

# Previsão de Queimadas: estudo de caso utilizando kNN, Random Forest e SVM

1<sup>st</sup> João Pedro Neigri Heleno *Engenharia da Computação*  
*Universidade Tecnológica Federal do Paraná*  
 Apucarana, Brasil  
 heleno@alunos.utfpr.edu.br

**Abstract**—This case study investigates the application of kNN (K-Nearest Neighbors), Random Forest, and SVM (Support Vector Machines) techniques to predict wildfires using the Nesterov Index. The main objective is to implement each model to forecast wildfire occurrences and analyze their efficiency, comparing their performance.

**Index Terms**—Wildfires, Nesterov Index, KNN, Random Forest, SVM

## I. INTRODUÇÃO

Atualmente, o avanço da inteligência artificial tem permitido atender diversas demandas sociais de forma mais eficiente, facilitando a realização de tarefas e a previsão de eventos. Diante desse cenário, este projeto foi motivado pela possibilidade de aplicar algoritmos de inteligência artificial para prever e mitigar um problema que afeta cada vez mais nossa sociedade: as queimadas.

A região de Castro, no Paraná, foi escolhida para investigar as queimadas devido à sua significativa incidência desses eventos no estado. Além disso, essa área abriga diversas unidades de conservação, como o Parque Nacional dos Campos Gerais, que protegem remanescentes de Floresta Ombrófila Mista, campos nativos e capões com Floresta de Araucárias, habitats de várias espécies ameaçadas de extinção.

O estudo tem como objetivo, por meio do uso do Índice de Nesterov, treinar e comparar os algoritmos *k-Nearest Neighbors* (kNN), *Random Forest* e *Support Vector Machines* (SVM). O Índice de Nesterov é uma métrica amplamente utilizada para avaliar o risco de incêndios florestais, considerando fatores como temperatura e umidade relativa, que são essenciais para prever a probabilidade de queimadas.

Os algoritmos escolhidos desempenham papéis específicos nesse processo:

- *k-Nearest Neighbors* (kNN): Um modelo baseado em instâncias, que classifica os dados com base na proximidade a pontos conhecidos, sendo útil em situações onde os dados apresentam estrutura clara [1].
- *Random Forest*: Um método baseado em árvores de decisão que combina múltiplas árvores para aumentar a precisão e a robustez nas previsões [3].
- *Support Vector Machines* (SVM): Um algoritmo que utiliza hiperplanos para separar classes de dados de forma eficiente, especialmente útil para problemas de classificação binária [2].

Ao comparar os resultados obtidos por cada algoritmo, o estudo busca avaliar qual deles apresenta melhor desempenho na previsão de queimadas, considerando métricas como precisão, sensibilidade e capacidade de generalização. Essa análise comparativa contribuirá para identificar a abordagem mais eficaz na utilização de inteligência artificial para mitigação de riscos ambientais.

## II. TRABALHOS RELACIONADOS

Durante o desenvolvimento deste estudo, foram identificados diversos trabalhos que contribuem para a previsão de incêndios florestais utilizando índices meteorológicos e técnicas de aprendizado de máquina. A relação entre condições climáticas e ocorrência de incêndios tem sido amplamente estudada, e abordagens como o uso do Índice de Nesterov e modelos supervisionados de aprendizado de máquina têm demonstrado eficácia na identificação de áreas de risco.

Os trabalhos apresentados a seguir servem como base para este projeto, permitindo um aprofundamento na relação entre fatores ambientais e a previsão de incêndios, além de possibilitar a comparação da eficácia de diferentes modelos preditivos na análise do risco de queimadas.

### A. Nesterov Ignition Index

O Índice de Nesterov é um dos métodos mais utilizados na avaliação do risco de incêndios florestais. Ele considera a temperatura do ar e o déficit de umidade para estimar a probabilidade de ignição, ignorando diretamente o impacto da precipitação acumulada. Sua simplicidade e eficiência tornaram esse índice amplamente aplicado em diferentes regiões, servindo como um dos principais parâmetros para monitoramento de incêndios e planejamento de estratégias preventivas. Este estudo incorpora o Índice de Nesterov como um dos principais atributos para alimentar os modelos de aprendizado de máquina na previsão de incêndios.

### B. A Forest Fire Prediction Model Based on Meteorological Factors and Machine Learning

Este artigo apresenta um modelo preditivo baseado em fatores meteorológicos e técnicas de aprendizado de máquina, com destaque para o Random Forest. O estudo analisa variáveis como temperatura, umidade relativa, velocidade do vento e precipitação para estimar a probabilidade de incêndios.

