



**MODELAGEM PREDITIVA DE VISIBILIDADE UTILIZANDO O ALGORITMO  
RANDOM FOREST: UMA ABORDAGEM BASEADA EM DADOS  
METEOROLÓGICOS DE REGIÕES COSTEIRAS**

***PREDICTIVE MODELING OF VISIBILITY USING THE RANDOM FOREST  
ALGORITHM: A DATA-DRIVEN APPROACH BASED ON METEOROLOGICAL  
DATA FROM COASTAL REGIONS***

**FLAVIA APARECIDA DO VALLE ESCORCIO** (FATEC  
RUBENS LARA)  
flavia.escorcio@fatec.sp.gov.br

**IGOR SILVA DE CARVALHO** (FATEC RUBENS LARA)  
igor.carvalho27@fatec.sp.gov.br

**LUCAS GABRIEL DE OLIVEIRA LAUS** (FATEC  
RUBENS LARA)  
lucas.laus@fatec.sp.gov.br

**LUIS HENRIQUE MOREIRA BARBOSA** (FATEC  
RUBENS LARA)  
luis.barbosa15@fatec.sp.gov.br

**PROF. DR. JOÃO PAULO DE MELLO** (FATEC RUBENS  
LARA)  
joao.mello12@fatec.sp.gov.br

**RESUMO**

A baixa visibilidade nas regiões costeiras é uma problemática constante, a qual reflete na segurança e eficiência das operações aeroportuárias. O presente trabalho procurou desenvolver um modelo de Random Forest visando estimar eventos de baixa visibilidade nessas regiões por meio de variáveis meteorológicas históricas. Desta maneira, a pesquisa foi realizada através de uma abordagem quantitativa e exploratória cujos dados foram oriundos da REDEMET (Departamento de Controle do Espaço Aéreo - DECEA) e da reanálise ERA5 (European Centre for Medium-Range Weather Forecasts -ECMWF). Os resultados preliminares apontaram que para um hora de previsão, o modelo atingiu o  $R^2$  de 0.689.

**PALAVRAS-CHAVE:** Visibilidade; *Random Forest*; Dados Meteorológicos; Aeroportuária; Aprendizado de Máquina.

**ABSTRACT**

*Low visibility in coastal regions represents a persistent challenge, impacting the safety and efficiency of airport operations. This study aimed to develop a Random*



*Forest model to estimate low visibility events in these areas using historical meteorological variables. Accordingly, the research followed a quantitative and exploratory approach, utilizing data from REDEMET (Airspace Control Department – DECEA) and the ERA5 reanalysis (European Centre for Medium-Range Weather Forecasts – ECMWF). Preliminary results indicated that for a one-hour-ahead forecast, the model achieved an  $R^2$  of 0.689.*

**KEYWORDS:** *Visibility, Random Forest, Meteorological Data, Port and airport, Machine Learning*

## 1 INTRODUÇÃO

A crescente ocorrência de fenômenos climáticos extremos, como o nevoeiro, impacta negativamente setores da infraestrutura global, incluindo portos e aeroportos (GU, 2024). A cidade de Santos, uma das áreas costeiras mais vulneráveis às mudanças climáticas (KIGUTI e AVANCINI, 2022), sofre com as consequências desse fenômeno. No Porto de Santos, o maior da América Latina, o nevoeiro causou 195 horas de paralisação em 2024, um recorde que gerou prejuízos financeiros e afetou a eficiência logística (SOARES, 2024).

Além do impacto econômico, a baixa visibilidade representa um risco significativo para a segurança. Em setores como a aviação, a visibilidade restrita pode comprometer a segurança de aeronaves e tripulações, especialmente durante pousos e decolagens (RODRIGUES et al., 2022).

Diante desse cenário, este estudo aplicou um modelo de previsão baseado no algoritmo *Random Forest* para dados meteorológicos de regiões costeiras. A escolha deste algoritmo justifica-se por sua capacidade de generalização, escalabilidade e interpretabilidade (LOUPPE, 2014).

Para a modelagem, dados do Aeroporto do Galeão e seu entorno foram utilizados, uma vez que o Porto de Santos carece de uma base de dados extensa sobre visibilidade. As fontes de dados, como REDEMET (Departamento de Controle do Espaço Aéreo - DECEA) e a Reanálise ERA5 (*European Centre for Medium-Range Weather Forecasts* - ECMWF), foram selecionadas por sua confiabilidade e reconhecimento internacional. Esta pesquisa, de natureza quantitativa e exploratória, empregou as linguagens de programação Python e R para a análise e modelagem.

