Fundamentação teórica:

Aqui está uma versão mais detalhada sobre a Inteligência Artificial (IA), abrangendo os tipos de aprendizado.

**1. Inteligência Artificial (IA)**

A Inteligência Artificial refere-se à simulação de processos de inteligência humana por máquinas, especialmente sistemas computacionais.

Ela envolve o desenvolvimento de algoritmos e modelos que permitem que as máquinas realizem tarefas que exigiriam a intervenção de um ser humano,

como raciocínio, reconhecimento de padrões, resolução de problemas, aprendizado e adaptação.

Russell, S., & Norvig, P. (2016). "Artificial Intelligence: A Modern Approach."

**Principais Áreas da IA:**

**Sistemas Especialistas:** São sistemas que replicam o processo de tomada de decisão de um especialista humano em uma área específica.

Eles usam regras baseadas em conhecimento e heurísticas para tomar decisões.

Rich, E., & Knight, K. (1991). "Artificial Intelligence."**Jurafsky, D., & Martin, J. H. (2021). "Speech and Language Processing."**

**Processamento de Linguagem Natural (PLN):** Envolve a capacidade de uma máquina de entender, interpretar e responder à linguagem

humana de maneira significativa. É amplamente utilizado em chatbots, assistentes virtuais e traduções automáticas.

**Visão Computacional:** Permite que uma máquina interprete e compreenda o mundo visual, como a detecção de objetos,

reconhecimento facial e análise de imagens.

Szeliski, R. (2010). "Computer Vision: Algorithms and Applications.

**Robótica:** A robótica envolve a criação de máquinas físicas ou virtuais que podem executar tarefas complexas no mundo

real de maneira automatizada, com autonomia e controle preciso.

Siciliano, B., & Khatib, O. (2016). Springer Handbook of Robotics.

### 2. Tipos de Aprendizado na IA

A Inteligência Artificial (IA) inclui a capacidade das máquinas de aprender com dados, um campo conhecido como aprendizado de máquina (machine learning) (Goodfellow et al., 2016).

Existem três categorias principais de aprendizado de máquina:

#### a) Aprendizado Supervisionado

No aprendizado supervisionado, o algoritmo é treinado com um conjunto de dados rotulado, ou seja, as entradas e saídas corretas são fornecidas ao modelo durante o treinamento. O objetivo é que a máquina aprenda um padrão e possa prever a saída correta para novos dados não vistos (Bishop, 2006).

**Exemplos de Algoritmos**: Regressão Linear, Regressão Logística, Árvores de Decisão, Redes Neurais (Hastie et al., 2009).

#### b) Aprendizado Não Supervisionado

No aprendizado não supervisionado, o modelo recebe dados sem rótulos e tenta encontrar padrões ou agrupamentos dentro deles. O objetivo é explorar a estrutura dos dados para entender sua organização e as relações entre os exemplos (Jain et al., 1999). Essa abordagem é útil para detecção de anomalias, agrupamento de dados e redução de dimensionalidade.

**Exemplos de Algoritmos**: Análise de Componentes Principais (PCA), K-Means, Algoritmo de Agrupamento Hierárquico (Kaufman & Rousseeuw, 2005).

#### c) Aprendizado por Reforço

No aprendizado por reforço, o algoritmo aprende por meio de interações com o ambiente. A máquina toma decisões e recebe feedback na forma de recompensas ou penalidades, com o objetivo de maximizar as recompensas ao longo do tempo. Este tipo de aprendizado é amplamente utilizado em robótica e em jogos, onde o agente explora ações para encontrar a melhor estratégia (Sutton & Barto, 2018).

**Exemplos de Algoritmos**: Q-Learning, Deep Q-Networks (DQN), Algoritmos Baseados em Políticas (Policy Gradient) (Mnih et al., 2015).