Os resultados demonstram que o uso de aprendizado de máquina melhora significativamente a precisão das previsões quando comparado a métodos tradicionais. O modelo Random Forest, em particular, se mostrou eficaz na identificação de padrões complexos nos dados, o que reforça sua utilidade na previsão de incêndios.

### C. Predicting Wildfires in Algerian Forests Using Machine Learning Models

Neste estudo, diferentes modelos de aprendizado de máquina são testados para prever incêndios em florestas da Argélia, com foco em Support Vector Machines (SVM) e Random Forest. O artigo destaca a importância de fatores meteorológicos na predição de incêndios e demonstra que o SVM pode ser uma abordagem promissora para a classificação do risco de queimadas. A análise compara o desempenho dos algoritmos e ressalta como modelos baseados em aprendizado de máquina podem superar índices meteorológicos tradicionais ao incorporar dados históricos e variações sazonais.

## III. METODOLOGIA

### A. Base de Dados e Aquisição dos Dados

Para a realização deste estudo, foram utilizados dados meteorológicos históricos disponibilizados pelo Instituto Nacional de Meteorologia (INMET). O período analisado compreende os anos de 2007 a 2020, permitindo a identificação de padrões climáticos relevantes para a previsão de queimadas na região dos Campos Gerais, Paraná.

A base de dados contempla diversas variáveis meteorológicas, incluindo precipitação, pressão atmosférica, radiação solar, temperatura de bulbo seco, temperatura de ponto de orvalho, temperatura média e umidade relativa do ar. Além dessas variáveis, algumas métricas derivadas foram incorporadas para enriquecer a análise preditiva. A precipitação acumulada foi calculada utilizando uma janela deslizante de cinco dias, permitindo que o modelo capture padrões recentes de chuva ao longo do tempo. O déficit de umidade foi obtido a partir da diferença entre a temperatura média e a temperatura do ponto de orvalho, garantindo uma melhor estimativa da secura do ambiente. Além disso, a precipitação acumulada foi ajustada para considerar reduções progressivas em períodos de alta precipitação, minimizando efeitos de valores extremos.

Uma variável categórica adicional foi incluída para representar a estação do ano, permitindo avaliar a influência sazonal na ocorrência de queimadas. Essa variável foi codificada numericamente conforme a Tabela I, onde cada estação recebe um valor específico.

Tabela I: Relação entre valores numéricos e estações do ano.

Valor numérico	Estação do ano
0	Verão
1	Outono
2	Inverno
3	Primavera

Fonte: Autoria própria (2025).

### B. Índice de Nesterov e sua Aplicação

O Índice de Nesterov é um indicador amplamente utilizado para avaliar o risco de incêndios florestais. Ele considera variáveis meteorológicas, como a temperatura e o déficit de umidade, para estimar a probabilidade de ignição do fogo [8]. Diferentemente de outros índices, o Índice de Nesterov ignora diretamente a precipitação acumulada, focando apenas na capacidade atmosférica de promover a secagem da vegetação e, consequentemente, aumentar o risco de incêndios.

A fórmula do Índice de Nesterov ajustado utilizada neste estudo é definida como:

$$N = D \cdot (P + 1) \cdot F_m \cdot C \quad (1)$$

Nesta equação,  $N$  representa o Índice de Nesterov ajustado. O valor de  $D$  corresponde ao déficit de umidade, definido como a diferença entre a temperatura média do ar ( $T_{med}$ ) e a temperatura de ponto de orvalho ( $T_{orvalho}$ ), sendo limitado a um valor mínimo de zero para evitar interpretações negativas:

$$D = \max(T_{med} - T_{orvalho}, 0) \quad (2)$$

O termo  $P$  refere-se à precipitação acumulada ao longo de uma janela deslizante de cinco dias, permitindo suavizar variações bruscas e evitar que eventos de chuva intensa reduzam drasticamente o valor do índice. Para capturar a influência sazonal, o fator  $F_m$  ajusta o índice conforme o mês do ano, refletindo variações climáticas típicas de cada período. O fator  $C$  atua como um parâmetro de escala, garantindo que os valores do índice estejam dentro de uma faixa apropriada para análise.

Com o índice calculado, foi necessário categorizar os valores em diferentes faixas de risco. Essas categorias foram estimadas com base nos quartis da distribuição dos valores históricos do índice, garantindo uma segmentação estatisticamente coerente. Os dados indicam que aproximadamente 25% das observações estão abaixo de 400, enquanto 50% dos valores ficam abaixo de 1000 e 75% não ultrapassam 1894. Assim, os limites foram definidos para refletir de forma mais realista o perigo de incêndios. A classificação adotada segue a Tabela II.