## 2 FUNDAMENTAÇÃO TEÓRICA

A névoa marítima é um fenômeno meteorológico de baixa visibilidade que ocorre quando o ar quente e úmido se move sobre águas mais frias, resultando na condensação do vapor d'água. Também conhecida como névoa de advecção, essa formação pode atingir áreas terrestres adjacentes, representando riscos à navegação e ao tráfego. Adicionalmente, a névoa de radiação, que se origina sobre o continente, pode se mover em direção a zonas costeiras e corpos d'água, também comprometendo a visibilidade. A emissão de alertas é recomendada quando a

visibilidade horizontal é reduzida para 1,61 quilômetros ou menos, devido ao seu impacto potencial nas operações marítimas e costeiras (NOAA, 2025).

A compreensão dos processos meteorológicos e físicos que levam à formação da névoa é essencial para aprimorar modelos de previsão. Sua gênese, associada a mecanismos como radiação ou advecção, dificulta o desenvolvimento de um modelo numérico único e plenamente eficaz. O processo de formação da névoa marítima ainda não está totalmente elucidado, com desafios persistentes em equações de previsão que envolvem variáveis como umidade, turbulência, propriedades da superfície e topografia local. Tais fatores influenciam tanto a formação quanto a dissipação da névoa, tornando indispensável um entendimento aprofundado para a construção de modelos mais robustos e precisos (TRÉMANT, 1987).

## 2.1 Conceitos e Formação da Névoa Marítima

A névoa marítima se caracteriza pela condensação do vapor de água próximo à superfície do oceano, formando uma camada de gotículas suspensas no ar. Sua ocorrência está associada a processos que resfriam o ar ou aumentam a umidade relativa até o ponto de orvalho. Como uma névoa de advecção, sua formação se dá quando uma massa de ar se move sobre águas com temperatura distinta, gerando trocas de calor que resultam na condensação (KORAČIN *et al.*, 2013).

Entre os tipos de névoa de advecção, o nevoeiro de mar frio se forma quando massas de ar quente e úmida se deslocam sobre águas frias. Em contraste, o nevoeiro de mar quente ocorre quando o ar frio ou seco avança sobre águas mais quentes, aquecendo e absorvendo umidade até atingir o ponto de orvalho. Em ambos os casos, a interação térmica mediada pela advecção é fundamental para a gênese da névoa marítima (KORAČIN *et al.*, 2013).

Já o nevoeiro de radiação se forma, sobretudo, à noite, sob uma forte inversão térmica superficial gerada pelo resfriamento do solo, que diminui a temperatura do ar até o ponto de orvalho, promovendo a condensação. Sua dissipação ocorre geralmente de baixo para cima (TRÉMANT, 1987).

## 2.2 Impactos da Névoa

A névoa representa um fator crítico para a segurança e eficiência de operações de transporte marítimo e aéreo. No Brasil, os nevoeiros são os principais causadores de limitações de visibilidade no transporte aéreo, com maior incidência nas regiões Sul e Sudeste, resultando em riscos à segurança e prejuízos econômicos por meio de atrasos e cancelamentos de voos (FRANÇA, 2008). No contexto marítimo, estima-se que um terço dos acidentes globais ocorra sob condições de névoa densa (TRÉMANT, 1987). Diante desses impactos, o desenvolvimento de modelos preditivos torna-se uma prioridade para alcançar maiores índices de segurança e eficiência operacional.

### 2.3 Modelagem Preditiva de Névoa e o Algoritmo Random Forests

A previsão de névoa é um desafio significativo na meteorologia devido à complexa interação entre variáveis atmosféricas, oceânicas e de superfície. Em ambientes marítimos, essa complexidade é ampliada pela atuação de processos como radiação, advecção e turbulência, que muitas vezes ocorrem em escalas não resolvidas por modelos numéricos. A precisão da previsão também é limitada pela qualidade e disponibilidade dos dados, já que parâmetros essenciais apresentam rápida variabilidade (TRÉMANT, 1987).

