

UNIVERSIDADE ESTADUAL DO MARANHÃO
CENTRO DE CIÊNCIAS E TECNOLOGIAS
CURSO DE FORMAÇÃO DE OFICIAIS BOMBEIRO MILITAR

ANTONIO MATHEUS VILAR GUIMARÃES BARBOSA

INCÊNDIOS EM VEGETAÇÃO NO MUNICÍPIO DE MIRADOR-MA: O estudo de
modelos preditivos baseados em dados atmosféricos.

São Luís

2021

ANTONIO MATHEUS VILAR GUIMARÃES BARBOSA

INCÊNDIOS EM VEGETAÇÃO NO MUNICÍPIO DE MIRADOR-MA: O estudo de modelos preditivos baseados em dados atmosféricos.

Trabalho de Conclusão de Curso apresentado à disciplina Monografia do Curso de Formação de Oficiais Bombeiro Militar da Universidade Estadual do Maranhão como pré-requisito para aprovação na disciplina.

Orientador: Prof. Dr. Mauro Sérgio Silva Pinto

São Luís

2021

Barbosa, Antonio Matheus Vilar Guimarães.

Incêndios em vegetação no município de Mirador - MA: o estudo de modelos preditivos baseados em dados atmosféricos / Antonio Matheus Vilar Guimarães Barbosa. – São Luís, 2021.

58f

Monografia (Graduação) – Curso de Formação de Oficiais BM-MA, Universidade Estadual do Maranhão, 2021.

Orientador: Prof. Dr. Mauro Sérgio Silva Pinto.

ANTONIO MATHEUS VILAR GUIMARÃES BARBOSA

INCÊNDIOS EM VEGETAÇÃO NO MUNICÍPIO DE MIRADOR-MA: O estudo de modelos preditivos baseados em dados atmosféricos.

Monografia apresentada ao Curso de Formação de Oficiais da Universidade Estadual do Maranhão, como requisito parcial para a obtenção de título de Bacharel em Segurança Pública e do Trabalho pela a Universidade Estadual do Maranhão.


Aprovado em: 28 / 07 / 2021

BANCA EXAMINADORA



Prof. Dr. Mauro Sérgio Silva Pinto (Orientador)

Doutor em Engenharia Elétrica
Universidade Federal do Maranhão



Ricardo Yvan de La Cruz Cueva
Matrícula: 00844834-00

Prof. Dr. Ricardo Yvan de La Cruz Cueva

Doutor em Geofísica Espacial
Instituto Nacional de Pesquisas Espaciais



Cap PM Cassio Koitsi Hashizume da Luz

Bacharel em Ciências Policiais de Segurança e Ordem Pública
Polícia Militar de São Paulo

AGRADECIMENTOS

Eu, agradeço primeiramente a Deus, por ter me proporcionado toda vida que vivi, todo conhecimento que obtive e me dar oportunidades para conquistar as vitórias e me apoiar nas minhas derrotas que tive até hoje.

Agradeço a minha mãe, a senhora Lucinda de Jesus Vilar Guimarães, pois além de ter me dado a vida, ajudou-me em toda minha jornada, estando do meu lado nos dias luta e tendo orgulho de mim nos dias de glórias, me auxiliando em tudo que precisava desde do início até hoje na minha vida.

Agradeço a meus avós, a senhora Maria José Rocha Vilar e o senhor Ernildo França Ferreira, estando ao lado de minha mãe, dando-me o necessário para que eu continuasse minha jornada.

“A imaginação é mais importante que o conhecimento, porque o conhecimento é limitado, ao passo que a imaginação abrange o mundo inteiro.”

Albert Einstein

RESUMO

Neste trabalho foi realizado um estudo com técnicas de ciência de dados com fins de construir modelos para a previsão de focos de incêndio no Município de Mirador no Maranhão. O objetivo deste trabalho foi obter, tratar e analisar os dados de variáveis atmosféricas associados com o de queimadas e aplicar os métodos de classificação com inteligência artificial para construção dos modelos de previsão. A problemática presente nesta pesquisa era responder ao seguinte questionamento: é possível utilizar modelos preditivos com inteligência artificial para previsão de focos de incêndio em vegetação com base em parâmetros históricos de variáveis atmosféricas? Para se obter uma resposta, primeiramente foram obtidos dados de variáveis atmosféricas a partir do Instituto nacional de Meteorologia e os dados de incêndios em vegetação no Instituto de pesquisas espaciais e posteriormente, foi feita uma análise exploratória destes com fins de buscar correlações matemáticas e as similaridades entre elas, para se entender a dinâmica relacional das variáveis em questão, em seguida foi aplicado alguns métodos de classificação de inteligência artificial com fins de construir os modelos preditos, das quais foram as redes neurais artificiais, k-vizinhos mais próximos e floresta de decisão aleatória, após construção dos modelos, foram testados e avaliados e chegou-se à conclusão de que eles tiveram resultado satisfatórios, apresentando acurácias acima de 80%, diante disso, este trabalho pode contribuir para implementação de um sistema de previsão de incêndios, para assim melhorar-se o tempo resposta do serviço do corpo de bombeiros militar.

Palavras Chaves: Inteligência Artificial. Modelos Preditivos. Incêndios em vegetação. Variáveis Atmosféricas

ABSTRACT

In this work, a study was carried out with data science techniques in order to build models for the prediction of fire outbreaks in the Municipality of Mirador in Maranhão. The objective of the work was to obtain, treat and analyze the data of atmospheric variables associated with the fires and apply the classification methods with artificial intelligence to build the prediction models. The problem presented in this research was to answer the following question: is it possible to use predictive models of artificial intelligence to predict fires in vegetation based on variable parameters of atmospheric variables? To obtain an answer, an exploratory analysis of the data was necessary in order to search for mathematical correlations and the similarities between them, to understand how they influence each other, after which some classification methods were used in order to build the predicted models, of which they were as artificial neural networks, k-nearest neighbors and random decision forest, after construction of the models, were tested and elaborated and it was reached the conclusion that they had satisfactory results, increased an accuracy above 80%, this work can be implemented for the implementation of a fire forecasting system, in order to improve the response time of the service of the military fire department.

Keywords: Artificial Intelligence. Predictive Models. Fires in vegetation. Atmospheric Variables

LISTA DE FIGURAS

Figura 1 - Incêndio Florestal em superfície	16
Figura 2 - Incêndio Florestal Subterrâneo	17
Figura 3 -Incêndio Florestal de Copas.....	17
Figura 4 - Dados históricos de focos de queimadas no Brasil de 1998 até 14/03/2021	18
Figura 5 - Comparativo dos dados de queimadas no Maranhão com os valores máximos, médios e mínimos, no período de 1998 até 15/03/2021.	19
Figura 6 - dados históricos do total de focos de queimadas no estado do Maranhão detectados por satélite, no período de 1998 até 15/03/2021	20
Figura 7 - Termômetro de máxima e mínima.....	24
Figura 8 -Relação entre chuvas e focos de queimadas.....	25
Figura 9 - Imagem Via Satélite do Município de Mirador mostrando os focos de incêndios vistos pelo satélite AQUA_M-T entre 1 de janeiro de 2020 até 31 de dezembro de 2020	31
Figura 10 - Imagem de queimada captada pelo satélite TERRA, através do sensor MODIS, o mesmo sensor presente no satélite AQUA.....	36
Figura 11 - Heatmap da correlação linear	36
Figura 12 - Dendograma	38
Figura 13 - Movimentação dos centros	40
Figura 14 - Análise de componentes principais.....	41
Figura 15 - Neurônio Artificial	43
Figura 16 - Rede Neural Artificial.....	43
Figura 17 - Exemplo de k-Vizinhos mais próximos.....	44
Figura 18 - Árvore de decisão	44

LISTA DE TABELAS

Tabela 1 - Matriz de correlação linear.....	36
Tabela 2 - Dados padronizados	37
Tabela 3 - Tabela de distâncias	38
Tabela 4 - Agrupamentos formados pelo método k-means	40
Tabela 5 - Matriz Confusão do Modelo de rede neural artificial	48
Tabela 6 - Resultados do Método de rede neural artificial.....	48
Tabela 7 - Matriz Confusão do Modelo de K-Vizinhos mais próximos	48
Tabela 8 - Resultados do Método de k-Vizinhos mais próximos.....	49
Tabela 9 - Matriz Confusão do Modelo de Floresta Aleatória	49
Tabela 10 - Resultados do Floresta Aleatória	49
Tabela 11 - Comparação entre os modelos.....	50

LISTA DE QUADROS

Quadro 1 - Dados de focos de incêndios	31
Quadro 2 - Dados Atmosféricos nulos.....	32
Quadro 3 - Banco de dados para o modelo.....	33
Quadro 4 - Softwares para Análise Multivariada	34
Quadro 5 - Classificação da correlação de Pearson	35
Quadro 6 - Softwares utilizados.....	42
Quadro 7 - Métodos de treinamento e testagem	45
Quadro 8 - Matriz Confusão	46

SUMÁRIO

1 INTRODUÇÃO	14
2 INCÊNDIOS FLORESTAIS	16
2.1 Incêndios Florestais de Superfície	16
2.2 Incêndio Subterrâneo	17
2.3 Incêndio de copas	17
2.4 Incêndios florestais no Brasil	18
2.5 Incêndios florestais no Maranhão	19
2.5.1 Incêndio em vegetação no município de Mirador	20
2.6 Estratégia de Combates a Incêndios Florestais	21
2.7 Variáveis meteorológicas	22
2.7.1 Variáveis atmosféricas e influências em incêndios florestais	22
2.7.2 Evapotranspiração	23
2.7.3 Temperatura do ar	23
2.7.4 Precipitação pluvial	24
2.7.5 Umidade relativa do ar	26
2.8 Inteligência Artificial	26
3 METODOLOGIA	29
3.1 Quanto à natureza	29
3.2 Quanto aos objetivos	29
3.3 Quanto aos procedimentos	29
3.4 Quanto à abordagem do problema	29
3.5 Local da pesquisa	30
3.6 Materiais e métodos	30
3.6.1 Obtenção de dados	30
3.6.2 Análise Multivariada	34
3.6.2.1 Matriz de correlação	34
3.6.2.2 Dendograma	37
3.6.2.3 Método K-Means	39
3.6.2.4 Método dos componentes principais	40
3.6.3 Modelagem	42
3.6.3.1 Redes Neurais artificiais	42
3.6.3.2 K-vizinhos mais próximos	43
3.6.3.3 Floresta de decisão aleatória	44
3.6.4 Avaliação dos modelos	45
3.6.4.1 Treinamento e testagem	45

3.6.4.2 Matriz confusão	46
3.6.4.3 Métricas	47
4 RESULTADOS E DISCUSSÕES.....	48
5 CONSIDERAÇÕES FINAIS.....	51
REFERÊNCIAS	53
APÊNDICE A - DECLARAÇÃO DE ORIGINALIDADE	58

1 INTRODUÇÃO

Os incêndios florestais certamente são uma preocupação constante, especificamente para o Corpo de Bombeiros Militar que todo o ano faz campanhas para que estes eventos sejam mitigados. Muito embora este tipo de ocorrência não apareça sistematicamente nos noticiários quando comparados a outras ocorrências como roubos e assaltos, sua ocorrência não é rara e em períodos mais secos, podem se tornar críticas. O que pode ocasionar de mais grave é a sua evolução para um incêndio residencial ou mesmo de grandes proporções com vítimas fatais e enormes prejuízos financeiros e impactos ambientais. Durante todo o ano, nos meses mais críticos, o Corpo de Bombeiros sempre faz um alerta de conscientização para que o risco seja diminuído em várias partes do estado.

Um dos principais problemas para o combate de queimadas é o deslocamento das guarnições de combate para o local afetado, isso aumenta muito o tempo resposta para o atendimento de tais ocorrências, então redução desse tempo é relevante para evitar o agravamento desses desastres e pode ser feita de algumas formas, uma delas é alocar equipes de combate a incêndios em locais onde ocorrem as suas maiores frequências e com isso descobrir os períodos em que a frequência dos incêndios florestais é mais intensa, ou seja, conhecer, por exemplo os dias (ou horas) em que as probabilidades de ocorrer sejam mais altas.

Com o avanço de novas tecnologias, hoje é possível mapear em tempo real os focos de incêndios florestais pelo site do Instituto Nacional de Pesquisas espaciais, uma vez que através de imagens de satélites esse foco é verificado e, assim, os órgãos competentes podem agir a fim de prevenir um mal maior. O grande problema é que ele verifica esses focos quando já estão acontecendo, então, até acionar as equipes e direcioná-las ao local, grande parte da vegetação já foi devastada. Dessa forma, para poder melhorar esse combate da vegetação, a melhor forma seria prevendo os locais onde tem maior probabilidade de acontecer.

No Maranhão, a maior parte dos incêndios ocorrem no cerrado, sendo que esse bioma é castigado por fortes estiagens em boa parte do ano, e a vegetação é muito afetada pelo clima. Assim, quando está em um período sem chover e umidade baixa, fica muito propenso a focos de incêndios se alastrarem. Destaca-se que maior parte das queimadas que ocorrem no cerrado tem origem antrópicas (causados por intervenção humana), mas os parâmetros meteorológicos são elementos bem

relevantes quando se está falando de fogo no cerrado, pois eles deixam a vegetação muito propícia a fogo e assim espalhar com facilidade um incêndio florestal.

O município de Mirador no Maranhão é um dos lugares no estado mais afetados por incêndios em todo o estado, sendo que sua vegetação é inteiramente cerrado e sempre apresenta altos número de focos, ao passo que nos últimos anos sempre ocupa o primeiro lugar nessa classificação.

