

Relatório CARD 23 - Prática: Métricas e Validação de Modelos de Aprendizado de Máquina (III)

Willian Augusto Soder de Souza

O objetivo deste relatório é apresentar os principais conhecimentos adquiridos ao assistir aos dois vídeos propostos no card, nos quais os apresentadores discutem diversas métricas para modelos de machine learning, abordando tanto a teoria quanto a prática. Abaixo, segue um resumo dos principais conceitos ensinados, acompanhado de um breve resumo.

- **Matriz de confusão:** é uma ferramenta utilizada para avaliar o desempenho de um modelo de machine learning, especialmente em problemas de classificação. Ela fornece uma visão detalhada das previsões do modelo, comparando as previsões feitas com os valores reais dos dados.

A matriz de confusão é uma tabela com quatro combinações possíveis de valores previstos e reais:

- **Verdadeiro Positivo (VP):** O modelo previu a classe positiva corretamente.
 - **Falso Positivo (FP):** O modelo previu a classe positiva incorretamente.
 - **Falso Negativo (FN):** O modelo previu a classe negativa incorretamente.
 - **Verdadeiro Negativo (VN):** O modelo previu a classe negativa corretamente.
- **Acurácia:** é uma métrica amplamente utilizada para avaliar o desempenho de modelos de machine learning, especialmente em problemas de classificação. Ela mede a proporção de previsões corretas feitas pelo modelo em relação ao total de previsões realizadas. A acurácia é definida como a razão entre o número de previsões corretas (tanto positivas quanto negativas) e o número total de previsões.
 - **Precisão:** é uma métrica usada em machine learning para avaliar a exatidão das previsões positivas feitas por um modelo de classificação. Ela é particularmente útil em situações onde o custo de falsos positivos é alto. A precisão é a proporção de previsões positivas corretas em relação ao total de previsões feitas para a classe positiva.
 - **Recall:** é uma métrica usada em machine learning para avaliar a capacidade de um modelo de identificar corretamente as instâncias positivas de uma determinada classe. Ele mede a proporção de exemplos positivos reais que foram corretamente classificados pelo modelo. O recall é definido como a razão entre o número de verdadeiros positivos e a soma dos verdadeiros positivos com os falsos negativos.
 - **Especificidade:** é uma métrica usada em machine learning para avaliar a capacidade de um modelo de identificar corretamente as instâncias negativas de uma determinada classe. Ela mede a proporção de exemplos negativos reais que foram corretamente classificados como negativos pelo modelo. A especificidade é definida como a razão entre o número de verdadeiros negativos e a soma dos verdadeiros negativos com os falsos positivos.
 - **ROC curve:** é uma ferramenta gráfica usada em machine learning e estatística para avaliar o desempenho de modelos de classificação binária. Ela permite visualizar a relação entre o recall e a especificidade do modelo em diferentes limiares de decisão.

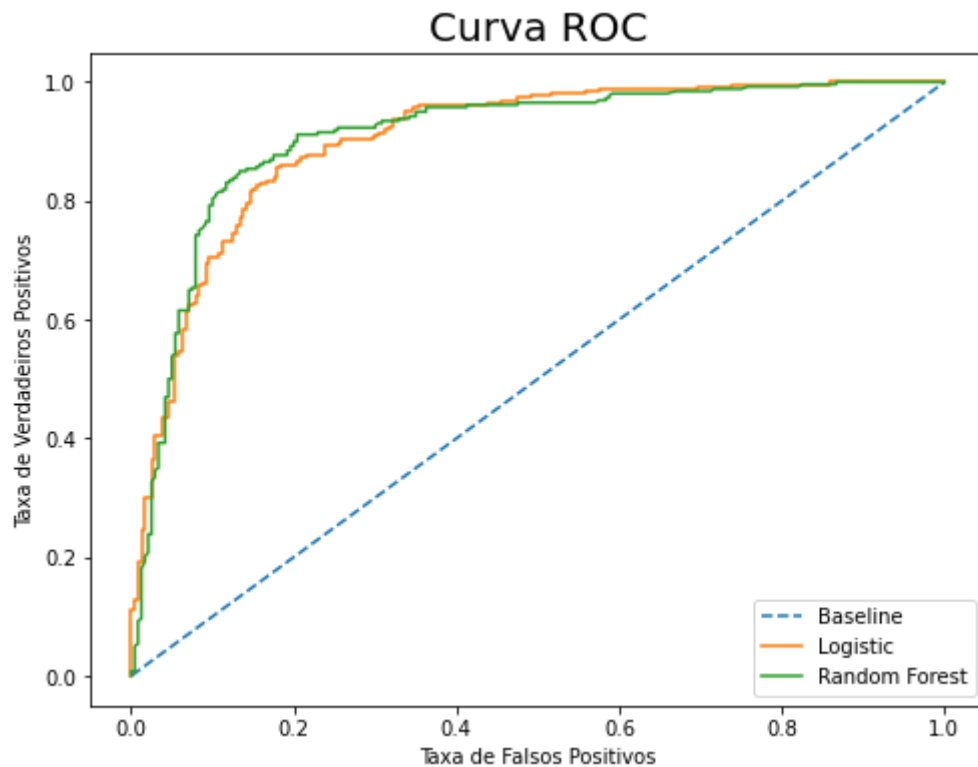


Figura 1 - Gráfico de curva ROC

- **AUC:** é uma métrica derivada da curva ROC usada para avaliar o desempenho de modelos de classificação binária. A AUC quantifica a área total sob a curva ROC, oferecendo uma medida única e agregada da capacidade do modelo em distinguir entre as classes positivas e negativas.