Nesse contexto, técnicas de aprendizado de máquina se destacam como alternativas aos métodos numéricos tradicionais, permitindo lidar com grandes volumes de dados e identificar padrões complexos de forma eficiente. O algoritmo *Random Forest* se sobressai por sua robustez e capacidade de generalização. Baseado em conjuntos de árvores de decisão, ele combina os resultados para melhorar a precisão e reduzir o *overfitting* (sobreajuste). Além de fornecer previsões, o *Random Forest* permite avaliar a importância relativa das variáveis, oferecendo uma interpretação sobre os fatores que mais influenciam o fenômeno (LOUPPE, 2014). Essas características o tornam especialmente útil para modelar e prever a ocorrência de névoa marítima.

## 3 PROCEDIMENTOS METODOLÓGICOS

Este estudo adotou uma abordagem quantitativa e exploratória, utilizando métodos de Análise Exploratória de Dados (EDA) para a coleta, exploração e modelagem de dados meteorológicos. O objetivo foi aplicar o algoritmo de *Random Forest* preliminar para a previsão de visibilidade. Os dados amostrais, de natureza não-probabilística, foram coletados hora a hora no entorno da Ilha do Governador, local do Aeroporto Internacional do Galeão, e complementados com estimativas da temperatura da superfície da água.

### 3.1 Coleta e Seleção de Dados

Para esta pesquisa, foram utilizadas duas bases de dados principais. A primeira é a API REDEMET do Departamento de Controle do Espaço Aéreo (DECEA), que fornece os Meteorological Aerodrome Reports (METARs) do Aeroporto do Galeão (RJ). A segunda fonte é o reanálise ERA5 do *Copernicus Climate Change Service* (C3S), implementado pelo *European Centre for Medium-Range Weather Forecasts* (ECMWF).

A escolha dos dados do Aeroporto do Galeão, em vez de dados do Porto de Santos, baseou-se em três razões:

- I. A ausência de dados de visibilidade do Porto de Santos.
- II. A pouca extensão temporal dos dados disponíveis em estações e boias da Marinha.
- III. A premissa de que os dados climáticos de aeroportos, dada sua importância operacional, são altamente acurados e de caráter oficial.

Inicialmente, o Aeroporto de Guarujá foi considerado como o melhor candidato para representar as condições de Santos. No entanto, a base de dados do local apresentava uma lacuna significativa de dados entre 22h e 10h, o que inviabilizaria a captura de padrões meteorológicos típicos e comprometeria a coesão temporal do estudo. A solução encontrada foi utilizar os dados do Aeroporto do Galeão, que se encontra em uma localização costeira de características geográficas que remetem às do Porto de Santos.

### 3.2 Variáveis

Os dados do Aeroporto de Galeão foram extraídos em formato de *Meteorological Aerodrome Report* (METAR), - informes meteorológicos regulares medidos e registrados em aeródromos (DECEA, 2021a) - acessados por meio do API REDEMET do DECEA. Após sua coleta os dados foram formatados em arquivos tabulares *Comma Separated Values* (CSV) por meio da biblioteca *Python* “Metar”. Segue abaixo os dados escolhidos e suas descrições, assim como consta no *site* do DECEA (2021b):

- I. **Report time**: Dia e hora do informe;
- II. **Visibilidade horizontal** (variável alvo): medida em metros (m), variando de 0 m à 10.000m;
- III. **Pressão do ar**: medida em milibar (MB);
- IV. **Temperatura do ar e do ponto de orvalho**: ambas medidas em graus Celsius;
- V. **Direção do vento**: A primeira medida em graus (°) de 0 a 360°;
- VI. **Velocidade do vento**: medida em nós (knots);

Como complemento, utilizou-se os dados de **umidade relativa** também acessíveis pelo REDEMET API para o aeroporto de Galeão.

Como dito anteriormente, decidiu-se utilizar os dados de estimativas - de alta precisão - da reanálise do ERA5 para um quadrante contendo Galeão (RJ). O uso do ERA5 se demonstrou admissível levando-se em conta a sua autoridade e seu grande uso nacional e internacional. A variável extraída foi a *Lake mix-layer temperature* (LMLT), variável representando, nesse quadrante, a **temperatura da superfície de água** costeira de ambas regiões (ECMWF, 2025).

Após requisitados, os dados foram recebidos em formato *General Regularly Distributed Information in Binary Form* (GRIB). Conforme a ECMWF (2023) GRIB é um formato de arquivos binários utilizados primariamente para armazenamento de dados meteorológicos. Este formato de dados é a escolha padrão de extensão de arquivos meteorológicos para instituições como o *World Meteorological Center* (WMO) e o próprio ECMWF.