As tecnologias presentes na ciência de dados e inteligências artificiais estão cada vez mais presentes no mundo, elas auxiliam humanos em decisões, analisando dados e observando padrões onde humanos são limitados, existem muitas pesquisas em diversas áreas se utilizando dessas ferramentas, seja para prever padrões de compras de clientes, no reconhecimento de imagens, automação de tarefas, no caso de incêndios florestais, podem ser utilizados usados os parâmetros meteorológicos, como umidade, temperatura, radiação, precipitação, associados com as ocorrências de incêndios, tudo isso dentro de um banco de dados e assim com o auxílio de um algoritmo prevê ocorrências de incêndios com uma certa precisão a partir de dados passados.

Neste trabalho o objetivo é um estudo de modelos preditivos utilizando métodos e técnicas da ciência de dados, aplicando isso para a obtenção, tratamento, análise de parâmetros meteorológicos e aplicação de técnicas de classificação para a previsão focos de incêndio no município de Mirador, no estado do Maranhão, assim poderá se ter uma ferramenta que se auxilia a diminuir o tempo resposta de ocorrências de queimadas, diminuindo assim os impactos ambientais e econômicos acometidos por esses desastres.

2 INCÊNDIOS FLORESTAIS

Incêndio florestal, popularmente chamado também de queimada, é o descontrole do fogo que incide sobre qualquer tipo de vegetação, e pode ser causado tanto por ação antrópica quanto por causas naturais (GOIÁS, 2017).

Esses incêndios causam inúmeros prejuízos econômicos, paisagísticos e ecológicos, podendo ocorrer em unidades de conservação, áreas de preservação, fazendas, margens de estradas, proximidades de aglomerados urbanos, entre outras localidades. Queimadas perto de aglomerações urbanas são uma grande preocupação, uma vez que têm riscos de causar danos a bens, ao passo que a fumaça, por ser rica em substâncias nocivas, pode causar variados problemas a saúde humana, (RIBEIRO; ASSUNÇÃO, 2002).

Um exemplo é o Hospital Infantil Cosme e Damião, que atendeu a todo o estado de Rondônia e relatou que foram realizados 120 atendimentos de crianças com problemas respiratórios entre 1 a 10 de agosto de 2019, e 380, entre 11 a 20 de agosto de 2019, segundo informado pelo diretor-adjunto Daniel Pires de Carvalho, e o motivo foi de a fumaça de incêndios florestais próximos de centros urbanos (TURBIANI, 2019). Incêndios Florestais podem ser classificados em 3 (três) tipos observando a variável de qual parte do combustível está se propagando o fogo: superfície, subterrâneo e de copas (GOIÁS, 2017).

2.1 Incêndios Florestais de Superfície

São aqueles em que a queima se propaga no material combustível sobre o solo, sendo material vivo morto. Pode ser considerado se for até uma altura de 1,80 do solo, ele se caracteriza por um crescimento rápido e costuma ser a causa dos outros tipos de queimadas (GOIÁS, 2017).

Figura 1 - Incêndio Florestal em superfície



FONTE: Sansuy (2021)

2.2 Incêndio Subterrâneo

São aqueles que se propagam na estrutura vegetal presente abaixo do solo, ocorrem com maior frequência em locais com maior abundância de húmus nas camadas abaixo do solo, eles se propagam com maior lentidão por causa do baixo contato com a atmosfera, causam mortes de raízes, sendo que é muito constante nos cerrados, e bem difícil de ser avistado, por causa de ser subterrâneo e ter pouca fumaça (GOIÁS, 2017).

Figura 2 - Incêndio Florestal Subterrâneo



Fonte: G1 (2019)

2.3 Incêndio de copas

É um tipo de incêndio florestal que se espalha pela copa das árvores, são as queimadas que ocorrem a mais de 1,80m de altura, é caracterizado pela grande liberação de calor, e pela dificuldade do combate, normalmente são originários de incêndios de superfície (GOIÁS, 2017).

Figura 3 -Incêndio Florestal de Copas



Fonte: Dimensão Incêndio (2019)

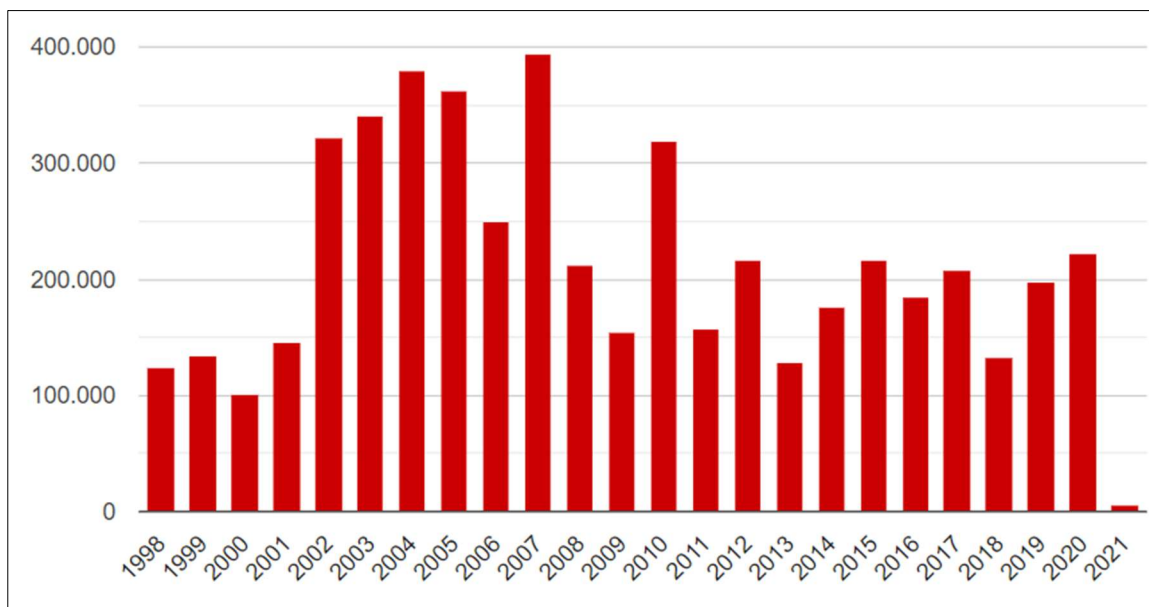
Dentro da discussão de queimadas, pode-se chegar aos quais fatores são relevantes tanto para as ocorrências quanto para a propagação destes tipos de incêndios e se faz relevante entender as variáveis meteorológicas, as variações topográficas e o tipo de vegetação (TORRES, 2011).

2.4 Incêndios florestais no Brasil

No Brasil, incêndios florestais, são um problema constante, como por exemplo no ano de 2020, durante o período que se estende do começo do ano até agosto, foram registrados 44.013 focos de queimadas e no ano anterior, foi um número um pouco menor durante este mesmo período, sendo detectados 30.900 focos (MADEIRO, 2020).

Ao longo dos anos, os dados obtidos via satélite de focos de queimadas se mantiveram altos em números, e revelam dados preocupantes, uma vez que é um problema enfrentado a muitos anos e mesmo assim continua numerosamente grande. Estas informações podem ser constatadas no gráfico abaixo:

Figura 4 - Dados históricos de focos de queimadas no Brasil de 1998 até 14/03/2021



Fonte: INPE (2021)

Assim, à medida que esses números continuam constantes ou crescem, podem haver alguns efeitos negativos à saúde pública, ocasionando sérios prejuízos econômicos. Estima-se que 2.135 pessoas foram hospitalizadas devido aos efeitos da

poluição do ar em virtude dos incêndios florestais ocorridos no bioma amazônico durante o todo o ano de 2019 e que isso levou a um gasto hospitalar aproximado de R\$ 5,64 milhões (USD 1,4 milhão). Dessa forma, espera-se ainda em 2020 que, devido ao aumento do desmatamento e da pandemia de COVID-19, aumentando muito os casos de doenças respiratórias, esses números de internações por causa dos efeitos de incêndios florestais sejam ainda maiores (SANT'ANNA; ROCHA, 2020)

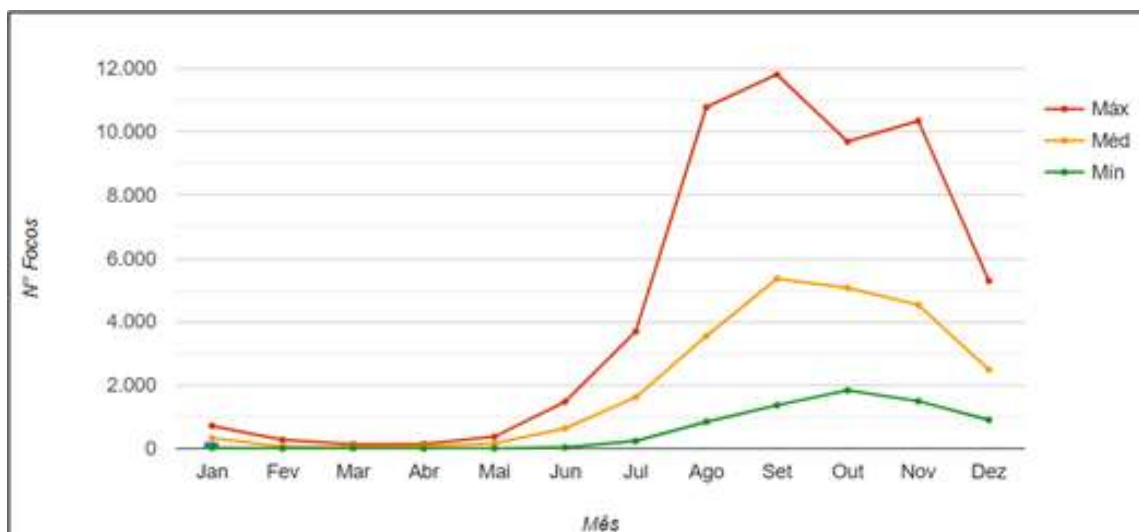
2.5 Incêndios florestais no Maranhão

No estado do Maranhão, os biomas são compostos basicamente de 64% de cerrado, 35% de floresta amazônica e 1% de caatinga, (SPINELLI-ARAÚJO et al., 2016).

Assim, sabe-se que, diferentemente do bioma amazônico, em que a principal origem dos incêndios florestais é o desmatamento, as ocorrências de incêndios no cerrado estão diretamente ligadas ao clima e a existência da agricultura que se utiliza de técnicas utilizando o fogo como ferramenta agrícola.

Destaca-se que os meses de agosto e setembro são os períodos onde se tem maiores índices de ocorrências de queimadas, por causa da escassez de chuvas, que deixam a vegetação suscetível a esses desastres, ou seja, quanto menor o índice de chuvas mensais, maior a probabilidade de incêndios florestais nesta região (SILVA JÚNIOR et al., 2018)

Figura 5 - Comparativo dos dados de queimadas no Maranhão com os valores máximos, médios e mínimos, no período de 1998 até 15/03/2021.

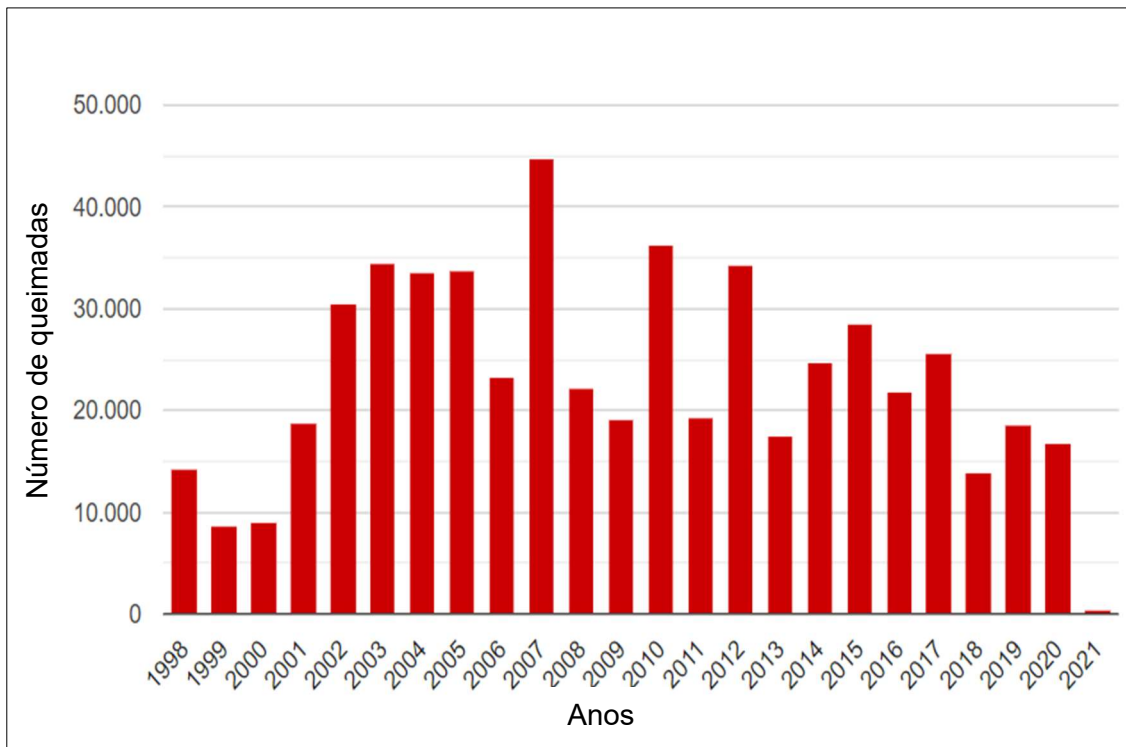


Fonte: INPE (2021)

Pode-se confirmar a partir do gráfico acima que durante o 2º semestre dos anos, o número de focos de incêndios no Maranhão é bem maior do que o 1º semestre, e isso porque coincide com a época que tem os menores índices de chuvas.

