



ANTONIO MENEGHETTI FACULDADE
Inteligência Artificial II

Trabalho de IA - Modelos Tradicionais

Hirruá Silva.

1. Proposta:

Problema: A saúde mental é uma preocupação crescente no ambiente de trabalho moderno, especialmente no setor de tecnologia, conhecido por sua alta competitividade e pressão por performance. Existe uma lacuna no entendimento sobre o que efetivamente leva um profissional a buscar ajuda para sua saúde mental, dificultando a criação de iniciativas de apoio eficazes e direcionadas.

Objetivo: O objetivo principal deste projeto é construir e avaliar um modelo de classificação supervisionada para identificar os principais fatores (preditores) que influenciam a decisão de um profissional de tecnologia em buscar tratamento formal para uma condição de saúde mental. Os objetivos secundários incluem:

- Analisar a importância relativa de fatores pessoais (ex: histórico familiar) versus fatores do ambiente de trabalho (ex: benefícios, cultura da empresa).
- Fornecer insights baseados em dados que possam guiar empresas na criação de ambientes de trabalho mais saudáveis e programas de apoio mais eficazes.

Dataset: O projeto utilizará o dataset "Mental Health in Tech Survey 2014", disponível publicamente na plataforma Kaggle. A escolha deste dataset se justifica por sua relevância para o problema em questão: [acesse aqui](#).

- Dados demográficos: idade, gênero.
- Fatores pessoais: histórico familiar de doenças mentais (*family_history*).
- Fatores do ambiente de trabalho: Tamanho da empresa (*no_employees*), se é uma empresa de tecnologia (*tech_company*), benefícios oferecidos (*benefits*), conhecimento sobre as opções de cuidado (*care_options*), e percepção sobre o estigma (*mental_health_consequence*, *coworkers*, etc.).

Avaliação: Dado que o problema é de classificação binária (o indivíduo buscou ou não tratamento), a acurácia é uma métrica inicial útil. No entanto, para garantir uma avaliação mais robusta e evitar conclusões enviesadas por um possível desbalanceamento de classes, a métrica principal de avaliação será o *F1-Score*. O *F1-Score* é a média harmônica entre a Precisão e o *Recall*, fornecendo um balanço entre os dois. Ele é particularmente útil em cenários onde tanto os falsos positivos (prever que alguém buscará tratamento quando não busca) quanto os falsos negativos (não prever que alguém buscará tratamento quando na verdade busca) são importantes.

$$Acurácia = \frac{Previsões\ Corretas}{Total\ de\ Previsões}$$

$$F1\ Score = 2 * \frac{Precisão * Recall}{Precisão + Recall}$$

Riscos: O dataset possui valores nulos em colunas importantes (ex: *work_interfere*, *self_employed*) e dados inconsistentes inseridos por usuários (ex: dezenas de variações na coluna *Gender*). Será realizada uma etapa de limpeza que inclui a padronização de categorias (como em *Gender*) e a imputação de valores faltantes. A imputação (ex: preenchimento pela moda) será realizada de forma a não causar vazamento de dados.

- O número de pessoas que buscaram tratamento pode ser significativamente diferente daquelas que não buscaram, o que pode enviesar o modelo a favor da classe majoritária.

2. Dados

A fase inicial consistiu em uma análise exploratória para compreender a estrutura do dataset. Foram inspecionados os tipos de dados, colunas com dados aleatórios ou diferentes, presença de valores nulos (*NaN*), distribuição de valores únicos em cada coluna categórica e a distribuição da variável alvo (*treatment*) para verificar o balanceamento das classes.

Outliers: A coluna *Age* apresentou diversos valores inconsistentes e não realistas (-1726, 5, 99999999999), sendo assim, aplicado um filtro para manter registros com idade entre 18 e 72 anos.

Inconsistências: A variável *Gender*, por ser um campo de texto livre, continha desenhos de entradas distintas. Realizada uma normalização para agrupar em 3 categorias: “Male”, “Female”, “Other”.

Dados faltantes: Valores ausentes em colunas categóricas como *self_employed* e *work_interfere*, foram tratados utilizando moda (valor mais frequente) como estratégia de preenchimento.

Engenharia de atributos: Para realizar a conversão dos dados brutos em formato numérico adequado para os algoritmos de *Machine Learning*.

