

# Análise orientada a dados como auxílio para tomada de decisão em Gestão de Pesquisa<sup>1</sup>

*Data-driven analysis as an aid for decision-making in Research Management*

*Análisis orientado a datos como ayuda para la toma de decisiones en Gestión de Investigación*

*Adrielson Ferreira Justino, Fabiano Paulo Nhoatto, Éfren Lopes de Souza e Fábio Manoel França Lobato*

<https://doi.org/10.36428/revistadacgu.v14i27.565>

**Resumo:** Considerando que as Universidades têm como base os pilares de ensino, pesquisa e extensão, faz-se imperativo o uso de métodos para avaliação de sua produção técnico-científica. Por meio da cientometria, mensura-se a contribuição técnico-científica em determinadas áreas, identificando potencialidades e oportunidades institucionais. Com o estudo de caso à Universidade Federal do Oeste do Pará (UFOPA), observaram-se os seguintes problemas na gestão de suas pesquisas: i) dificuldade de coleta e organização dos dados de pesquisa da instituição; ii) falta integração com sistemas institucionais; e iii) pouca transparência das produções técnico-científicas. Tais fatos impactam na eficiência da gestão da pesquisa. Visando resolver os problemas supracitados, o presente trabalho compreende na análise de dados para auxiliar na tomada de decisão referente aos editais de pesquisa da instituição. Por meio das análises foi identificado que alguns institutos possuem produção técnico-científica relevante e competitiva dentro do cenário nacional, enquanto outros têm docentes com baixa produção, mas com acesso ao programa de bolsas de iniciação científica. Além disso, verificou-se que os principais tópicos dos projetos e planos de trabalho estão relacionados ao desenvolvimento regional sustentável, com o objetivo de compreender as particularidades da região e promover avanços econômicos e sociais. Os resultados obtidos possibilitaram uma maior transparência, inserção social da universidade e capacidade de tomada de decisão orientada a dados, fornecendo auxílio à gestão da pesquisa e inovação e avaliação de políticas tecnocientíficas. Espera-se que o produto possibilite uma maior eficiência na gestão de recursos públicos, fomentando parcerias inter e intra institucional e aumento do impacto das pesquisas. A abordagem adotada é escalável, podendo ser utilizada por outras Instituições de Ciência e Tecnologia.

**Palavras chave:** Ciência de Dados, Cientometria, Gestão de Pesquisa, Políticas de Pesquisa, Universidade, transparência

1. Artigo submetido em 18/07/2022 e aceito em 01/03/2023.

**Abstract:** Considering that Universities are based on the pillars of teaching, research, and extension, it is imperative to use methods to evaluate their technical-scientific production. Through scientometrics, the technical-scientific contribution in certain areas is measured, identifying institutional potentialities and opportunities. Through a case study at the Federal University of West Pará (UFOPA), the following problems were observed in the management of its research: i) difficulty in collecting and organizing the institution's research data; ii) lack of integration with institutional systems; and iii) little transparency in technical-scientific productions. Such facts impact the efficiency of research management. In order to solve the aforementioned problems, this study involves data analysis to assist in decision-making regarding the institution's research calls. Through the analysis, it was identified that some institutes have relevant and competitive technical-scientific production within the national scenario, while others have teachers with low production but access to the scientific initiation scholarship program. In addition, it was found that the main topics of projects and work plans are related to sustainable regional development, with the aim of understanding the region's particularities and promoting economic and social advances. The results obtained allowed greater transparency, social insertion of the university, and data-driven decision-making capacity, providing assistance to research and innovation management and evaluation of techno-scientific policies. It is hoped that the product will enable greater efficiency in the management of public resources, fostering inter and intra-institutional partnerships, and increasing the impact of research. The approach adopted is scalable and can be used by other Science and Technology Institutions.

**Keywords:** Data Science, Scientometrics, Research Management, Research Policies, University, transparency.

**Resumen:** Considerando que las Universidades tienen como base los pilares de enseñanza, investigación y extensión, es imperativo el uso de métodos para la evaluación de su producción técnico-científica. Mediante la ciencimetría, se mide la contribución técnico-científica en determinadas áreas, identificando potencialidades y oportunidades institucionales. Con el estudio de caso de la Universidad Federal del Oeste de Pará (UFOPA), se observaron los siguientes problemas en la gestión de sus investigaciones: i) dificultad en la recolección y organización de los datos de investigación de la institución; ii) falta de integración con sistemas institucionales; y iii) poca transparencia en las producciones técnico-científicas. Estos hechos afectan la eficiencia en la gestión de la investigación. Con el objetivo de resolver los problemas mencionados anteriormente, el presente trabajo comprende el análisis de datos para ayudar en la toma de decisiones referente a las convocatorias de investigación de la institución. Mediante los análisis se identificó que algunos institutos tienen una producción técnico-científica relevante y competitiva dentro del escenario nacional, mientras que otros tienen docentes con baja producción, pero con acceso al programa de becas de iniciación científica. Además, se verificó que los principales temas de los proyectos y planes de trabajo están relacionados con el desarrollo regional sostenible, con el objetivo de comprender las particularidades de la región y promover avances económicos y sociales. Los resultados obtenidos posibilitaron una mayor transparencia, inserción social de la universidad y capacidad de toma de decisiones orientada a datos, brindando ayuda a la gestión de la investigación e innovación y evaluación de políticas tecnocientíficas. Se espera que el producto posibilita una mayor eficiencia en la gestión de recursos públicos, fomentando asociaciones inter e intra institucionales y aumento del impacto de las investigaciones. La aproximación adoptada es escalable, pudiendo ser utilizada por otras Instituciones de Ciencia y Tecnología.

**Palabras clave:** Ciencia de Datos, Cientometría, Gestión de Investigación, Políticas de Investigación, Universidad, transparencia

## 1. INTRODUÇÃO

Considerando que as Universidades têm como base os pilares ensino, pesquisa e extensão e que os professores devem desenvolver atividades nestes três eixos, faz-se imperativo o uso de métodos para avaliação de sua produção técnico-científica (HICKS et al., 2015; CARPENTER et al., 2014). Neste sentido a análise, extração de dados de produção científica, identificação de padrões bibliométricos e mensuração da ciência são tópicos relevantes na área de Bibliometria e Cientometria (MENA-CHALCO; CESAR JR, 2013, MINGERS; LEYDESDORFF, 2015, NHOATTO et al., 2021). Dentre a bibliometria e a cientometria pode-se afirmar que a primeira é focada em análises voltadas à gestão de bibliotecas quantificando os livros, literatura, documentos e revistas científicas. Enquanto a segunda aplica análises à produção científica, principalmente artigos, a circulação e o consumo da produção científica, visando a dinâmica da ciência: utilidade científica e impacto (PARRA et al., 2019).

A cientometria tem a finalidade de investigar a atividade científica como fenômeno humano, social e mediante parâmetros e indicadores baseados em modelos matemáticos (PARRA et al., 2019). Por meio desses indicadores que se guia o processo de investigação e mensuração da ciência (PARRA et al., 2019). Eles são úteis no planejamento e avaliação de políticas públicas voltadas para pesquisas (SANTOS; KOBASHI, 2009). Existem inúmeras formas de utilização das técnicas cientométricas, e segundo Spinak (1998), a importância pode ser vista analisando a seguinte lista de possíveis aplicações:

1. Identificar tendências e crescimento do conhecimento em todas as disciplinas;
2. Identificar autores e tendências em diferentes disciplinas;
3. Mensurar a utilidade de disseminação seletiva de informações;
4. Prever tendências;
5. Formular políticas ajustadas às aquisições de orçamento;
6. Adaptar as políticas públicas.

A aplicação de indicadores é uma necessidade inequívoca na otimização de recursos, quase sempre limitados, que se destinam ao fomento da

ciência e tecnologia (SILVA; BIANCHI, 2001). Conforme relata Lopes (2012), a busca pela excelência em áreas de pesquisa e a competição pelos recursos financeiros de agências de fomento são fatores que impulsionam a realização destes estudos e a adoção de métricas cientométricas.

