

# O Uso da Técnica de Conformal Prediction Para Melhoria de Classificação em Algoritmos de Machine Learning

João Vinicius Martins

joao2014viniciusm@gmail.com

Marcos Vinicius Carvalho de Oliveira

markyankg@gmail.com

19 de janeiro de 2026

Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE)  
Fortaleza – CE – Brasil

## Resumo

Os algoritmos de Machine Learning são amplamente utilizados para classificação, mas modelos tradicionais operam como classificadores pontuais, falhando em comunicar a incerteza de suas previsões. Em áreas críticas, como o diagnóstico médico, essa limitação pode comprometer a segurança da tomada de decisão. Este trabalho aplica a técnica de Predição Conforme (*Conformal Prediction*) como uma camada de calibração sobre três classificadores distintos: Regressão Logística, k-Nearest Neighbors (K-NN) e Support Vector Machine (SVM), utilizando o conjunto de dados *Heart Disease Prediction*. A metodologia adotou a abordagem indutiva (*Split Conformal*), avaliando métricas de cobertura marginal, tamanho médio dos conjuntos preditivos e taxas de ambiguidade. Os resultados evidenciaram que, embora a precisão pontual seja semelhante, os algoritmos apresentam dinâmicas de calibração divergentes. A Regressão Logística demonstrou maior estabilidade na manutenção da cobertura marginal, ao passo que o SVM e o K-NN apresentaram variações esperadas devido à sensibilidade da calibração em amostras reduzidas. Conclui-se que a Predição

Conforme enriquece a interpretabilidade dos modelos, transformando previsões incertas em conjuntos de decisão com garantias estatísticas de validade.

**Palavras-chave:** Machine Learning. Predição Conforme. Quantificação de Incerteza. Classificação. Diagnóstico Médico.

## Abstract

Machine Learning algorithms are widely used for classification, yet traditional models operate as point classifiers, failing to communicate the uncertainty of their predictions. In critical areas such as medical diagnosis, this limitation can compromise decision-making safety. This work applies the Conformal Prediction technique as a calibration layer over three distinct classifiers: Logistic Regression, k-Nearest Neighbors (K-NN), and Support Vector Machine (SVM), using the Heart Disease Prediction dataset. The methodology adopted the inductive approach (Split Conformal), evaluating metrics of marginal coverage, average size of prediction sets, and ambiguity rates. The results evidenced that, although point accuracy is similar, the algorithms present divergent calibration dynamics. Logistic Regression demonstrated greater stability in maintaining marginal coverage, whereas SVM and K-NN presented expected variations due to calibration sensitivity in reduced samples. It is concluded that Conformal Prediction enhances model interpretability, transforming uncertain predictions into decision sets with statistical validity guarantees.

**Keywords:** Machine Learning. Conformal Prediction. Uncertainty Quantification. Classification. Heart Disease.

## 1 Introdução

Os algoritmos de Machine Learning se tornaram ferramentas essenciais para a análise classificatória, demonstrando capacidade de identificar padrões em domínios complexos. Contudo, uma limitação inerente à maioria dos modelos de classificação é a sua incapacidade de fornecer uma medida de confiança estatisticamente robusta junto às suas previsões. Eles operam como classificadores pontuais, informando a classe mais provável, mas sem comunicar o quanto incerta ou ambígua essa decisão pode ser. Em áreas de alto risco, como o diagnóstico clínico, essa lacuna representa uma barreira para a adoção segura e responsável dessas tecnologias.

Identificando essa dificuldade, este artigo se aprofunda na técnica da Predição Conforme (Conformal Prediction), um framework agnóstico a modelos proposto por Vovk, Gammerman e Shafer (2005) para quantificar a incerteza. A técnica opera como uma camada sobre um modelo de Machine Learning já treinado, utilizando um conjunto de dados de calibração para aprender um limiar de "estranhamento" ou não-conformidade. A partir disso, para uma nova amostra, em vez de fornecer uma única resposta, a técnica gera um conjunto de previsão. A principal vantagem deste método é a sua garantia teórica de cobertura: o usuário pode especificar um nível de

confiança (por exemplo, 95%), e o framework garante que o resultado verdadeiro estará contido no conjunto de predição em 95% das vezes, a longo prazo.

A abordagem deste artigo para explorar a Predição Conforme será prática e aplicada. Primeiramente, apresenta-se a fundamentação teórica da técnica, detalhando o funcionamento das medidas de não-conformidade e do processo de calibração. Em seguida, aplica-se o framework sobre algoritmos de classificação distintos treinados com o dataset "Heart Disease Prediction". A performance será avaliada sob uma dupla perspectiva: por um lado, utilizando métricas de classificação tradicionais como acurácia e por outro, analisando as métricas específicas da Predição Conforme, como a cobertura efetiva e o tamanho médio dos conjuntos de predição.

O objetivo final é demonstrar como a Predição Conforme enriquece as previsões não por alterar a performance intrínseca dos modelos, mas por adicionar uma camada de confiabilidade. Com isso, buscamos evidenciar o potencial da técnica para transformar um classificador padrão em uma ferramenta de apoio à decisão mais segura e transparente, especialmente em domínios onde o custo de um erro é alto.