- Mapeamento binário: colunas com respostas binárias (*yes/no*) foram mapeadas para 1 e 0, respectivamente.
- *Ordinal Encoding*: para variáveis categóricas com uma ordem clara, como *work_interfere* (*Never* < *Rarely* < *Sometimes* < *Often*) e *no_employees* (faixas de tamanho crescentes), foi aplicado um mapeamento manual para números inteiros (0, 1, 2, ...).
- *One-Hot Encoding*: para todas as demais variáveis categóricas nominais (sem ordem lógica), utilizando a função *pd.get_dummies*.

Divisão e validação: de modo a garantir integridade dos resultados e evitar *data leakage*, a divisão dos dados foi a primeira etapa principal do fluxo de trabalho. O dataset foi dividido em conjuntos de treino (80%) e teste (20%). A função *train_test_split* foi utilizada com parâmetro *stratify=y* para manter a mesma proporção da variável alvo em ambos os conjuntos. Todas as etapas de pré-processamento que dependem de cálculos agregados (como a imputação pela moda) foram "aprendidas" exclusivamente no conjunto de treino e, então, aplicadas ao conjunto de testes. Adicionalmente, *Cross-Validation* de 5 *folds* (*k-fold Cross-Validation*, *k*=5) foi aplicada sobre o conjunto de treino para obter uma estimativa robusta e estável da performance dos modelos.

O problema foi modelado como uma classificação binária, onde o objetivo é prever se um indivíduo buscou ou não tratamento para uma condição de saúde mental (*treatment*).

3. Modelos (IA Tradicional)

Para construção do modelo preditivo, o algoritmo selecionado foi o *RandomForestClassifier*, um modelo de *ensemble* conhecido por bom desempenho em dados tabulares e interpretabilidade nativa. A performance dos modelos foi avaliada utilizando um conjunto de teste isolado (20% dos dados). Adicionalmente, para garantir a robustez da avaliação do modelo principal, foi empregada uma Validação Cruzada (*Cross-Validation*) de 5 *folds* no conjunto de treino.

Hiperparâmetros: O modelo foi ajustado para evitar *overfitting* e otimizar a generalização, utilizando os seguintes hiperparâmetros:

- *n_estimators*: 300
- *n_jobs*:-1
- *max_depth*: 10
- *min_sample_leaf*: 10
- *min_sample_split*: 20
- *max_features*: “sqrt”
- *random_state*: 42

Baseline: foi utilizado um modelo *baseline* (*DummyClassifier*) com a estratégia *most_frequent*, em que classifica todas as instâncias com a classe majoritária do conjunto do treino, antes do treinamento com o *RandomForestClassifier*, onde o *DummyClassifier* atingiu 50,5% de acurácia.

4. Avaliação

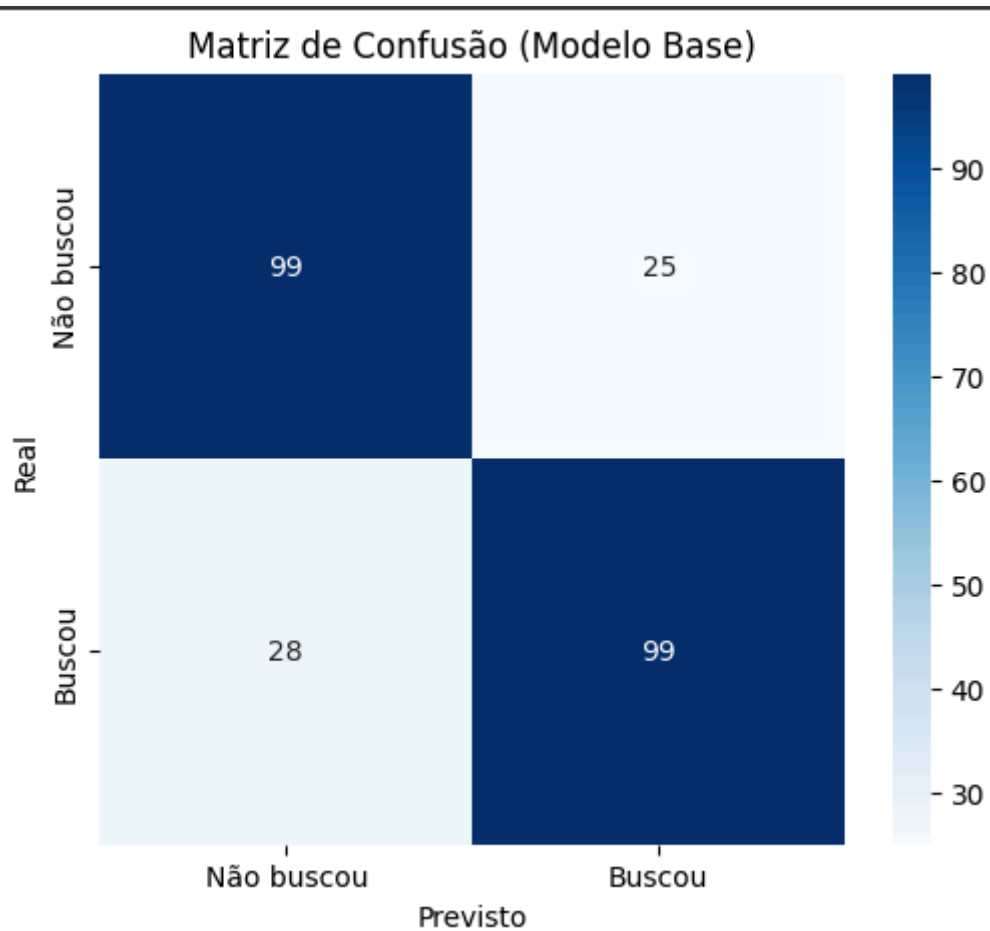
A tabela 1 apresenta um resumo comparativo da performance do modelo *RandomForest* contra o *DummyClassifier*. Foram utilizadas três métricas principais: Acurácia, *F1-Score*, *ROC-AUC*.