De acordo com Dias et al. (2018), editais para financiamento de projetos de pesquisa por agências de amparo à produção de pesquisas no Brasil consideram dados dos currículos cadastrados na Plataforma Lattes (PL) como uma das formas de avaliação. A PL possui uma vasta fonte de informação para a criação e análise de redes de pesquisadores, tornando-a uma fonte extremamente rica para análise da produção científica brasileira (BALANCIERI et al., 2005). Neste contexto, este trabalho surge como demanda espontânea da UFOPA, que apresentou suas dificuldades na gestão da pesquisa e inovação. Assim, todo o estudo é baseado no contexto da instituição, sendo essa aplicada como estudo de caso deste trabalho.

A complexidade de realização dessas análises baseia-se na multitudine de produções técnico-científicas e docentes envolvidos, especificamente no caso da UFOPA que conta com 485 docentes ativos, mais de 70 mil produções e média 4.769 produções anuais nos últimos dez anos<sup>2</sup>. Frente ao exposto, nota-se que seja patente a necessidade de filtros temporais no cálculo de índices de produtividade dos docentes para distribuição de recursos de editais de fomento, além de outros filtros como área ou seleção de docentes pertencentes a um grupo de pesquisa etc. Tal funcionalidade é difícil de ser articulada com o uso de planilhas eletrônicas, uma vez que os dados ficam dispersos (e.g.: artigos publicados por pesquisadores da instituição em uma planilha e membros de grupo de pesquisa em outra).

Considerando a necessidade de avaliação da produção técnico-científica de pesquisadores com intuito de fomentar a pesquisa na instituição, juntamente com a inviabilidade de análise manual dado o volume de informações, e ainda, a multitudine de produções como: artigos científicos publicados em periódicos, livros, capítulos de livros, projetos de pesquisa, patentes, programas de computador, entre outras; verificou-se a imperatividade da utilização de

2. Disponível em: <http://www.ufopa.edu.br/media/file/site/proplan/documentos/2020/766ea1d5a36f6bde3acdc4b46199218e.pdf>. Acesso em 25 de outubro de 2022.

ferramentas computacionais para auxiliar a coleta e análise de dados da PL. A análise deve conter indicadores quantitativos e visualização da produtividade e desempenho de instituições e pesquisadores. Destaca-se ainda que análises de grandes grupos e de docentes de uma instituição são inviáveis de serem feitas manualmente, pois seria tarefa demorada e altamente suscetível a erros (MENA-CHALCO; CESAR-JR, 2013, NHOATTO et al., 2021). Nesse contexto pretende-se responder às seguintes perguntas de pesquisa:

- Pergunta de Pesquisa 1 (PP1): Qual o perfil dos pedidos e concessões de bolsas de iniciação científica da instituição?
- Pergunta de Pesquisa 2 (PP2): Quais as temáticas mais prevalentes nos projetos e nos planos de trabalho envolvidos nos editais de fomento da instituição?

## 2. OBJETIVOS

À luz dos fatos apresentados, o objetivo do presente trabalho é otimizar a gestão da pesquisa e inovação na UFOPA, conduzindo uma análise orientada a dados sobre as produções técnico-científicas dos pesquisadores da instituição. Espera-se que os resultados obtidos permita auxiliar na tomada de decisões referente aos editais de pesquisa da instituição. Visando adotar prática de desenvolvimento ágil ao projeto de ciência de dados, definiram-se objetivos menores e que pudessem ser validados pelos gestores, estes foram:

1. Identificar os perfis e tendências de pesquisa da Instituição a partir de dados extraídos da plataforma Lattes, utilizando algoritmos de mineração de textos e de cientometria;
2. Automatizar a obtenção dos dados de produção dos pesquisadores e suas relações de cooperação;
3. Prover painel com informações para auxiliar no planejamento e na avaliação de políticas voltadas para pesquisa e inovação.

## 3. METODOLOGIA

O processo de análise adotado no trabalho segue o fluxo de mineração de dados apresentado em Domingues e Rezende (2005), concebido em forma o ciclo composto de três grandes etapas: Pré-Processamento, Extração de Padrões e Pós-Processamento, proposta em Baranauskas (2002) e adotada em Rezende (2003), conforme pode ser observado na Figura 01.

FIGURA 01 - ETAPAS DO PROCESSO DE MINERAÇÃO DE DADOS



A esse modelo é incluída uma fase anterior ao processo de Mineração de Dados, referente à Identificação do Problema e uma fase posterior ao processo, que se refere à Utilização do Conhecimento obtido. Nas subseções a seguir são descritas a fase de etapas de Identificação do Problema, as etapas do processo de Mineração de Dados, e a fase Utilização do Conhecimento obtido, seja em um sistema inteligente ou como apoio em processos de tomada de decisão.

### 3.1 Identificação do Problema

Nessa fase são definidos os objetivos a serem alcançados e que tipo de informação se deseja extrair dos dados. Desta forma, foi realizada a identificação do problema, também chamado de entendimento do domínio de aplicação. Neste passo foi necessário compreender os objetivos do projeto,

problematizar esse conhecimento tomando como base a mineração e análise de dados e, por fim, a definição do plano para alcançar os objetivos.

Para o presente estudo, o início foi centrado no problema que foi apresentado pela Diretoria de Pesquisa (DPE) ligada à Pró-Reitoria de Pesquisa, Pós-Graduação e Inovação Tecnológica (PROPPIT). A fase inicial visou a identificação e a motivação do problema, juntamente com setor responsável pela pesquisa da UFOPA. Considerando que eram vários desafios apresentados, selecionou-se o mais relevante, tal como destacado na subseção 3.1.2. A partir do estudo da relevância e viabilidade, passou-se para a definição dos objetivos, já apresentados na Seção 2.

### 3.1.1 Sobre a instituição

A Universidade Federal do Oeste do Pará é a primeira Instituição Federal de Ensino Superior (IFES) sediada no interior da Amazônia brasileira, no oeste do Pará. Criada com uma abordagem regional e multicampi, com sede na cidade de Santarém e com campi localizados nos municípios de Alenquer, Itaituba, Juruti, Monte Alegre, Óbidos e Oriximiná. Foi criada pela Lei nº 12.085, de 5 de novembro de 2009, por desmembramento e integração dos campi da Universidade Federal do Pará (UFPA) e da Universidade Federal Rural da Amazônia (UFRA), envolvida no Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais (REUNI).

Segundo o Plano de Desenvolvimento Institucional (PDI) 2019 a 2023 da UFOPA<sup>3</sup>, sua estrutura é composta por 7 institutos e 9 campi, espalhados por 7 municípios da região. Desenvolve um processo de formação de 7.909 alunos ativos, destes 6.924 são da graduação e 985 da pós-graduação, excluindo alunos trancados. Seu quadro funcional é constituído por 1.095 servidores efetivos, sendo 607 técnicos administrativos em educação, distribuídos nas atividades-meio e atividades-fim, além dos 485 docentes conforme consta na Seção 1. Oferece para a sociedade formação em 42 cursos de graduação<sup>4</sup> e 16 programas de pós graduação<sup>5</sup>. Envolve 49 linhas

de pesquisas e 22 áreas cadastradas na plataforma Sucupira<sup>6</sup>. Dentre elas: biotecnologia, biodiversidade, matemática aplicada e propriedade intelectual e transferência de tecnologia para inovação. E seus pesquisadores coordenam 88 grupos de pesquisa<sup>7</sup>.

A gestão da pesquisa na instituição é gerenciada pela PROPPIT, mais especificamente pela DPE. Esta unidade gerencia a pesquisa por meio de editais de fomento e políticas públicas.