## 2 Fundamentação Teórica

Os modelos de machine learning consistem em métodos computacionais capazes de estimar funções a partir de padrões extraídos de dados, geralmente por meio de técnicas estatísticas. Nesse contexto, este trabalho utiliza os modelos de Regressão Logística, KNN e SVM, bem como o método de Conformal Prediction. Inicialmente, apresenta-se a formulação matemática e os princípios que fundamentam esses modelos como técnicas de classificação. Em seguida, descreve-se o funcionamento do Conformal Prediction, destacando sua capacidade de quantificar incertezas e fornecer previsões mais confiáveis.

### 2.1 Algoritmo de Regressão Logística

A Regressão Logística é um método estatístico amplamente empregado em tarefas de classificação binária. Seu objetivo é modelar a probabilidade de uma amostra pertencer à classe positiva a partir de uma combinação linear dos atributos de entrada. Para isso, aplica-se a função logística (sigmoide), que restringe a saída ao intervalo [0,1]. Seja  $x$  o vetor de atributos de uma amostra, a probabilidade do rótulo  $y = 1$  é estimada por:

$$P(y = 1|x) = \sigma(w^T x + b),$$

em que  $w$  representa o vetor de pesos,  $b$  o termo de viés e  $\sigma(z)$  é a função sigmoide, definida por:

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

Onde  $z$  é uma combinação linear das variáveis independentes:

$$z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

$\beta_0, \beta_1, \beta_2, \dots, \beta_n$  representam os parâmetros do modelo e precisam ser estimados durante o processo de treinamento. Na Regressão Logística, a estimação é realizada por meio do método de Máxima Verossimilhança (Maximum Likelihood Estimation — MLE). Esse método busca encontrar o conjunto de parâmetros que maximiza a probabilidade dos dados observados. Para isso, a função de verossimilhança é convertida em sua forma logarítmica — a log-verossimilhança — e algoritmos iterativos de otimização ajustam os valores de  $\beta$  de modo a maximizar essa função, resultando na melhor estimativa dos parâmetros do modelo.

A Regressão Logística apresenta vantagens importantes para este estudo: é um modelo interpretável, possui fronteira de decisão linear e fornece probabilidades calibradas, características que a tornam adequada como baseline probabilístico na avaliação da técnica de Predição Conforme.

## 2.2 K-Nearest Neighbors

O algoritmo k-Nearest Neighbors (K-NN) é um método de classificação supervisionado e não paramétrico, cuja decisão se baseia exclusivamente na proximidade entre as amostras no espaço das features. Diferentemente de modelos que aprendem parâmetros explícitos, o K-NN armazena o conjunto de treinamento e realiza a classificação de uma nova amostra comparando-a com seus  $k$  vizinhos mais próximos.

O processo de decisão ocorre por meio de votação majoritária: cada vizinho contribui com seu rótulo, e a classe mais frequente entre os  $k$  vizinhos determina a classificação final. Assim, o desempenho do K-NN depende diretamente da distribuição espacial dos dados e da escolha adequada de  $k$ .

Para identificar os vizinhos mais próximos, o algoritmo utiliza uma métrica de distância. Neste trabalho, adotou-se a distância Euclidiana, definida por:

$$d(x, y) = \sqrt{\sum_{i=1}^n (y_i - x_i)^2}$$

O KNN é utilizado como baseline para analisar seu comportamento antes e depois da calibração por meio do Conformal Prediction. Dessa forma, é possível compreender como modelos não paramétricos se comportam em análises de confiabilidade de seus resultados.

## 2.3 Support Vector Machine

As Máquinas de Vetores de Suporte (Support Vector Machines — SVM) são algoritmos de aprendizado supervisionado amplamente utilizados para tarefas de classificação. Seu princípio fundamental consiste em encontrar um hiperplano de separação que divide o espaço de atributos de modo a maximizar a margem, definida como a distância entre o hiperplano e os pontos mais próximos de cada classe, denominados vetores de suporte. A maximização dessa margem tende a produzir modelos com boa capacidade de generalização.

Quando as classes não são linearmente separáveis no espaço original, o SVM utiliza funções kernel para projetar os dados em espaços de maior dimensionalidade, onde a separação se torna possível. No scikit-learn, o kernel padrão é o RBF (Radial Basis Function), adequado para capturar padrões não lineares e fronteiras de decisão mais complexas.

De forma simplificada, a função de decisão de um SVM pode ser expressa como:

$$f(x) = w * x + b$$

Onde  $w$  representa o vetor de pesos associado ao hiperplano e  $b$  é um termo de viés. A predição final é determinada pelo lado do hiperplano no qual o ponto  $x$  se encontra.

Neste trabalho, o SVM é empregado como um dos modelos baseline para avaliar o impacto da calibração via Conformal Prediction. Como o SVM não gera probabilidades de forma nativa, o modelo foi configurado com `probability=True`, o que ativa o método interno de Platt Scaling — responsável por produzir as estimativas probabilísticas necessárias para o cálculo dos nonconformity scores durante o processo de calibração.

## 2.4 Predição Conforme

A Predição Conforme (Conformal Prediction — CP) é um framework estatístico desenvolvido para fornecer estimativas de incerteza com garantias formais de cobertura, independentemente do modelo base utilizado. Diferentemente das abordagens tradicionais de classificação, que retornam previsões pontuais, a Predição Conforme produz conjuntos preditivos, possibilitando quantificar o grau de confiança associado a cada decisão do modelo.

