

**UNIVERSIDADE FEDERAL DE JUIZ DE FORA  
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM  
COMPUTACIONAL**

**Raíla Alves de Andrade**

**Ciência de Dados aplicada na sistematização da avaliação do potencial  
esportivo**

Juiz de Fora

2026

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF  
com os dados fornecidos pelo(a) autor(a)

Alves de Andrade, Raíla.

Ciência de Dados aplicada na sistematização da avaliação do potencial  
esportivo / Raíla Alves de Andrade. – 2026.

145 f. : il.

Orientadora: Priscila Vanessa Zabala Capriles Goliatt

Coorientadora: Flávia de Souza Bastos

Dissertação (Mestrado) – Universidade Federal de Juiz de Fora, . Pro-  
grama de Pós-Graduação em Modelagem Computacional, 2026.

1. Talento Esportivo. 2. Machine Learning. 3. Gold Score. I. Goliatt,  
Priscila Vanessa Zabala Capriles, orient. II. Título.

**Raíla Alves de Andrade**

**Ciência de Dados aplicada na sistematização da avaliação do potencial  
esportivo**

Dissertação apresentada ao Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Mestre em Modelagem Computacional

Orientadora: Prof<sup>ª</sup>. Dr<sup>ª</sup>. Priscila Vanessa Zabala Capriles Goliatt

Coorientadora: Prof<sup>ª</sup>. Dr<sup>ª</sup>. Flávia de Souza Bastos

Juiz de Fora

2026

**Raíla Alves de Andrade**

**Ciência de Dados aplicada na sistematização da avaliação do potencial esportivo**

Dissertação apresentada ao Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Mestra em Modelagem Computacional. Área de concentração: Modelagem Computacional

Aprovada em 26 de março de 2026.

BANCA EXAMINADORA

**Prof<sup>a</sup>. Dr<sup>a</sup>. Priscila Vanessa Zabala Capriles Goliatt** - Orientadora

Universidade Federal de Juiz de Fora

**Prof. Dr. Rodrigo Cesar Pedrosa Silva**

Universidade Federal de Ouro Preto

**Prof. Dr. Francisco Zacaron Werneck**

Universidade Federal de Juiz de Fora

Juiz de Fora, 29/03/2026.



Documento assinado eletronicamente por **Francisco Zacaron Werneck, Professor(a)**, em 31/03/2026, às 07:55, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Priscila Vanessa Zabala Capriles Goliatt, Professor(a)**, em 08/04/2026, às 21:49, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Rodrigo César Pedrosa Silva, Usuário Externo**, em 09/04/2026, às 12:14, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no Portal do SEI-Ufjf ([www2.ufjf.br/SEI](http://www2.ufjf.br/SEI)) através do ícone Conferência de Documentos, informando o código verificador **2925590** e o código CRC **007CEBA2**.

Dedico este trabalho a todos os pesquisadores e educadores que, movidos pela curiosidade e pelo amor ao saber, transformam o conhecimento em uma ponte para um mundo melhor. Aos atletas e talentos em descoberta, que, com cada passo, salto e sonho, inspiram a ciência a compreender o que há de mais belo no potencial humano.

## AGRADECIMENTOS

Agradeço, de forma muito especial, ao Prof. Jorge Perrou, por todos os ensinamentos dentro e fora das pistas, pela orientação nos treinos, pelas valiosas contribuições científicas e, sobretudo, pela amizade construída ao longo dessa caminhada. Sua presença foi fundamental para meu crescimento pessoal e profissional.

Ao meu irmão, que esteve ao meu lado em todos os momentos desses últimos anos, meu agradecimento mais profundo. Sua parceria, apoio e incentivo foram essenciais para que eu pudesse chegar até aqui.

Aos meus pais, pelo amor incondicional, pelo apoio constante e por sempre acreditarem em mim, mesmo nos momentos mais difíceis. Essa conquista também é de vocês.

À minha orientadora, Prof.<sup>a</sup> Priscila Vanessa Zabala Capriles Goliatt, pela orientação cuidadosa, paciência e por compartilhar seus conhecimentos ao longo de todo o processo deste trabalho.

À minha coorientadora, Prof.<sup>a</sup> Flávia de Souza Bastos, pela atenção, pelas contribuições valiosas e pelo incentivo constante.

Aos professores da Educação Física Francisco Zacaron Werneck e Emerson Filipino Coelho, pelo apoio e pela colaboração na obtenção e organização dos dados utilizados nesta pesquisa.

Ao Colégio Militar de Juiz de Fora (CMJF), pela parceria institucional e por viabilizar a realização das coletas de dados junto aos escolares, fundamentais para o desenvolvimento deste estudo.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e à Universidade Federal de Juiz de Fora (UFJF), pelo apoio financeiro por meio da concessão da bolsa de estudos, essencial para a realização desta pesquisa.

A todos os amigos que fizeram parte dessa trajetória: os da família, os da escola, os da FAEFID e os da primeira graduação. Cada conversa, gesto de apoio e momento compartilhado foram fundamentais para que eu chegasse até aqui.

Ao Rafael, que chegou há pouco tempo, mas já trouxe tanto carinho, apoio e companheirismo à minha vida.

Aos demais amigos, pela companhia, compreensão e palavras de encorajamento nos momentos mais desafiadores.

*“Sonhe, mas trabalhe duro. O talento é natural, mas sem disciplina ele não  
significa nada.”*

— Usain Bolt

## RESUMO

O talento esportivo é um constructo complexo e multifatorial, cuja identificação precoce é fundamental para o desenvolvimento de atletas. Nesse contexto, o Projeto Atletas de Ouro, desenvolvido pelo Laboratório de Estudos e Pesquisa do Exercício e Esporte (LABESPEE), propôs o *Gold Score*, um índice preditivo que utiliza uma abordagem multidimensional para estimar o potencial esportivo de jovens escolares. No entanto, o modelo original depende de avaliações subjetivas realizadas por treinadores, o que pode introduzir vieses e limitar sua aplicabilidade em contextos com escassez de profissionais especializados.

Diante disso, o objetivo desta dissertação foi desenvolver e avaliar modelos de aprendizado de máquina capazes de estimar o *Gold Score*, reduzindo a dependência de avaliações subjetivas do treinador. Para isso, foram realizados quatro experimentos progressivos, nos quais foram testados 13 algoritmos de aprendizado de máquina, incluindo técnicas de *boosting* (CatBoost, XGBoost, Gradient Boosting, LightGBM), regressões lineares e métodos baseados em árvores e distância. Os modelos foram otimizados por meio do algoritmo PSO (*Particle Swarm Optimization*) e avaliados com validação cruzada de 10 *folds*, utilizando uma base de dados com mais de 3.000 avaliações de escolares coletadas entre 2015 e 2023.

Os resultados demonstraram que os algoritmos baseados em *gradient boosting* (especialmente CatBoost e XGBoost) apresentaram o melhor desempenho preditivo. No cenário completo (com variáveis subjetivas), os modelos alcançaram  $R^2 > 0,99$  e identificaram todos os medalhistas entre os 20 escolares com maior potencial previsto. Mesmo após a remoção das avaliações subjetivas do treinador, os modelos mantiveram poder explicativo satisfatório ( $R^2 \approx 0,87$ ) e preservaram a capacidade de diferenciar selecionados de não selecionados (tamanhos de efeito grandes,  $d > 0,88$ ).

Além de otimizar o direcionamento de recursos para a avaliação aprofundada dos indivíduos, os achados ampliam a discussão sobre o uso de métodos computacionais como ferramentas de apoio à decisão em processos de identificação de talentos esportivos. A abordagem proposta pode ser replicada em diferentes contextos escolares, contribuindo para a equidade na avaliação ao reduzir a dependência de fatores subjetivos. Estes resultados apontam caminhos promissores para futuras investigações, incluindo a validação prospectiva dos modelos, a expansão da base de dados para outras instituições e o desenvolvimento de ferramentas computacionais acessíveis (planilhas, aplicativos web) para uso por professores e treinadores.

**Palavras-chave:** talento esportivo; identificação de talentos; aprendizado de máquina; ciência de dados; desempenho esportivo; *Gold Score*.

## ABSTRACT

Sports talent is a complex and multifactorial construct, and its identification is fundamental for athlete development. In this context, the Projeto Atletas de Ouro (Gold Athletes Project), developed by the Laboratory for Exercise and Sport Studies and Research (LABESPEE), proposed the *Gold Score*, a predictive index that adopts a multidimensional approach to estimate the sports potential of young individuals. However, the current model may introduce biases and limit its applicability in certain contexts due to its reliance on subjective evaluations performed by coaches. Therefore, the aim of this dissertation was to develop and evaluate machine learning models capable of estimating the *Gold Score*, reducing or eliminating the reliance on subjective assessments. To achieve this, four progressive experiments were conducted wherein 13 machine learning algorithms were tested, including boosting techniques (CatBoost, XGBoost, Gradient Boosting, LightGBM), linear regression, and tree- and distance-based methods. The models were optimized using the Particle Swarm Optimization (PSO) algorithm and evaluated via 10-fold cross-validation on a dataset of over 3,000 student assessments collected between 2015 and 2023. The results demonstrated that gradient boosting algorithms (especially CatBoost and XGBoost) exhibited the best predictive performance. In the complete scenario (including subjective variables), the models achieved an  $R^2 > 0.99$  and successfully identified all medalists among the 20 students with the highest predicted potential. Even after removing the coaches subjective evaluations, the models maintained satisfactory explanatory power ( $R^2 \approx 0.87$ ) and preserved the ability to differentiate selected from non-selected individuals (large effect size,  $d > 0.88$ ). In addition to optimizing resource allocation for in-depth individual assessments, the findings broaden the discussion on the use of computational methods as decision-support tools in sports talent identification processes. The proposed approach can be replicated in different school contexts, contributing to assessment equity by reducing reliance on subjective factors. These results point to promising directions for future research, including the prospective validation of the models, the expansion of the dataset to include other institutions, and the development of accessible computational tools (e.g., spreadsheets and web applications) for teachers and coaches. **Keywords:** sports talent; talent identification; machine learning; data science; sports performance; *Gold Score*.

## LISTA DE ILUSTRAÇÕES

Figura 1 – Analogia do Alvo: Visualização dos quatro cenários possíveis do <i>Trade-off</i> Viés-Variância. . . . .	33
Figura 2 – Ajuste do Modelo aos Dados: Cenários de Underfitting, Overfitting e Ajuste Ideal. . . . .	34
Figura 3 – Regressão Linear Simples: Visualização do ajuste da reta de regressão e destaque dos Resíduos (Erro Vertical). . . . .	35
Figura 4 – Support Vector Regression (Máquina de Vetores de Suporte para Regressão (SVR)): A Banda de Erro $\epsilon$ . Erros que caem dentro da margem de $\epsilon$ (região laranja) não são penalizados. Somente os pontos fora dessa banda contribuem para o custo do modelo, indicando a função de perda $\epsilon$ -insensível. . . . .	36
Figura 5 – Suavidade no SVR: Comparação do efeito do hiperparâmetro $C$ na suavidade da função de regressão. A Curva Verde ( $C = 1$ ) demonstra alta suavidade e baixa variância, enquanto a Curva Vermelha ( $C = 1000$ ) demonstra baixa suavidade e alta variância (sobreajuste). . . . .	38
Figura 6 – Support Vector Regression (SVR): Os Vetores de Suporte. Os pontos de treinamento destacados com círculos pretos e um preenchimento roxo são os vetores de suporte, que estão na borda da banda de erro $\epsilon$ ou a excedem. Estes pontos são fundamentais para definir o modelo de regressão. . . . .	39
Figura 7 – $k$ -NN para Regressão: Visualização do zoom do processo de predição para $k = 7$ . A predição $\hat{y}$ (linha laranja) é a média dos valores $Y$ dos 7 vizinhos mais próximos (em verde), selecionados com base na distância Euclidiana (horizontal) ao ponto de teste $X$ . . . . .	40
Figura 8 – Exemplo da estrutura de uma Árvore de Decisão Binária ilustrando o Particionamento Binário Recursivo. . . . .	42
Figura 9 – Esquema de funcionamento do método Bagging. . . . .	44
Figura 10 – Esquema do Random Forest: Múltiplas árvores treinadas em amostras bootstrap com seleção aleatória de features. . . . .	44
Figura 11 – Visualização geométrica do Gradient Descent. . . . .	45
Figura 12 – Fluxograma representativo do processo de aprendizagem sequencial no Boosting. . . . .	45
Figura 13 – Comparação das estratégias de crescimento: (A) <i>Level-wise</i> (tradicional); (B) <i>Leaf-wise</i> (LightGBM); (C) Árvore Simétrica (CatBoost). . . . .	48
Figura 14 – Fluxograma geral do percurso experimental. . . . .	55
Figura 15 – Fluxograma do Processo de Pré-processamento. . . . .	58
Figura 16 – Matriz de correlação de Pearson da base feminina. . . . .	67

Figura 17 – Matriz de correlação de Pearson da base masculina. . . . .	68
Figura 18 – Distribuição etária por ano (Masculina). . . . .	78
Figura 19 – Distribuição etária por ano (Feminina). . . . .	78
Figura 20 – Gráfico de dispersão entre valores reais e preditos pelo modelo Categorical Boosting (CatBoost) - Experimento I (Dados Femininos) . . . . .	82
Figura 21 – Gráfico de dispersão entre valores reais e preditos pelo modelo Extreme Gradient Boosting (XGBoost) - Experimento I (Dados Femininos) . . . . .	83
Figura 22 – Gráfico de dispersão entre valores reais e preditos pelo modelo CatBoost - Experimento I (Dados Masculinos) . . . . .	84
Figura 23 – Gráfico de dispersão entre valores reais e preditos pelo modelo Gradient Boosting - Experimento I (Dados Masculinos) . . . . .	85
Figura 24 – Importância das variáveis segundo valores SHapley Additive exPlanations (SHAP) - CatBoost (Experimento I, Dados Femininos) . . . . .	85
Figura 25 – Importância das variáveis segundo valores SHAP - XGBoost (Experimento I, Dados Femininos) . . . . .	86
Figura 26 – Importância das variáveis segundo valores SHAP - CatBoost (Experimento I, Dados Masculinos) . . . . .	86
Figura 27 – Importância das variáveis segundo valores SHAP - Gradient Boosting (Experimento I, Dados Masculinos) . . . . .	87
Figura 28 – Gráfico de dispersão - CatBoost (Experimento II, Dados Femininos) . . . . .	92
Figura 29 – Gráfico de dispersão - XGBoost (Experimento II, Dados Femininos) . . . . .	92
Figura 30 – Gráfico de dispersão - CatBoost (Experimento II, Dados Masculinos) . . . . .	93
Figura 31 – Gráfico de dispersão - XGBoost (Experimento II, Dados Masculinos) . . . . .	93
Figura 32 – Importância das variáveis segundo valores SHAP - CatBoost (Experimento II, Dados Femininos) . . . . .	94
Figura 33 – Importância das variáveis segundo valores SHAP - XGBoost (Experimento II, Dados Femininos) . . . . .	94
Figura 34 – Importância das variáveis segundo valores SHAP - CatBoost (Experimento II, Dados Masculinos) . . . . .	95
Figura 35 – Importância das variáveis segundo valores SHAP - XGBoost (Experimento II, Dados Masculinos) . . . . .	95
Figura 36 – Gráfico de dispersão - CatBoost (Experimento III, Dados Femininos) . . . . .	100
Figura 37 – Gráfico de dispersão - XGBoost (Experimento III, Dados Femininos) . . . . .	101
Figura 38 – Gráfico de dispersão - CatBoost (Experimento III, Dados Masculinos) . . . . .	101
Figura 39 – Gráfico de dispersão - XGBoost (Experimento III, Dados Masculinos) . . . . .	102
Figura 40 – Importância das variáveis segundo valores SHAP - CatBoost (Experimento III, Dados Femininos) . . . . .	103

Figura 41 – Importância das variáveis segundo valores SHAP - XGBoost (Experimento III, Dados Femininos) . . . . .	103
Figura 42 – Importância das variáveis segundo valores SHAP - CatBoost (Experimento III, Dados Masculinos) . . . . .	104
Figura 43 – Importância das variáveis segundo valores SHAP - XGBoost (Experimento III, Dados Masculinos) . . . . .	104
Figura 44 – Gráfico de dispersão - XGBoost (Experimento IV, Dados Femininos)	110
Figura 45 – Gráfico de dispersão - CatBoost (Experimento IV, Dados Femininos)	111
Figura 46 – Gráfico de dispersão - XGBoost (Experimento IV, Dados Masculinos)	111
Figura 47 – Gráfico de dispersão - CatBoost (Experimento IV, Dados Masculinos) . . . . .	112
Figura 48 – Importância das variáveis segundo valores SHAP - XGBoost (Experimento IV, Dados Femininos) . . . . .	112
Figura 49 – Importância das variáveis segundo valores SHAP - CatBoost (Experimento IV, Dados Femininos) . . . . .	113
Figura 50 – Importância das variáveis segundo valores SHAP - XGBoost (Experimento IV, Dados Masculinos) . . . . .	113
Figura 51 – Importância das variáveis segundo valores SHAP - CatBoost (Experimento IV, Dados Masculinos) . . . . .	114
Figura 52 – Comparação do $R^2$ do melhor modelo ao longo dos experimentos	117

## LISTA DE TABELAS

Tabela 1 – Indicadores de validade do Gold Score nos Jogos da Amizade. . . . .	27
Tabela 2 – Regras de Decisão Derivadas da Árvore Exemplo . . . . .	43
Tabela 3 – Principais variáveis com dados ausentes . . . . .	59
Tabela 4 – Variáveis presentes nas bases de dados por sexo . . . . .	60
Tabela 5 – Validação da implementação do cálculo do <i>Gold Score</i> . . . . .	62
Tabela 6 – Quantidade de registros inferidos por coluna (Base Masculina) . . . . .	64
Tabela 7 – Quantidade de registros inferidos por coluna (Base Feminina) . . . . .	65
Tabela 8 – Impacto da inferência por Cópulas no <i>Gold Score</i> . . . . .	66
Tabela 9 – Espaço de busca de hiperparâmetros por algoritmo . . . . .	72
Tabela 10 – Resumo Estatístico do Diagnóstico de Viés (ANOVA e Tukey HSD). . . . .	77
Tabela 11 – Síntese das configurações experimentais . . . . .	79
Tabela 12 – Ranking de performance dos modelos no conjunto de teste - Experimento I (Dados Femininos) . . . . .	81
Tabela 13 – Ranking de performance dos modelos no conjunto de teste - Experimento I (Dados Masculinos) . . . . .	81
Tabela 14 – Comparação entre ranking real e ranking previsto pelo CatBoost para os 20 maiores valores de <i>Índice de potencial esportivo (Gold Score)</i> no conjunto de teste (dados femininos) . . . . .	87
Tabela 15 – Comparação entre ranking real e ranking previsto pelo XGBoost para os 20 maiores valores de <i>Gold Score</i> no conjunto de teste (dados femininos) . . . . .	88
Tabela 16 – Comparação entre ranking real e ranking previsto pelo CatBoost para os 20 maiores valores de <i>Gold Score</i> no conjunto de teste (dados masculinos) . . . . .	89
Tabela 17 – Comparação entre ranking real e ranking previsto pelo Gradient Boosting para os 20 maiores valores de <i>Gold Score</i> no conjunto de teste (dados masculinos) . . . . .	90
Tabela 18 – Ranking de performance dos modelos no conjunto de teste - Experimento II (Dados Femininos) . . . . .	90
Tabela 19 – Ranking de performance dos modelos no conjunto de teste - Experimento II (Dados Masculinos) . . . . .	91
Tabela 20 – Distribuição da idade e do <i>Gold Score</i> dos participantes (masculino) por ano de avaliação . . . . .	96
Tabela 21 – Distribuição da idade e do <i>Gold Score</i> dos participantes (feminino) por ano de avaliação . . . . .	96
Tabela 22 – Ranking de performance dos modelos no conjunto de teste - Experimento III (Dados Masculinos) . . . . .	99

Tabela 23 – Ranking de performance dos modelos no conjunto de teste - Experimento III (Dados Femininos) . . . . .	100
Tabela 24 – Comparação Top 20 - CatBoost (Experimento III, Dados Femininos)	105
Tabela 25 – Comparação Top 20 - XGBoost (Experimento III, Dados Femininos)	106
Tabela 26 – Comparação Top 20 - CatBoost (Experimento III, Dados Masculinos) . . . . .	107
Tabela 27 – Comparação Top 20 - XGBoost (Experimento III, Dados Masculinos)	108
Tabela 28 – Ranking de performance dos modelos no conjunto de teste - Experimento IV (Dados Femininos) . . . . .	109
Tabela 29 – Ranking de performance dos modelos no conjunto de teste - Experimento IV (Dados Masculinos) . . . . .	109
Tabela 30 – Comparação Top 20 - CatBoost (Experimento IV, Dados Femininos)	115
Tabela 31 – Comparação Top 20 - XGBoost (Experimento IV, Dados Femininos)	116
Tabela 32 – Comparação Top 20 - CatBoost (Experimento IV, Dados Masculinos) . . . . .	117
Tabela 33 – Comparação Top 20 - XGBoost (Experimento IV, Dados Masculinos)	118
Tabela 34 – Síntese comparativa dos experimentos - Melhores modelos . . .	118
Tabela 35 – Evolução da capacidade de ordenação dos dois melhores modelos ao longo dos experimentos . . . . .	119
Tabela 36 – Comparação do desempenho treino vs teste para os dois melhores modelos . . . . .	121
Tabela 37 – Comparação das estatísticas descritivas entre escores reais e preditos	122
Tabela 38 – Comparação entre selecionadas e não selecionadas (dados femininos)	123
Tabela 39 – Comparação entre selecionados e não selecionados, e entre medalhistas e não medalhistas (dados masculinos) . . . . .	123
Tabela 40 – Dicionário de Variáveis do Projeto Atletas de Ouro . . . . .	135
Tabela 41 – Comparação entre ranking real e ranking previsto pelo CatBoost para os 20 maiores valores de Gold Score no conjunto de teste (Experimento II, Dados Femininos) . . . . .	139
Tabela 42 – Comparação entre ranking real e ranking previsto pelo XGBoost para os 20 maiores valores de Gold Score no conjunto de teste (Experimento II, Dados Femininos) . . . . .	140
Tabela 43 – Comparação entre ranking real e ranking previsto pelo CatBoost para os 20 maiores valores de Gold Score no conjunto de teste (Experimento II, Dados Masculinos) . . . . .	141
Tabela 44 – Comparação entre ranking real e ranking previsto pelo XGBoost para os 20 maiores valores de Gold Score no conjunto de teste (Experimento II, Dados Masculinos) . . . . .	142
Tabela 45 – Resultados completos de todos os modelos por experimento e sexo	143

## LISTA DE ABREVIATURAS E SIGLAS

- AM** Aprendizado de Máquina. 28, 30–32, 39, 49, 51, 52, 55, 125, 126, 128
- CatBoost** Categorical Boosting. 9–12, 28, 48, 49, 80, 82–89, 91–95, 98, 100–107, 111–115, 117, 139, 141
- CMJ** Countermovement Jump. 57, 68, 79
- CMJF** Colégio Militar de Juiz de Fora. 23, 56, 126
- GBM** Gradient Boosting Machine. 46, 47
- Gold Score** Índice de potencial esportivo. 11, 12, 15, 16, 20, 22, 23, 25–27, 56, 62, 63, 66, 68–70, 75, 76, 78, 80, 82, 83, 87–90, 96–98, 105, 114, 116, 121, 124–126, 128, 139–142
- GOSS** Gradient-based One-Side Sampling. 48
- IA** Inteligência Artificial. 29
- KNN** K-Nearest Neighbors. 39–41, 82
- LightGBM** Light Gradient Boosting Machine. 48, 49, 80
- MAE** Erro Médio Absoluto. 49, 50, 80
- MSE** Erro Quadrático Médio. 42
- PLN** Processamento de Linguagem Natural. 29
- PROESP-BR** Projeto Esporte Brasil. 20
- PSO** Particle Swarm Optimization. 70–73, 75, 79, 80
- RMSE** Raiz do Erro Quadrático Médio. 50, 80, 81, 90, 91, 99, 100, 108, 109
- RSS** Soma dos Quadrados dos Resíduos. 34, 42, 50
- SCMB** Sistema Colégio Militar do Brasil. 25
- SHAP** SHapley Additive exPlanations. 9, 10, 53, 54, 83, 85–87, 91, 94–96, 102–104, 107, 110, 112–114, 124–127
- SVM** Máquina de Vetores de Suporte. 35, 37
- SVR** Máquina de Vetores de Suporte para Regressão. 8, 35–39