### 3.3 Ferramentas e Transformações

Para a extração inicial utilizou-se uma biblioteca de ferramentas disponibilizada pelo próprio ECMWF para a leitura e posterior conversão dos dados: *ecCodes*, sendo este descrito pelo ECMWF (2023) como sendo uma biblioteca que

oferece interface para Python, C, Fortran e ferramentas de *command-line interface* (CLI) para lidar com os dados em formato GRIB e análogos. Por fim, houve a unificação dos arquivos do ERA5 de LMLT e do arquivo dos METARs de Galeão em um *conjunto de dados* definitivo, de onde as análises foram então realizadas.

Para a aplicação do modelo de *Random Forest* foram utilizadas as bibliotecas “randomForest” e “ranger” - via “caret” - para a linguagem R. Já na linguagem Python, utilizou-se o módulo “Scikit-learn” para o treino e avaliação do modelo.

Na confecção deste trabalho utilizaram-se ferramentas de Inteligência Artificial (IA), nominalmente, o *ChatPDF* para a pesquisa no texto, que está em francês, de Trèmant (1987) e também o *Gemini* da *Google* para na revisão de partes do texto.

O conjunto de dados utilizado no treino do modelo preliminar continha 126686 registros de 21 variáveis preditoras, destas 7 foram extraídas diretamente dos conjuntos de dados utilizados, enquanto as demais foram derivadas a partir da engenharia de atributos. No total, gerou-se 14 variáveis derivadas (cíclicas, interações, polinomiais e *lags*), algumas em escala logarítmica, com objetivo de melhorar o desempenho do modelo.

Por fim, descrições mais detalhadas de como os dados foram obtidos e suas especificações, além de como, foram processados e derivados, além de todas as ferramentas e bibliotecas - incluindo auxiliares - utilizadas podem ser encontradas por meio da documentação e códigos do GitHub do projeto (SILVA, 2025).

## 4 RESULTADOS E DISCUSSÕES

Nesta seção, apresentam-se os principais resultados obtidos a partir da construção e avaliação do algoritmo *Random Forest* para dados meteorológicos de regiões costeiras com o objetivo principal de detectar padrões entre as variáveis e a ocorrência de baixa visibilidade. Os resultados enfatizam o treinamento e ajuste do modelo, os indicadores de desempenho e a relevância das variáveis preditoras, assim como o desempenho alcançado e a base de dados utilizada.

O modelo foi treinado com um conjunto de dados previamente tratado e balanceado, composto por 126.686 registros e 21 variáveis preditoras. Dentre essas variáveis, parte foi extraída diretamente de observações meteorológicas históricas, enquanto outra parte foi obtida por meio da “engenharia de atributos”, técnica empregada para ampliar a capacidade preditiva do modelo. Neste caso, ampliaram-se as variáveis preditoras incluindo um “lag” de 1 hora. Assim, o modelo foi capaz de prever a variável alvo (visibilidade) com até uma hora de antecedência.

### 4.1 Treinamento e Teste do Modelo

A configuração do modelo foi realizada utilizando o pacote “ranger”, coordenado pelo pacote “caret”, o que possibilitou a integração entre ajuste de hiperparâmetros, treinamento e avaliação. Diferentemente de algoritmos que requerem normalização, padronização ou transformações matemáticas nos dados, o *Random Forest* pode operar diretamente com as variáveis em seus valores originais, uma vez que cada árvore da floresta utiliza limiares próprios de divisão, eliminando a necessidade de pré-processamentos complexos através do *framework* caret.

No intuito de assegurar a validade da consistência das estimativas geradas, a técnica de *cross-validation* (validação cruzada) foi utilizada para separar o *dataset* em vários subconjuntos, onde o modelo é treinado em uma parte e testado na outra de maneira iterativa, evitando a dependência da divisão única entre treino e teste. Devido à natureza preliminar dessa análise, 5 *folds* foram alocados, isto é, 5 subconjuntos, cada um contendo aproximadamente 101.349 registros. Com isso, a configuração é de 80% e 20% para treino e teste respectivamente.

## 4.2 Métricas de Desempenho do Modelo

Durante as etapas de validação cruzada, o modelo baseado em árvores de decisão foi avaliado sob diferentes combinações de hiperparâmetros, com ênfase no ajuste do hiperparâmetro “*mtry*” (*Number of randomly sampled predictors* - número e preditores aleatoriamente escolhidos), e do hiperparâmetro “*splitrule*” (regra de divisão). O hiperparâmetro *mtry* representa o número de variáveis candidatas consideradas em cada divisão da árvore, sendo testados os valores que correspondem ao número de variáveis preditoras selecionadas por árvore, variando de no mínimo 2 até 21 que abrangem desde uma seleção mais restrita até a utilização do conjunto completo de preditores disponíveis no *dataset*.