A principal propriedade do método é a sua validade marginal, que assegura que o rótulo verdadeiro pertence ao conjunto preditivo gerado com probabilidade mínima de  $1 - \alpha$ , onde  $\alpha$  representa o nível de erro estabelecido pelo usuário.

Neste trabalho, adota-se a variante Split Conformal Prediction, que divide o conjunto de dados em três partes: treinamento, validação e calibração.

1. Treinamento: utilizado para ajustar o modelo base.
2. Validação: opcional, para ajustes finos do modelo.
3. Calibração: reservado exclusivamente para calcular os nonconformity scores, garantindo que o processo seja independente do ajuste inicial do modelo.

Calibração é feita calculando os escores de não conformidade (*nonconformity scores*) definido como:

$$S_i = 1 - \hat{P}_{y_i}(x_i)$$

O escore de não conformidade mede o quão “atípica” é uma previsão feita pelo modelo. Onde  $\hat{P}_{y_i}(x_i)$  representa a probabilidade prevista pelo modelo para a classe verdadeira para o elemento  $i$ . Quanto maior  $S_i$ , menos confiante o modelo estava na classe correta.

O limiar  $q_\alpha$  desempenha um papel central no método de Split Conformal Prediction, pois determina o ponto de corte necessário para garantir que o conjunto preditivo resultante apresente a cobertura estatística desejada. Depois de calcular todos os escores de não conformidade no conjunto de calibração, esses valores são organizados em ordem crescente para formar uma distribuição empírica da “não conformidade” do modelo.

Em seguida, define-se o limiar  $q_\alpha$  como o quantil de ordem  $1 - \alpha$  dessa distribuição. Esse quantil corresponde ao valor que delimita o maior escore de não conformidade aceitável para que, no limite, a verdadeira classe seja incluída no conjunto preditivo com probabilidade mínima de  $1 - \alpha$ . Em termos práticos, representa a tolerância máxima de incerteza observada no conjunto de calibração: quaisquer previsões futuras cujo grau de não conformidade excede esse limiar são consideradas insuficientemente confiáveis e são descartadas do conjunto preditivo.

Assim, para uma nova entrada  $x_{new}$ , o conjunto preditivo é definido como:

$$C(x_{new}) = \{y : 1 - \hat{p}_y(x_{new}) \leq \hat{q}_\alpha\}$$

Esse conjunto contém todas as classes consideradas plausíveis dado o nível de confiança escolhido. Por exemplo, para  $\alpha=0.1$ , o método garante teoricamente que o conjunto conterá a verdadeira classe em pelo menos 90% dos casos.

### 3 Metodologia

A metodologia adotada neste trabalho fundamenta-se na estrutura da Predição Conforme (*Conformal Prediction*), aplicada sobre um *pipeline* de aprendizado de máquina supervisionado. O objetivo experimental é validar a hipótese de que a camada de calibração confere garantias estatísticas de cobertura marginal aos classificadores, independentemente da sua arquitetura subjacente.

#### 3.1 Aquisição e Tratamento dos Dados

O estudo utilizou o conjunto de dados *Heart Disease Prediction* (Cleveland Database). Este conjunto de dados é composto por 303 amostras e 14 atributos, incluindo idade, sexo, tipo de dor no peito, pressão arterial em repouso, colesterol, entre outros. A variável-alvo, denominada target, é binária, indicando a presença (1) ou ausência (0) de doença cardíaca.

O pré-processamento dos dados seguiu etapas para evitar o vazamento de informações (*data leakage*):

1. **Codificação:** Variáveis categóricas foram mantidas em sua representação numérica original conforme a documentação do dataset.
2. **Normalização:** Dado que algoritmos baseados em distância (como K-NN) e margem (como SVM) são sensíveis à escala dos atributos, aplicou-se a padronização Z-score

através da função:

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

Onde  $\mu$  e  $\sigma$  foram calculados **apenas** sobre o subconjunto de treinamento e aplicados subsequentemente aos conjuntos de calibração e teste.

### 3.2 Estratégia de Particionamento

Para satisfazer a premissa de intercambialidade exigida pela Predição Conforme, os dados foram particionados aleatoriamente em três subconjuntos disjuntos:

- **Conjunto de Treinamento ( $D_{train}$  - 70%)**: Utilizado para a otimização dos parâmetros  $\theta$  dos modelos de classificação base ( $\hat{f}$ ). Com 212 amostras.
- **Conjunto de Calibração ( $D_{cal}$  - 15%)**: Um conjunto de validação reservado estritamente para o cálculo dos scores de não-conformidade. Este conjunto não foi visto pelo modelo durante o treinamento. Com 46 amostras.
- **Conjunto de Teste ( $D_{test}$  - 15%)**: Utilizado para a avaliação final das métricas de generalização e verificação empírica da cobertura. Com 45 amostras.

### 3.3 Arquitetura dos Classificadores Base

Foram instanciados três modelos distintos para investigar o comportamento da calibração em diferentes fronteiras de decisão:

1. **Régressão Logística**: Utilizada como *baseline* linear. A estimativa de probabilidades foi baseada na função Sigmoidal aplicada ao logit.
2. **k-Nearest Neighbors (K-NN)**: Configurado com  $k = 5$  vizinhos e métrica de distância Euclidiana. Como o K-NN não gera probabilidades nativamente, a estimativa  $P(y|x)$  foi calculada pela proporção de votos entre os  $k$  vizinhos.
3. **Support Vector Machine (SVM)**: Configurado com kernel RBF. Para obter estimativas de probabilidade (necessárias para o cálculo do score), utilizou-se o método de *Platt Scaling*, que ajusta uma regressão logística sobre as distâncias das amostras ao hiperplano separador.

