as tecnologias, sejam de q áreas forem, devam trabalhar o viés E S G – Ações ambientais, sociais e de governança. Sim, o ESG com a forte inclinação socioambiental e alinhado com a economia circular e, muito importante, com a pegada da mitigação de carbono. Enfim, em todos os rumos e atalhos da inovação q VC está ou

estará, será primordial a democratização da ciência em todos os seus aspectos para a melhoria da qualidade da vida no planeta terra.

Por fim, vale enfatizar que ambientes como esse da Campus, trabalham cooperando na transformação para uma sociedade empreendedora e aprendedora, disseminando que empreender não é apenas correr riscos para ganhar dinheiro, mas sim uma atitude psicológica para iniciar, desenvolver e concretizar um negócio, mas alicerçado tb. num SONHO realizável, claro. E ainda: o empreendedor não nasce, se faz. Ele aprende a ser, com seus erros, com os erros dos outros, com seus acertos e com os acertos dos outros e daí a importância do aumento dos ambientes onde a inovação e o empreendedorismo acontecem entre pessoas e grupos com formações e negócios distintos. O q importa é ser extremamente apaixonado por solucionar problemas e melhorar as vidas das pessoas.

Nessa linha, vale o recado de Malcolm Gladwell no livro Outliers, quando afirma não existir fórmulas mágicas de autoajuda, e que o meio onde vamos crescer é que vai motivar e criar a nossa história. Arremato com uma das máximas que carrego: “Só QI não resolve a vida de ninguém”. O fator emocional conta muito mais do que somente a inteligência ou talento. E emocional é soft skills? Sim, pq são aquelas habilidades q permitem à pessoa gerenciar emoções, lembrando q hard skills são as capacitações que o profissional tem pelos diplomas e certificados de qualificação conseguidos.

Portanto, criar e turbinar espaços onde o fator EMOCIONAL também joga a favor dos negócios é tarefa EMOCIONANTE e extremamente gratificante para nós Amigos / Parceiros / frequentadores e torcedores fanáticos desse rico espaço da Campus Party e agora também dessa Revista que chega para colocar mais luzes e conhecimento nas teses acadêmicas, mas com a pegada do projeto rumo ao mercado que significa INOVAÇÃO gerando notas fiscais e BUSINESS.

swr.

07.11.22

Editorial - Revista Campus Party e a carta de abertura	
Sergio Risola	03
Avanços recentes em inteligências artificiais sintetizadoras de imagens:Relato de experiência educacional com o uso de dall-e e dream	
Filipe Albuquerque Ito Russo	
Nanahira de Rabelo e Sant'Anna	
Rodrigo Hideki Imai	06
SSDPH: A Hybrid Approach for Top-K Subgroup Discovery in High Dimensional Databases	
Giordano Araújo Régis Toscano	
Tarcísio Daniel Pontes Lucas	19
Internet das coisas na gestão educacional: Inovação da gestão acadêmica e desenvolvimento tecnológico na rede itego - Goiás - Brasil	
Douglas Araújo Falcão	33
Gestão financeira de empreendedores rurais do Município de Pires do Rio/GO	
Auto de Paula Rodrigues Neto	
Ezequiel de Paula Alves Rodrigues	43
Análise de dados da saúde com foco na síndrome gripal e síndrome respiratória aguda	
André Almeida Luz	
Ivo Sócrates Moraes de Oliveira	
Jeferson Moraes da Costa	
Ary Henrique Moraes de Oliveira	
Glenda Michele Botelho	
André Luis Korzenowski	49

Avanços recentes em inteligências artificiais sintetizadoras de imagens: relato de experiência educacional com o uso de dall-e e dream

Filipe Albuquerque Ito Russo (Universidade de São Paulo – USP; ORCID: <https://orcid.org/0000-0003-4583-5733>)

Nanahira de Rabelo e Sant’Anna (Universidade de Brasília – UnB; Ministério da Ciência, Tecnologia e Inovações – MCTI; ORCID: <https://orcid.org/0000-0002-4072-7250>)

Rodrigio Hideki Imai (Escola Técnica Integrada ao Médio Martin Luther King – ETIM MLK; ORCID: <https://orcid.org/0000-0003-0306-7166>)

Resumo

O presente trabalho analisa potencialidades em síntese criativa a partir do uso de inteligências artificiais (IAs) sintetizadoras de imagens, com base em experimentações das ferramentas Midjourney, Craiyon, Dream e DALL-E, realizadas durante debate interativo online com estudantes de Ensino Médio Técnico. Apresenta-se um relato da experiência educacional, a qual incluiu divulgação científica, letramento digital e computacional e educação midiática, a partir de discussões e reflexões sobre os conceitos de inteligência, IA e inteligência natural (IN). O artigo exhibe um resumo do evento, das impressões e produções do público participante, assim como propõe o conceito de síntese criativa como a produção, por sistemas inteligentes, de respostas adaptativas frente às adversidades postas. A educação sobre e com IA se mostra necessária e urgente, uma vez que as disrupções provocadas por seus usos já impactam todos os setores estratégicos das nações. Os resultados da experiência relatada indicam interesse da juventude em compreender e se apropriar das tecnologias de IA.

Palavras-chave: inteligência artificial; inteligência natural; síntese criativa; educação computacional; educação midiática.

Introdução

O presente artigo apresenta uma análise sobre o potencial de síntese criativa a partir do uso de inteligências artificiais (IAs) sintetizadoras de imagens. Sua investigação articula conceitos de IAs com uma experiência educacional online realizada com estudantes de Ensino Médio Técnico, em que foram apresentados estudos de caso sobre o uso das ferramentas Midjourney, Craiyon, DALL-E e Dream, e realizada experimentação ao vivo de DALL-E e Dream pelo público participante (FILIFE RUSSO, 2022).

A inteligência artificial (IA) enquanto área do conhecimento é uma ciência, ou melhor, uma amálgama de ciências que estuda e desenvolve sistemas inteligentes, a emergência de inteligência em sistemas artificiais e outros conceitos relativos a sistemas adaptativos que demonstram comportamentos ou desempenhos ditos inteligentes, ao compará-los com a inteligência humana.

Pei Wang (2019) recupera três definições que impactaram profundamente por décadas a trajetória de grandes pesquisas em IA, dadas por quatro pesquisadores que entre outros cunharam o termo “Inteligência Artificial” na reunião de Dartmouth e lideraram três centros de pesquisas na área. Segundo suas investigações, Newell e Simon, em 1976, definem “ação inteligente geral” como o escopo da inteligência como visto na ação humana: agir em qualquer situação real de acordo com os fins do sistema e adaptável às demandas do ambiente, sob limites de velocidade e complexidade. Já Minsky, em 1985, define inteligência como a habilidade de resolver problemas difíceis. Por outro lado, McCarthy, em 1988, afirma que a IA se preocupa com métodos para alcançar metas em situações onde a informação disponível possui uma certa complexidade, dessa forma os métodos a serem usados se relacionam com o problema em questão e são similares seja o resolvidor de problemas um ser humano, um extraterrestre ou um computador. Por fim, Wang (2019) propõe que inteligência diz respeito à adaptação mediante recursos e conhecimento insuficientes e enfatiza que qualquer definição operacional de IA possui potências e limites, entendê-los garante uma melhor compreensão dos caminhos possíveis a partir de cada conceito.

O progresso tecnológico em IA já tem impactado o mundo de forma dramática, influenciando processos e resultados em medicina, engenharia, direito, economia, educação, cultura e arte, sendo essas

três últimas as áreas de enfoque deste artigo. Em especial, iremos nos deter nos avanços recentes em IAs sintetizadoras de imagem.

A síntese de imagem melhorou significativamente nos últimos anos devido ao crescimento dos algoritmos de Aprendizado Profundo, mais especificamente, Redes Generativas Adversárias (GANs, na sigla em inglês). O framework das GANs combina representação de características de alto nível dos dados, provida pelas redes neurais profundas, e um procedimento de treinamento adversário para cobrir a lacuna entre distribuições complexas. Com o uso de GANs, qualquer distribuição desconhecida (por exemplo, imagens de pássaros) pode ser mapeada a partir de distribuições conhecidas (por exemplo, Gaussiana) sem depender de qualquer supervisão adicional além de uma quantidade limitada de amostras da distribuição alvo. Sem uma representação matemática explícita, novos dados indistinguíveis dos reais podem ser sintetizados (LOPES, 2019, p. 25, tradução nossa).

Das tantas IAs sintetizadoras de imagens a partir de textos, vale enfatizar Midjourney, Dream, Craiyon e DALL-E, pois são algumas das mais atuais, encontram-se disponíveis na internet e foram utilizadas em estudos de caso no evento escrutinado por este artigo.

Midjourney é um laboratório independente de pesquisa que explora novas formas de pensamento e busca expandir as habilidades imaginativas da humanidade. O laboratório dispõe de duas maneiras de se experimentar suas ferramentas, sen-

do a primeira pelo Midjourney Bot, robô virtual com o qual se pode gerar imagens a partir de textos em um canal ou chat do Discord. A segunda ferramenta é um aplicativo web onde se hospeda as galerias de imagens dos usuários (MIDJOURNEY, 2022).

Dream, por sua vez, é um aplicativo web e mobile que cria obras de arte usando o poder da IA; digita-se um dado de entrada, seleciona-se um estilo de arte e observa-se o WOMBO Dream transformar com inteligência artificial essa ideia numa obra de arte em segundos (WOMBO, 2022).

Craiyon, anteriormente DALL·E mini, é um modelo de IA capaz de desenhar imagens a partir de qualquer texto de entrada (CRAIYON, 2022). DALL-E é um sistema de IA desenvolvido pela OpenAI que consegue criar imagens originais e realistas a partir de uma breve descrição textual, podendo fazer edições realistas e sensíveis ao contexto, o que inclui retocar, remover e inserir regiões específicas de uma imagem a partir de uma descrição em linguagem natural, suas capacidades também lhe permitem tomar uma imagem base e produzir novas e criativas variações baseadas na original (OPENAI, 2022).

A síntese de imagem a partir de texto é uma tarefa de gerar imagens a partir de linguagem natural. A linguagem natural provê uma forma fácil, para usuários, gerarem ou manipularem imagens a gosto. É uma restrição geral que pode guiar diferentes aspectos de uma imagem. Com descrições textuais, pode-se guiar o pro-

cesso de síntese a fim de se obter objetos específicos com atributos específicos tais como cores, texturas, posição relativa, entre outros (LOPES, 2019, p. 25, tradução nossa).

Na sequência desta Introdução, são apresentadas as justificativas e objetivos do presente trabalho. O relato sobre a experiência educacional, a qual se pautou na experimentação das ferramentas Dream e DALL-E pelo público participante, está contemplado nas seções referentes à metodologia e aos resultados alcançados.

Justificativa

A inteligência artificial (IA) pode ser considerada uma das tecnologias mais disruptivas da contemporaneidade. Soluções de IA envolvem agrupamentos de tecnologias distintas, as quais conseguem simular capacidades humanas relativas à cognição. Recentes estudos e discussões sobre as aplicações de IA indicam tendências no sentido de uma maior capilaridade e penetração dessa tecnologia nos mais diversos setores estratégicos, assim como em atividades operacionais e em momentos lúdicos do dia a dia.

A novidade é que esses avanços têm implicações profundas para a economia, para a elevação da produtividade, para o emprego e o desenvolvimento dos países. Mais ainda, a IA tem potencial para comandar os processos de inovação e remodelar toda a Pesquisa e Desenvolvi-

mento (P&D) nas empresas. Ou seja, a IA que se desenvolve hoje, distinta do passado, exhibe características semelhantes às que marcaram a computação digital, a eletricidade e a máquina a vapor, que re- viraram o modo de se produzir, consumir, comercializar e viver (AGRAWAL, GANS e GOLDFARB, 2018 apud ARBIX, 2021, p. 263).

Discussões técnicas e jurídicas acerca das potencialidades e transversalidades no uso e aplicações de IA suscitaram a instituição de políticas como a Estratégia Brasileira de Inteligência Artificial (EBIA) em 2021, cujo objetivo é “potencializar o desenvolvimento e a utilização da tecnologia com vistas a promover o avanço científico e solucionar problemas concretos do País”. A elaboração da EBIA baseou-se em um diagnóstico que considerou políticas, estratégias ou planos similares adotados em outros países. Os principais eixos identificados nessas iniciativas incluem “pesquisa e desenvolvimento” e “competências profissionais e o futuro do trabalho”, que remetem à importância dos componentes educacional e profissionalizante. Um dos objetivos estratégicos da EBIA consiste em “capacitar e formar profissionais para o ecossistema da IA”, e dois de seus eixos são denominados “educação” e “força de trabalho e capacitação” (BRASIL, 2021, p. 5-15).¹

¹ A EBIA foi elaborada em seguimento à priorização conferida ao tema da IA no âmbito da Estratégia Brasileira para a Transformação Digital (E-Digital), de 2018. O documento contemplou “educação e capacitação profissional” como um de seus eixos, com o objetivo de “formar a sociedade para o mundo digital, com novos conhecimentos e tecnologias avançadas, e prepará-la para o trabalho do futuro” (BRASIL, 2018, p. 58).

Segundo a EBIA, o desafio que se coloca no campo da educação é “preparar as gerações atuais e futuras para conviver com as mudanças e os impactos da IA, muitos dos quais ainda não são plenamente compreendidos” (BRASIL, 2021, p. 29). Ainda nesse eixo, o documento faz referência ao Consenso de Beijing sobre Inteligência Artificial e Educação, cujas recomendações incluem atenção à natureza multidisciplinar da IA e seus impactos, adoção de abordagens governamentais completas, intersetoriais e multissetoriais ao planejamento e governança da IA na educação, e definição de prioridades estratégicas com base nos desafios locais (UNESCO, 2019).

Já no eixo de força de trabalho e capacitação, menciona-se a tendência de criação de novos empregos e deslocamento ou transformação de outros. O crescente desenvolvimento de novas tecnologias visa solucionar contradições sociais, ensejando novos panoramas e, conseqüentemente, “a criação de novas profissões associadas a supervisionar, manter e incrementar as tecnologias recém-introduzidas” (BRASIL, 2021, p. 33). Contudo, informa-se que o número de profissionais da área de IA no Brasil é de apenas 4.429 especialistas, comparando-se aos 415.166 empregos em TI no setor empresarial do país.

A realização de um debate interativo sobre IA, com foco em tecnologias sintetizadoras de imagens, as quais se encontram em rápida difusão global nas últimas

semanas, oportuniza o compartilhamento de conhecimentos e estimula reflexões sobre impactos presentes e potenciais da IA. Trata-se de uma iniciativa de divulgação científica, letramento digital e computacional e educação midiática, voltada à inclusão da sociedade civil, em especial as juventudes, no debate sobre um novo setor estratégico com impacto na soberania nacional, na economia e na sociedade brasileira, com particular ênfase na economia criativa do país.

Objetivos

O principal objetivo do presente trabalho é apresentar possibilidades em síntese criativa com o uso de inteligências artificiais (IAs) sintetizadoras de imagens, identificadas a partir de articulação entre conceitos de IA e uma experiência educacional online envolvendo o uso de ferramentas de IA. Definimos síntese criativa como a produção, por sistemas inteligentes, de respostas adaptativas frente às adversidades postas.

Outro objetivo consiste em apresentar recentes avanços em termos de IAs sintetizadoras de imagens, especialmente no que se refere às ferramentas Midjourney, Dream, Craiyon e DALL-E, as quais já se encontram disponíveis na internet para o uso público e, assim, possibilitam novas experiências pedagógicas, sociais, culturais e artísticas.

Ademais, objetiva-se enfatizar as demandas sociais por uma educação midiática atual e amplamente acessível à população brasileira, uma vez que, com o ad-

vento da internet, às mídias físicas foram acrescentadas as mídias digitais, depois as mídias sociais digitais e atualmente as mídias digitais sintetizadas por IAs. O desconhecimento das características peculiares desses avanços tecnológicos, comunicacionais e sociais pode agravar o descompasso e a desarmonia intergeracional, além de submeter a população aos riscos imbricados nas novas mídias das *fake news* (FN) e das *deep fakes*.

As FN atingem, diariamente, milhões de pessoas, tumultuando a cultura democrática e desacreditando o jornalismo. [...] Antes de tudo, a expressão *fake news* não poder ser traduzida ao pé da letra, porque se são fakes não são news, pois, no Jornalismo, em princípio, não existe notícia falsa, tanto que uma das normas é a checagem dos fatos antes da publicação, conforme evidenciado em estudos anteriores (Prado, 2021b, p. 48). Mas é assim que são conhecidas as informações, ou melhor dizendo, mensagens fraudulentas proliferadas na atual era da pós-verdade pela qual o mundo vem passando de forma descontrolada. [...] A expressão “*fake news*” (FN) abrange diversas categorias. Todas ameaçam a qualidade do jornalismo, seus meios de subsistência e, por conseguinte, a formação da opinião coletiva. (PRADO, 2022, p. 25-26)

Ainda, este trabalho tem por objetivo contribuir para maior divulgação e popularização da ciência mediante exposição de relato de experiência educacional envolvendo debate e experimentação de ferramentas de IA por parte de estudantes de Ensino Médio Técnico.

Metodologia

No contexto das celebrações do Dia Internacional da Juventude, data comemorada anualmente em 12 de agosto, conforme estabelecido pela Organização das Nações Unidas (ONU) em 1999 (ONU, 2000)², as pessoas autoras do presente artigo organizaram o evento online denominado “Inteligências artificial e natural: debate sobre síntese criativa usando DALL-E e Dream”.

O evento, realizado em 20 de agosto de 2022, foi apoiado pelo IYD (*International Youth Day*) Brasil, movimento brasileiro organizado por jovens que, desde 2019, apoia a realização de atividades de impacto em todo o país durante o mês de agosto, voltadas ao protagonismo das juventudes e alinhadas com os 17 Objetivos de Desenvolvimento Sustentável (ODS) da ONU (IYD BRASIL, 2022).

O debate interativo alinou-se com o ODS 4, denominado como Educação de Qualidade, cujo objetivo é “assegurar a educação inclusiva e equitativa e de qualidade, e promover oportunidades de aprendizagem ao longo da vida para todos” (ONU, 2022). Na plataforma do IYD Brasil, o evento foi inscrito entre os realizados com o propósito de promover o “Eixo 2 - Educação”, relativo ao direito dos jovens à educação assegurado pelo Estatuto da Juventude, conforme Seção II do Capítulo II da Lei nº 12.852, de 5 de

agosto de 2013 (BRASIL, 2013)³. A realização do debate foi divulgada pelo IYD Brasil e nas redes sociais e profissionais das pessoas organizadoras, que autoram este trabalho. Nas peças de divulgação, foi ressaltado o foco no público jovem.

O evento foi realizado na plataforma StreamYard e transmitido ao vivo pelo canal Filipe Russo da plataforma YouTube. Contou com a participação de 11 pessoas, em sua maioria estudantes do curso de Mecatrônica da Escola Técnica Integrada ao Médio Martin Luther King (ETIM MLK). Somadas às pessoas organizadoras que apresentaram o evento, contabilizou-se um total de 14 pessoas.