Tabela II: Classificação do risco de incêndio com base no Índice de Nesterov

Faixa do Índice de Nesterov	Classificação do Risco
$N < 400$	Baixo Risco
$400 \leq N \leq 1000$	Médio Risco
$1000 < N \leq 2500$	Alto Risco
$N > 2500$	Risco Crítico

Fonte: Autoria própria (2025).

A categorização do risco permite que os modelos de aprendizado de máquina operem sobre uma variável categórica, facilitando a interpretação dos resultados e tornando a previsão mais acessível para análise e tomada de decisão. A inclusão dessas faixas melhora a compreensão sobre quais períodos apresentam maior potencial para incêndios e possibilita a implementação de estratégias preventivas mais eficazes.

### C. Análise Exploratória

Antes da aplicação dos modelos, foi realizada uma análise exploratória dos dados para compreender melhor o comportamento das variáveis. Foram gerados diversos gráficos para visualizar a relação entre temperatura, umidade e Índice de Nesterov ao longo dos meses e anos. Essa análise permitiu identificar padrões sazonais e compreender como as condições meteorológicas influenciam o risco de incêndios, fornecendo informações importantes para o ajuste dos modelos de aprendizado de máquina [5].

### D. Tratamento de Valores Ausentes

Durante a análise exploratória dos dados, foi identificado que algumas variáveis meteorológicas apresentavam valores ausentes em determinados períodos. Para evitar que esses dados incompletos comprometessem o desempenho dos modelos de aprendizado de máquina, foi adotado um método de preenchimento baseado na média histórica.

A estratégia utilizada considera o comportamento da variável ao longo dos anos, preenchendo os valores ausentes com a média registrada para o mesmo dia e horário em anos anteriores. Esse método garante que os valores preenchidos preservem a sazonalidade e as tendências climáticas da região, evitando vieses que poderiam distorcer as previsões.

A utilização desse método permitiu que os modelos de aprendizado de máquina operassem com um conjunto de dados mais completo e confiável. Ao substituir os valores ausentes por médias históricas coerentes, foi possível minimizar distorções nos padrões climáticos e melhorar a generalização das previsões de risco de incêndio, tornando o modelo mais robusto e adaptado às variações sazonais.

### E. Implementação dos Modelos de Aprendizado de Máquina

Para realizar a previsão das faixas de risco de incêndio com base no Índice de Nesterov, foram testados três modelos de aprendizado de máquina: *k-Nearest Neighbors (kNN)*, *Random Forest* e *Support Vector Machines (SVM)*. A implementação foi realizada utilizando a biblioteca *Scikit-Learn*.

O conjunto de dados foi carregado a partir de um arquivo CSV e passou por uma fase de pré-processamento, onde foram removidas as colunas irrelevantes para a previsão, como data e hora. A variável alvo utilizada foi a *faixa de risco*, que classifica o risco de incêndio em quatro categorias.

Os dados foram divididos em 80% para treinamento e 20% para teste, mantendo a distribuição das classes com a técnica de estratificação. Além disso, foi aplicada a normalização dos dados preditores com o método *StandardScaler*, garantindo que todas as variáveis estivessem na mesma escala, um requisito essencial para modelos baseados em distância, como *kNN* e *SVM*.

1) *k-Nearest Neighbors (kNN)*: O modelo *k-Nearest Neighbors (kNN)* foi escolhido devido à sua simplicidade e capacidade de capturar padrões locais nos dados. Como o algoritmo calcula a distância entre os pontos para classificar as amostras, foi aplicada normalização para evitar que variáveis com escalas diferentes influenciassem o resultado.

Para encontrar o valor ideal de  $k$ , foram realizados testes variando  $k$  entre 7 e 11, avaliando o desempenho do modelo para cada valor. O objetivo foi identificar um ponto de equilíbrio entre *overfitting* (valores muito pequenos de  $k$ ) e *underfitting* (valores muito grandes de  $k$ ). O modelo final foi treinado com o valor de  $k$  que apresentou melhor desempenho nos testes. A avaliação foi realizada utilizando a acurácia e a matriz de confusão, permitindo analisar o desempenho do modelo em cada classe de risco.

2) *Random Forest*: O modelo *Random Forest* foi escolhido devido à sua robustez e capacidade de lidar com dados não lineares. Como algoritmos baseados em árvores de decisão não são sensíveis à escala das variáveis, os dados foram utilizados sem necessidade de normalização.