### 3. Outras Abordagens

Existem algumas técnicas derivadas que combinam aspectos de vários tipos de aprendizado:

#### Aprendizado Semi-Supervisionado

O aprendizado semi-supervisionado utiliza uma pequena quantidade de dados rotulados e uma grande quantidade de dados não rotulados para melhorar a precisão do modelo. Isso é particularmente útil quando rotular dados é caro ou demorado (Zhu & Goldberg, 2009).

#### Aprendizado por Transferência

O aprendizado por transferência aproveita o conhecimento adquirido em uma tarefa anterior para melhorar a performance em uma nova tarefa. É comum em redes neurais profundas, onde os pesos aprendidos em uma rede podem ser reutilizados para tarefas semelhantes (Pan & Yang, 2010).

### Referências

* Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.
* Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
* Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. Springer.
* Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data Clustering: A Review. ACM Computing Surveys, 31(3), 264-323.
* Kaufman, L., & Rousseeuw, P. J. (2005). Finding Groups in Data: An Introduction to Cluster Analysis. Wiley-Interscience.
* Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2015). Human-Level Control through Deep Reinforcement Learning. Nature, 518(7540), 529-533.
* Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345-1359.
* Sutton, R. S., & Barto, A. G. (2018). Reinforcement Learning: An Introduction. MIT Press.
* Zhu, X., & Goldberg, A. B. (2009). Introduction to Semi-Supervised Learning. Synthesis Lectures on Artificial Intelligence and Machine Learning, 3(1), 1-130.

### 4. Conexão com o Aprendizado Supervisionado

O aprendizado supervisionado é especialmente eficaz quando temos grandes quantidades de dados rotulados. Essa abordagem permite que o modelo estabeleça correlações entre as variáveis de entrada e a saída, criando previsões ou classificações com base nos padrões aprendidos (Hastie et al., 2009). É essa categoria de aprendizado que será abordada em mais detalhes ao falarmos sobre árvores de decisão, que são uma forma popular de aprendizado supervisionado (Breiman et al., 1986).

#### Escolha do Tipo de Aprendizado de Máquina

Para explicar a escolha do tipo de aprendizado de máquina, é importante considerar os diferentes tipos de variáveis:

1. **Variáveis Discretas** Variáveis discretas assumem valores finitos ou contáveis, representando contagens ou classificações. Uma característica importante é que não podem assumir valores intermediários entre os valores observados (Mendenhall et al., 2013).

**Exemplos**:

* + Número de filhos: Pode ser 0, 1, 2, mas não pode ser 1,5.
  + Quantidade de itens vendidos: Pode ser 10, 15, 20, mas não pode ser 12,3.

1. **Variáveis Contínuas** Variáveis contínuas podem assumir qualquer valor dentro de um intervalo, incluindo números fracionários. Elas são usadas para representar medições, onde qualquer valor em uma escala contínua é possível (Bliss, 2018). A principal característica é que existem infinitos valores possíveis entre quaisquer dois pontos.

**Exemplos**:

* + Altura de uma pessoa: Pode ser 170,2 cm, 170,25 cm, ou qualquer valor entre 170 e 171 cm.
  + Temperatura: Pode ser 25,4°C, 30,1°C, etc.

1. **Variáveis Categóricas** Variáveis categóricas, também conhecidas como qualitativas, representam grupos ou categorias. Elas não têm um valor numérico natural e são usadas para classificar ou categorizar elementos em diferentes grupos. As categorias podem ou não ter uma ordem entre si (Agresti, 2015).

**Tipos de Variáveis Categóricas**:

* + **Nominais**: Não têm uma ordem ou hierarquia entre as categorias. Exemplo: cor dos olhos (azul, verde, castanho), gênero (masculino, feminino).
  + **Ordinais**: Têm uma ordem ou hierarquia, mas a distância entre as categorias não é uniformemente mensurável. Exemplo: nível de escolaridade (fundamental, médio, superior), grau de satisfação (ruim, bom, ótimo).