**TCLE** Termo de Consentimento Livre e Esclarecido. 23

**UFJF** Universidade Federal de Juiz de Fora. 56

**UFOP** Universidade Federal de Ouro Preto. 23

**XGBoost** Extreme Gradient Boosting. 9–12, 28, 47–49, 80, 82–84, 86, 88, 91–95, 98, 101–106, 108, 110–113, 116, 118, 140, 142

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO . . . . .</b>	<b>18</b>
1.1	Definição e Complexidade do Talento Esportivo . . . . .	18
1.2	Modelos de Identificação e Desenvolvimento de Talentos . . . . .	18
1.2.1	Objetivos . . . . .	22
1.2.2	Objetivo Geral . . . . .	22
1.2.2.1	Objetivos Específicos . . . . .	22
<b>2</b>	<b>O ÍNDICE DE POTENCIAL ESPORTIVO (<i>Gold Score</i>) . . .</b>	<b>23</b>
2.1	Caracterização da Amostra e Ética . . . . .	23
2.2	Procedimentos e Instrumentos de Medida . . . . .	24
2.3	Cálculo do Índice de Potencial Esportivo . . . . .	25
2.4	CrITÉrios de Validação e Sucesso Esportivo . . . . .	25
2.5	Evidências de Validação do Gold Score . . . . .	26
2.5.1	Acurácia e Poder Preditivo . . . . .	26
2.5.2	Análise de Risco e Viés Maturacional . . . . .	27
<b>3</b>	<b>REFERENCIAL TEÓRICO . . . . .</b>	<b>28</b>
3.1	Inteligência Artificial . . . . .	29
3.2	Aprendizado de Máquina . . . . .	29
3.3	Categorias de Aprendizado . . . . .	30
3.4	Algoritmos de Regressão . . . . .	32
3.4.1	Modelos Lineares e Regularização . . . . .	33
3.4.2	Máquina de Vetores de Suporte para Regressão (SVR) . . . . .	35
3.4.3	K-Nearest Neighbors (KNN) . . . . .	39
3.4.4	Árvores de Decisão . . . . .	41
3.4.5	Métodos de <i>Ensemble</i> . . . . .	43
3.5	Métricas de Avaliação de Desempenho . . . . .	49
3.5.1	Erro Médio Absoluto (MAE) . . . . .	49
3.5.2	Raiz do Erro Quadrático Médio (RMSE) . . . . .	50
3.5.3	Coefficiente de Determinação ( $R^2$ ) . . . . .	50
3.6	Cópulas . . . . .	50
3.7	Otimização de Hiperparâmetros . . . . .	51
3.8	Validação de Modelos e Divisão de Dados . . . . .	52
3.9	Explicabilidade de Modelos: Importância de variáveis . . . . .	53
3.9.1	Importância Global via SHAP Médio . . . . .	53
3.9.2	Correlação de Pearson e Spearman . . . . .	54
<b>4</b>	<b>MATERIAL E MÉTODOS . . . . .</b>	<b>55</b>
4.1	Base de Dados . . . . .	56
4.2	Pré-processamento e Tratamento da Base de Dados . . . . .	57

4.2.1	Estratificação por Sexo . . . . .	60
4.2.2	Validação da fórmula do <i>Gold Score</i> . . . . .	62
4.2.3	Tratamento de Valores Ausentes . . . . .	62
4.2.4	Recálculo do <i>Gold Score</i> após a inferência dos dados . . . . .	63
4.2.5	Exploração e Redução de Variáveis . . . . .	66
4.3	Construção de Modelos Preditivos . . . . .	70
4.3.1	Otimização de Hiperparâmetros . . . . .	70
4.3.2	Avaliação dos Modelos: Validação Cruzada e Divisão Treino-Teste . . . . .	73
4.4	Delineamento dos Experimentos . . . . .	76
4.4.1	Experimento I: Modelo de Referência (Conjunto Integral) . . . . .	76
4.4.2	Experimento II: Modelo sem as variáveis subjetivas do treinador . . . . .	77
4.4.3	Experimento III: Remoção da variável <i>Ano_Avaliação</i> . . . . .	77
4.4.4	Experimento IV: Adição de Variáveis de Percentil . . . . .	78
<b>5</b>	<b>RESULTADOS E DISCUSSÃO . . . . .</b>	<b>80</b>
5.1	Experimento I: Modelo de Referência . . . . .	80
5.1.1	Desempenho Preditivo dos Modelos . . . . .	80
5.1.2	Análise Gráfica das Previsões . . . . .	82
5.1.3	Análise de Importância das Variáveis . . . . .	83
5.1.4	Capacidade de Ordenação dos Escolares com Melhor <i>Gold Score</i> . . . . .	83
5.2	Experimento II: Remoção das Variáveis Subjetivas do Treinador . . . . .	88
5.2.1	Desempenho Preditivo dos Modelos . . . . .	89
5.2.2	Análise Gráfica das Previsões . . . . .	91
5.2.3	Análise de Importância das Variáveis . . . . .	91
5.2.4	Capacidade de Ordenação dos Escolares com Melhor <i>Gold Score</i> . . . . .	98
5.3	Experimento III: Remoção da Variável Temporal . . . . .	98
5.3.1	Desempenho Preditivo dos Modelos . . . . .	98
5.3.2	Análise Gráfica das Previsões . . . . .	99
5.3.3	Análise de Importância das Variáveis . . . . .	102
5.3.4	Capacidade de Ordenação dos Escolares com Melhor <i>Gold Score</i> . . . . .	105
5.3.5	Síntese do Experimento III . . . . .	106
5.4	Experimento IV: Inclusão de Percentis . . . . .	107
5.4.1	Desempenho Preditivo dos Modelos . . . . .	108
5.4.2	Análise Gráfica das Previsões . . . . .	110
5.4.3	Análise de Importância das Variáveis . . . . .	110
5.4.4	Capacidade de Ordenação dos Escolares com Melhor <i>Gold Score</i> . . . . .	114
5.5	Síntese Comparativa dos Experimentos . . . . .	115
5.5.1	Análise da Capacidade de Ordenação . . . . .	119
5.5.2	Análise de <i>overfitting</i> . . . . .	120
5.6	Validação Externa dos Escores Preditos . . . . .	121

5.6.1	Comparação das Estatísticas Descritivas . . . . .	121
5.6.2	Diferenciação de Seleccionados e Medalhistas . . . . .	122
5.6.3	Síntese dos Principais Achados . . . . .	124
<b>6</b>	<b>CONCLUSÃO . . . . .</b>	<b>125</b>
6.1	Contribuições do Estudo . . . . .	125
6.2	Limitações do Estudo . . . . .	126
6.3	Recomendações para Trabalhos Futuros . . . . .	127
6.4	Considerações . . . . .	128
	<b>REFERÊNCIAS . . . . .</b>	<b>129</b>
	<b>APÊNDICE A – DICIONÁRIO DE VARIÁVEIS DO ESTUDO</b>	<b>135</b>
	<b>APÊNDICE B – Top 20 - Experimento II . . . . .</b>	<b>139</b>
B.1	Resultados para o Sexo Feminino . . . . .	139
B.2	Resultados para o Sexo Masculino . . . . .	139
	<b>APÊNDICE C – RESULTADOS COMPLETOS DOS EXPE-</b>	
	<b>RIMENTOS . . . . .</b>	<b>143</b>

# 1 INTRODUÇÃO

Antes de discutir o conceito de talento esportivo, é importante reconhecer que a excelência no esporte desperta interesse e questionamentos em diferentes públicos. O desempenho esportivo de alto nível chama a atenção e levanta dúvidas sobre os fatores que o explicam e o sustentam. Nesse contexto, o talento esportivo ganhou destaque global, não apenas na mídia, mas também na pesquisa científica, por seu caráter complexo e multifatorial. Ele vai além das capacidades físicas e esportivas, englobando criatividade, cultura e processos de aprendizagem, evidenciando a necessidade de compreender como esse potencial pode ser identificado e desenvolvido (ISSURIN, 2017).

## 1.1 Definição e Complexidade do Talento Esportivo

Não há consenso na literatura sobre o que caracteriza talento no esporte, uma vez que o termo assume diferentes significados conforme o contexto e a perspectiva adotada. Cobley et al. (2020) destacaram que definições populares do senso comum pouco contribuem para esse entendimento, pois frequentemente associam talento a habilidades inatas, consideradas estáticas ao longo do tempo.

No contexto esportivo, o termo é utilizado tanto para descrever o desempenho global de um atleta em relação aos seus pares quanto para se referir a capacidades específicas de aprender e executar habilidades técnicas com excepcionalidade, como a qualidade do passe, a tomada de decisão ou ações técnicas particulares. Dessa forma, talento representa conceitos distintos para diferentes pessoas, o que dificulta sua definição unificada (COBLEY et al., 2020).

Em uma tentativa inicial de esclarecer o conceito, Cobley et al. (2012) definiram talento esportivo como a qualidade — ou conjunto de qualidades — identificada em um momento anterior que promove ou prediz a excepcionalidade em um momento futuro. Essa definição foi retomada e discutida pelos autores em publicação posterior (COBLEY et al., 2020).

## 1.2 Modelos de Identificação e Desenvolvimento de Talentos

Para sistematizar a compreensão do talento esportivo, a literatura apresenta diferentes modelos que variam de acordo com o objetivo e com o ambiente em que o processo ocorre. Em relação ao objetivo, os modelos podem estar orientados para o rendimento esportivo, com foco na preparação de atletas para competições de alto nível, ou para o contexto educacional, buscando desenvolver habilidades motoras, hábitos saudáveis e a participação em atividades físicas. Quanto ao ambiente, o desenvolvimento pode ocorrer em clubes esportivos, com treinamento estruturado e acompanhamento técnico e científico,

ou em escolas, por meio das aulas de educação física e de intervenções pedagógicas, como programas de aprendizagem motora, jogos educativos e projetos esportivos extracurriculares que incentivam a participação e o engajamento dos estudantes. No contexto do esporte de alto rendimento, destaca-se o modelo clássico proposto por Williams e Reilly (2000), que divide o processo em:

1. **Detecção:** Descoberta de potenciais talentos que ainda não estão no sistema esportivo;
2. **Identificação:** Reconhecimento dos indivíduos com maior probabilidade de sucesso entre os praticantes;
3. **Desenvolvimento:** Oferta de treinamento adequado para potencializar as capacidades;
4. **Seleção:** Escolha dos melhores atletas para competições ou avanço de categoria.

Embora eficaz para o esporte de elite, essa abordagem seletiva pode não ser adequada ao ambiente escolar. Nesse sentido, Bailey e Morley (2006) propuseram uma adaptação desse processo para o apoio ao crescimento de todos os alunos com potencial na Educação Física. O modelo de Bailey e Morley para o contexto escolar estrutura-se em três etapas principais:

- **Identificação:** Reconhecimento dos estudantes que demonstram potencial em diferentes habilidades motoras e esportivas;
- **Prática e Desenvolvimento:** Garantia de oportunidades para que esses estudantes exercitem e aprimorem essas habilidades;
- **Suporte:** Oferta de apoio e recursos necessários (físicos e pedagógicos) para sustentar o aproveitamento desse talento.

Dessa forma, o modelo escolar assume um caráter mais amplo e inclusivo, valorizando o papel da Educação Física no desenvolvimento integral do aluno, e não apenas na formação de atletas. Entretanto, independentemente do modelo adotado, a fase de identificação constitui um denominador comum e essencial. Em ambos os contextos, faz-se necessário reconhecer as características e o potencial dos indivíduos para direcionar adequadamente as intervenções pedagógicas.

Além da proposta de Bailey e Morley (2006), outros autores também defenderam a importância da Educação Física escolar na fase de identificação e desenvolvimento de talentos esportivos (PION, 2015; MIRANDA et al., 2019; PRIETO-AYUSO et al., 2020; FABER et al., 2022; XIANG et al., 2022). Historicamente, os processos de identificação de

talentos no esporte basearam-se em baterias de testes físicos e antropométricos, visando selecionar indivíduos com perfis similares a atletas de elite (GARAY et al., 1974; HOARE, 1995). Contudo, essas baterias estimam o desempenho atual e frequentemente desconsideram variáveis importantes para medir o real potencial esportivo (BAILEY; MORLEY, 2006).

Segundo Werneck et al. (2022), potencial esportivo é a aptidão do indivíduo para o esporte. Trata-se de um conceito mensurável, multidimensional e dinâmico (que muda com o tempo), indicando a projeção de evolução do atleta no futuro.

Diversas metodologias tentaram quantificar essa aptidão. No cenário internacional, destacam-se sistemas inteligentes como o *Sport Interactive* no Reino Unido (ABBOTT; COLLINS, 2002), o sistema croata de identificação (PAPIĆ et al., 2009) e o *Flemish Sports Compass* na Bélgica (PION, 2015). No Brasil, iniciativas pioneiras pavimentaram esse caminho, como a Estratégia Z-Celafiscs na década de 1980 (MATSUDO, 1996) e, posteriormente, Projeto Esporte Brasil (PROESP-BR) (GAYA, 2002), focado na detecção de talentos motores em escolares (MELLO et al., 2016).

Alinhado à visão de identificar potenciais esportivos no ambiente escolar (BAILEY; MORLEY, 2006), foi criado o o Projeto Atletas de Ouro® (WERNECK et al., 2022). O projeto não se limita à análise isolada de desempenho físico ao adotar uma abordagem multidimensional — integrando aspectos biológicos, motores e a percepção de treinadores — para gerar o índice de potencial esportivo denominado *Gold Score* (WERNECK et al., 2024).

Tal índice visa identificar o jovem atleta que possui elevado potencial para se tornar um atleta profissional, apresentando desempenho superior aos seus pares e desenvolvendo suas habilidades em um ambiente propício. Essa abordagem multidimensional refletiu a conclusão de autores como Bailey e Morley (2006), que observaram uma mudança nos paradigmas de identificação: o foco deslocou-se apenas do desempenho para considerar aspectos contextuais, como saúde, maturação e fatores psicológicos.

Na construção do índice, as amostras foram consideradas independentes. Dessa forma, um mesmo aluno avaliado em diferentes anos foi tratado como indivíduos distintos na base de dados, uma vez que cada avaliação corresponde a um novo conjunto de características associadas ao seu estágio de desenvolvimento naquele momento.

Vale ressaltar que os aspectos metodológicos, as variáveis constituintes e a validação estatística do *Gold Score* serão detalhados oportunamente no Capítulo 2.

Embora sistemas multidimensionais, como o *Gold Score*, representem um avanço significativo ao integrar diferentes dimensões do potencial esportivo, a literatura tem alertado para os riscos associados à dependência excessiva de julgamentos subjetivos no processo de identificação de talentos. Entre esses julgamentos, destaca-se a avaliação do avaliador — amplamente conhecida como o “olho do treinador” — que, embora valorizada

na prática esportiva, apresenta limitações importantes do ponto de vista científico (LATH et al., 2021; HARTIGH et al., 2018; ROBERTS et al., 2019).

Nesse contexto, o que é o “olho do treinador”? O estudo de Lath et al. (2021) buscou definir esse termo e chega à conclusão de que ele pode ser compreendido como um processo intuitivo e holístico de avaliação e seleção de atletas, fundamentado na experiência e no julgamento subjetivo do treinador, e não exclusivamente em métricas objetivas de desempenho. A partir do conceito de senso prático, desenvolvido por Bourdieu (1998) no âmbito da teoria do habitus, Christensen (2009) argumentou que treinadores identificam talentos com base em impressões globais e experiências visuais acumuladas.

Ainda que a experiência prática dos treinadores seja valiosa, a observação é suscetível a vieses cognitivos e influências externas. Estudos indicam que treinadores tendem a confundir maturidade biológica com talento, favorecendo jovens que amadurecem mais cedo e possuem vantagens físicas momentâneas, em detrimento daqueles com maior potencial técnico ou coordenativo a longo prazo (BAKER et al., 2017; FURLEY; WOOD, 2010). Além disso, a subjetividade pode ser afetada por expectativas prévias e estereótipos, comprometendo a precisão do diagnóstico de potencial esportivo.

Conforme o estudo de Roberts et al. (2020), que propôs que treinadores avaliassem atletas sob condições idênticas ao longo de um período de observação, os níveis de concordância entre suas avaliações são baixos, indicando limitada confiabilidade interavaliadores.

Segundo Böhme (2000), não é possível prever o talento esportivo com exatidão. A autora destacou que experts desempenham um papel relevante nesse processo, porém as escolhas devem ser sustentadas por evidências científicas. Além disso, apontou a necessidade de reorganização do sistema esportivo brasileiro, de modo a possibilitar que um maior número de jovens tenha oportunidades de desenvolvimento ao longo do tempo. Embora o julgamento de especialistas seja amplamente reconhecido como um componente importante nos processos de identificação de talentos, a formação desses profissionais demanda longos períodos de prática deliberada e exposição a contextos competitivos específicos, o que limita sua disponibilidade e dificulta sua reprodução em diferentes realidades esportivas (LARKIN et al., 2019).

Nesse cenário, o uso de técnicas de aprendizado de máquina mostra-se pertinente como ferramenta complementar no processo de identificação de talentos esportivos. Ao permitir a análise simultânea de múltiplas variáveis e a identificação de padrões complexos nos dados. Importante destacar que essa abordagem não busca substituir a atuação do treinador, cuja experiência prática e conhecimento contextual são fundamentais, mas sim oferecer suporte à tomada de decisão. Em contextos nos quais não há disponibilidade de profissionais altamente experientes — como em ambientes escolares ou programas iniciais de detecção —, ferramentas baseadas em dados podem auxiliar na triagem e na identificação inicial de potenciais talentos.

### 1.2.1 Objetivos

#### 1.2.2 Objetivo Geral

Desenvolver e avaliar modelos de aprendizado de máquina para estimar o índice de potencial esportivo *Gold Score* em escolares, a partir de variáveis objetivas e sem a utilização de informações subjetivas do treinador, tendo em vista uma ferramenta que apoie a identificação de talentos esportivos.

##### 1.2.2.1 Objetivos Específicos

- Realizar o pré-processamento e a análise exploratória dos dados, utilizando técnicas de ciência de dados para preparação da base;
- Construir e treinar modelos de aprendizado de máquina para estimativa do *Gold Score* em diferentes cenários de entrada de variáveis e avaliar seus desempenhos;
- Comparar as estimativas geradas pelos modelos com os valores originais do *Gold Score* e com a classificação dos indivíduos no ranking (Top 20) no conjunto de teste;
- Analisar o impacto das variáveis na predição dos modelos, identificando aquelas com maior contribuição para a estimativa do potencial esportivo;
- Investigar a consistência das predições em termos de ordenação dos indivíduos (ranking), além da acurácia das estimativas numéricas;
- Verificar a possibilidade de uma ferramenta que funcione como apoio à triagem de talentos esportivos, com potencial implementação em ambiente computacional, voltada principalmente a contextos com limitada disponibilidade de avaliação especializada.

## 2 O ÍNDICE DE POTENCIAL ESPORTIVO (*Gold Score*)

O Projeto Atletas de Ouro® foi criado na Universidade Federal de Ouro Preto (UFOP), com estudo-piloto implantado em 2015 no Colégio Militar de Juiz de Fora Colégio Militar de Juiz de Fora (CMJF) (WERNECK et al., 2022). Essa iniciativa desenvolveu uma avaliação multidimensional que, por meio de um algoritmo, gerou o *Gold Score* original — um índice de potencial esportivo baseado em variáveis antropométricas, neuromotoras e na percepção subjetiva de treinadores.

Com a maturação do projeto, houve movimentos distintos de evolução. O primeiro foi a especialização, criando-se índices para modalidades específicas, como o *Gold Score Judo* (RIBEIRO, 2022), *Gold Score Athletics* (KEULEN et al., 2024) e o *Gold Score Basketball* (JÚNIOR et al., 2021). O segundo movimento, fundamental para a viabilidade escolar em larga escala, foi a otimização da bateria de testes. Conforme estabelecido por Werneck et al. (2024), foi consolidado o *Gold Score Reduzido*. Esta versão simplificada diminuiu o número de variáveis necessárias sem perder a validade estatística do índice original.

Considerando a validação da versão otimizada por Werneck et al. (2024), esta dissertação adotou o *Gold Score Reduzido* como a variável alvo (*target*). Para fins de padronização textual, ressalta-se que, adiante, todas as menções ao termo *Gold Score* referem-se exclusivamente a esta versão reduzida aplicada a escolares.

### 2.1 Caracterização da Amostra e Ética

A amostra foi composta por 1.064 estudantes (627 do sexo masculino e 437 do sexo feminino), com idades variando entre 13,0 e 17,9 anos, matriculados no CMJF. A coleta de dados ocorreu entre agosto de 2015 e junho de 2017. Para a avaliação do potencial esportivo e de variáveis subjetivas (aspectos intangíveis), contou-se com a expertise de dez docentes de Educação Física da instituição, especialistas em diversas modalidades e com média de experiência profissional de 12,5 anos.

O protocolo seguiu rigorosos critérios de elegibilidade, exigindo a assinatura do Termo de Consentimento Livre e Esclarecido (TCLE) pelos responsáveis e o assentimento dos escolares. O estudo original foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal de Ouro Preto (CAAE: 32959814.4.1001.5150) (WERNECK et al., 2024).

Os escolares foram submetidos à bateria de testes multidimensional do Projeto Atletas de Ouro®. O protocolo englobou a aplicação de questionários, coleta de medidas antropométricas, realização de testes físico-motores e a avaliação da resistência cardiorrespiratória. Todas as etapas foram conduzidas por uma equipe de profissionais treinados, sob a supervisão dos professores de Educação Física.

## 2.2 Procedimentos e Instrumentos de Medida

As variáveis do estudo foram organizadas em três dimensões principais: indicadores antropométricos, desempenho físico-motor e variáveis psicométricas. Abaixo, detalham-se as definições operacionais e os procedimentos de coleta para cada dimensão.

**Definições operacionais.** Para facilitar a compreensão, apresentam-se os conceitos técnicos adotados:

**Antropometria:** Mensuração das dimensões físicas (massa, estatura, dobras cutâneas).

**Maturação Somática:** Indicador da “idade biológica”, representando o avanço físico em direção à fase adulta.

**Estatura Adulta Prevista (EAP):** Estimativa matemática da altura final do escolar baseada nos pais.

**Força Explosiva (Potência):** Capacidade de exercer energia máxima em movimento curto e rápido.

**Medicine-ball:** Bola de peso calibrado (2 kg) para testes de potência.

**Sistema de Fotocélulas:** Cronômetro de precisão via sensores infravermelhos.

**Teste de Léger:** Teste de resistência aeróbica com ritmo sonoro progressivo (“bips”).

**Escala Likert:** Escala de concordância/avaliação (ex: 1 a 5 pontos).

**Aspectos Intangíveis:** Características subjetivas não mensuráveis fisicamente.

**Indicadores antropométricos e maturação.** Procedimentos para caracterização morfológica e biológica:

- **Medidas físicas:** massa corporal, estatura, envergadura e dobras cutâneas (equação de Slaughter et al. (1988)).
- **Maturação somática:** percentual da EAP pelo método de Khamis e Roche (1994).

**Indicadores físico-motores.** Avaliação do desempenho motor por testes padronizados:

- **Arremesso de *medicine-ball*:** força explosiva de membros superiores.
- **Corrida de 20 m:** velocidade (fotocélulas).
- **Teste de Léger:** resistência cardiorrespiratória.

**Variáveis subjetivas e psicométricas.** Instrumentos baseados na percepção do aluno e do treinador:

- **Competência percebida:** escala de Sheldon e Eccles (2005).
- **Olho do treinador:** avaliação subjetiva (Likert 1–5).
- ***Intangibles Checklist:*** Brown (2001).