Quando configurado com valores mais baixos de *mtry*, o modelo introduz maior aleatoriedade no processo de construção das árvores, com vistas a reduzir a correlação entre elas. Isso tende a aumentar o viés, porém diminui a variância do modelo agregado, resultando em maior robustez contra o *overfitting*. Por outro lado, valores elevados de *mtry* levam à construção de árvores mais fortes e precisas individualmente, no entanto, mais correlacionadas, entre si. Quando as árvores de um modelo de Random Forest apresentam alta correlação entre si, sua capacidade de contribuição independente é reduzida. Isso limita a diversidade do conjunto e, conseqüentemente, diminui a efetividade do processo de agregação, uma vez que os erros tendem a se repetir entre as árvores. Por outro lado, maior heterogeneidade entre os classificadores individuais favorece a generalização do modelo.

Além disso, foram avaliadas duas estratégias distintas para a escolha dos pontos de divisão: (i) o critério *variance*, que busca a minimização da variância residual em cada nó, maximizando a pureza dos subconjuntos e reduzindo o viés do modelo; e (ii) o critério *extratrees*, no qual os pontos de divisão são selecionados de forma aleatória, promovendo árvores mais diversas e com menor correlação entre si. Enquanto a abordagem *variance* tende a construir modelos mais ajustados aos dados de treinamento, contribuindo para diminuir o erro de previsão imediato, a estratégia *extratrees* prioriza a redução da variância do modelo agregado, favorecendo uma maior capacidade de generalização em cenários de dados futuros.

A Figura 1 apresenta o relatório da modelagem com os resultados obtidos para a configuração ótima identificada durante a validação cruzada. O modelo ótimo reflete uma solução estável, capaz de generalizar para novos dados e evidência o equilíbrio entre viés e variância alcançado pelos hiperparâmetros selecionados. Na figura, são exibidas as principais métricas de desempenho: o Erro Quadrático Médio da Raiz (RMSE), calculado na escala logarítmica da variável-alvo transformada, que indica a magnitude média dos erros de previsão; o Coeficiente de Determinação

( $R^2$ ), que quantifica a proporção da variância explicada pelo modelo; e o Erro Absoluto Médio (MAE), que fornece uma medida de erro médio consistente, com o objetivo de validar a precisão das estimativas em diferentes condições de visibilidade.

**Figura 1 – Relatório Modelo Random Forest**

```
Random Forest
126686 samples
 21 predictor

No pre-processing
Resampling: Cross-Validated (5 fold)
Summary of sample sizes: 101349, 101349, 101349, 101349, 101348
Resampling results across tuning parameters:

mtry  splitrule  RMSE      Rsquared  MAE
2     variance  0.1829680  0.6872165  0.08122299
2     extratrees 0.1867808  0.6854891  0.08543665
11    variance  0.1794603  0.6876239  0.07321871
11    extratrees 0.1788743  0.6894315  0.07322479
21    variance  0.1826545  0.6772839  0.07440526
21    extratrees 0.1805512  0.6839338  0.07392490

Tuning parameter 'min.node.size' was held constant at a value of 5
RMSE was used to select the optimal model using the smallest value.
The final values used for the model were mtry = 11, splitrule = extratrees
and min.node.size = 5.
```

**Fonte:** elaborado pelos autores (2025)

O desempenho do modelo apresentado na Figura 1 revelou que a solução ótima com  $mtry = 11$ , aproximadamente metade do total de 21 preditores, conseguiu equilibrar o viés e a variância, capturando padrões relevantes da visibilidade atmosférica. Vale destacar que a seleção do subconjunto de variáveis otimizou a capacidade de generalização do modelo, evidenciada pela estabilidade das previsões. A opção pela regra de divisão *extratrees* diminuiu a similaridade entre as árvores, ao passo que o valor do *min.node.size* igual a 5 garantiu estabilidade nas decisões de cada árvore e a floresta composta por 500 árvores asseguraram consistência e confiabilidade nas estimativas.