### 3.4 Modelagem dos Classificadores

Para validar a generalização da Predição Conforme, foram selecionados três algoritmos baseados em paradigmas de aprendizado distintos: linear (Régressão Logística), baseado em instâncias (K-NN) e baseado em margem (SVM). A seguir, detalha-se a fundamentação teórica e a configuração de cada modelo.

### 3.4.1 Regressão Logística (Baseline Probabilístico)

A Regressão Logística foi adotada como o modelo base (*baseline*) devido à sua natureza intrinsecamente probabilística. O algoritmo modela a probabilidade de uma instância pertencer à classe positiva ( $y = 1$ ) ajustando uma função linear aos dados, que é posteriormente transformada pela função logística (Síntese).

- **Configuração:** O modelo foi treinado utilizando o otimizador padrão *lbfgs*. Devido à sensibilidade do algoritmo à convergência, definiu-se um limite de iterações suficiente para garantir a estabilização dos coeficientes. Como a saída do modelo já é uma probabilidade direta  $P(y|x)$ , ela foi utilizada diretamente para o cálculo dos scores de não-conformidade.

### 3.4.2 k-Nearest Neighbors (K-NN)

O algoritmo K-NN (*k-Vizinhos Mais Próximos*) foi escolhido para representar métodos não-paramétricos, que não assumem uma distribuição fixa dos dados. O K-NN classifica uma nova amostra baseando-se na similaridade com as instâncias presentes no conjunto de treino.

- **Configuração:** Definiu-se o hiperparâmetro  $k = 5$ , utilizando a distância Euclidiana como métrica de similaridade.
- **Adaptação para Predição Conforme:** Como o K-NN é originalmente um classificador de "voto majoritário", a estimativa de probabilidade necessária para o score de não-conformidade foi calculada pela frequência relativa das classes entre os 5 vizinhos (ex: se 4 vizinhos são da classe 1,  $P(y = 1|x) = 0.8$ ).

### 3.4.3 Support Vector Machine (SVM)

O SVM foi utilizado para avaliar o comportamento da técnica em fronteiras de decisão complexas e não-lineares. O objetivo teórico do SVM é encontrar o hiperplano que maximiza a margem de separação entre as classes.

- **Configuração:** Utilizou-se o *kernel RBF* (Radial Basis Function) para mapear os dados em dimensões superiores.
- **Calibração de Probabilidades (Platt Scaling):** O SVM padrão retorna apenas a distância da amostra ao hiperplano, não uma probabilidade. Para viabilizar a aplicação da Predição Conforme, o modelo foi instanciado com o parâmetro *probability=True*. Isso ativou internamente o método de *Platt Scaling*, que ajusta uma regressão logística sobre as distâncias geradas pelo SVM, convertendo-as em probabilidades calibradas ( $P(y|x)$ ) aptas para o cálculo do score.

### 3.5 Aplicação da Predição Conforme (Camada de Incerteza)

Após o treinamento individual de cada modelo descrito acima, aplicou-se o framework de Predição Conforme de maneira uniforme:

1. **Cálculo de Scores:** Para cada modelo, calculou-se o grau de "estranheza" (*strangeness score*) das amostras de calibração subtraindo a probabilidade predita de 1.
2. **Definição do Quantil:** Estabeleceu-se o limiar de corte  $\hat{q}$  com base na distribuição desses scores para um nível de confiança alvo de 90% ( $\alpha = 0.1$ ).
3. **Construção dos Conjuntos:** Na fase de teste, para cada nova amostra, o modelo gerou um conjunto de predição contendo todas as classes cujos scores ficaram abaixo desse limiar  $\hat{q}$ .

### 3.6 Métricas de Avaliação de Confiabilidade

Além das métricas tradicionais (Acurácia, F1-Score), a avaliação focou em métricas específicas de Predição Conforme:

- **Cobertura Marginal:** A taxa observada de inclusão do rótulo verdadeiro nos conjuntos preditivos.
- **Ineficiência (Tamanho Médio):** A média da cardinalidade dos conjuntos  $|C(x)|$ . Conjuntos menores indicam maior precisão informativa.
- **Taxa de Ambiguidade e Vazios:** Percentual de conjuntos contendo múltiplas classes ou nenhuma classe (conjunto vazio).
- **Quantidade de amostras que tiveram determinada classe em seu conjunto resposta:** O somatório de quantas amostras que tiveram determinada classe em seu conjunto resposta a partir da incerteza do modelo.

