**Trabalho de Conclusão de Curso**

PÓS-GRADUAÇÃO EM CIÊNCIA DE DADOS E INTELIGÊNCIA ARTIFICIAL

**ALUNO: Faustino Firmino de Carvalho Junior**

**ORIENTADOR: Nome do Orientador**



Sumario

´

[**1. RESUMO 2**](#_Toc83580266)

[**2. INTRODUÇÃO 3**](#_Toc83580267)

[**3. TRABALHOS RELACIONAIS 4**](#_Toc83580268)

[**4. METODOLOGIA 5**](#_Toc83580269)

[**5. RESULTADOS 6**](#_Toc83580270)

[**6. DISCUSSÃO 7**](#_Toc83580271)

[**7. CONCLUSÃO E TRABALHOS FUTUROS 8**](#_Toc83580272)

**ARTIGO CIENTÍFICO**

# RESUMO

Apresentar uma síntese do trabalho com uma breve contextualização, incluindo objetivos e resultados ( o resumo deve ter 150 a 500 palavras).

# INTRODUÇÃO

O Brasil está listado entre os países que mais sofrem com as atividades associadas a queimada, realizadas tanto em áreas de pastagens como floretais. As causas da queima estão relacionadas a eventos naturais e, pricipalmente, pela ação antropogênica. Os impactos humano-ambientais assumem um teor delicado de preocupação que demanda ao máximo pesquisas científicas, produtos, tecnologias de informática, debates entre governantes e sociedade cívil afim de combater tal problemática.

A preocupação com a quantidade de áreas dos biomas brasileiros destruída pelo fogo não está ligadada a fatos atuais. Há décadas comunidades científicas mundial e local vem comprovando que milhares de quilômetros da vegetação tupniquim está sendo devastada, sem precedentes, devido incêdicios descontralodados (SETZER, FERREIRA. et al, 2021). De acordo com levantamento feito pelo Projeto MapBiomas[[1]](#footnote-2), entre 1985 e 2020, o Brasil teve quase 20% do seu território queimado, sendo que um pouco mais da métade do espaço afetado foi de vegetação nativa, principalmente em estados do Centro-Oeste e Norte do País (Neumam, CNN BRASIL, 2020).

São tantos os motivos responsáveis pela queimadas, contudo eles são classificados em dois tipos: naturais e artificial (causado pelo homem). Apesar de eventos da natureza – descarga elétrica por exemplo – iniciarem um incêndio, as atos humanos ligados à monetização da floresta associado o *agrobusiness* são os principais resposáveis pela grande maioria das queimadas. Ao tratar sobre a ação antrópica no meio ambiente, Barbosa (BRASIL DE FATO, 2020) afirma que “[Em geral, esse ato ilegal responde a uma dessas três intenções: renovação de pasto; para desmatamento; ou para terminar de desmatar”.](http://www.ibama.gov.br/prevfogo)

Diante do desafio, iniciativas como o Sistema Nacional de Prevenção e Combate aos Incêndios Floretais - PrevFogo[[2]](#footnote-3) e o Programa Queimadas do Instituto Nacional de Pesquisas Espaciais (Inpe)[[3]](#footnote-4) surgiram para realizar detecção diária, quase imediata de focos de incêndios em imagens de satélites e combater a queima das florestais em todo o território nacional. Tais projetos trazem consigo pesquisas, artefatos, desenvolvimento e inovação tecnológica para coletar, processar, disponibilizar e prever riscos de fogo.

Além de trazer consequências a fauna, flora, emitir gases poluentes na atmosfera e elevar os casos relacionados a doenças respiratórias (SBMFC, 2020), as queimadas *podem* afetar negativamente no desenvolvimento das necessidades humanas básicas, acesso ao conhecimento e inclusão social da população afetada. A situação se agrava em municípios do bioma amazônico. Castigados pelo desmatamento e fogo descontrolado, eles pussem, comparado com a média nacional, pior qualidade de vida de acordo com pesquisa liderada pelo Imazon[[4]](#footnote-5) (IMAZON, 2021).