Alguns desafios operacionais e de gestão foram elencados por servidores atuantes na DPE a fim de que às análises propostas pudessem auxiliá-los, dentre os quais, foi destacado a busca de agilidade e viabilidade por meio da automatização de processos. O primeiro desafio apresentado pela PROPPIT foi a mensuração de índices de produtividade personalizáveis dos docentes para distribuição de recursos de pesquisa por meio de editais. Sua justificativa se baseia na grande quantidade de produções e pesquisadores que participam dos processos de seleção. Assim, almeja-se maior eficácia no processo de distribuição e alocação de recursos.

### 3.1.2 Desafio: Edital Pibic

A necessidade de uma nova abordagem de análise de dados surgiu com uma demanda espontânea da instituição ao levantar dados de forma ágil e eficiente sobre as produções desenvolvidas por seus pesquisadores. O problema inicial foi a dificuldade em se quantificar um índice de produtividade para distribuição de bolsas de pesquisa por meio de editais de fomento interno. Tal qual outras IFES, para distribuir recursos de pesquisas, a universidade estabeleceu um índice de produtividade dos pesquisadores, em que determinados tipos de produção têm pontuações específicas e ainda, classes de produtos têm limitação de pontuação.

Desta forma, as análises conduzidas foram apresentadas para a DPE e aplicadas no Edital de pesquisa 02 de 2020<sup>8</sup>. O Edital tinha por finalidade despertar vocação científica e incentivar talentos potenciais entre estudantes de graduação universi-

3. Disponível em: <http://www.ufopa.edu.br/media/file/site/proplan/documentos/2020/766ea1d5a36f6bde3acdc4b46199218e.pdf>. Acesso em 25 de outubro de 2022.

4. Disponível em: <http://www.ufopa.edu.br/ufopa/ensino/cursos-de-graduaao-2/>. Acesso em 25 de outubro de 2022.

5. Disponível em: <http://www.ufopa.edu.br/proppit/programas/>. Acesso em 25 de outubro de 2022.

6. Disponível em: <https://sucupira.capes.gov.br/sucupira/public/consultas/coleta/linhaPesquisa/listaLinhaPesquisa.xhtml>. Acesso em 29 de julho de 2021.

7. Disponível em: <http://www.ufopa.edu.br/proppit/diretoria-de-pesquisa/>. Acesso em 29 de julho de 2021.

8. Disponível em: <http://www.ufopa.edu.br/media/file/site/proppit/documentos/2021/578384e5368d4d14f2696814c6ef2864.pdf>. Acesso em 24 de fevereiro de 2023.

tária, mediante participação em projeto de pesquisa, orientados por pesquisador qualificado em cotas disponíveis pela agência de fomento e pela estrutura orçamentária da UFOPA aprovada no Conselho Superior. A partir do entendimento do problema e dos dados disponíveis, passou-se para a etapa subsequente, de preparação dos dados.

### 3.2 Pré-Processamento

A primeira etapa do fluxo de mineração de dados é chamada de “Pré-Processamento”, na qual além da aquisição dos dados foi possível identificar problemas nos mesmos e fornecer *insights* iniciais sobre os dados disponíveis que seriam úteis para as análises. Visto que geralmente os dados selecionados para o processo de Mineração de Dados não estão em um formato adequado para a extração de conhecimento.

Para esta fase primeiramente foi definido a linguagem de programação Python, em sua versão 3<sup>9</sup>, para o desenvolvimento do projeto. Além de ser uma linguagem bastante utilizada no meio científico, dispõe de muitas bibliotecas úteis para análise de dados e aprendizado de máquina. Algumas dessas bibliotecas utilizadas neste projeto foram a biblioteca *Pandas*<sup>10</sup> para manipulação de dados e para visualização de dados estatísticos a biblioteca *Seaborn*<sup>11</sup>.

As etapas de pré-processamento do conteúdo textual aplicadas aos dados foram realizadas utilizando o *pipeline* convencional da mineração de textos, a saber: remoção de sinais de pontuação, remoção de caracteres numéricos, remoção de espaços duplos, conversão para caracteres em minúsculo, remoção de stopwords e remoção de acentuação (CIRQUEIRA et al., 2018; COSTA et al., 2022).

#### 3.2.1 Coleta de dados

A etapa de coleta de dados permitiu a seleção das informações relevantes para a formação do banco de fonte de dados necessária para o processamento das análises e visualizações propostas neste estudo. Desta forma, aqui são apresentadas as fontes de dados utilizadas na composição do estudo

e na sequência um resumo do conjunto de dados construído.

#### 3.2.1.1 Plataforma Lattes

A Plataforma Lattes foi disponibilizada pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) com a finalidade de unificar o cadastro nacional de pesquisadores. Em 2012, a Plataforma Lattes contava com mais de dois milhões de currículos cadastrados (ALVES et al., 2012). A PL tornou-se a maior e mais relevante base de dados da produção acadêmica brasileira, por armazenar essencialmente informações dos currículos dos pesquisadores (LANE, 2010, MEGLIORANSI, 2019). O currículo Lattes inclui identificação, formação acadêmica, prêmios científicos/artísticos, produções bibliográficas, orientações, entre outras informações (AMORIN, 2003).

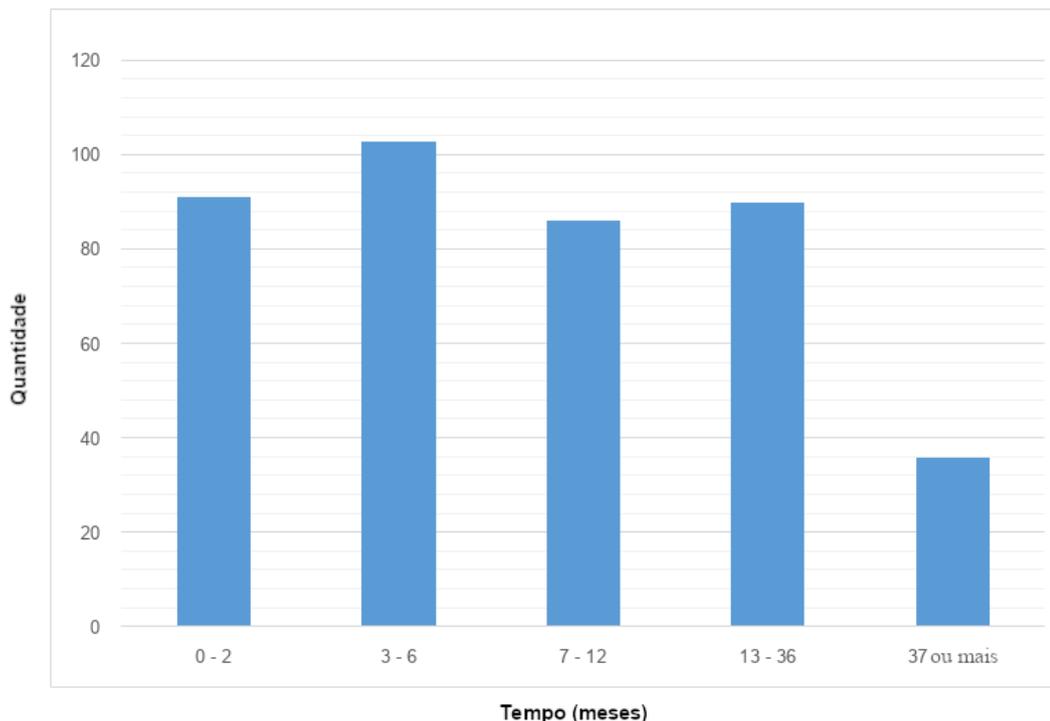
Devido a ser uma plataforma de inserção obrigatória para pesquisadores que necessitam participar de editais de fomento à pesquisa, se torna uma fonte valiosa de informações unificadas. O CNPq possibilita às instituições a extração de dados da Plataforma Lattes, incluindo a base de Currículos Lattes e o Diretório dos Grupos de Pesquisa, por meio da ferramenta Lattes Extrator<sup>12</sup>. Isto viabilizou o desenvolvimento deste projeto uma vez que a instituição já possuía permissão do CNPq para coletar dados da PL por meio de seu *web service* (LEITE et al., 2011, MUGNAINI et al., 2011). Destaca-se que a instituição precisou atender os seguintes requisitos: i) Ser Instituição de Ensino e Pesquisa; ii) Possuir um IP dedicado e iii) Desenvolver solução tecnológica para integração com o *web service* do Lattes Extrator. Após efetuar a coleta de dados, foi produzido um histograma, conforme disposto na Figura 02, baseado na data da última atualização dos currículos dos docentes da Instituição.