Imagem 1 - Abertura do debate interativo



Fonte: elaboração das pessoas autoras (2022)

Após a abertura e de modo introdutório, foi apresentado um conceito geral de IA, definida, a partir das abordagens de Newell e Simon (1976), Minsky (1985) e McCarthy (1988), tal como apresentadas por Wang (2019), como “um sistema de computador projetado para interagir com

2 A data foi criada por meio da Resolução 54/120 da Assembleia Geral da ONU, que declarou o 12 de agosto como o Dia Internacional da Juventude. No Brasil, a data foi instituída pela Lei nº 10.515, de 11 de julho de 2002 (BRASIL, 2002)

3 Para os efeitos da Lei nº 12.852/2013, conforme art. 1º, § 1º, são consideradas jovens as pessoas com idade entre 15 (quinze) e 29 (vinte e nove) anos de idade. Para os fins deste trabalho, entretanto, a juventude não é limitada pelo critério etário, mas considerada em sua diversidade, tendo em vista a heterogeneidade dos contextos e experiências dos sujeitos juvenis.

o mundo através de capacidades e comportamentos considerados inteligentes e, portanto, comumente humanos” (FILIPE RUSSO, 2022). Informou-se que a IA, o Aprendizado de Máquina e a Visualização de Dados pertencem à área do conhecimento chamada Ciência de Dados, conforme proposto por Henry E. Brady (2019), uma vez que esta é uma das categorias mais abrangentes.

Métodos da Ciência de Dados têm origem na ciência da computação, estatística, ciência da informação e biblioteconomia, com algumas raízes nos esforços de biólogos em modelar as conexões entre neurônios no cérebro humano e no trabalho de cientistas da cognição (tais como o político polímata Herbert Simon) no desenvolvimento de inteligência artificial (BRADY, 2019, p. 313, tradução nossa).

Por sua vez, a síntese criativa, considerando o foco do evento nas IAs sintetizadoras de imagens, foi definida como um sistema em que a ferramenta de IA recebe uma entrada de dados, chamada input, e sintetiza uma imagem com base no seu treinamento e nos dados inseridos, a qual representa o output.

Em seguida, passou-se a estudos de caso sobre o uso das ferramentas Midjourney, Craiyon, Dream e DALL-E. Após breves explicações sobre cada ferramenta, foram apresentadas imagens produzidas com o uso delas. As pessoas participantes foram convidadas a responder e comentar sobre quais possíveis textos de entrada (input) teriam gerado os resulta-

dos imagéticos (output). Posteriormente, na sessão chamada “Mão na Máquina”, experimentaram o uso das ferramentas Craiyon, Dream e DALL-E, com diversas sugestões para textos de entrada. As imagens produzidas foram verificadas, comentadas e analisadas pelas pessoas apresentadoras e participantes.

Após o experimento das mencionadas ferramentas de IA, apresentou-se o seguinte conceito geral de inteligência natural (IN): “capacidade dos sistemas biológicos de interagir com o mundo de modo a suprir suas demandas metabólicas, psicológicas e/ou culturais” (FILIPE RUSSO, 2022). O conceito foi elaborado com inspiração na abordagem de GARDNER (2000), para quem a inteligência é “um potencial biopsicológico para processar informações que pode ser ativado num cenário cultural para solucionar problemas ou criar produtos que sejam valorizados numa cultura”. No tocante à síntese criativa, destacou-se que as INs realizam-na tendo as adversidades como elemento de entrada e o processo evolutivo como elemento de saída.

Na sequência, o público foi convidado a apresentar suas concepções sobre o termo “inteligência”, as quais foram incorporadas, ao vivo, aos slides da apresentação em realização. Houve mais uma sessão de experimentação de DALL-E antes da conclusão do debate interativo. Hiperlinks de interesse sobre as ferramentas de IAs e suas documentações

foram listados na apresentação realizada durante o evento, e todo o material apresentado foi disponibilizado às pessoas participantes mediante interesse expresso em formulário de participação.

Resultados alcançados

O evento “Inteligências artificial e natural: debate sobre síntese criativa usando DALL-E e Dream” foi marcado por significativa interatividade. Todas as 11 pessoas participantes apresentaram contribuições na forma de perguntas, comentários e/ou sugestões de textos de input para experimentação nas ferramentas Dream e DALL-E.

Nos comentários, o público considerou intrigantes as formas e cores das imagens geradas. As pessoas participantes apresentaram referências a seres, objetos e lugares reais e imaginários, bem como propuseram textos com carga ora filosófica, ora poética.

Como exemplo dessas interações, em resposta ao que poderia ter sido o texto de entrada (input) que gerou a imagem de saída (output) a seguir na ferramenta Midjourney, foram sugeridas as entradas textuais “as borboletas iluminam a cidade noturna”, “o povo da cidade deixa suas ideias brilharem” e “ideias brilhantes iluminam um mundo de trevas”:

Imagem 2 – Output gerado na IA Midjourney



Fonte: elaborado por Filipe Russo (2022)

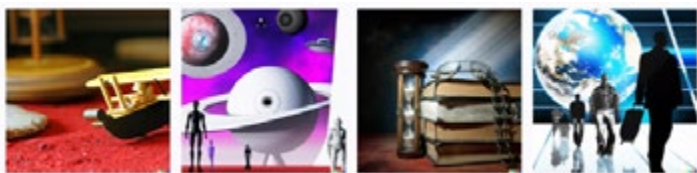
Revelou-se que o texto de input que havia gerado a imagem de output havia sido: “Mil fadas chegando em São Paulo”.

Um participante indagou se as imagens seriam geradas a partir da concepção humana ou da interpretação da IA a partir do comando. Em resposta, explicou-se que os textos propostos são transformados em espaços vetoriais multidimensionais pela IA, que busca referências próximas ao texto proposto dentro do que já está disponível na Internet e/ou no banco de dados da própria IA, e recombina as formas de simbolização.

Na experimentação com DALL-E, durante a sessão “Mão na Massa”, as quatro imagens a seguir foram produzidas, dispostas lado a lado, a partir do texto de

input “viagem no tempo”, o qual foi proposto por um dos participantes. O texto foi inserido no idioma inglês – “time travel” – em razão de as ferramentas serem estrangeiras e otimizadas para esse idioma.

Imagem 3 – Output gerado na IA DALL-E



Fonte: elaboração das pessoas autoras, com contribuições do público participante (2022)

Comentou-se que a primeira imagem do canto esquerdo remete a um retorno ao tempo de infância, resultado artístico não imaginado pelas pessoas apresentadoras e participantes. Notou-se que as três outras imagens têm em comum planos de fundo com referências espaciais. Na segunda, da esquerda para a direita, aparecem bonecos humanoides e planetas com aparência de olhos gigantes. Já na terceira, chamou a atenção uma ampulheta com três segmentos e uma corrente ao redor de livros que parecem feitos de pedra ou areia. Por sua vez, a quarta e última imagem apresenta uma fila de seres humanoides que parecem embarcar em uma viagem.

Na sessão de debate sobre os possíveis sentidos para o termo “inteligência”, as pessoas participantes apresentaram suas concepções, que foram inseridas na apresentação realizada ao vivo, conforme imagem a seguir:

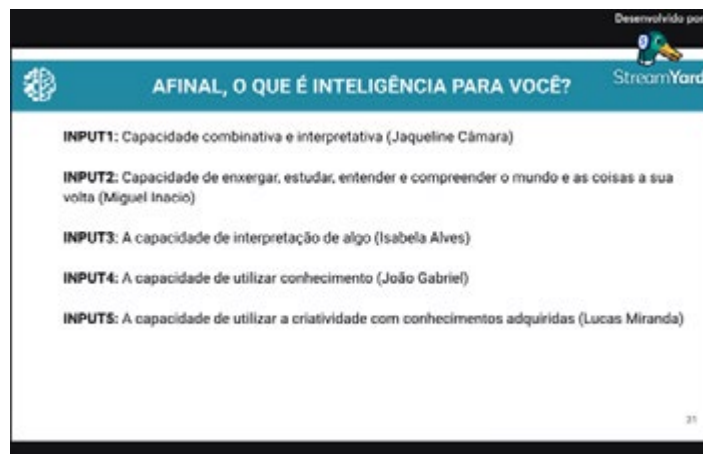


Imagem 4 – Conceitos de inteligência

Fonte: elaboração das pessoas autoras, com contribuições do público participante (2022)

Notou-se o uso comum da palavra “capacidade” no início das cinco definições apresentadas e foi destacada a relevância de um ponto de partida, até mesmo arbitrário, para processos numéricos, estatísticos e computacionais.

Com base nas definições dos inputs 1, 2 e 3, propôs-se um entendimento dos termos “enxergar”, “estudar”, “entender” e “compreender” como subprocessos ou inputs do processo de “interpretar”. O entender e o compreender podem equivaler a inserir o que já se conhecia em um modelo interpretativo do mundo, como fazem as IAs a partir de modelos previamente construídos. Aplicando-se ao debate criativo, entender e compreender representariam a leitura dos modelos interpretativos das pessoas apresentadoras e participantes, ao processarem algo que foi inserido e reorganizado em seus sistemas e mapas internos, de modo a gerar sentido.

A partir do sentido utilitarista presente

nas definições dos inputs 4 e 5, remeteu-se à questão de demandas a serem supridas com o uso da inteligência. Frisou-se que a existência de IAs sintetizadoras de imagem se deve a demandas sociais por novos processos criativos, por inteligências artificiais que possam modelar a imaginação humana, acessar e ativar o universo simbólico da comunicação social, demandas por instrumentos computacionais com potências de realizar a síntese criativa apresentada no evento, promovida e consumida por pessoas que desejam a construção conjunta de conhecimentos sobre inteligência, inteligência artificial (IA) e inteligência natural (IN).

Em seguida, foi realizada uma breve discussão sobre as relações entre IA e conhecimento. O conhecimento de uma IA corresponderia ao banco de dados e à quantidade de informações que a ela foram atribuídas. Aspectos a serem pensados sobre inteligência em sistemas computacionais podem dizer respeito a como usar inteligência para representar e expressar conhecimento, como formar bases de conhecimento para compreender o domínio de questões diversas, como agregar conhecimento com o tempo e como realizar inferências a partir de determinado conjunto de conhecimentos. A questão da utilidade foi retomada no que se refere à resolução de vários problemas no mundo, os quais demandam novas, criativas e críticas articulações das inteligências e das bases de conhecimento, sejam essas máquinas ou humanas.

Antes de o evento ser concluído, o público participou de mais uma sessão de experimentação de DALL-E, propondo diversos inputs e comentando os outputs, sendo comparadas expectativas e imagens resultantes. Os temas dos textos de entrada variaram entre religiosidades, ditados populares, alimentos, heróis de quadrinhos, personagens da cultura pop japonesa, esportes, lugares concretos e abstratos, realidade virtual e a própria IA. Notou-se significativo engajamento, com participações em tom animado e divertido. A gravação do evento, em vídeo com duração de duas horas, cinquenta e três minutos e quarenta segundos, foi disponibilizada no canal Filipe Russo da plataforma Youtube (FILIPE RUSSO, 2022).

O material apresentado durante o evento foi disponibilizado às pessoas participantes que manifestaram interesse, mediante preenchimento de formulário de participação. O formulário também continha um campo para que pudessem compartilhar seus comentários e sugestões. Nas palavras do público respondente – sete (7) do total de 11 pessoas – a apresentação foi considerada informativa, interativa e convidativa; quatro (4) pessoas manifestaram interesse em aprofundar conhecimentos sobre IAs.

Considerações finais

Com este relato de experiência, apresentamos possíveis desenvolvimentos em síntese criativa com o uso de inteli-

gências artificiais (IAs) sintetizadoras de imagens. Nesse contexto, observou-se a realização de um diálogo criativo entre as “bolhas imaginativas” dos participantes e o “modelo de imaginação” das IAs sintetizadoras de imagens, uma vez que cada output impactou na seleção do próximo input.

Pessoas possuem redes intersubjetivas que acessam para produzir conteúdo – trata-se da imaginação, que só existe no atual grau de sofisticação porque existe a imaginação de outras pessoas, sendo que a imaginação de uma pessoa afeta e impacta a imaginação de outras a partir de produções e ações sociais. A bolha imaginativa de uma IA é uma rede gigantesca e prontamente acessível, estimulável e produtiva, diferentemente da bolha imaginativa de cada pessoa que permanece individual e inexoravelmente privada, íntima e inefável. Em outras palavras, as IAs sintetizadoras de imagens modelam a imaginação humana em seus aspectos de simbolismo visual a partir de produções sociais neste domínio, publicadas no ciberespaço, possibilitando o acesso e o uso da nossa intersubjetividade ou de um modelo operacional desta. A imaginação humana encontra na imaginação máquina um amplificador e um qualificador de seus sinais comunicacionais.

Tendo em vista que os objetivos do evento foram alcançados mediante significativa interatividade entre as pessoas apresentadoras e o público participante, sobretudo

do nas sessões de experimentação das IAs Dream e DALL-E, considera-se que o debate contribuiu tanto para a divulgação científica, o letramento digital e computacional e a educação midiática, quanto para favorecer o protagonismo das pessoas jovens que realizaram e participaram do evento.

Espera-se que o evento inspire iniciativas semelhantes que possam contribuir para a popularização da ciência, o debate tecnológico e o pensamento crítico sobre mídias digitais sintetizadas por inteligências artificiais.

Referências

ARBIX, Glauco. “Algoritmos não são inteligentes nem têm ética, nós temos”: a transparência no centro da construção de uma IA ética. In: COZMAN, Fabio G.; PLONSKI, Guilherme Ary; NERI, Hugo. *Inteligência artificial: avanços e tendências*. São Paulo: Instituto de Estudos Avançados, 2021, p. 262-286.

BRADY, Henry E. The Challenge of Big Data and Data Science. *Annual Review of Political Science*, n. 22, p. 297-323, 2019. Disponível em:

<<https://doi.org/10.1146/annurev-polisci-090216-023229>>. Acesso em 10 set. 2022.

BRASIL. Ministério da Ciência, Tecnologia e Inovações. *Estratégia Brasileira de Inteligência Artificial – EBIA – 2021*. Brasília: MCTI, 2021. Disponível em <https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivos/inteligenciaartificial/ebia-diagramacao_4-979_2021.pdf>. Acesso em 11 set. 2022.

_____. Ministério da Ciência, Tecnologia, Inovações e Comunicações. *Estratégia Brasileira para a Transformação Digital*. 2018. Brasília, MCTIC, 2018. Disponível em: <<https://www.gov.br/mcti/pt-br/centrais-de-conteudo/comunicados-mcti/estrategia-digital-brasileira/estrategiadigital.pdf>>. Acesso em 11 set. 2022.

- _____. Presidência da República. Casa Civil. Subchefia para Assuntos Jurídicos. Lei nº 10.515, de 11 de julho de 2002. Institui o 12 de agosto como Dia Nacional da Juventude. Disponível em <http://www.planalto.gov.br/ccivil_03/leis/2002/l10515.htm>. Acesso em 10 set. 2022.
- _____. Presidência da República. Casa Civil. Subchefia para Assuntos Jurídicos. Lei nº 12.852, de 5 de agosto de 2013. Institui o Estatuto da Juventude e dispõe sobre os direitos dos jovens, os princípios e diretrizes das políticas públicas de juventude e o Sistema Nacional de Juventude – SINAJUVE. Disponível em <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2013/lei/l12852.htm>. Acesso em 11 set. 2022.
- CRAIYON. Frequently asked questions. Disponível em: <<https://www.craiyn.com/#faq>>. Acesso em 11 set. 2022.
- FILIPPE RUSSO. Inteligências artificial e natural: debate interativo sobre síntese criativa usando DALL-E e Dream. YouTube, 20 de agosto de 2022. Disponível em <https://www.youtube.com/watch?v=e0Mej4QXOyE&list=PLY4NWojCWTCNYtImp8ssaEBpdbo2kKs_c&index=2>. Acesso em 10 set. 2022.
- GARDNER, H. Inteligência – um conceito reformulado. Rio de Janeiro: Objetiva, 2000.
- IYD BRASIL. Quem somos. Disponível em <<https://www.diadajuventude.com.br/quem-somos>>. Acesso em 9 set. 2022.
- LOPES, M. A. Fully-disentangled text-to-image synthesis. Dissertação (Mestrado em Ciência da Computação) – Escola Politécnica, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2019.
- MCCARTHY, J. Mathematical logic in artificial intelligence. *Dædalus*, n. 117 (1), p. 297-311, 1988.
- MIDJOURNEY. Midjourney documentation. Disponível em <<https://midjourney.gitbook.io/docs/>>. Acesso em 11 set. 2022.
- MINSKY, M. The society of mind. New York: Simon and Schuster, 1985.
- NEWELL, A.; SIMON, H. A. Computer science as empirical inquiry: symbols and search. *Communications of the ACM*, n. 19 (3), p. 113-126, 1976.
- ONU – ORGANIZAÇÃO DAS NAÇÕES UNIDAS. Nações Unidas Brasil. Objetivo de desenvolvimento sustentável 4: educação de qualidade. Disponível em: <<https://brasil.un.org/pt-br/sdgs/4>>. Acesso em 10 set. 2022.
- _____. Resolução adotada pela Assembleia Geral [sobre o relatório do 3º Comitê (A/54/595)] 54/120. Políticas e programas envolvendo os jovens. 20 de janeiro de 2000. Disponível em <<https://daccess-ods.un.org/access.nsf/Get?OpenAgent&DS=A/RES/54/120&Lang=E>>. Acesso em 10 set. 2022.
- OPENAI. About. Disponível em: <<https://labs.openai.com/about>>. Acesso em 11 set. 2022.
- PRADO, M. Fake news e inteligência artificial: o poder dos algoritmos na guerra da desinformação. São Paulo: Edições 70, 2022.
- UNESCO – Organização das Nações Unidas para a Educação, a Ciência e a Cultura. Consenso de Beijing sobre a inteligência artificial e a educação. Documento final da Conferência Internacional sobre Inteligência Artificial e Educação “Planejando a educação na era da IA: liderar o avanço”, 16-18 de maio de 2019, Beijing, República Popular da China. Beijing: UNESCO, 2019. Disponível em: <<https://unesdoc.unesco.org/ark:/48223/pf0000372249>>. Acesso em 11 set. 2022.
- WANG, Pei. On Defining Artificial Intelligence. *Journal of Artificial General Intelligence* 10 (2), p. 1-37, 2019. Disponível em <<https://doi.org/10.2478/jagi-2019-0002>>. Acesso em 10 set. 2022.
- WOMBO. Dream. Disponível em <<https://www.wombo.art/>>. Acesso em 11 set. 2022.

SSDPH: A Hybrid Approach for Top-K Subgroup Discovery in High Dimensional Databases

Giordano Araújo Régis Toscano, Tarcísio Daniel Pontes Lucas UNICAP ICAM TECH – Universidade Católica de Pernambuco (Unicap) - 50.050-900 – Recife – PE – Brasil / giordano.2018108372@unicap.br, tarcisio.lucas@unicap.br

Resumo

Esse trabalho apresenta um novo algoritmo para a tarefa de mineração de subgrupos com foco em bases de dados de cidades inteligentes e de alta dimensionalidade e com grande número de exemplos: O SSDPH. Trata-se de uma abordagem híbrida, top-k e mono-objetivo, proveniente da integração do SSDP+, seu antecessor, com a busca exaustiva. Além disso, ele contém um módulo para a tarefa de pré-processamento dos dados. Um de nossos objetivos é avaliar o impacto dessa hibridização com relação a certos critérios como qualidade, diversidade e tempo de processamento. Os experimentos foram feitos com 10 bases de dados (incluindo bases de bioinformática e bases que contemplam problemas relevantes no Brasil) e os resultados indicam que a solução proposta é promissora em problemas envolvendo bases de dados de alta dimensionalidade e viável em bases com muitos exemplos.