### 2.3 Cálculo do Índice de Potencial Esportivo

O cálculo do *Gold Score* seguiu etapas metodológicas rigorosas de padronização dos dados, conforme descrito a seguir:

**Padronização e normalização.** Inicialmente, calculou-se o Escore Z para as variáveis contínuas, utilizando como referência os dados normativos publicados por Miranda et al. (2024). Para a variável velocidade, o Escore Z foi invertido, de modo que menores tempos representassem maior desempenho. Em seguida, os Escores Z foram convertidos em percentis (P), variando de 0 a 100.

Para a variável “Olho do Treinador”, realizou-se uma conversão direta: Ruim = 20, Regular = 40, Bom = 60, Muito Bom = 80 e Excelente = 100.

**Equação do índice.** O *Gold Score* final consiste na média aritmética dos percentis das seis variáveis, conforme apresentado na Equação 2.1:

$$Gold\ Score = \frac{P_{Força} + P_{Velocidade} + P_{Resistência} + P_{Competência} + P_{Intangíveis} + P_{Treinador}}{6} \quad (2.1)$$

onde  $P$  representa o valor em percentil de cada variável. O critério para classificação de “Talento Esportivo” foi estabelecido como o percentil 90 ( $P_{90}$ ), correspondente a uma pontuação superior a 75.

### 2.4 Critérios de Validação e Sucesso Esportivo

Para assegurar que o *Gold Score* efetivamente identifica potenciais talentos, a métrica foi validada utilizando como “padrão-ouro” o desempenho nos Jogos da Amizade.

Esta competição é o maior evento esportivo da Secretaria de Educação e Cultura do Exército, ocorrendo anualmente em diferentes sedes pelo Brasil. A competição reúne cerca de 1.500 atletas-alunos dos 15 Colégios Militares que compõem o Sistema Colégio Militar do Brasil (SCMB). Devido ao alto nível de organização e à seleção interna das equipes, os Jogos da Amizade são considerados importantes no contexto da performance esportiva escolar militar.

A validação da métrica estruturou-se em dois pilares estatísticos fundamentais:

- **Validade de Construto:** Investigada pela comparação entre alunos selecionados para representar o colégio nos Jogos e os não selecionados. O objetivo foi testar se o *Gold Score* diferenciava efetivamente os atletas de elite escolar (os top 10% da amostra).
- **Validade de Critério (Preditiva):** Verificada pela comparação entre alunos medalhistas e não medalhistas na competição nacional. Esta etapa testou a capacidade do índice em prever o sucesso competitivo real.

As análises comparativas utilizaram o teste t de Student para amostras independentes ( $p < 0,05$ ). Para a avaliação da validade de construto, foram considerados os grupos de alunos selecionados e não selecionados para os Jogos da Amizade, enquanto a validade de critério (preditiva) foi avaliada a partir da comparação entre medalhistas e não medalhistas na mesma competição. A relevância prática foi quantificada pelo tamanho do efeito ( $d$  de Cohen), classificado conforme os critérios de Cohen (1992): pequeno ( $d < 0,50$ ), médio ( $0,50 \leq d < 0,80$ ) e grande ( $d \geq 0,80$ ).

Adicionalmente, empregou-se a Razão de Chances, com intervalo de confiança de 95% (IC95%), para estimar a probabilidade de um aluno classificado como “talento” pelo índice tornar-se um medalhista nos Jogos da Amizade.

## 2.5 Evidências de Validação do Gold Score

A eficácia do *Gold Score* como métrica diagnóstica de talento foi confirmada por meio de análises de suas propriedades psicométricas. O modelo demonstrou consistência interna satisfatória ( $\alpha > 0,70$ ) e elevada estabilidade temporal em avaliações longitudinais, mantendo coeficientes de correlação intraclasse ( $CCI > 0,75$ ) em reavaliações após 12 e 24 meses (WERNECK et al., 2024).

### 2.5.1 Acurácia e Poder Preditivo

A validade do índice é sustentada por sua capacidade de discriminar grupos com diferentes níveis de rendimento. Estudantes selecionados para representar a instituição nos Jogos da Amizade apresentaram escores significativamente superiores aos seus pares, com magnitudes de efeito consideradas grandes ( $d \geq 0,80$ ) (WERNECK et al., 2024). Além da seleção, o índice mostrou-se um preditor para o sucesso competitivo real, especificamente na obtenção de medalhas (WERNECK et al., 2024).

A Tabela 1 sintetiza os indicadores de validade de construto e de critério observados na validação do modelo:

Tabela 1 – Indicadores de validade do Gold Score nos Jogos da Amizade.

<b>Grupo</b>	<b>Comparação</b>	<b>Gold Score (M ± DP)</b>	<b>p-valor</b>	<b>d</b>
<i>Masculino</i>	Selecionados vs. Não Sel.	69,1 ± 12,6 vs. 50,2 ± 16,4	< 0,001	1,15
<i>Feminino</i>	Selecionados vs. Não Sel.	69,1 ± 15,1 vs. 48,4 ± 16,3	< 0,001	1,27
<i>Masculino</i>	Medalhistas vs. Não Med.	74,0 ± 13,3 vs. 66,4 ± 11,5	0,005	0,66
<i>Feminino</i>	Medalhistas vs. Não Med.	76,4 ± 12,2 vs. 67,0 ± 15,3	0,047	0,61

Fonte: Elaborado pela autora.

### 2.5.2 Análise de Risco e Viés Maturacional

Ao aplicar o ponto de corte no percentil 90 (escore > 75), o modelo revelou que escolares classificados como talentos possuem uma probabilidade elevada de êxito competitivo. No sexo masculino, essa classificação resultou em uma razão de chances (*odds ratio*) **3,76 vezes maior** para a conquista de medalhas (WERNECK et al., 2024).

Um aspecto fundamental para a fidedignidade do *Gold Score* é que sua eficácia não sofreu interferência direta do estágio de maturação biológica dos avaliados ( $p > 0,05$  para ambos os sexos) (WERNECK et al., 2024).

### 3 REFERENCIAL TEÓRICO

O Aprendizado de Máquina (AM) tem se consolidado como uma ferramenta promissora na identificação de talentos esportivos, oferecendo vantagens sobre modelos tradicionais ao permitir a análise de interações complexas entre múltiplas variáveis (TANG et al., 2026). Essa abordagem justifica-se por três pilares fundamentais:

- **Modelagem de Relações Não-Lineares:** O talento esportivo é um fenômeno multidimensional e não-linear, no qual variáveis antropométricas, motoras e biológicas interagem de forma complexa (ABBOTT; COLLINS, 2002; MIRANDA et al., 2019). Diferentemente da estatística clássica, o AM — especialmente algoritmos baseados em árvores (como XGBoost e CatBoost) — captura essas interações e identifica padrões de potencial que não seriam evidentes em análises isoladas ou lineares (PION, 2015; JAMES et al., 2013).
- **Redução de Vieses e Escalabilidade:** Algoritmos de AM automatizam a triagem, reduzindo a influência de julgamentos subjetivos e estereótipos (LATH et al., 2021; ROBERTS et al., 2019). Isso possibilita a aplicação objetiva e escalável do modelo, inclusive em contextos escolares ou de iniciação esportiva, sem depender exclusivamente de especialistas.
- **Equidade e Análise por Subgrupos (Masculino e Feminino):** O AM permite criar modelos que consideram diferenças biológicas e de maturação entre os sexos (WILLIAMS; REILLY, 2000), garantindo precisão preditiva equilibrada para atletas masculinos e femininos (SHELDON; ECCLES, 2005; KEULEN et al., 2024).

Apesar do crescente interesse, em revisão bibliográfica recente, Tang et al. (2026) destacam que a aplicação de AM na identificação de talentos ainda é uma área em expansão, com grande parte dos estudos concentrados em modalidades específicas, especialmente futebol (TANG et al., 2026). Essa fragmentação evidencia a necessidade de pesquisas que explorem contextos mais amplos e diversas modalidades esportivas.

Para fundamentar a abordagem proposta neste estudo, torna-se necessário apresentar os conceitos teóricos que sustentam a utilização de técnicas de AM na análise de dados relacionados à identificação de talentos esportivos. Assim, este capítulo apresenta os principais fundamentos de inteligência artificial e aprendizado de máquina, bem como os algoritmos e métodos de avaliação de modelos utilizados para a predição do índice de talento considerado nesta pesquisa.

### 3.1 Inteligência Artificial

A definição de Inteligência Artificial (IA) é um tema complexo e polêmico, não havendo um consenso único na literatura. Conforme apontou Engelbrecht (2007), a dificuldade começa na própria definição de “inteligência”. Analisando definições de dicionário, o autor destacou a inteligência como a habilidade de compreender, aprender com a experiência e ter a capacidade de pensar e raciocinar. Contudo, ele também ressaltou a existência de aspectos subjetivos da inteligência, como criatividade, emoção e intuição, cuja réplica computacional é dificultada.

Engelbrecht (2007) também abordou a questão central: “Os computadores podem ser inteligentes?”. O autor destacou que Alan Turing, um dos pioneiros da área, defendia a ideia de que um computador bem projetado poderia, em tese, replicar as funções de um cérebro humano. Esta é uma visão que, no entanto, exclui ou simplifica diversos aspectos da cognição humana.

Buscando uma maior precisão terminológica, alguns autores como Poole et al. (1998) preferiram o termo “Inteligência Computacional”. Os autores argumentaram que a nomenclatura Inteligência Artificial pode gerar confusão, pois “artificial” é frequentemente interpretado como “não real” ou “falso”. Poole et al. (1998) também estabeleceram uma distinção crucial entre o propósito da área que é entender como o comportamento inteligente é possível e sua metodologia, que consiste em projetar, construir e experimentar com sistemas computacionais que realizam tarefas comumente associadas à inteligência.

Admitindo a dificuldade de definir o que é IA e buscando organizar diferentes visões, Russell e Norvig (2020) propuseram uma organização das diferentes abordagens em categorias, que historicamente se dividiram entre aquelas focadas na fidelidade ao desempenho humano (sistemas que pensam ou agem como humanos) e aquelas que preferiram uma definição mais abstrata e formal, focada na racionalidade (sistemas que pensam ou agem racionalmente, fazendo a “coisa certa”).

Essa abordagem focada em “agir racionalmente” é a base de uma definição de IA, que vê a inteligência através do conceito de agentes. Russell e Norvig (2020) definem um agente como qualquer entidade capaz de perceber seu ambiente através de sensores e agir sobre esse ambiente por meio de atuadores.

Em resumo, a IA pode ser vista como uma área multidisciplinar dedicada à criação de sistemas inteligentes, capazes de abordar e resolver problemas de complexidade crescente.

### 3.2 Aprendizado de Máquina

Como uma área multidisciplinar, a IA engloba diversas subáreas de pesquisa e aplicação (RUSSELL; NORVIG, 2020). Entre os ramos mais notáveis, destacam-se o Processamento de Linguagem Natural (PLN), a Visão Computacional, a Robótica, os

Sistemas Especialistas e o AM.

Dentre essas abordagens, o presente trabalho foca no AM.

A habilidade de aprender é um dos elementos mais importantes do comportamento inteligente (LUGER, 2008). Sem essa capacidade, um sistema computacional não age de forma inteligente: quando exposto ao mesmo problema pela segunda vez, ele não possui memória dos passos da solução e repete uma mesma sequência de cálculos.

O AM surge como uma solução para essa limitação. Trata-se de uma abordagem que permite aos programas aprenderem por conta própria — seja por meio da experiência, da analogia, de exemplos, de instruções supervisionadas ou ainda através do reforço (recompensas e penalidade) com base em seus resultados (LUGER, 2008).

Enquanto a definição anterior estabelece o conceito de aprendizado, na definição proposta por Mitchell (1997), o AM é definido formalmente: diz-se que um programa de computador aprende da experiência, com respeito a alguma classe de tarefas e medida de desempenho, se seu desempenho nas tarefas melhora com a experiência.

É importante notar que o termo “Experiência”, na definição de Mitchell, é um conceito amplo que representa a fonte de informação para o aprendizado. Dessa forma, a Experiência engloba os diferentes métodos, como o aprendizado por “exemplos” (no caso supervisionado) ou a interação com um ambiente (no aprendizado por reforço).

Formalmente, um problema de AM assume um conjunto de preditores  $X$  (variáveis de entrada) e uma resposta  $Y$  (variável de saída), ligados pela relação descrita na Equação 3.1.

$$Y = f(X) + \epsilon \quad (3.1)$$

onde,  $f$  é função fixa, porém desconhecida, que representa a relação entre  $X$  e  $Y$ . O termo  $\epsilon$  é o erro irreduzível, um ruído aleatório que representa toda a variação em  $Y$  que não pode ser explicada por  $X$ .

O objetivo central do AM não é encontrar a  $f$  verdadeira, mas sim estimá-la. Os Algoritmos de AM encontram a melhor aproximação  $\hat{f}$  para a função  $f$  aprendendo com os dados de treinamento. O modelo final é a função  $\hat{f}$ , selecionada de um vasto espaço de hipóteses por ser aquela que melhor generaliza a relação entre  $X$  e  $Y$  (JAMES et al., 2013).

### 3.3 Categorias de Aprendizado

Os métodos de AM são diferenciados pelo tipo de dados e *feedback* que recebem durante o treinamento (RUSSELL; NORVIG, 2020). Essas abordagens podem ser organizadas nas seguintes categorias:

**Aprendizado Supervisionado:** No aprendizado supervisionado, o valor de uma variável de resultado (ou de saída) é estimado ou previsto com base em um conjunto de variáveis de entrada (HASTIE et al., 2009). As variáveis de entrada podem ser chamadas de preditores ou, classicamente, variáveis independentes. Da mesma forma, a variável de resultado também é conhecida como resposta ou variável dependente. Os problemas supervisionados de AM podem ser subdivididos pela natureza da saída desejada do modelo (RUSSELL; NORVIG, 2020):

- **Classificação:** São problemas onde a saída esperada pertence a um conjunto finito de valores discretos, ou seja, uma categoria. Por exemplo, a identificação de e-mails (spam ou não spam) ou o reconhecimento de objetos em imagens (ex: “gato”, “cachorro”).
- **Regressão:** Envolve problemas onde a saída desejada é um valor numérico contínuo ou um vetor de valores. Problemas de regressão têm como objetivo prever um número, como o preço de uma ação no mercado financeiro ou a pontuação em um teste de aptidão.

**Aprendizado Não Supervisionado:** O algoritmo recebe os dados de entrada sem dispor de rótulos (*labels*) ou saídas correspondentes que sirvam como *feedback* (HASTIE et al., 2009). A tarefa do modelo é, portanto, descobrir estruturas, padrões ou associações ocultas nos próprios dados. Esta abordagem torna-se essencial diante das limitações do aprendizado supervisionado, que depende de dados previamente rotulados. Conforme aponta Alpaydin (2010), o processo de obtenção desses rótulos, que se assemelha a ter um “professor” que provê a resposta correta, é frequentemente difícil e oneroso.

Para lidar com a complexidade de dados não rotulados e a “Maldição da Dimensionalidade”, o aprendizado não supervisionado abrange um vasto conjunto de técnicas, como o agrupamento (*clustering*), regras de associação, detecção de anomalias e redução de dimensionalidade (PCA e Autoencoders) (NAEEM et al., 2019).

**Aprendizado Semissupervisionado:** O aprendizado semissupervisionado representa uma perspectiva híbrida, entre o aprendizado supervisionado e o não supervisionado. É utilizado quando existe uma pequena quantidade de dados rotulados e um grande volume de dados não rotulados (ZHU, 2005).

O aprendizado semissupervisionado parte do princípio de que os dados não rotulados carregam informações valiosas sobre a estrutura de distribuição dos dados. Baseando-se na hipótese de agrupamento (se dois pontos estão no mesmo cluster, provavelmente pertencem à mesma classe), o modelo utiliza essa estrutura para melhorar a precisão da classificação que seria obtida usando apenas os poucos dados rotulados (CHAPELLE et al., 2006).

**Aprendizado por Reforço:** O Aprendizado por Reforço foca em como um agente deve tomar ações em um ambiente para maximizar alguma noção de recompensa cumulativa (SUTTON; BARTO, 2018).

O agente não é treinado com pares de entrada-saída corretos. Em vez disso, o único *feedback* que recebe são “reforços”: sinais de recompensa ou punição que o ambiente retorna após cada ação. A finalidade é aprender uma política, isto é, uma estratégia que define a melhor ação a ser tomada em um determinado estado para maximizar a recompensa total ao longo do tempo (RUSSELL; NORVIG, 2020).

### 3.4 Algoritmos de Regressão

Os problemas de regressão supervisionada, cujo objetivo é prever um valor numérico contínuo, podem ser abordados por diversos algoritmos. Cada algoritmo possui uma abordagem metodológica distinta para estimar a função  $f$  que mapeia as variáveis de entrada  $X$  a uma saída  $Y$ .

Nesta seção, os algoritmos de regressão são descritos usando duas notações matemáticas equivalentes para um modelo linear:

- **Notação Estatística Clássica:**

$$f(\mathbf{X}) = \beta_0 + \sum_j \beta_j X_j,$$

onde  $\beta_j$  são os coeficientes (incluindo o intercepto  $\beta_0$ ).

- **Notação de AM (usada em SVMs e Redes Neurais):**

$$f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b,$$

onde  $\mathbf{w}$  é o vetor de pesos (equivalente aos  $\beta_j$ ),  $b$  é o *bias* (equivalente ao  $\beta_0$ ), e  $\langle \mathbf{w}, \mathbf{x} \rangle$  denota o produto escalar entre os pesos e as entradas.

**Trade-off Viés-Variância:** Um conceito fundamental na seleção e avaliação de algoritmos de AM é o equilíbrio entre a complexidade do modelo e sua capacidade de generalizar (JAMES et al., 2013).

O viés é o erro introduzido por um modelo excessivamente simples (subajuste ou *underfitting*), enquanto a variância é o erro introduzido por um modelo excessivamente complexo (sobreajuste ou *overfitting*). O objetivo é encontrar o meio-termo que minimize o erro total em dados de teste.

Para ilustrar este conceito, a Analogia do Alvo (Figura 1) visualiza a dispersão dos acertos como a Variância e a distância do centro como o Viés.

Em complemento, a Figura 2 compara os cenários de ajuste do modelo aos dados.

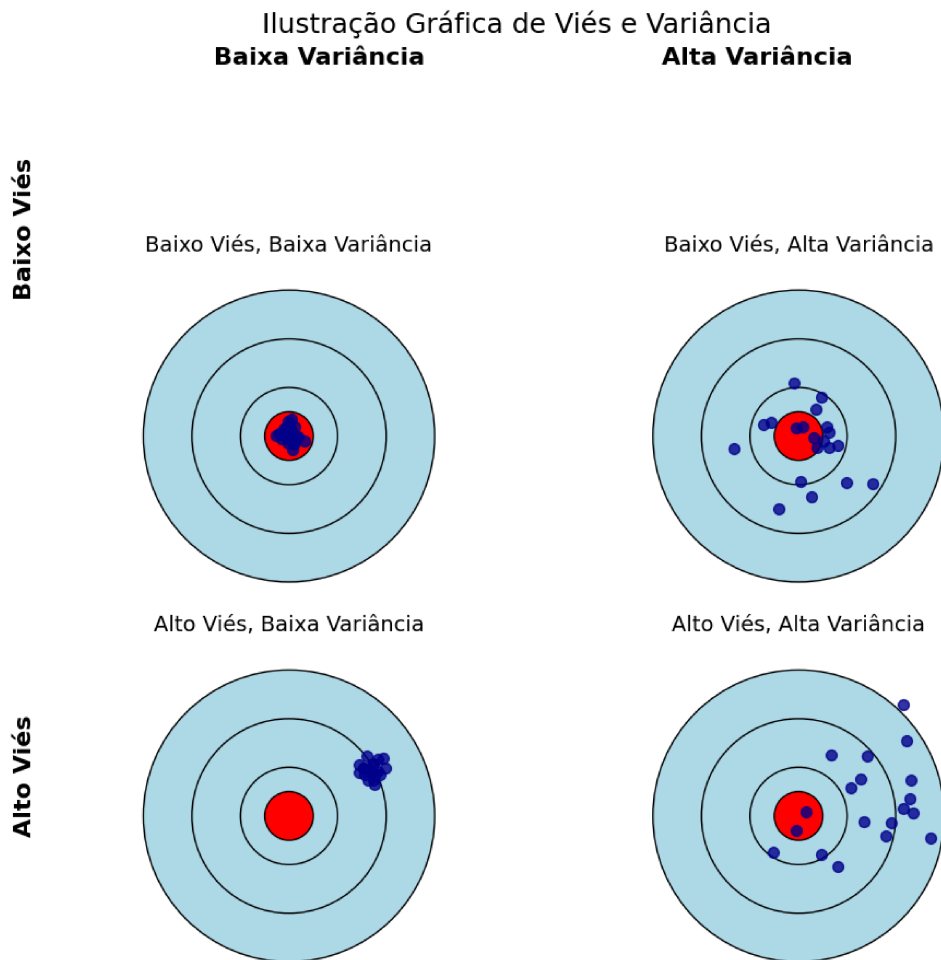


Figura 1 – Analogia do Alvo: Visualização dos quatro cenários possíveis do *Trade-off* Viés-Variância.

Fonte: Elaborado pela autora.

A seguir, os principais algoritmos de regressão são apresentados, agrupados por suas famílias conceituais.

### 3.4.1 Modelos Lineares e Regularização

A família dos modelos lineares representa a abordagem mais comum e interpretável para problemas de regressão. O objetivo desses modelos é estimar a resposta  $Y$  como uma combinação linear dos preditores  $X$ .

**Regressão Linear (Modelo Base):** O método fundamental desta família assume que a relação entre as variáveis é linear, seguindo a forma descrita por Hastie et al. (2009):

$$f(X) = \beta_0 + \sum_{j=1}^p X_j \beta_j \quad (3.2)$$

Onde  $\beta_0$  é o intercepto (ou *bias*) e  $\beta_j$  são os coeficientes (ou pesos) que representam

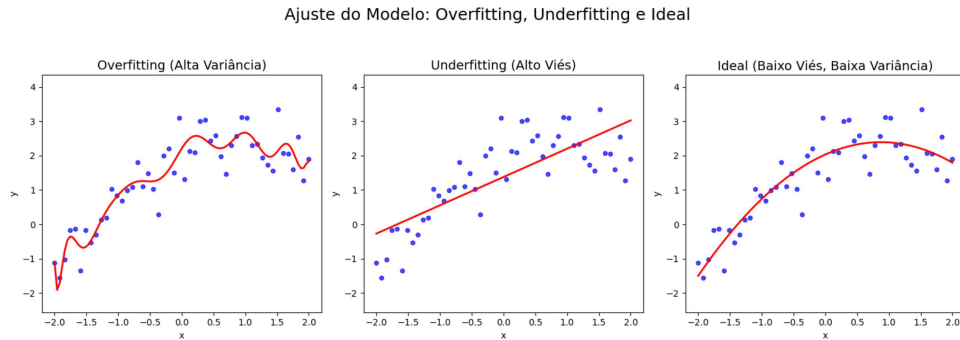


Figura 2 – Ajuste do Modelo aos Dados: Cenários de Underfitting, Overfitting e Ajuste Ideal.

Fonte: Elaborado pela autora.

a contribuição de cada *feature*  $X_j$  para a predição.

O objetivo do algoritmo é estimar os coeficientes ( $\hat{\beta}$ ) com base nos dados de treinamento. Para encontrar os “melhores” valores, utiliza-se o método de Mínimos Quadrados (*Least Squares*), que busca minimizar a Soma dos Quadrados dos Resíduos (RSS):

$$\text{RSS}(\beta) = \sum_{i=1}^N \left( y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 \quad (3.3)$$

A Figura 3 ilustra este conceito, onde a linha vermelha representa a reta ajustada e as linhas pontilhadas representam a distância que o método busca minimizar.