Um dos principais índice utilizado para avaliar o progresso social e ambiental das nações é o Índice de Progresso Social (IPS) mantido pela organização Social Progress Imperative (SPI). Apoiado por estudiosos e especialistas mudiais em política pública, sua estrutura é formada 143 indicadores em 12 componentes e 3 dimensões[[5]](#footnote-6) sem qualquer relação com aqueles relacionados ao desenvovimento ecônomico como Produto interno Bruto (PIB) e o Índice de desenvolvimento Humano (IDH).

Em Terra Tupiniquim, liderado pelo Imazon, nasce o IPS Amazônia em 2014 como proposta original para apreciar a performance social e ambiental nos 772 municípios que fazem parte da Amazônia Legal. Dessa forma, através dos mesmos métodos estatísticos do IPS global, pode-se assumir os indicadores que espelham a realidade social do território da região.

Tendo em vista essa temática, o objetivo deste trabalho visa demonstrar a aplicabilidade da Inteligência Artificial, através de modelos de *Machine Learning e Deep Learning,* para identificar correlações entre as regiões afetadas pela queimada e os indicadores sociais, auxiliar estados e municípios na tomada decisões eficazes

# **TRABALHOS RELACIONAIS**

Apresentar um breve resumo sobre o estado da arte comentando e comparando trabalhos relacionados ao descrito.

# **METODOLOGIA**

Nessa sessão será apresentada, sistematicamente, a estrutura do projeto bem como os passos utilizados para coletar e preparar os dados.

**4.1. Coleta dos dados**

Os dados utilizados nesse projeto foram coletados de duas fontes distintas: Programa Queimadas do Instituto Nacial de Pesquisas Espaciais (INPE) e IPS Amazônia. Ambos oferecemacesso aberto pela internet na qual pode ser feito o *download* das informações em formato CSV e XLXS respectivamente.

Fatos referentes a focos de calor e risco fogo precisam ser exportados pelo Sistema de Monitoramento BDQueimadas[[6]](#footnote-7). Apesar de ser possível aplicar filtros mais abrangentes, para o interesse desse trabalho, competirá somente os dados relativos aos estados e municípios brasileiros próprios do bioma amazônico, entre 01/01/2012 e 31/12/2021. O Apêndice – A, figura 1, mostra as colunas e as primeiras linhas contidas no arquivo CSV.

Já o conteúdo pertecentes ao progresso social pode ser obtido pelo próprio site do IPS Amazônia no formato de planilha eletrônica. Nela as dimensões, componentes e indicadores, até a conclusão dessa obra, foram avaliados para os anos 2014, 2018, 20121 e contempla os 772 municípios da Amazônia Legal.

**4.2.** **Preparação e transformação dos dados**

**4.2.1. Planilha Progresso Social**

Os dados cedidos pelo IPS Amazônia, estão em um arquivo na extensão XLSX (planilha Excel) no qual separa em abas o resultados coletados para cada ano. Visando facilitar as etapas seguintes do pré-processamento, fez-se necessário converter a planilha eletrônica para o formato CSV.

Para tanto, foram utilizados métodos da ferramenta Pandas para carregar o arquivo original, adicionar uma coluna com o ano, corrigir nomes das colunas que estavam com espaços em branco no final e por último salvar o produto final em novos arquivos, uma para cada ano.

**4.2.2.** **Limpeza dos dados**

No estudo preliminar dos dados, identificou-se inconsistências, discrepâncias e ausência nos valores em ambas fontes de dados, também chamadas de *datasets*[[7]](#footnote-8). Por exemplo, foram encontrados registros contendo dias sem chuva e risco fogo com valores negativos e deveras elevado o que leva a crer em problemas na coleta feita pelos sensores de monitoramento orbitais. A estratégia adotada tanto para os *outliers* (dados distantes da normalidade) quanto campos vazios foi preenchê-los com a correspondência mais genérica. Pegando o caso anterior, as *features*[[8]](#footnote-9) diasemchuva e riscofogo que sinalizavam valor mínimo igual a -999 foram ajustadas para zero.

Dois campos importantes que precisaram sofrer transformações para estarem aderentes em ambos conjuntos de dados são o estado e município. No primeiro caso, em hotspot\_df, o nome está por extenso e no spi\_df estava abreviado para sua sigla. Nesse caso, optou-se por manter o nome ao invés do acrônimo. Já no segundo cenário foi necessário, também deixar a nomenclatura dos municípios em spi\_df igual ao do hotspot\_df, ou seja caixa alta e sem acentuação ou qualquer outro caracter especial.