As métricas  $R^2$ , RMSE, e MAE foram utilizadas por tratar-se de medidas convencionais para variáveis-alvo contínuas, permitindo quantificar a proporção da variância explicada, conforme apresentado nos resultados da Figura 1. O  $R^2 = 0,689$  indicou que o modelo explicou cerca de 69% da variância da variável logarítmica de visibilidade e apontou um elevado grau de ajuste do modelo com os dados, apesar de se tratar de um fenômeno atmosférico naturalmente imprevisível. O Erro Quadrático Médio da Raiz (RMSE = 0,179) apresentou desvios absolutos menores em condições de baixa visibilidade, aproximando-se de 39 m para visibilidade de 200 m, reforçando a precisão do modelo em situações críticas de nevoeiro. O Erro Absoluto Médio (MAE = 0,0732) confirmou a confiabilidade das estimativas, fornecendo uma métrica de fácil interpretação e aplicação direta para a gestão de risco operacional e utilidade prática na antecipação de eventos de baixa visibilidade.

### 4.3 Otimização e Ajuste de Hiperparâmetros

Nesta etapa da modelagem, o foco concentrou-se na otimização e ajuste de hiperparâmetros a fim de maximizar a capacidade preditiva e a generalização do modelo. Os hiperparâmetros definem aspectos estruturais e operacionais do algoritmo, incluindo o número de variáveis por divisão (*mtry*), o critério de divisão dos nós (*splitrule*). Dessa forma, ajustes adequados possibilitam equilibrar viés e variância, reduzir correlações entre árvores e a probabilidade de *overfitting*, além de assegurar que o modelo capture padrões relevantes do fenômeno estudado, neste caso, a incidência de baixa visibilidade. O resultado do crescimento das árvores evidenciou progressos, alcançando 48% e 95%, conforme ilustrado na Figura 2.

**Figura 2 – Ajuste de Hiperparâmetros**

Número de amostras: 126686  
Número de preditores: 21

```
Aggregating results
Selecting tuning parameters
Fitting mtry = 11, splitrule = extratrees, min.node.size = 5 on full training set
Growing trees.. Progress: 24%. Estimated remaining time: 1 minute, 40 seconds.
Growing trees.. Progress: 48%. Estimated remaining time: 1 minute, 8 seconds.
Growing trees.. Progress: 71%. Estimated remaining time: 37 seconds.
Growing trees.. Progress: 95%. Estimated remaining time: 6 seconds.
Computing permutation importance.. Progress: 45%. Estimated remaining time: 37 seconds.
Computing permutation importance.. Progress: 90%. Estimated remaining time: 6 seconds.
```

**Fonte:** elaborado pelos autores (2025)

O relatório de ajuste do modelo indicou que, após a agregação dos resultados e a seleção dos hiperparâmetros considerados ótimos, o *Random Forest* foi calibrado para fornecer previsões consistentes e generalizáveis. O modelo foi treinado com *mtry* = 11, *splitrule* = *extratrees* e *min.node.size* = 5 sobre o conjunto completo de treinamento, configurando-se como a solução ótima identificada na fase de ajuste de hiperparâmetros. Essa configuração otimizada demonstrou capacidade de capturar padrões associados à visibilidade atmosférica e gerou resultados acerca da importância das variáveis estudadas, com aplicabilidade prática na previsão de condições críticas de baixa visibilidade.