9. Disponível em: <https://www.python.org/>. Acesso 02 em de julho de 2021.

10. Disponível em: <https://numpy.org/>. Acesso 02 em de julho de 2021.

11. Disponível em: <https://seaborn.pydata.org/>. Acesso 02 em de julho de 2021.

12. Disponível em: : <http://memoria.cnpq.br/web/portal-lattes/extracoes-de-dados>. Acesso 02 em de julho de 2021.

**FIGURA 02 - HISTOGRAMA DA ÚLTIMA ATUALIZAÇÃO DOS CURRÍCULOS LATTES DOS PESQUISADORES DA UFOPA**

Fonte: Elaborado pelos autores.

Conforme visto na Figura 02, observa-se que a grande maioria dos docentes atualizou seus currículos em menos de 12 meses da data da coleta de dados. Apesar de sua completude, alguns dados não estão disponíveis no Lattes, como o número de citações, por este motivo, buscou-se no Google Scholar uma alternativa, tal como descrito a seguir.

### 3.2.1.2 Google Scholar

O *Google Scholar*, é uma ferramenta gratuita, que permite localizar trabalhos acadêmicos de vários tipos, além de determinar a frequência com que foram citados em outras publicações acadêmicas (CAREGNATO, 2011). Assim, tem possibilidade de ser utilizada como referência principal em índices de citações para estudos bibliométricos (NORUZI, 2005), em mensuração do impacto das pesquisas (VAUGHAN; SHAW, 2008) e periódicos de acesso aberto (MUKHERJEE, 2009).

O processo de coleta de dados foi realizado por meio de um *software* que rastreia a *web* e é capaz de reconhecer automaticamente os campos que compõem os documentos científicos e suas referências.

Esse método inovador resultou no desenvolvimento de uma ferramenta sem precedentes, que disponibiliza gratuitamente dados bibliográficos e milhões de artigos completos (JACSÓ, 2010).

Nesta etapa foi utilizada a seção *Profiles*<sup>13</sup> da plataforma *Google Scholar* que indexa uma lista de pesquisadores associados a uma determinada instituição, neste caso pesquisadores com *webmail @ufopa.edu.br* validados. Nessa página, é possível visualizar os perfis dos pesquisadores, incluindo suas publicações, citações, índice-h, colaborações e outras informações relacionadas à pesquisa. Diante do exposto, optou-se por utilizar dados provenientes do *Google Scholar* para obter índices de impacto e citações das pesquisas dos docentes da instituição. Permitindo assim um gerenciamento e visualização de forma rápida e unificada das produções e pesquisadores mais relevantes na instituição. Para identificar os pesquisadores tanto no Lattes quanto no *Google Scholar*, utilizam-se os dados armazenados nos sistemas institucionais, que serão descritos a seguir.

13. Disponível em: [https://scholar.google.com/citations?view\\_op=view\\_org&hl=en&org=2173319500607027007](https://scholar.google.com/citations?view_op=view_org&hl=en&org=2173319500607027007). Acesso em de julho de 2021.

### 3.2.1.3 Sistemas Institucionais

Alguns dados que não se pode obter pelas fontes anteriormente citadas foram obtidos por meio de sistemas institucionais. A UFOPA utiliza um sistema acadêmico bem difundido pelas IFES brasileiras, o Sistema Integrado de Gestão de Atividades Acadêmicas (SIGAA), desenvolvido pela Universidade Federal do Rio Grande do Norte (UFRN). Segundo o site da UFRN, 73 instituições utilizam seus sistemas.

A base de dados do SIGAA fornece dados sobre os docentes, servidores técnicos e discentes da instituição, suas unidades de lotação, inícios e terminos de vínculo, facilitando assim a alimentação de maneira automática da ferramenta utilizada para coleta de dados. O SIGAA permite, por meio do módulo pesquisa, o cadastro e gerenciamento de editais de pesquisa. Assim, a coleta de dados buscou

os inscritos nestes editais para produzir resultados de pontuações. É por meio do Sistema Intellectus<sup>14</sup>, subsistema do SIGAA, que se obtém os dados dos currículos Lattes importados usando o Lattes Extrator.

### 3.2.1.4 Conjunto de dados

A partir da coleta obteve-se a construção do conjunto de dados com o resumo das solicitações dos pesquisadores inscritos no edital de pesquisa 2020 da UFOPA e dados resultantes da fusão das plataformas mencionadas nas subseções anteriores. O conjunto de dados foi estruturado em um arquivo no formato *Comma-Separated Values (CSV)*. Com o total de 227 planos inscritos e 13 colunas conforme a Tabela 01.

14. Disponível em: <https://portalcooperacao.info.ufrn.br/pagina.php?a=parceiros>. Acesso em 29 de julho de 2021.

TABELA 01 - RESUMO DOS DADOS COLETADOS.

ATRIBUTO	DESCRIÇÃO
nome	Nome do pesquisador
email	Email do pesquisador
genero	Identidade de gênero do pesquisador
lotacao	Lotação se refere ao instituto responsável pelo plano de pesquisa
titulo_pt1	Título do primeiro plano de trabalho
opcao1_modalidade_bolsa_pt1	Se refere à primeira opção da modalidade da bolsa para o primeiro plano de trabalho.
opcao1_modalidade_bolsa_pt12	Se refere à segunda opção da modalidade da bolsa para o primeiro plano de trabalho.
titulo_pt2	Título do segundo plano de trabalho
opcao1_modalidade_bolsa_pt2	Se refere à primeira opção da modalidade da bolsa para o segundo plano de trabalho.
opcao2_modalidade_bolsa_pt22	Se refere à segunda opção da modalidade da bolsa para o segundo plano de trabalho.
area_conhecimento	Se refere à área do conhecimento do pesquisador
producao_academica	Produção acadêmica que foi usada para cálculo da Pontuação de acordo com resolução vigente
ano_formacao	Se refere ao ano em que o pesquisador

Por meio deste conjunto de dados foi possível analisar, por exemplo, a distribuição das modalidades de bolsa escolhidas pelos pesquisadores para cada plano de trabalho, identificar padrões de preferência em relação às diferentes opções e o cálculo da pontuação da produção acadêmica. Além disso, é possível realizar outras análises exploratórias para identificar tendências ou *insights* relevantes.

### 3.3 Extração de Padrões

A etapa de Extração de Padrões é direcionada a cumprir os objetivos definidos na fase de Identificação do Problema. Esta etapa abrangeu as atividades que objetivaram a construção do conjunto de dados final utilizado para o desenvolvimento deste trabalho.

Foram definidas as técnicas de análises de dados utilizadas no projeto. Nesta fase definiu-se que o presente estudo de caso consistiria na combinação de análise cientométrica e mineração de textos, tendo como base os dados obtidos a partir da Plataforma Lattes. Além disso, em sincronia com os sistemas institucionais – SIGAA e Intellectus e com o *Google Scholar*.

Com os dados pré-processados, passou-se para a etapa de modelagem ainda por meio do pipeline convencional da mineração de textos. Sobretudo, foi um trabalho de mineração de textos, usando as bibliotecas *Natural Language Toolkit (NLTK)*<sup>15</sup>, (BIRD; KLEIN; LOPER, 2009), *re*<sup>16</sup> e *scikit-learn*<sup>17</sup>. A modelagem de tópicos foi realizada utilizando a abordagem não supervisionada *Non-Negative Matrix Factorization (NMF)*. Por ser mais eficiente em tarefas de mineração de textos curtos consegue produzir tópicos mais coerentes quando comparado a outras abordagens (CHEN et al., 2019; COSTA et al., 2022; NUGROHO et al., 2020). E como representação *Term Frequency-Inverse Document Frequency (TF-IDF)* (RAMOS, 2003). Na modelagem de tópicos a avaliação qualitativa foi baseada nos Títulos do primeiro plano de trabalho dos projetos inscritos no edital que nos levou à escolha de 10 tópicos e 20 principais palavras por tópico.