Tabela 1: DummyClassifier (baseline) vs Modelo RandomForest (base)

Métrica	DummyClassifier	RandomForestClassifier
Acurácia	50,6%	78,9%
F1-Score	67,2%	78,9%
ROC-AUC	0.5000	0.8549

Os resultados demonstram uma diferença do modelo *RandomForest*. A acurácia aumentou em mais em 28,3%, indicando um avanço. O *RandomForest* alcançou um F1-Score de 78.9% de forma equilibrada, conforme detalhado na Tabela 2 e na Matriz de Confusão (Figura 1). A pontuação ROC-AUC de 0.8549 confirma a excelente capacidade do modelo em distinguir entre as duas classes.

Figura 1: Matriz de Confusão (RandomForest)



A Matriz de Confusão ilustra que o modelo previu corretamente 99 casos para cada classe, com um número de erros (25 Falsos Positivos e 28 Falsos Negativos) bem distribuído.

$$Acurácia = \frac{198}{251} = 0.7888446215139442$$

$$Precisão = \frac{TP}{TP + FP} \quad Precisão = \frac{99}{99+25} = \frac{99}{124} \approx 0.7984$$

$$Revocação = \frac{TP}{TP + FN} \quad Revocação = \frac{99}{99+28} = \frac{99}{127} \approx 0.7795$$

$$F1 - Score = 2 * \frac{0.7984 * 0.7795}{0.7984 + 0.7795} = 2 * \frac{0.6223528}{1.5779} \approx 0.7888$$

A partir da Matriz de Confusão, é possível calcular as métricas de performance detalhadas para a classe de interesse ("Buscou tratamento"). Com 99 verdadeiros positivos, 25 falsos positivos e 28 falsos negativos, podemos derivar a Precisão e a Revocação do modelo. A Precisão, que mede a acurácia das previsões positivas, foi de aproximadamente 79.8%. A Revocação, mede a capacidade do modelo de encontrar todos os casos positivos, foi de 78.0%. O *F1-Score* representa a média entre essas duas métricas, resultando em aproximadamente 0.79. Este valor, consistente com o apresentado no Relatório de Classificação (tabela 2), confirma o bom equilíbrio do modelo entre não fazer previsões positivas incorretas e, ao mesmo tempo, ser eficaz em identificar os casos de interesse.