### Referências

* Agresti, A. (2015). Foundations of Statistics. Wiley.
* Bliss, C. I. (2018). Statistics in Biology. Springer.
* Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1986). Classification and Regression Trees. Wadsworth.
* Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. Springer.
* Mendenhall, W., Beaver, R. J., & Beaver, B. M. (2013). Introduction to Probability and Statistics. Cengage Learning.

### Aprendizado Supervisionado e Variáveis Categóricas

No contexto específico de análise de incêndios, temos acesso a tabelas com históricos de incêndios que incluem a localização do ponto inicial e as datas e horários em que ocorreram. No entanto, para aplicar modelos de regressão mais completos, precisaríamos de dados adicionais, como umidade relativa do ar, índice de insolação, velocidade do vento e rajadas de vento, coletados precisamente nos locais onde os incêndios ocorreram. Infelizmente, essas informações não estão disponíveis na forma necessária.

Uma análise simples das estações de monitoramento atmosférico do Estado de São Paulo revela que são muito poucas as estações que coletam todos os dados essenciais para permitir o uso de variáveis contínuas. Portanto, optamos por utilizar variáveis categóricas, adotando índices padrão para identificar regiões onde incêndios podem ocorrer. Atribuímos um valor de 1 para a presença de condições necessárias e 0 para a ausência dessas condições.

Essa escolha nos leva à segunda parte da seleção do modelo a ser aplicado. Dado que trabalhamos com variáveis categóricas, o aprendizado supervisionado, especificamente por meio de árvores de decisão, se mostra como o modelo mais adequado. Essa abordagem é ainda mais pertinente em um projeto didático. Mesmo assim, as conclusões deste estudo demonstrarão que podemos utilizar este modelo para alcançar resultados válidos. Além disso, a melhoria na coleta de dados e na tecnologia de abstração de novos pontos de medição pode enriquecer significativamente o modelo, possibilitando a aplicação de redes neurais profundas para análise. Essa evolução poderia resultar em resultados excelentes e, se adotado em larga escala, poderia mitigar significativamente os incêndios e suas consequências para a sociedade.

Portanto, considerando que os dados disponíveis para os eventos não são contínuos, justificamos a escolha da árvore de decisão no aprendizado supervisionado como uma solução viável para a análise dos dados de incêndios.

### Referências

* Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1986). Classification and Regression Trees. Wadsworth.
* Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. Springer.

### Árvore de Decisão

Uma árvore de decisão é uma ferramenta de modelagem preditiva que utiliza uma estrutura em forma de árvore para tomar decisões. Ela é composta por nós de decisão (que representam perguntas ou condições), galhos (que representam as respostas possíveis) e folhas (que representam o resultado final, como uma classificação ou predição) (Breiman et al., 1986).

#### Exemplo Prático

Considere uma empresa de confecção de moda da Suécia que deseja otimizar o transporte de suas vendas de roupas de inverno. As variáveis que influenciam a decisão de venda incluem:

* **Estação do ano** (verão, inverno, etc.)
* **Temperatura média** (baixa, média, alta)
* **Demanda histórica** (alta, média, baixa)

Cada nó da árvore de decisão faz uma pergunta com base em uma dessas variáveis. Dependendo da resposta, a árvore divide os dados e continua a fazer perguntas até chegar a uma decisão final (vender ou não vender).

### Funcionamento de uma Árvore de Decisão

#### a) Geração de Dados

Os dados são obtidos a partir de registros históricos de vendas, clima e sazonalidade, além de informações sobre a demanda de roupas de inverno em diferentes regiões do mundo. A empresa pode coletar dados como:

* Localização (Hemisfério Sul, Norte ou Central)
* Estação do ano
* Temperatura média
* Demanda de roupas de inverno

Esses dados são organizados em uma tabela, onde cada linha representa uma observação (uma combinação de condições em um local específico).