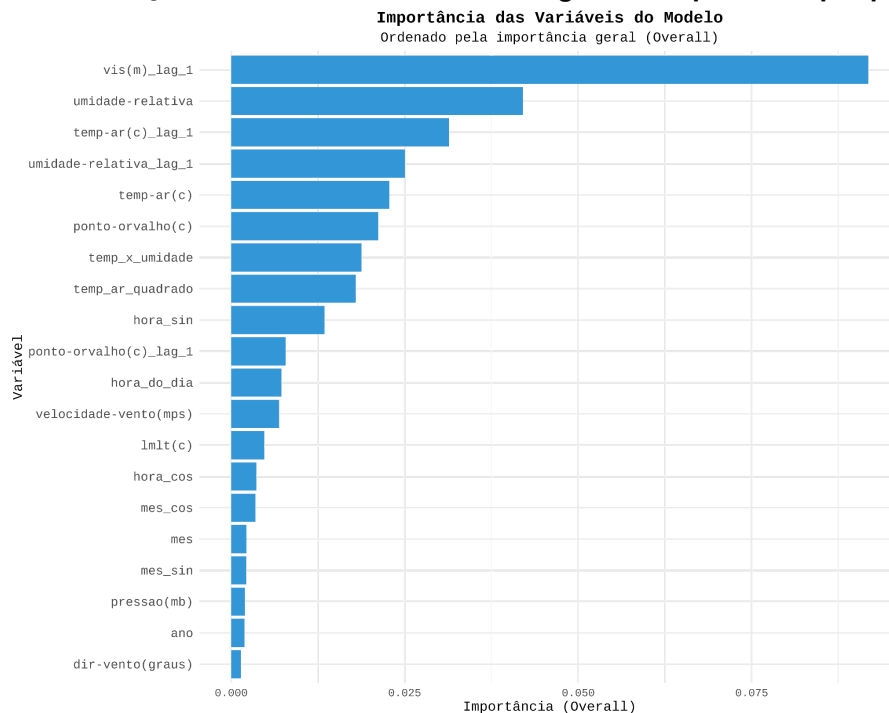
Paralelamente, o processo de crescimento das árvores ocorreu de forma progressiva, com *checkpoints* indicando 24%, 48%, 71% e 95% do progresso, concluindo com um tempo total de treinamento de aproximadamente 8 minutos quando treinado em computador com 32gb de RAM e CPU Ryzen 3600x de 12 núcleos. A execução do modelo permitiu atestar a eficiência computacional do *Random Forest* e sua capacidade de escalabilidade diante do grande volume de dados. Na perspectiva aplicada, o modelo tem potencial para viabilizar sua utilização em sistemas de apoio à decisão, com vistas a prover alertas de baixa visibilidade dentro de prazos compatíveis com a necessidade de intervenção.

### 4.4 Importância das Variáveis Preditoras

A análise da importância das variáveis preditoras foi conduzida por meio da métrica de importância por permutação, que verifica a degradação do desempenho

do modelo quando cada preditor é embaralhado aleatoriamente. Através dessa abordagem, foi possível estabelecer uma hierarquia clara entre os preditores, identificando os fatores que exerceram maior influência sobre a estimativa da visibilidade atmosférica. A Figura 3 apresenta o ranqueamento das variáveis de acordo com seu grau de relevância por permutação.

**Figura 3 - Ordenação das variáveis conforme seu grau de importância por permutação**



**Fonte:** elaborado pelos autores (2025)

A análise revelou que quanto maior a degradação do desempenho do modelo após o embaralhamento de uma variável, maior foi a relevância atribuída a essa variável. Para sistematizar a interpretação da importância das variáveis preditoras, os resultados foram organizados em quatro blocos principais: Dominância do Contexto Temporal, Influência Combinada de Condições Atuais e Passadas, Validação da Engenharia de Atributos e Preditores de Menor Influência.

Conclui-se, portanto que, os resultados da avaliação da importância das variáveis mais influentes na previsão da ocorrência de baixa visibilidade, conduzida por meio do método de permutação não apenas validam a robustez da modelagem, mas também alinham suas descobertas com os princípios físicos da meteorologia. Além disso, o modelo aprendeu a priorizar a persistência do fenômeno e a dinâmica de atributos meteorológicos auxiliares, que exerceram maior impacto na acurácia das estimativas, confirmando o êxito da abordagem de engenharia de atributos.

## 5 CONSIDERAÇÕES FINAIS



Observa-se que a aplicação do modelo *Random Forest* preliminar evidenciou uma capacidade explicativa considerável do modelo à variável alvo ( $R^2 = 0,689$ ). Sendo que o modelo em si prioriza uma série de fatores em sua previsão: persistência temporal (indicado pelas variáveis *lag*); umidade; temperatura; proximidade ao ponto de orvalho, etc. Sendo, portanto, consistente com a bibliografia acerca de nevoeiros e suas variáveis de influência.

O intervalo da previsão é relativamente disperso ao redor do valor real, mas ainda assim, há uma baixa divergência e anormalidade dos valores de visibilidade, sendo um bom preditor mesmo que não possa expor a visibilidade exata.

Retomando a questão central deste estudo, a aplicação do modelo de *Random Forest* demonstra-se promissora para a previsão de eventos climáticos de baixa visibilidade. Contudo, leva-se em conta o fato de se tratar de uma modelagem preliminar, passível de mudanças e aprimorações, sendo o objeto dessa pesquisa um possível modelo para algoritmos semelhantes.