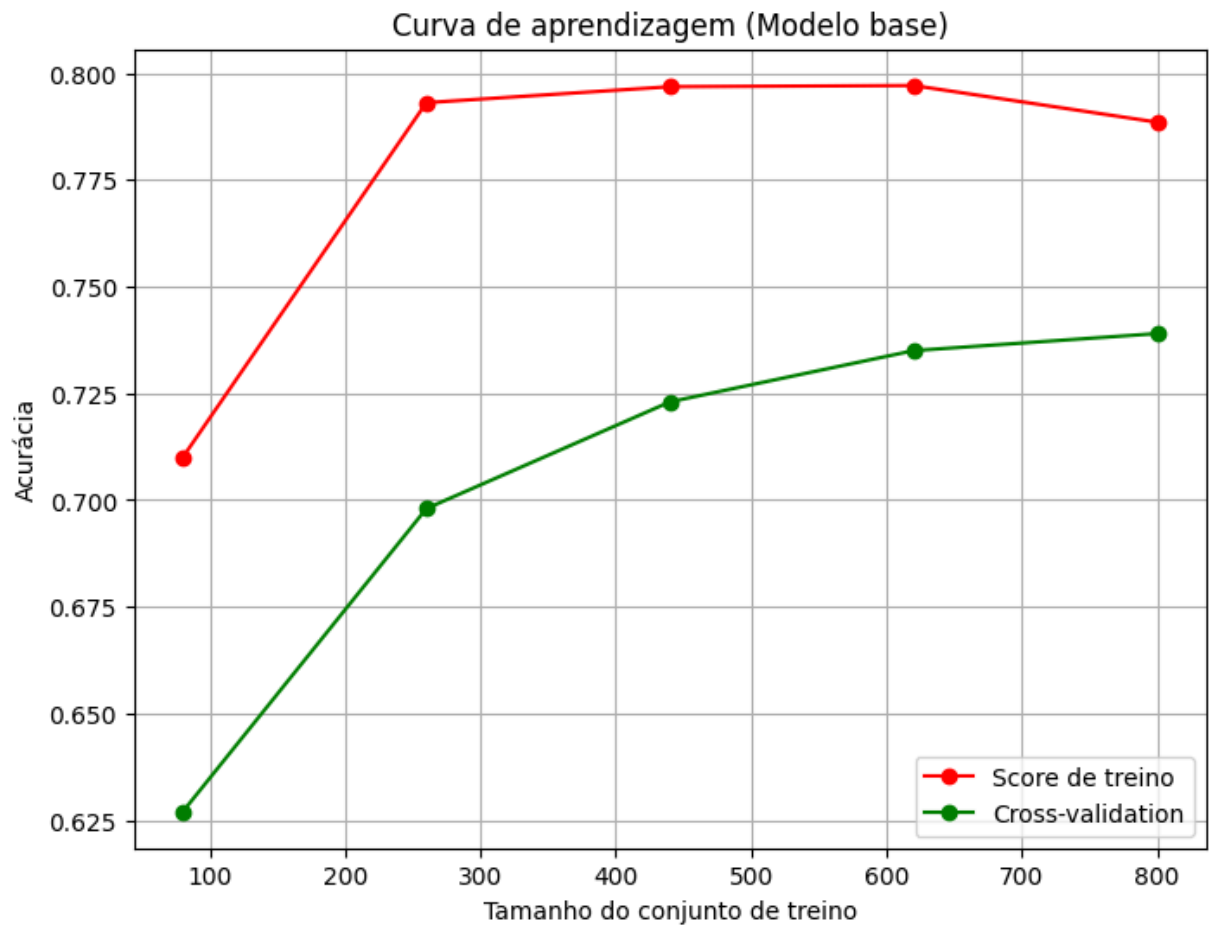
Tabela 2: Relatório de classificação detalhado (RandomForest)

	Precision	Recall	F1-Score	Support
Não buscou (0)	0.78	0.80	0.79	124
Buscou (1)	0.80	0.78	0.79	127

Com um F1-Score de 0.79 para ambas as classes, indicando um excelente equilíbrio entre precisão e revocação. Adicionalmente, o modelo alcançou uma pontuação ROC-AUC de 0.8549, reforçando sua forte capacidade de distinguir entre indivíduos que buscaram e não buscaram tratamento.

Overfitting e generalização: a robustez do modelo foi confirmada pela análise de overfitting, que comparou o desempenho de treino (0.7890 \approx 78.90%) com o de teste (0.7888 \approx 78.88%), mostrando uma diferença insignificante (0.0002 = 0.02%). A Curva de Aprendizagem (Figura 2) reforça visualmente este achado, com as curvas de treino e validação convergindo, o que indica um modelo bem generalizado.