Para otimizar os hiperparâmetros, foi utilizada a técnica de *Grid Search*, variando a quantidade de árvores na floresta entre 100 e 500 e ajustando a profundidade máxima entre 10 e 50. O objetivo foi encontrar a melhor configuração para maximizar a capacidade de generalização do modelo. A avaliação foi realizada utilizando acurácia e matriz de confusão, garantindo uma análise detalhada do desempenho do classificador.

3) *Support Vector Machines (SVM)*: O modelo *Support Vector Machines (SVM)* foi implementado com o objetivo de capturar padrões complexos na separação das classes [2]. Como o desempenho do *SVM* pode ser afetado por escalas diferentes entre as variáveis, os dados foram normalizados antes do treinamento.

Foram testados diferentes tipos de *kernels* (linear, polinomial e radial) e ajustado o parâmetro de regularização  $C$ , buscando um equilíbrio entre margem de separação e erro de classificação. O modelo final foi escolhido com base nos resultados obtidos durante os testes. Assim como nos demais modelos, a avaliação foi realizada utilizando a acurácia e matriz de confusão, permitindo identificar possíveis dificuldades do modelo em classificar corretamente as faixas de risco mais altas.

## IV. RESULTADOS OBTIDOS

### A. Análise Exploratória dos Dados

A análise exploratória dos dados (EDA) foi realizada para compreender o comportamento das variáveis meteorológicas ao longo dos anos, destacando padrões sazonais e tendências relacionadas ao Índice de Nesterov. Foram gerados gráficos para visualizar a evolução da temperatura média, umidade relativa e Índice de Nesterov, permitindo uma melhor compreensão dos fatores que influenciam o risco de queimadas.

Fig. 1: Regressão linear da temperatura média ao longo dos anos

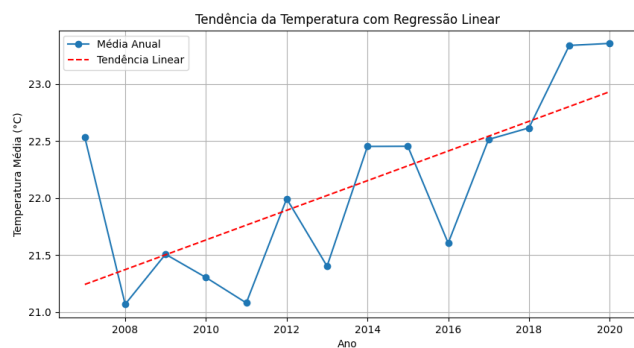
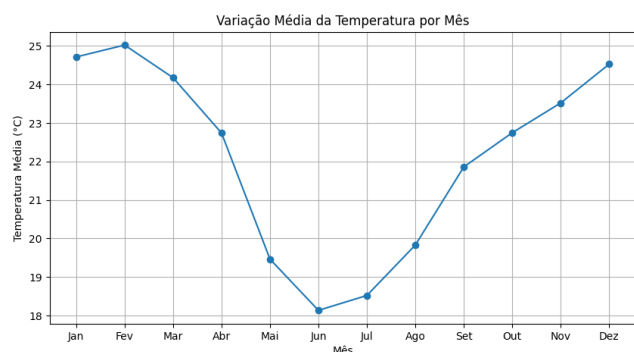


Fig. 2: Variação média do Índice de Nesterov por mês

Observa-se uma tendência de aumento da temperatura, com uma inclinação estimada de aproximadamente  $0.1300^{\circ}\text{C}$  por ano. Esse crescimento gradual indica uma possível influência de mudanças climáticas ou variações sazonais ao longo do tempo, reforçando a necessidade de monitoramento contínuo das condições meteorológicas para entender seus impactos sobre o ambiente e o risco de incêndios florestais.

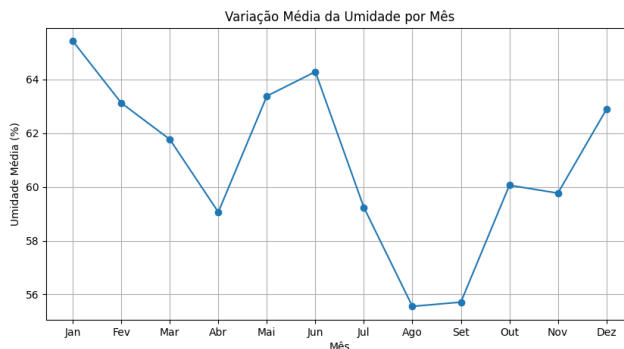
Fig. 3: Variação média da temperatura ao longo do ano



Fonte: Autoria própria (2025).