## 4 Resultados

Após o treinamento de cada modelo, o teste de acurácia e precisão da categorização foi feito com nossas 45 amostras reservadas. Obtendo assim os dados a seguir:

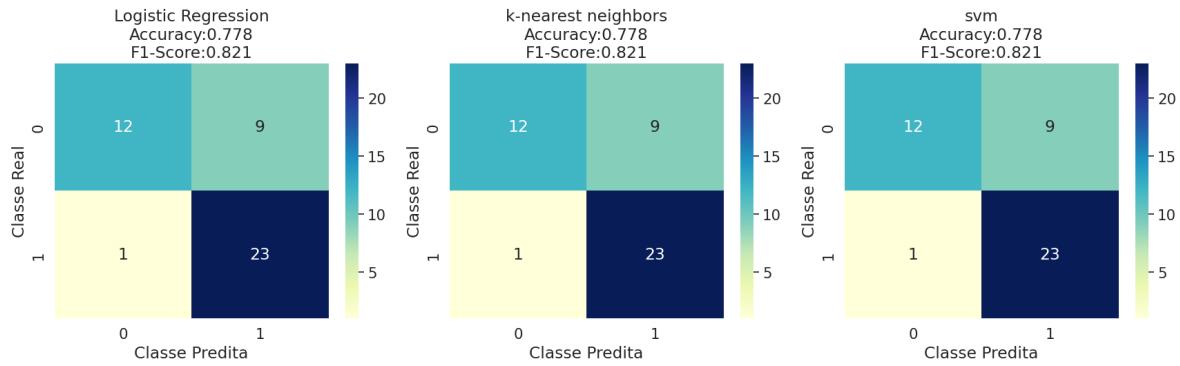


Figura 1 – Matrizes de confusão obtida após treinar a Regressão Logística, KNN e SVM, respectivamente.

A partir dessas matrizes, observa-se que os três modelos apresentaram desempenhos muito semelhantes para o conjunto de dados utilizado, não havendo diferenças significativas entre eles no que diz respeito à capacidade de classificação. Dessa forma, pode-se assumir que, para este cenário experimental, a escolha do algoritmo não alterou substancialmente o resultado final da classificação.

No entanto, uma vez que os modelos foram ajustados, torna-se possível aplicar o método de Conformal Prediction (CP) para avaliar propriedades adicionais relacionadas à confiabilidade das previsões. A calibração por CP permite obter métricas complementares que aprofundam a análise dos modelos, indo além da avaliação tradicional baseada apenas na acurácia

MÉTRICAS DE CONFORMAL PREDICTION DA RL	MÉTRICAS DE CONFORMAL PREDICTION DO KNN	MÉTRICAS DE CONFORMAL PREDICTION DO SVM
Cobertura (CP): 0.911	Cobertura (CP): 0.911	Cobertura (CP): 0.889
Tamanho médio do conjunto: 1.267	Tamanho médio do conjunto: 1.289	Tamanho médio do conjunto: 1.156
Ambiguidade: 0.267	Ambiguidade: 0.289	Ambiguidade: 0.156
Singletons: 0.733	Singletons: 0.711	Singletons: 0.844
Conjuntos vazios: 0.000	Conjuntos vazios: 0.000	Conjuntos vazios: 0.000
amostras no conjunto 0: 24	amostras no conjunto 0: 21	amostras no conjunto 0: 18
amostras no conjunto 1: 33	amostras no conjunto 1: 37	amostras no conjunto 1: 34

Figura 2 – Métricas da calibração do conformal prediction para Regressão Logística, KNN e SVM, respectivamente.

A partir da análise das métricas apresentadas na Figura 2, é possível sintetizar os seguintes resultados. Primeiramente, em relação à Cobertura Marginal (CP), considerando o nível de confiabilidade de 90%, observa-se que os modelos de Regressão Logística e KNN alcançaram aproximadamente 91,1% de cobertura. Isso indica que, para esses modelos, cerca de 91% das amostras testadas continham o rótulo verdadeiro dentro de seus respectivos conjuntos preditivos. Por outro lado, o modelo SVM apresentou desempenho ligeiramente inferior, com cerca de 88,9% de cobertura.

No que diz respeito à Ineficiência, medida pelo tamanho médio dos conjuntos preditivos, verificou-se que as cardinalidades médias foram de 1,267 para a Regressão Logística, 1,289 para o KNN e 1,156 para o SVM. Assim, nota-se que o SVM produziu os conjuntos mais enxutos,

enquanto o KNN gerou os conjuntos mais amplos, refletindo diretamente o comportamento intrínseco de cada algoritmo.

Ao analisar a Taxa de Ambiguidade e a ocorrência de conjuntos vazios, constatou-se que nenhum dos modelos retornou conjuntos vazios, o que é desejável para aplicações práticas. Quanto à ambiguidade — definida como a presença de mais de um rótulo no conjunto preditivo —, o SVM apresentou a menor taxa, com 15,6% dos conjuntos contendo múltiplas respostas. Em contraste, o KNN exibiu a maior taxa de ambiguidade, alcançando 28,9%.

Por fim, quanto à distribuição das amostras classificadas nos conjuntos preditivos, considerando o conjunto de teste composto por 45 indivíduos, observou-se que a Regressão Logística resultou em 57 rótulos atribuídos (dos quais 12 ambíguos), o KNN em 58 rótulos (13 ambíguos) e o SVM em 52 rótulos (7 ambíguos). Essas diferenças refletem de maneira consistente as variações no tamanho médio dos conjuntos e nas taxas de ambiguidade previamente discutidas.

A taxa de incerteza pode ser analisada por meio dos nonconformity scores obtidos durante o processo de calibração. Como esses scores refletem o grau de “desajuste” entre a previsão do modelo e o rótulo verdadeiro, a análise de sua distribuição fornece uma visão direta sobre o comportamento de cada modelo em relação à incerteza. Dessa forma, o gráfico apresentado a seguir ilustra como cada modelo distribui seus nonconformity scores, permitindo comparar de forma mais detalhada o nível de confiabilidade entre eles.