Sobre os dados históricos de incêndios florestais no Maranhão, podemos observar que seguiu o mesmo padrão do gráfico presente na Figura 04, como se pode verificar abaixo:

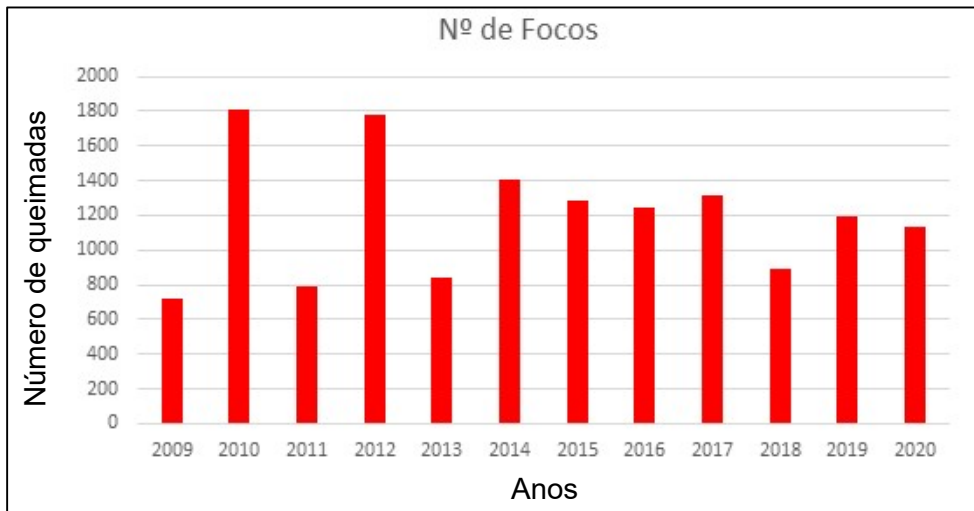
Figura 6 - dados históricos do total de focos de queimadas no estado do Maranhão detectados por satélite, no período de 1998 até 15/03/2021



Fonte: INPE (2021)

2.5.1 Incêndio em vegetação no município de Mirador

Mirador é um município a 426 quilômetros da capital São Luís, possui uma área territorial de 8521 quilômetros quadrados e tem cerca de 21015 habitantes (IBGE, 2021). O município, em média, é um dos que mais sofre de queimadas no Estado, destacando-se que no ano de 2020, foi o município que mais registrou focos de incêndio em todo Maranhão (INPE, 2021). Em 2015, o município foi classificado com a 2ª cidade do país com mais focos de incêndio em vegetação. (G1, 2021)

Gráfico 1 - Número de focos de incêndio por ano em Mirador

Fonte: Elaborado pelo autor com base nos dados do INPE

O número de focos de incêndio em Mirador vem seguindo a uma certa constância, como mostra o gráfico 1 que reapresenta os focos de incêndio em Mirador dentre os anos de 2009 a 2020, dados obtidos no Instituto Nacional de Pesquisas Espaciais.

2.6 Estratégia de Combates a Incêndios Florestais

O Maranhão atualmente passa metade do ano em estiagem e metade do ano em período de chuva, de forma que até mesmo a maior parte dos incêndios em vegetação são causados por intervenção humana, os fatores meteorológicos são uma grande influência para a propagação destes tipos de desastres, então por isso ocorre muito mais ocorrências durante o período de estiagem. (PACHECO, SILVA, 2018).

A melhor forma de combater é prevenindo e para a prevenção é necessário investimentos na área de educação ambiental, na redução de riscos de propagação de fogo, monitoramento, treinamento de pessoal, investimento em equipamentos e o desenvolvimento de um banco de dados. (RIBEIRO, 2004)

No Maranhão existe, ainda, o programa Maranhão Sem Queimadas, que é um programa destinado a realizar a fiscalização, monitoramento, conscientização da sociedade em geral, com o objetivo da prevenção de incêndios em vegetação. (MARANHÃO, 2021).

2.7 Variáveis meteorológicas

Há uma grande diferença entre clima e tempo, usualmente as pessoas acabam confundindo estes dois conceitos, mas quando está se estudando mais profundamente isso tem que ficar bem claro. O tempo é o estado momentâneo da atmosfera em um referido lugar, ao passo que o clima é a sucessão dos estados do tempo durante um prolongado período.

Esses dados podem ser obtidos através de estações meteorológicas e a organização mundial de meteorologia (WMO) convencionou este período de tempo como um período que se estende por 30 anos, segundo (REBOITA et al., 2015), e, para este trabalho, será estudado apenas parâmetros meteorológicos que se consideram influentes dentro do âmbito de queimadas.

2.7.1 Variáveis atmosféricas e influências em incêndios florestais

Dentre algumas variáveis que precisam ser apresentadas neste trabalho, aqui entra na discussão as variáveis meteorológicas, uma vez que são de suma importância quando se está no assunto dos incêndios florestais, dado que elas estão fortemente ligadas às ocorrências e propagação destes tipos de desastres, tendo em vista que podem definir a intensidade e a sua velocidade, variáveis como precipitação, temperatura, umidade relativa do ar e velocidade do vento, como disse (NUNES; SOARES; BATISTA, 2006)

O Instituto nacional de meteorologia (INMET) atualmente faz o registro destes dados através de estações meteorológicas convencionais (EMC) e estações meteorológicas automáticas (EMA), estes dados podem ser acessados no próprio site do INMET, o www.inmet.gov.br.

As estações meteorológicas convencionais são formadas por alguns sensores e esses dados são registrados por uma pessoa 3 vezes ao dia, às 9h, às 15h e às 21h. Dentre esses sensores, eles mensuram variáveis atmosféricas como umidade relativa do ar, temperatura, insolação, pressão atmosférica do ar, precipitação e várias outras (RIBEIRO, 2017).

As estações meteorológicas automáticas, como o próprio nome já diz, atuam de forma automática, elas registram os parâmetros meteorológicos e funcionam sem precisarem ser supervisionados toda hora, elas são bastante confiáveis por causa da tecnologia avançada de novos sensores e diminuem os erros humanos,

estes dados são enviados a cara hora para o INMET e este, por sua vez, disponibiliza estes dados em seu site (RIBEIRO, 2017)

2.7.2 Evapotranspiração

A evapotranspiração está entre as variáveis com bastante correlação quando se está estudando sobre incêndios em vegetação (TORRES, 2011), possuindo acentuada relevância dentro dos parâmetros influentes as queimadas, por isso cabe também uma breve exposição sobre ela.

É uma forma de descrever o fenômeno da evaporação e da transpiração, em que o primeiro, a água presente é transferida para a atmosfera proveniente de solos úmidos, em rios, mares, lagos e a transpiração das atividades fisiológicas de seres vivos (VAREJÃO-SILVA, 2013)

Esta variável é importante para se verificar a situação da vegetação na região, assim verificar o consumo de água destes seres vivos, sendo que ela é estimada a através do balanço energético solicitado pela atmosfera e a reserva de água no solo para as plantas (RIBEIRO, 2016). A importância dessa variável para esse estudo é que modifica diretamente a estrutura das plantas, quanto maior a evaporação, menos água tem esse vegetal e assim estas plantas estão mais secas e mais propícias a incêndios, (TORRES, 2011).

2.7.3 Temperatura do ar

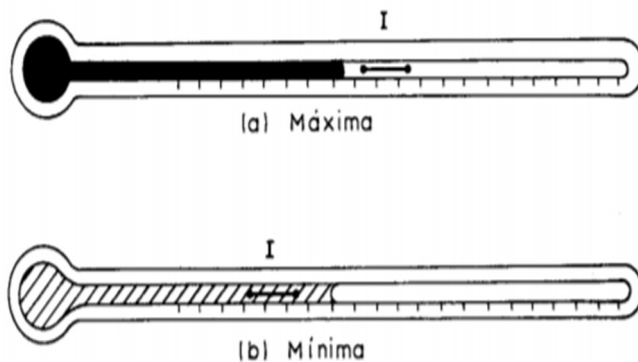
A temperatura do ar é uma variável atmosférica que apresenta pouca correlação em relação aos incêndios florestais, como demonstrado no trabalho de Pacheco e Silva (2018), foi-se apresentado que a variação da temperatura em São Luís (MA) varia muito pouco ao longo do ano.

A temperatura está diretamente relacionada ao grau de agitação das moléculas e pode ser medida no sistema internacional de medidas pela escala Kelvin(K), mas existem outras unidades de medidas também utilizadas como o Fahrenheit (°F) e o Celsius (°C), (SEARS, 2008), mas um fator muito importante para o seguinte estudo, que faz relevante a apresentação desta variável neste trabalho, é que a temperatura do ar afeta de forma significativa a fisiologia da vegetação presente no local (LIMA, RIBEIRO, 1998).

Uma coisa interessante que é notável ser apresentado é a relação entre essas variáveis, uma que vez que a umidade serve como regulador de temperatura do ambiente, então quanto menor a umidade, mais fácil de a radiação solar aumentar a temperatura do ar (TORRES, 2006)

A temperatura do ar pode ser medida através de um instrumento chamado de termômetro meteorológico, este é utilizado para medir a temperatura do ambiente, ele exibe tanto temperatura máxima quanto mínima dentro de um determinado período (TIPOS, 2018).

Figura 7 - Termômetro de máxima e mínima



Fonte: AYOADE (1996)

O termômetro de máxima é composto por uma espécie de vidro e no seu interior encontra-se mercúrio, ao passo que quando a temperatura sobe, o mercúrio se expande e quando a temperatura diminui, o mercúrio diminui seu volume, o termômetro de mínima é parecido, mas no lugar do mercúrio é composto de álcool, no qual ocorre o mesmo fenômeno físico da dilatação térmica (AYOADE, 1996)

É necessário durante as medições aferir tanto as máximas e mínimas, pois ocorre muito a variação da temperatura do ar durante a mudança de horários, e cada vez que se tem estações automáticas trabalhando a este favor, se tem uma maior possibilidade de trabalho, tanto para a amostragem de dados, quanto para difíceis localidades (STRASSBURGER et al., 2011).

2.4.7 Precipitação pluvial

A precipitação se inclui dentre os fatores importantes para as ocorrências de incêndio principalmente no bioma do cerrado, uma vez que a falta de chuvas coloca

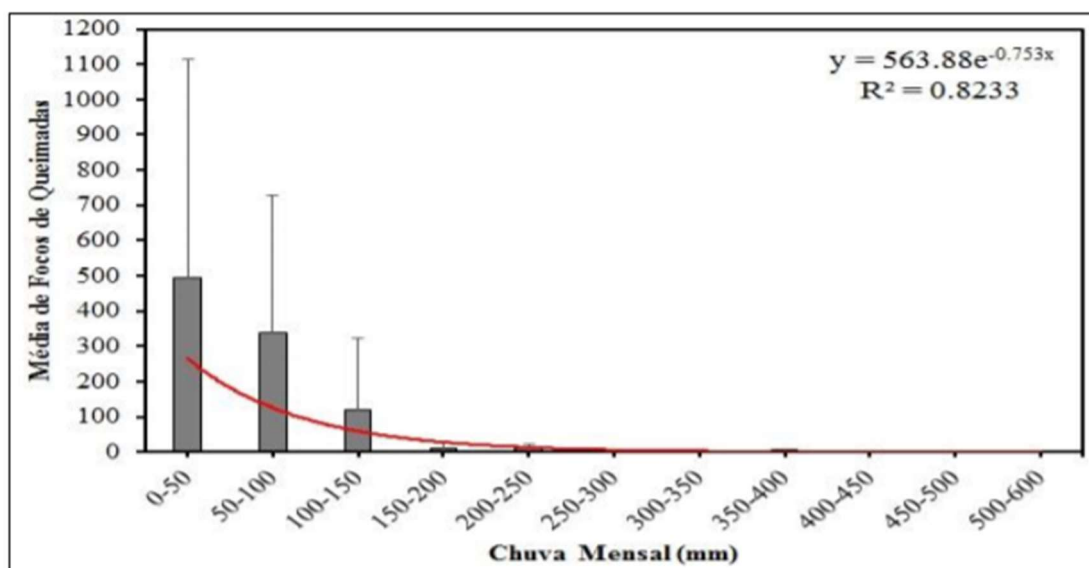
a vegetação em um estado propício a se tornar combustível do fogo (SILVA JUNIOR et al., 2018). Este parâmetro está ligado a quantidade de chuvas, em meteorologia, a precipitação se conceitua como qualquer forma de depósito, em forma líquida ou sólida, oriundo da atmosfera (AYOADE, 1996), mas a critério deste trabalho que se limita ao cerrado, será tratado apenas da precipitação pluvial.

Para se medir a precipitação pluvial é necessário o auxílio do pluviômetro, onde ele mede o volume, o peso e sua duração, tudo isso de acordo com seu alcance, após isso, é possível se tirar o índice de precipitação pluviométrica, e ela é medida em milímetros de acordo com determinado período (AYOADE, 1996).

O cerrado maranhense apresenta hoje dentre todo o estado a menor taxa de precipitação anual em comparação ao restante do estado, (SILVA et al., 2014), e ainda, não é dividido de uma forma proporcional, uma vez que o período chuvoso se concentra no primeiro semestre do ano e o outro semestre sofre com baixíssimas taxas de precipitação pluviométrica (CHUVA, 2021).

Dentro dos incêndios em vegetação no estado do Maranhão, pode-se observar que a variação do clima ao longo do ano no Maranhão, ainda mais quando o parâmetro que mais varia é a precipitação, que as queimadas estão diretamente relacionadas à falta de chuvas, de acordo com o entendimento de Silva Júnior (2016) em estudo sobre a dinâmica dos incêndios em vegetação na baixada maranhense.