Palavras-chave: mineração de subgrupos; bases de dados de alta dimensionalidade, cidades inteligentes.

1. Introdução

1.1 Justificativa

A mineração de subgrupos é uma área de Inteligência Artificial que utiliza algoritmos e dados para encontrar os conjuntos de características que concentram a presença de um grupo alvo (ex.: mulheres vítimas de violência, acidentes de trânsito, alunos que evadem da escola) (HERRERA, 2011). A mineração de subgrupos tem sido aplicada na descoberta de conhecimento em diversas áreas, como bioinformática (LUCAS, 2019), medicina (CARMONA, 2013), marketing (JESUS, 2007) e acidentes de carros (KAVSEK, 2002).

No entanto, a mineração de subgrupos é uma tarefa desafiadora sob diferentes olhares. Bases de dados de alta dimensionalidade (número de atributos na ordem de 10^2 ou 10^3), por exemplo, representam um espaço de busca desafiador para os algoritmos de mineração de subgrupos. Já as bases de dados com grande número de registros/exemplos, torna a tarefa de mineração de subgrupos desafiadora em relação ao custo de processamento e memória. Um tipo de base de dados que pode concentrar todos os desafios apresentados são as advindas de aplicações de cidades inteligentes.

Dessa forma, essa pesquisa visa testar e adaptar algoritmos de mineração de subgrupos para lidar com parte dos desafios das bases de dados de cidades inteligentes. A pesquisa terá como foco propor um algoritmo de mineração de subgrupos eficiente para bases de dados diversificadas, com grande ou pequeno número de instâncias, bem como alta ou baixa dimensionalidade.

1.2 Objetivos

1.2.1 Objetivo Geral

Propor algoritmo de mineração de subgrupos para o contexto de cidades inteligentes com foco em bases de dados de alta dimensionalidade e grande número de instâncias.

1.2.2 Objetivos específicos

Estudar alternativas promissoras de algoritmos para aplicações de mineração de subgrupos em bases e dados de Cidades Inteligentes.

Propor algoritmo eficiente para mineração de subgrupos no contexto de bases de dados de cidades inteligentes.

Este artigo está organizado da seguinte forma. Seção 2 define a tarefa da mineração de subgrupos e algumas métricas de avaliação. A Seção 3 fornece uma visão geral acerca dos trabalhos relevantes na

área. Na Seção 4 o SSDPH é descrito em detalhes. As Seções 5 e 6 mostram, respectivamente, os experimentos e os resultados. Por fim, a Seção 7 apresenta a conclusão

2. Mineração de Subgrupos

Mineração de Subgrupos (HERRERA, 2011) é uma técnica da mineração de dados focada em descobrir relações/padrões interessantes entre diferentes objetos de uma base de dados com relação a uma propriedade de interesse do usuário, a variável alvo. Os padrões extraídos são normalmente representados em forma de regras e são chamados de subgrupos.

Subgrupos podem ser classificados de acordo com sua relevância, que é definida pelas métricas de avaliação. Uma das métricas mais usadas é a acurácia relativa ao peso (WR_{Acc}), dada pela Equação 1, onde TP e FP são os verdadeiros e falsos positivos do subgrupo s , $|D|$ é o número de exemplos da base de dados e $|D^+|$ o número de exemplos que é alvo da investigação (Flach, 2003).

$$WR_{Acc}(s) = \frac{TP + FP}{|D|} \left(\frac{TP}{TP + FP} - \frac{|D^+|}{|D|} \right)$$

Equação 1

Outra métrica de avaliação conhecida é o Qg , dada pela Equação II, onde g é um parâmetro de generalização (HERRERA, 2011). Existem também algumas métricas globais, cuja finalidade é avaliar um conjunto de subgrupos. Uma delas é o suporte positivo global SUPP+, que mede a porcentagem de exemplos positivos $D+$ cobertos pelo conjunto de top-k subgrupos (Lucas, 2017).

$$Qg(s) = \frac{TP}{FP + g}$$

Equação II

3. Trabalhos Relacionados

Dependendo do número de atributos, testar todas as possibilidades de conjunções de itens (busca exaustiva) pode se tornar computacionalmente custoso ou inviável. Por isso, várias abordagens heurísticas foram desenvolvidas para a tarefa de mineração de subgrupos. Dentre os algoritmos heurísticos, os que são baseados em beam search e computação evolucionária tem se destacado.

Os algoritmos baseados em beam search são inicializados a partir de uma lista de subgrupos de tamanho w (que corresponde ao valor do parâmetro `beamWidth`). Novos subgrupos são gerados a partir dos w subgrupos da iteração ante-

rior. Portanto, abordagens baseadas em beam search restringem o uso de memória explorando apenas parte do espaço de busca. Um dos primeiros e mais proeminentes algoritmos para mineração de subgrupos baseado em beam search é o SD (Gamberger, 2002).

Já o algoritmo FSSD (Belfodil, 2019) usa a abordagem de cobertura sequencial e um esquema guloso para resolver o problema de mineração de conjuntos de padrões. Seu objetivo é fornecer um pequeno conjunto de padrões que seja diversificado e bem discriminante para a classe alvo. No entanto, os testes foram realizados com bases de dados de baixa dimensão (10 atributos ou menos). O MEFASD-BD (Pulgar-Rubio, 2017) usa o paradigma MapReduce para particionar o conjunto de dados e uma abordagem de sistema fuzzy evolucionário para a tarefa de mineração de subgrupos. Embora possa lidar com mais atributos que o FSSD, ainda não pode fornecer um desempenho satisfatório com conjuntos de dados com mais de 50 atributos.

Posteriormente, algoritmos foram desenvolvidos com foco em bases de alta dimensionalidade (muitos atributos), como o SSDP e o SSDP+. O SSDP (Lucas, 2017) é um algoritmo evolucionário, top-k e mono-objetivo adaptado para problemas de mineração de subgrupos envolvendo bases de dados de alta di-

mensionalidade e que possui apenas dois parâmetros externos facilmente ajustáveis: o número k de subgrupos que são retornados ao fim de sua execução e a métrica de avaliação. O SSDP faz uso de um conceito semelhante ao do SD para os k melhores subgrupos. No algoritmo, a população P_k mantém apenas os melhores k subgrupos considerados relevantes a fim de minimizar redundância entre os top- k subgrupos.

Já o SSDP+ (Lucas, 2018) é uma evolução do SSDP com os seguintes diferenciais: (i) lida de forma flexível com o problema de redundância, (ii) reduz o risco de perda de informação relevante; e (iii) gera um conjunto de subgrupos mais informativos para o usuário. Uma descrição mais detalhada do SSDP+ será fornecida na seção a seguir.

No SSDP+, parte dos subgrupos considerados redundantes são armazenados para minimizar o risco de informações relevantes serem descartadas e, ao fim, fornecer um conjunto de subgrupos mais informativos para o usuário. Para isso, cada subgrupo s no top- k tem uma cache $s.cache$, cujo tamanho é definido pelo usuário. A $s.cache$ armazena os melhores subgrupos considerados redundantes em relação a s . No SSDP+, dois subgrupos s e s' são considerados redundantes se $sim(s, s') > \min_similarity$, onde sim é uma métrica de similaridade e \min_simi-

larity é um limiar definido pelo usuário. A similaridade entre dois subgrupos é dada pela Equação III, $c^+(s)$ são os exemplos alvo cobertos por um subgrupo s .

Equação III

$$sim_J(s, s') = \frac{|c^+(s) \cap c^+(s')|}{|c^+(s) \cup c^+(s')|}$$

O **Algoritmo 1** é um pseudocódigo do SSDP+. O algoritmo começa gerando a população inicial com todos os possíveis subgrupos $s \in I$ de tamanho $size(s) = 1$, ou seja, todos os subgrupos formados por um item (linha 1).

Algoritmo 1: SSDP+ (pseudocódigo)

```

00 Require: k, metricEvaluation, kc, min_similarity
01  $P \leftarrow \{\{i1\}, \{i2\}, \dots, \{i|I|\}\}$ 
02 for  $i \leftarrow 0$  to  $k$  do
03    $Pk[i] \leftarrow \{\}$ 
04   for  $j \leftarrow 0$  to  $kc$  do
05      $Pk[i].cache.ADD(\{\})$ 
06   end for
07 end for
08 for all  $P[i] \in P | \text{evaluation}(P[i]) > \text{evaluation}(Pk[k-1])$  do
09    $updateTopK(Pk, P[i], min\_similarity)$ 
10 end for
11  $reinitializationCount \leftarrow 0$ 
12  $mutationRate \leftarrow 0.4$ 
13  $crossoverRate \leftarrow 0.6$ 
14 while  $reinitializationCount < 2$  do
15   while  $Pk$  not improve three consecutive generations keeping  $mutationRate == 1.0$  do
16      $Pnew \leftarrow evolutionaryOperator(P, mutationRate, crossoverRate)$ 
17      $P \leftarrow best(P, Pnew)$ 
18     for all  $P[i] \in P | \text{evaluation}(P[i]) > \text{evaluation}(Pk[k-1])$  do
19        $updateTopK(Pk, P[i], min\_similarity)$ 
20     end for
21      $update(mutationRate, crossoverRate)$ 
22   end while
23  $reinitializationCount++$ 
24  $P \leftarrow restart(P)$ 
25 end while
26 return  $Pk$ 

```

Então, o P_k e suas caches são preenchidas com subgrupos vazios (linhas 2-7). Nas linhas 8 a 10, os subgrupos candidatos de P passam pela seleção para entrar no P_k através da função $updateTopK(P_k, P[i], min_similarity)$. Linhas 11 a 13 inicializam as variáveis que controlam o número de reinicializações ($reinitializationCount$) e as taxas de crossover ($crossoverRate$) e mutação ($mutationRate$). As taxas de crossover e mutação são definidas dinamicamente e seus valores são alterados por meio da função $update(mutationRate,$

$crossoverRate)$ (linha 21). A taxa de mutação aumenta e a de crossover diminui quando não há melhora nos subgrupos do top-k. Caso contrário, a taxa de crossover aumenta e a de mutação diminui (o valor dessas variações é de 0.2). As duas taxas podem assumir valores entre 0 e 1, sendo a soma das duas igual a 1.

Das linhas 15 a 22, o algoritmo genético realiza a busca gerando novos subgrupos candidatos ao P_k , atualizando-o a

cada geração. Quando não há mudanças no P_k (nem nas suas caches) por três gerações consecutivas tendo $\text{mutation-Rate} = 1$, a população é resetada (linha 24). O critério de parada do SSDP+ é a estabilização do top-k após a população ter sido resetada duas vezes. Finalmente, na linha 26, os top-k subgrupos (bem como suas caches) são retornados para o usuário.

4. SSDPH: Uma abordagem híbrida

O SSDPH surge para incorporar a busca exaustiva ao SSDP+ até uma dimensão M , escolhida pelo usuário. Dessa forma trata-se de um algoritmo híbrido, que garante o melhor resultado até uma dimensão M , e passa a usar a heurística SSDP+ apenas em dimensões de busca mais complexas. Outro ponto de flexibilização do SSDPH foi poder considerar apenas parte dos exemplos no processo de busca. Ou seja, o usuário tem a opção buscar os melhores subgrupos considerando, por exemplo, 10% dos exemplos de uma base de dados. O objetivo nesse segundo ponto é viabilizar buscas que em bases de dados completas seria computacionalmente inviável.

Assim o algoritmo SSPH opera executando primeiro a busca Exaustiva e, em seguida, o algoritmo do SSDP+. A primeira etapa consiste em executar o algoritmo exaustivo até uma dada dimensão M retornando, ao fim, um top-k que servirá como um dos componentes para gerar a população inicial (formada por indivíduos

com número de itens igual a $M+1$) que será utilizada pelo SSDP+. Essa população inicial é formada por indivíduos obtidos de 3 maneiras : (i) 80% através da geração aleatória de indivíduos utilizando quaisquer itens da base; (ii) 5% através da geração aleatória de indivíduos utilizando apenas os itens do top-k; (iii) 15% através da geração aleatória de indivíduos utilizando itens escolhidos aleatoriamente do top K ou da base completa.

Algoritmo 2: SSDPH (pseudocódigo)

```

00 Require: k, metricEvaluation, kc, min_similarity, maxDimension
01 listOfPks = {}
02 for each partition in Partitions do
03   if partition exists and it's not the whole base do
04     switchAndLoadPartition(partition)
05   end if
06   Pk ← GulosoD.run(k, metricEvaluation, min_similarity, maxDimension)
07   P ← generateInitialPopulation(metricEvaluation, |I|, maxDimension + 1, Pk)
08   for all P[i] ∈ P | evaluation(P[i]) > evaluation (Pk[k-1]) do
09     updateTopK(Pk, P[i], min_similarity)
10   end for
11   reinializationCount ← 0
12   mutationRate ← 0.4
13   crossoverRate ← 0.6
14   while reinializationCount < 2 do
15     while Pk not improve three consecutive generations keeping mutationRate == 1.0 do
16       Pnew ← evolutionaryOperator(P, mutationRate, crossoverRate)
17       P ← best(P, Pnew)
18       for all P[i] ∈ P | evaluation(P[i]) > evaluation(Pk[k-1]) do
19         updateTopK(Pk, P[i], min_similarity)
20       end for
21       update(mutationRate, crossoverRate)
22     end while
23     reinializationCount++
24     P ← restart(P)
25   end while
26   listOfPks += Pk
27 end for
28 return makeOneFinalTopK(listOfPks, k)

```

Após a formação da população inicial, e considerando que nenhum particionamento foi feito, o algoritmo do SSDP+ é executado da mesma forma como foi explicado anteriormente. A única diferença seria ao fim, na linha 26, onde o top-k final é adicionado em um array listOfPks. Contudo, como nenhuma partição foi criada e só existe um top-k, na linha 28 esse top-k será retornado normalmente

Além disso, para a tarefa de pré-processamento, no SSDPH existem os módulos de amostragem e particionamento de base de dados, que são particularmente interessantes para problemas que envolvem bases com muitos exemplos. O primeiro módulo consiste em gerar uma amostra da base conforme um valor percentual de-

terminado pelo usuário. Nesse processo, o usuário configura o tamanho da amostra através de uma porcentagem da base original, podendo também inserir a porcentagem de presença de exemplos positivos na amostra. Na Figura I, temos o exemplo de uma base D com 150 exemplos ($|D|$), sendo 30 positivos ($|D+|$) e 120 negativos ($|D-|$). A partir dessa base, foi extraída uma amostra A de tamanho $|A| = |D| * 0.6 = 90$, isto é, equivalente a 60% do tamanho da base D . Para essa amostra A , foi considerada uma porcentagem de presença de exemplos positivos é de 30%, ou seja, $|A+| = 90 * 0.3 = 27$.



Figura I: exemplos de uso do operador de amostragem

A amostra A é armazenada no array `Partitions`, fazendo com que o loop da linha 02 seja executado uma única vez. Após esse processo, o algoritmo do SSDPH é executado nessa amostra da mesma forma como foi explicado anteriormente. Ao fim da execução, o top-k resultante é adicionado no array `listOfPk` (linha 26) e, no método `makeOneFinalTopK` (linha 28), todos os subgrupos do top-k são avaliados em relação à base completa e, finalmente, retornados.

O segundo módulo é o de particionamento, no qual a base de dados de entrada é dividida em N partições (sendo N um valor fornecido pelo usuário) de forma

que todas as partições sejam conjuntos disjuntos e que tenham a mesma proporção entre exemplos positivos e negativos (ver Figura II). A Figura II ilustra um exemplo considerando a mesma base do exemplo da Figura I e $N = 3$, o que indica que a base original D é dividida em 3 partições: $P1$, $P2$ e $P3$; cada uma delas com quantidade de exemplos igual a $|D|/3 = 50$. A partição $P1$, por exemplo, possui 50 exemplos ($|P1|$) coletados da base original, sendo 10 positivos ($|P1+|$) e 40 negativos ($|P1-|$).

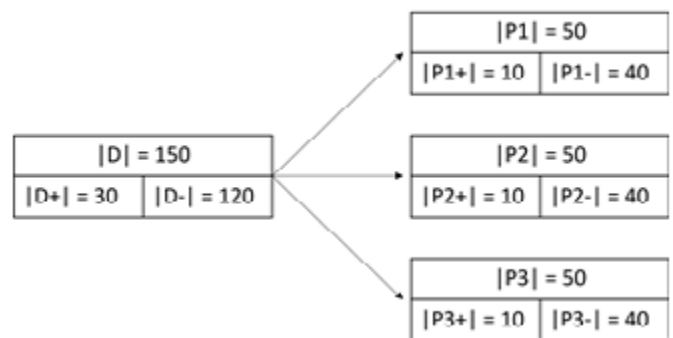


Figura II: exemplo do uso do operador de divisão da base de dados

Todas as N partições são adicionadas ao array `Partitions`, fazendo com que o loop da linha 02 seja executado N vezes. Após esse processo, o algoritmo do SSDPH é executado para cada partição da mesma forma como foi explicado anteriormente. Ao final da primeira iteração, por exemplo, teríamos o $Pk1$, que é o top-k resultante da execução do SSDPH sobre a partição $P1$, sendo adicionado ao array `listOfPk` (linha 26). Da mesma forma, ao final da segunda iteração, teríamos o $Pk2$

também sendo adicionado ao listOfPk, e assim por diante até que o PkN seja adicionado ao listOfPk. Assumindo que o usuário queira somente os 10 melhores subgrupos ($k = 10$), o método makeOneFinalTopK irá avaliar todos os subgrupos em listOfPk em relação à base completa e retornar um top-k final com os 10 melhores desses subgrupos.

5. Experimentos

Para nossos experimentos, usamos 10 bases de dados de alta dimensionalidade, as quais estão descritas na Tabela I. O grupo de bases de Bioinformática foi obtido a partir do pacote datamicroarray (Ramey, 2016) do software R. O grupo de Humanidades/Ciências Sociais consiste em 4 bases, duas sobre educação e duas sobre saúde, todas relacionadas ao contexto de cidades inteligentes. Para todas as bases, os atributos numéricos foram discretizados e o valor majoritário de cada classe foi considerado como alvo.