### 3.4 Pós-Processamento e Utilização do conhecimento

No Pós-Processamento o processo de Mineração de Dados na qual o conhecimento extraído pode ser simplificado, avaliado, visualizado ou simplesmente documentado para o usuário final. Com isso, esta etapa consistiu na divulgação das análises dos dados coletados. Para isso, neste estudo será realizada a apresentação dos resultados da modelagem de tópicos produzidos, revelando as principais temáticas que representam as vocações da instituição, destacando sua inserção regional. E também a construção de um *dashboard* produzido durante o edital de pesquisa 2020.

A construção do *dashboard* foi baseada em boas práticas para mostrar conhecimento por meio de dados, fornecendo visualizações eficientes para a tomada de decisões, conforme Knafllic (2019) orienta. Algumas destas orientações foram aplicadas para entregar conhecimento por meio de infográficos simples e objetivos, evitando problemas comuns que dificultam a transmissão da mensagem. Por exemplo, evitar gráficos em formato de pizza quando existem muitos dados ou pouca diferença entre os valores, evitar gráficos 3D caso a terceira dimensão não apresente uma nova informação e eliminar elementos apenas visuais que não colaboram com a passagem da mensagem.

Este infográfico produzido foi publicado na página institucional da diretoria de pesquisa<sup>18</sup>, assim como mostra a Figura 03. Convém destacar que o edital PIBIC foi avaliado internamente pela diretoria de pesquisa por meio de pesquisa de satisfação com os docentes participantes. A descrição do *dashboard* e dos resultados da da modelagem de tópicos encontram-se descritos na Seção a seguir.

15. Disponível em: <https://www.nltk.org/>. Acesso em 29 de julho de 2022.

16. Disponível em: <https://docs.python.org/3/library/re.html#module-re>. Acesso em 29 de julho de 2022.

17. Disponível em: <https://scikit-learn.org/stable/>. Acesso em 29 de julho de 2022.

18. Disponível em: <http://www.ufopa.edu.br/media/file/site/proppit/documentos/2020/5b4ac5b5dbed47c76d67b207d0223125e.pdf>. Acesso em 24 de fevereiro de 2023.

## 4. RESULTADOS

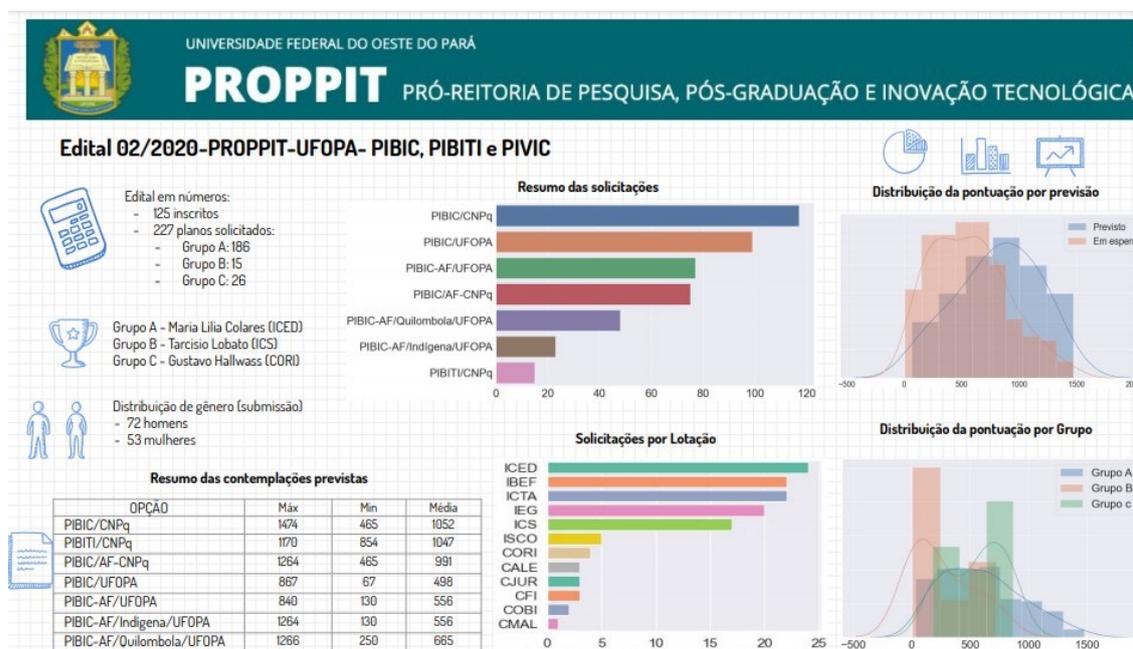
Nesta Seção serão apresentados os resultados obtidos a partir da metodologia descrita anteriormente, perpassando pela análise do infográfico produzido (Figura 03) e os resultados da modelagem de tópicos.

### 4.1 Análise do infográfico

A análise do infográfico traz a baila o perfil dos pedidos e concessões de bolsas de iniciação científica da instituição satisfazendo à questão PP1.

Neste processo 125 docentes estavam concorrendo às cotas. A Figura 03 mostra os pesquisadores mais bem pontuados, a distribuição das bolsas por setores e a distribuição da pontuação por grupo dos pesquisadores. De acordo com o gráfico “Resumo das Solicitações” a modalidade PIBIC/CNPq possui o maior número de solicitações. Também podemos obter um *insight* quanto à inclinação por políticas públicas, em que a maior parte dos professores preferem quotas ligadas aos discentes quilombolas em detrimento de discentes indígenas.

FIGURA 03 - INFOGRÁFICO DO EDITAL 02/2020 DE PESQUISA



Fonte: Elaborado pelos autores.

Ao observarmos as pontuações no “Resumo das contemplações previstas”, a pontuação máxima foi obtida na modalidade PIBIC/CNPq, enquanto a pontuação mínima cabe à modalidade PIBIC/UFOPA. Nota-se uma discrepância de mais de mil pontos entre o pesquisador melhor pontuado (1474 pontos) e o pesquisador com menor pontuação (67 pontos), todavia também possui acesso a bolsa de iniciação científica. Ressalta-se que este fato pode estar relacionado por a UFOPA ser uma universidade nova, com jovens doutores, professores em processo de doutoramento e até mesmo docentes especialistas. A partir dessa percepção a instituição pode avaliar melhor como está sendo realizada a distribuição de

recursos. Além disso, também pôde ser observado o comportamento das solicitações pelos institutos, na qual o Instituto de Ciências da Educação (ICED) possui maior número de solicitações. O ICED é um dos institutos com maior quantidade de docentes, sendo sua origem a fusão do *campus* da UFPA e do *campus* da UFRA, localizados em Santarém – Pará. Em que cursos de Licenciatura herdados da UFPA, bem como toda a estrutura física, passaram a compor o ICED. Este comportamento também se reflete nos dois institutos que seguem sendo eles Instituto de Biodiversidade e Florestas (IBEF) e Instituto de Ciências e Tecnologia das Águas (ICTA), caracterizando-os os três institutos mais sólidos. Entretanto,

vale salientar que o Instituto de Engenharia e Geociências (IEG), um dos mais novos, também possui boa competitividade no cenário institucional.

No infográfico da Figura 03 também pode ser observada a distribuição de cotas por 3 faixas, sendo elas grupos A, B e C. Os grupos A e B consistem em docentes da sede com titulação máxima doutorado e mestrado, respectivamente. O grupo C representa docentes dos campi fora da sede, com mestrado ou doutorado. O gráfico da “Distribuição da pontuação por Grupo” evidencia que o grupo A formado por professores com maior titularidade têm maior competitividade. Seguidos pelos professores do grupo C que também possuem doutores em sua composição, porém têm acesso a menos recursos, perante o exposto de estarem fora da sede.