Para o estudo em questão, a data exata que os satélites capturaram os focos de incêndios não é muito importante. O que mais interessa desse atributo é o ano, pois é a partir dele que serão calculadas as médias anuais dos principais fatores responsáveis pela queimadas em cada município no dataset hotspot\_df. Por tanto, a partir da data, foi extraído o ano e adicionado como nova feature no cojuento de dados.

Outro ponto que precisou ser padronizado foi o nome de alguns municípios, pois estavam com escritas diferentes. Por exemplo, São Valério, Poxoréu, Eldoradoo dos Carajás e Rio Branco (Acre) estavam definidos como São Valério da Natividade, Poxoréo, Eldorado do Carajás e Rio Branco respectivamente. Como critério utilizado, fez-se uma pesquisa sobre os nomes corretos e ajustou-se aqueles com a escrita errada.

Constatou-se também que o dataset hotspot\_df tinha dados referentes a todos os 217 municípios maranhense, contudo, apenas 181 são contabilizados como pertencente da Amazônia Legal. Ou seja 36 cidades deveriam ser desconsiderada na ánalise e, portanto, todas as ocorrências presente no conjunto de dados foram removidas.

Um tratamento feito foi quanto aos anos. Para deixar o anos equiparados, ou seja 2014, 2018 e 2021, no conjunto de dados hotspot\_df foi ajustados os dados coletados para um dos três anos em questão. Por exemplo: 2012 e 2013 passaram a ser dados relativos a 2014, 2016, 2017 foram atualizados para 2018 e assim por diante. Ou seja para refletir no levantamento dos dados sociais de um determinado ano levou em consideração o ano de lavantamento e os anos anteriores imediatamente superior ao levantamento social anterior.

Também foi adicionada uma nova feature no dataset hotspot\_df chamada *fococalor[[9]](#footnote-10)* para totalizar todas as vezes que o monitoramento espacial captou um ponto em uma região. O somatório levou em consideração o agrupamento do ano, estado e município. Dessa forma pode-se avaliar as regiões que mais sofreram com possíveis queimadas e incêndio florestais.

**4.2.3. Geração dos dados**

Após preparar os dados, a etapa seguinte foi selecionar as features utilizadas na ánalise exploratória dos dados e modelos preditivos de machine learning.

No dataset relacionado a queimada, as colunas mantidas foram *estado, municipio diasemchuva, precipatacao riscofogo, frp e o ano*.

Já no conjunto de dados ligado aoprogresso social, foram mantidas as features ligadas as 3 dimensões e as 12 dimensões.

Por fim após de juntar os dois datasets levando em consideração o ano, estado e município, gerou-se um arquivo csv.

Descrever a metodologia usada para resolver o problema sob análise, incluindo arquitetura desenvolvida, recursos e materiais utilizados. Descrever os tipos de experimentos feitos sempre visando um detalhamento que viabilize a reprodução desse estudo. Aplicar os métodos e técnicas vistos ao longo do curso.

# RESULTADOS

Apresentar os resultados de forma clara, adequada e coerente, procurando explorar os recursos e técnicas de comunicação e visualização vistos no curso.

# DISCUSSÃO

Discutir os resultados do estudo, questionando sua significância, relevância e limitações.

# CONCLUSÃO E TRABALHOS FUTUROS

Apresentar as considerações finais sobre o trabalho e melhorias que poderiam ser realizadas como passos futuros de desenvolvimento.

**APÊNDICE A – DADOS COLETADOS**

Apresentar as considerações finais sobre o trabalho e melhorias que poderiam ser realizadas como passos futuros de desenvolvimento.

|  |
| --- |
| Figura 1 – Primeiras 5 linhas do conjunto de dados queimadas |
|  |
| Fonte: Elaborada pelo autor. |

Legenda das colunas:

1. datahora:
2. satelite
3. pais
4. estado
5. municipio
6. bioma
7. diasemchuva
8. precipitacao
9. riscofogo
10. latitude
11. longitude
12. frp

|  |
| --- |
| Figura 2 – Primeiras 5 linhas do conjunto de dados SPI |
|  |
| Fonte: Elaborada pelo autor. |

**REFERÊNCIAS**

BARBOSA, Catarina. **Quais são os tipos de queimadas ilegais mais utilizadas pelo agronegócio na Amazônia**. Brasil de Fato, Belém (PA), 19 de ago. de 2020. Disponível em: https://www.brasildefato.com.br/2020/08/19/quais-sao-os-tipos-de-queimadas-ilegais-utilizadas-pelo-agronegocio-na-amazonia. Acesso em: 18 de mai. de 2022.