Tabela I: descrição das bases de dados utilizadas nos experimentos

Nome	D	D ⁺	D ⁻	$\frac{ D^+ }{ D }$	A
<i>Humanidades / Ciências Sociais</i>					
Depressão	34704	2800	31904	0,0807	112
ENEM	10000	2483	7517	0,2483	114
Evasao	118755	18528	100227	0,1560	177
Expectativa de Vida	11619	4969	6650	0,4277	909
<i>Bioinformática</i>					
alon	62	40	22	0,6452	1000
burczynski	127	59	68	0,4646	11142
chin	118	75	43	0,6356	11108
christensen	217	113	104	0,5207	707
gordon	181	150	31	0,8287	6267
matrix binária global	372	68	304	0,1828	7794

Foram testados os seguintes algoritmos e parâmetros nas bases de dados da Tabela 2:

- ED ▸ implementação da busca exaustiva, quando ele for executado para $M = 1$, e ED2, quando for executado para $M = 2$.
- SSDP+ tradicional
- SSDPHD1 e SSDPHD2 ▸ SSDPH com busca exaustiva executada até a dimensão $M = 1$ e $M = 2$, respectivamente.
- SSDPHp20aD1 ▸ SSDPH com amostra aleatória de 20% da base original e $M = 1$
- SSDPHp60aD1 ▸ SSDPH com amostra aleatória de 60% da base original e $M = 1$
- SSDPHp20b50D1 ▸ SSDPH com amostra 20%, sendo que 50% dos exemplos são positivos e $M = 1$
- SSDPHp60b50D1 ▸ SSDPH com amostra 60%, sendo que 50% dos exemplos são positivos e $M = 1$
- SSDPHpn5D1 ▸ SSDPH com 5 partições da base original e $M = 1$
- SSDPHpn10D1 ▸ SSDP com 10 partições da base original e $M = 1$

Os experimentos consistem em executar cada um dos algoritmos nas 10 bases de dados para $k = 10$, $\text{min_similarity} = 0.5$, $k_s = 3$, métricas de avaliação WRAcc e Qg. Cada experimento foi repetido 10 vezes.

Para nossa análise, direcionamos nossa atenção para os seguintes critérios: qualidade média dos subgrupos, suporte global positivo e tempo de processamento. Todos os experimentos foram realizados na mesma máquina com a seguinte configuração: Processador Intel® Core™ i7-6500U CPU @ 2.50GHz; Memória RAM de 16 GB; Sistema operacional de 64 bits; Windows 10 Home Single Language.

6. Resultados

As tabelas II e III mostram os valores médios de Qg, WRAcc, SUPP+, desvio padrão (DP) e tempo dos algoritmos testados. Em ambos os grupos de bases de dados, o uso de amostras, e não do conjunto de dados todo, reduz de forma significativa o tempo de processamento, mas frequentemente representa uma perda na qualidade da busca. Ainda em ambos os grupos de bases, o tempo de processamento ED2 e SSDPH com M=2 são consideravelmente maiores que as demais abordagens. Já a divisão da base de dados em N grupos de dados normalmente resultou em uma piora na qualidade média dos subgrupos, embora nas bases de Humanidades/Ciências Sociais tenha havido uma melhoria para a métrica WRAcc como pode ser vista na Tabela III.

Tabela II: resultados dos algoritmos testados nas bases de Bioinformática

Bioinformática									
Algoritmo	WRAccN	SUPP+	Tempo(s)	DP	Qg	SUPP+	Tempo(s)	DP	
ED1	0.352	0.987	0.202	0.121	26.089	0.900	0.235	21.216	
ED2	0.382	0.966	2288.229	0.118	35.550	0.903	2279.370	22.815	
SSDP+	0.426	0.989	210.128	0.062	33.972	0.871	14.048	22.814	
SSDPHD1	0.418	0.996	332.385	0.060	32.957	0.893	13.115	22.513	
SSDPHD2	0.431	0.971	2595.652	0.062	35.570	0.906	2272.094	22.776	
SSDPHp20aD1	0.230	0.985	12.764	0.090	9.251	0.921	2.905	9.933	
SSDPHp60aD1	0.358	0.987	81.025	0.059	21.387	0.909	6.574	18.167	
SSDPHp20b50D1	0.204	0.983	5.673	0.086	11.228	0.938	3.351	13.696	
SSDPHp60b50D1	0.350	0.988	98.003	0.066	23.702	0.889	7.712	20.265	
SSDPHp5D1	0.323	0.988	32.865	0.083	18.346	0.897	14.594	16.062	
SSDPHp10D1	0.305	0.980	23.762	0.096	17.005	0.906	16.020	16.541	

Tabela III: resultados dos algoritmos testados nas bases de Humanidades

Humanidades/Ciências Sociais									
Algoritmo	WRAccN	SUPP+	Tempo(s)	DP	Qg	SUPP+	Tempo(s)	DP	
ED1	0.112	0.995	11.445	0.080	7.675	0.249	10.660	10.665	
ED2	0.163	0.908	2497.915	0.118	20.578	0.042	2519.408	23.622	
SSDP+	0.175	0.881	400.611	0.134	31.958	0.038	285.858	34.808	
SSDPHD1	0.171	0.885	419.042	0.129	27.265	0.034	257.150	29.146	
SSDPHD2	0.172	0.883	2886.367	0.130	31.793	0.040	2780.652	36.058	
SSDPHp20aD1	0.169	0.875	89.700	0.129	7.351	0.087	63.827	8.520	
SSDPHp60aD1	0.170	0.881	231.405	0.128	17.083	0.047	184.039	17.797	
SSDPHp20b50D1	0.166	0.867	76.810	0.125	5.127	0.087	84.996	5.719	
SSDPHp60b50D1	0.169	0.887	261.928	0.128	16.090	0.051	205.178	18.858	
SSDPHp5D1	0.176	0.882	461.016	0.134	13.325	0.083	322.749	16.231	
SSDPHp10D1	0.177	0.888	567.172	0.135	11.908	0.094	390.834	13.435	

Ainda de uma forma geral, os resultados mostram que é comum um algoritmo ser melhor em qualidade (ex. Qg) e pior em diversidade (SUPP+). Por fim, com relação a qualidade dos subgrupos, a abordagem proposta SSDPH com parâmetro M=2 obteve os subgrupos com qualidade média maior ou próximo do melhor em relação ao SSDP+ tradicional e as demais abordagens, embora com tempo de processamento consideravelmente alto.

Figura III: Box Plot – Valores de Qg x Algoritmos

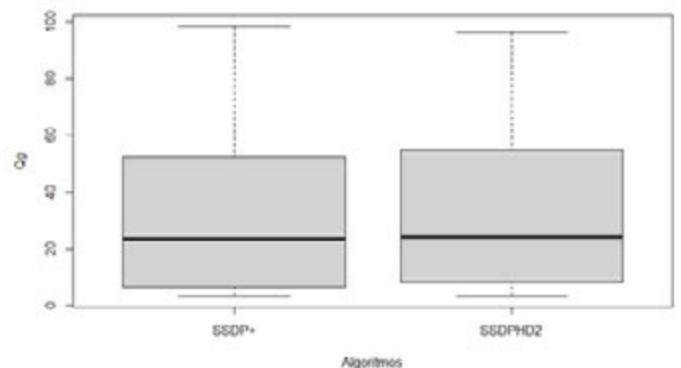
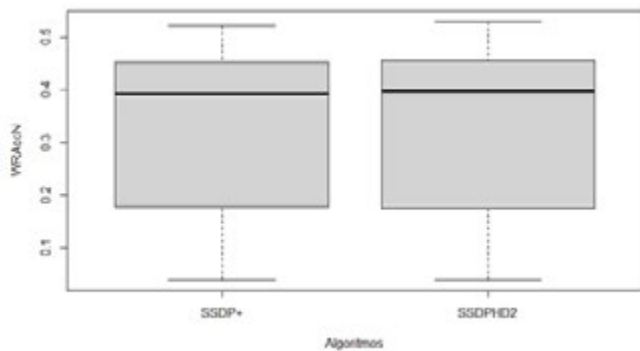


Figura IV: Box Plot – Valores de WRAcc x Algoritmos



Os gráficos das figuras III e IV também ilustram uma pequena vantagem do SSDPH ($M = 2$) em relação ao SSDP+, considerando as métricas Qg e WRAcc respectivamente. Aplicamos o teste de Wilcoxon a fim de saber se essas diferenças são estatisticamente significantes. O teste de Wilcoxon é um teste não-paramétrico que foi indicado e usado para análise de performance entre dois algoritmos (DEM-SAR, 2006). A hipótese nula é a de que o desempenho do SSDP+ é tão bom quanto o do SSDPH ($M = 2$). Tal hipótese foi rejeitada tanto considerando o Qg quanto o WRAcc. O teste de Wilcoxon foi aplicado com 95% de confiança ($\alpha = 0.05$) e os p-values retornados foram de $2.494e-05$, para a métrica Qg, e de 0.04773 , para a métrica WRAcc, comprovando uma diferença estatística na performance entre os dois algoritmos.

Figura V: Box Plot – Tempo x Algoritmos (Qg)

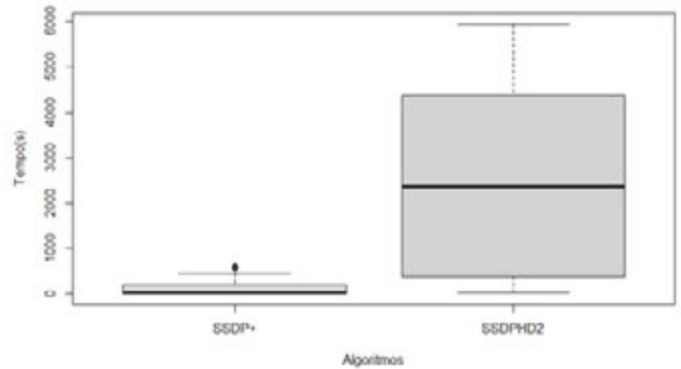
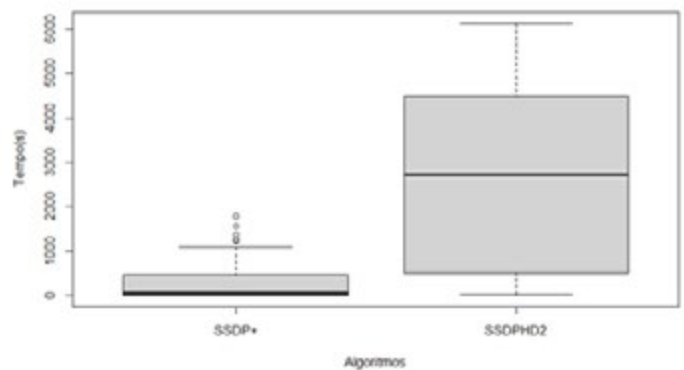


Figura VI: Box Plot – Tempo x Algoritmos (WRAcc)



Por outro lado, os gráficos das figuras V e VI mostram uma desvantagem do SSDPH ($M=2$) quando comparado com o SSDP+, considerando o tempo de execução para a métrica Qg. A aplicação de teste Wilcoxon ($\alpha = 0.05$) considerando o tempo dos algoritmos (para Qg) confirma a significância dessa diferença, retornando um p-value de $2.2e-16$.

7. Conclusão

Esse trabalho apresentou o SSDPH uma nova abordagem para o problema de mineração de subgrupos considerando o contexto de bases de dados com muitos exemplos e dimensões. O SSDPH é uma evolução do SSDP+ por trazer mais flexibilidade ao já bem-sucedido algoritmo. Com o SSDPH o usuário pode escolher entre: (1) lidar com todo o conjunto de dados ou com parte dele, (2) lidar com todos os dados de uma vez ou realizar uma busca top-k para cada uma das N partes de uma base de dados e (3) decidir realizar busca exaustiva até a dimensão M e só depois aplicar o SSDP+.

Dessa forma, é normal que o modelo proposto seja competitivo ou mesmo supere o SSDP+ tradicional com maior tempo de simulação por um lado, e que seja rápido ao utilizar menos dados, mas perdendo na qualidade dos subgrupos por outro. O modelo proposto buscou flexibilizar um algoritmo SSDP+ antes focado em bases de dados de alta dimensionalidade, ser uma solução eficiente também em contextos antes não aplicáveis (bases de dados muito grandes) e, no caso de base mais simples, trazer a possibilidade de minerar subgrupos de forma exaustiva até a dimensão 2 ou 3, por exemplo.

Sendo assim, o SSDPH se mostra um projeto promissor, mas que possui algumas limitações (como o fato de não suportar

atributos numéricos) e várias possibilidades de melhorias. Para implementações futuras desse projeto, é necessária uma investigação acerca dos melhores parâmetros para a formação da população inicial, que é um problema não trivial e que tem muito efeito sobre a qualidade do conjunto final de subgrupos.

8. Referências

- BELFODIL, A., BELFODIL, A., BENDIMERAD, A., LAMARRE, P., ROBARDET, C., KAYTOUE, M., & PLANTEVIT, M. (2019, October). Fssd-a fast and efficient algorithm for subgroup set discovery. In 2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA) (pp. 91-99). IEEE.
- CARMONA, C. J., CHRYSOSTOMOU, C., SEKER, H., & DEL JESUS, M. (2013). Fuzzy rules for describing subgroups from influenza A virus using a multi-objective evolutionary algorithm. *Applied Soft Computing*, 13(8):3439–3448
- DEL JESUS, M. J., GONZALEZ, P., HERRERA, F., & MESSONERO, M. (2007). Evolutionary fuzzy rule induction process for subgroup discovery: a case study in marketing. *IEEE Transactions on Fuzzy Systems*, 15(4):578–592
- FLACH, P. A. (2003). The geometry of ROC space: understanding machine learning metrics through ROC isometrics. In *Proceedings of the 20th international conference on machine learning (ICML-03)* (pp. 194-201).
- GARCÍA-VICO, Á. M., González, P., Carmona, C. J., & del Jesus, M. J. (2019). Study on the use of different quality measures within a multi-objective evolutionary algorithm approach for emerging pattern mining in big data environments. *Big Data Analytics*, 4(1), 1-15.
- GAMBERGER, D., & LAVRAC, N. (2002). Expert-guided subgroup discovery: Methodology and application. *Journal of Artificial Intelligence Research*, 17, 501-527.
- HERRERA, F., CARMONA, C. J., GONZÁLEZ, P., & DEL JESUS, M. J. (2011). An overview on subgroup discovery: foundations and applications. *Knowledge and information systems*, 29(3):495–525
- KAVŠEK, B., LAVRAC, N., & BULLAS, J. C. (2002). Rule induction for subgroup discovery: a case study in mining UK traffic accident data. In *Proceedings of the international multi-conference on information society*, 127–130
- LUCAS, T., VIMIEIRO, R., & LUDERMIR, T. (2018, July). SSDP+: a diverse and more informative subgroup discovery approach for high dimensional data. In 2018 IEEE Congress on Evolutionary Computation (CEC) (pp. 1-8). IEEE.
- LUCAS, T., SILVA, T. C., VIMIEIRO, R., & LUDERMIR, T. B. (2017). A new evolutionary algorithm for mining top-k discriminative patterns in high dimensional data. *Applied Soft Computing*, 59:487–499
- PULGAR-RUBIO, F., RIVERA-RIVAS, A. J., PÉREZ-GODOY, M. D., GONZÁLEZ, P., CARMONA, C. J., & DEL JESUS, M. J. (2017). MEFASD-BD: multi-objective evolutionary fuzzy algorithm for subgroup discovery in big data environments-a mapreduce solution. *Knowledge-Based Systems*, 117, 70-78.
- RAMEY, J. (2016). “The datamicroarray r package” [Online] Available: <https://github.com/ramhiser/datamicroarray>
- DEMSAR, J. (2006). Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine learning research*, 7, 1-30.

Internet das coisas na gestão educacional: inovação da gestão acadêmica e desenvolvimento tecnológico na rede itego – Goiás - Brasil

Douglas Araújo Falcão (Mestrando em Inovação e Empreendedorismo - Escola de Negócios Europeia de Barcelona – ENEB, E-mail: msdouglasfalcao@gmail.com, Orcid: <https://orcid.org/0000-0003-1191-0221>)

Abstract. This work reports the results of the experience of digital transformation using the Internet of Things and technological resources through the implementation of modern management to achieve quantitative and qualitative goals in the ITEGO Network - Network of Technological Institutes of the State of Goiás, Brazil - in the cities of Anápolis and Catalão. It was possible to prove the relevance of planning, monitoring and evaluation to reach qualitative and quantitative goals in Technological Development and Innovation actions. Recent historical data obtained from SIGA - computerized academic management system, Q-LINK SENSE - BI - computerized system for monitoring quantitative goals and monitoring system for qualitative goals were analyzed. This work concludes with the success of the shared management process, even though it brings together competitive efforts for Vocational and Technological Education in parallel with Technological Development and Innovation actions.

Keywords: educational management, planning, internet of things, technological development and innovation.

Resumo. Este trabalho relata os resultados da experiência da transformação digital com uso da Internet das Coisas e recursos tecnológicos por meio da implantação de uma gestão moderna para alcance de metas quantitativas e qualitativas na Rede ITEGO - Rede de Institutos Tecnológicos do Estado de Goiás, Brasil - nas cidades de Anápolis e Catalão. Foi possível comprovar a relevância do planejamento, monitoramento e da avaliação, para o alcance de metas qualitativas e quantitativas nas ações de Desenvolvimento Tecnológico e Inovação. Foram analisados dados históricos recentes obtidos do SIGA - sistema informatizado de gestão acadêmica, Q-LINK SENSE – BI – sistema informatizado de monitoramento de metas quantitativas e sistema de monitoramento de metas qualitativas. Este trabalho se conclui pelo sucesso no processo de gestão compartilhada, ainda que reúna esforços competitivos para a Educação Profissional e Tecnológica paralela às ações de Desenvolvimento e Inovação Tecnológica.

Palavras-chave: gestão educacional, planejamento, internet das coisas, desenvolvimento e inovação tecnológica.

1. Introdução

A Educação Profissional e Tecnológica (EPT), prevista na Lei de Diretrizes e Bases da Educação Nacional (LDB) (BRASIL, 1996), deve preparar o cidadão para o exercício profissional e, para isso, busca capacitá-lo e profissionalizá-lo para seu ingresso e permanência no mundo do trabalho e da vida em sociedade, mas em sintonia com as demandas da industrialização e da era da informação (BRASIL, 2020).

Moura, Filho e Silva (2015) discutem a expansão dos institutos federais, e aponta a presença do Estado brasileiro por meio de instituições de educação profissional nas periferias das capitais e em regiões mais afastadas dos grandes centros urbanos, ampliando as possibilidades de muitos brasileiros para acessar as habilidades necessárias.

Mauro Sérgio (DE et al., 2020) afirma que:

“A ampliação da rede e, por consequência, a ampliação da oferta de ensino profissionalizante e superior se baseou na premissa de fornecer instrumentos renovados no desenvolvimento local e regional”.

Neste caso de estudo, o Governo do Estado de Goiás sempre implementou políticas públicas de EPT por meio da Secretaria de Desenvolvimento e Inovação (SEDI) até 2014, por meio de seus Centros de Formação Profissional - CEPA. Em julho

de 2015, por meio da Lei nº 18.931, o Governo do Estado de Goiás criou os Institutos Tecnológicos do Estado de Goiás - ITEGOS e Escolas Tecnológicas - CO-TECS, uma rede com 30 (trinta) ITEGOS.

O objetivo deste artigo é analisar a eficácia dos processos de Planejamento e Avaliação Educacional, na implementação da gestão compartilhada em ações de DIT, nas unidades da REDE ITEGO, nas cidades de Anápolis e Catalão, em Goiás, Brasil.