Considerando as políticas de capacitação do corpo docente da instituição ao longo de sua existência, a análise da pontuação dos docentes foi cruzada com quantitativo dos que já alcançaram a titulação máxima doutorado por meio destas capacitações. Desta forma, verificou-se a não necessidade do grupo B, visto que este também adiciona uma complexidade para o edital. Além disso, foi atribuída pontuação mínima exigida para cada grupo nos editais subsequentes, conforme a Portaria Nº 30 / 2022 - PROPPIT<sup>19</sup>. Destaca-se que este foi um dos impactos diretos do estudo, balizando a construção de uma política pública mais adequada à realidade institucional.

## 4.2 Modelagem de Tópicos

A modelagem de tópicos revela as temáticas mais prevalentes nos projetos e nos planos de tra-

19. Disponível em: <https://www.ufopa.edu.br/media/file/site/proppit/documentos/2022/7d36f3adbbae895c736ef07f808f6392.pdf>. Acesso em 25 de outubro de 2022.

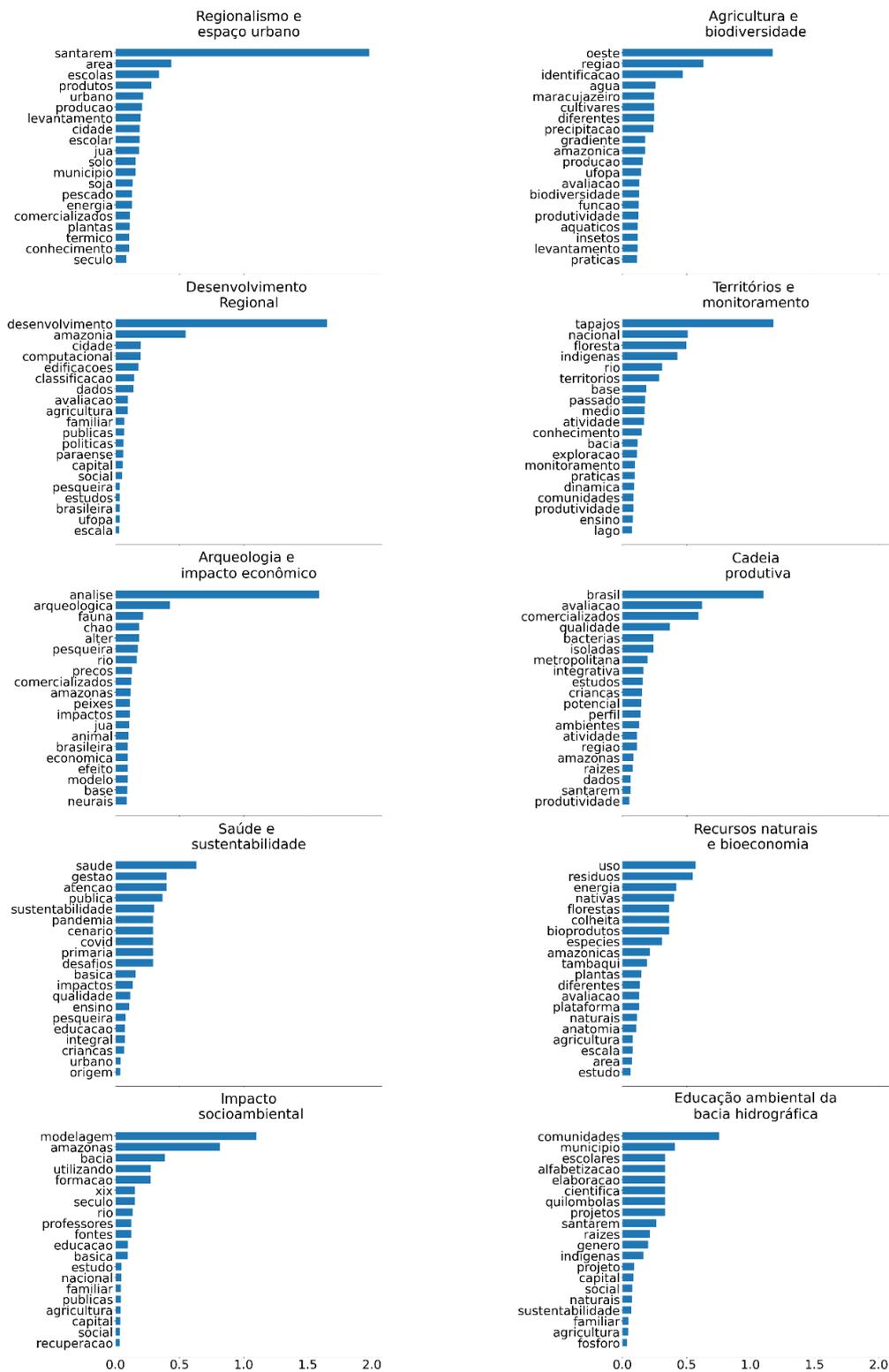
balho envolvidos nos editais de fomento da instituição, atendendo à PP2. Por meio da Figura 04 se observa que a UFOPA tem como principais vocações temas relacionados ao desenvolvimento regional sustentável, como agricultura, biodiversidade, cadeia produtiva, saúde e sustentabilidade, entre outros. Os projetos da instituição visam entender a região, promover avanços econômicos e sociais, valorizando o comércio e preservando o patrimônio cultural e ambiental.

Ainda de acordo com a Figura 04, na área de regionalismo e espaço urbano, a UFOPA se destaca em pesquisas buscando entender o desenvolvimento regional a partir das especificidades da região, produção de conhecimento sobre a cidade de Santarém, a área urbana e a produção de produtos comerciais, como soja e pescado, além de levantamento do solo e vegetação.

Outro tema bastante trabalhado por pesquisadores da UFOPA é Agricultura e Biodiversidade. Neste ponto, destacam-se estudos que visam identificar práticas agrícolas sustentáveis e a conservação da biodiversidade local. Esse tema está diretamente relacionado ao de Desenvolvimento Regional, onde a instituição tem pesquisas sobre políticas públicas para agricultura familiar, políticas para a cidade de Santarém e região, além de estudos computacionais e avaliação de impacto econômico e social.

Na temática de Territórios e Monitoramento são realizados estudos sobre a bacia do rio Tapajós, floresta e dinâmicas de comunidades indígenas e ribeirinhas, além de práticas de monitoramento e ensino na região. Por sua vez, no tema de Arqueologia e Impacto Econômico, a universidade realiza análises arqueológicas, avaliação de preços de produtos comercializados e estudos sobre impactos econômicos, incluindo modelos neurais e fauna local.

FIGURA 04 - TÓPICOS MODELADOS DAS PRINCIPAIS TEMÁTICAS DOS PLANOS DE TRABALHOS



Fonte: Elaborado pelos autores.

A universidade também se destaca em estudos sobre a Cadeia Produtiva, incluindo avaliação de qualidade de produtos comercializados, potencial de criação de ambientes metropolitanos e estudos integrativos de dados em diferentes regiões, como Santarém e o estado do Amazonas. Na temática de Saúde e Sustentabilidade às pesquisas são voltadas para gestão e atenção à saúde, incluindo desafios da pandemia, impactos na saúde pública e qualidade de atendimento, além de educação integral e desafios da saúde urbana e rural.

A universidade também tem pesquisas sobre Recursos Naturais e Bioeconomia, incluindo uso de resíduos, energia, colheita de espécies nativas, como o tambaqui, e avaliação de anatomia e agricultura em diferentes escalas e áreas. Também é investigado o Impacto Socioambiental com estudos de modelagem e formação socioambiental em diferentes regiões, como a bacia do rio Amazonas, com foco na recuperação social e ambiental.