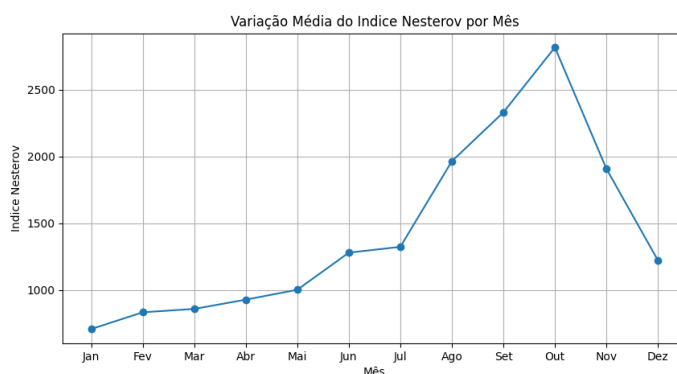
A variação da temperatura ao longo do ano apresenta um comportamento cíclico, com os meses de inverno exibindo temperaturas mais baixas, enquanto os meses de verão possuem temperaturas significativamente mais elevadas. Esses padrões refletem as mudanças sazonais características da região analisada.

Fig. 4: Variação média da umidade ao longo do ano



Fonte: Autoria própria (2025).

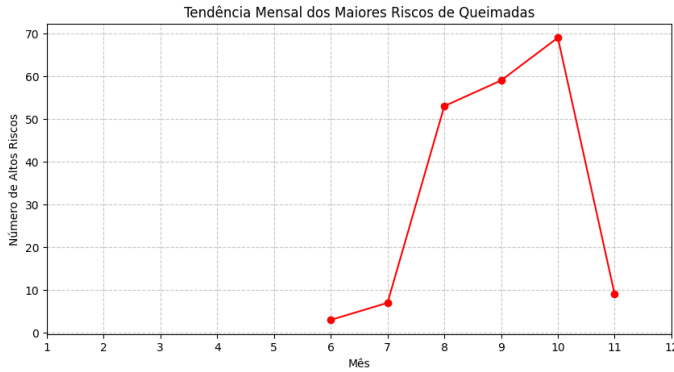
Fig. 5: Variação média do Índice de Nesterov por mês



Fonte: Autoria própria (2025).

O comportamento do Índice de Nesterov ao longo dos meses apresenta uma clara variação sazonal, refletindo a influência das condições meteorológicas na estimativa do risco de incêndios. Durante os primeiros meses do ano, os valores do índice permanecem relativamente baixos, acompanhando períodos de maior umidade e temperaturas moderadas. Conforme a transição para os meses mais secos ocorre, especialmente a partir de julho, observa-se um aumento expressivo do índice, atingindo seu pico entre setembro e outubro, quando a combinação de temperaturas elevadas e baixa umidade cria condições mais favoráveis para a ocorrência de queimadas. Nos meses finais do ano, os valores do índice tendem a diminuir, refletindo a recuperação da umidade e a redução das temperaturas, fatores que contribuem para a diminuição do risco de incêndios florestais.

Fig. 6: Distribuição mensal dos 200 maiores Índices de Nesterov



Fonte: Autoria própria (2025).

A análise dos 200 maiores valores do Índice de Nesterov ao longo do período observado revela um forte padrão sazonal, com a maioria dos eventos de alto risco concentrados entre os meses de agosto e outubro. Esse comportamento está diretamente relacionado à redução da umidade relativa e ao aumento das temperaturas nesses meses, criando condições propícias para a ocorrência de incêndios. Além disso, observa-se que os maiores valores do índice ocorrem de forma recorrente em determinados períodos do ano, reforçando a importância do monitoramento contínuo e da adoção de estratégias preventivas durante as épocas mais críticas.

#### B. Desempenho do $k$ -Nearest Neighbors ( $kNN$ )

Foram testados diferentes valores de  $k$  entre 7 e 11 para identificar a melhor configuração para o modelo. A seguir, são apresentados os principais resultados obtidos.

Tabela III: Acurácia para diferentes valores de  $k$

Valor de $k$	Acurácia
7	0.7840
8	0.7810
9	0.7791
10	0.7781
11	0.7781

Fonte: Autoria própria (2025).

A Tabela III apresenta a acurácia obtida para cada valor de  $k$ . O maior valor de acurácia foi obtido com  $k$  igual a 7, atingindo 78,40%, enquanto valores maiores de  $k$  apresentaram uma leve redução no desempenho.