Figura 2: Curva de Aprendizado (RandomForest)

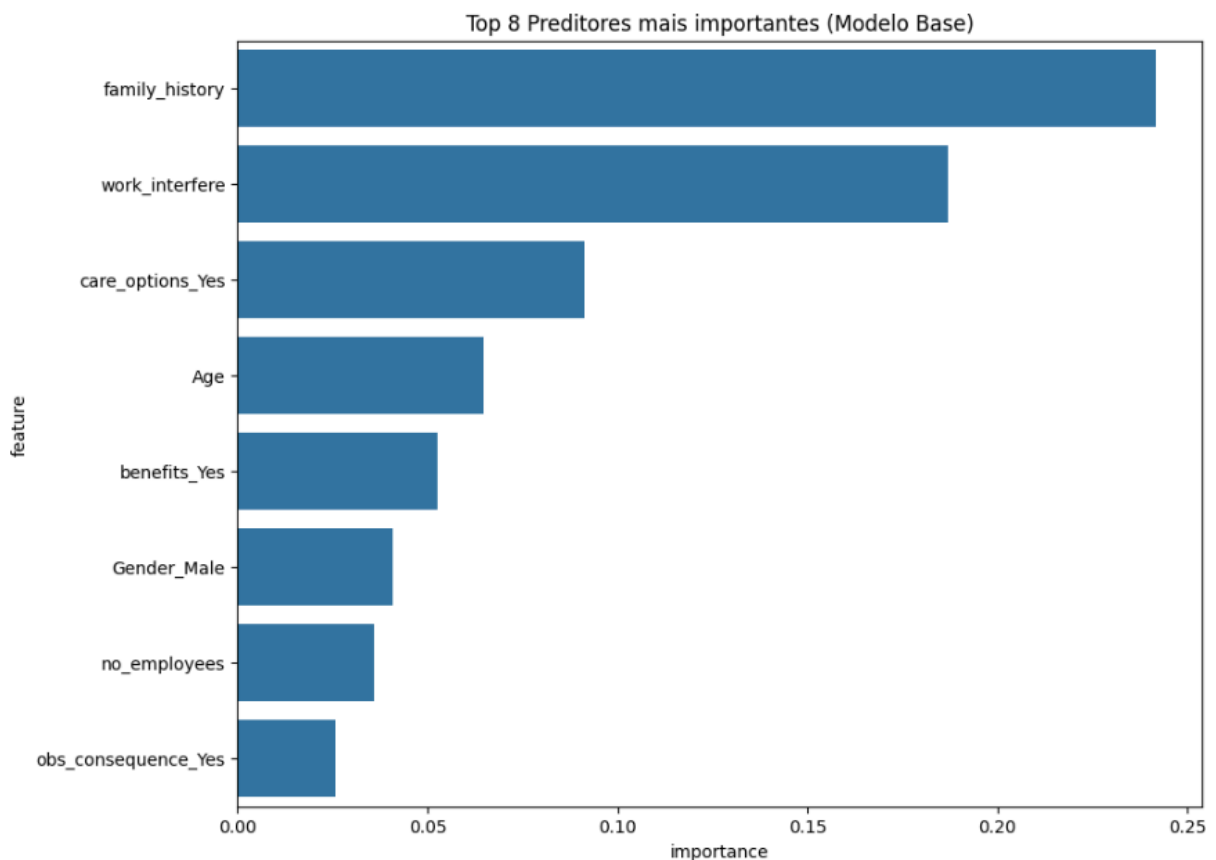


5. Interpretação

Após a validação da performance e da robustez do modelo RandomForest, a etapa final consiste em interpretar suas decisões para extrair insights e responder à pergunta central do projeto: "Quais são os principais fatores que influenciam a decisão de um profissional de tecnologia em buscar tratamento para sua saúde mental?".

Importância de features: calcula a contribuição média de cada variável para a redução do erro de classificação em todas as árvores do *ensemble*, fornecendo um *ranking* claro dos preditores mais influentes, como mostra a figura 3.