Por fim, em relação à Educação Ambiental da Bacia Hidrográfica, a universidade tem projetos sobre atividades escolares, alfabetização, gênero, sustentabilidade e raízes culturais em comunidades municipais, quilombolas e indígenas. Além de estudos que buscam fomentar a educação ambiental sobre a região da bacia hidrográfica.

## 5. CONSIDERAÇÕES FINAIS

Visando resolver problemas relacionados à organização dos dados de pesquisa da instituição e transparência das produções técnico-científicas, este estudo consistiu em uma análise orientada a dados de professores inscritos em editais de pesquisa da instituição, para auxiliar na tomada de decisão na gestão de pesquisa. As análises conduzidas permitem responder às perguntas de pesquisa, revelando importantes *insights* sobre perfil dos pedidos e concessões de bolsas e tendências de pesquisas da instituição.

Por exemplo, em relação à PP1 – foram obtidos *insights* por meio do infográfico produzido visando maior eficiência no processo de distribuição e alocação de recursos (Figura 03). Destaca-se que não era de conhecimento do órgão gestor de pesquisa padrões de distribuição de cotas, em termos de qual instituto é contemplado com mais bolsas, qual instituto tem a maior demanda e qual instituto possui

melhor pontuação quanto a produção acadêmica, por exemplo.

Com o auxílio das análises verificou-se questões relacionadas à preferência por políticas públicas, em que a maioria dos docentes são propensos a realizarem mais projetos com alunos quilombolas do que com alunos indígenas. Foi observada também a configuração dos institutos com produção técnico-científica relevante.

Além disso, verificou-se docentes com baixa produção técnico-científicas, no entanto, com acesso a bolsas de iniciação científica. Ainda, a não necessidade da faixa de docentes da sede com titulação máxima mestrado, considerando o amadurecimento do corpo docente por meio de processos de capacitação. A retirada deste grupo conferiu maior simplicidade aos certames posteriores. Dessa forma, conclui-se que as análises conduzidas impactaram a política pública, pois outros editais seguiram as novas diretrizes. Com o auxílio dessas análises foi possível avaliar os resultados para otimizar modelos que mais se adequam aos perfis de pesquisa da instituição.

Sobre à questão PP2 - verificou-se que a UFOPA possui diversas vocações que buscam entender e promover o desenvolvimento regional de forma sustentável. Com sua inserção no interior da Amazônia, a instituição se destaca em pesquisas sobre temas como regionalismo e espaço urbano, agricultura e biodiversidade, territórios e monitoramento, arqueologia e impacto econômico, cadeia produtiva, saúde e sustentabilidade, recursos naturais e bioeconomia, impacto socioambiental e educação ambiental da bacia hidrográfica. Todas essas áreas de pesquisa visam contribuir para a promoção do desenvolvimento socioeconômico e ambientalmente responsável da região e suas comunidades, valorizando sua cultura e conhecimentos tradicionais.

Além disso, por meio dos resultados obtidos, sobretudo no que tange à visualização da distribuição dos grupos, institutos e das suas respectivas pontuações, foi possível entender melhor os perfis de pesquisa da instituição. Tal fato, foi particularmente importante para a atualização da política de pesquisa da instituição. Ademais, utilizaram-se técnicas de análises baseadas em combinação de análise cientométrica e modelagem de tópicos tendo como base os dados obtidos a partir da Plataforma Lattes para prover informações com maior acuidade

e de fácil entendimento, possibilitando a tomada de decisões baseadas em dados. Desta forma, às análises conduzidas também subsidiaram as comissões de pesquisa na avaliação dos projetos de pesquisa da instituição e foi fundamental para a construção da nova política de pesquisa da UFOPA<sup>20</sup>.

Dos impactos observados, vale mencionar a melhora na eficiência da gestão por meio do processo de mapeamento das áreas e tendências de pesquisa, integração com dados institucionais, acompanhamento da produtividade e possibilidade de avaliação e direcionamento das pesquisas e políticas públicas da instituição. Algumas dificuldades foram enfrentadas no decorrer do projeto, como o entendimento do domínio de aplicação tão amplo que é a cionometria, aliada a multiplicidade e dinamicidade das fontes de dados. Este foi o quesito que mais consumiu o tempo deste projeto. Os resultados obtidos possibilitam uma maior transparência, inserção social da universidade e capacidade de tomada de decisão orientada a dados, fornecendo auxílio à gestão da pesquisa e inovação e avaliação de políticas técnico-científicas.

## AGRADECIMENTOS

Este trabalho foi parcialmente financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) - DT - 308334/2020; pela Fundação Amazônia de Amparo a Estudos e Pesquisas (FAPESPA) - PRONEM-FAPESPA/CNPq nº 045/2021. Agradecemos também aos revisores(as) pelas sugestões que muito auxiliaram na melhora do trabalho.

20. Disponível em: <http://www.ufopa.edu.br/media/file/site/proppit/documentos/2021/1dbf9114fd2b586a041fa1461c607e1d.pdf>. Acesso em 25 de outubro de 2022.

## REFERÊNCIAS

- ALVES, A. D.; YANASSE, H. H.; SOMA, N. Y. Lattesminer: uma linguagem de domínio específico para extração automática de informações da plataforma lattes. XII Workshop de Computação Aplicada - WORCAP, 2012.
- AMORIN, C. V. Organização do currículo: plataforma Lattes. Pesquisa Odontológica Brasileira, v. 17, n. suppl 1, p. 18–22, 2003.
- BALANCIERI, R. et al. A análise de redes de colaboração científica sob as novas tecnologias de informação e comunicação: um estudo na plataforma lattes. Ciência da informação, IBICT, v. 34, n. 1, 2005.
- BIRD, S; KLEIN, E; LOPER, E. Natural language processing with Python: analyzing text with the natural language toolkit. “ O’Reilly Media, Inc.”, 2009.
- BARANAUSKAS, José Augusto et al. Extração automática de conhecimento por múltiplos indutores. 2002.
- CAREGNATO, S. E. Google acadêmico como ferramenta para os estudos de citações: avaliação da precisão das buscas por autor. Pontodeacesso, v. 5, n. 3, p. 72–86, 2011.
- CARPENTER, C. R.; CONE, D. C.; SARLI, C. C. Using publication metrics to highlight academic productivity and research impact. Academic emergency medicine, Wiley Online Library, v. 21, n. 10, p. 1160–1172, 2014.
- CHEN, Y. et al. Experimental explorations on short text topic mining between LDA and NMF based Schemes. Knowledge-Based Systems, v. 163, p. 1-13, 2019.
- CIRQUEIRA, D. et al. A literature review in preprocessing for sentiment analysis for Brazilian Portuguese social media. In: 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI). IEEE. p. 746-749, 2018.
- COSTA, G. D. S. et al. Feminismo e Redes Sociais Online: uma Análise de Tweets sobre o Dia Internacional da Mulher. In: Anais do XI Brazilian Workshop on Social Network Analysis and Mining. SBC, p. 169-180. 2022.
- DIAS, T. M.; MOREIRA, T. H.; DIAS, P. M. Caracterização e análise das redes de colaboração científica dos bolsistas de produtividade em pesquisa do cnpq. In: SBC. Anais do VII Brazilian Workshop on Social Network Analysis and Mining. [S.l.], 2018.
- DOMINGUES, M. A; REZENDE, S. O. Pós-processamento de Regras de Associação usando Taxonomias. INFOCOMP Journal of Computer Science, v. 4, n. 2, p. 26-37, 2005.
- HICKS, D. et al. Bibliometrics: the leiden manifesto for research metrics. Nature News, v. 520, n. 7548, p. 429, 2015.
- JACSÓ, P. Metadata mega mess in google scholar. Online Information Review, Emerald Group Publishing Limited, 2010.
- KNAFLIC, C. N. Storytelling com Dados: Um guia sobre visualização de dados para profissionais de negócios. [S.l.]: Alta Books, 2019
- LANE, J. Let’s make science metrics more scientific. Nature, Nature Publishing Group, v. 464, n. 7288, p. 488–489, 2010.
- LEITE, P; MUGNAINI, R.; LETA, J. A new indicator for international visibility: exploring brazilian scientific community. Scientometrics, Akadémiai Kiadó, co-published with Springer Science+ Business Media BV, v. 88, n. 1, p. 311–319, 2011.