Considerando a complexidade da gestão e operacionalização desta rede ITEGOS, o Governo de Goiás optou por realizar uma Chamada Pública, para que a gestão seja compartilhada entre o Parceiro Público - SEDI e o Parceiro Privado - Organização Social devidamente credenciada pelo Estado de Goiás. Assim, a “Chamada Pública 07-2016”, regida pela Lei Estadual de Goiás 15.505, de 28 de dezembro de 2005, estabeleceu os critérios norteadores dessa operação, que consistiu na implantação do modelo de gestão compartilhada entre a SEDI e uma Organização Social.

Além disso, esse modelo de gestão compartilhada foi baseado em critérios claros e estabelecidos, como metas quantitativas de cursos a serem ofertados nas categorias de formação, qualificação, técnico e de tecnologia superior nas modalidades de educação presencial e a distância (EaD).

Em relação à educação a distância, a NASU (2020) destaca que:

“É preciso considerar também o contexto da sociedade contemporânea em que ocorreram as mudanças na forma de comunicação representadas pelo fenômeno da cibercultura e do ciberespaço em que as tecnologias digitais de informação e comunicação (TDIC) sustentam novas formas de os indivíduos se relacionarem em sociedade, também possibilitar novas formas de trabalhar, ensinar e aprender”.

Portanto, o escopo do projeto também estabeleceu as Atividades Práticas Acadêmicas Complementares (APA) como um diferencial no projeto pedagógico do ITEGO, visando a formação integral do aluno, alinhando a teoria prevista nos planos de curso, com a experiência com no mundo técnico-científico e no trabalho.

Os ITEGOs também devem prestar serviços de DIT, por meio da educação, empreendedorismo, pesquisa e extensão das atividades de inovação e transferência de tecnologia para o setor produtivo. Isso inclui atividades de inovação e transferência de tecnologia, bem como a prestação de serviços tecnológicos, com o objetivo de promover e fortalecer ambientes de inovação em Incubadoras de Empresas, Arranjos Produtivos Locais (APLs) e outros segmentos do setor produtivo (FALCÃO, 2021).

SANTOS (2019) realizou relevante estudo da REDE ITEGO na cidade de Catalão, e observou que:

“A transferência de tecnologia ocorre por meio do conhecimento, do desenvolvimento e da tecnologia pelo ambiente que não a gerou. Em decorrência dessa perspectiva, é necessário que a inovação seja transferida por meio da ação de transferência de tecnologia, e os mecanismos de inovação oferecem estratégias competitivas para o desenvolvimento das empresas, tornando-as mais competitivas para o mercado. As interações entre os diferentes atores (instituições de pesquisa, universidades, governo e indústria) são fundamentais para o sucesso do modelo de inovação”.

Novas tecnologias, tanto para aulas presenciais quanto para educação a distância, também começaram a ser empregadas devido à demanda por cursos mais flexíveis, executados a distância e orientados de forma assíncrona, como evidencia Do Nascimento (2020.):

“[...] uma nova era nos processos educativos, com o crescente desenvolvimento de recursos midiáticos para aplicação no processo de ensino-aprendizagem, até que se desenvolvessem as condições necessárias para que o próprio processo se tornasse plenamente mediado por ferramentas tecnológicas, com o advento de Educação a Distância (DE) ”.

Dentre todos os elementos de monitoramento e controle das ações de DIT, destacam-se:

- i. As projeções dos planos para atingir metas quantitativas e qualitativas;
- ii. Avaliação periódica para atingir metas quantitativas e qualitativas.

As metas quantitativas estabelecidas para acompanhamento diário, semestral e anual computam: matrícula de alunos, horas de treinamento de alunos, alunos atendidos na APAS, horas de DIT.

Para que a execução das metas quantitativas ocorra de forma sustentável e se torne a cultura organizacional da rede REDE ITEGO, foram estabelecidas algumas dimensões e aspectos de avaliação ou indicadores.

Para garantir que o monitoramento e a fidedignidade dos dados referente à metas, sejam de qualidade ou de quantidade, devido ao seu volume a serem cumpridos anualmente percebeu-se a necessidade da implementação de ferramentas que possibilitassem não apenas a consolidação de dados inicialmente geridos em planilhas em Excel, mas sistemas que fossem inteligentes para colaborar com a tomada de decisão na gestão e nos processos do projeto.

Nesse contexto percebeu-se que implantar sistemas baseado no conceito da Internet das Coisas traria um diferencial, considerando que por se tratar de um gargalo que era o preenchimento de diário de aula por professores por exemplo tal ação pu-

desse ser realizada por smartphone. Outro aspecto considerado foi o fato da alta demanda das matrículas e cursos realizados em EaD que também deveriam ser utilizadas em dispositivos móveis. Como destaca Sundmaeker et al. 2010:

Ao conectar objetos com diferentes recursos a uma rede, potencializa-se o surgimento de novas aplicações. Neste sentido, conectar esses objetos à Internet significa criar a Internet das Coisas. Na IoT, os objetos podem prover comunicação entre usuários, dispositivos. Com isto emerge uma nova gama de aplicações, tais como coleta de dados de pacientes e monitoramento de idosos, sensoriamento de ambientes de difícil acesso e inóspitos, entre outras.

Assim foram desenvolvidas e implantadas as ferramentas:

- SIGA – Sistema informatizado de Gestão Acadêmica:
- QLink Sense - BI de Monitoramento de Metas Quantitativas.

2. Metodologia

A metodologia de pesquisa baseou-se no método hipotético-dedutivo, por meio da construção de conjecturas (MARCONI & LAKATOS, 2017). Foi construído a partir de dados relacionados ao DIT, coletados no portal da transparência e SER/SEDI, no período de 2017 a 2020, com abrangência limitada às cidades polo de Anápolis e Catalão, ambas no interior de Goiás, Brasil.

Os dados finais foram coletados nos relatórios anuais de avaliação de metas quantitativas e qualitativas que emitido pela Comissão de Avaliação das Secretárias de Estado Desenvolvimento e Inovação e Retomada com base nos dados que foram alimentados pelos ITEGOS/COTECs no SIGA/Q-LINK, e partir de Documentos Institucionais, como:

- i. Número atual de funcionários;
- ii. Número de alunos matriculados;
- iii. Quantidades do Curso;
- iv. Quantidade de alunos atendimentos em Atividades Prático Acadêmicas – APA;
- v. Quantidade de Horas de DIT.

Relativo ao primeiro ano de celebração do contrato de gestão, em 2017, foi analisado o Plano de Assunção Imediata, apresentado ao Parceiro Público aquando da assinatura do “Contrato de Gestão 01-2017-SED”. Este Plano foi o primeiro planejamento que orientou como seria a gestão operacional.

Seguindo o Plano de Assunção Imediata, foi elaborado um Planejamento Educacional e de Avaliação, que também foi objeto de análise para este trabalho.

Outro instrumento utilizado para este trabalho foi a Pesquisa de Satisfação com os alunos, constituindo o instrumento de coleta elaborado com base nas dimensões e aspectos de avaliação ou indicadores já estabelecidos pelo parceiro público no

“Chamada Pública 07-2016”.

Para melhor acompanhamento, foi estabelecido um fluxo de coleta de informações, e a utilização do SIGA (sistema informatizado de gestão acadêmica), para obter dados diários do professor e do aluno, e o acompanhamento da evasão escolar (DE REZENDE GUEDES & DA VEI-GA JARDIM FILHO, 2020).

Por fim, os dados estatísticos foram extraídos do SIGA, que armazena e processa os históricos acadêmicos de cursos, eixos, componentes, turmas e alunos.

Para facilitar o entendimento desses objetivos, foi seguido o método Balanced Scored Card - BSC. Assim, foi elaborado um Manual de Indicadores para capacitar colaboradores e gestores, e avaliado seu Mapa Estratégico (DE REZENDE GUEDES, 2020).

3. Resultados Alcançados

A partir da análise dos Documentos Institucionais, constatou-se que o planejamento quantitativo da DIT tinha, de fato, um viés qualitativo, por exemplo: para cada curso planejado, havia uma elaboração de atividades e materiais que eram gerados na percepção de alunos, professores e a comunidade.

Constatou-se que as atividades planejadas da DIT buscaram preencher lacunas

no ecossistema de inovação local, formado por todas as associações que atuam na promoção, apoio, financiamento, execução de atividades que resultem em produtos, processos e negócios inovadores.

Com base nos critérios de avaliação, foi possível extrair, do SIGA, o atingimento de metas quantitativas e qualitativas, conforme tabelas abaixo.

Tabela 1: Metas Quantitativas e Qualitativas de Dez/2017 a Jul/2021.

Fonte: SIGA/ Q-Link Sense, 2021

Ano	Pesquisa de Satisfação	Metas Qualitativas Alcançadas	Metas Quantitativas Alcançadas
2017	89,00%	85,05%	82,24%
2018	82,45%	85,31%	79,12%
2019	86,35%	88,57%	90,20%
2020	87,00%	70,60%	97,16%
2021	92,00%	81,01%	100%

Tabela 2: Metas Quantitativas e Qualitativas de Dez/2017 a Jul/2021.

Fonte: SIGA/Q-Link Sense, 2021

Modalidade	Categoria	Metas	Execução	Percentual de Atingimento %
Presencial	Superior	811	689	85%
	Técnico	3608	1827	51%
	Qualificação	14.317	12555	88%
	Capacitação	21.486	22189	103%
EaD	Técnico	864	893	103%
	Qualificação	8.745	9249	106%
	Capacitação	22.325	23128	104%
TOTAL		72156	70530	98%
Outras Ofertas	DIT	34560	34041	98%
	APA	6720	7925	118%

É importante destacar que, apesar dos esforços do Estado de Goiás e da Entidade Parceira para atingir as metas do EPT, que desafiaram os esforços do DIT; houve correspondência em percentuais de metas alcançadas, tanto para EPT quanto para DIT, inclusive com destaque para as atividades da APA.

Percebeu-se que outro desafio das ações da DIT, foi a sensibilização para o empreendedorismo dos alunos e da comunidade atendida, que tinha como finalidade a transformação social e econômica dessas sociedades.

Como parte importante dos processos de gestão a inovação tecnológica alinhada a IoT – Internet das Coisas que foram os sistemas implantados SIGA – Sistema informatizado de Gestão Acadêmica, que não somente foi possível registrar os dados de produção acadêmica e de alunas mas também gerar algoritmos para tomadas de decisão quanto ao planejamento de dimensionamento de sala de aula, demanda de professores, modelo e formato de eventos, perfil dos alunos, essa ferramenta de forma inteligente ajudou melhorar a performance de trabalho da equipe pedagógica e administrativa para a tomada de decisões estratégica, principalmente para a redução da evasão escolar é um dos maiores problemas deste país, pois os modelos matemáticos demonstravam as tendências.

Figura 1 – Tela de Login SIGA – Sistema Informatizado de Gestão Acadêmica

Fonte: Tela login SIGA, 2021



Figura 2– Tela de Turmas SIGA – Sistema Informatizado de Gestão Acadêmica

Fonte: SIGA, 2021

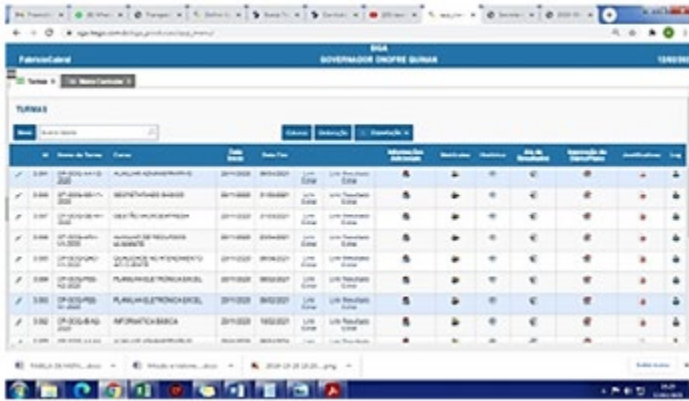
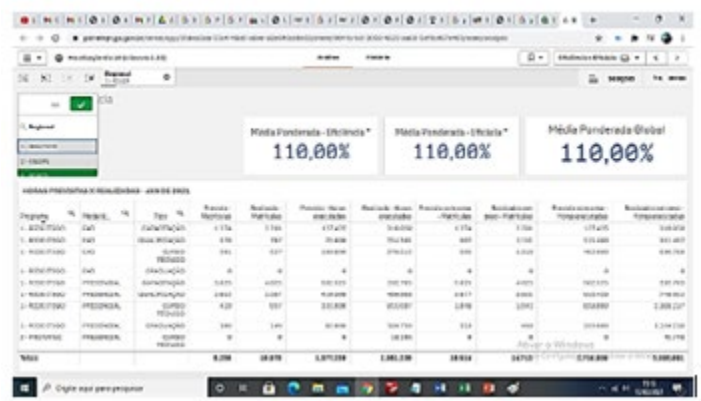


Figura 4 – Tela QLink Sense - BI de Monitoramento de Metas Quantitativas - Painel de Avaliação de Metas

Fonte: Tela o QLink Sense



Já a ferramenta QLink Sense - BI de Monitoramento de Metas Quantitativas, além do API de integração com o SIGA, ela era alimentada diariamente possibilitando aos gestores o padrão de comportamento das metas quantitativas. Por ser um ferramenta com possibilidade de ser acessada de qualquer tipo de dispositivo, garantiu um acompanhamento sistêmico e integral de toda a equipe tática e estratégica, oportunizando encontrar possíveis causa raiz, subsidiando a tomada decisão para mitigar qualquer desvio de rota traçado no atingimento das metas propostas no Pacto pela Educação Profissional e Desenvolvimento Tecnológico em Goiás.

Figura 3 – Tela o QLink Sense - BI de Monitoramento de Metas Quantitativas

Fonte: Tela o QLink Sense



Constatou-se que foi estabelecido um cronograma de atividades de treinamento para colaboradores e técnicos administrativos focados. Oficinas e rodadas de planejamento de forma sistemática e periódica, pois a princípio devido as metas quantitativas serem altas para o pleno cumprimento. A Política de Educação Permanente e Desenvolvimento de Recursos Humanos foi fundamental, para que os colaboradores estivessem inseridos nas diretrizes e programas a desenvolvidos, objetivos a serem alcançados e também a atualização permanente nas novas tecnologias que foram implementadas durante o projeto.

4. Considerações Finais

A partir do estudo dos Documentos Institucionais, constatou-se que a REDE ITEGO promoveu ações e serviços de DIT, de forma efetiva e eficiente, para o alcance das metas estabelecidas, com foco no potencial e vocação de cada Arranjo Produtivo Local (APL) dentro de Goiás, por meio de três mecanismos:

i. Transferência de Tecnologia e Inovação: Envolveu atividades de interação e troca de conhecimento voltadas às empresas dos respectivos APLs, com foco nos gargalos para o desenvolvimento tecnológico e a inovação;

ii. Prestação de Serviços Tecnológicos: Oferecer serviços tecnológicos raros ou inexistentes nas regiões atendidas, o que permitiu aos empresários, principalmente micro e pequenos, acessar equipamentos e laboratórios de alta tecnologia para realizar atividades produtivas colaborativas, testes e ensaios;

iii. Fortalecimento de Ambientes de Inovação: Estabelecer espaços propícios à geração e fortalecimento de negócios inovadores nascentes, como Incubadoras de Empresas, Startups e Parques Tecnológicos.

Ressalta-se que, diante de todas as avaliações, os aspectos fundamentais para o sucesso do modelo de gestão compartilhada para a DIT foram:

i. Implantação do sistema informatizado de gestão acadêmica SIGA, desenvolvido por empresa contratada e customizado especificamente para uso na rede da REDE ITEGO;

ii. Implantação do sistema Q-Link Sense de monitoramento de metas quantitativas, desenvolvido por empresa contratada e customizado especificamente para uso na REDE ITEGO;

iii. Padronização de processos e procedimentos, atividades de gestão participativa formando Governança;

iv. Treinamento e formação continuada de colaboradores;

v. Padronização do acompanhamento das metas quantitativas e qualitativas de horas de DIT e alunos da APA por meio de sistemas informatizados e dashboard;

vi. Ações corretivas pela gerência de controle e avaliação.

Bibliografia

BRASIL. Lei 9.394, de 20 de dezembro de 1996. Estabelece as diretrizes e bases da educação nacional.

. Ministro da Educação. Educação Profissional e Tecnológica (EPT).

DE REZENDE GUEDES, Leonardo Guerra; DA VEIGA JARDIM FILHO, José Leopoldo. On the Evaluation of Companies Maturity for Implementing Strategic Information Systems Planning. *Brazilian Journal of Development*, v. 6, n. 11, p. 84371-84383, 2020.

DE REZENDE GUEDES, Leonardo Guerra et al. On the Evaluation of Companies' Managerial Maturity. *Brazilian Journal of Business*, v. 2, n. 3, p. 2473-2483, 2020.

FALCÃO, DOUGLAS. Desenvolvimento Regional Por Meio de Transferência de Tecnologia: Avaliação do Arranjo Produtivo Local de Confeção de Catalão Vinculado d Rede COTEC, Goiás, Brasil. *Revista de Desenvolvimento e Inovação Tecnológica* v.2 n.1. 2021.

GOIÁS. Lei 15.505, de 28 de dezembro de 2005. Dispõe sobre a qualificação de entidades como organizações sociais estaduais.

. Lei 18.931, de 08 de julho de 2015. Cria e nomeia os Institutos Tecnológicos do Estado de Goiás - ITEGOS.

DE, M. S. P. dos S. et al. The expansion of vocational and higher education in Rio de Janeiro's federal network of vocational education. *Brazilian Journal of Development*, v. 6, n. 1, p. 624-637, 2020.

MARCONI, M. de A.; LAKATOS, E. M. *Scientific methodology*. São Paulo: Ed. Atlas, 2017.

MOURA, D. H.; LIMA FILHO, D. L.; SILVA, M. R. Polytechnic and integrated training: conceptual confrontations, political projects and historical contradictions in Brazilian education. *Brazilian Journal of Education*, v. 20, n. 63, p. 1057-1080, 2015.

SANTOS, V. B. dos. ITEGO network of innovative entrepreneurship and creation of new businesses in the municipality of Catalão-GO. Master's Dissertation, Federal University of Goiás, 2019.

Sundmaeker, H., Guillemin, P., Friess, P., and Woelfflé, S. (2010). *Vision and challenges for realising the Internet of Things*, volume 20. EUR-OP.



Gestão financeira de empreendedores rurais do Município de Pires do Rio/GO

Auto de Paula Rodrigues Neto (Economista pela UEG; ndp.netodepaula@gmail.com)

Ezequiel de Paula Alves Rodrigues (Empreendedor pela UNOPAR; ezequieldepaula1@hotmail.com)

Resumo: O agronegócio é um setor de atividade econômica com muito impacto no Brasil, movimentando quase 25% do PIB nacional, colocando o Estado de Goiás em 9º lugar entre os maiores PIB nacional, mostrando a força do setor agrícola no Estado. Sendo assim, existe uma necessidade para que os empreendedores rurais tenham em suas propriedades uma atenção especial com a gestão financeira da propriedade rural, com ferramentas que possam contribuir para a boa administração. Este trabalho tem como objetivo realizar um levantamento das ferramentas básicas de gestão financeira que são utilizadas nas pequenas propriedades rurais do município de Pires do Rio/GO. Foi aplicado um questionário online e presencial para identificar a gestão financeira das propriedades rurais da região e para que o questionário alcançasse o maior número de propriedades, utilizou-se o aplicativo do WhatsApp para a circulação do link do questionário. Os resultados mostraram que apenas 40% das propriedades rurais têm um gerente ou um responsável pela propriedade e que as ferramentas mais utilizadas são o fluxo de caixa e o controle de custos, mostrando que a gestão financeira é carente e que há espaço para a qualificação profissional e pessoal dos responsáveis pela gestão.