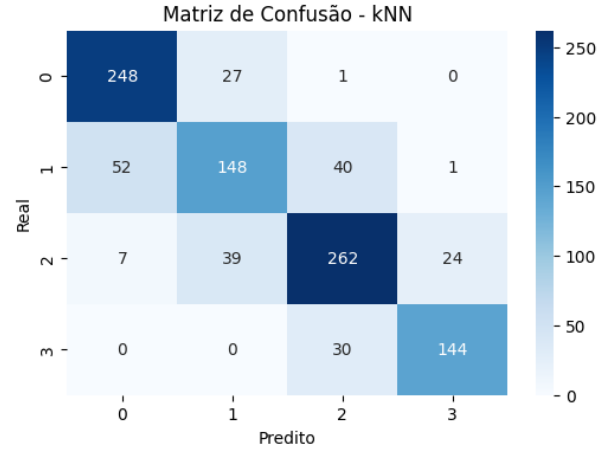
Tabela IV: Métricas de avaliação para  $k = 7$

Classe	Precisão	Recall	F1-score
Baixo risco	0.81	0.90	0.85
Médio risco	0.69	0.61	0.65
Alto risco	0.79	0.79	0.79
Risco crítico	0.85	0.83	0.84

Fonte: Autoria própria (2025).

A Tabela IV apresenta as métricas de desempenho (Precisão, Recall e F1-score) para cada classe de risco no modelo para  $k$  com o maior acurácia.

Fig. 7: Matriz de confusão para o modelo  $kNN$  ( $k$ )



Fonte: Autoria própria (2025).

A Figura 7 apresenta a matriz de confusão do modelo para  $k$  igual a 7, evidenciando as taxas de erro entre as classes.

O modelo  $kNN$  com  $k$  igual a 7 apresentou melhor desempenho na classificação da classe de baixo risco, com um F1-score de 0.85 e recall de 90%. Por outro lado, a classe de médio risco teve o pior desempenho, com recall de 61%, indicando que o modelo teve dificuldade em diferenciá-la das demais.

A matriz de confusão mostra que a maioria dos erros ocorre na transição entre médio risco e alto risco, sugerindo que essas categorias possuem características meteorológicas semelhantes, tornando a distinção mais difícil para o algoritmo.

#### C. Desempenho do Random Forest

O modelo de Random Forest foi testado com diferentes quantidades de árvores para identificar a melhor configuração para previsão do risco de incêndios. O modelo final foi treinado com 200 árvores, profundidade máxima de 10 e um mínimo de 5 amostras para divisão dos nós, garantindo um bom equilíbrio entre complexidade e generalização [3].

Tabela V: Acurácia do modelo Random Forest otimizado

Modelo	Acurácia
Random Forest	0.8641

Fonte: Autoria própria (2025).

A Tabela V apresenta a acurácia obtida para o modelo otimizado, que atingiu 86,41%, um desempenho superior ao obtido pelo  $kNN$ .

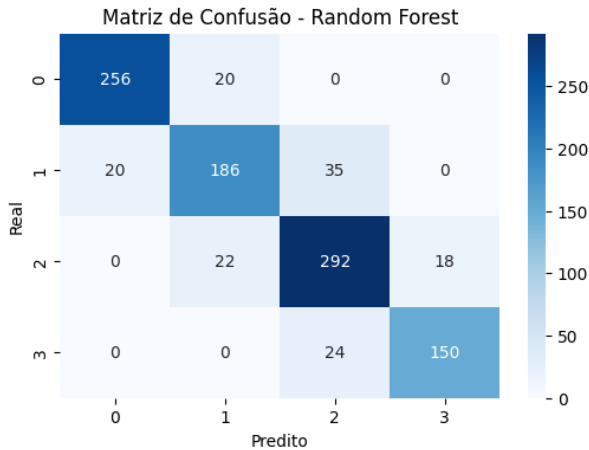
Tabela VI: Métricas de avaliação do Random Forest

Classe	Precisão	Recall	F1-score
Baixo risco	0.93	0.93	0.93
Médio risco	0.82	0.77	0.79
Alto risco	0.83	0.88	0.86
Risco crítico	0.89	0.86	0.88

Fonte: Autoria própria (2025).

A Tabela VI apresenta as métricas de desempenho (Precisão, Recall e F1-score) para cada classe de risco no modelo de Random Forest. Observa-se que a classe baixo risco apresentou o melhor desempenho, com F1-score de 0.93, enquanto a classe de médio risco teve um recall de 0.77, sugerindo que alguns exemplos dessa classe foram classificados incorretamente.