- LOPES, G. R. Avaliação e recomendação de colaborações em redes sociais acadêmicas. 2012.
- MEGLIORANSI, M. D. Um sistema para coleta e apresentação de indicadores bibliométricos da plataforma lattes. Universidade Federal do Pampa, 2019.
- MENA-CHALCO, J. P.; JÚNIOR, C. Prospecção de dados acadêmicos de currículos lattes através de scriptlattes. *Bibliometria e Cientometria: reflexões teóricas e interfaces*. São Carlos: Pedro & João, p. 109–128, 2013.
- MINGERS, J.; LEYDESDORFF, L. A review of theory and practice in scientometrics. *European Journal of Operational Research*, v. 246, n. 1, p. 1–19, 2015. ISSN 0377-2217. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S037722171500274X>>.
- MUGNAINI, R.; LEITE, P.; LETA, J. Fontes de informação para análise de internacionalização da produção científica brasileira. *PontodeAcesso*, v. 5, n. 3, p. 87–102, 2011.
- MUKHERJEE, B. Do open-access journals in library and information science have any scholarly impact? a bibliometric study of selected open-access journals using google scholar. *Journal of the American Society for Information Science and Technology, Wiley Online Library*, v. 60, n. 3, p. 581–594, 2009.
- NHOATTO, F. et al. Prospecção de ferramentas para Gestão De Pesquisa e desenvolvimento em Instituições de Ciência e Tecnologia. *Revista E-Tech: Tecnologias para Competitividade Industrial-ISSN-1983-1838*, v. 15, n. 1, 2021.
- NORUZI, A. Google scholar: The new generation of citation indexes. Walter de Gruyter GmbH & Co. KG, 2005.
- NUGROHO, R. et al. A survey of recent methods on deriving topics from Twitter: algorithm to evaluation. *Knowledge and information systems*, v. 62, n. 7, p. 2485-2519, 2020.
- PARRA, M. R.; COUTINHO, R. X.; PESSANO, E. F. C. Um breve olhar sobre a cienciometria: origem, evolução, tendências e sua contribuição para o ensino de ciências. *Revista Contexto & Educação*, v. 34, n. 107, p. 126–141, 2019.
- SANTOS, R. N. M. d.; KOBASHI, N. Y. *Bibliometria, cientometria, infometria: conceitos e aplicações. Tendências da Pesquisa Brasileira em Ciência da Informação*, 2009.
- SILVA, J. A. d.; BIANCHI, M. d. L. P. *Cientometria: a métrica da ciência*. Paidéia (Ribeirão Preto), SciELO Brasil, v. 11, n. 21, p. 5–10, 2001.
- RAMOS, Juan et al. Using tf-idf to determine word relevance in document queries. In: *Proceedings of the first instructional conference on machine learning*. p. 29-48, 2003.
- REZENDE, Solange Oliveira. *Sistemas inteligentes: fundamentos e aplicações*. Editora Manole Ltda, 2003.
- SPINAK, E. Indicadores cientométricos. *Ciência da informação, SciELO Brasil*, v. 27, n. 2, p. nd–nd, 1998.
- VAUGHAN, L.; SHAW, D. A new look at evidence of scholarly citation in citation indexes and from web sources. *Scientometrics, Springer*, v. 74, n. 2, p. 317–330, 2008.

**Adrielson Ferreira Justino**[adrielferreira28@gmail.com](mailto:adrielferreira28@gmail.com)ORCID: <https://orcid.org/0000-0002-0760-1947>

Universidade Federal do Oeste do Pará

Graduando do curso de Ciência da Computação pelo Instituto de Engenharia e Geociências da Universidade Federal do Oeste do Pará (UFOPA). Bolsista do Programa de Educação Tutorial (PET) do Ministério da Educação (MEC). Atualmente é membro do Grupo de Estudo e Pesquisa do Laboratório de Computação Aplicada (LACA-UFOPA). Possui experiência nas áreas de Análise de Mídias Sociais e Mineração de Textos.

**Fabiano Paulo Nhoatto**[fabiano.nhoatto@ufopa.edu.br](mailto:fabiano.nhoatto@ufopa.edu.br)ORCID: <https://orcid.org/0000-0003-1405-9081>

Universidade Federal do Oeste do Pará

Mestre em Propriedade Intelectual e Transferência de Tecnologia para Inovação (PROFNIT) pelo Instituto de Engenharia e Geociências (IEG) da Universidade Federal do Oeste do Pará (UFOPA). Graduado em 2014 em Sistemas de Informação pela Universidade Federal do Pará (UFPA). Servidor público no cargo de Analista de TI da UFOPA, atuando principalmente em desenvolvimento web e mobile, mineração de dados e ciência de dados.

**Éfren Lopes de Souza**[efren.lopez@ufopa.edu.br](mailto:efren.lopez@ufopa.edu.br)ORCID: <https://orcid.org/0000-0002-9718-9422>

Universidade Federal do Oeste do Pará

Possui Doutorado (2014) e Mestrado (2010) em Informática pela Universidade Federal do Amazonas (UFAM) e Graduação (2006) em Processamento de Dados pela Universidade Federal do Pará (UFPA). Atualmente é Professor Adjunto da Universidade Federal do Oeste do Pará (UFOPA). Tem experiência na área de Ciência da Computação, com ênfase em Redes de Sensores Sem Fio, atuando principalmente nos seguintes temas: algoritmos distribuídos, fusão de dados, localização e rastreamento de alvos.

**Fábio Manoel França Lobato**[fabio.lobato@ufopa.edu.br](mailto:fabio.lobato@ufopa.edu.br)ORCID: <https://orcid.org/0000-0002-6282-0368>

Universidade Federal do Oeste do Pará

Professor adjunto do Instituto de Engenharia e Geociências da Universidade Federal do Oeste do Pará (IEG). Pesquisador do Programa de Pós-Graduação em Propriedade Intelectual e Transferência de Tecnologia para Inovação (PROFNIT-UFOPA) e do Programa de Pós-Graduação em Engenharia da Computação e Sistemas da Universidade Estadual do Maranhão (PECS-UEMA). Colaborou com o Programa de Pós-Graduação em Pesquisa e Teoria do Comportamento da Universidade Federal do Pará (PPGPTC-UFPA). Pertence ao quadro de pesquisadores do Social CRM Research Center (SCRC), vinculado à Universidade de Leipzig, Alemanha. Possui graduação em Engenharia da Computação (2010), mestrado (2011) e doutorado (2016) pela Universidade Federal do Pará e University of Kent (período sanduíche). Realizou pesquisas de pós-doutoramento na Universidad Del País Vasco (2018), envolvendo a geração de morfologias neuronais utilizando aprendizagem profunda. Tem experiência de pesquisa, desenvolvimento tecnológico e inovação na área de Computação, com ênfase em Sistemas de Informação, atuando principalmente nos seguintes temas: Inteligência Computacional, Análise de Redes Sociais, Gestão de Relacionamento com Clientes e Tecnologias Sociais. É representante da UFOPA junto a Sociedade Brasileira de Computação (SBC). Atuou como Coordenador de Projetos de Pesquisa (2020-2021) e Coordenador de Projetos de Inovação Tecnológica (2021-2022), vinculadas à Pró-Reitoria de Pesquisa, Pós-Graduação e Inovação Tecnológica. É líder do Grupo de Estudo e Pesquisa em Computação Aplicada (GREP.ComPA-UFOPA) e coordena o Laboratório de Computação Aplicada. É como editor associado ao periódico Electronic Markets e atuou como revisor em mais de 10 periódicos e conferências indexadas pelo Qualis da CAPES.