#### b) Organização dos Dados

A tabela de dados pode se parecer com isso:

| Localização | Estação | Temperatura | Demanda | Venda |
| --- | --- | --- | --- | --- |
| Norte | Inverno | Baixa | Alta | 1 |
| Sul | Verão | Alta | Baixa | 2 |
| Central | Primavera | Média | Média | 2 |
| Norte | Verão | Alta | Baixa | 2 |
| Sul | Inverno | Baixa | Alta | 1 |

A variável "Venda" (1 para vender, 2 para não vender) é a variável alvo que a árvore de decisão tentará prever.

#### c) Construção da Árvore de Decisão

1. **Divisão Inicial**: A árvore começa no "nó raiz", que contém todas as observações. A primeira tarefa é escolher a variável que mais "divide" os dados, criando subconjuntos mais homogêneos em relação à variável alvo. A métrica usada para determinar isso pode ser o ganho de informação ou o índice Gini (Hastie et al., 2009).
   * Por exemplo, a árvore pode escolher a variável "Estação do ano" para a primeira divisão, observando que, em regiões onde é inverno, há uma maior probabilidade de venda.
2. **Divisões Recursivas**: Para cada subconjunto resultante da primeira divisão, o processo se repete. A árvore seleciona outra variável que melhor divide os dados, criando mais "nós" e "galhos".
   * Por exemplo, pode-se dividir os dados de "Inverno" com base na "Temperatura Média".
3. **Decisões Finais**: Ao final desse processo de divisão, a árvore terá folhas que correspondem às decisões finais de vender ou não vender. Por exemplo, se estivermos no Hemisfério Norte, no inverno, e a temperatura for baixa, a árvore pode prever que a empresa deve vender as roupas.

### Exemplo de Árvore Final

A árvore de decisão poderia ter a seguinte estrutura:

Estação

/ \

Inverno Verão

/ / \

Temperatura Não Vende Demanda

/ \ / \

Baixa Alta Alta Não Vende

/ \ \

Vende Não Vende Vende

### Outros Conceitos Importantes

#### a) Divisão dos Dados

O processo de construção da árvore começa com todo o conjunto de dados no "nó raiz". A primeira tarefa do algoritmo é dividir esse conjunto de dados em subconjuntos com base em uma variável preditiva que maximiza a separação ou purificação das classes (Breiman et al., 1986).

#### b) Escolha da Melhor Divisão

Para cada possível divisão, o algoritmo calcula uma métrica de qualidade da divisão, como o ganho de informação ou o índice Gini. A divisão que resulta na maior melhoria na pureza é escolhida, e esse processo é repetido recursivamente (Hastie et al., 2009).

#### c) Parada do Crescimento

O processo de divisão continua até que um critério de parada seja atendido. Isso pode acontecer quando:

* Todos os nós são "puros"
* O número de amostras em um nó cai abaixo de um limite mínimo
* A profundidade da árvore atinge um limite máximo

Esses critérios ajudam a evitar que a árvore cresça demais, o que poderia levar ao overfitting (quando o modelo se ajusta excessivamente aos dados de treinamento, perdendo capacidade de generalização) (James et al., 2013).

### Ganho de Informação

O ganho de informação quantifica a redução da incerteza após a divisão dos dados. É usado para determinar a qualidade de uma divisão ao escolher a variável preditiva mais relevante.

Fórmula:  de InformaGanho de Informac¸​a˜o=H(S)−∑i​∣S∣∣Si​∣​H(Si​)

Onde:

* H(S) é a entropia do conjunto de dados antes da divisão.
* H(Si​) é a entropia dos subconjuntos Si​ resultantes da divisão.

### Índice Gini

O índice Gini mede a impureza de um nó, ou seja, a probabilidade de que uma amostra seja classificada incorretamente se for selecionada aleatoriamente.

Fórmula:  GiniIˊndice Gini=1−∑k​pk2​

Onde pk​ é a proporção de elementos da classe k no nó.