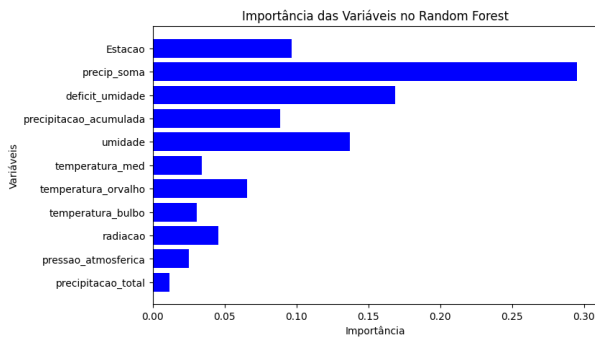
Fig. 8: Matriz de confusão para o modelo Random Forest



Fonte: Autoria própria (2025).

A Figura 8 apresenta a matriz de confusão do modelo Random Forest, evidenciando as taxas de erro entre as classes. Em comparação ao kNN, o modelo reduziu os erros na classificação de alto risco e risco crítico, demonstrando melhor capacidade de separação entre essas categorias.

Fig. 9: Importância das variáveis no modelo Random Forest



Fonte: Autoria própria (2025).

A Figura 9 apresenta a importância relativa das variáveis utilizadas pelo modelo Random Forest. Observa-se que a precipitação ajustada foi a variável mais relevante para a previsão do risco de incêndios, seguido por fatores como deficit de umidade e umidade.

#### D. Desempenho do Support Vector Machine (SVM)

O modelo de Support Vector Machine (SVM) foi configurado com o kernel radial básico (RBF), parâmetro de regularização C igual a 1.0 e fator gamma = "scale",

garantindo um ajuste equilibrado entre complexidade e capacidade de generalização. O modelo foi treinado com random\_state igual a 42.

Tabela VII: Acurácia do modelo SVM

Modelo	Acurácia
SVM	0.8534

Fonte: Autoria própria (2025).

A Tabela VII apresenta a acurácia obtida para o modelo, que atingiu 85,34%.

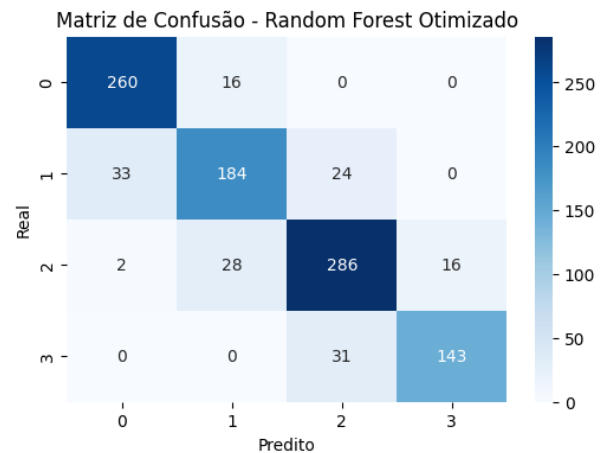
Tabela VIII: Métricas de avaliação do SVM

Classe	Precisão	Recall	F1-score
Baixo risco	0.88	0.94	0.91
Médio risco	0.81	0.76	0.78
Alto risco	0.84	0.86	0.85
Risco crítico	0.90	0.82	0.86

Fonte: Autoria própria (2025).

A Tabela VIII apresenta as métricas de desempenho (Precisão, Recall e F1-score) para cada classe de risco no modelo de SVM. A classe baixo risco teve o melhor desempenho, com um F1-score de 0.91 e recall de 94%, enquanto a classe médio risco obteve um recall menor, indicando que algumas instâncias dessa classe foram classificadas incorretamente.

Fig. 10: Matriz de confusão para o modelo SVM



Fonte: Autoria própria (2025).

A Figura 10 apresenta a matriz de confusão do modelo SVM, evidenciando as taxas de erro entre as classes.

O desempenho do modelo SVM mostra que ele consegue capturar padrões complexos nos dados, diferenciando bem as classes de baixo risco e alto risco. No entanto, a classe médio risco apresentou dificuldades, possivelmente devido à proximidade de seus valores com outras classes.

#### E. Comparação dos Modelos

Após a análise individual de cada modelo de aprendizado de máquina, esta seção apresenta uma comparação direta entre *k-Nearest Neighbors* (kNN), *Random Forest* e *Support Vector Machines* (SVM), considerando suas métricas de desempenho.