Palavras-chave: agronegócio, administração rural, fluxo de caixa.

Introdução

Com o crescimento do agronegócio no mundo e a importância indiscutível para o comércio nacional com a sua produção de alimentos para abastecer a grande necessidade alimentícias, os governos têm investido no setor de agronegócio. Segundo o Ministério da Agricultura e Abastecimento (2017), os grandes e pequenos produtores do Brasil movimentaram 24% do PIB (Produto Interno Bruto) brasileiro em 2016, com perspectivas para ultrapassar um terço do PIB em 2020.

Os dados econômicos mostram que uma gestão rural é uma necessidade para o setor, sabendo que grande parte dos pequenos e grandes produtores não têm escolaridade ou um conhecimento técnico em gestão de propriedades rurais, principalmente financeira. Segundo Rosa e Garcia (2015) há uma necessidade do aprendizado e conhecimento de ferramentas de gestão para obter a maximização do resultado de produção e produtividade dentro da atividade exercida.

Segundo Gräf (2016), a administração rural nada mais é do que o controle e gerenciamento das atividades rurais em que buscam manter o negócio ativo e lucrativo a fim de cobrir os gastos e gerar reservas para a renovação de investimentos necessários à realidade econômica do setor.

Na administração rural uma das partes que se deve levar em consideração é a questão da utilização de ferramentas financeiras. Segundo Frühanf (2014), devido às grandes mudanças ocorridas no setor agropecuário nos últimos anos, o empreendedor rural precisou passar a se preocupar com questões que vão além do simples cultivo agrícola ou da produção pecuarista. A alta competitividade, as relações trabalhistas, as questões ambientais, a política tributária e as margens de lucro cada vez mais reduzidas, fizeram com que este setor voltasse sua atenção para a gestão da atividade.

No Estado de Goiás a administração rural é uma realidade apenas para as grandes propriedades onde há exigências contábeis para prestação de contas de sócios e cotistas, exigindo eficiência na gestão, entretanto, para as pequenas propriedades, em sua maioria são compostas pela agricultura familiar.

Para Goiás, o setor agrícola é um fator importante na geração de empregos, renda e principalmente, contribui com a balança comercial do Estado. Para os próximos anos, o cenário tende a voltar a ser su-

peravitário, pois de acordo com o mistério da agricultura, Pecuária e Abastecimento, espera-se um aumento na produção dos principais produtos agrícolas de Goiás, no entanto a uma concentração das principais atividades agropecuárias no Estado deixando os demais com um baixo grau de produtividade e de geração de empregos e renda oriundos do setor agrícola. (ARAÚJO. et al, 2013)

Justificativa

O empreendedor rural goiano, por cultura regional e dificuldades no sistema nacional para acesso a informação, tem como característica que essa categoria empreendedora possui pouca escolaridade, situação que pode ter sido gerada pelos familiares que não tiveram acesso a informação e que vão replicando essa situação ao longo do tempo, chegando a filhos e netos, mesmo nos dias de hoje.

Independente do setor, categoria, nível empreendedor ou situação econômica, existem algumas ferramentas básicas que podem auxiliam o empreendedor rural na gestão do seu cultivo. Muitos empreendedores não sabem ao certo qual a situação do seu negócio em que gerencia, não sabe se está gerando lucro, não sabe se está gerando prejuízo, não sabem se está próximo da falência ou até acham que está indo muito bem, mas a palavra “ACHA” é uma armadilha, principalmente para as pessoas que não fazem o mínimo de registros financeiros. O primeiro

registro que é um mandamento na área financeira, é o controle de ENTRADAS e SAÍDAS. Essas anotações permitem com que a pessoa comece a ter uma noção real dos valores que entram para o seu bolso e dos valores que saem dele. Pessoas que não se atentam para esse detalhe estão fadadas ao prejuízo, lembrando que uma situação financeira difícil não é iniciada da noite para o dia, sua trajetória para o abismo já começou há 18, 24 ou 36 meses antes de realmente revelar o colapso. O Movimento de Caixa tem a finalidade de registrar as entradas e saídas de dinheiro com um único objetivo, saber o saldo no final do período (dia, semana, mês ou até ano). Essa ferramenta é o passo inicial para uma boa administração financeira pessoal, porque através das simples anotações de cada centavo que entra e que sai, é que o empreendedor rural poderá chegar ao final do mês e fazer a devida avaliação se sobrou ou não dinheiro na conta para futuras ações.

Uma outra ferramenta que pode auxiliar o empreendedor é o “Contas a Pagar”, essa ferramenta é importantíssima para que possa ter em mãos o controle das contas a serem pagas no dia, no dia seguinte, na semana seguinte, no mês seguinte e até nos próximos 12 meses que ainda estão por vir. Essa ferramenta permite ao empreendedor rural, a oportunidade de visualizar as obrigações do seu negócio e da sua vida pessoal em determinados períodos e para que isso aconteça é necessário que suas anotações estejam em plena sintonia, como os seus gastos, que

de fato é o principal fornecedor de informações para a sua gestão de contas a pagar. Assim como a ferramenta de Contas a Receber é importantíssima, existem executivos que determinam que ela é mais importante que a ferramenta anterior pelo simples argumento: “Se não tiver o que receber, não tem como pagar”.

A terceira ferramenta fundamental para a boa administração financeira do empreendedor rural é a de “Contas a Receber”, porque nela será possível identificar pontos importantíssimos para a gestão do dinheiro que está nas mãos dos clientes, fato a ser trabalhado e analisado com muito cuidado, para não prejudicar a sua saúde financeira do negócio. O modelo de contas a receber, é a ferramenta mais particular (personalizada) entre as já citadas no trabalho. Ela não segue um padrão pré-determinado, isso porque as anotações de contas a receber é de acordo com a política de recebimento e pagamento de pessoa.

A quarta, última e mais importante ferramenta de gestão financeira para o empreendedor rural é o fluxo de caixa, porque é a menos operacional ou a junção de todas as outras ou simplesmente, o “Fluxo de Caixa” de todo o dinheiro que circula no negócio. Muitos empreendedores confundem, até com uma certa razão o que vem a ser a diferença do fluxo de caixa com o movimento de caixa, por fim, acabam chamando um pelo outro e consideram

sendo a mesma coisa, mas não é. Pois então, existem grandes diferenças entre a primeira ferramenta e a quarta ferramenta. É importantíssimo diferenciar o movimento de caixa de fluxo de caixa. Para simplificar, o movimento de caixa são registros de entradas e saídas do dinheiro no tempo presente e o fluxo de caixa é o registro das possíveis entradas e possíveis saídas de dinheiro no tempo futuro, se transformando em uma ferramenta de planejamento estratégico.

Com base no que foi citado sobre as ferramentas financeiras básicas para a boa administração do empreendedor rural, é fundamental que as suas decisões de compra, venda, preço, desconto, ampliação e outros, sejam tomadas com base em números através do uso de ferramentas financeiras básicas, elevando assim, o seu negócio a outros níveis empresariais.

Objetivos

O trabalho tem como objetivo levantar a gestão financeira dos empreendedores rurais nas propriedades em que existem produção e comercialização de produtos, bem como, identificar quantas propriedades possuem gestores capacitados para tal administração e gestão financeira do empreendimento,

Metodologia

O presente trabalho foi uma pesquisa classificada como descritiva, objetivando identificar as ferramentas financeiras

utilizadas pelos gestores de propriedade rurais, baseado na proposta de Zanella (2014).

Para execução da pesquisa, foi criado um questionário e aplicado focado nas ferramentas de gestão financeira das pequenas propriedades rurais. O questionário possui 20 questões, obtendo-se informações sobre “o nome da propriedade”, “telefone”, “e-mail”, “área total da propriedade”, “principal atividade da propriedade”, “informações sobre os gerentes ou responsáveis da propriedade”, questões quanto as “ferramentas de gestão financeiras empregadas na fazenda” e a “forma que são trabalhadas” essas ferramentas.

O questionário foi aplicado aleatoriamente chegando aos gestores de pequenas propriedades rurais de duas formas: com entrevista presencial e com visitas nas fazendas, onde coletou-se as respostas dos responsáveis presentes. Essas visitas foram sem aviso prévio, de forma que o entrevistado não pudesse manipular as respostas.

Outro método empregado foi através de um questionaria on-line, criado através de um aplicativo do endereço de e-mail do gmail.com chamado “google formulário” com acesso ao link em: <https://goo.gl/forms/GxU4gG47nzq2UzGh2>, na qual foi distribuído através dos próprios empreendedores rurais e em grupos do aplicativo WhatsApp, alcançando 350 gestores.

As perguntas estabeleceram um perfil de gerenciamento financeiro das pequenas propriedades rurais.

O questionário foi dividido em 4 sessões. Na primeira sessão foram estabelecidas questões sobre a propriedade e o responsável que gerencia a mesma; a segunda sessão foi estabelecido um perfil para o ambiente de trabalho dentro da fazenda e o número de funcionários existentes; a terceira sessão foi estabelecido um perfil das principais atividades da propriedade e na quarta e última sessão do questionário, foi possível estabelecer o perfil da administração da propriedade as questões abordaram de como é alimentado as ferramentas de gestão financeira utilizadas na propriedade rural.

Os resultados apresentados foram tabulados e analisados, não passando por nenhum programa específico de análise de dados estatísticos e representação gráfica.

Resultados alcançados

A aplicação do questionário mostrou que 40% das propriedades possuem um gestor financeiro, sendo 40% desses gestores com ensino superior completo e que 80% das pequenas propriedades rurais possuem de 1 a 3 funcionários.

As propriedades não possuem apenas uma atividade, mas a principal que gera

faturamento é voltada para o gado com 60%, seguido pelo plantio de soja com 15% e de milho com 15%, os outros 10% estão distribuídos entre o sorgo, cana-de-açúcar e hortaliça, porém, nenhuma pequena propriedade rural tem como sua atividade principal o suíno, aves, equinos e apicultura.

Quanto a ferramenta financeira “Movimento de Caixa”, 70% fazem uso e registro das movimentações de dinheiro diariamente e 30% não a utilizam; as ferramentas “Contas a Pagar” e “Contas a Receber” são utilizadas por 50% das pequenas propriedades, sendo que 60% registram em caderno ou ata, 25% por meio eletrônico e 15% de outras maneiras. A ferramenta que as propriedades rurais mais utilizam é o “Controle de Custo”, sendo validada por 85% dos entrevistados.

Considerações finais

O questionário aplicado mostra que as pequenas propriedades rurais do município de Pires do Rio/GO negligenciam a importância da gestão financeira em suas próprias atividades diárias, impedindo o crescimento da produção e geração de receita para a propriedade e consequentemente, a oferta de emprego e renda para os moradores do município, fortalecendo as grandes propriedades rurais, denominadas como empresas do setor rural de grande porte.

Os dados afirmam a carência de conhecimento técnico e noção financeira dos gestores e dos colaboradores das propriedades rurais, mostrando que existe uma demanda grande para qualificação profissional dos colaboradores envolvidos com as pequenas propriedades rurais.

Referências

ARAÚJO, L. A. Planejamento de propriedades rurais: livro didático. design instrucional Marina Melhado Gomes da Silva. – Palhoça. Unisul Virtual, 2013.

FRÜHAUF, A. R. Gestão financeira e produtividade do empreendimento rural: uma análise da propriedade Frühauf. Trabalho de Graduação. Centro Universitário Univates. Lajeado, 2014.

GRÄF, L. V. Gestão de propriedades rural: um estudo a autonomia do jovem na gestão da propriedade rural. Trabalho de Graduação, Centro Universitário Univates. Lajeado, 2016.

Ministério da Agricultura e Abastecimento. Agropecuária puxa o PIB de 2017. Disponível em: <<http://www.agricultura.gov.br/noticias/agropecuaria-puxa-o-pib-de-2017>> . Acesso em: 09 out. 2017.

ROSA, W. F.; GARCIA, P. A importância da gestão financeira para pequenos produtores rurais de Pântano Grande. 2015. Disponível em: <<http://www.domalberto.edu.br/wp-content/uploads/2015/12/A-IMPORTANCIA-DA-GESTÃO-FINANCEIRA-PARA-PEQUENOS.pdf>>. acesso em 22 out. 2018.

ZANELLA, F. S; BARICHELLO, R.; Gestão financeira na agricultura familiar: um estudo de casos nas micros e pequenas empresas ligadas a cooperativa alternativa de Chapecó-SC. Dissertação de Especialização. Universidade Comunitária Regional de Chapecó, Chapéco, 2014.

Análise de dados da saúde com foco na síndrome gripal e síndrome respiratória aguda

André Almeida Luz (Instituto Federal do Tocantins, almeidaandreluz@gmail.com, <https://orcid.org/0000-0001-6539-6235>)

Ivo Sócrates Moraes de Oliveira (Instituto Federal do Tocantins e Universidade do Vale do Rio dos Sinos, ivo@ifto.edu.br, ivosocrates@edu.unisinos.br, <https://orcid.org/0000-0001-6539-6235>)

Jeferson Morais da Costa (Universidade Estadual do Tocantins, jeferson.mc@unitins.br, <https://orcid.org/0000-0001-7605-3174>)

Ary Henrique Morais de Oliveira (Universidade Federal do Tocantins, aryhenrique@uft.edu.br, <https://orcid.org/0000-0001-5484-870X>)

Glenda Michele Botelho (Universidade Federal do Tocantins, glendabotelho@uft.edu.br, <https://orcid.org/0000-0003-1334-0939>)

André Luis Korzenowski (Universidade do Vale do Rio dos Sinos, akorzenowski@unisinos.br, <https://orcid.org/0000-0002-7640-2845>)

Resumo: A tecnologia da informação pode representar uma aliada à saúde humana, além das diversas outras utilidades é possível implementar ferramentas robustas de análise de dados para possibilitar decisões assertivas e em tempo real. Logo, o objetivo desta pesquisa é propor uma solução de Business Intelligence para análise de dados da saúde do Estado do Tocantins, com ênfase em dados públicos acerca de Síndrome Gripal e Síndrome Respiratória Aguda Grave. Tal proposta vai de encontro as necessidades dos gestores e médicos das organizações públicas e privadas da área da saúde, que precisam tomar decisões para expandir o atendimento, adquirir e prescrever remédios e vacinas, realizar campanhas de orientação, entre outras. Para realização da proposta dentre as várias soluções de *Business Intelligence*, a que melhor se adequou foi o Microsoft Power BI, sendo aplicadas ações de ETL e esquemas de visualização voltados as características das percepções humanas. Como resultados foram apresentadas in-

formações da estruturação dos painéis e os insights com informações da Síndrome Gripal e Síndrome Respiratória Aguda Grave no Tocantins entre 2019 e 2020, como: número de mortes, principais sintomas dos pacientes, taxa de pacientes vacinados; informações que nem sempre estão acessíveis aos gestores e médicos. Como trabalho futuro, pode-se incluir a aplicação de técnicas e ferramentas da Inteligência Artificial para predizer e prescrever possíveis acontecimentos, antecipando eventuais ações, uma vez que tal análise fornece base para isto.

Palavras-chave: *Business Intelligence*, Análise de Dados, Painel, Monitoramento, Saúde

1. Introdução

A saúde é um dos assuntos mais relevantes da atualidade e que precisa ser tratado com extrema atenção. A tecnologia da informação pode representar uma aliada à saúde humana, fornecendo ferramentas

robustas, garantindo uma melhor performance dos médicos e trabalhadores da área da saúde, proporcionando rapidez nas consultas e cirurgias, proporcionando maior assertividade nos diagnósticos e tratamento, prolongando a vida dos pacientes. Os avanços da tecnologia permitiram a criação de diversas ferramentas, inclusive as ferramentas de análise de dados.

Sabe-se que dados podem gerar informações valiosas, e garantir tomadas de decisões assertivas quando utilizadas com ferramentas capazes de processar grandes volumes de dados. Visto isso, as ferramentas de *Business Intelligence* (BI - Inteligência de Negócios) voltado à área da saúde possibilitam a coleta e análise dos dados possibilitando gerar relatórios e painéis com análises baseadas em parâmetros observados a partir da extração e cruzamentos dos dados. Assim, informações e dados estatísticos possibilitam um entendimento mais preciso sobre o cenário encontrado, permitindo tomadas de decisões estratégicas a respeito.

Esta pesquisa sustenta-se na coleta e análise de dados relacionados a Síndrome Gripal (SG) e na Síndrome Respiratória Aguda Grave (SRAG) gerando informações importantes sobre o tema e possibilitando aos gestores da área da saúde uma visão panorâmica do cenário referentes ao contexto estabelecido, isto com base em dados públicos oferecidos

pelo Ministério da Saúde do Governo Federal Brasileiro.

O Ministério da Saúde (2021) define Síndrome Gripal em indivíduos com quadro respiratório agudo, que apresentam sintomas como febre, calafrios, dor de cabeça, dor de garganta, tosse, coriza, distúrbios olfativos ou distúrbios gustativos. Já a Síndrome Respiratória Aguda Grave é definida em casos de pacientes que apresentam dispneia/desconforto respiratório ou pressão persistente no Tórax ou saturação de O₂ menor que 95% em ar ambiente ou colorização azulada dos lábios ou rosto. Síndromes essas que estão ligadas diretamente ao COVID-19, que possuem sintomas semelhantes, como febre, tosse seca, dor de garganta e cansaço.

A síndrome gripal e a síndrome respiratória aguda foram responsáveis por várias mortes ao longo dos anos, e no cenário atual torna-se importante o desenvolvimento e aplicação de estudos utilizando-se de técnicas, com objetivo de agregar valor estratégico para a geração de conhecimentos sobre as situações causadoras de tais síndromes, buscando proporcionar avanços nos tratamentos e prevenções.

A partir do contexto apresentado esta pesquisa foi alicerçada no seguinte questionamento: É possível por meio da análise

de dados e ferramentas de *Business Intelligence* disponibilizar informações para decisões estratégicas, a fim de auxiliar profissionais da saúde no conhecimento sobre as doenças abordadas?

Os dados aliados as ferramentas analíticas possibilitam aos gestores tomadas de decisões estratégicas por meio de relatórios, gráficos, painéis (*dashboards*), tabelas de uma forma simples e dinâmica, e na área da saúde essa técnica pode agregar aos médicos e trabalhadores da área uma visão panorâmica e mais compreensível das informações coletadas através de dados relacionados as doenças em questão.