A AUC varia entre 0 e 1, e sua interpretação está relacionada à capacidade do modelo de classificar corretamente as instâncias em duas classes:

- **AUC = 1:** Indica um modelo perfeito que consegue separar as classes positivas e negativas sem cometer erros. Ou seja, para qualquer par de instâncias (uma positiva e outra negativa), o modelo sempre atribui uma pontuação mais alta à instância positiva.
 - **AUC = 0,5:** Representa um modelo sem capacidade discriminativa, equivalente a uma classificação aleatória. Se o modelo tiver uma AUC de 0,5, ele não consegue distinguir entre as classes positivas e negativas.
 - **AUC < 0,5:** Indica um modelo que, de forma consistente, classifica as instâncias negativas como positivas e as positivas como negativas, ou seja, está fazendo um trabalho pior que o aleatório. Se isso acontecer, geralmente significa que o modelo pode estar invertendo as previsões, ou que há algo fundamentalmente errado com a forma como o modelo foi treinado.
- **SHAP (Shapley Additive exPlanations):** é uma técnica baseada na teoria dos valores de Shapley, usada para explicar de forma interpretável as previsões de modelos de machine learning. Ele atribui a cada variável de entrada uma contribuição específica para o resultado da predição, garantindo que essa atribuição seja justa e consistente. Sua principal função é tornar modelos complexos, como redes neurais e random forests, mais transparentes, permitindo entender como cada variável afeta individualmente ou em combinação o resultado final. Além disso, o SHAP pode ser usado tanto para explicações locais, de uma predição específica, quanto para análises globais, revelando o comportamento geral do modelo.

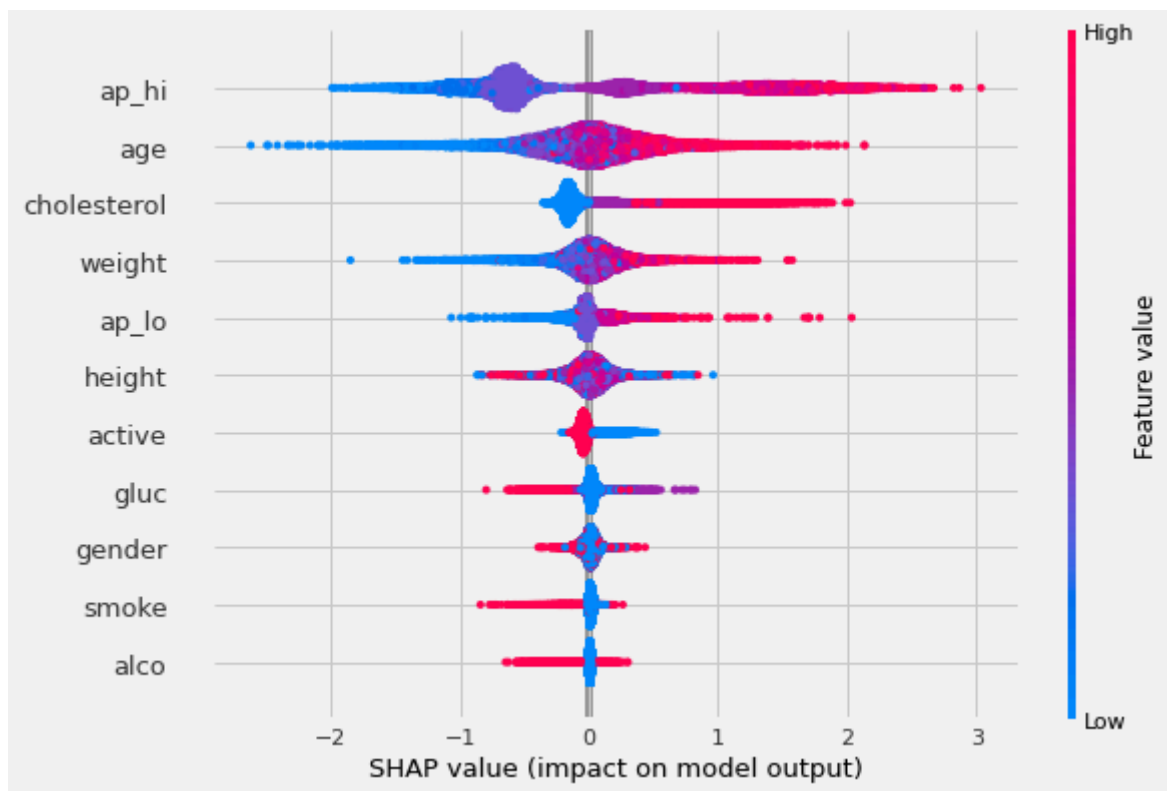


Figura 2 - Gráfico com valores de SHAP

CONCLUSÃO:

Compreender e aplicar métricas de validação como acurácia, precisão, recall, especificidade, curva ROC e AUC é crucial para avaliar e otimizar modelos de machine learning. Essas métricas oferecem uma visão detalhada do desempenho do modelo, permitindo identificar pontos fortes e limitações em diferentes contextos. O conhecimento dessas ferramentas, aliado a técnicas explicativas como o SHAP, que permite entender a contribuição de cada variável para as previsões, é essencial para construir modelos robustos, interpretáveis e ajustados ao cenário específico, garantindo resultados mais precisos, relevantes e transparentes.