Figura 8-Relação entre chuvas e focos de queimadas



Fonte: Junior, (2016)

2.7.5 Umidade relativa do ar

Dentre todas as variáveis atmosféricas, uma das que apresenta maior correlação matemática com as queimadas é a umidade relativa do ar, como visto em trabalhos como Torres (2011) e PACHECO e SILVA (2018), onde foram calculados correlações baseadas em incêndios florestais em área urbana de Juiz de Fora (MG) e área urbana de São Luís (MA).

Para se entender sobre umidade relativa primeiramente deve-se voltar a alguns conceitos da física clássica, entendendo a composição do ar atmosférico. O ar atmosférico é composto por diversos gases, tais como nitrogênio (78%), o oxigênio (21%) e certas quantidades de outros gases, como dióxido de carbono, vapor de água, entre outros (TIPLER, 1995).

Esse número de vapor de água presente no ar atmosférico varia muito com o local e a condição climática, esse número é expresso como a umidade absoluta, as partículas de água, quando forem evaporadas em um local fechado, aumentam em quantidade, crescendo assim a pressão parcial e a umidade absoluta, existe um equilíbrio entre a evaporação e condensação, quando este equilíbrio está sendo obedecido, é considerado que este local está saturado de vapor e assim, a pressão é chamada de pressão de saturação, essa pressão é alterada diretamente proporcional a temperatura do ambiente (ESGANZELA, 2014).

A umidade relativa do ar (UR), é expressa como a razão, entre a umidade absoluta e a umidade quando está no ponto de saturação, também pode ser determinada na proporção da pressão parcial de vapor e a pressão de saturação, na mesma temperatura, assim como mostra a equação abaixo. (COSTA, 2003)

$$UR = \frac{p}{P} \quad \text{OU} \quad UR = \frac{ua}{us}$$

2.8 Inteligência Artificial

A tecnologia está cada vez mais presente nos últimos anos no dia a dia das pessoas, os processadores estão exponencialmente mais rápidos e a internet está cada dia mais rápida e presente na vida da população, umas dessas nossas inovações a serem apresentada são as Inteligências Artificiais (IA) que, por causa da evolução

do processamento, estão sendo utilizadas nas mais variadas áreas como forma de auxiliar na decisão dos humanos (TEIXEIRA, 2019).

No processo de coleta de dados, data mining e filtros, a inteligência artificial tem se tornado cada vez mais criteriosa e eficaz para extrair as informações relevantes dentro de inúmeros dados de uma base. E o diferencial será o resultado proveniente da interação da inteligência humana com a artificial, cada uma delas com suas vantagens sendo utilizadas na construção de um resultado que trará várias soluções para o contexto atual, originando uma inteligência ampliada (TEIXEIRA, 2019).

Entrando no contexto histórico, as IA's começaram a tentar ser produzidas a partir dos anos 40, quando se buscava descobrir novas formas de como se utilizar os computadores e, assim, tentando se assemelhar com os estudos de como funciona o raciocínio humano. (JESS, 2004)

Desde sempre, o termo inteligência artificial soou de uma forma presunçosa, uma vez que para a época, como a limitação de processamento, seus objetivos eram muito exagerados e certamente decepcionantes, mas com o passar das décadas, cada vez surgindo novas tecnologias de processamentos, esses objetivos não soavam um tanto utópicos e começou-se a surgir aplicações e desenvolvimento destes modelos, assim com a consolidação das redes neurais.

Inteligências artificiais tem inúmeras definições, pois suas diferentes formas de aplicação deixam difícil achar um conceito geral, isso porque são variadas formas onde elas trabalham com objetivos diferentes, sejam para buscar padrões em variáveis, ou agrupar semelhantes e podem ser vistas em diversas formas (TEIXEIRA; 2019).

A inteligência artificial é uma área de estudos ainda não muito bem delimitada que reúne a Ciência da Computação, a Psicologia, a Linguística e a Filosofia. Ela possui duas vertentes principais representadas pelo Projeto de Simulação Cognitiva e pelo Projeto de Inteligência Artificial num sentido restrito do termo. (TEIXEIRA; GONZALEZ, 1983).

Uma inteligência artificial pode ser comparada a uma criança, essa por sua vez antes de entrar dados se encontra vazia, os tipos de algoritmos serão a forma que essas IA's irão utilizar essas entradas de dados, seja para perceber um padrão nos dados passados e assim se prever o futuro, agrupar coisas que se parecem, ou tentar aperfeiçoar um resultado de algo pelo erro, assim como uma criança aprende, quando está segurando uma caneca e ela solta essa caneca, ela aprende que se soltar a

caneca ela cai e quebra, então ela não irá mais soltar. Essas formas de aprendizado podem ser do tipo supervisionado e não supervisionado.

a) Aprendizado Supervisionado

Ele tem esse nome pois as entradas e saídas de dados produzidas pelo algoritmo são fornecidas pelo supervisor, assim os dados são fornecidos, o supervisor ajusta o modelo vendo se está sendo efetivo ou não assim podendo regular a saída, (BRAGA, CARVALHO, LUDEMIR, 2000)

b) Aprendizado não supervisionado

No caso do aprendizado não supervisionado, a inteligência artificial irá trabalhar sem um supervisor, então o algoritmo forma uma espécie de padrões estatísticos para criar agrupamentos, só é possível quando essas entradas de dados possuem redundância, ou seja, atributos que permitam que sejam encontrados padrões característicos nesse banco de dados (BRAGA, CARVALHO, LUDEMIR, 2000).

3 METODOLOGIA

3.1 Quanto à natureza

Esta produção científica teve como alicerces a pesquisa aplicada que se baseia na busca de conhecimentos para se chegar na resposta de problemas encontrados pelo investigador. Dessa forma, essa pesquisa buscou informações com a finalidade de serem investidas em circunstâncias específicas (GIL, 2017).

3.2 Quanto aos objetivos

Este trabalho teve por base a pesquisa exploratória. Essa metodologia nos estabelece um primeiro contato com o tema, assim se pôde conhecer a respeito de queimadas e fatores atmosféricos que os influenciam, assim como métodos de inteligência artificial que possam ser utilizados na pesquisa (ZAMBELLO, et. al., 2018).

Além disso, foi realizado também, um estudo voltado nos princípios da pesquisa descritiva, da qual é elaborada com a finalidade de descrever fenômenos, como relações entre variáveis, e objetivo de descrever características presente nos diferentes métodos utilizados nessa pesquisa (GIL, 2017).

3.3 Quanto aos procedimentos

Este estudo teve como procedimentos o descrito nos princípios da pesquisa bibliográfica, os quais buscam em banco de dados, artigos, livros, teses dissertações, monografias e outros documentos publicados (SOUZA, 2019).

Esta pesquisa teve também como recurso a pesquisa experimental, que é constituída em após ser determinado um objeto de pesquisa, será selecionado parâmetros que possam influenciá-lo e assim o investigador observará que efeitos ocorrem no objeto (GIL, 2017).

3.4 Quanto à abordagem do problema

Neste projeto foi empregado predominantemente a pesquisa na modalidade quantitativa que se fundamenta na obtenção de dados numéricos e estes são tratados com ferramentas matemáticas para se moldar na forma que seja útil para

o pesquisador poder analisá-los e chegar a certas conclusões, essas ferramentas matemáticas podem ser equações, método numéricos, transformações de grandezas, porcentagens e outras (PEREIRA, 2018).

3.5 Local da pesquisa

Este estudo está delimitado às queimadas que ocorreram no município de Mirador do estado do Maranhão no espaço de tempo que compreendem os anos de 2009 a 2020.

3.6 Materiais e métodos

3.6.1 Obtenção de dados

Os dados de variáveis atmosféricas utilizados nesta pesquisa foram obtidos através da estação meteorológica convencional (EMC) de Colinas, cidade do Estado do Maranhão com coordenadas de latitude -6,03305554, longitude -44,23333333 e altitude de 175,56 com o código da estação 82765 que foram obtidos pelo Instituto Nacional de Meteorologia (INMET), a EMC de Colinas é a estação mais próxima ao município de Mirador, que fica a 47,4 quilômetros de distância do município.

Os dados estão disponíveis no período de 1º de janeiro de 2009 a 31 de dezembro de 2020 e estão dispostos em dados diários de evaporação de piche(mm), velocidade média do vento(m/s), precipitação (mm), temperatura máxima (em °C), temperatura média (em °C), insolação (horas por dia), umidade relativa do ar (%).

Os dados de incêndios em vegetação no município de Mirador foram adquiridos com base em imagens de satélite pelo Instituto Nacional de Pesquisas Espaciais (INPE) para o período de 1º de janeiro de 2009 a 31 de dezembro de 2020. A figura 9, abaixo, ilustra 1136 manchas de incêndios florestais ocorridas neste período, o quadro 1 mostra como foi obtido o conjunto de dados de incêndio, que estão dispostos por data e horários e em coordenadas do tipo decimal, assim como mostra o satélite do qual foram obtidos os dados, sendo utilizado apenas dados de um satélite neste trabalho para não ocorrer a duplicidade de dados.

O Satélite AQUA foi lançado em 2002 e ainda continua em operação até a atualidade, de acordo com os ensinamentos de Tomzhinski (2011), em estudos

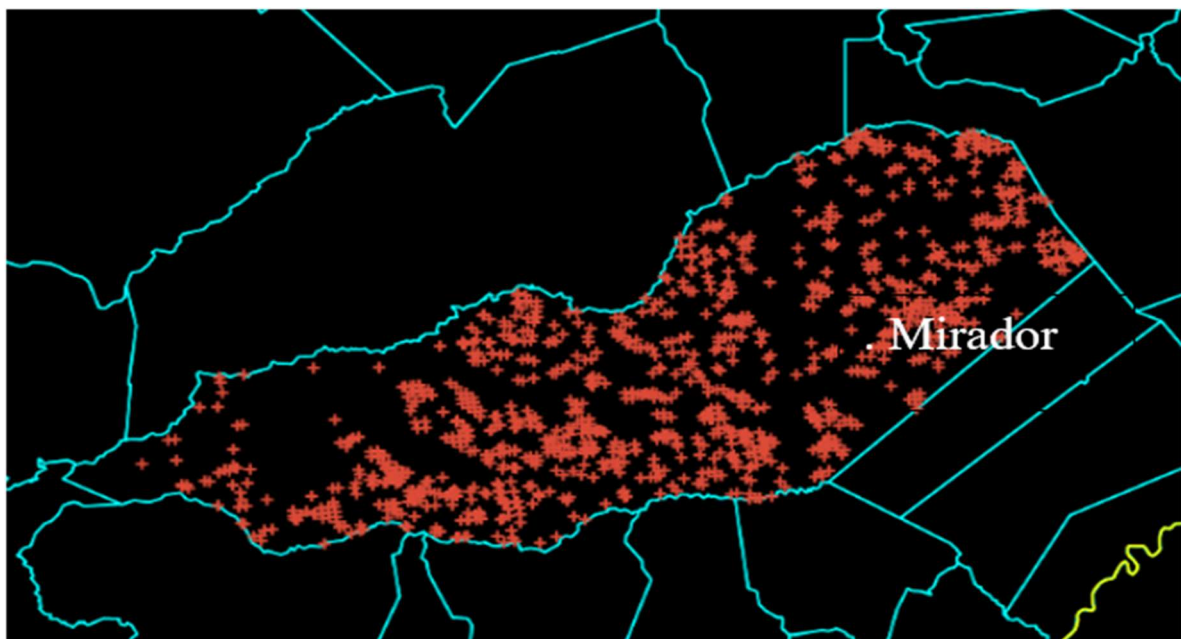
mostra que ele é o que tem os melhores resultados nas detecções de focos de incêndios, é um satélite americano que está aproximadamente a uma altitude de 705 quilômetros e mesmo tendo pixels de 1x1km, consegue detectar focos com frente de fogo com cerca de 30 metros de extensão por 1 metro de largura, na figura 10 mostra um exemplo da imagem de satélite de um foco de incêndio.

Quadro 1 - Dados de focos de incêndios

Data/hora	Satélite	Pais	Estado	Município	Bioma	Latitude	Longitude
21/04/2009 16:27	AQUA_M-T	Brasil	MARANHAO	MIRADOR	Cerrado	-66,910	-458,230
13/06/2009 16:45	AQUA_M-T	Brasil	MARANHAO	MIRADOR	Cerrado	-64,060	-449,710
13/06/2009 16:45	AQUA_M-T	Brasil	MARANHAO	MIRADOR	Cerrado	-66,960	-451,310
15/06/2009 16:33	AQUA_M-T	Brasil	MARANHAO	MIRADOR	Cerrado	-63,620	-449,830
17/06/2009 16:20	AQUA_M-T	Brasil	MARANHAO	MIRADOR	Cerrado	-66,640	-450,730
17/06/2009 16:20	AQUA_M-T	Brasil	MARANHAO	MIRADOR	Cerrado	-66,600	-458,170
17/06/2009 16:20	AQUA_M-T	Brasil	MARANHAO	MIRADOR	Cerrado	-66,670	-450,910
20/06/2009 16:51	AQUA_M-T	Brasil	MARANHAO	MIRADOR	Cerrado	-64,430	-455,040
20/06/2009 16:51	AQUA_M-T	Brasil	MARANHAO	MIRADOR	Cerrado	-64,340	-455,050
22/06/2009 16:38	AQUA_M-T	Brasil	MARANHAO	MIRADOR	Cerrado	-66,880	-449,160

Fonte: INPE, 2021

Figura 9 - Imagem Via Satélite do Município de Mirador mostrando os focos de incêndios vistos pelo satélite AQUA_M-T entre 1 de janeiro de 2020 até 31 de dezembro de 2020



Fonte: INPE (2021)

Figura 10 – Imagem de queimada captada pelo satélite TERRA, através do sensor MODIS, o mesmo sensor presente no satélite AQUA



Fonte: NUNES (2020)

Foi necessário fazer um tratamento dos dados atmosféricos, pois apresentava dados nulos, estes dados se encontram dispostos nessa situação quando ocorre alguma falha nos sensores (INMET, 2021), utilizando biblioteca PANDAS na linguagem Python, que é uma recurso que possui a função “*isnull()*”, que procura dados nulos em um banco de dados, foi-se verificado como mostra o quadro 2, o número de dados nulos.