Para isso existem as ferramentas de *Business Intelligence*, que além de permitir todo o processo de análise dos dados, ainda garante informações importantes que ajudam a entender melhor sobre os dados. Para Negash e Gray (2008) e Davenport (2006), *Business Intelligence* consiste em um sistema de coleta e armazenamento de dados gerenciados por ferramentas analíticas cujo objetivo é transformar dado em conhecimento de modo a disponibilizar informações complexas para dar suporte ao planejamento e à tomada de decisões estratégicas nas organizações.

Logo, *Business Intelligence* pode ser compreendido como um conjunto de con-

ceitos e metodologias voltadas ao apoio e tomada de decisão, em que é possível transformar dado em informação e da informação em conhecimento (CÔRTEZ, 2002). Até chegar ao estado de informação os dados passam por processos conhecidos como ETL (Extração, Transformação e Carga de Dados). “O ETL é um processo para extrair dados de um sistema de Bases de Dados (BD), sendo esses dados processados, modificados, e posteriormente inseridos numa outra BD” (FERREIRA et al., 2010).

Para quem deseja realizar uma amostragem de dados, deve levar em consideração a criação de painéis (*dashboards*) e gráficos para o processo de visualização. Painéis quando bem estruturados fornecem um melhor e rápido entendimento na visualização dos dados, possibilitando uma avaliação tanto geral, como específica de determinado contexto (FEW, 2005).

Com o aumento de casos de pessoas com síndrome gripal e síndrome respiratória aguda, e levando em consideração o cenário atual envolvendo a pandemia causada pelo novo coronavírus, identificou-se que seria viável realizar pesquisas com o intuito de demonstrar a eficácia do uso de ferramentas de BI aplicada à saúde e sua utilidade à gestores e médicos para identificar possíveis informações valiosas que podem estar ocultas no meio de tantos dados desorganizados, e até

mesmo contribuir para melhores tomadas de decisões estratégicas de médicos e afins.

2. Objetivos

Propor uma solução de Business Intelligence para análise de dados da saúde do Estado do Tocantins, com ênfase em dados públicos acerca de Síndrome Gripal e Síndrome Respiratória Aguda.

3. Metodologia

Nesta pesquisa buscou-se utilizar os conceitos de pesquisa aplicada, com suporte da utilização de pesquisa bibliográfica e pesquisa exploratória. Em primeiro momento utilizou-se como base a pesquisa bibliográfica, que conforme Macedo (1995) constitui em buscar fundamentações teóricas já existentes relacionado ao tema proposto, que abrangeu a coleta de conteúdos relacionados ao tema deste trabalho, que foram pesquisados em livros, monografias, teses, dissertações, artigos na plataforma Scielo e através de pesquisas no Google Acadêmico relacionados aos conceitos de BI (Inteligência de Negócios), Saúde e Visualização de Dados. Em seguida iniciou-se a coleta de dados quantitativos e qualitativos, que foram coletados a partir das bases nacionais do Ministério da Saúde e da Secretaria da Saúde do Tocantins, este processo se deu por meio da aplicação da pesquisa exploratória, que segundo Gil (2002), consiste em conhecer sobre o problema apresentado, realizar estudos e análises,

e em cima disso construir hipóteses ou intuições. Com a aplicação da pesquisa exploratória foi possível encontrar dados disponibilizados em formatos de tabelas e gráficos, o que possibilitou a manipulação para utilização nas ferramentas de *Business Intelligence*. Para Piovesan e Temporini (1995, p.321) a pesquisa exploratória “(...) Tem por objetivo conhecer a variável de estudo tal como se apresenta, seu significado e o contexto onde ela se insere”. A partir disso foi coletado dados qualitativos e quantitativos relacionadas a síndrome gripal e a síndrome respiratória aguda, como proposta de realizar análises e identificar informações de valor para ajuda na tomada de decisões utilizando ferramenta de *Business Intelligence*. Com os dados coletados foi feita a análise de ferramenta para a análise dos dados, para este trabalho optou-se por utilizar a versão gratuita da ferramenta Microsoft Power BI, criada em 2014 pela empresa Microsoft, voltada à inteligência de negócios e proporciona ao usuário realizar o tratamento de dados com uma visão simples e intuitiva dos resultados encontrados. Foram encontradas diversas ferramentas para análise de dados, uma delas foi a IBM Watson Analytics, uma das grandes ferramentas e das mais conceituadas no mercado de BI. Trata-se de uma ferramenta gratuita, também contendo a versão paga, voltada para a análise de dados em massa. O Google também possui uma ferramenta para análise de dados chamada de Google Data Studio, basta ter uma conta no Google para começar a utilizá-

-la. O Data Studio é uma ferramenta mais simples, gratuita, leve e com uma curva de aprendizado relativamente pequena, porém possuem algumas limitações como a demora para carregar painéis complexos e também o fato da ferramenta está disponível somente na versão Web. A escolha pelo Microsoft Power BI dentre tantas outras ferramentas de *Business Intelligence* foi pelo fato ser uma ferramenta que pode ser usada gratuitamente, e fornece todos os recursos necessários para trabalhar com os dados, rapidez ao mostrar os painéis, possibilita atualizações dinâmicas, além de ser bem-conceituada e de simples manipulação.

4. Resultados

Os dados utilizados nesta pesquisa foram coletados no site do *OpenDataSUS*, departamento de informática do Sistema Único de Saúde, onde são disponibilizadas diversas bases de dados do Ministério da Saúde e nos diferentes tipos e formatos. Criado em 1991, o *DataSUS* fornece dados da saúde que ao serem analisados podem ajudar no suporte à tomada de decisões de gestores da saúde e servir para pesquisadores e estudantes a encontrar possíveis informações relevantes que agreguem algum valor posteriormente (DATASUS, 2020).

Para esta pesquisa foram utilizadas a base de dados da Síndrome Respiratória Aguda Grave (SRAG) dos anos de

2019 e 2020 respectivamente, e a base de dados da Síndrome Gripal do ano de 2019. As bases de dados da SRAG estão disponibilizadas em formato CSV, e vem acompanhada com um dicionário de dados, que consiste em um documento cujo objetivo é descrever as variáveis da base de dados exportada em DBF, e uma ficha de notificação, que é um documento onde são preenchidos dados dos pacientes que são hospitalizados com caso da Síndrome Respiratória Aguda Grave, ambos em formato PDF.

A base de dados da Síndrome Gripal (SG) é proveniente do sistema e-SUS NOTIFICA, um sistema que foi criado para o cadastramento de casos de paciente com suspeita de COVID-19. Os dados da SG estão disponibilizados em formato CSV e contam com dados de todos os estados brasileiros separadamente, mas para este trabalho focou-se somente no Estado do Tocantins.

Assim como nas bases de dados da SRAG, a base de dados da SG vem acompanhada de um dicionário de dados e uma ficha de notificação, ambos em formato PDF, e uma API dos dados. O critério para a escolha das colunas que foram analisadas pela ferramenta baseou-se na quantidade de dados e na relevância, as colunas não utilizadas foram classificadas como “inativas”, sendo possível usá-las de acordo com a necessidade do gestor.

Para o cruzamento dos dados, foram criadas diversas “Medidas”, uma técnica presente no Microsoft Power BI, em que é possível criar cálculos lógicos e aritméticos de acordo com a coluna desejada. Foram feitos cálculos para realizar o somatório de óbitos, curados, taxa de pacientes vacinados contra gripe, taxa de pacientes que utilizaram algum tipo de antiviral, taxa de pacientes internados em UTIs, quantidade de pacientes que possuem fatores de risco, casos nosocomiais, casos provenientes da Síndrome Gripal, classificação final do caso, quantidade de pacientes que tiveram contato direto com aves ou suínos, evolução do caso e contagem de pacientes de acordo com o seu sexo. Para melhor descrição dos resultados foram criadas 3 seções: Análise descritiva das bases de dados, Processo de ETL e Painéis e visualizações, que são apresentadas a seguir.

Análise descritiva das bases de dados

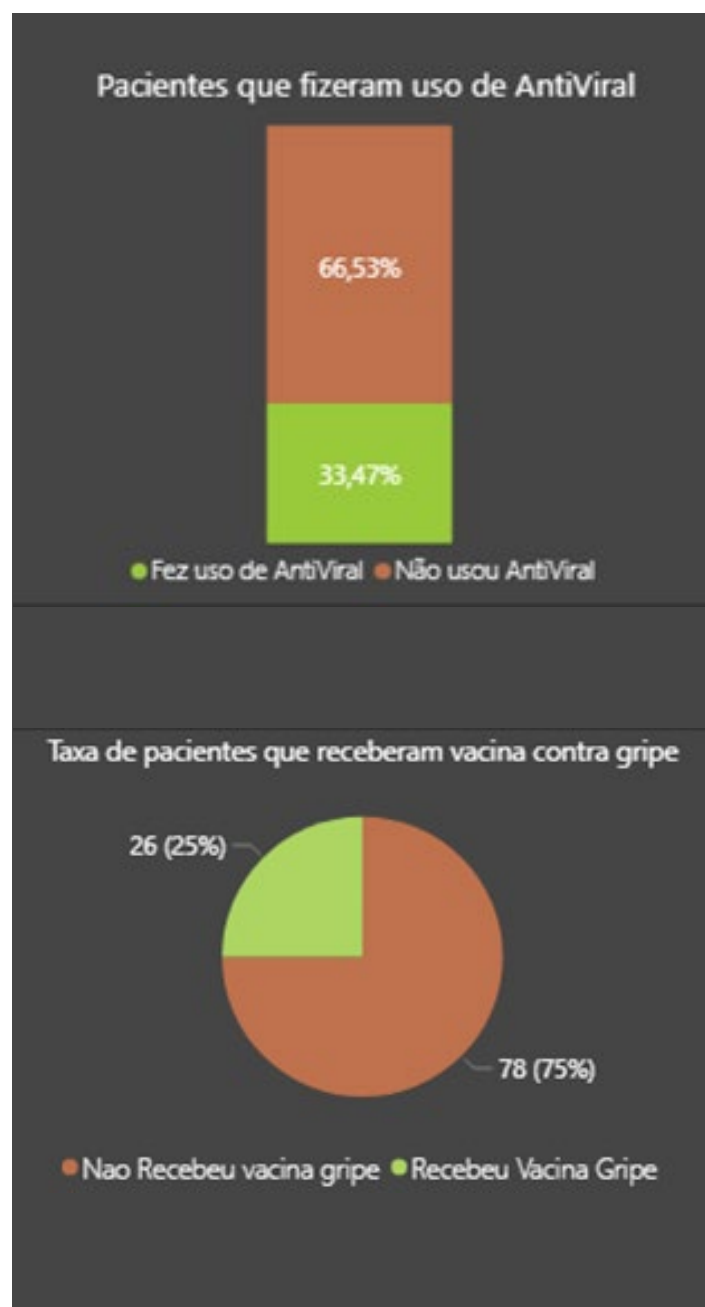
Com base na análise dos dados dos pacientes que tiveram SRAG, foi observado que no ano de 2019 houve um maior número de pacientes curados em relação à óbitos, pacientes do sexo masculino obtiveram um maior índice de mortes em relação ao sexo feminino. O mês de abril foi o mês em que foi notabilizado o maior índice de pacientes curados, por data de internação, já o mês de março se notabilizou como o mês que tiveram mais óbitos. A faixa etária do paciente é um fator a se levar em consideração, pois quanto mais avançada a idade do paciente, maior o ris-

co de morte por complicações da SRAG.

Nos Gráficos da Figura 1 (a) e (b) são apresentados os resultados dos pacientes hospitalizados no Tocantins por SRAG no ano de 2019 que fizeram e não fizeram uso de antivirais e vacinas para gripe.

Figura 1: Percentual de pacientes hospitalizados no Tocantins por SRAG em 2019 (a) que fizeram e não fizeram uso de antiviral e (b) que tomaram e não tomaram a vacina contra gripe

Fonte: Elaborada pelo autor (2022)



Assim como apresentado na Figura 1 (a), cerca de 66,5% dos pacientes não fizeram uso de nenhum tipo de antiviral, e a Figura 1 (b) mostra que 75% não tomaram vacina para gripe, conseqüentemente, ligados a este fato, o número de óbitos foi maior em pacientes que não utilizaram nem antivirais e nem vacina para gripe.

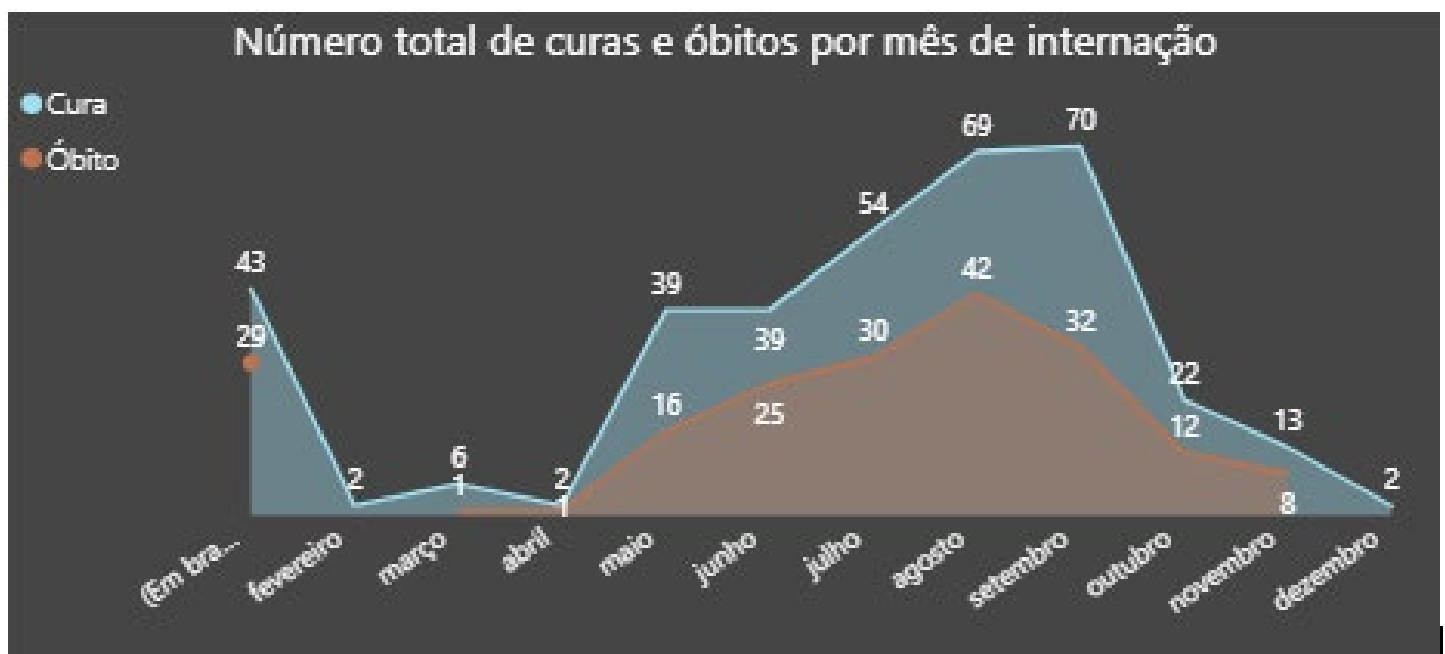
Dentre as cidades do Tocantins mais afetadas, Palmas-TO foi a que teve mais óbitos e o maior número de curados. Quanto a classificação do caso, no Tocantins o maior número de casos foram de SRAG não especificado com 63% dos casos, seguido por SRAG por influenza, casos de SRAG por outro vírus respiratório ficou com 14,94% dos casos, casos de SRAG por COVID-19 não foram descobertos.

Em relação aos fatores de risco que pacientes poderiam ter, foram selecionados asma, diabetes e doença cardiovascular crônica para a análise. Através da análise é possível perceber que 7% dos pacientes

que foram internados por SRAG possuem asma, 17% possuem diabetes e 38% possuem doença cardiovascular. O número de pacientes que precisaram ser internados em UTI (Unidade de Terapia Intensiva) foi de 37%, porém pacientes cuja faixa etária entre 0 e 14 anos foram internados cerca de 51%. Também foi notar que cerca de 99% dos casos não foram provenientes da Síndrome Gripal. Apenas 6% dos casos da SRAG trata-se de infecção adquirida no hospital e somente 3,21% dos pacientes tiveram contato com aves ou suínos.

Quanto aos casos de SRAG no ano de 2020 no Tocantins houve um total de 361 curados e 196 mortos, sendo grande parte dos óbitos do sexo masculino e os casos de SRAG 62% foram ocasionados pelo vírus da COVID-19. O gráfico apresentado na Figura 2 demonstra o número de óbitos e curas de acordo com o mês de internação dos pacientes com SRAG no ano de 2020 no Tocantins.

Figura 2: Número de óbitos e curas por mês de internação dos pacientes com SRAG no ano de 2020 no Tocantins / Fonte: Elaborada pelo autor (2022)



Conforme a Figura 2, o mês de agosto foi o mês com mais mortes em decorrência da SRAG e o mês de setembro foi o mês com mais pacientes curados. A cidade de Araguaína-TO foi a que possuiu o maior número de mortes decorrente da SRAG em 2020, seguida por Palmas-TO. No geral, somente 18,12% dos pacientes receberam vacina para gripe e 4,38% utilizaram algum tipo de antiviral.

Dos pacientes internados por SRAG em 2020 no Tocantins que possuem algum tipo de fator de risco, 7,2% informaram ter asma, 44,4% diabetes e 52,7% doença cardiovascular crônica. Foi possível notar um grande número de pacientes que precisaram ser internados em UTIs por complicações da SRAG em relação ao ano de 2019. Dos 26,61% dos casos de SRAG internados em UTI, 65,12% evoluíram para óbito e somente 34,88% foram curados. Já sobre casos de pacientes que tiveram SRAG provenientes da Síndrome Gripal, 64,35% não foram ocasionados por SG e 35,65% de casos provenientes da SG. Assim como no ano de 2019, casos nosocomial tiveram índices baixos, e somente 2,3% dos pacientes que tiveram algum tipo de contato com aves e suínos antes de serem diagnosticados com SRAG.

Os resultados referentes a análise dos casos de Síndrome Gripal (SG) em 2019 no Tocantins são mostrados através dos gráficos apresentados nas Figuras 3 e 4.