### Vantagens e Desvantagens das Árvores de Decisão

**Vantagens**:

* **Interpretação fácil**: Intuitivas e facilmente visualizáveis.
* **Capacidade de lidar com variáveis categóricas e contínuas**: Não requerem transformação.
* **Sem necessidade de normalização**: Não precisam de normalização ou escalonamento dos dados.

**Desvantagens**:

* **Sensibilidade ao overfitting**: Podem se ajustar excessivamente aos dados de treinamento.
* **Instabilidade**: Pequenas alterações nos dados podem resultar em grandes mudanças na estrutura da árvore.
* **Menor desempenho em relação a outros algoritmos**: Podem não ser tão precisas quanto métodos mais complexos, como florestas aleatórias ou modelos de gradiente boosting.

### Referências

* Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1986). Classification and Regression Trees. Wadsworth.
* Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. Springer.
* James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning. Springer.

### Aplicações ao Projeto

Nosso projeto tem a intenção de gerenciar os recursos de combate a incêndios florestais por meio da análise de dados. A principal dificuldade que encontramos foi a obtenção de dados atmosféricos referentes à situação no momento em que os incêndios ocorreram. Isso foi o principal motivo pelo qual optamos pela Árvore de Decisão, devido à possibilidade de usar variáveis categóricas.

Nesse sentido, organizamos a tabela da seguinte forma:

**DADOS PARA ENTRADA**

* id\_municipio
* nome\_municipio
* latitude
* longitude
* ano
* sigla\_uf
* bioma
* data\_hora
* Precipitação Total < 10 mm
* Pressão Atmosférica entre 1015 e 1020 hPa
* Temperatura do Bulbo Seco ACIMA DE 30°C
* Temperatura do Ponto de Orvalho abaixo de 10°C
* Vento com velocidade maior que 30 km/h
* Rajada máxima > 10 m/s
* Umidade relativa do ar < 30%
* Radiação Solar acima de 4 kWh/m²
* Fogo

**EXEMPLO DE DADOS DE SAÍDA**

* 3501707
* Américo Brasiliense
* 21.7260; 48.0650
* 2013
* SP
* Cerrado
* 9/8/2013 17:01
* 1; Possibilidade de incêndio
* 1; Possibilidade de incêndio
* 1; Possibilidade de incêndio
* 1; Possibilidade de incêndio
* 1; Possibilidade de incêndio
* 1; Possibilidade de incêndio
* 1; Possibilidade de incêndio
* 1; Possibilidade de incêndio
* 1; Possibilidade de incêndio

Observamos que o valor 1 indica uma condição favorável ao incêndio, enquanto 0 indica uma condição desfavorável.

### Justificativas

**Precipitação Total < 10 mm (FAVORÁVEL AO INCÊNDIO)**

* Fonte: National Interagency Fire Center (NIFC). (2021). "Fire Weather: The Role of Weather in Wildfire." Este documento discute como a falta de precipitação, especialmente valores baixos como 10 mm, pode criar condições secas que aumentam a suscetibilidade a incêndios.
* Outra referência útil: Bradshaw, L. S., & McCormick, J. (2000). "Fire Weather and Fire Behavior." In Fire Management: A Comprehensive Review. Este capítulo analisa como diferentes níveis de umidade e precipitação impactam o comportamento do fogo.

**Pressão Atmosférica entre 1015 e 1020 hPa (FAVORÁVEL AO INCÊNDIO)**

* Fonte: Burgan, R. E., & Rothermel, R. C. (1984). "BEHAVE: Fire Behavior Prediction System." U.S. Department of Agriculture, Forest Service. Este documento discute como diferentes condições atmosféricas, incluindo a pressão, podem influenciar o comportamento do fogo e a propagação de incêndios.
* Fonte: Westerling, A. L., et al. (2006). "Warming and Earlier Spring Increase Western U.S. Forest Wildfire Activity." Science, 313(5789), 940-943. Este estudo analisa como variações na pressão atmosférica, junto com outros fatores climáticos, estão relacionadas ao aumento da atividade de incêndios florestais.