Tabela IX: Comparação das métricas dos modelos

Modelo	Acurácia	Macro Recall	Macro F1-score
kNN	0.7840	0.78	0.78
Random Forest	0.8641	0.86	0.86
SVM	0.8534	0.85	0.85

Fonte: Autoria própria (2025).

A Tabela IX apresenta um resumo das métricas de desempenho dos três modelos. O *Random Forest* obteve a melhor acurácia, atingindo 86,41%, seguido pelo *SVM* com 85,34% e pelo *kNN*, que apresentou a menor taxa de acertos, com 78,40%.

A análise das matrizes de confusão revelou que a classe médio risco foi a mais difícil de ser corretamente classificada em todos os modelos, sendo frequentemente confundida com as classes adjacentes. Esse comportamento sugere que os dados dessa categoria podem compartilhar características similares com as demais, dificultando a separação exata pelos algoritmos.

## V. CONCLUSÃO

Este estudo investigou a aplicação de algoritmos de aprendizado de máquina para prever o risco de incêndios florestais utilizando o Índice de Nesterov como principal variável meteorológica. Foram testados três modelos distintos: *k-Nearest Neighbors* (kNN), *Random Forest* e *Support Vector Machines* (SVM), com o objetivo de avaliar sua eficácia na classificação do risco de incêndios em diferentes faixas.

Os resultados mostraram que o modelo *Random Forest* apresentou o melhor desempenho geral, atingindo uma acurácia de 86,41%. Esse modelo demonstrou alta robustez na diferenciação entre as classes, especialmente para os níveis de alto risco e risco crítico. O *SVM* também obteve um bom desempenho, com acurácia de 85,34%, enquanto o *kNN* apresentou a menor taxa de acertos, com acurácia de 78,40%. Observou-se que os modelos baseados em árvores, como o *Random Forest*, foram mais eficazes na separação das classes, o que pode estar relacionado à sua capacidade de capturar relações complexas nos dados.

A análise da matriz de confusão indicou que a classe de médio risco foi a mais difícil de classificar corretamente, sendo frequentemente confundida com as classes adjacentes. Esse problema pode estar associado à similaridade dos valores meteorológicos entre essas faixas de risco, tornando a distinção mais desafiadora para os algoritmos.

Embora os resultados tenham sido satisfatórios, há espaço para aprimoramentos. Trabalhos futuros podem incluir a aplicação de técnicas de balanceamento de classes para reduzir a assimetria entre as categorias, bem como o uso de modelos mais sofisticados, como redes neurais. Além disso, a incorporação de outros índices de incêndios como fórmula de Monte Alegre e possivelmente variáveis como velocidade do vento e índice de seca, pode fornecer uma visão mais abrangente das condições ambientais que influenciam o risco de incêndios.

Dessa forma, este estudo reforça a viabilidade do uso de aprendizado de máquina na previsão de queimadas e destaca

a importância do Índice de Nesterov como uma métrica relevante para essa aplicação. A implementação desses modelos pode contribuir significativamente para estratégias de monitoramento e prevenção de incêndios florestais.

## REFERÊNCIAS

- [1] MARS LAND, Stephen. Machine Learning: An Algorithmic Perspective. 2. ed. Boca Raton: CRC Press, 2015. ISBN 978-1-4665-8333-7.
- [2] CORTES, Corinna; VAPNIK, Vladimir. Support-vector networks. Machine Learning, v. 20, n. 3, p. 273–297, 1995.
- [3] BREIMAN, Leo. Random forests. Machine Learning, v. 45, n. 1, p. 5–32, 2001.
- [4] BROCKWELL, Peter J.; DAVIS, Richard A. Introduction to Time Series and Forecasting. 3. ed. Cham: Springer, 2016. ISBN 978-3-319-29852-8.
- [5] NIELSEN, Aileen. Practical Time Series Analysis: Prediction with Statistics and Machine Learning\*. Beijing: O'Reilly, 2019.
- [6] CHOI, Seungcheol; SON, Minwoo; KIM, Changgyun; KIM, Byungsik. A Forest Fire Prediction Model Based on Meteorological Factors and the Multi-Model Ensemble Method. Forests, v. 15, n. 1981, 2024.
- [7] ZAIDI, Abdelhamid Taieb. Predicting wildfires in Algerian forests using machine learning models. Heliyon, v. 9, e18064, 2023.
- [8] BORGES, Karla M.; SOARES, Rômulo V.; BATISTA, Antônio C. Análise da eficiência de índices de perigo de incêndios florestais para o estado do Espírito Santo. Revista Árvore, v. 35, n. 3, p. 601–608, 2011.