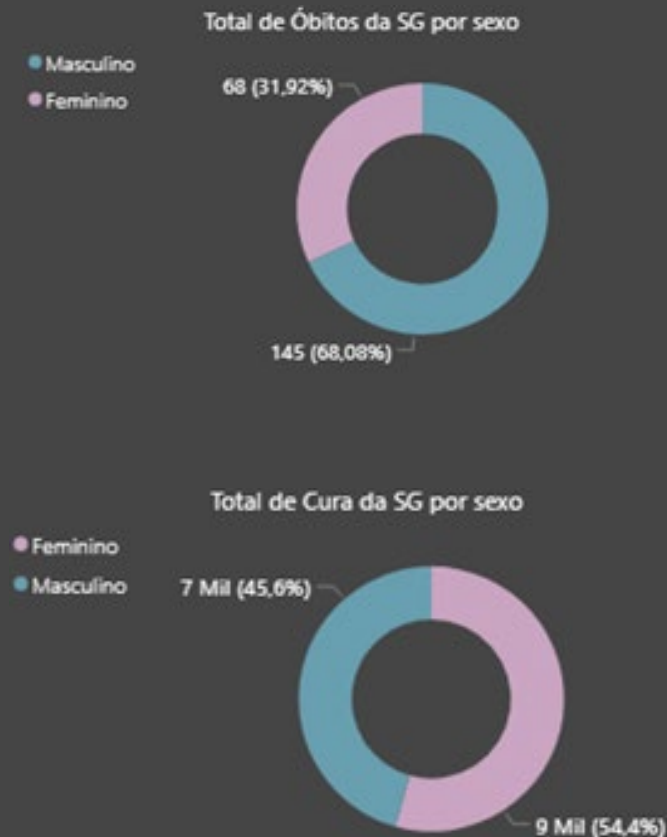


Figura 3: Total de óbitos e curados da SG no ano de 2019 no Tocantins

Fonte: Elaborada pelo autor (2022)

Figura 4: Gráficos com (a) quantidade e percentual de óbitos e (b) quantidade e percentual de curados da SG no ano de 2019 no Tocantins

Fonte: Elaborada pelo autor (2022)



O gráfico da Figura 3 apresenta um total de 213 mortos e cerca de 16 mil curados por SG em 2019 no Tocantins, o percentual de mortes representam 1,33% do total. Já o gráfico da Figura 4 (a) apresenta que cerca de 68% das mortes relacionadas SG em 2019 no Tocantins foram de pacientes do sexo masculino, já os curados, Figura 4 (b), houve quase que uma neutralidade nos resultados por sexo, sendo 54,4% dos curados da SG foram do sexo feminino e 45,6% do sexo masculino.

Os gráficos das Figuras 5 e 6 apresentam os principais sintomas da Síndrome Gripal sentidos pelos pacientes de acordo com o sexo.

Figura 5: Total de óbitos e curados da SG no ano de 2019 no Tocantins

Fonte: Elaborada pelo autor (2022)

8 principais sintomas sentidos pelo sexo masculino

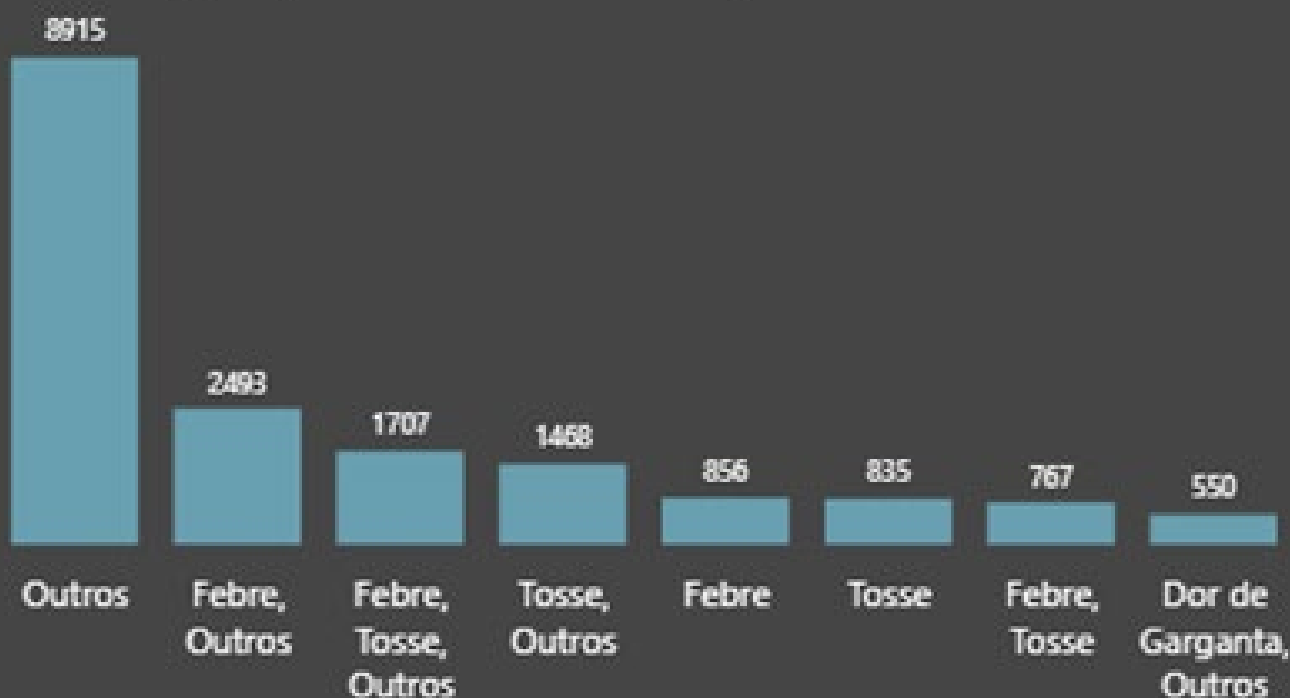


Figura 6: Quantidade e Percentual de óbitos por SG no ano de 2019 no Tocantins

Fonte: Elaborada pelo autor (2022)

8 principais sintomas sentidos pelo sexo feminino



De acordo com o gráfico da Figura 5 é possível notar que: febre, tosse, e outros sintomas não especificados estão entre os primeiros sentidos pelo sexo masculino, já de acordo com a Figura 6, em pacientes do sexo feminino é mostrada uma alteração: sintomas como dor de garganta aparece em 5º lugar, diferente do sexo masculino, em que dor de garganta aparece em 8º lugar.

A Figura 7 apresenta uma lista contendo a quantidade de óbitos por SG em 2019 no Estado do Tocantins de acordo com o município.

Figura 7: Quantidade de óbitos por SG em 2019 no Tocantins de acordo com o município
Fonte: Elaborada pelo autor (2022)

Óbitos por cidade	
município	Óbitos
Araguaína	70
Palmas	23
Tocantinópolis	10
Paraíso do Tocantins	8
Gurupi	7
Porto Nacional	7
São Miguel do Tocantins	7
Xambioá	7
Aguiarnópolis	6
Formoso do Araguaia	5
Total	213

Já nas Figuras 8 e 9 apresentam uma lista contendo a quantidade de óbitos por SG no Tocantins com base em gráficos de área expondo números tanto de óbitos quanto de curados de acordo com mês no ano de 2019, no primeiro semestre.

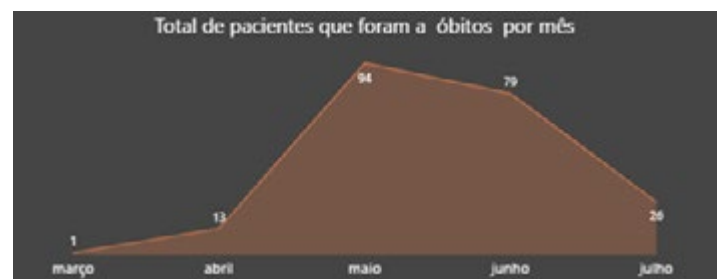
Figura 8: Número de pacientes curados da SG em 2019 no Tocantins por mês

Fonte: Elaborada pelo autor (2022)



Figura 9: Número de óbitos pela SG em 2019 no Tocantins por mês

Fonte: Elaborada pelo autor (2022)



Como é possível notar na Figura 7, o município de Araguaína-TO foi identificado como a cidade com mais óbitos em decorrência da SG em 2019 no Tocantins, seguida de Palmas-TO e Tocantinópolis-TO. Na Figura 8 é possível observar que o mês de junho foi o mês com mais pacientes curados da SG em 2019 no Tocantins e de acordo com a Figura 9, maio foi o mês com mais mortes no Tocantins.

O propósito de trazer uma amostragem de dados em uma ferramenta de BI é justa-

mente por ela dar a liberdade do usuário a filtrar os resultados da maneira que ele necessita naquele momento, por exemplo, caso o gestor deseje descobrir qual a faixa etária de pacientes internados por SRAG que mais tomaram antiviral no ano de 2020: é necessário somente ir ao filtro de idade, que já foi criado a partir da coluna de idade, e selecionar a faixa etária até encontrar o resultado.

O Microsoft Power BI possibilita a criação desses painéis (dashboards) de filtros, gráficos de pizza, dentre tantos outros modelos que se encaixe no contexto desejado. Ela conta com um editor chamado Power Query, que possibilita a realização do que é chamado de ETL, onde é realizado todo um processo de “preparação dos dados” para, assim, poder analisá-los em painéis. Um fator que vale a pena ressaltar a respeito da ferramenta é a capacidade de atualização dinâmica dos dados que estão sendo tratados, ou seja, a partir do momento em que esta base de dados que foi extraída é alimentada com mais dados, conseqüentemente esses dados são automaticamente importados pela ferramenta, atualizados dinamicamente o painel, possibilitando que as informações sejam visualizadas em tempo real.

Processo de ETL (Extração, Transformação, Carregamento)

Assim como apresentado anteriormente ETL é o processo que os dados são extraídos das bases de dados, processados, e logo após são inseridos em uma nova

base de dados. Vale ressaltar a importância de realizar a transformação dos dados, pois é nesta etapa em que os dados passam pelo processo de limpeza, onde são identificados possíveis erros, como dados nulos, incorretos ou de formato diferente do especificado, para garantir uma integridade e confiabilidade dos dados antes de serem utilizados para produzir informações.

Nesta pesquisa foi-se aplicado o processo de ETL nas bases de dados utilizadas, com o objetivo de filtrar e utilizar somente campos essenciais para gerar as informações relevantes ao visualizador. Inicialmente, foram identificados erros em algumas linhas, colunas com os campos completamente nulos, campos vazios que não foram preenchidos anteriormente na ficha de notificação antes de serem cadastrados na base de dados e colunas com dados que foram julgados como desnecessárias foram removidas. Colunas que obtiveram uma taxa abaixo de 20% dos dados válidos foram descartadas. Nomes de algumas colunas foram renomeadas para melhor identificação do campo, para esta atividade foi necessário fazer uso do dicionário de dados acompanhado da ficha de notificação. O tipo de dado de cada coluna não necessitou aplicar quaisquer mudanças, pois todos foram obtidos no formato adequado.

A base de dados da SG ao ser baixada possuía um total de 33 colunas, com as exclusões das colunas “yid”, “cbo”, “PaisO-

rigem”, “estadoIBGE”, “municipioIBGE”, “cep”, “origem”, “cnes”, “EstadoNotificação”, “EstadoNotificaçãoIBGE”, “municipioNotificação”, “municipioNotificaçãoIBGE”, “numeroNotificacao”, “excluído” e “validado”, esse valor abaixou para 18 colunas, o nome da tabela que antes era “dados-to” foi alterada para SG-to-2019. Já a base de dados da SRAG 2019 possuía um total de 140 colunas, e foi diminuído para o total de 66 colunas, o nome da tabela foi renomeada de “influd19_limpo-27.04.2020-final” para “SRAG-2019”. A base de dados da SRAG 2020 possuía um total de 140 colunas, e foi diminuída para 78 e renomeada “influd19_limpo-27.04.2020-final” para “SRAG-2020”.

Um dos motivos primordiais que ocasionou na exclusão de inúmeras colunas foi no intuito de diminuir o grande volume de dados que não seriam necessários para demonstrar os resultados, ou seja, dados que não seriam utilizados. Com essas quantidades de colunas reduzidas tornou-se possível trabalhar de forma centralizada com os dados essenciais nos insights desejados.

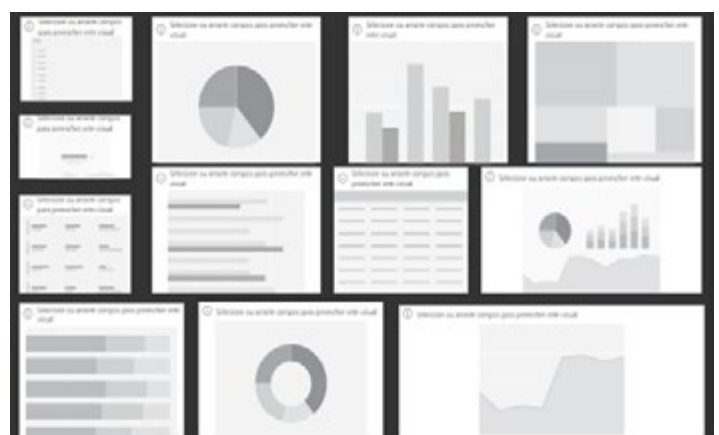
Painéis e visualizações

A maneira como os dados são mostrados em um painel é indispensável para se ter uma boa análise de dados, logo, focou-se veemente na melhor forma de transmitir as informações para possibilitar a compreensão de forma clara e objetiva sobre o que está sendo transmitido independentemente do grau de conhecimen-

to técnico do gestor. Inicialmente, para trabalhar com os painéis na ferramenta de BI, foi escolhido como tela de fundo a cor preta, pois a partir de testes realizados com os gráficos e por pesquisas relacionadas, conclui-se que cores radiantes como branco e azul tendem a forçar mais a visão, causando desconforto e cansaço para o visualizador. Portanto, foi importado o tema denominado de “Nowalls Analytics Theme” em formato JSON através da própria biblioteca do Microsoft Power BI. A paleta de cores deste tema conta com um fundo de tela preto e para a coloração dos gráficos possuem tons de branco, preto, cinza, verde, azul, vermelho e roxo. Os painéis que foram utilizados para a amostragem de dados foram: filtros, KPI, gráficos de colunas 100% empilhadas, gráfico de pizza, tabela, cartões de linha múltipla, gráfico de coluna clusterizado, gráfico de área, gráfico de rosca, TreeMap, Timeline 2.1.1, gráfico de barras clusterizado. A Figura 10 destaca, de forma genérica, os tipos de gráficos aplicados nos painéis para a amostragem dos dados.

Figura 10: Tipos de gráficos aplicados nos painéis

Fonte: Elaborada pelo autor (2022)



A escolha por cada um desses tipos de gráficos baseou-se em critérios como: facilidade na visualização do receptor, simplicidade e dinamismo. As Informações que trazem dados “negativos” como óbitos, pacientes que não tomaram vacina e outros semelhantes foram coloridos em tonalidades vermelhas, e informações positivas em tonalidades verdes. Informações que são comparativos entre sexo do paciente foram definidas como: tons de azul para masculino e tons de roxo para feminino, gráficos de filtros foram coloridos com tons de roxo escuro.

A Figura 11, no canto superior esquerdo mostram 2 dos 14 filtros presentes nos painéis criados, sendo 8 filtros por idade e 4 filtros por sexo. A Figura 11 também apresenta de forma ampla um dos painéis relacionados a Síndrome Respiratória Aguda Grave do ano de 2019 no Tocantins, em que há diversos gráficos, todos padronizados de acordo com as regras apresentadas anteriormente.

Figura 11: Uma das páginas dos painéis propostos

Fonte: Elaborada pelo autor (2022)



A padronização das cores nos painéis ajuda na organização dos painéis, tornando a visualização mais compreensível, ajudando o cérebro no processo de interpretação e abstração das informações apresentadas.

5. Considerações finais

Esta pesquisa teve como objetivo analisar os dados referentes a Síndrome Gripal e Síndrome Respiratória Aguda Grave nos anos de 2019 e 2020 no Estado do Tocantins, e possibilitar aos gestores da saúde e pesquisadores da área uma visão mais clara e abrangente dos casos e notificações dessas duas síndromes. A escolha de analisar esses dados pelos anos de 2019 e 2020 foi devido a pandemia do novo coronavírus, que demandou muitas informações compiladas para monitoramento dos pacientes entre o período pré-pandemia e o período pandêmico.

De acordo com os resultados obtidos através dos painéis foi demonstrado um elevado índice de óbitos em decorrência da que evoluíram para Síndrome Respiratória Aguda Grave (SRAG) no ano de 2020 com relação ao ano de 2019. Outro ponto a se destacar é que no ano de 2020 os casos de SRAG em decorrência do vírus da COVID-19 foram cerca de 62,5%, diferentemente do ano de 2019, onde os índices foram praticamente nulos. Casos de Síndrome Gripal (SG) que evoluíram para SRAG no ano de 2020 cresceram absurdamente em relação ao ano de 2019.

Deste modo, conclui-se que o elevado número de casos e óbitos decorrentes da SRAG no ano de 2020 no Estado do Tocantins tem como principal fator o início da pandemia ocasionada pela COVID-19.

Diante de todos os resultados encontrados e de todas as etapas realizadas, concluiu-se com êxito que dados ao serem trabalhados com uma ferramenta de Business Intelligence entregam informações de aspectos descritivos relevantes para gestores. Trazendo para o contexto do trabalho, os dados das SGs e SRAGs utilizando o Microsoft Power BI trouxeram informações, como: número de mortes, principais sintomas sentidos pelos pacientes, taxa de pacientes vacinados; informações essas que não são encontradas antes de passar pelo processo de tratamento de dados até chegar nos painéis, possibilitando gestores da saúde e pesquisadores da área resultados íntegros, assertivos e uma forma racional para a tomada de decisões. Como trabalho futuro, pode-se incluir a aplicação de técnicas e ferramentas da Inteligência Artificial para prever e prescrever possíveis acontecimentos, antecipando eventuais ações, uma vez que tal análise fornece base para isto.

Referências

CÔRTEZ, S. BI, Data Warehouse e Data Mining - Como a Tecnologia aumenta a Inteligência do Negócio. Casos, Soluções e Tendências em Inteligência de Negócios e Gestão do Conhecimento, SECESU, Rio de Janeiro, 2002.

DATASUS - DEPARTAMENTO DE INFORMÁTICA DO SUS. Ministério da Saúde Datasus. 2020. Disponível em: <<https://datasus.saude.gov.br/sobre-o-datasus/>>. Acesso em: 23/01/2021.

DAVENPORT, T. H. Competing on analytics. Harvard business review, v.84, n.1, p.98, 2006.

FERREIRA, J.; MIRANDA, M.; ABELHA, A.; MACHADO, J. O Processo ETL em Sistemas Data Warehouse. In: IN-Forum, 2010. p.757-765.

FEW, S. Dashboard design: Beyond meters, gauges, and traffic lights. Business Intelligence Journal, v. 10, n. 1, p. 18-24, 2005.

GIL, A. C. Como classificar as pesquisas. Como elaborar projetos de pesquisa, v.4, p.44-45, 2002.

MACEDO, N. D. de. Iniciação à pesquisa bibliográfica. Edições Loyola, 1995.

Ministério da Saúde. Saiba como é feita a definição de casos suspeitos de COVID-19 no Brasil. 2021. Disponível em: <<https://www.gov.br/saude/pt-br/coronavirus/artigos/definicao-e-casos-suspeitos>>. Acesso em: 23/07/2022.

NEGASH, S.; GRAY, P. Business Intelligence. In: Handbook on decision support systems 2. Springer, Berlin, Heidelberg, 2008. p.175-193.

PIOVESAN, A.; TEMPORINI, E. R. Pesquisa exploratória: procedimento metodológico para o estudo de fatores humanos no campo da saúde pública. Revista de saúde pública, v.29, p.318-325, 1995.

PREFEITURA DE
SÃO PAULO
INOVACAO
E TECNOLOGIA

IMPRESA

LOUNGE
OFFLINE

PALCO
TMT

BRASIL

 **campus
party**

realização **GOUVÊA**
experience



**REVISTA
CIENTÍFICA
CAMPUS PARTY**