**Modelos Lineares Regularizados:** Apesar da simplicidade conceitual, o modelo padrão é suscetível a overfitting e instabilidade em casos de multicolinearidade. Para contornar isso, foram criadas versões regularizadas que adicionam uma “penalidade” à função de custo RSS para restringir a magnitude dos coeficientes.

As principais variações são:

- **Regressão Ridge (Regularização L2):** Adiciona o quadrado da magnitude dos coeficientes ( $\alpha \sum \beta_j^2$ ) à função de custo. É eficaz contra a multicolinearidade.
- **Lasso (Regularização L1):** Adiciona o valor absoluto da magnitude dos coeficientes ( $\alpha \sum |\beta_j|$ ). Além de reduzir o overfitting, é capaz de zerar coeficientes de features irrelevantes (seleção automática).
- **ElasticNet:** Uma combinação híbrida que aplica ambas as penalizações, L1 e L2, permitindo um balanço entre os benefícios do Ridge e do Lasso.

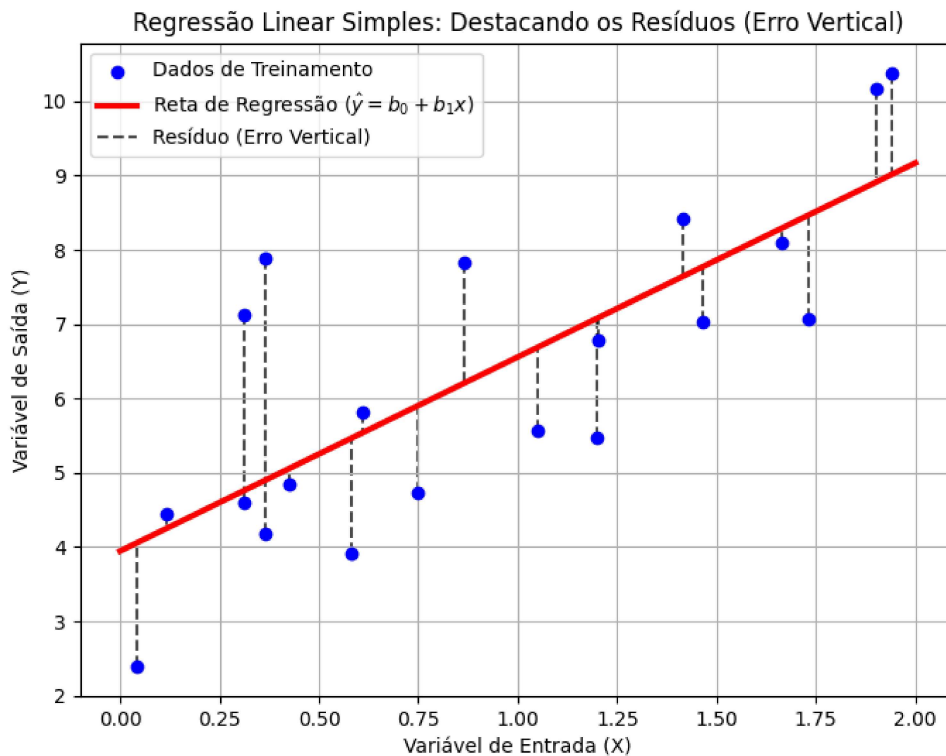


Figura 3 – Regressão Linear Simples: Visualização do ajuste da reta de regressão e destaque dos Resíduos (Erro Vertical).

Fonte: Elaborado pela autora.

### 3.4.2 Máquina de Vetores de Suporte para Regressão (SVR)

A SVR é uma adaptação do algoritmo de Máquina de Vetores de Suporte (SVM). O algoritmo de SVM foi originalmente proposto por Vapnik (1995) como um método de classificação, e a adaptação específica para problemas de regressão foi introduzida por Drucker et al. (1996).

A lógica do algoritmo de SVR, detalhada no tutorial de Smola e Schölkopf (2004), é diferente da Regressão Linear, que busca minimizar o erro para todos os pontos. Enquanto a Regressão Linear busca minimizar o erro quadrático entre os pontos e a linha de regressão, o SVR adota uma filosofia diferente. Seu objetivo é encontrar uma função  $f(X)$  que desvie dos valores reais  $y_i$  em no máximo  $\epsilon$  (épsilon) para a maioria dos dados de treinamento. Qualquer erro dentro desta margem  $\epsilon$  é considerado aceitável e não gera penalidade na função de custo.

Os principais componentes do SVR são descritos a seguir:

- **Função de Perda  $\epsilon$ -Insensível e a Banda de Erro**

A característica central do SVR é a função de perda  $\epsilon$ -insensível. Esta função define uma “banda de erro” simétrica de largura  $2\epsilon$  ao redor da função de regressão  $f(X)$ , como ilustrado na Figura 4.

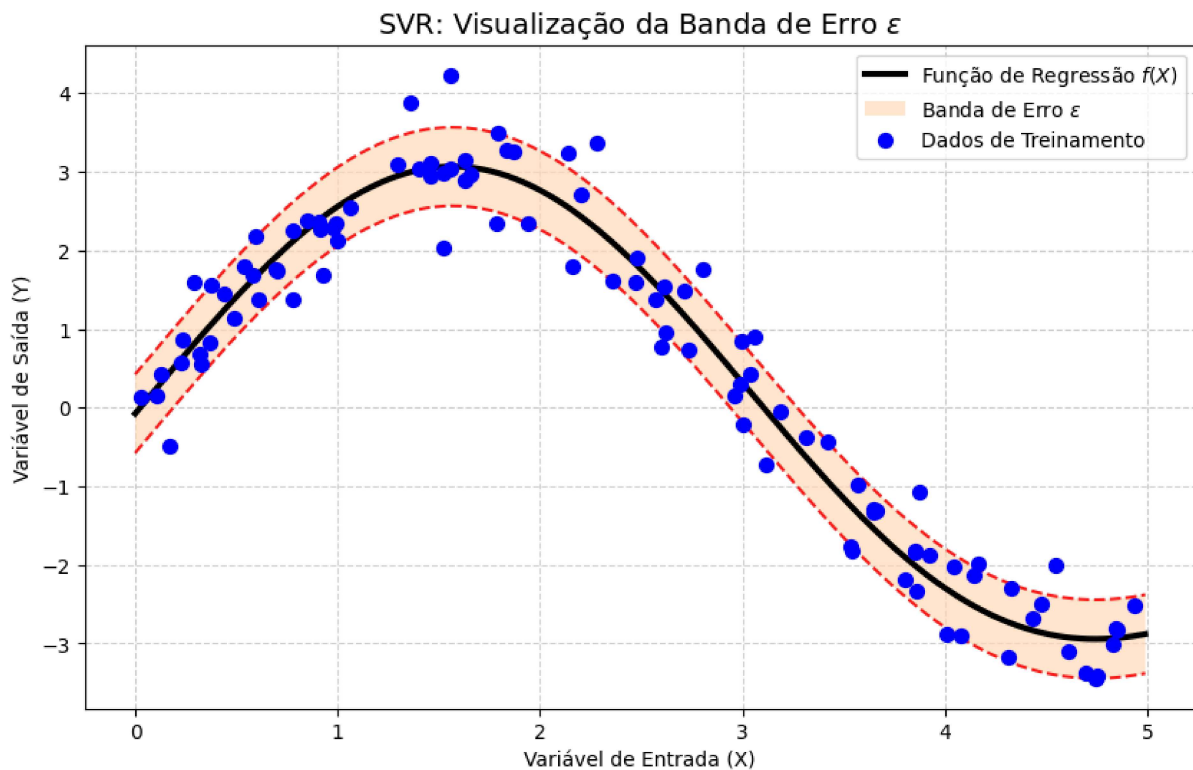


Figura 4 – Support Vector Regression (SVR): A Banda de Erro  $\epsilon$ . Erros que caem dentro da margem de  $\epsilon$  (região laranja) não são penalizados. Somente os pontos fora dessa banda contribuem para o custo do modelo, indicando a função de perda  $\epsilon$ -insensível.

Fonte: Elaborado pela autora.

Os pontos dentro da banda (laranja) não contribuem para a função de custo do modelo. O SVR está satisfeito com esses pontos. Entretanto, os pontos fora da banda (penalizados), contribuem para a função de custo, e o SVR tenta minimizar a soma desses desvios, ao mesmo tempo que mantém a suavidade da função  $f(X)$  (minimizando os coeficientes  $w$ ). Dessa forma, o objetivo do  $\epsilon$ -SVR é duplo:

1. Encontrar uma função  $f(x)$  que tenha, no máximo, um desvio  $\epsilon$  (epsilon) dos valores reais  $y_i$ .
2. Garantir que essa função  $f(x)$  seja o mais plana possível.

#### • Minimização e Variáveis de Folga

No caso de um modelo linear,  $f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b$ , o objetivo de encontrar uma função o mais suave ou plana possível é alcançado através da minimização do vetor de pesos  $\mathbf{w}$ . Essa minimização, que funciona como um termo de regularização, é formalizada pelo cálculo da norma Euclidiana quadrada do vetor de pesos, representada por  $\frac{1}{2}\|\mathbf{w}\|^2$  (SMOLA; SCHÖLKOPF, 2004). A minimização de  $\|\mathbf{w}\|^2$  é essencial para alcançar a suavidade desejada.

O problema inicial assume que é possível encontrar uma função  $f(x)$  que se ajuste a todos os dados de treino dentro dessa margem  $\epsilon$ . Contudo, em problemas reais, isso é frequentemente impossível (inviável).

Para permitir erros e tornar o modelo aplicável, são introduzidas variáveis de folga,  $\xi_i$  e  $\xi_i^*$ , análogas à “margem suave” da classificação SVM (SMOLA; SCHÖLKOPF, 2004). Essas variáveis medem o quanto cada ponto de dado  $i$  excede a margem de tolerância  $\epsilon$ .

O problema de otimização final torna-se, então, um *trade-off* controlado por dois termos:

$$\min_{w,b,\xi,\xi^*} \frac{1}{2}\|w\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \quad (3.4)$$

Sujeito a:

$$\begin{aligned} y_i - \langle w, x_i \rangle - b &\leq \epsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i &\leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* &\geq 0 \end{aligned}$$

## • O Hiperparâmetro $C$ e Regularização

Este problema de otimização busca o equilíbrio entre:

- Minimizar  $\frac{1}{2}\|w\|^2$ : Manter o modelo “plano” (simples, com baixa complexidade).
- Minimizar  $C \sum (\xi_i + \xi_i^*)$ : Penalizar os pontos de dados que caem fora da “rua”  $\epsilon$ -insensível.

O hiperparâmetro  $C > 0$  é o parâmetro de regularização que controla esse *trade-off*: um  $C$  alto penaliza duramente os erros (forçando um ajuste mais rigoroso), enquanto um  $C$  baixo prioriza a suavidade do modelo.

A importância da regularização é ilustrada na Figura 5, que compara o impacto do hiperparâmetro de penalidade  $C$  na suavidade da função de regressão  $f(X)$ . A Curva Verde (Alta Suavidade,  $C = 1$ ) resulta de um forte termo de regularização ( $\|w\|^2$  é minimizado com prioridade). Este modelo é mais suave e menos propenso a adaptar-se ao ruído, apresentando alto viés e baixa variância. A Curva Vermelha (Baixa Suavidade,  $C = 1000$ ), por outro lado, prioriza a minimização do erro de predição. Esta função é mais irregular e exibe maior suscetibilidade às flutuações dos pontos de treinamento. Ela demonstra baixo viés (ajuste próximo aos dados de treino) e alta variância (propenso ao overfitting), falhando na generalização da função. Portanto, o ajuste ideal do SVR requer o balanceamento de  $C$  para encontrar a função  $f(X)$  mais suave possível que ainda minimize os erros  $\epsilon$ -insensíveis.

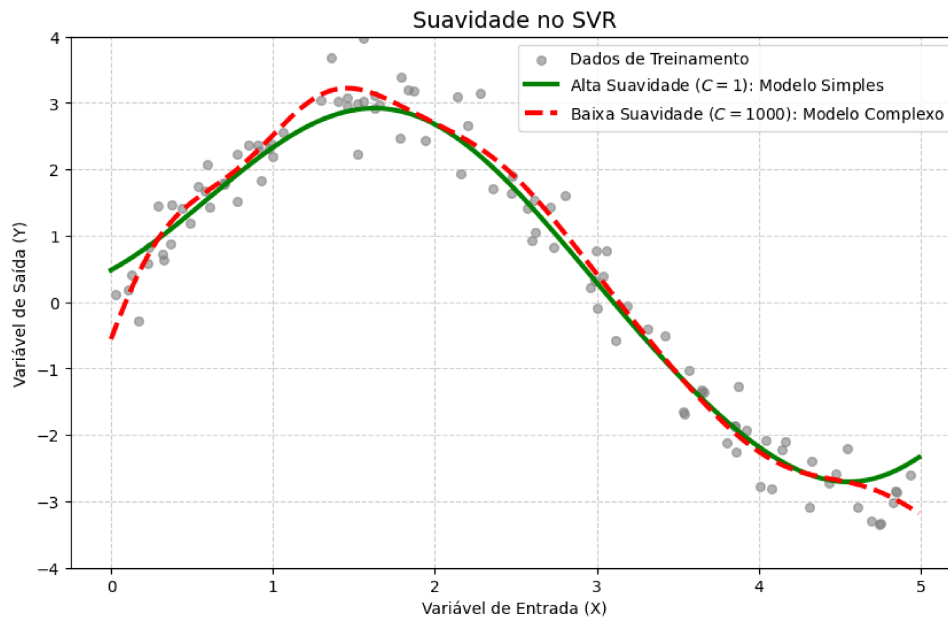


Figura 5 – Suavidade no SVR: Comparação do efeito do hiperparâmetro  $C$  na suavidade da função de regressão. A Curva Verde ( $C = 1$ ) demonstra alta suavidade e baixa variância, enquanto a Curva Vermelha ( $C = 1000$ ) demonstra baixa suavidade e alta variância (sobreajuste).

Fonte: Elaborado pela autora.

A característica central do SVR é a sua função de perda  $\epsilon$ -insensível. Ela estabelece que qualquer erro (resíduo) que seja menor ou igual a  $\epsilon$  é considerado zero e completamente ignorado pelo modelo. A penalidade só é aplicada aos erros que ultrapassam esse limiar  $\epsilon$ .

### • Vetores de Suporte

Um conceito crucial no SVR são os vetores de suporte. Conforme ilustrado na Figura 6, estes são os pontos de dados de treinamento que:

- Estão exatamente na borda da banda de erro  $\epsilon$ .
- Estão fora da banda de erro  $\epsilon$ .

Somente os vetores de suporte contribuem para a definição da função de regressão  $f(X)$ . Isso significa que o SVR é um modelo, onde a maioria dos pontos de treinamento (aqueles bem dentro da banda de erro) não afeta o resultado final. Essa característica torna o SVR robusto e eficiente, especialmente em conjuntos de dados grandes.

O SVR também pode utilizar funções de kernel (como RBF, polinomial, sigmoide) para mapear os dados para um espaço de maior dimensão, permitindo modelar relações não-lineares de forma eficaz.

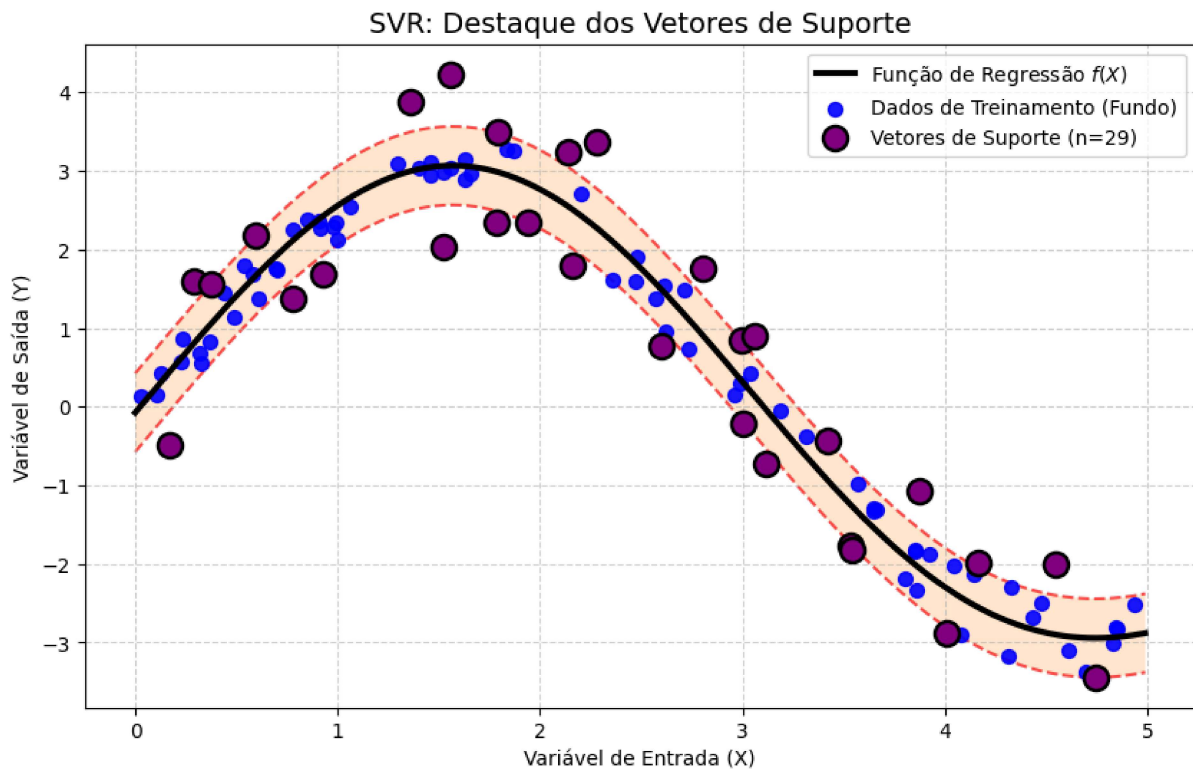


Figura 6 – Support Vector Regression (SVR): Os Vetores de Suporte. Os pontos de treinamento destacados com círculos pretos e um preenchimento roxo são os vetores de suporte, que estão na borda da banda de erro  $\epsilon$  ou a excedem. Estes pontos são fundamentais para definir o modelo de regressão.

Fonte: Elaborado pela autora.

### 3.4.3 K-Nearest Neighbors (KNN)

Um algoritmo não paramétrico é um tipo de algoritmo de AM que não faz suposições prévias sobre a forma da função que ele está tentando aprender. Um exemplo clássico desta abordagem é K-Nearest Neighbors (KNN), cuja introdução formal na literatura de classificação de padrões foi feita por Cover e Hart (1967).

A lógica do algoritmo é prever a resposta de um novo ponto identificando os  $k$  pontos de dados mais próximos a ele no conjunto de treinamento. Ele assume que observações similares devem ter respostas  $Y$  similares. Para prever a saída de um novo ponto de dados ( $x_0$ ), o algoritmo se baseia na “vizinhança” desse ponto no conjunto de treinamento. Conforme explicam James et al. (2013), a “proximidade” é comumente definida usando uma métrica de distância, como a distância Euclidiana (medida mais comum e intuitiva de distância entre dois pontos no espaço de *features*  $X$ ).

O processo para regressão, conforme descrito por James et al. (2013), é o seguinte:

1. Primeiro, identifica-se os  $k$  pontos de dados no conjunto de treinamento que estão mais “próximos” de  $x_0$  no espaço de entrada.

2. Em seguida, a previsão  $\hat{y}_0$  é estimada simplesmente calculando a média dos valores de resposta  $y_i$  desses  $k$  vizinhos.

Formalmente, o KNN para regressão estima a função  $\hat{f}(x_0)$  como:

$$\hat{f}(x_0) = \frac{1}{k} \sum_{x_i \in \mathcal{N}_0} y_i \quad (3.5)$$

Onde  $\mathcal{N}_0$  é o conjunto (a vizinhança) que contém os  $k$  pontos de treinamento  $x_i$  que são mais próximos de  $x_0$ .

A Figura 7 ilustra este processo para  $k = 7$ . A linha tracejada vertical vermelha indica a coordenada  $X$  do ponto avaliado ( $x_0$ ). Os vizinhos mais próximos (em verde) são selecionados com base na distância horizontal a  $x_0$ . A linha horizontal laranja representa a média dos valores  $Y$  desses  $k$  vizinhos, e o  $X$  vermelho é a previsão final  $\hat{y}$  neste ponto. As linhas pontilhadas verticais entre os vizinhos e a linha laranja demonstram graficamente o cálculo da média.

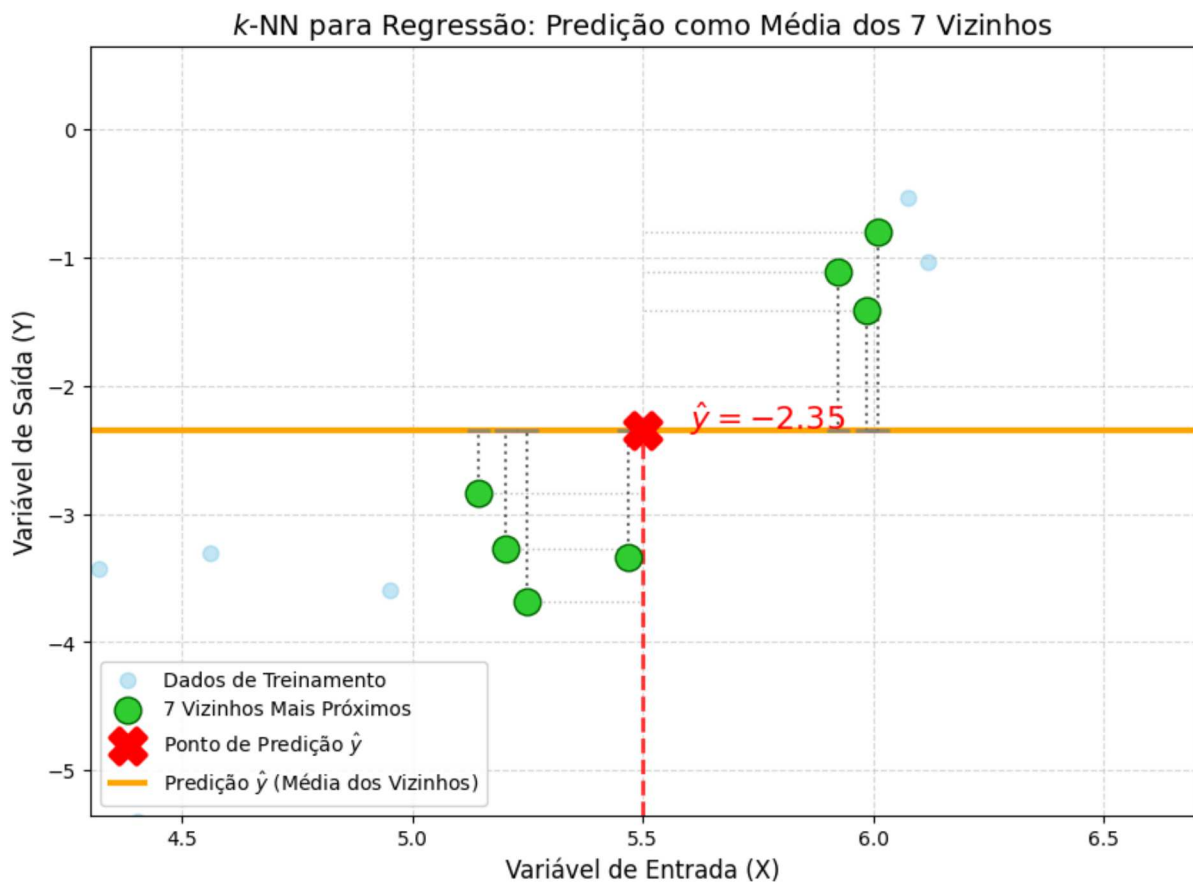


Figura 7 –  $k$ -NN para Regressão: Visualização do zoom do processo de predição para  $k = 7$ . A predição  $\hat{y}$  (linha laranja) é a média dos valores  $Y$  dos 7 vizinhos mais próximos (em verde), selecionados com base na distância Euclidiana (horizontal) ao ponto de teste  $X$ .