Quadro 2 - Dados Atmosféricos nulos

Variável	Quantidade de dados Nulos
Evaporação do Pinche Diária(mm)	1163
Insolação total, diário(h)	180
Precipitação total, diário(mm)	69
Temperatura máxima, diário (°C)	128
Temperatura média, diário (°C)	206

Temperatura mínima, diário (°C)	467
Umidade Relativa do Ar, média diária (%)	110
Vento, velocidade média diária(m/s)	1775

Fonte: Elaborado pelo autor

Para se resolver tal situação, primeiramente, pode-se observar que evaporação, velocidade média do vento e temperatura mínima possuem muitos dados nulos, por este motivo foram removidos desta pesquisa, após isso, para cada dado nulo foi substituído por uma média entre dados de um dia anterior e o dado do próximo dia que não tiver dados nulos.

Para a criação dos modelos, o banco de dados de 4383 linhas, foi disposto em dados diários de variáveis atmosféricas, foi inserido uma coluna com a variável “dias sem chuva” onde são valores calculados a partir de uma contagem de dias anteriores em sequência de quando a variável "precipitação" era igual a 0 e foi inserido também uma coluna com uma variável chamada “Focos ativos”, em que nela foi inserido valores categóricos entre “sim” e “não”, onde quando tem o valor “sim”, foi detectado incêndio no referido dia e onde tem o valor “não”, não foi observado incêndios no dia, como mostrado no quadro 3

Quadro 3 - Banco de dados para o modelo

Data	Insolação	Precipitação	Temperatura Máxima	Temperatura Média	Umidade	Dias sem chuva	Focos Ativos
04/11/2009	1,4	31,7	28,7	24,36	92,75	5	nao
05/11/2009	6	5	30,8	25,82	84,25	0	nao
06/11/2009	10,2	2,2	34,2	27,04	77,75	0	sim
07/11/2009	7,1	0	33,7	27,14	78,25	0	nao
08/11/2009	6,9	0	32,9	27,56	78	1	nao
09/11/2009	11	0	35	28,16	72,75	2	nao
10/11/2009	8	0	33,2	27,72	74,75	3	nao
11/11/2009	2,2	0	31,7	26,68	84,5	4	nao
12/11/2009	9,3	0	35,2	28,02	80,25	5	nao

13/11/2009	6,6	0	33	28,12	75,75	6	sim
------------	-----	---	----	-------	-------	---	-----

Fonte: Elaborado pelo autor

3.6.2 Análise Multivariada

Fazer uma análise multivariada dos dados que se quer trabalhar nos modelos preditivos é de suma importância quando se está trabalhando com diversos dados, auxilia no dimensionamento do grau de relação entre essas variáveis (SCHIMITT, 2008).

A análise multivariada é um conjunto de ferramentas estatísticas em que realizam medições simultâneas de várias variáveis do banco de dados (AFONSO, 2001). Os métodos da análise multivariada que foi utilizado nesta pesquisa é a matriz de correlação, o agrupamento em dendograma, o agrupamento K-means e análise de componentes principais:

Os métodos multivariados são escolhidos de acordo com os objetivos da pesquisa, pois sabe-se que a análise multivariada é uma análise exploratória de dados, prestando-se a gerar hipóteses, e não tecer confirmações a respeito dos mesmos, o que seria uma técnica confirmatória, como nos testes de hipótese, nos quais se tem uma afirmação a respeito da amostra em estudo. (PAIXÃO, 2017)

Quadro 4 - Softwares para Análise Multivariada

Método	Software
Matriz de correlação	Linguagem Python, biblioteca PANDAS
Dendograma	Linguagem Python, biblioteca SciPy
K-means	Linguagem Python, biblioteca Sckit Learn
Análise de componentes principais	STATISTICA—desenvolvido pela StatSof.

Fonte: Elaborado pelo autor

3.6.2.1 Matriz de correlação

No estudo das relações entre variáveis, é necessário se observar a correlação existente, que no caso é a interdependência entre elas, muitas vezes é imperceptível ao fazer uma observação direta, pois o conjunto de dados é muito grande e disperso, então existem formas de se calcular tal correlação

matematicamente (VICINI, 2005). Nesta pesquisa foi utilizado o coeficiente de correlação de Pearson.

O coeficiente de correlação de Pearson é um método de correlação linear entre duas variáveis, onde se encontra o valor de “ ρ ”, esse valor pode estar entre o “-1” e “+1”, onde 0 representa que não há correlação, “-1” quantas mais próximos se chegar de “-1” maior será a correlação negativa e quanto mais próximo se aproximar de “+1” indica uma forte correlação positiva (MORETTIN, BUSSAB, 2010), a fórmula para obtenção deste índice é a presente abaixo:

$$\rho = \frac{\Sigma(x-\underline{x}).(y-\underline{y})}{\sqrt{\Sigma(x-\underline{x})^2.\Sigma(y-\underline{y})^2}}$$

A interpretação deste do valor “ ρ ” pode ser feita dividindo intervalos de valores e qualificando como fraca, moderada, forte e muito forte, (CALLEGARI-JACQUES, 2009), assim como apresentado no quadro abaixo:

Quadro 5 - Classificação da correlação de Pearson

“ ρ ” de Pearson	Classificação
$0 < \rho \leq 0,3 $	Fraca
$ 0,3 < \rho \leq 0,6 $	Moderada
$ 0,6 < \rho \leq 0,9 $	Forte
$ 0,9 < \rho < 1 $	Muito forte

Fonte: CALLEGARI-JACQUES, 2009.

A partir os dados apresentados na tabela 5 onde está disposto os dados de correlação entre todos os dados trabalhados na tabela 3, no caso da variável “Focos Ativos”, para a utilização da correlação, se utilizou o valor 1 representando “sim” e o valor “0” representando “não”.

Pode-se observar que os parâmetros meteorológicos possuem correlação linear com os focos de incêndio de classificação entre “fraca” e “moderada”, de acordo com a relação da tabela 2, como por exemplo a maior correlação dos focos de

incêndios foi com a umidade, atingindo o índice de -0,574394, se classificando como uma correlação moderada, juntamente com insolação, temperatura máxima e dias sem chuva, variáveis como precipitação, temperatura média se classificaram como correlação linear fraca.

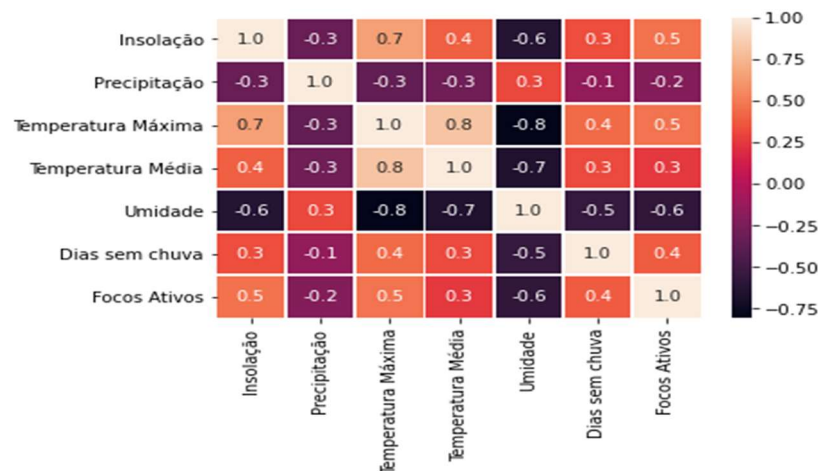
Tabela 1 - Matriz de correlação linear

	Insolação	Precipitação	Temperatura Máxima	Temperatura Média	Umidade	Dias sem chuva	Focos Ativos
Insolação	1,000000	-0,313729	0,651243	0,396788	-0,624881	0,329894	0,467848
Precipitação	-0,313729	1,000000	-0,346696	-0,287422	0,313799	-0,142891	-0,204342
Temperatura Máxima	0,651243	-0,346696	1,000000	0,812987	-0,807552	0,438126	0,475424
Temperatura Média	0,396788	-0,287422	0,812987	1,000000	-0,650701	0,316891	0,253777
Umidade	-0,624881	0,313799	-0,807552	-0,650701	1,000000	-0,548127	-0,574394
Dias sem chuva	0,329894	-0,142891	0,438126	0,316891	-0,548127	1,000000	0,381994
Focos Ativos	0,467848	-0,204342	0,475424	0,253777	-0,574394	0,381994	1,000000

Fonte: Elaborado pelo autor

Este método, mostrando que possuem correlação fraca e moderada, sugere que modelos preditivos com carácter linear não são recomendados e sim deve-se buscar métodos de modelos preditivos de carácter não linear, outra forma de se observar a matriz de correlação linear, ainda de forma mais didática é através de mapas de cores (heatmap) disposto na figura 11.

Figura 11 - Heatmap da correlação linear



Fonte: Elaborado pelo autor

Uma coisa interessante que pôde-se observar é que o valor de “ρ” entre as variáveis umidade e temperatura máxima foi de -0,807552 e entre temperatura máxima e temperatura média foi de 0,812987, demonstrando que tais combinações de variáveis possuem um forte grau de correlação linear, a primeira negativamente e a segunda positivamente.

3.6.2.2 Dendograma

Um dendograma é uma espécie de gráfico que se verifica a semelhança de agrupamentos, é muito utilizado para se estudar clusters hierárquicos para poder se tomar decisões a respeito dos parâmetros que serão utilizados na pesquisa, ele é gerado a partir do cálculo de distâncias entre os parâmetros que estão sendo estudados, existem algumas formas para se calcular essa distância, para efeitos deste estudo será utilizado a distância euclidiana, onde se calcula este valor através do teorema de Pitágoras (VICINI, 2009).

$$d_{ij} = \left[\sum_{j=1}^p (x_{ij} - x_{ij})^2 \right]^{\frac{1}{2}}$$

A distância euclidiana possui um problema, pois quando se trata de variáveis que possuem escalas diferentes, seus resultados podem ter alguma alteração, então é necessário se fazer a padronização dessas variáveis da pesquisa (MANLY, 1986), uma forma de se fazer a essa normalização é utilizando a fórmula z-score, em que se calcula Z, o valor padronizado, onde μ é a média e σ é o desvio padrão, esta ferramenta matemática ajuda os dados a serem redimensionados, tendo desvio padrão igual a 1 e a média com o valor 0 (SCHIMIT, 2009), assim como se pode observar na tabela 2.

$$z = \frac{x - \mu}{\sigma}$$

Tabela 2 - Dados padronizados

Data	Insolação	Precipitação	Temperatura Máxima	Temperatura Média	Umidade	Dias sem chuva	Focos Ativos
2009-01-01	-0.347936	-0.322191	0.022199	-0.399007	0.714841	-0.578917	0
2009-01-02	-0.874571	0.813923	-0.628303	-0.175136	0.678259	-0.524980	0
2009-01-03	0.581420	-0.301345	-0.016066	0.683035	-0.217988	-0.578917	0
2009-01-04	0.767292	-0.322191	-0.092595	0.732784	-0.126534	-0.578917	0
2009-01-05	-0.719678	-0.322191	-0.283919	0.260168	0.111246	-0.524980	0

Fonte: Elaborado pelo autor

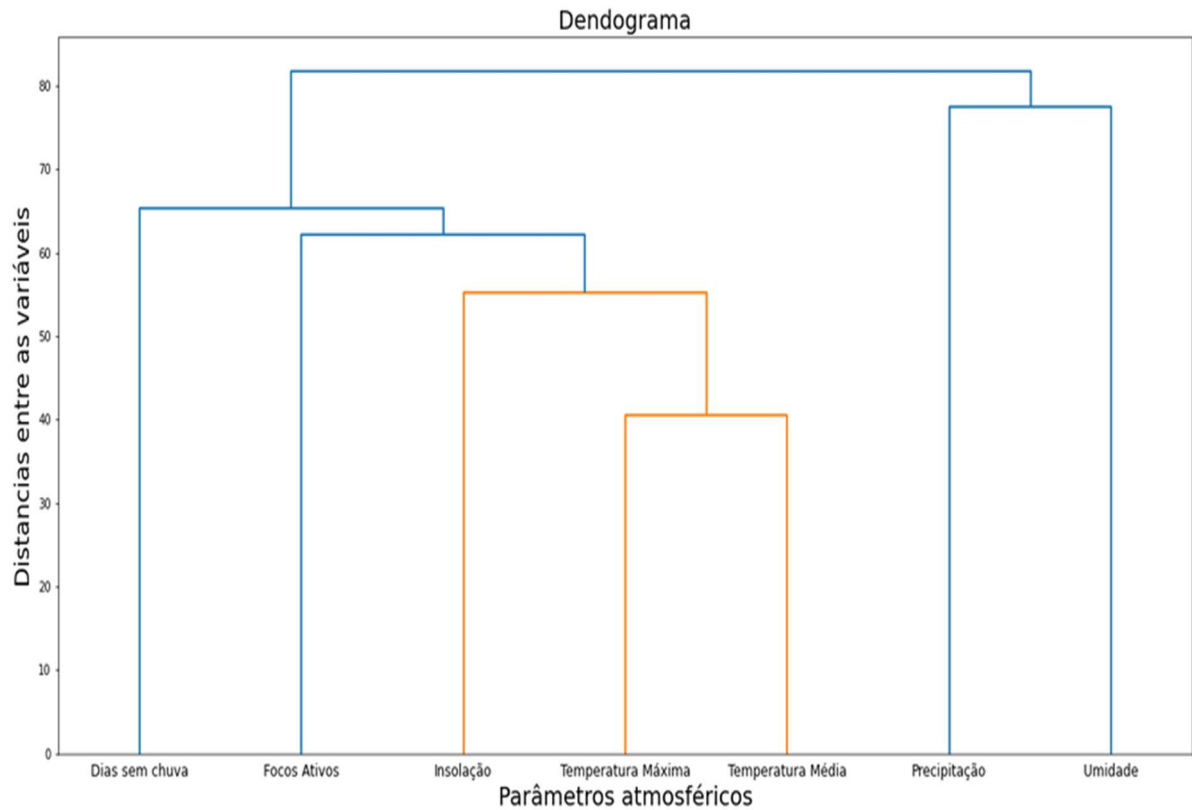
Pode-se gerar a tabela de distâncias onde se pode observar as distâncias entre cada variável, onde o método utilizado foi de ligação simples, que se baseia pela distância mínima, como pode-se apreciar na tabela 3, e a partir destes dados pode-se gerar um dendrograma e verificar as relações entre tais parâmetros presente na figura 12, onde pode-se observar que foi feito 2 tipos de agrupamentos.