Figura 3: Top 8 Preditores mais importantes (RandomForest)



- *family_history* (≈ 0.24): preditor dominante, com uma importância consideravelmente maior que todos os outros. O resultado indica que indivíduos com histórico familiar de doença mental são significativamente mais propensos a buscar tratamento. Isso pode ser atribuído tanto a uma predisposição genética quanto a uma maior conscientização e menor estigma pessoal sobre o tema.
- *work_interfere* (≈ 0.19): O segundo fator mais crítico atua como um indicador da severidade da condição. Quando um problema de saúde mental começa a impactar a performance profissional, a busca por ajuda se torna uma necessidade mais imediata e concreta.
- *care_options_yes* (≈ 0.09) e *benefits_Yes* (≈ 0.05): dois fatores mostram a importância do suporte explícito oferecido pela empresa. É notável que saber das opções de cuidado (*care_options_yes*) é mais influente do que simplesmente saber que existem benefícios (*benefits_Yes*). Isso sugere que a comunicação e a clareza sobre os recursos

de saúde mental são mais cruciais do que sua mera existência.

- *Age*, *Gender_Male*, *no_employees*: fatores demográficos e estruturais também desempenham um papel. A idade (*Age*), o gênero (*Gender_Male*) e o tamanho da empresa (*no_employees*) são utilizados pelo modelo para refinar suas previsões, embora com um peso menor que os fatores pessoais e de suporte direto.

A análise de interpretabilidade revela que a decisão de buscar tratamento é multifatorial, mas com uma hierarquia clara. Fatores intrínsecos ao indivíduo (histórico e severidade) são os mais fortes, seguidos por fatores diretamente ligados ao suporte explícito e à cultura da empresa.

- Comunicação: as empresas devem focar não apenas em oferecer benefícios, mas em comunicá-los de forma proativa e clara, garantindo que os funcionários saibam quais são suas opções de cuidado (*care_options*).
- Ambiente seguro: a importância de *work_interfere* e *obs_consequence* sugere que um ambiente onde os funcionários se sintam seguros para admitir dificuldades sem medo de retaliação é fundamental.
- Fatores pessoais: as iniciativas de saúde mental devem reconhecer que muitos funcionários já chegam à empresa com predisposições (*family_history*), reforçando a necessidade de um suporte abrangente e desestigmatizar desde o primeiro dia.

6. Refinamento - Engenharia de atributos

Com base nos insights gerados pelo Modelo base, uma etapa de refinamento foi conduzida para testar a hipótese de que a criação de novas features de interação poderia melhorar a performance preditiva. Foram criadas variáveis sintéticas como *high_risk_support* e *support_gap* para capturar relações mais complexas entre os preditores. Para avaliar o impacto desta iteração, foi realizado um experimento controlado, comparando o Modelo base com um novo Modelo refinado. Ambos os modelos utilizaram o mesmo algoritmo (*RandomForestClassifier*) e os mesmos hiperparâmetros para uma comparação justa.

Tabela 3: Performance antes e depois - Engenharia de atributos

Métrica	Modelo base	Modelo refinado
Acurácia	78,9%	76,5%
F1-Score	78,9%	76,3%
ROC-AUC	0.8549	0.8526

Os resultados da Tabela 3 mostram que a inclusão das novas features resultou em uma pequena, mas consistente queda de performance em todas as métricas de avaliação. Isso sugere que, para este dataset, o *RandomForestClassifier* já extrai eficientemente as informações das variáveis originais, e as features de interação manual não adicionaram novo poder preditivo, possivelmente introduzindo ruído.

7. Conclusão

O objetivo deste projeto foi construir um modelo para identificar os principais preditores da busca por tratamento de saúde mental em tecnologia. Após um processo de pré-processamento, treinamento e iteração, a descoberta central é que o modelo base, sem engenharia de atributos adicional, demonstrou ser a solução mais performática e robusta. As tentativas de refinamento, embora metodologicamente válidas, não superaram a performance do modelo inicial, reforçando sua eficácia. A resposta final para a pergunta de pesquisa: "Quais são os principais fatores que influenciam a decisão de um profissional de tecnologia em buscar tratamento para sua saúde mental?". Ordem de importância:

- Histórico familiar (*family_history*)
- A interferência da condição no trabalho (*work_interfere*)
- Conhecimento sobre opções de cuidado (*care_options*)

Este projeto valida o princípio da parcimônia (Navalha de Ockham) no século XVII. De acordo com Matioli, a solução mais simples e direta, quando bem executada, pode ser a mais poderosa. O trabalho resultou em um modelo confiável com uma acurácia de 78.9% e insights claros e acionáveis.

8. Referência

MATIOLI, R., S. **Princípios gerais do método de parcimônia.** Disponível em: <<https://srmatioli.ib.usp.br/sbg2k/prinpar.html>>. Acesso em: 20 de set. 2025.