### O *Trade-off* de $k$

O número de vizinhos  $k$  é o hiperparâmetro central do KNN. A escolha de seu valor é crucial por definir a complexidade do modelo e, com isso, controlar o equilíbrio do *trade-off* de viés-variância (discutido em 3.4).

- Um valor de  $k$  pequeno (ex:  $k = 1$ ) resulta em um modelo de baixo viés (bias) mas alta variância. O modelo se torna muito flexível e “irregular”, pois a previsão depende de um único ponto. Isso leva ao overfitting.
- Um valor de  $k$  grande (ex:  $k = N$ ) resulta em um modelo de alto viés (bias) e baixa variância. A previsão para qualquer ponto  $\hat{f}(x_0)$  se aproxima da média de todos os dados de treinamento, resultando em um modelo excessivamente simples que falha em capturar a estrutura local (underfitting).

Em espaços com um grande número de features, a distância entre quaisquer dois pontos se torna muito grande, um fenômeno conhecido como “maldição da dimensionalidade”. O “vizinho mais próximo” pode, na prática, estar muito distante da observação  $x_0$ , tornando a previsão pouco confiável.

#### 3.4.4 Árvores de Decisão

A Árvore de Decisão é um algoritmo de aprendizado supervisionado não paramétrico e um dos métodos mais populares de inferência indutiva (MITCHELL, 1997). A função aprendida é representada por uma estrutura hierárquica, semelhante a um fluxograma, que utiliza regras simples baseadas em *features* para particionar recursivamente o espaço de dados em regiões (ALPAYDIN, 2010). As árvores podem ser expressas como conjuntos de regras “se-então”, melhorando a legibilidade e a transparência do modelo. Esse método constitui a base fundamental para algoritmos de *ensemble* de alto desempenho, como *Random Forest* e *Gradient Boosting*.

A construção baseia-se no princípio da Divisão e Conquista (QUINLAN, 1986), implementado através do Particionamento Binário Recursivo (JAMES et al., 2013). O processo inicia com todo o conjunto de dados na raiz e, a cada passo, divide-o em duas sub-regiões menores até que um critério de parada seja atingido.

#### Mecanismo de Particionamento

O componente fundamental da árvore é o nó de decisão. Matematicamente, uma divisão em um nó  $t$  é definida por uma regra binária que particiona o conjunto de dados  $S_t$  em dois subconjuntos disjuntos,  $S_t^E$  (esquerda) e  $S_t^D$  (direita). Essa divisão é caracterizada pela tupla  $(X_j, \tau)$ , onde:

- $X_j$  é a  $j$ -ésima variável preditora disponível (feature);
- $\tau$  é um valor limiar (*threshold*) no domínio de  $X_j$ .

Um dado  $x_i$  é direcionado para o nó filho à esquerda se  $x_{ij} \leq \tau$ , e para a direita caso contrário. O algoritmo opera de forma gananciosa (*greedy*), testando todas as combinações possíveis de variáveis e limiares para encontrar a divisão que maximiza a homogeneidade nos nós resultantes. Para problemas de **regressão**, o critério busca minimizar a RSS ou o Erro Quadrático Médio (MSE) nas duas regiões resultantes (BREIMAN et al., 1984).

O funcionamento pode ser visualizado na Figura 8, que ilustra um particionamento sequencial por Idade e Altura.

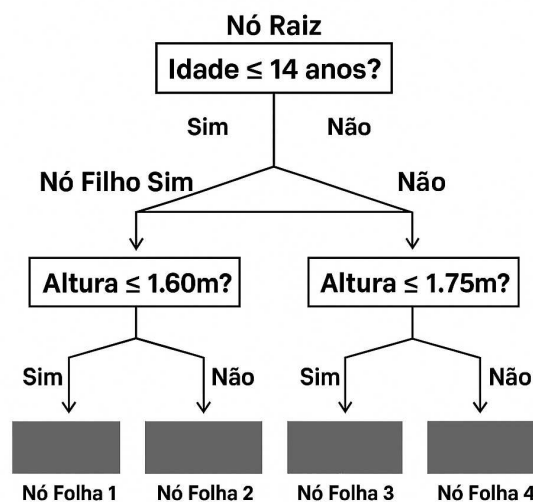


Figura 8 – Exemplo da estrutura de uma Árvore de Decisão Binária ilustrando o Particionamento Binário Recursivo.

Fonte: Elaborado pela autora.

## Regras de Decisão

A estrutura final pode ser representada por regras lógicas (Tabela 2). A previsão para um novo dado é a média dos valores de treinamento que caíram na mesma folha (região terminal).

## O Problema da Variância

Apesar de sua simplicidade, árvores de decisão individuais sofrem de alta variância. Elas tendem ao overfitting se crescerem profundamente, ajustando-se ao ruído dos

Tabela 2 – Regras de Decisão Derivadas da Árvore Exemplo

Regra (Caminho da Raiz à Folha)	Previsão ( $\hat{Y}$ )
SE Idade $\leq 14$ E Altura $\leq 1.60m$	Média da Folha 1
SE Idade $\leq 14$ E Altura $> 1.60m$	Média da Folha 2
SE Idade $> 14$ E Altura $\leq 1.75m$	Média da Folha 3
SE Idade $> 14$ E Altura $> 1.75m$	Média da Folha 4

Fonte: Elaborado pela autora.

dados de treinamento (ALPAYDIN, 2010). Essa instabilidade justifica o uso de métodos de *ensemble*.

### 3.4.5 Métodos de *Ensemble*

Métodos *Ensemble* combinam as previsões de múltiplos modelos individuais para formar um sistema de decisão unificado (POLIKAR, 2006). A premissa é que a combinação de diversos “especialistas” (modelos), ponderando suas opiniões, resulta em uma previsão final mais precisa, estável e robusta do que a de qualquer modelo isolado, explorando a diversidade dos erros individuais.

As duas principais estratégias de *textitensemble* são o *Bagging* (paralelo) e o *Boosting* (sequencial).

#### 1. Abordagem Paralela: Bagging e Random Forest

O Bagging (*Bootstrap Aggregating*) (BREIMAN, 1996) visa reduzir a variância criando várias versões de um mesmo modelo através de amostragem *bootstrap* (amostragem com reposição). Cada réplica do modelo é treinada em um conjunto de dados ligeiramente diferente, gerando diversidade (Figura 9). As previsões são agregadas por média (regressão) ou votação (classificação).

Uma evolução direta desse conceito é o Random Forest (BREIMAN, 2001). Embora utilize a estrutura de nós e divisões descrita anteriormente, este algoritmo introduz uma camada extra de aleatoriedade para garantir a decorrelação entre os modelos.

Diferente do *bagging* padrão, o Random Forest restringe a busca a um subconjunto aleatório de  $m$  variáveis ( $m \leq p$ ) em cada divisão. Essa estratégia combate o problema dos “preditores dominantes” e força as árvores a explorarem caminhos alternativos (Figura 10).

#### 2. Otimização via Gradient Descent

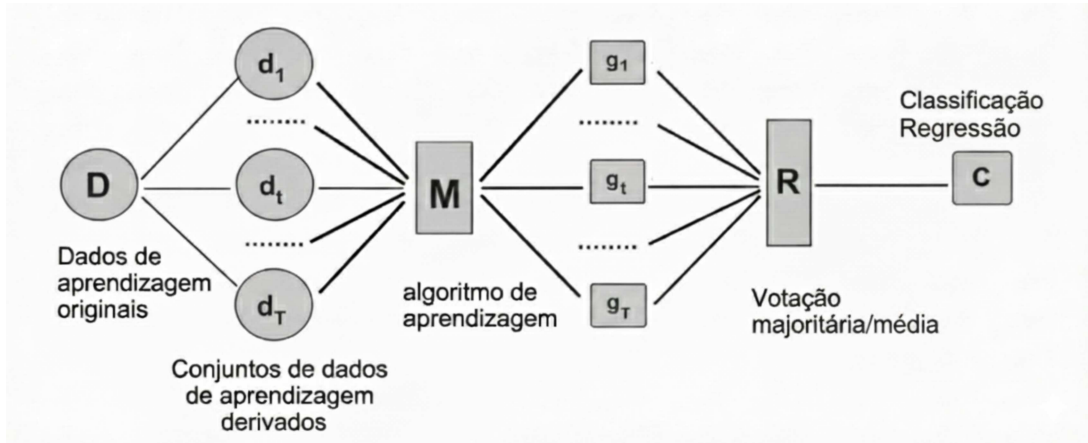


Figura 9 – Esquema de funcionamento do método Bagging.

Fonte: Adaptado de Ciaccio e Giorgi (2016).

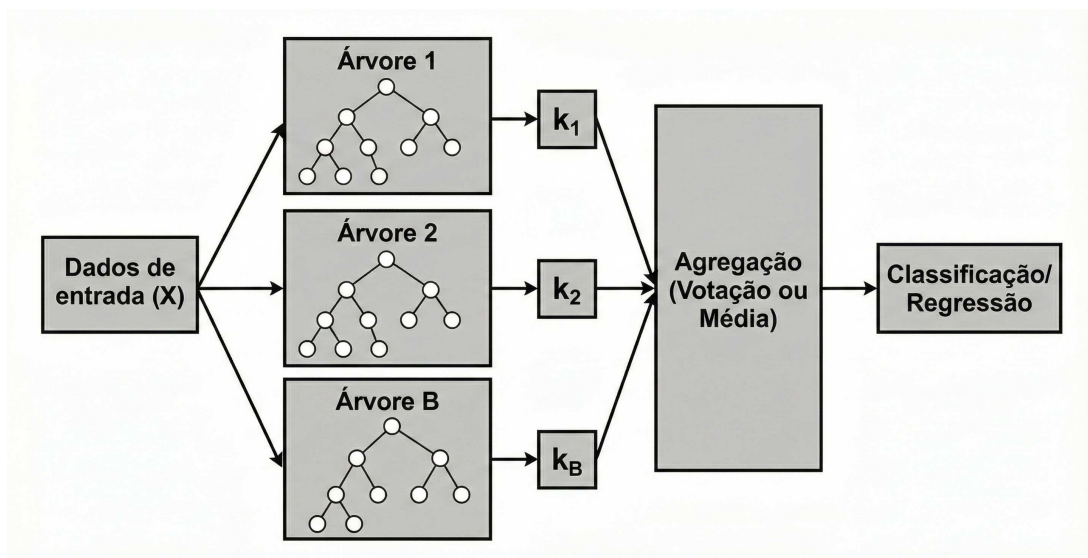


Figura 10 – Esquema do Random Forest: Múltiplas árvores treinadas em amostras bootstrap com seleção aleatória de features.

Fonte: Elaborado pela autora.

Para compreender a segunda família de *ensembles* (o Boosting), é necessário entender primeiramente o conceito de otimização iterativa. O Gradient Descent é utilizado para minimizar uma função de custo  $J(\theta)$  (GOODFELLOW et al., 2016).

A intuição é análoga a descer uma montanha no nevoeiro: sente-se a inclinação do terreno e dá-se um passo na direção da descida mais íngreme (o gradiente negativo). A atualização dos parâmetros  $\theta$  ocorre iterativamente:

$$\theta_{t+1} = \theta_t - \eta \cdot \nabla J(\theta_t) \quad (3.6)$$

Onde  $\eta$  (taxa de aprendizado) determina o tamanho do passo. Se  $\eta$  for muito grande, o modelo pode divergir; se muito pequeno, a convergência é lenta. O Boosting utiliza

exatamente esse princípio, mas ao invés de atualizar pesos de uma equação, ele atualiza o modelo adicionando novas árvores que corrigem o erro (gradiente) das anteriores.

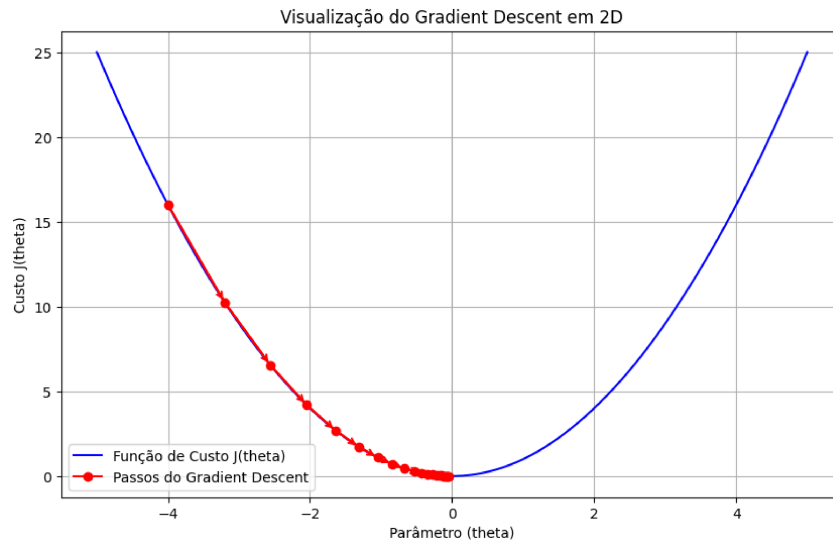


Figura 11 – Visualização geométrica do Gradient Descent.

Fonte: Elaborado pela autora.

### 3. Abordagem Sequencial: A Família Boosting

Diferente do Bagging, o *boosting* constrói modelos sequencialmente. Cada novo modelo é treinado para corrigir os erros (resíduos) cometidos pelo modelo anterior (FRIEDMAN, 2001). O modelo final é um “aprendiz forte”, composto pela soma ponderada de vários “aprendizes fracos”.

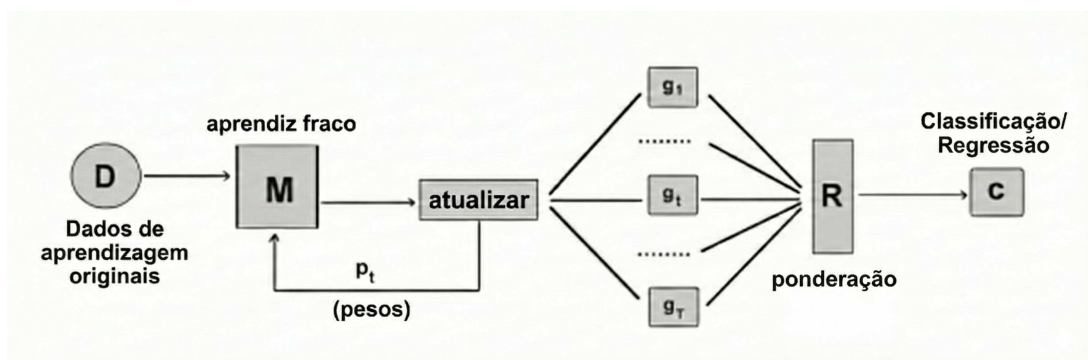


Figura 12 – Fluxograma representativo do processo de aprendizagem sequencial no Boosting.

Fonte: Adaptado de Ciaccio e Giorgi (2016).

Abaixo, detalha-se a evolução dos algoritmos desta família:

- **Gradient Boosting Machine (GBM)** O Gradient Boosting Machine (GBM), proposto originalmente por (FRIEDMAN, 2001), utiliza a técnica de *boosting*, construindo um *ensemble* de forma sequencial para reduzir o viés do modelo. A ideia central do algoritmo é que cada nova árvore fraca é treinada para aproximar os resíduos (erros) produzidos pelas árvores previamente ajustadas, em vez de modelar diretamente a variável resposta original (HASTIE et al., 2009). O modelo tenta aprender a direção do gradiente negativo da função de perda, comportando-se como o algoritmo de Gradient Descent descrito anteriormente.

O processo é iterativo e pode ser resumido da seguinte forma:

1. O modelo começa com uma predição constante (geralmente a média dos valores alvo);
2. Em cada etapa, calcula-se o pseudo-resíduo, que indica a direção na qual o modelo deve ser ajustado para reduzir o erro;
3. Uma nova árvore de decisão é treinada para prever esse pseudo-resíduo;
4. A predição dessa árvore é multiplicada por uma taxa de aprendizado (*learning rate*,  $\nu$ ) e somada ao modelo atual.

O procedimento formal é detalhado no Algoritmo 1.

---

**Algoritmo 1** Gradient Boosting para Regressão

---

- 1: **Inicializar** o modelo com um valor constante:

$$F_0(x) = \arg \min_{\gamma} \sum_{i=1}^N L(y_i, \gamma)$$

- 2: **para**  $m = 1$  até  $M$  (número de iterações) **faça**

- 3:     Calcular os *pseudo-resíduos* para cada observação  $i = 1, \dots, N$ :

$$r_{im} = - \left[ \frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)}$$

- 4:     Ajustar uma árvore de decisão  $h_m(x)$  aos resíduos  $r_{im}$ .

- 5:     Calcular o multiplicador  $\gamma_m$  resolvendo:

$$\gamma_m = \arg \min_{\gamma} \sum_{i=1}^N L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i))$$

- 6:     Atualizar o modelo com a taxa de aprendizado  $\nu$  ( $0 < \nu \leq 1$ ):

$$F_m(x) = F_{m-1}(x) + \nu \gamma_m h_m(x)$$

- 7: **fim para**

- 8: **Saída:** modelo final  $F_M(x)$ .

Fonte: Elaborado pela autora.

---

Embora a notação de derivada no passo 2 do algoritmo pareça complexa, ela representa uma generalização para qualquer função de custo. Para problemas de regressão comuns, onde se utiliza o Erro Quadrático Médio ( $L(y, F) = \frac{1}{2}(y - F)^2$ ), a derivada simplifica-se para a diferença direta entre o valor real e o predito:

$$r_{im} = y_i - F_{m-1}(x_i) \quad (3.7)$$

Portanto, intuitivamente, o GBM está simplesmente treinando novas árvores para corrigir o erro residual ( $y - \hat{y}$ ) deixado pelo modelo anterior.

Uma distinção crucial do GBM é o hiperparâmetro  $\nu$  (*learning rate*). Ele controla a contribuição de cada árvore para a soma final. Valores menores de  $\nu$  (ex: 0.01) exigem um número maior de árvores  $M$ , mas geralmente resultam em melhor generalização, evitando o *overfitting* (FRIEDMAN, 2002).

- **Extreme Gradient Boosting (XGBoost)** XGBoost, proposto por (CHEN; GUESTRIN, 2016), é uma evolução otimizada do algoritmo GBM, discutido anteriormente. O algoritmo mantém o princípio de criar árvores sequenciais para corrigir erros, mas implementa algumas mudanças. O XGBoost se destaca por ser mais rápido e, também, por possuir mecanismos internos que evitam que o modelo “decore” os dados de treinamento.

As principais mudanças no algoritmo são:

- \* **Controle de Complexidade (Regularização):** No GBM tradicional, as árvores podem crescer indefinidamente e se ajustar demais aos ruídos dos dados (*overfitting*). O XGBoost resolve isso adicionando um termo de “punição” ( $\Omega$ ) à sua equação.

Em outras palavras, o algoritmo penaliza árvores muito complexas. É como um professor que prefere uma explicação simples e direta a uma explicação longa e confusa, mesmo que ambas cheguem à resposta correta. A função a ser minimizada considera tanto o erro quanto a simplicidade do modelo:

$$\mathcal{L}^{(t)} = \underbrace{\sum_{i=1}^n L(y_i, \hat{y}_i)}_{\text{Erro da Predição}} + \underbrace{\Omega(f_t)}_{\text{Punição por Complexidade}} \quad (3.8)$$

- \* **Otimização Mais Inteligente (Segunda Ordem):** Para encontrar o melhor ajuste, os algoritmos “descem a montanha” do erro. Enquanto o GBM decide o passo olhando apenas para a inclinação do terreno (gradiente), o XGBoost olha também para a **curvatura** do terreno (o Hessiano).

Isso permite que o XGBoost perceba se o buraco está perto ou longe, ajustando o tamanho do passo de forma mais precisa. Matematicamente, isso é feito através de uma aproximação de segunda ordem, o que torna a convergência para o resultado ideal muito mais rápida e segura do que no método tradicional.

- **LightGBM:** Desenvolvido pela Microsoft, o Light Gradient Boosting Machine (LightGBM) (KE et al., 2017) surgiu com o objetivo principal de aumentar a eficiência computacional em grandes volumes de dados, onde implementações tradicionais como o XGBoost poderiam se tornar lentas. A grande diferença do LightGBM para os outros algoritmos de boosting está na estratégia de crescimento das árvores.

As duas principais características que diferenciam o LightGBM são:

- \* **Crescimento Vertical:** A maioria dos algoritmos tradicionais (como *Random Forest* e XGBoost) utiliza uma estratégia de crescimento horizontal, tentando manter a árvore equilibrada e expandindo todos os ramos simultaneamente. O LightGBM, por outro lado, adota uma estratégia vertical. Ele identifica a folha que possui o maior erro global e a divide, aprofundando a árvore apenas naquela direção específica.
- \* **Amostragem Inteligente (GOSS):** O algoritmo utiliza o Gradient-based One-Side Sampling (GOSS). Em termos intuitivos, o GOSS assume que dados com erros pequenos já foram “aprendidos” e os descarta aleatoriamente, focando o treinamento apenas nas instâncias com erros grandes.

A Figura 13 ilustra essa diferença estrutural, comparando o crescimento vertical do LightGBM (B) com o crescimento tradicional (A) e a abordagem simétrica do CatBoost (C), que será detalhada no item a seguir.

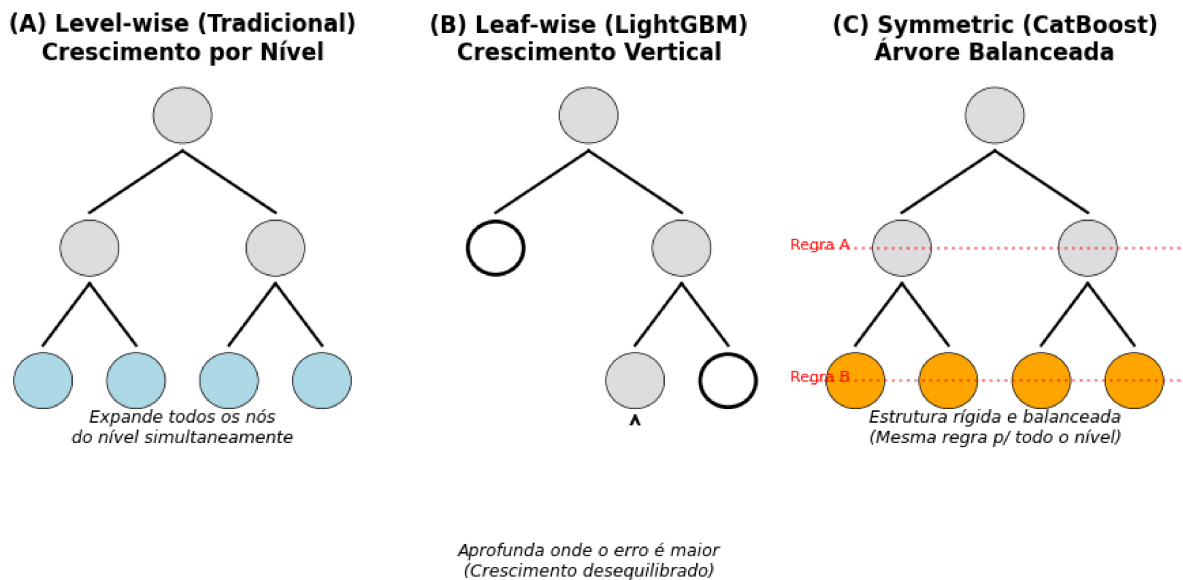


Figura 13 – Comparação das estratégias de crescimento: (A) *Level-wise* (tradicional); (B) *Leaf-wise* (LightGBM); (C) Árvore Simétrica (CatBoost).

Fonte: Elaborado pela Autora

- **CatBoost:** CatBoost, desenvolvido pela Yandex (DOROGUSH et al., 2018), destaca-se por resolver nativamente um dos desafios mais persistentes em modela-

gem de dados tabulares: o processamento eficiente de variáveis categóricas (como “Sexo”, “Modalidade” ou “Escola”) sem a necessidade de pré-processamento manual extenso.