Tabela 3 - Tabela de distâncias

	Insolação	Precipitação	Temperatura Máxima	Temperatura Média	Umidade	Dias sem chuva	Focos Ativos
Insolação	0,000000	34,968549	13,081231	16,491635	27,922694	29,937546	29,340478
Precipitação	34,968549	0,000000	39,110590	24,672988	47,885573	13,952893	38,555495
Temperatura Máxima	13,081231	39,110590	0,000000	16,310756	21,826976	34,136805	32,313898
Temperatura Média	16,491635	24,672988	16,310756	0,000000	30,551445	20,725754	30,208604
Umidade	27,922694	47,885573	21,826976	30,551445	0,000000	46,217762	46,195043
Dias sem chuva	29,937546	13,952893	34,136805	20,725754	46,217762	0,000000	32,794049
Focos Ativos	29,340478	38,555495	32,313898	30,208604	46,195043	32,794049	0,000000

Fonte: Elaborado pelo autor

Figura 102 - Dendrograma



Fonte: Elaborado pelo autor

3.6.2.3 Método K-Means

O método *k-means* é uma ferramenta de agrupamento do tipo não-hierárquico, ele busca separar dados função de centros, que são chamados de centroide e isso é chamado de *clustering*, o objetivo é selecionar argumentos de dados similares em determinado número de grupos que são fornecidos como entrada (MACHADO, FARINHAKI, SILVA, 2013).

Durante seu funcionamento, gera pontos aleatórios que se tornam centros de cada agrupamento, que se chama de cluster, ou seja, o centro do clusters, o número desses pontos é a entrada que é definida pelo usuário, então são calculados as distâncias dos dados para esses centros, através da distância euclidiana, então o dado será adicionado ao cluster mais próximo, depois dessa etapa é calculado o ponto dos centros através da média dos pontos de cada parâmetro, e isso vai ocorrendo a cada ciclo até os centroides não se movimentarem mais, isso ocorre quando nenhum dado se movimenta de cluster(LI, Wu, 2021), assim como ocorre na figura abaixo.

Figura 13 - Movimentação dos centros



Fonte: MINERANDO DADOS (2017)

Nessa pesquisa o *k-means* foi utilizado como forma de se agrupar as variáveis da tabela 6, e a entrada passada foi em 2 grupos, e o algoritmo tratou de fazer o agrupamento, assim como demonstrado na tabela 4, o que se pôde observar é que tanto na divisão dos agrupamentos pelo algoritmo de *k-means* quanto no dendrograma produzido a partir das distância médias entre variáveis ocorre a mesma divisão de relação entre estes parâmetros, onde as variáveis Umidade e precipitação se colocam em um grupo e as restantes em outro agrupamento.

Tabela 4 - Agrupamentos formados pelo método *k-means*

Grupo 1		Grupo 2	
Variável	Distância do centro	Variável	Distância do centro
Precipitação	38	Temperatura máxima	30
Umidade	38	Insolação	43
		Temperatura média	44
		Focos Ativos	48
		Dias sem Chuva	50

Fonte: Elaborado pelo autor (2021)

3.6.2.4 Método dos componentes principais

Essa ferramenta matemática possibilita análises de um número muito grande de dados, possibilitando uma investigação sem perdas de informações significativas,

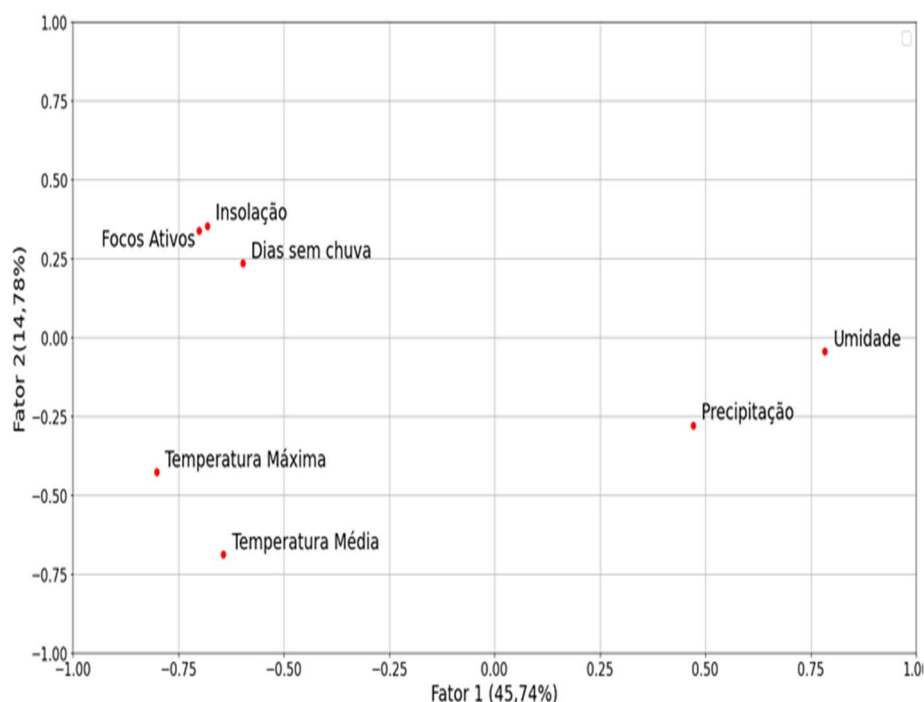
ela busca transformar um conjunto de dados de uma dimensão grande e transformar em outro tipo de conjunto (HONGYU, 2016).

A técnica de componentes principais tem como objetivo geral a redução da dimensionalidade e interpretação de um conjunto de dados. Obter esta redução num conjunto de variáveis, consiste em encontrar combinações lineares das variáveis, que irão gerar um outro conjunto de variáveis com novas coordenadas e não correlacionadas entre si. Geometricamente, os componentes principais representam um novo sistema de coordenadas, obtidas por uma rotação do sistema original, que fornece as direções de máxima variabilidade, e proporciona uma descrição mais simples e eficiente da estrutura de covariância dos dados. (SCHIMIT, 2005)

Este método, transforma um conjunto de dados de muitas dimensões e transforma para interpretação em poucas dimensões, fazendo algumas combinações lineares, assim se pode tirar análises relacionais entre as variáveis, essas informações nós chamamos de fatores, estes novos dados representam a variância dos mesmos dados anteriores e assim pode-se verificar a relação entre estes parâmetros (VARELA, 2008)

O software STATISTICA, da *StatSoft*, auxilia nesta tarefa, onde colocamos os dados no programa e ele faz essas combinações lineares e dividem em fatores, em que cada fator explica uma certa porcentagem da variância dos dados, na figura 14, se tem um gráfico com relação a 2 fatores que explicam mais de 50% da variância dos dados.

Figura 14 - Análise de componentes principais



Fonte: Elaborado pelo autor (2021)

Através deste método pode-se perceber que também se aproximou bastante dos resultados obtidos mediante as análises feitas através do dendrograma e *do k-means*, onde os focos ativos de incêndios estão relacionados com a insolação e dias sem chuva, a temperatura máxima está relacionada a temperatura média e a umidade com a precipitação, onde uma coisa que pode-se observar é que os altos índices de insolação aliados com dias sem chuva aumentam os focos ativos de incêndios, assim como também estudado por (TORRES, 2011).

3.6.3 Modelagem

Para esta pesquisa, foram utilizados 3 métodos para a construção de modelos preditivos, as redes neurais artificiais (RNA), k-vizinhos mais próximos e o Florestas de decisão Aleatória onde no quadro 6 mostra os softwares utilizados foram:

Quadro 6 - Softwares utilizados

Método	Software
Redes neurais artificiais	Linguagem Python, biblioteca <i>Sckit Learn</i>
k-Vizinhos mais próximos	Linguagem Python, biblioteca <i>Sckit Learn</i>
Floresta de Decisão Aleatória	Linguagem Python, biblioteca <i>Sckit Learn</i>

Fonte: Elaborado pelo autor

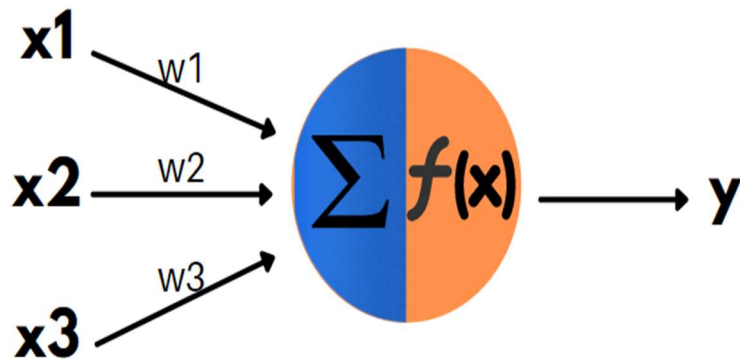
3.6.3.1 Redes Neurais artificiais

As redes neurais artificiais são estruturas matemáticas que se inspiram na estrutura das conexões nervosas dos cérebros, onde existem os neurônios e esses neurônios se conectam entre si através das ligações sinápticas, em que se tem uma entrada, que são multiplicados por um peso para chegar nos neurônios, nos neurônios ocorre uma operação matemática para se ter determinada saída, a conexão de vários neurônios que são chamados de redes neurais. (DAUFEMBACK, 2008)

Neste estudo, para a modelagem será utilizado a rede neural do tipo perceptron de multicamadas, onde se tem uma camada de entrada, uma camada oculta, bem como uma rede de neurônios e uma camada de saída, onde os dados ao

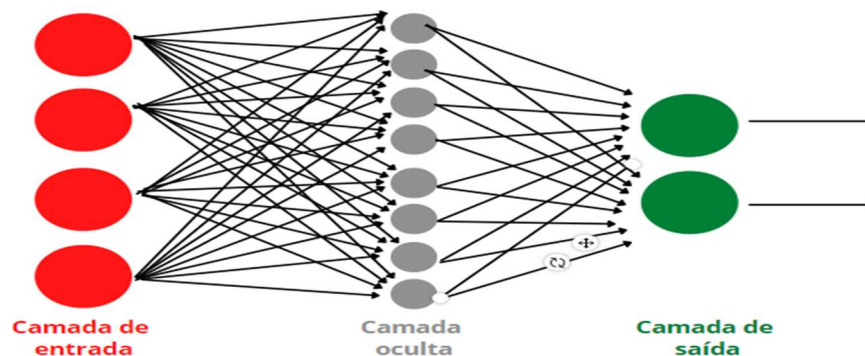
chegarem nos neurônios, depois serem ponderados (w_i), ocorre uma combinação linear dos valores e passam por uma função de ativação, essa função de ativação gera o valor da saída do neurônio, assim como mostra a figura abaixo, esse tipo de algoritmo é treinado através do método de retro propagação de erros. (VIGANÓ, 2017).

Figura 15 - Neurônio Artificial



Fonte: Elaborado pelo Autor

Figura 16 - Rede Neural Artificial



Fonte: Elaborado pelo autor

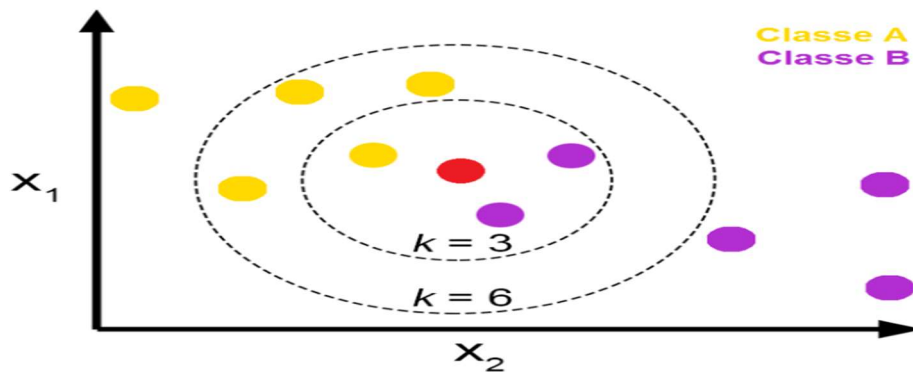
A retro propagação de erros é um tipo de treinamento onde se tem a primeira etapa onde se tem a entrada de dados, o processamento na rede neural e a saída, após isso a saída é comparada com o resultado que se esperava, então é calculado o erro, após isso o valor é processado de volta na rede pela camada de saída reajustando os pesos das conexões sinápticas, uma rede é considerada estar treinada quando não houver mais necessidade desses pesos serem alterados (VIGANÓ, 2017).