**Temperatura do Bulbo Seco ACIMA DE 30°C (FAVORÁVEL AO INCÊNDIO)**

* Fonte: Dimitrakopoulos, A. P., & Mitrakos, P. (2001). "Weather and Fire Danger in Greece: A Historical Perspective." International Journal of Wildland Fire, 10(3), 283-292. Este estudo analisa como altas temperaturas contribuem para a secagem da vegetação, aumentando a suscetibilidade a incêndios florestais.
* Fonte: McKenzie, D., et al. (2004). "Climate Change, Wildfire, and Conservation." Conservation Biology, 18(4), 880-892. Este artigo discute a relação entre o aumento das temperaturas e a frequência e intensidade de incêndios florestais.

**Temperatura do Ponto de Orvalho abaixo de 10°C (FAVORÁVEL AO INCÊNDIO)**

* Fonte: Boulain, N., et al. (2007). "Influence of Temperature and Humidity on Fire Behavior." International Journal of Wildland Fire, 16(5), 541-552. Este estudo analisa como a umidade do ar e o ponto de orvalho afetam a combustibilidade da vegetação.
* Fonte: Schroeder, M. J. (1971). "The Influence of Weather on Wildfire Occurrence." Journal of Forestry, 69(12), 847-851. Este artigo discute a relação entre a umidade relativa e o ponto de orvalho.

**Vento com velocidade maior que 30 km/h (FAVORÁVEL AO INCÊNDIO)**

* Fonte: National Interagency Fire Center (NIFC). (2019). "Fire Weather." Este documento descreve como ventos fortes podem aumentar a taxa de propagação de incêndios, tornando-os mais difíceis de controlar.
* Fonte: Rothermel, R. C. (1972). "A Mathematical Model for Fire Spread Rate in Wildland Fuels." USDA Forest Service Research Paper. Este trabalho apresenta um modelo matemático que demonstra como a velocidade do vento influencia a propagação do fogo.

**Rajada máxima > 10 m/s (FAVORÁVEL AO INCÊNDIO)**

* Fonte: National Wildfire Coordinating Group (NWCG). (2006). "Fire Weather." Este documento enfatiza que rajadas de vento intensas podem aumentar a taxa de propagação do fogo.
* Fonte: Anderson, H. E. (1982). "Aids to Determining Fuel Models for Estimating Fire Behavior." USDA Forest Service General Technical Report INT-GTR-122.

**Umidade relativa do ar < 30% (FAVORÁVEL AO INCÊNDIO)**

* Fonte: National Fire Protection Association (NFPA). (2018). "Fire Weather." Este documento menciona que uma umidade relativa abaixo de 30% pode indicar condições secas que aumentam a suscetibilidade à ignição.
* Fonte: Bradshaw, L. S., & McCormick, K. (2000). "Fire Weather." In "Fire Behavior." O estudo discute como baixos níveis de umidade relativa favorecem a secagem da vegetação.

**Radiação Solar acima de 4 kWh/m² (FAVORÁVEL AO INCÊNDIO)**

* Fonte: Rothermel, R. C. (1972). "A Mathematical Model for Predicting Fire Spread in Wildland Fuels." U.S. Forest Service. Este estudo aborda como a intensidade da radiação solar pode influenciar a temperatura da vegetação.
* Fonte: Jenkins, M. J., et al. (2004). "Fuel Moisture Dynamics in a Pinus ponderosa Forest." In "Fire Ecology."

Com essas variáveis, construímos uma Árvore de Decisão, que irá definir se, em um determinado local, existe a possibilidade de haver incêndio. Outro elemento importante é o fato de a região já ter sido atingida por incêndio, o que revela dois aspectos relevantes: primeiro, que o agente humano pode ter sido responsável pelo início do incêndio e pode estar presente novamente; segundo, que uma vegetação já atingida por incêndio resiste menos a novos incêndios. Assim, a proximidade de pontos de incêndio é um elemento crucial e está contemplada em nosso projeto.