Diferente de abordagens tradicionais que exigem a conversão de categorias em matrizes numéricas esparsas (como o *One-Hot Encoding*, que pode aumentar a dimensionalidade dos dados), o CatBoost implementa inovações algorítmicas:

- \* **Codificação Ordenada (*Ordered Boosting*):** O algoritmo converte categorias em valores numéricos utilizando estatísticas do próprio alvo.

No entanto, a codificação pelo alvo tradicional sofre de vazamento de dados, pois utiliza a informação da própria linha para prever ela mesma. O CatBoost resolve isso introduzindo uma permutação aleatória artificial nos dados durante o treinamento. Ele calcula a média da categoria considerando apenas as observações “passadas” na ordem dessa permutação. Isso impede que o modelo “veja o futuro” e reduz o overfitting causado pela predição enviesada.

- \* **Árvores Simétricas (*Oblivious Trees*):** Ao contrário do LightGBM e XGBoost, que constroem árvores com estruturas irregulares, o CatBoost impõe que toda a árvore seja perfeitamente balanceada (consulte a Figura 13-C).

Isso significa que a mesma regra de decisão (ex: *Idade > 15*) é aplicada a todos os nós de um mesmo nível de profundidade. Essa restrição estrutural atua como um regularizador natural, tornando o modelo menos sensível a ruídos nos dados. Além disso, essa simetria permite uma execução computacional mais eficiente: na hora da predição, o modelo funciona como uma tabela de indexação direta, tornando-o significativamente mais rápido.

### 3.5 Métricas de Avaliação de Desempenho

A avaliação de modelos preditivos é uma etapa crucial no AM. Para problemas de regressão, onde o objetivo é prever uma variável contínua, as métricas quantificam a distância entre os valores previstos pelo modelo ( $\hat{y}$ ) e os valores reais observados ( $y$ ). A escolha da métrica correta depende da sensibilidade desejada para o modelo e da interpretabilidade do erro (CHICCO et al., 2021).

A seguir, são apresentadas as métricas normalmente consideradas:

#### 3.5.1 Erro Médio Absoluto (MAE)

O Erro Médio Absoluto (MAE) calcula a média das diferenças absolutas entre as predições e os valores reais. Por utilizar o módulo, essa métrica trata erros positivos e negativos de forma igual e é menos sensível a valores discrepantes (*outliers*).

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3.9)$$

Onde  $n$  é o número de observações. O MAE fornece o erro médio na mesma unidade da variável resposta.

### 3.5.2 Raiz do Erro Quadrático Médio (RMSE)

A Raiz do Erro Quadrático Médio (RMSE) é uma das métricas mais utilizadas em regressão. O RMSE eleva os erros ao quadrado antes de tirar a média, o que penaliza grandes erros.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3.10)$$

Essa característica torna o RMSE preferível em situações onde cometer erros grandes é particularmente indesejável para o problema resolvido pelo modelo.

### 3.5.3 Coeficiente de Determinação ( $R^2$ )

O Coeficiente de Determinação ( $R^2$ , ou R-quadrado) difere das métricas anteriores por não medir o erro absoluto, mas sim a proporção da variância da variável resposta que é explicada pelo modelo. Ele compara o modelo treinado com um modelo base (que prevê sempre a média  $\bar{y}$ ).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\text{RSS}}{\text{TSS}} \quad (3.11)$$

Onde RSS é a Soma dos Quadrados dos Resíduos e TSS é a Soma Total dos Quadrados. Um  $R^2$  de 1.0 indica um ajuste perfeito, enquanto um valor de 0 indica que o modelo não é melhor do que simplesmente prever a média dos dados.

## 3.6 Cópulas

Enquanto os métodos de regressão tradicionais focam na modelagem da média condicional de uma variável resposta, a teoria das Cópulas oferece uma abordagem flexível para modelar a estrutura de dependência entre variáveis aleatórias, independentemente de suas distribuições marginais.

Introduzidas formalmente por (SKLAR, 1959), as cópulas permitem decompor uma distribuição de probabilidade conjunta multivariada em dois componentes separados: as distribuições marginais de cada variável (univariadas) e uma função de cópula que descreve a dependência entre elas.

O **Teorema de Sklar** é o fundamento desta teoria. Ele estabelece que para qualquer função de distribuição conjunta  $H(x_1, \dots, x_d)$  com marginais  $F_1(x_1), \dots, F_d(x_d)$ , existe uma função de cópula  $C$  tal que:

$$H(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)) \quad (3.12)$$

Essa separação é extremamente poderosa em modelagem estatística e inferência. Diferente da correlação de Pearson, que captura apenas dependências lineares e assume normalidade, as cópulas conseguem capturar dependências não lineares e assimetrias, como a dependência de cauda (a probabilidade de eventos extremos ocorrerem simultaneamente).

Existem diversas famílias de cópulas, sendo as mais comuns:

- **Cópula Gaussiana:** Baseada na distribuição normal multivariada, útil para dependências simétricas sem dependência de cauda forte.
- **Cópulas Arquimedianas:** Como as cópulas de *Clayton*, *Gumbel* e *Frank*, que possuem formas funcionais explícitas e são capazes de modelar dependências assimétricas (ex: maior correlação em eventos negativos extremos do que em positivos).

A inferência baseada em cópulas é frequentemente utilizada para gerar dados sintéticos realistas ou para realizar inferência de dados faltantes, preservando a estrutura de correlação original entre as variáveis.

### 3.7 Otimização de Hiperparâmetros

A eficácia da maioria dos algoritmos de AM depende intrinsecamente da escolha adequada de hiperparâmetros, os quais influenciam tanto o desempenho preditivo do modelo quanto o custo computacional (tempo de processamento e uso de memória). Existem duas abordagens principais para a definição desses parâmetros: a manual, que exige profundo conhecimento teórico e intuição do especialista, e a automática, que visa mitigar a necessidade de intervenção humana, embora possa elevar o custo computacional inicial (GOODFELLOW et al., 2016).

Dentre as principais estratégias de otimização automática, destacam-se:

**Busca em Grade (*Grid Search*):** Método exaustivo que avalia todas as combinações possíveis de hiperparâmetros dentro de um espaço de busca discretizado. Embora garanta encontrar o melhor valor dentro da grade definida, sofre da “maldição da dimensionalidade” — seu custo computacional cresce exponencialmente com o número de parâmetros, tornando-se inviável para problemas com muitos hiperparâmetros (BERGSTRA; BENGIO, 2012).

**Busca Aleatória (*Random Search*):** Abordagem que seleciona combinações aleatórias de hiperparâmetros segundo uma distribuição definida. Conforme demonstrado por Bergstra e Bengio (BERGSTRA; BENGIO, 2012), é estatisticamente mais eficiente que a busca em grade em espaços de alta dimensão, pois concentra os recursos computacionais nas regiões mais promissoras do espaço de busca.

**Otimização por Enxame de Partículas (PSO):** Metaheurística inspirada no comportamento social de bandos de aves ou cardumes de peixes (KENNEDY; EBERHART, 1995). No contexto de otimização de hiperparâmetros, uma população de “partículas” navega pelo espaço de busca contínuo, onde cada partícula representa uma configuração candidata. O movimento é guiado pela melhor posição visitada por cada partícula (memória individual) e pela melhor posição encontrada pelo enxame (conhecimento coletivo), equilibrando exploração de novas regiões e refinamento de soluções promissoras.

A escolha da estratégia adequada depende de fatores como a dimensionalidade do espaço de busca, o custo computacional de cada avaliação e os recursos disponíveis.

### 3.8 Validação de Modelos e Divisão de Dados

Em problemas de AM, a avaliação do desempenho de modelos preditivos é um aspecto importante para garantir que o modelo tenha boa capacidade de generalização. A generalização é a habilidade de um modelo em produzir boas previsões para dados que não foram utilizados durante o processo de treinamento (HASTIE et al., 2009; JAMES et al., 2013).

Uma das abordagens mais utilizadas para avaliar modelos de aprendizado de máquina consiste na divisão do conjunto de dados disponível em subconjuntos distintos, geralmente denominados conjunto de treinamento, conjunto de validação e conjunto de teste. O conjunto de treinamento é utilizado para ajustar os parâmetros do modelo, enquanto o conjunto de validação auxilia na escolha de hiperparâmetros e na comparação entre diferentes modelos. Já o conjunto de teste é reservado exclusivamente para a avaliação final do modelo. A aplicação do modelo nesse conjunto permite obter uma estimativa imparcial de seu desempenho, uma vez que os dados de teste não foram utilizados durante o processo de treinamento ou ajuste do modelo. (HASTIE et al., 2009)

Uma técnica amplamente empregada para validação de modelos é a validação cruzada (*cross-validation*). Nesse procedimento, o conjunto de dados é particionado em múltiplos subconjuntos de tamanho semelhante, chamados de *folds* (HASTIE et al., 2009). No método conhecido como validação cruzada *k-fold*, os dados são divididos em  $k$  partes. Em cada iteração,  $k - 1$  subconjuntos são utilizados para treinamento e o subconjunto restante é utilizado para validação. Esse processo é repetido  $k$  vezes, de modo

que cada subconjunto seja utilizado uma vez como conjunto de validação, procedimento que, conforme destacam Faceli et al. (2021), visa obter uma estimativa mais confiável do desempenho do modelo ao reduzir a variabilidade associada a uma única partição dos dados. O desempenho do modelo é então estimado pela média das métricas obtidas ao longo das iterações.

Além da validação cruzada, outras estratégias também podem ser empregadas para avaliação de modelos, como o método *hold-out*, em que os dados são divididos uma única vez em subconjuntos de treinamento e teste, e o método *bootstrap*, que utiliza reamostragem com reposição para estimar o desempenho do modelo (EFRON; TIBSHIRANI, 1994). A escolha da estratégia de validação depende de fatores como o tamanho do conjunto de dados, o custo computacional do treinamento e os objetivos da análise (KOHAVI, 1995).

### 3.9 Explicabilidade de Modelos: Importância de variáveis

Modelos de alto desempenho, como os baseados em árvores de decisão e *ensembles*, possuem uma lógica interna que leva a uma previsão que não é visualmente direta. Para amenizar esse problema e fornecer interpretabilidade, utilizam-se técnicas de importância de variáveis.

Uma das técnicas é o SHAP (LUNDBERG; LEE, 2017). O método SHAP baseia-se na Teoria dos Jogos Cooperativos, mais especificamente nos valores de Shapley (SHAPLEY, 1953). Ele calcula a contribuição de cada variável de entrada para o resultado final, tratando cada variável como um “jogador” e a previsão como o “prêmio”. A técnica distribui esse prêmio de forma justa, considerando a contribuição marginal de cada variável em todas as combinações possíveis com as demais.

#### 3.9.1 Importância Global via SHAP Médio

Diferente das métricas de importância nativas de algoritmos específicos, o SHAP é uma métrica agnóstica ao modelo (LUNDBERG; LEE, 2017). Isso permite que diferentes arquiteturas de algoritmos sejam comparadas sob uma mesma escala de importância, garantindo consistência na análise de quais variáveis são mais influentes para o fenômeno estudado.

Para obter uma visão geral do comportamento do modelo, utiliza-se o SHAP Médio. Enquanto os valores SHAP individuais explicam previsões locais, a média da magnitude dos valores SHAP em todo o conjunto de dados fornece um ranking de importância global, conforme a Equação 3.13:

$$I_j = \frac{1}{n} \sum_{i=1}^n |\phi_j^{(i)}| \quad (3.13)$$

Onde  $\phi_j^{(i)}$  representa o valor SHAP da variável  $j$  para a observação  $i$ , e  $n$  é o número total de observações no conjunto de dados.

### 3.9.2 Correlação de Pearson e Spearman

A correlação é uma medida estatística utilizada para avaliar o grau de associação entre duas variáveis. Existem diferentes coeficientes de correlação, cada um adequado a tipos específicos de dados e relações, diferenciando-se principalmente pela natureza da relação que capturam.

O **coeficiente de correlação de Pearson** ( $r$ ) mede a força e a direção da relação linear entre duas variáveis contínuas. É calculado pela covariância das variáveis dividida pelo produto de seus desvios padrão, conforme a Equação 3.14:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3.14)$$

O coeficiente de Pearson assume que os dados seguem distribuição aproximadamente normal e que a relação entre as variáveis é linear, sendo sensível à presença de *outliers* (PEARSON, 1895; MORETTIN; BUSSAB, 2017).

O **coeficiente de correlação de Spearman** ( $\rho$ ), também conhecido como correlação de postos, é uma medida não paramétrica baseada nos *ranks* (postos) das observações. Em vez de utilizar os valores brutos, Spearman ordena as observações e calcula a correlação de Pearson sobre os postos, conforme a Equação 3.15:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (3.15)$$

Onde  $d_i$  é a diferença entre os postos das duas variáveis para a observação  $i$ , e  $n$  é o número de observações. Spearman é indicado quando os dados são ordinais, não atendem ao pressuposto de normalidade, ou quando a relação entre variáveis não é linear, mas monotônica (SPEARMAN, 1904; HOWELL, 2010; FIELD et al., 2012).

## 4 MATERIAL E MÉTODOS

Este capítulo descreve o percurso experimental adotado para alcançar os objetivos deste estudo. A visão geral do método está sintetizada no fluxograma apresentado na Figura 14, o qual ilustra as principais etapas da pesquisa, desde o tratamento inicial dos dados até a avaliação final dos modelos preditivos. Cada bloco representado no fluxograma será detalhado nos tópicos seguintes deste capítulo, onde se discutirá a fundo a composição da base de dados, as técnicas de pré-processamento empregadas e as estratégias de AM selecionadas para os experimentos.

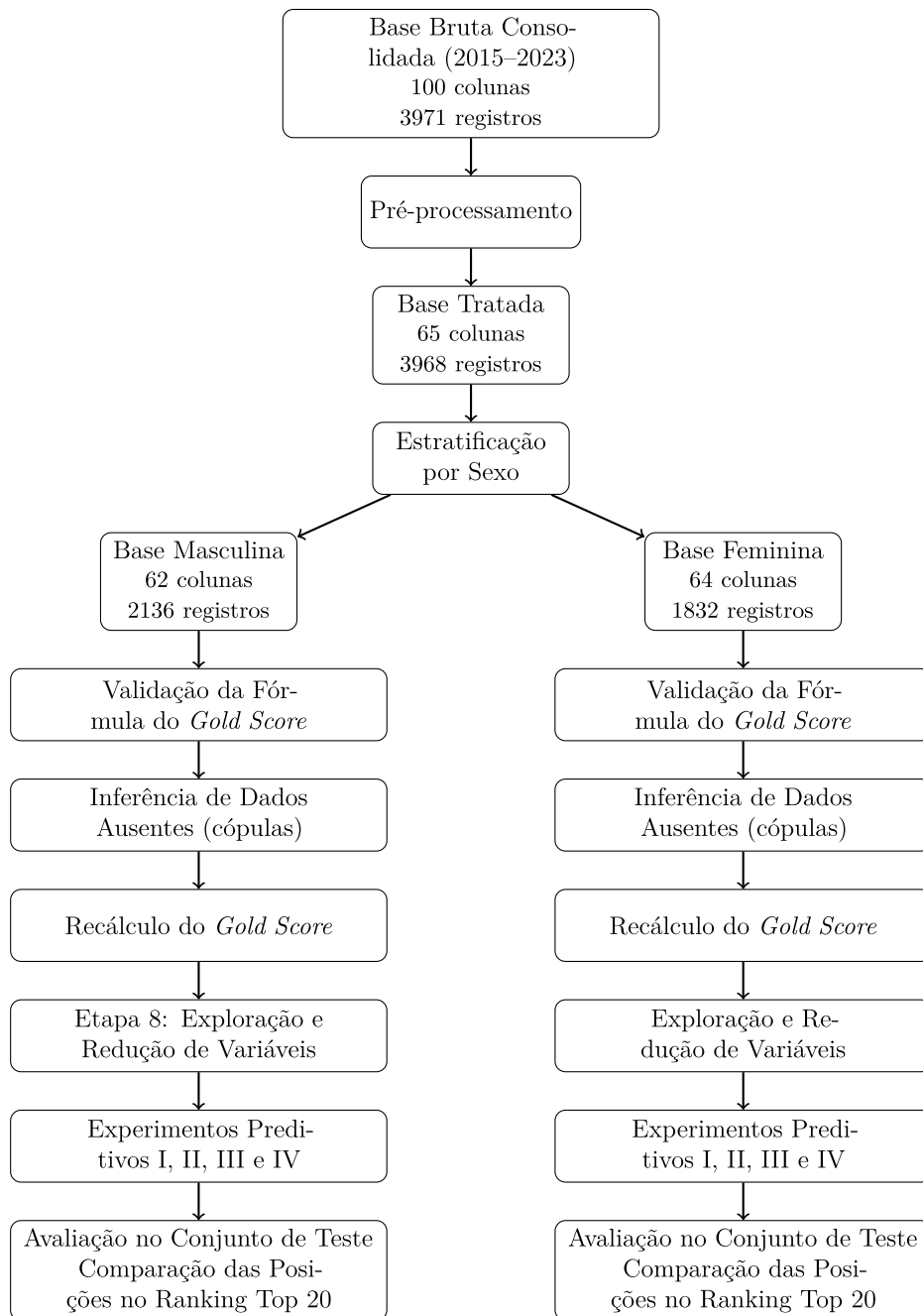


Figura 14 – Fluxograma geral do percurso experimental.

Fonte: Elaborado pela autora.

#### 4.1 Base de Dados

Foram utilizados dados oriundos de uma base constituída por registros de escolares participantes do Projeto Atletas de Ouro: Avaliação Multidimensional e Longitudinal do Potencial Esportivo de Jovens Atletas (WERNECK et al., 2022). A base utilizada é uma extensão daquela apresentada no estudo de validação do *Gold Score* (WERNECK et al., 2024), compreendendo agora um horizonte temporal mais amplo, com coletas realizadas entre 2015 e 2023.

É importante salientar que o fluxo de coletas não foi contínuo devido às restrições sanitárias impostas pela pandemia de COVID-19; conseqüentemente, não houve coleta de dados nos anos de 2020 e 2021, sendo as atividades retomadas em 2022.

Um aspecto metodológico crucial refere-se à definição da unidade amostral. Considerando a natureza do desenvolvimento infanto-juvenil, no qual características físicas e capacidades motoras sofrem alterações significativas ao longo do tempo, optou-se por tratar cada avaliação anual como uma observação independente. Dessa forma, embora um mesmo escolar possa ter participado do projeto em múltiplos anos (por exemplo, avaliado em 2016 e novamente em 2017), cada avaliação foi considerada uma nova entrada no banco de dados.

Essa decisão metodológica teve como objetivo aumentar o número de registros disponíveis para a análise, respeitando as mudanças biológicas próprias do desenvolvimento infanto-juvenil. Cada avaliação representa um momento específico do processo maturacional do aluno, no qual ocorrem variações relevantes nas características físicas e nas capacidades motoras. A literatura indica que essas mudanças não acontecem de forma linear ao longo do tempo, o que torna inadequada a suposição de que o desempenho permaneça estável entre diferentes idades cronológicas (MALINA et al., 2004; LLOYD; OLIVER, 2012).

Além disso, do ponto de vista da modelagem preditiva, a ampliação e diversificação do conjunto de dados contribuem para o treinamento de modelos com maior capacidade de generalização, especialmente em contextos nos quais o desempenho é fortemente dependente da idade e do estágio de desenvolvimento (HASTIE et al., 2009).

A utilização da base de dados para fins de pesquisa foi formalmente autorizada pelos pesquisadores responsáveis e está em total conformidade com os protocolos éticos, aprovados pelo Comitê de Ética em Pesquisa da Universidade Federal de Juiz de Fora (UFJF). Ressalta-se que todos os registros utilizados foram devidamente anonimizados.

A base original recebida é formada por aproximadamente 100 campos categóricos ou numéricos, com informações detalhadas relativas a 1866 escolares, do CMJF. O conjunto de dados abrange 3971 registros de testes, coletados ao longo do período de avaliação do projeto (2015-2023), e engloba:

- **Dados Cadastrais e Cronológicos:** Identificação, data de nascimento, sexo e

indicadores de idade relativa (`QuartilNascimento`).

- **Variáveis Antropométricas e Maturacionais:** Medidas corporais, composição corporal (% de Gordura), e indicadores de desenvolvimento biológico (Maturity Offset - `MO_Moore` e Estatura Adulta Prevista - `EAP`).
- **Variáveis de Aptidão Físico-Motora:** Resultados objetivos em testes de força (`Handgrip`, `Countermovement Jump (CMJ)`), velocidade (`Vel10m`, `Vel120m`), resistência (`Léger_Distância`) e arremesso.
- **Variáveis Psicológicas e Socioesportivas:** Indicadores de competência percebida, motivação, fatores ambientais, histórico de prática e o impacto de eventos competitivos (`Seleção Jogos da Amizade`).

A descrição detalhada de cada uma das variáveis contidas na base de dados original pode ser consultada no **Apendice A**, intitulado “Dicionário de Variáveis do Estudo”.

## 4.2 Pré-processamento e Tratamento da Base de Dados

O tratamento inicial dos dados foi crucial para garantir a validade estatística e a confiabilidade dos resultados. O fluxograma com as etapas de pré-processamento realizadas está apresentado na Figura 15, e cada uma delas será detalhada a seguir:

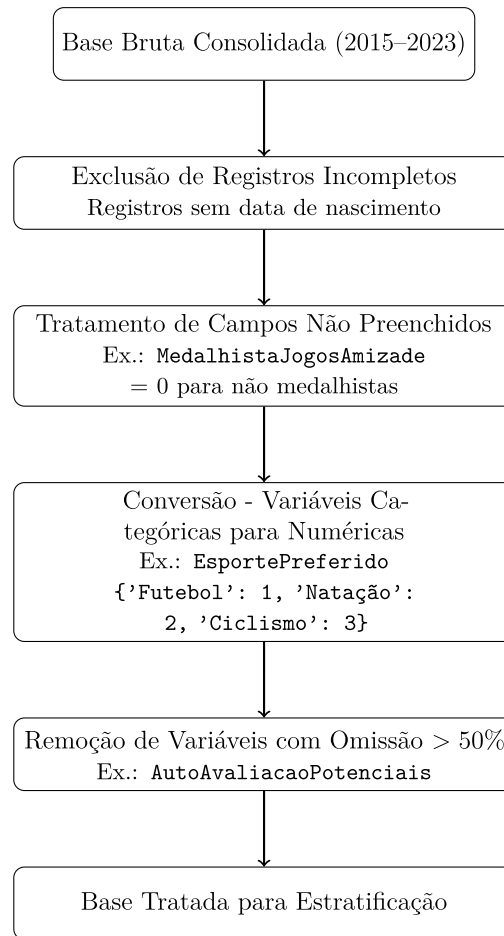


Figura 15 – Fluxograma do Processo de Pré-processamento.  
Fonte: Elaborado pela autora.

- **Exclusão de Registros Incompletos:** Registros (linhas) correspondentes a escolares sem informação na coluna `DataNascimento` foram removidos da base de dados. Esta variável é considerada essencial para o cálculo da idade e a realização das análises subsequentes.
- **Tratamento de Campos não Preenchidos:**
  - **Identificação de Registros:** Códigos de identificação foram gerados para três registros que estavam sem a devida identificação do escolar na base de dados.
  - **Medalhista:** Valores vazios na coluna `MedalhistaJogosAmizade` foram preenchidos com 0 (zero), assumindo-se que a ausência de registro indicava que o escolar não foi medalhista naquela competição.
  - **Maturação Biológica:** As colunas `Idade Menarca` e `Menarca` foram preenchidas com 0 (zero) para todos os escolares do sexo masculino, visto que são variáveis biologicamente exclusivas do sexo feminino.
- **Conversão de Variáveis Categóricas:** Colunas categóricas nominais, como `Esporte Preferido` e `Esporte Indicado`, foram submetidas à codificação para

serem representadas como variáveis numéricas, procedimento necessário para a aplicação de algoritmos de aprendizado de máquina e análises quantitativas.