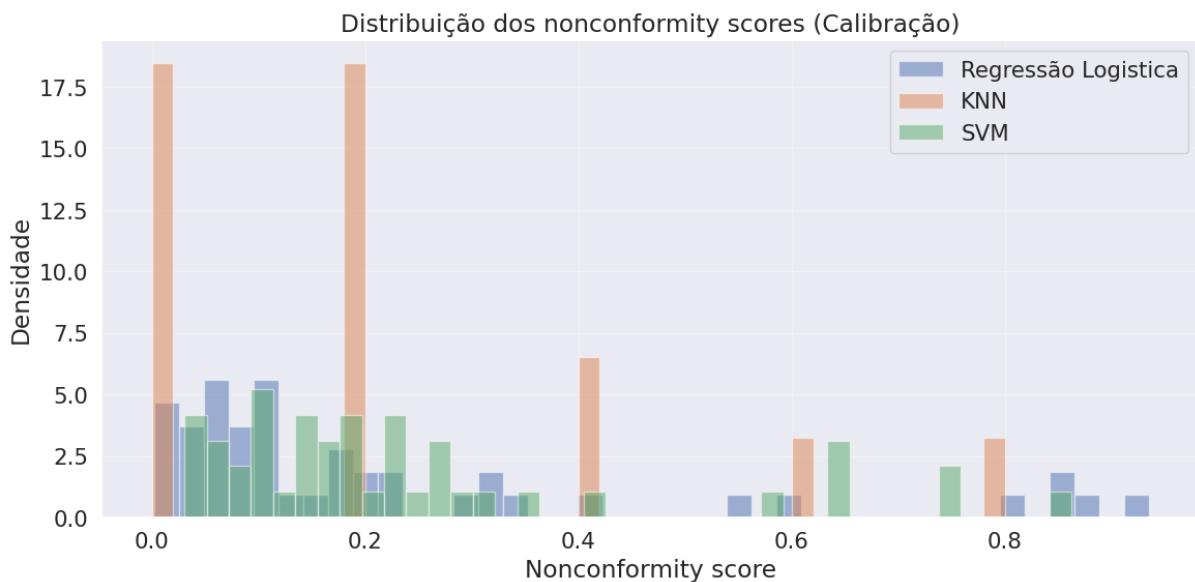


Figura 3 – Distribuição de não conformidade dos nossos modelos.

Pela Figura 3, observamos que o KNN apresenta uma distribuição discreta dos nonconformity scores, resultado direto do fato de suas probabilidades serem derivadas da proporção de vizinhos pertencentes a cada classe. Como utilizamos  $k=5$ , as probabilidades possíveis são quantizadas em poucos valores fixos. A Regressão Logística, por outro lado, exibe maior variabilidade nas extremidades da distribuição. Isso sugere que o modelo frequentemente atribui probabilidades

muito altas ou muito baixas às classes, indicando previsões feitas com alto grau de confiança, seja para acertos ou erros. O modelo SVM apresenta uma distribuição mais uniforme e espalhada dos scores, refletindo uma maior diversidade nos níveis de confiança atribuídos às previsões. Esse comportamento indica um modelo mais equilibrado, que raramente produz probabilidades extremas e tende a ser mais conservador na atribuição de certeza.

Uma análise particularmente relevante para o nosso cenário é observar o comportamento dos modelos conforme variamos o nível de confiança. Nesse contexto, quanto maior o nível de confiança desejado, mais restritivo o método se torna ao formar os conjuntos de predição, exigindo maior garantia de que a classe verdadeira esteja incluída. Essa variação impacta diretamente tanto o tamanho médio dos conjuntos quanto a cobertura obtida pelos modelos. Dessa forma, analisar essas curvas permite compreender como cada modelo responde ao ajuste do parâmetro  $\alpha$  e qual deles mantém o melhor equilíbrio entre cobertura e tamanho dos conjuntos.