3.6.3.2 K-vizinhos mais próximos

O método de k-vizinhos mais próximos utiliza a lógica simples de que ele observa um certo número de dados próximos ao dado de entrada (número de vizinhos)

e o classifica com a classe da maioria, ou seja, como por exemplo na figura –, onde se escolhe um número de vizinhos(K) que irá ser analisado, após isso é verificada a classe que possui o maior número de indivíduos entre os vizinhos, então a entrada é classificada como esta classe. Essa proximidade pode ser calculada assim como no agrupamento *k-means* por distancia euclidiana (ZHANG, 2016).

Figura 17 - Exemplo de k-Vizinhos mais próximos



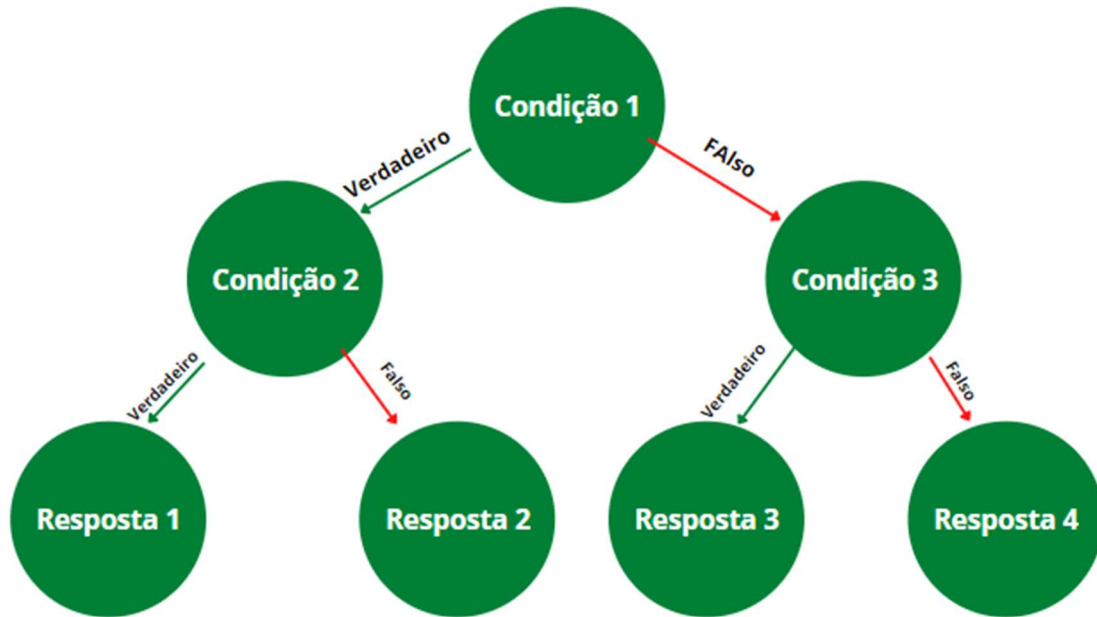
Fonte: JOSÉ, 2008

Nesta pesquisa foi-se utilizado para análise o total de 100 vizinhos, onde o dado de entrada é um vetor com 6 dados, onde cada um destes dados representa um valor de cada variável atmosférica e as classes são focos ativos e focos não ativos.

3.6.3.3 Floresta de decisão aleatória

Uma árvore de decisão é uma espécie de mapa de possibilidades onde permite que chegue a um através de combinações de condições diferentes em suas ramificações, e a partir de seus dados de treinamento, vai criando nós condicionais com fins de chegar a um determinado valor (SANTANA, 2020)

Figura 18 - Árvore de decisão



Fonte: Elaborado pelo autor (2021)

A floresta de decisão aleatória é um conjunto de árvores de decisão onde elas trabalham em conjunto para se chegar a uma previsão com maior acurácia, quando está sendo treinado o modelo, ele cria um certo número de árvores que são passadas como entrada, onde o número de árvores aliado com os fatores de aleatoriedade dos fatores nas árvores busca sempre os melhores resultados, isso ajuda a criar um classificador forte mesmo tendo uma correlação fraca. (SANTANA, 2020)

3.6.4 Avaliação dos modelos

3.6.4.1 Treinamento e testagem

Para avaliação desses modelos é necessário existir formas para o treinamento e testagem dele, no quadro 7, é demonstrada uma explicação de alguns métodos, dentre eles a ressubstituição, *holdout* e validação cruzada.

Quadro 7 - Métodos de treinamento e testagem

Método	Funcionamento
--------	---------------

Resubstituição	O conjunto de dados de dados que é utilizado para treinamento também é utilizado para fazer a testagem.
<i>Holdout</i>	O banco de dados é dividido em 2 grupos, um para a treinamento e outro para a testagem, os dados de cada grupo são selecionados de forma aleatória e normalmente o grupo de treinamento é maior que o grupo para o teste.
Validação cruzada	É um método onde os dados são divididos em um certo número de partições ou subconjuntos e é se escolhe 1 subconjunto para a testagem e o restante é utilizado para o treinamento, é isso várias vezes até todos os subconjuntos serem utilizados como dados de teste.

Fonte: COSTA, 2008

Para treinamento e testagem dos modelos, foi-se utilizado o método *holdout*, bem como foi usado 70% dos dados para o treinamento dos modelos e 30% para ser feito o teste, ou seja, dos 4383 dados, 3068 deles foram utilizados para treinamento e 1315 foram utilizados para a testagem do modelo

3.6.4.2 Matriz confusão

Para se mensurar a qualidade e assertividade se utiliza a matriz confusão, onde pode-se comparar os dados verdadeiros com os dados preditivos dos modelos , assim pode-se ter uma ideia de o quanto o modelo acertou e o quanto errou e saber exatamente onde ele foi mais assertivo e onde ele errou mais, (COSTA, 2008), como demonstrado no quadro 8.

Quadro 8 - Matriz Confusão

	Dados Previstos pelo modelo	
Dados reais	Focos não ativos	Focos Ativos
Focos não ativos	Previu corretamente	Previu incorretamente
Focos ativos	Previu incorretamente	Previu corretamente

Fonte: Elaborado pelo autor (2021)

3.6.4.3 Métricas

São utilizadas algumas métricas para se calcular a taxa de acerto do modelo, como por exemplo a acurácia e precisão. A acurácia é a porcentagem de acertos em função de todos os dados e a precisão é a porcentagem de acertos de cada dado que é importante, (JUNIOR, 2018), como demonstrado nas equações abaixo:

$$\text{Acurácia} = \frac{\sum \text{Previsões corretas (Tanto de focos ativos quanto de focos não ativos)}}{\text{Total de dados de previsão}}$$

$$\text{Precisão de Focos Ativos} = \frac{\text{Previsões corretas de focos ativos}}{\text{Focos ativos reais} + \text{Focos ativos previstos}}$$

$$\text{Precisão de Focos não Ativos} = \frac{\text{Previsões corretas de focos não ativos}}{\text{Focos não ativos reais} + \text{Focos não ativos previstos}}$$

4 RESULTADOS E DISCUSSÕES

No método utilizando redes neurais artificiais, foi-se utilizado a configuração padrão da função “*MLPRegressor()*” da biblioteca Scikit Learn, onde foi construído uma estrutura com uma camada oculta composta apenas de 1 camada de 100 neurônios. Pode-se observar na tabela 5, a matriz confusão referente a este modelo.

Tabela 5 - Matriz Confusão do Modelo de rede neural artificial

	Focos não ativos	Focos Ativos
Focos não ativos	738	132
Focos ativos	123	322
Total de amostras	1315	

Fonte: Elaborado pelo autor (2021)

Através destes fatos, obter as informações de que a acurácia deste modelo é de 80,60%, a precisão com relação ao acerto de focos ativos foi de 72,36% e a precisão com relação ao acerto de focos não ativos foi de 84,82%, como mostra a tabela 6.

Tabela 6 - Resultados do Método de rede neural artificial

	Focos não ativos	Focos ativos
Precisão	84,82%	72,36%
Acurácia do modelo	80,60%	

Fonte: Elaborado pelo Autor (20210)

No modelo obtido a partir do método K-Vizinhos mais próximos, se pode adquirir os resultados demonstrados na tabela abaixo.

Tabela 7 - Matriz Confusão do Modelo de K-Vizinhos mais próximos

	Focos não ativos	Focos Ativos
Focos não ativos	761	122
Focos ativos	127	319
Total de amostras	1315	

Fonte: Elaborado pelo autor

A partir desses dados, pode-se adquirir os dados relativos com as métricas do modelo, então a acurácia dele é igual a 77,71%, a precisão com relação aos acertos de focos ativos se tem o valor de 50,68% e a precisão com referência aos focos não ativos foi de 91,12%, assim como mostra a tabela 8.

Tabela 8 - Resultados do Método de k-Vizinhos mais próximos

	Focos não ativos	Focos ativos
Precisão	86,18%	71,52%
Acurácia do modelo	82,12%	

Fonte: Elaborado pelo Autor (2021)

Sobre o método de Floresta Aleatória, a matriz confusão da mesma encontra-se disposta na tabela 9, para assim poder-se analisar seus resultados.

Tabela 9 - Matriz Confusão do Modelo de Floresta Aleatória

	Focos não ativos	Focos Ativos
Focos não ativos	778	121
Focos ativos	119	297
Total de amostras	1315	

Fonte: Elaborado pelo autor (2021)

Então, mediante esses resultados, pode-se calcular uma acurácia de 81,74%, uma precisão nos focos ativos de 62,39% e uma precisão de focos não ativos de 86,54%, assim como mostra a tabela a seguir.

Tabela 10 - Resultados do Floresta Aleatória

	Focos não ativos	Focos ativos
Precisão	86,54%	71,39%
Acurácia do modelo	81,74%	

Fonte: Elaborado pelo Autor (2021)

Tabela 11 - Comparação entre os modelos

Modelos	Métricas		
	Acurácia	Precisão	
		Focos não ativos	Focos ativos
Rede Neural Artificial	80,60%	84,82%	72,36%
K-Vizinhos mais próximos	82,12%	86,18%	71,52%
Floresta Aleatória	81,74%	86,54%	71,39%
Melhor Modelo	k-vizinhos mais próximos	Floresta aleatória	Rede neural artificial

Fonte: Elaborado pelo autor (2021)

Mediante a análise contida nos dados trazidos no decorrer do trabalho, pode-se chegar à conclusão que todos os modelos obtiveram resultados satisfatórios, entre estes números pode verificar que, dentre a métrica de acurácia, o modelo que teve o melhor desempenho, foi o de K-Vizinhos mais próximos, dentre a precisão, o modelo se destacou na previsão dias com focos não ativos foi de Floresta Aleatória e a rede neural se destacou na previsão de dias com focos ativos.

5 CONSIDERAÇÕES FINAIS

Os incêndios florestais são uma preocupação constante, pois o não controle deles causa diversos danos, tanto a fauna e a flora e prejuízos econômicos, pois para seu controle é necessário o gasto de muitos recursos, por isso a importância do controle rápido das chamas. Em vegetações densas é muito difícil acessar locais de focos em incêndio e até as guarnições chegarem, boa parte já foi consumida, por isso a importância de se encontrar métodos para prevenir incêndios.

Neste trabalho fez-se análise dos parâmetros atmosféricos e como relacionam-se, a partir disso foi construído modelos de predição com 3 métodos diferentes e estes modelos alcançaram resultados satisfatórios, o objetivo inicial era a obtenção de dados em intervalos de tempo de 1 hora, pois o INMET fornece estes dados, mas os INPE não fornece dados horários de queimadas, pois o satélite passa apenas 1 vez por dia no seu ciclo orbital, então com estes escasso dados não poderia ser treinado um modelo, então optou-se por dados diários.

Para uma futura pesquisa, pode-se fazer analisando métodos diferentes, fazer uma categorização melhor dos dados com incêndio, como por exemplo, nesta pesquisa foi dividido em 2 categorias, focos ativos e focos não ativos, então poderia se criar mais classe discretas represento probabilidades de ocorrência de incêndios como crítico, moderado, baixo, pouco provável e assim poder-se construir um sistema usual que auxilie o combate das queimadas.

Pode-se também utilizar dados de mais satélites para detecção de incêndios, assim como outras fontes de dados atmosféricos com fins de melhorar o modelo construído, também é possível se estudar outros métodos preditivos e utilizá-los em conjunto para poder se ter uma melhor previsão.

Pode-se concluir que a construção de um sistema utilizado tais modelos é viável e pode ser melhorados ainda mais, como por exemplo, se tivessem mais estações meteorológicas na região, poderia ser feito previsões com espaços geográficos menores, pois os dados de focos de incêndio se tem as coordenadas geográficas e na pesquisa foi utilizado apenas uma estação meteorológica, que ela é a única da região, então os dados atmosféricos ficam limitados serem os mesmos para uma grande região.

Com base nessas discussões, um sistema para prever incêndios seria muito importante, pois com base nas informações repassadas por esse sistema, viabilizaria para os órgãos competentes para mobilizar recursos para a localidade, tais como brigadistas, caminhões pipas e também o apoio de Corpo de Bombeiros, com fins de diminuir o tempo resposta para o atendimento dessas ocorrências.