BORMA, Laura de Simone; NOBRE, Carlos Afonso. **Seca na Amazônia: causas e consequências**. São Paulo: Oficina de Texto, 2013.

GÉRON, [Aurelie](https://www.google.com/search?q=aurelien+geron&stick=H4sIAAAAAAAAAOPgE-LVT9c3NEw2N0nLic_NU-IBc43jLZMNsky0ZLKTrfST8vOz9cuLMktKUvPiy_OLsq0SS0sy8osWsfIllhal5mSm5imkpxbl5-1gZdzFzsTBAAAJc3qfVgAAAA&sa=X&ved=2ahUKEwiu6IGCxOf3AhVoBbkGHaCWAZYQmxMoAXoECEQQAw)n. **Mãos à Obra: Aprendizado de Máquina com Scikit-Learn & TensorFlow**. Rio de Janeiro: Alta Books, 2019.

IMAZON - Instituto do Homem e Meio Ambiente da Amazônia. **Mnunicípios que mais desmatam têm pior qualidade de vida na Amazônia**. IMAZON, Belém, 06 de dez. de 2021. Disponível em: [https://imazon.org.br/imprensa/municipios-que-mais-desmatam-tem-pior-qualidade-de-vida-na-amazonia](https://imazon.org.br/imprensa/municipios-que-mais-desmatam-tem-pior-qualidade-de-vida-na-amazonia/). Acesso em: 22 de mai. de 2022.

KNAFLIC, Cole Nussbaumer. **Storytelling com dados: Um guia sobre visualização de dados para profissionais de negócios**. Rio de Janeiro: Alta Books, 2018.

MCKINNEY, Wes. **Python Para Análise de Dados: Tratamento de Dados com Pandas, NumPy e Ipython**. São Paulo: Novatec Editora Ltda., 2018.

NEUMAM, **Camila. Em 36 anos, Brasil teve quase 20% de seu território queimado, diz levantamento**. CNN Brasil, São Paulo, 03 de set. de 2020. Disponível em: https://www.cnnbrasil.com.br/nacional/em-36-anos-brasil-teve-quase-20-de-seu-territorio-queimado-diz-levantamento. Acesso em: 18 de mai. de 2022.

SBMFC - Sociedade Brasileira de Medicina de Família e Comunidade. **Queimadas e doenças respiratórias**. SBMFC, Rio de Janeiro, 10 de set. de 2020. Disponível em: https://www.sbmfc.org.br/noticias/queimadas-e-doencas-respiratorias. Acesso em: 18 de mai. de 2022.

SETZER, Alberto S; FERREIRA Nelson J. et al. **Queimadas e incêncios florestais: mediante monitoramento orbital**. São Paulo: Oficina de Texto, 2021.

1. https://mapbiomas.org [↑](#footnote-ref-2)
2. https://www.ibama.gov.br/prevfogo [↑](#footnote-ref-3)
3. https://queimadas.dgi.inpe.br/queimadas/portal [↑](#footnote-ref-4)
4. Instituto do Homem e Meio Ambiente da Amazônia [↑](#footnote-ref-5)
5. Anexo A [↑](#footnote-ref-6)
6. https://queimadas.dgi.inpe.br/queimadas/bdqueimadas#exportar-dados [↑](#footnote-ref-7)
7. *Datasets* refere-se ao conjunto de dados utilizados em projetos de *Data Science* e treinamento de algoritimos de *machine learning*. [↑](#footnote-ref-8)
8. *Feature* refere-se a uma coluna no conjunto de dados. Ela é usados como de entrada para um modelo preditivo. [↑](#footnote-ref-9)
9. Um foco de calor representa uma região com temperatura acima de 47°C e não equivale ao conceito de queimada ou incêndio florestal. Um incêndio ou uma ocorrência podem gerar um ou vários focos de calor, dependendo da extensão da linha de fogo. [↑](#footnote-ref-10)