- **Remoção de Redundância:** Colunas que continham valores de percentis (representações derivadas) foram removidas. A base de dados manteve-se apenas com os valores brutos, garantindo a primariedade e a precisão dos dados utilizados na análise.
- **Exclusão de Variáveis com Alta Omissão:** Foi elaborado um resumo da proporção de dados ausentes em cada coluna como exemplificado na Tabela 3. Colunas que apresentavam mais de 50% de dados faltantes foram removidas (*Idade\_Inicio*, *TempoPrática\_Idade* e *AutoAvaliacaoPotenciais*). Essa filtragem prévia foi necessária para garantir que apenas variáveis com densidade de informações suficiente fossem mantidas. A manutenção dessas colunas e a tentativa de inferir tantos valores ausentes poderiam produzir um viés e ruído excessivo nos modelos, comprometendo a confiabilidade dos resultados.

Optou-se por não separar as bases antes dessa filtragem porque a remoção por alta omissão, se feita separadamente, poderia manter uma variável em uma base e excluí-la em outra, criando assimetria entre os grupos. Ao aplicar o filtro de forma unificada, garantimos que as mesmas variáveis estejam disponíveis tanto na base masculina quanto na feminina. As variáveis que eventualmente apresentarem comportamento distinto por sexo — como *Menarca* — foram tratadas posteriormente.

Tabela 3 – Principais variáveis com dados ausentes

Variável	% Ausentes
<b>Variáveis excluídas (&gt;50%)</b>	
AutoAvaliacaoPotenciais	73.56
Idade_Inicio	52.90
TempoPrática_Idade	52.24
<b>Variáveis com omissão moderada (10-50%)</b>	
Vencedor	39.69
Determinado	39.69
Competitivo	36.77
[... mais 14 variáveis nesta faixa ...]	
<b>Demais variáveis</b>	
As demais variáveis apresentaram menos de 10% de dados ausentes, sendo a maioria com menos de 5%.	

Fonte: Elaborado pela autora.

#### 4.2.1 Estratificação por Sexo

Para otimizar as análises subsequentes e o desenvolvimento de modelos preditivos específicos, os dados foram separados por sexo. Nessa etapa, os registros inconsistentes com o perfil biológico de cada base foram filtrados e a coluna **sexo** foi removida, uma vez que se tornou uma variável constante e redundante após a segmentação. A base foi, então, exportada em dois arquivos distintos no formato **.csv** (masculino e feminino). A partir deste ponto, todas as etapas do trabalho foram conduzidas de forma independente e paralela para cada uma das duas bases de dados, visando modelos preditivos específicos e personalizados para cada sexo. A Tabela 4 descreve as variáveis presentes nas bases de dados por sexo.

Tabela 4 – Variáveis presentes nas bases de dados por sexo

<b>Variável</b>	<b>Masculino</b>	<b>Feminino</b>
Codigo	Sim	Sim
Ano_Avaliacao	Sim	Sim
DataNascimento	Sim	Sim
QuartilNascimento	Sim	Sim
SemestreNascimento	Sim	Sim
IdadeDecimal	Sim	Sim
FaixaEtaria	Sim	Sim
Repetencia	Sim	Sim
Dominancia_Maos	Sim	Sim
Dominancia_Pes	Sim	Sim
NivelAtividadeFisica	Sim	Sim
Treinamento	Sim	Sim
Esporte_Preferido	Sim	Sim
Participacao_Competicacao	Sim	Sim
Nivel_Competivo	Sim	Sim
QuerserAtleta	Sim	Sim
PossuiAtleta_Familia	Sim	Sim
ApoioFamilia	Sim	Sim
MassaCorporal	Sim	Sim
Estatura	Sim	Sim
IMC	Sim	Sim
AlturaSentado	Sim	Sim
Envergadura	Sim	Sim
ComprimentoMMII	Sim	Sim
SomaDC	Sim	Sim

*Continua na próxima página*

*Continuação da Tabela 4*

<b>Variável</b>	<b>Masculino</b>	<b>Feminino</b>
Gordura_%	Sim	Sim
PotencialTreinador	Sim	Sim
AltoPotencial_Treinador	Sim	Sim
Esporte_Indicado	Sim	Sim
Intangíveis	Sim	Sim
CompetenciaPercebida	Sim	Sim
Competitivo	Sim	Sim
Vencedor	Sim	Sim
Determinado	Sim	Sim
LidarAdversidades	Sim	Sim
DesempenhoPressao	Sim	Sim
MetasPreparoMental	Sim	Sim
Concentracao	Sim	Sim
LivrePreocupacao	Sim	Sim
ConfiancaMotivacao	Sim	Sim
Treinabilidade	Sim	Sim
Coping	Sim	Sim
Flexibilidade	Sim	Sim
Handgrip	Sim	Sim
CMJ	Sim	Sim
Arremesso	Sim	Sim
Vel10m	Sim	Sim
Vel20m	Sim	Sim
Leger_Distancia	Sim	Sim
EstaturaAdultaPrevistaEAP	Sim	Sim
EAP_atingida	Sim	Sim
EscoreZ_EAPatingida	Sim	Sim
Classif_StatusMaturacao	Sim	Sim
MO_Moore	Sim	Sim
IdadePVC_Moore	Sim	Sim
Menarca	–	Sim
Idade_Menarca	–	Sim
GoldScore_GS	Sim	Sim
SelecaoJogosdaAmizade	Sim	Sim
MedalhistaJogosAmizade	Sim	Sim
Nivel_VITORIA	Sim	Sim

*Continua na próxima página*

Continuação da Tabela 4

Variável	Masculino	Feminino
Codigo_Esporte_Preferido	Sim	Sim
Codigo_Esporte_Indicado	Sim	Sim

Fonte: Elaborado pela autora.

#### 4.2.2 Validação da fórmula do *Gold Score*

Após as etapas iniciais de limpeza, foi realizada a validação da principal métrica do estudo. O *Gold Score*, teve seus percentis e escores calculados. Este procedimento consistiu na validação da aplicação dos dados normativos da literatura, essencial para assegurar a precisão da métrica na amostra.

- **Recálculo:** Os percentis e o *Gold Score* foram calculados para toda a amostra, utilizando as fórmulas e os dados normativos (masculinos e femininos) descritos por (MIRANDA et al., 2024).
- **Validação da Fórmula:** A acurácia dos escores calculados foi verificada através da comparação com os escores originais como demonstrado na Tabela 5.

Tabela 5 – Validação da implementação do cálculo do *Gold Score*.

Estatística	Base Masculina	Base Feminina
Correlação de Pearson ( $r$ )	1,00	1,00
Média da Diferença	$\sim 0,00^*$	$\sim 0,00^*$
Desvio Padrão da Diferença	$\sim 0,00^*$	$\sim 0,00^*$

\*Nota: Os valores observados são desprezíveis em escala prática, indicando identidade numérica entre os escores original e recalculado.

Fonte: Elaborado pela autora.

#### 4.2.3 Tratamento de Valores Ausentes

Para garantir a robustez da análise e reduzir a perda de informações, os dados faltantes remanescentes foram tratados por meio de métodos de inferência estatística, com foco na preservação das relações multivariadas da base de dados. Todas as etapas de inferência foram conduzidas separadamente para as bases feminina e masculina.

- **Estratégia de Inferência:**

Inicialmente, foram consideradas abordagens baseadas na técnica de Min-Max, realizando o preenchimento por meio de uma amostragem uniforme entre os limites de cada variável. Contudo, a estratégia baseada em Cópulas apresentou desempenho

superior. Enquanto o método Min-Max tratava as variáveis de forma isolada, o uso de Cópulas permitiu modelar a estrutura de dependência entre elas. Assim, a inferência preservou as correlações originais, garantindo que os dados gerados fossem mais fiéis às correlações naturais entre os dados biológicos e de desempenho (SKLAR, 1959).

- **Procedimento:** As cópulas foram ajustadas separadamente para os conjuntos de dados feminino e masculino, respeitando as particularidades estatísticas de cada base. Os valores ausentes foram então inferidos a partir das distribuições condicionais estimadas para cada variável, garantindo a preservação das relações de dependência existentes entre os atributos.

O detalhamento das variáveis que passaram por esse processo, bem como o volume de dados recuperados por coluna, está apresentado nas Tabelas 6 e 7. Ao todo, a metodologia permitiu a inferência de 7.728 valores para a base masculina e 7.086 para a feminina. Essa abordagem foi fundamental para evitar o descarte de registros e manter a representatividade da amostra original.

#### 4.2.4 Recálculo do *Gold Score* após a inferência dos dados

Após a inferência dos valores ausentes, o *Gold Score* foi recalculado com base na base de dados estratificada completa. Esse recálculo foi realizado porque o índice depende da combinação de várias variáveis, e a presença de dados faltantes poderia afetar seu valor final. Ao incorporar os valores inferidos, garantiu-se que todas as informações disponíveis fossem consideradas de forma consistente, permitindo que o *Gold Score* representasse de maneira mais fiel o desempenho de cada indivíduo. Assim, o escore recalculado tornou-se mais adequado para as análises posteriores.

Para avaliar o impacto da inferência sobre a variável-alvo, comparou-se o novo escore com aquele obtido antes do procedimento (a partir dos dados observados). A diferença, também calculada como *Gold Score* (Pós-Inferência) - *Gold Score* (Pré-Inferência), reflete o ajuste introduzido pelo método.

Os resultados da Tabela 8 indicam que o método de inferência preservou a estrutura do *Gold Score*. As correlações ( $r = 1,00$ ) mostram que a ordem relativa dos indivíduos não foi alterada. As médias das diferenças positivas e baixas (0,31 e 0,71) sugerem um leve ajuste positivo nos escores, coerente com o fato de a inferência ter preenchido valores ausentes. Os desvios padrão das diferenças (0,92 e 1,65) representam a magnitude média desses ajustes individuais. Considerando que o desvio padrão amostral do *Gold Score* pré-inferência é de 18,64 pontos para a base masculina e 18,01 pontos para a feminina, os desvios padrão das diferenças (DPD = 0,92 e 1,65, respectivamente) representam apenas

Tabela 6 – Quantidade de registros inferidos por coluna (Base Masculina)

Variável	Valores Ausentes	
	n	%
Competitivo	887	41,5
Vencedor	887	41,5
Determinado	887	41,5
Vel10m	682	31,9
NívelAtividadeFísica	502	23,5
Flexibilidade	444	20,8
CMJ	234	10,9
Handgrip	226	10,6
AlturaSentado	220	10,3
ComprimentoMMII	220	10,3
EstaturaAdultaPrevistaEAP	234	10,9
EAP_atingida	234	10,9
EscoreZ_EAPatingida	234	10,9
Classif_StatusMaturação	234	10,9
Repetência	222	10,4
Leger_Distância	180	8,4
LivrePreocupação	120	5,6
Nível_Competitivo	119	5,6
DesempenhoPressão	106	5,0
Treinabilidade	106	5,0
Coping	105	4,9
Participação_Competição	100	4,7
ApoioFamilia	100	4,7
Concentração	99	4,6
MetasPreparoMental	97	4,5
ConfiançaMotivação	94	4,4
QuerserAtleta	93	4,4
LidarAdversidades	92	4,3
Envergadura	88	4,1
PossuiAtleta_Familia	80	3,7
CompetênciaPercebida	76	3,6
Dominância_Pés	72	3,4
Dominância_Mãos	63	2,9
Vel20m	53	2,5
Arremesso	20	0,9
SomaDC	14	0,7
Gordura_%	14	0,7
MassaCorporal	4	0,2
Estatura	4	0,2
IMC	4	0,2
PotencialTreinador	4	0,2
Intangíveis	1	0,05
AltoPotencial_Treinador	1	0,05
<b>TOTAL</b>	<b>7.728 registros inferidos</b>	

Fonte: Elaborado pela autora.

Tabela 7 – Quantidade de registros inferidos por coluna (Base Feminina)

Variável	Valores Ausentes	
	n	%
Vencedor	688	37,6
Determinado	688	37,6
Competitivo	572	31,2
Vel10m	614	33,5
Idade_Menarca	451	24,6
NívelAtividadeFísica	441	24,1
Flexibilidade	391	21,3
Leger_Distância	246	13,4
AlturaSentado	213	11,6
ComprimentoMMII	213	11,6
Handgrip	217	11,8
CMJ	214	11,7
Repetência	201	11,0
EAP_atingida	166	9,1
EscoreZ_EAPatingida	166	9,1
Classif_StatusMaturação	166	9,1
EstaturaAdultaPrevistaEAP	166	9,1
Menarca	134	7,3
ApoioFamilia	128	7,0
Nível_Competitivo	110	6,0
Participação_Competição	97	5,3
CompetênciaPercebida	81	4,4
Treinabilidade	79	4,3
QuerserAtleta	78	4,3
Vel20m	83	4,5
Dominância_Pés	72	3,9
Envergadura	72	3,9
LivrePreocupação	72	3,9
MetasPreparoMental	71	3,9
DesempenhoPressão	69	3,8
ConfiançaMotivação	68	3,7
Coping	68	3,7
LidarAdversidades	63	3,4
Dominância_Mãos	49	2,7
Arremesso	29	1,6
SomaDC	14	0,8
Gordura_%	14	0,8
IMC	13	0,7
Estatura	12	0,7
MassaCorporal	10	0,5
MO_Moore	6	0,3
IdadePVC_Moore	6	0,3
Intangíveis	3	0,2
PotencialTreinador	1	0,05
AltoPotencial_Treinador	1	0,05
<b>TOTAL</b>	<b>7.086 registros inferidos</b>	

Fonte: Elaborado pela autora.

4,94% e 5,11% da variabilidade total original do escore. Esta proporção – inferior a 5,2% em ambos os grupos – confirma que a inferência não introduziu ruído excessivo ou alterações substanciais na métrica-alvo. Dessa forma, mais de 94% da variabilidade original do *Gold Score* foi preservada, mantendo sua integridade para as análises preditivas.

Para fins de organização, a variável recalculada foi denominada *New\_GoldScore\_GS*, enquanto o valor obtido exclusivamente a partir dos dados observados permaneceu como *GoldScore\_GS*. O escore pós-inferência foi utilizado nas análises subsequentes.

Tabela 8 – Impacto da inferência por Cópulas no *Gold Score*.

<b>Estatística</b>	<b>Base Masculina</b>	<b>Base Feminina</b>
<b>Correlação de Pearson (<math>r</math>)</b>	1,00	1,00
<b>Diferença (média <math>\pm</math> DP)</b>	0,31 $\pm$ 0,92	0,71 $\pm$ 1,65

**Nota:** A diferença é calculada como  $Gold\ Score(\text{Pós-Inferência}) - Gold\ Score(\text{Pré-Inferência})$ . DP = Desvio Padrão.

Fonte: Elaborado pela autora.

#### 4.2.5 Exploração e Redução de Variáveis

O passo final de pré-processamento visou otimizar a dimensionalidade do modelo, removendo variáveis redundantes.

- **Seleção de Variáveis:** A seleção das variáveis foi conduzida com base nos seguintes critérios:
  - Padrão visual de associação com o *Gold Score* nas matrizes de correlação feminina e masculina, considerando a intensidade e a consistência das relações observadas;
  - Relevância teórica de cada variável dentro do modelo multidimensional de potencial esportivo adotado no estudo;
  - Redução de redundância entre preditores, por meio do controle da multicolinearidade.

Inicialmente, foram analisadas as matrizes de correlação de Pearson separadamente para a base feminina e para a base masculina (Figura 16 e Figura 17). A inspeção visual permitiu identificar a intensidade relativa das associações, a consistência dos padrões entre os sexos e a formação de blocos de variáveis fortemente correlacionadas. No mapa de calor, as correlações mais fortes foram evidenciadas pela coloração mais intensa, com coeficientes próximos a 1,0 (correlações positivas) ou -1,0 (correlações negativas). Como exemplo, observou-se uma correlação elevada entre as variáveis *Ve110m* e *Ve120m* (coeficiente próximo a 1,0), indicando forte associação entre medidas

de velocidade em distâncias curtas. Padrões semelhantes de coloração intensa foram identificados em outros blocos de variáveis, permitindo visualizar a estrutura de relações presente nos dados.

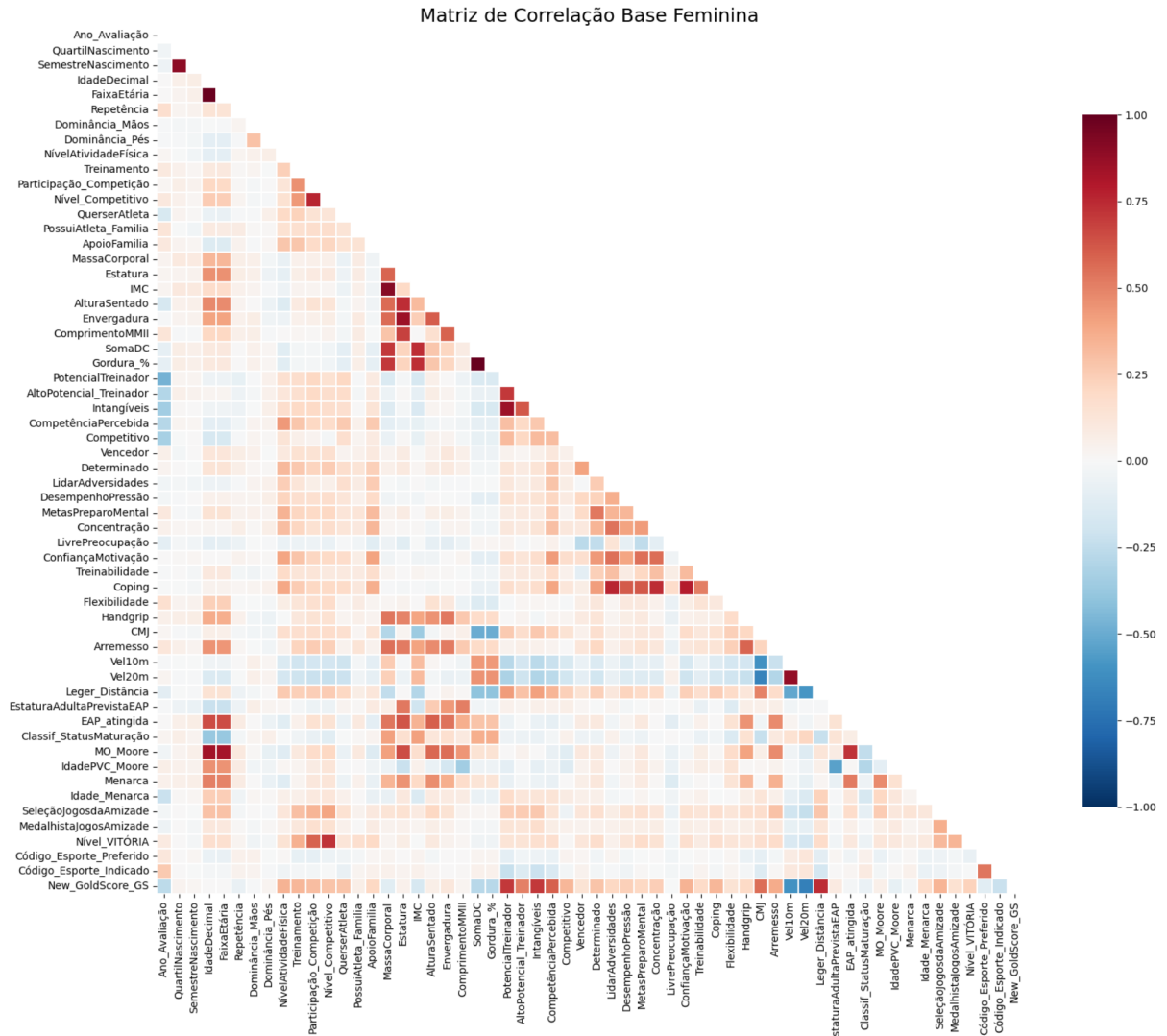


Figura 16 – Matriz de correlação de Pearson da base feminina.

Fonte: Elaborado pelo autor.

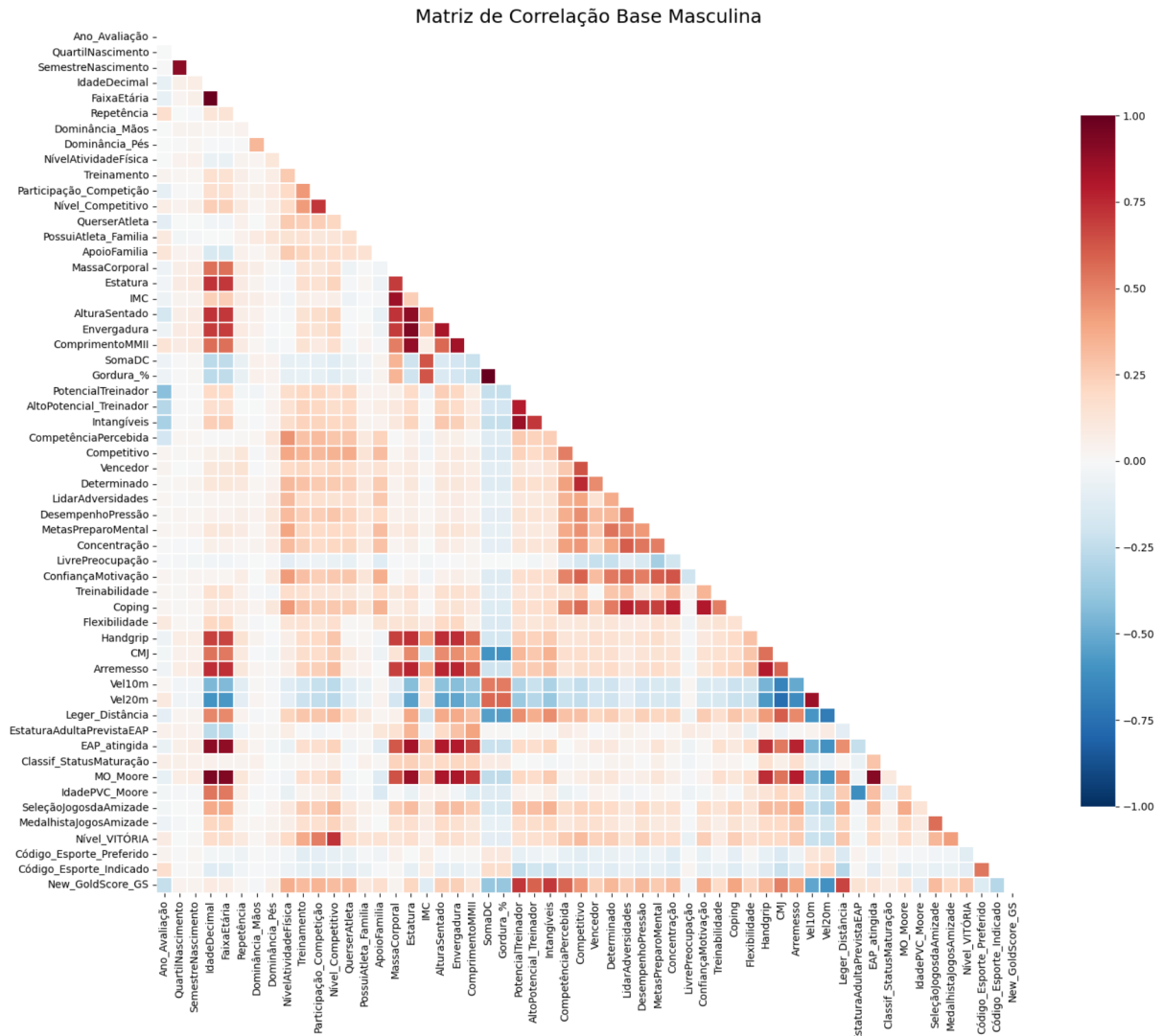


Figura 17 – Matriz de correlação de Pearson da base masculina.  
Fonte: Elaborado pelo autor.

A partir dessa análise, as seguintes variáveis foram mantidas no modelo:

**Indicadores avaliativos do treinador:** *PotencialTreinador*, *Intangíveis* e *AltoPotencial\_Treinador*. Essas variáveis apresentaram associação visualmente forte e consistente com o *Gold Score* nas duas bases, representando a dimensão observacional integrada do potencial esportivo.