REFERÊNCIAS

- AFONSO. **Aplicação da análise multivariada para classificação e previsão de avaliação do desempenho acadêmico dos alunos de engenharia mecânica do CEFET**. Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro Tecnológico. Programa de Pós-Graduação em Engenharia de Produção. Santa Catarina, 2001 Disponível em: <
https://repositorio.unesp.br/bitstream/handle/11449/154346/paixao_css_dr_jabo.pdf?sequence=6> Acesso em 10 de mar. 2021
- AGENCIA DE NOTICIAS. **Governo lança programa “Maranhão sem Queimadas”, em Imperatriz**. Disponível em: <https://www.ma.gov.br/agenciadenoticias/?p=286061>
 > Acesso em 11 de mar. 2021
- AYOADE, J. O. **Introdução à Climatologia para os Trópicos**. 4ª. ed. Rio de Janeiro: Bertrand Brasil, 1996.
- BRAGA, Antonio de Pádua; LUDERMIR, Teresa Bernarda; CARVALHO, André Carlos Ponce de Leon Ferreira. **Redes neurais artificiais: teoria e aplicações**. [S.l.: s.n.], 2000.
- CALLEGARI-JACQUES, S, M, **Bioestatística: princípios e aplicações**, Tradução, [s.l.] Artmed Editora, 2009.
- CHUVAS. **NUGEO**, 2021. **Núcleo GEO**. Disponível em: <
https://www.nugeo.uema.br/?page_id=111>. Acesso em: 20 de março de 2021
- COSTA, Eduardo de Paula. **Investigação de técnicas de classificação hierárquica para problemas de bioinformática. 2008**. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2008. doi:10.11606/D.55.2008.tde-09052008-144238. Acesso em: 2021-07-14.
- COSTA, E. V. **Medidas da umidade relativa do ar em um ambiente fechado**. Revista Brasileira de Ensino de Física, v. 25, n. 3, p. 346–348, 2003.
- ESGANZELA, João Américo. **Material de estudo para o ensino da umidade relativa do ar**. UFMT. Cuiabá, 2014.
- FIEDLER, N. C., RODRIGUES, T. O., MEDEIROS, M. D. **Avaliação das condições de trabalho, treinamento, saúde e segurança de brigadistas de combate a incêndios florestais em unidades de conservação do Distrito Federal: estudo de caso**. Revista Árvore, 30(1), 55-63. 2006.
- FONTES, L. B. **Análise de Aprendizado de Máquina Aplicada a Focos de Incêndio nos Estados Brasileiros**. 2020.
- G1. **Mirador, MA, é 2ª cidade do país com mais focos de incêndio em 2015**. Reportagem por Mauricio Araya. Disponível em:

<<http://g1.globo.com/ma/maranhao/noticia/2015/08/mirador-ma-e-2-cidade-do-pais-com-mais-focos-de-incendio-em-2015.html>. Acesso em 10 de mar. 2021

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 5. ed. São Paulo: Atlas, 2017.

GOIÁS, CORPO DE BOMBEIRO MILITAR DO. **Manual Operacional De Bombeiros Combate a Incêndio Urbano**. 1° ed. Goiânia: Estado de Goiás, 2016.

HONGYU, Kuang. SANDANIELO, Vera Lucia Martins. JUNIOR, Gilmar Jorge de Oliveira. **ANÁLISE DE COMPONENTES PRINCIPAIS: RESUMO TEÓRICO, APLICAÇÃO E INTERPRETAÇÃO**. In: *Engineering and Science 2015*. disponível em: https://www.researchgate.net/publication/305713088_ANALISE_DE_COMPONENTES_PRINCIPAIS_RESUMO_TEORICO_APLICACAO_E_INTERPRETACAO> Acesso em 08 de mar 2021

IBGE. **Mirador (MA)**. Disponível em: <<https://www.ibge.gov.br/cidades-e-estados/ma/mirador.html>>. Acesso em: 20 de março de 2021

INPE. **Satélite AQUA**. disponível em: <<http://www.dgi.inpe.br/documentacao/satelites/aqua> > Acesso em 08 de mar. 2021

JESS, G. **Inteligência Artificial E Tecnologias Da Inteligência-Um Repensar Segundo Os Processos De Elaboração Matemática**. p. 139, 2004.

JOSE, Italo. **KNN (K-Nearest Neighbors) #1**. Disponível em: <<https://medium.com/brasil-ai/knn-k-nearest-neighbors-1-e140c82e9c4e>> Acesso em 10 de mar. 2021.

LIMA, M.G.; RIBEIRO, V.Q. **Equações de estimativa da temperatura do ar para o Estado do Piauí**. Revista Brasileira de Agrometeorologia, Santa Maria, v.6, n.2, p.221-227, 1998.

MACHADO, Renan Kruchelski; FARINHAKI, Ricardo; SILVA, Thiago Avelino da. **Seta: ambiente de realidade aumentada para o ensino de algoritmos de aprendizagem de máquina**. 2013. 82 f. Trabalho de Conclusão de Curso (Graduação) – Universidade Tecnológica Federal do Paraná, Curitiba, 2013. Disponível em: <<http://repositorio.utfpr.edu.br:8080/jspui/handle/1/8150> > Acesso em 2 de jun. 2021.

MADEIRO, Carlos. **Agosto atinge recorde de focos de incêndio no ano; AC e pantanal preocupam**. Uol, 2020. Disponível em <<https://noticias.uol.com.br/meio-ambiente/ultimas-noticias/redacao/2020/09/01/agosto-atinge-recorde-de-focos-de-incendio-no-ano-ac-e-pantanal-preocupam.htm>>. Acesso em: 20 de março de 2021

NUNES, Mônica. **Nasa divulga imagens de focos de incêndio “fora de hora” no Pantanal**. Conexão Planeta, 2020. Disponível em <<https://conexaoplaneta.com.br/blog/nasa-divulga-imagens-de-focos-de-incendio-fora-de-hora-no-pantanal/#fechar>>

THOMAS, E., MANLY, Bryan F. J.: **Multivariate Statistical Methods: A Primer**. Chapman and Hall, London – New York 1986, 159 S. ISBN 0-412-28610-6, ISBN 0-

412-28620-3 (Paperback-Ausgabe). *Biom. J.*, 30: 874-875. 1988. Disponível em: <<https://onlinelibrary.wiley.com/action/showCitFormats?doi=10.1002%2Fbimj.4710300742>> Acesso em 10 de abr. 2021

MORETTIN,, Pedro Alberto. MORETTIN, Wilton., **Estatística Básica** – 6, ed, – São Paulo : Saraiva, 2010

NUNES, J. R. S.; SOARES, R. V.; BATISTA, A. C. **FMA+ - Um Novo Índice de Perigo de Incêndios Florestais para o Estado do Paraná – Brasil**. *Floresta*, Curitiba, v. 36, n. 1, p. 178-179, 2006.

PACHECO; Gabriel de Oliveira; SILVA, William Ronald Sousa; **Desenvolvimento de uma ferramenta matemática para correlação de incêndio com dados atmosféricos**. UEMA, São Luís, 2018.

PAIXÃO, Jabo. **ANÁLISE UNI E MULTIVARIADA APLICADA À QUALIDADE OPERACIONAL DA COLHEITA MECANIZADA DE SOJA**. 2017. Tese (doutorado) - Universidade Estadual Paulista – UNESP, Faculdade de Ciências Agrárias e Veterinárias, 2017. São Paulo, 2017. Disponível em: <https://repositorio.unesp.br/bitstream/handle/11449/154346/paixao_css_dr_jabo.pdf?sequence=6> Acesso em 2 de mar. 2021.

PEREIRA, Adriana et al. **Metodologia da pesquisa científica**. 1. ed., Santa Maria, RS: UFSM, NTE, 2018.

REBOITA, M. S. et al. **Entendendo o tempo e o clima na América do Sul**. *Terra e Didática*, v. 8, n. 1, p. 34, 2015.

RIBEIRO, A. A.; ANDRADE JÚNIOR, A. S.; SILVA E. M.; SIMEÃO, M.; BASTOS E. A. **Comparação entre dados meteorológicos obtidos por estações convencionais e automáticas no estado do Piauí, Brasil**. *Irriga*, v. 22, n. 2, p. 220-235, 2017

RIBEIRO, H.; ASSUNÇÃO, J. V. DE. **Efeitos das queimadas na saúde humana**. *Estudos Avançados*, v. 16, n. 44, p. 125–148, 2002.

SANT'ANNA, A. A.; ROCHA, R. **Impactos dos Incêndios Relacionados ao Desmatamento na Amazônia Brasileira Sobre Saúde**. p. 7, 2020.

SANTANA, Felipe Bachion. **Floresta aleatória para desenvolvimento de modelos multivariados de classificação e regressão em química analítica**. Campinas, SP : [s.n.], 2020

SANTANA, Felipe. **Entenda o Algoritmo K-means e Saiba como Aplicar essa Técnica**. Disponível em: <https://minerandodados.com.br/entenda-o-algoritmo-k-means/#aplicacoes_cenarios> Acesso em 08 mar.de 2021

SCHMITT, Jeovani. et al. **Pré-processamento para a mineração de dados: uso da análise de componentes principais com escalonamento ótimo**. **Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro Tecnológico. Programa de Pós-Graduação em Ciência da Computação**, Florianópolis, 2005.

Disponível em: < <https://repositorio.ufsc.br/handle/123456789/81750> > Acesso em 12 de mar. 2021

SEARS, Francis Weston; ZEMANSKY, Mark Waldo; YOUNG, Hugh D.; FREEDMAN, Roger A. **Física II: termodinâmica e ondas**. 12. ed. São Paulo, SP: Pearson Addison Wesley, c2008-2009 vol 2;

SILVA JUNIOR, C. H. L. et al. **Dinâmica das Queimadas na Baixada Maranhense**. InterEspaço: Revista de Geografia e Interdisciplinaridade, v. 2, n. 5, p. 355–375, 2016.

SILVA JUNIOR, C. H. L. et al. **Dinâmica das Queimadas no Cerrado do Estado do Maranhão, Nordeste do Brasil**. Geography Department University of Sao Paulo, v. 35, p. 1–14, 2018.

SILVA, F. B. et al. **Análise espaço-temporal da precipitação no estado do Maranhão no período de 2003 a 2012**. XIV SAFETY, HEALTH AND ENVIRONMENT WORLD CONGRESS. Anais... Cubatão: COPEC, 2014.

SOUZA, Evânia et al. **Metodologia da pesquisa: aplicabilidade em trabalhos científicos na área da saúde**. 2. ed., rev. e ampl. – Natal, RN : EDUFRRN, 2019.

SPINELLI-ARAÚJO, L. et al. **Conservação da biodiversidade do estado do Maranhão: cenário atual em dados geoespaciais**. Embrapa Meio Ambiente, p. 28, 2016.

STRASSBURGER, A. S. et al. **Comparação da temperatura do ar obtida por estação meteorológica convencional e automática**. Revista Brasileira de Meteorologia, v. 26, n. 2, p. 273–278, 2011.

TEIXEIRA J.; GONZALEZ M. **Inteligência artificial e teoria de resolução de problemas**. Trans/Form/Ação, Marília , v. 6, p. 45-52, 1983.

TEIXEIRA, Daniel Souza. **Inteligência artificial aplicada à pesquisa de mercado e comunicação**. *Dissertação Doutorado*. UNIFESP. São Paulo, 2019

TIPLER, Paul – **Física para cientistas e engenheiros**, vol. II, 1995 - Livros Técnicos e Científicos Editora.

INTRUSUL. **TIPOS de termômetros e suas utilidades**. 2018. Disponível em: < <http://blog.instrusul.com.br/tipos-de-termometros-e-suas-utilidades/> >. Acesso em: 20 de março de 2021

TORRES, F. T. P. **Relações entre fatores climáticos e ocorrências de incêndios florestais na cidade de Juiz de Fora (MG)**. Caminhos de Geografia, v. 7, n. 18, p. 162–171, 2006.

TORRES, Fillipe Tamiozzo Pereira et al . **Correlações entre os elementos meteorológicos e as ocorrências de incêndios florestais na área urbana de Juiz de Fora, MG**. Rev. Árvore, Viçosa , v. 35, n. 1, p. 143-150, 2011.

TURBIANI, Renata. **Fumaça de queimadas é ameaça a saúde pública, ameaçam médicos.** BBC, 2019. Disponível em: < <https://www.bbc.com/portuguese/brasil-49430367>>. Acesso em: 20 de março de 2021.

VAREJÃO-SILVA, M. A. Meteorologia E Climatologia. **Journal of Chemical Information and Modeling**, v. 53, n. 9, p. 1689–1699, 2013.

VARELA, Carlos Alberto. **Análise de componentes principais.** Universidade Federal Rural do Rio de Janeiro, Rio de Janeiro, 2008. Disponível em: <<http://www.ufrj.br/institutos/it/deng/varella/Downloads/multivariada%20aplicada%20as%20ciencias%20agrarias/Aulas/analise%20de%20componentes%20principais.pdf>> Acesso em 2 de abr. de 2021

VICINI, Lorena **Análise multivariada da teoria à prática.** Orientador Adriano Mendonça Souza, - Santa Maria : UFSM, CCNE, , 2005.

YOUGUO Li, Haiyan Wu A, Henan 464000, China **Clustering Method Based on K-Means Algorithm ,Department of Computer Science Xinyang Agriculture College Xinyang.**

ZAMBELLO, Aline V. et al., Mazucato, Thiago(org). **Metodologia da pesquisa e do trabalho científico.** Penápolis: FUNEPE, 2018.

ZHANG, Zhongheng. **Introduction to machine learning: k-nearest neighbors.” *Annals of translational medicine* vol. 4,1.: 218. doi:10.21037.** Disponível em: <<https://atm.amegroups.com/article/view/10170/11310> > Acesso em 07 de jun. 2021

APÊNDICE A - DECLARAÇÃO DE ORIGINALIDADE

Eu, Asp Of. BM Antonio Matheus Vilar Guimarães Barbosa declaro para todos os fins que meu trabalho de fim de curso intitulado "*Incêndios em vegetação no município de Mirador-MA: O estudo de modelos preditivos baseados em dados atmosféricos*" é um documento original elaborado e produzido por mim.

Dados do Orientador:

Nome/Grau/Hierarquia: Prof. Dr. Mauro Sérgio Silva Pinto

Filiação/Instituição: Universidade Estadual do Maranhão

E-mail: maurosergiospinto@gmail.com

Telefone: (98) 9 8896-4585

DISCENTE

CPF: 05186566322