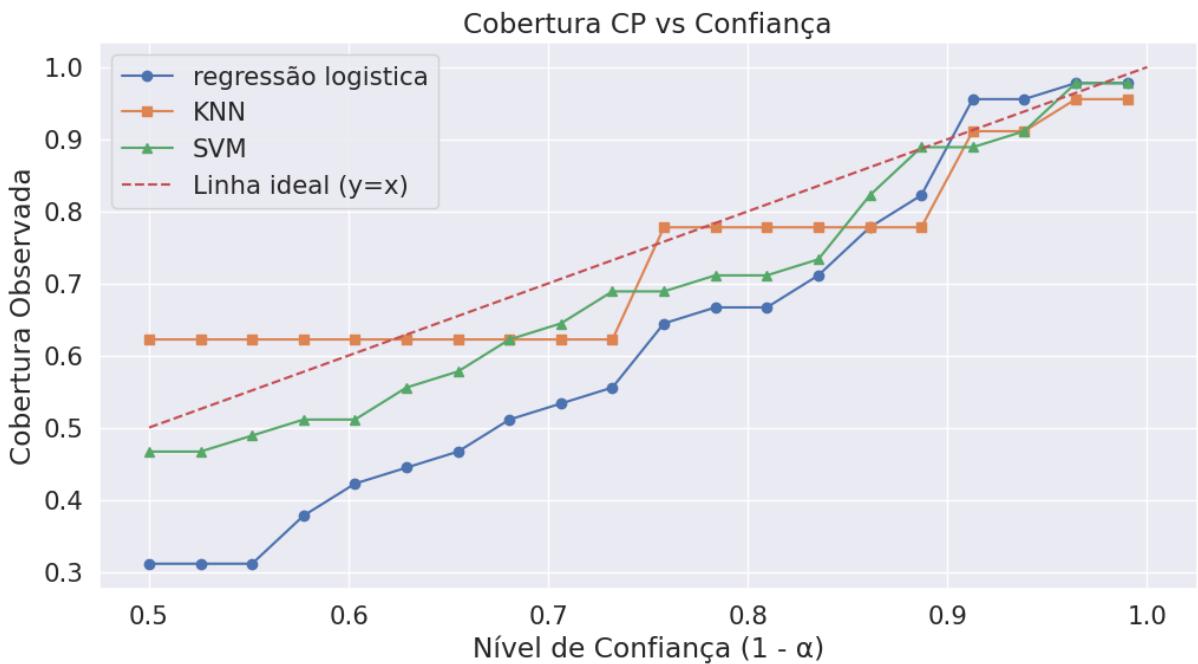


Figura 4 – Variação de cobertura do conjunto resposta pela variação da taxa de confiança.

A linha ideal representa o comportamento esperado teoricamente, no qual a cobertura cresce de forma linear conforme aumentamos o nível de confiança. Assim, se desejamos um nível de confiança de 50%, espera-se que a cobertura observada fique em torno de 50% ou acima desse valor, indicando que o método está calibrado. Ao analisar a variação da cobertura, observamos que o KNN — por apresentar previsões discretizadas — tende a produzir uma curva mais irregular, porém com tendências mais lineares em determinados intervalos, já que grandes variações na cobertura ocorrem apenas quando o valor de  $\alpha$  ultrapassa certos limiares associados aos seus nonconformity scores. O modelo SVM apresentou o crescimento de cobertura mais próximo da linha ideal ao longo da maior parte do intervalo de confiança,

superando o desempenho da Regressão Logística nesse aspecto. Entretanto, ao atingir níveis de confiança mais altos (como 90%), a Regressão Logística ultrapassa tanto o KNN quanto o SVM — e inclusive a própria linha ideal — alcançando sua melhor calibragem. Por outro lado, quando o nível de confiança se aproxima de 100%, todos os modelos passam a perder capacidade discriminativa: para garantir que a classe verdadeira esteja sempre presente, os conjuntos de predição tornam-se mais amplos, frequentemente incluindo todas as classes possíveis. Nesse cenário, a qualidade da análise diminui, já que o método deixa de fornecer informação útil para diferenciar as classes.

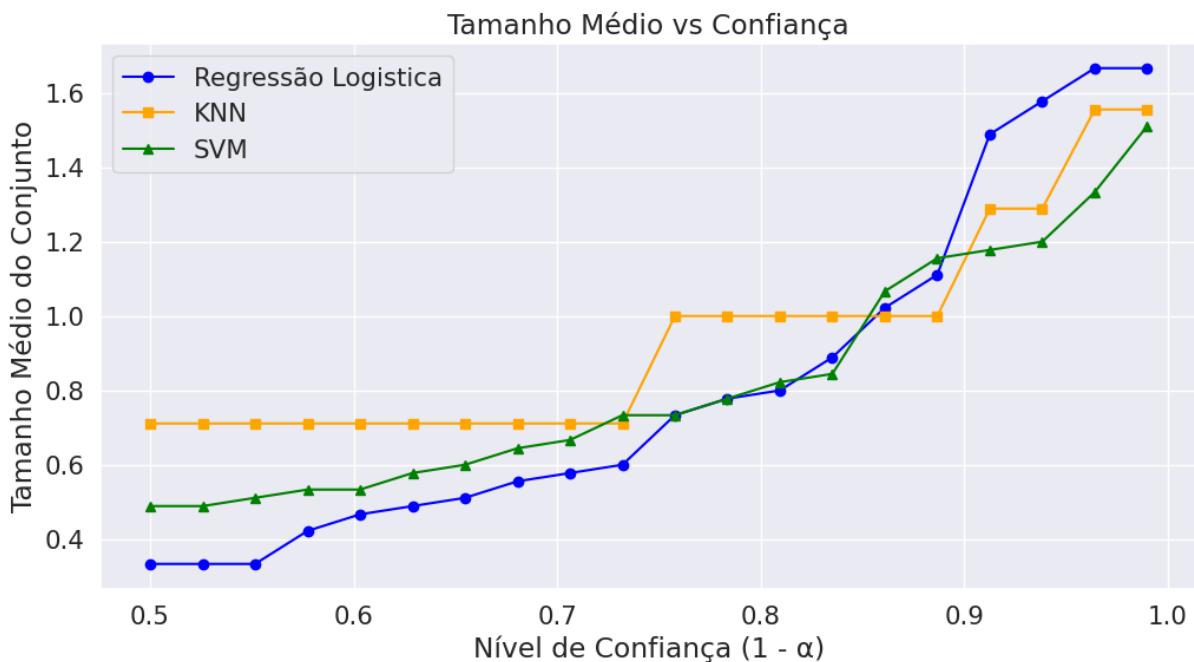


Figura 5 – Variação do tamanho do conjunto resposta dos nossos modelos pela variação da taxa de confiança.