Dito isso, esta pesquisa contém algumas limitações, em especial relacionadas às configurações dos “hiperparâmetros” do modelo, e como seus valores podem influenciar sua capacidade preditiva. Além disso, devido à alta demanda de recursos computacionais para o treino do modelo em versões mais refinadas, há a exigência de uma quantidade significativa de tempo para o treino e consequente teste do modelo. Pesquisas futuras voltadas à otimização da estruturação das variáveis e do ajuste dos hiperparâmetros poderão ser de grande valia para o problema de visibilidade, para a previsão de visibilidade, especialmente em horizontes superiores a uma hora.

## REFERÊNCIAS

CAMPANINI, M. A. **Recorde de paralisações no porto de Santos devido à neblina densa**. Disponível em: <https://11nq.com/BM822>. Acesso em: 19 jul. 2025.

DECEA (Brasil). **O que é a REDEMET? – Central de Ajuda DECEA**. Disponível em: <https://ajuda.decea.mil.br/base-de-conhecimento/o-que-e-a-redemet/>. 2021a. Acesso em: 22 jul. 2025.

DECEA (Brasil). **Como decodificar o METAR e o SPECI? – Central de Ajuda DECEA**. Disponível em: <https://ajuda.decea.mil.br/base-de-conhecimento/como-decodificar-o-metar-e-o-speci/>. 2021b. Acesso em: 19 jul. 2025.

ECMWF. **Lake mix-layer temperature - lmlt**. 2025. Disponível em: <https://codes.ecmwf.int/grib/param-db/228008>. Acesso em: 29 jun. 2025.

ECMWF **Confluence Wiki. 2023**. Disponível em: <https://tinu.be/NHUqA0tt1>. Acesso em: 18 jul. 2025.

FRANÇA, V. D. J. **Avaliação da metodologia de previsão de nevoeiro e visibilidade horizontal do modelo ETA**. INPE, São José dos Campos, 2008.



GU, L.-M., *et al.* **Uma simulação numérica de um evento de neblina na Baía de Sichuan, China: a sensibilidade às elevações do terreno.** *Atmosphere*,(2024)15 (12), 1546. <https://doi.org/10.3390/atmos15121546>.

KIGUTI, L. S. H.; AVANCINI, G. F. **Eventos Climáticos extremos e seus reflexos na vulnerabilidade social e ambiental em Santos/SP.** 36 f. Graduação em Engenharia Ambiental - Instituto do Mar, Universidade Federal de São Paulo, Santos, 2022.

KORAČIN, Darko; DORMAN, Clive E.; LEWIS, John M.; HUDSON, James G.; WILCOX, Eric M.; TORREGROSA, Alicia. **Marine fog: a review.** *Atmospheric Research*, [s.l.], v. 143, p. 142-175, 2014. DOI: 10.1016/j.atmosres.2013.12.012.

LOUPPE, G. **Understanding Random Forests: from theory to practice.** Faculdade de Ciências Aplicadas, Universidade de Liège, Liège, 2014.

NOAA. **Fog Over Water.** 2025. Disponível em: <https://www.weather.gov/safety/fog-water>. Acesso em: 22 jul. 2025.

SILVA, Gabriel Luiz dos Santos; CARVALHO, Igor Silva de; BARBOSA, Luis Henrique Moreira. **Port Visibility Analysis.** Versão 0.0.1. Repositório GitHub. 2025. Disponível em: <https://github.com/igorzeck/Port-Visibility-Analysis>. Acesso em: 25 ago. 2025.

SOARES, G. **Neblina fecha o maior porto da América do Sul por tempo recorde: veja os impactos e prejuízos milionários.** veja os impactos e prejuízos milionários. 2024. Disponível em: <https://g1.globo.com/sp/santos-regiao/porto-mar/noticia/2024/09/14/neblina-fecha-o-maior-porto-da-am>. Acesso em: 24 set. 2025.

TRÉMANT, M. **La prevision du brouillard en mer.** Centro de meteorologia marinha, Brest, 1987. Disponível em: <https://library.wmo.int/records/item/53704-la-prevision-du-brouillard-en-mer?offset=>. Acesso em: 22 jul. 2025.

Os conteúdos expressos no trabalho, bem como sua revisão ortográfica e adequação às normas ABNT, são de inteira responsabilidade dos autores. Declara-se que durante a preparação deste trabalho foram utilizados ChatPDF e o Gemini para tradução de um artigo do francês para o português. Após utilizar essas ferramentas/serviços, os autores editaram e revisaram o conteúdo conforme necessário e assumem total responsabilidade pela publicação.