Pela Figura 5, observamos que o tamanho médio dos conjuntos aumenta conforme elevamos o nível de confiança. Esse comportamento ocorre porque níveis mais altos de confiança exigem que o rótulo verdadeiro esteja contido no conjunto de predição; para atender a essa exigência, os modelos tendem a incluir mais classes no conjunto resposta, ampliando sua cardinalidade. Assim como em análises anteriores, o KNN apresenta um crescimento quase linear devido à natureza discreta de suas probabilidades.

O modelo SVM, por sua vez, exibe tamanhos médios ligeiramente maiores quando o nível de confiança é baixo, mas passa a ter o menor tamanho médio entre os três modelos quando a confiança se aproxima dos valores mais elevados. Já a Regressão Logística apresenta um aumento mais acentuado ao redor do nível de confiança de 90%, com a cardinalidade média subindo aproximadamente de 1,1 para 1,7, indicando um ajuste mais sensível do modelo nessa região.

## 5 Conclusão

Os resultados obtidos neste estudo demonstram que o uso de Conformal Prediction adiciona uma nova camada de interpretabilidade e segurança aos modelos de classificação, permitindo avaliar não apenas a acurácia, mas também o grau de incerteza associado a cada previsão. A análise conjunta das métricas tradicionais e das propriedades derivadas da calibração — como cobertura, tamanho médio dos conjuntos e distribuição dos nonconformity scores — evidenciou que os três modelos avaliados apresentam comportamentos distintos quando submetidos ao mesmo procedimento de conformalização.

Enquanto a Regressão Logística se destacou em níveis de confiança mais elevados, alcançando maior cobertura, o SVM apresentou um comportamento mais estável e próximo da linha ideal ao longo de todo o intervalo analisado. O KNN, por sua natureza discreta, mostrou variações mais abruptas, refletindo limitações estruturais do método quando calibrado no contexto do Conformal Prediction. Esses resultados reforçam que a escolha do modelo não deve se basear apenas na acurácia, mas também em como cada algoritmo responde às exigências de confiabilidade impostas pela conformalização.

A adição de mecanismos de quantificação de incerteza torna-se especialmente relevante em cenários nos quais a decisão tomada pelo modelo deve refletir com precisão a realidade observada. Em aplicações médicas, como no conjunto de dados utilizado neste estudo, é fundamental que potenciais casos de doenças sejam identificados antes de sua manifestação, reduzindo riscos e orientando intervenções preventivas.

Embora os três modelos tenham apresentado classificações semelhantes nas métricas tradicionais, a análise baseada em Conformal Prediction evidenciou diferenças importantes em termos de confiabilidade. Entre eles, a Regressão Logística demonstrou o comportamento mais consistente em níveis elevados de confiança, apresentando maior cobertura e melhor adaptação ao processo de calibração. Dessa forma, dentro do contexto analisado, este modelo se mostra o mais adequado para apoiar a avaliação da probabilidade de um paciente desenvolver ou não a condição estudada.

## Referências

IBM. O que é KNN (k-nearest neighbors)??. IBM Think. Disponível em: <https://www.ibm.com/br-pt/think/topics/knn>. Acesso em: 8 dez. 2025.

IBM. O que são máquinas de vetores de suporte (SVMs)??. IBM Think. Disponível em: <https://www.ibm.com/br-pt/think/topics/support-vector-machine>. Acesso em: 8 dez. 2025.

SCARABELO, Gabriel. Uma introdução prática à Previsão Conforme. Medium (Data Hackers), 26 jul. 2023. Disponível em: <https://medium.com/data-hackers/uma-introducao-pratica-a-previsao-conforme-de4c7479e021>. Acesso em: 4 out. 2025.

SCARABELO, Gabriel. Aplicando Previsão Conforme em modelos de classificação. Me-

dium (Data Hackers), 9 ago. 2023. Disponível em: <https://medium.com/data-hackers/aplicando-previsão-conforme-em-modelos-de-classificação-a26b2805ab0>. Acesso em: 4 out. 2025.

SCARABELO, Gabriel. Calibrando modelos de classificação binária com Previsão Conforme. Medium (Data Hackers), 30 ago. 2023. Disponível em: <https://medium.com/data-hackers/calibrando-modelos-de-classificação-binária-com-previsão-conforme-f547e68602ee>. Acesso em: 4 out. 2025.

SCARABELO, Gabriel. Quantificando incerteza de modelos de regressão com Previsão Conforme. Medium (Data Hackers), 23 ago. 2023. Disponível em: <https://medium.com/data-hackers/quantificando-incerteza-de-modelos-de-regressão-com-previsão-conforme-766fa03b2587>. Acesso em: 4 out. 2025.

SCHÄFER, J. et al. Conformal Prediction: A Data Perspective. *Frontiers in Big Data*, v. 6, 2023.

FOULDS, J. et al. Conformal predictions for probabilistically robust scalable machine learning classification. In: \*Proceedings of the 2020 SIAM International Conference on Data Mining\*. SIAM, 2020. p. 469–477.

TIBSHIRANI, R. Conformal Prediction. In: \*Advanced Topics in Statistical Learning\*. Pittsburgh: Carnegie Mellon University, 2023. Notas de aula.

WADOUX, A. M. J.-C. et al. Using Monte Carlo conformal prediction to evaluate the uncertainty of deep-learning soil spectral models. \*Geoderma\*, v. 433, p. 116428, 2023.