**Capacidades físicas e motoras:** *Leger\_Distância*, *CMJ*, *Arremesso*, *Handgrip* e *Vel20m*. Essas variáveis demonstraram associação consistente com o desfecho nas matrizes e representam diferentes componentes do desempenho físico: capacidade aeróbia, potência de membros inferiores, potência de membros superiores, força isométrica e velocidade linear. A variável *Vel20m* foi escolhida por apresentar associação visualmente mais expressiva em comparação com a medida de 10 metros.

**Dimensão psicológica:** *CompetênciaPercebida*, *QuererSerAtleta* e *Coping*.

A variável *Coping* foi mantida por reunir, em uma única medida, diferentes características psicológicas importantes, como competitividade, determinação, concentração e desempenho sob pressão. Em vez de utilizar várias variáveis separadas e muito semelhantes entre si, optou-se por utilizar esse indicador mais geral, tornando o modelo mais simples e evitando redundâncias.

As variáveis *CompetênciaPercebida* e *QuererSerAtleta* foram mantidas por apresentarem relação consistente com o *Gold Score* e por representarem aspectos importantes do engajamento e da motivação esportiva.

**Experiência esportiva:** *Participação\_Competição* e *Nível\_Competitivo*. Ambas refletem exposição competitiva e trajetória esportiva, apresentando associação visual consistente nas duas matrizes.

As variáveis *SeleçãoJogosdaAmizade* e *MedalhistaJogosAmizade* foram deliberadamente removidas, pois compõem o critério de validação externa do *Gold Score*.

**Variáveis antropométricas e composição corporal:** *Estatura* e *Gordura\_%*.

A variável *Estatura* foi mantida por sua relevância estrutural no desempenho esportivo, especialmente em modalidades nas quais características morfológicas influenciam o rendimento.

A variável *Gordura\_%* foi priorizada em relação à *SomaDC* por representar uma medida derivada mais diretamente interpretável da composição corporal. Embora ambas estejam relacionadas, o percentual de gordura expressa de forma mais clara o componente adiposo do corpo, além de apresentar associação visual consistente nas duas bases analisadas.

**Bloco maturacional e etário:** *MO\_Moore* e *IdadeDecimal*.

A única variável diferente entre as bases feminina e masculina foi *Idade\_Menarca*, pois ela é um indicador de maturação que existe apenas no sexo feminino. Entre as variáveis específicas das meninas que foram analisadas, foi a única que permaneceu no modelo final por ser importante para representar o desenvolvimento biológico das atletas. Por esse motivo, ela foi incluída apenas no modelo feminino. No modelo masculino, permaneceram somente as variáveis de maturação que podem ser aplicadas aos dois sexos.

A variável *IdadeDecimal* foi mantida porque representa a idade de forma contínua, sem dividir em faixas, o que evita perda de informação. Já *MO\_Moore* foi mantida por resumir o estágio de maturação biológica, podendo ser utilizada tanto para a base feminina quanto para a base masculina.

**Variáveis contextuais:** As variáveis *PossuiAtleta\_Familia*, *Ano\_Avaliação* e *Código\_Esporte\_Indicado* foram incluídas por sua relevância estrutural e potencial

influência ambiental e institucional sobre o desenvolvimento do potencial esportivo, apresentando padrões de associação visualmente identificáveis nas matrizes.

A decisão final de seleção das variáveis não se baseou exclusivamente em critérios estatísticos. Além dos padrões observados nas análises quantitativas, consideraram-se e a relevância prática de cada variável para a composição do índice final.

### 4.3 Construção de Modelos Preditivos

A construção dos modelos preditivos teve como objetivo estimar o *Gold Score* com precisão e capacidade de generalização, considerando separadamente as bases masculina e feminina para respeitar a estratificação por sexo adotada no estudo.

Todos os procedimentos foram implementados em *Python*, utilizando **pandas** e **NumPy** para manipulação e organização eficiente dos dados. Para a construção e avaliação dos modelos de aprendizado de máquina foram empregadas bibliotecas consolidadas, como **Scikit-learn** (`sklearn`), **XGBoost** (`xgboost`), **LightGBM** (`lightgbm`) e **CatBoost** (`catboost`).

Foram avaliados modelos de diferentes famílias, incluindo:

- **Modelos Lineares:** *Linear Regression, Ridge, Lasso, ElasticNet.*
- **Modelos Baseados em Árvores:** *Decision Tree, Extra Tree, Random Forest, Gradient Boosting.*
- **Modelos de Boosting Avançado:** *XGBoost, LightGBM, CatBoost.*
- **Modelos Baseados em Distância e Margem:** *KNN, SVR.*

#### 4.3.1 Otimização de Hiperparâmetros

Para a otimização dos hiperparâmetros dos modelos de regressão, empregou-se o algoritmo *Particle Swarm Optimization* (Particle Swarm Optimization (PSO)) implementado através da biblioteca `pyswarms` (MIRANDA, 2018).

#### Configuração do Algoritmo

O PSO foi configurado com os seguintes parâmetros:

- **Número de partículas:** 15
- **Número de iterações:** 10
- **Total de avaliações por modelo:** 150 (15 partículas × 10 iterações)

- **Coefficiente cognitivo ( $c_1$ ):** 0,5
- **Coefficiente social ( $c_2$ ):** 0,3
- **Fator de inércia ( $w$ ):** 0,9

A escolha de 15 partículas e 10 iterações reflete um equilíbrio entre a qualidade da otimização e o custo computacional, considerando que cada avaliação de *fitness* demanda o treinamento e validação cruzada de 10-folds de um modelo de aprendizado de máquina. Para o conjunto completo de 13 algoritmos, o processo totalizou 19.500 treinamentos com validação cruzada (13 modelos  $\times$  150 avaliações  $\times$  10 folds).

Os valores dos coeficientes cognitivo e social foram ajustados para favorecer sutilmente a componente cognitiva em relação à social ( $c_1 > c_2$ ), incentivando as partículas a explorarem regiões próximas às suas melhores descobertas individuais. O fator de inércia  $w = 0,9$ , próximo ao limite superior da faixa recomendada por Shi e Eberhart (SHI; EBERHART, 1998), promove adequada exploração inicial do espaço de busca.

#### Função Objetivo

A função objetivo minimizada pelo PSO foi o negativo do coeficiente de determinação ( $R^2$ ) médio obtido via validação cruzada com 10-folds:

$$f(\mathbf{x}) = -\frac{1}{k} \sum_{j=1}^k R_j^2(\mathbf{x}) \quad (4.1)$$

onde  $k = 10$  é o número de *folds* e  $R_j^2(\mathbf{x})$  é o coeficiente de determinação obtido no  $j$ -ésimo *fold* para a configuração de hiperparâmetros representada pelo vetor  $\mathbf{x}$ . A escolha pelo negativo do  $R^2$  deve-se à convenção de algoritmos de otimização minimizarem funções (quanto menor, melhor).

Para configurações que geram erros durante o treinamento (como combinações inválidas de hiperparâmetros), implementou-se um mecanismo de tolerância a falhas que retorna um valor penalizado de  $10^{10}$ , garantindo que o algoritmo PSO possa continuar sua execução mesmo diante de configurações problemáticas.

#### Espaço de Busca dos Hiperparâmetros

Para cada algoritmo avaliado, definiu-se um espaço de busca específico, contemplando seus hiperparâmetros mais relevantes. A Tabela 9 resume os intervalos explorados para cada modelo, construídos com base na literatura especializada e em experimentos preliminares.

Tabela 9 – Espaço de busca de hiperparâmetros por algoritmo

Algoritmo	Hiperparâmetro	Lim. Inferior	Lim. Superior
<i>XGBoost</i>	<i>learning_rate</i>	0,01	0,2
	<i>n_estimators</i>	100	400
	<i>max_depth</i>	3	10
	<i>subsample</i>	0,6	0,9
	<i>colsample_bytree</i>	0,6	0,9
<i>LightGBM</i>	<i>learning_rate</i>	0,01	0,2
	<i>n_estimators</i>	100	400
	<i>num_leaves</i>	15	64
	<i>max_depth</i>	5	15
<i>CatBoost</i>	<i>n_estimators</i>	200	500
	<i>learning_rate</i>	0,01	0,2
	<i>depth</i>	4	10
	<i>l2_leaf_reg</i>	3	10
<i>Random Forest</i>	<i>n_estimators</i>	200	500
	<i>max_depth</i>	5	25
	<i>min_samples_split</i>	2	10
	<i>min_samples_leaf</i>	1	5
<i>Gradient Boosting</i>	<i>n_estimators</i>	100	400
	<i>learning_rate</i>	0,01	0,2
	<i>max_depth</i>	3	8
	<i>min_samples_split</i>	2	10
	<i>min_samples_leaf</i>	1	5
<i>SVR</i>	<i>C</i>	0,1	10,0
	$\epsilon$	0,01	0,5
	<i>kernel*</i>	linear, rbf	
<i>KNN</i>	<i>n_neighbors</i>	3	10
	<i>weights*</i>	uniform, distance	
<i>ElasticNet</i>	$\alpha$	0,001	1,0
	<i>l1_ratio</i>	0,1	0,9
<i>Lasso</i>	$\alpha$	0,001	1,0
<i>Ridge</i>	$\alpha$	0,1	10,0
<i>Decision Tree</i>	<i>max_depth</i>	5	25
	<i>min_samples_split</i>	2	10
	<i>min_samples_leaf</i>	1	5
<i>Extra Trees</i>	<i>n_estimators</i>	200	500
	<i>max_depth</i>	5	25

Nota: (\*) Hiperparâmetros categóricos foram codificados como variáveis contínuas com limiar de 0,5: valores menores que 0,5 são mapeados para a primeira opção e valores maiores ou iguais a 0,5 para a segunda.

Fonte: Elaborado pela autora.

### Tratamento de Casos Especiais

Para o modelo de *Regressão Linear*, que não possui hiperparâmetros a serem otimizados, realizou-se apenas a avaliação direta via validação cruzada, sem a etapa de otimização PSO.

### 4.3.2 Avaliação dos Modelos: Validação Cruzada e Divisão Treino-Teste

Para garantir a robustez e a capacidade de generalização dos modelos preditivos, adotaram-se estratégias consolidadas de validação e divisão dos dados, conforme descrito a seguir.

#### Divisão Treino-Teste

Antes de qualquer procedimento de otimização ou treinamento, a base de dados estratificada por sexo foi particionada em dois subconjuntos independentes:

- **Conjunto de Treinamento (80% dos dados):** Utilizado para todas as etapas de desenvolvimento do modelo, incluindo a otimização de hiperparâmetros via PSO e o treinamento final dos algoritmos. Este subconjunto é o único que tem contato com o processo de ajuste do modelo.
- **Conjunto de Teste (20% dos dados):** Reservado exclusivamente para a avaliação final do desempenho dos modelos já treinados e otimizados. Este conjunto permaneceu completamente isolado durante todas as fases de desenvolvimento, sendo utilizado apenas uma única vez ao final do processo para reportar as métricas de desempenho.

A divisão foi realizada de forma aleatória. Para isso, utilizou-se `shuffle=True` para embaralhar os dados, evitando viés da ordem original. O `random_state=42` funciona como uma “semente” que fixa a aleatoriedade, garantindo que a mesma divisão seja obtida sempre que o código for executado novamente.

#### Codificação de Variáveis Categóricas

Após a divisão dos conjuntos de treinamento e teste, procedeu-se à codificação das variáveis categóricas presentes na base de dados. Em particular, duas variáveis demandaram tratamento especial:

- **Código\_Esporte\_Indicado:** Esta variável representa a modalidade esportiva indicada pelo treinador para cada jovem, contendo múltiplas categorias (futebol, natação, atletismo, judô, etc.). Devido à natureza nominal dessas categorias, sem qualquer relação de ordem entre si, a codificação por mapeamento simples (ex: 1, 2, 3) não seria adequada, pois introduziria uma relação ordinal artificial inexistente.
- **Ano\_Avaliação:** Embora seja uma variável numérica, o ano de avaliação foi tratado como categórico nos experimentos. Esta decisão metodológica baseou-se na hipótese de que poderiam existir efeitos sistemáticos associados a cada ano de coleta (como diferenças na amostra, condições de teste, etc.).

Para ambas as variáveis, empregou-se a técnica de *one-hot encoding* (codificação *one-hot*), implementada através da função `pd.get_dummies` da biblioteca `pandas`. Este procedimento cria variáveis *dummy* binárias (0 ou 1) para cada categoria presente na variável original.

É importante destacar que o *one-hot encoding* foi aplicado separadamente nos conjuntos de treinamento e teste, seguindo as boas práticas de modelagem preditiva.

1. Ajustou-se o codificador exclusivamente no conjunto de treinamento, identificando todas as categorias presentes.
2. Aplicou-se a transformação no conjunto de treinamento, criando as colunas *dummy* correspondentes.
3. Aplicou-se a mesma transformação no conjunto de teste, garantindo que as mesmas categorias fossem codificadas.

#### Validação Cruzada de *k-Folds*

Durante a fase de otimização de hiperparâmetros (descrita na Seção 4.3.1), o conjunto de treinamento foi submetido à técnica de validação cruzada com *k-folds* (*k-fold cross-validation*).

O procedimento adotado foi o seguinte:

1. O conjunto de treinamento (80% dos dados) é particionado aleatoriamente em  $k = 10$  subconjuntos (*folds*) de tamanhos aproximadamente iguais.
2. O modelo é treinado  $k$  vezes, cada vez utilizando  $k-1$  *folds* para treinamento e 1 *fold* diferente para validação.
3. A métrica de desempenho ( $R^2$ ) é calculada para cada um dos  $k$  *folds*.
4. A métrica final reportada para uma dada configuração de hiperparâmetros é a média dos resultados obtidos nos  $k$  *folds*.

A escolha por  $k = 10$  fundamenta-se em estudos clássicos da área, que demonstram que este valor oferece um bom equilíbrio entre viés e variância na estimativa do erro de generalização (KOHAVI, 1995; HASTIE et al., 2009). Valores muito baixos (como  $k = 5$ ) podem produzir estimativas com maior viés, enquanto valores muito altos (como  $k = n$ , *leave-one-out*) aumentam a variância e o custo computacional.

## Função Objetivo na Validação Cruzada

Conforme detalhado na Seção 4.3.1 (Equação 4.1), a função objetivo minimizada pelo PSO foi o negativo do  $R^2$  médio obtido via validação cruzada de 10-folds. Cada configuração de hiperparâmetros avaliada corresponde, portanto, a 10 treinamentos e validações do modelo, garantindo que a escolha dos parâmetros seja baseada em uma estimativa robusta e não em uma única divisão eventualmente favorável dos dados.

## Avaliação Final no Conjunto de Teste

Após a conclusão da otimização e seleção dos melhores hiperparâmetros para cada algoritmo, os modelos foram treinados novamente com o conjunto de treinamento completo (100% dos 80%) e finalmente avaliados no conjunto de teste isolado. As métricas reportadas nos resultados (Capítulo 5) referem-se exclusivamente a esta avaliação final, garantindo que não houve qualquer tipo de “vazamento” de informação do conjunto de teste durante o desenvolvimento.

## Avaliação Complementar: Poder de Identificação do Top 20 no Conjunto de Teste

Como métrica complementar de desempenho, específica para o contexto de identificação de talentos esportivos, avaliou-se a capacidade dos modelos em recuperar os indivíduos com maior *Gold Score* dentro do conjunto de teste. Esta análise busca responder a uma questão prática fundamental: “entre os atletas do conjunto de teste, aqueles apontados pelo modelo como mais promissores realmente estão entre os melhores?”.

O procedimento adotado foi o seguinte:

1. No conjunto de teste (20% dos dados), estabeleceu-se o *ranking* dos atletas com base nos valores **reais** do *Gold Score*, identificando o grupo dos 20 com maior pontuação (*Top 20 real do teste*).
2. Utilizando as predições de cada modelo otimizado para o mesmo conjunto de teste, construiu-se o *ranking* dos atletas com base na pontuação predita, identificando também os 20 primeiros (*Top 20 predito do teste*).
3. Calculou-se a interseção entre os dois grupos, ou seja, o número de atletas que estavam presentes no Top 20 predito e também pertenciam ao Top 20 real, ambos restritos ao conjunto de teste.

Este procedimento resultou em uma métrica intuitiva e diretamente interpretável: a quantidade de atletas corretamente identificados entre os mais promissores no conjunto de teste, expressa como “X/20”. Por exemplo, um resultado de 15/20 indica que, dos

20 atletas que o modelo considerou mais promissores no conjunto de teste, 15 realmente estavam entre os 20 melhores segundo o *Gold Score* real nesse mesmo conjunto.

É importante destacar que esta métrica não exige que o modelo acerte exatamente a ordem do *ranking*, mas sim que os indivíduos por ele selecionados no conjunto de teste sejam, de fato, integrantes do grupo de elite dentro da amostra avaliada. Esta abordagem é particularmente relevante para programas de identificação de talentos, nos quais o objetivo prático é selecionar um grupo limitado de jovens promissores para receber acompanhamento especializado, mesmo que a ordenação exata dentro desse grupo não seja crítica.

A escolha pelo Top 20 justifica-se pelo contexto prático da pesquisa e pelo tamanho do conjunto de teste. Considerando que o conjunto de teste representa 20% da base original, o Top 20 corresponde aproximadamente aos 10% melhores atletas dessa amostra (variando conforme o tamanho exato de cada base). Em programas de desenvolvimento de talentos esportivos, os recursos para acompanhamento especializado são tipicamente limitados, restringindo o número de jovens que podem ser monitorados intensivamente. Assim, mais importante do que prever o escore exato de todos os atletas é a capacidade de identificar corretamente aqueles com maior potencial dentro de um grupo seletivo.

Os resultados desta análise são apresentados e discutidos no Capítulo 5, juntamente com as demais métricas de desempenho dos modelos.

#### 4.4 Delineamento dos Experimentos

Para avaliar a capacidade do Aprendizado de Máquina em estimar o potencial esportivo, o percurso experimental foi estruturado em etapas progressivas. Esta organização permite comparar como diferentes conjuntos de variáveis influenciam a precisão final das predições, culminando na abordagem que apresentou os melhores resultados.

##### 4.4.1 Experimento I: Modelo de Referência (Conjunto Integral)

O primeiro teste foi desenhado para estabelecer um padrão de comparação ideal. O objetivo foi verificar o nível de precisão que o algoritmo consegue atingir quando possui acesso a todas as informações disponíveis na base de dados, inclusive aquelas que compõem indiretamente a lógica do índice.

- **Configuração:** O modelo foi treinado utilizando o conjunto total de variáveis que ficaram após o pré-processamento, o que inclui tanto os dados físicos objetivos quanto as avaliações subjetivas dos treinadores.
- **Propósito:** Obter um resultado de previsão que sirva como parâmetro de excelência. Este teste responde à pergunta: “Quão perto da fórmula original o modelo consegue

chegar?”. O resultado de desempenho aqui alcançado serviu de base para medir o impacto das restrições aplicadas nos testes seguintes.

#### 4.4.2 Experimento II: Modelo sem as variáveis subjetivas do treinador

Este teste representa o núcleo da pesquisa, simulando um cenário onde a predição depende exclusivamente de critérios mensuráveis, sem o auxílio do olhar clínico especializado.

- **Configuração:** Foram removidas todas as variáveis subjetivas (`PotencialTreinador`, `AltoPotencialTreinador` e `Intangíveis`), mantendo apenas preditores físicos, antropométricos e contextuais.
- **Propósito:** Investigar a viabilidade de um modelo autônomo. Durante esta fase, as análises de importância revelaram que a variável `Ano_Avaliação` exercia uma influência nos modelos. Notou-se que o modelo utilizava o ano como um “atalho” para a predição, uma vez que a média de idade dos escolares oscilava entre os anos de coleta e que os modelos não deveriam depender do ano da avaliação.

#### 4.4.3 Experimento III: Remoção da variável `Ano_Avaliação`

O terceiro experimento foi delineado como uma resposta direta ao diagnóstico estatístico realizado na etapa anterior. Ao analisar a importância das variáveis no Experimento II, observou-se que o modelo utilizava o `Ano_Avaliação` como um preditor, o que motivou uma investigação sobre a natureza dessa variável.

- **Diagnóstico de Viés:** Realizou-se uma análise de variância (ANOVA) para verificar a homogeneidade do *GoldScore* ao longo dos anos de coleta. Os resultados, sintetizados na Tabela 10, indicaram diferenças estatisticamente significativas ( $p < 0,001$ ) para ambos os sexos. O teste *post-hoc* de Tukey HSD revelou que as maiores disparidades ocorreram entre 2015 e 2023, com uma diferença de médias de 15,73 pontos para o sexo masculino e 13,34 pontos para o feminino.

Tabela 10 – Resumo Estatístico do Diagnóstico de Viés (ANOVA e Tukey HSD).

Grupo	Estatística F	Valor- $p$	Par de Anos (Máx. Dif)	Diferença de Médias
Masculino	26,222	< 0,001	2015 – 2023	-15,73
Feminino	24,598	< 0,001	2015 – 2023	-13,34

Fonte: Elaborado pelo autor.

- **Análise da Distribuição Etária:** O cruzamento dos resultados estatísticos com a distribuição etária (Figuras 18 e 19) revelou que o algoritmo identificava um

desbalanceamento na amostra em vez de padrões de performance. Observou-se que em anos com médias de *Gold Score* elevadas, a idade média dos escolares também era superior. Como o índice é sensível à maturação física, o modelo utilizava o ano como um “atalho” para a predição.

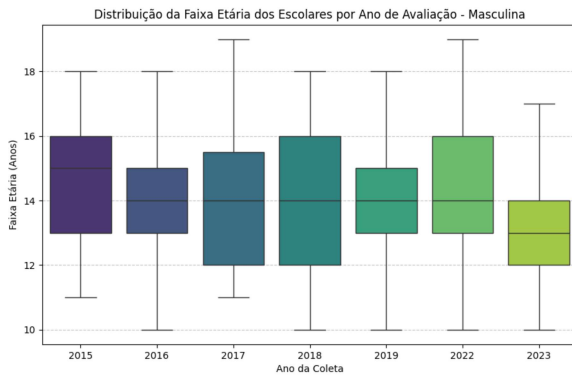


Figura 18 – Distribuição etária por ano (Masculina).

Fonte: Elaborado pelo autora.

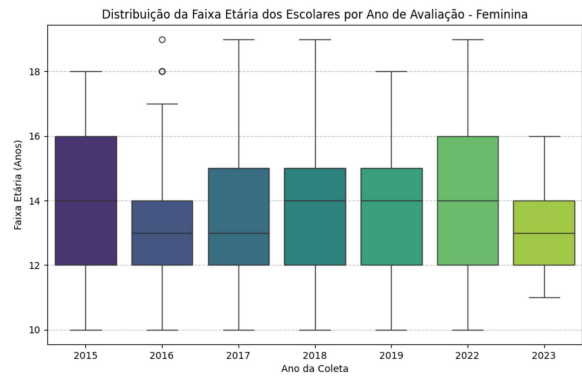


Figura 19 – Distribuição etária por ano (Feminina).

Fonte: Elaborado pelo autora.

- **Configuração:** Partiu-se da mesma base do Experimento II (sem variáveis subjetivas do treinador), com a exclusão adicional da variável `Ano_Avaliação`.
- **Propósito:** Avaliar a capacidade de generalização real do modelo. A intenção foi verificar se, ao remover o viés temporal, o modelo ainda seria capaz de sustentar um nível de precisão aceitável.
- **Resultado esperado:** Embora a exclusão pudesse reduzir o desempenho absoluto das métricas, o modelo ganharia em interpretabilidade e capacidade de generalização para novos dados.

#### 4.4.4 Experimento IV: Adição de Variáveis de Percentil

Após a realização do Experimento III, observou-se que variáveis relacionadas à idade cronológica continuavam exercendo influência relevante nas previsões do modelo. Esse comportamento é esperado em contextos de avaliação física de escolares, uma vez que diferenças no estágio maturacional e na idade cronológica impactam diretamente o desempenho em testes físicos e motores.

Além das diferenças maturacionais entre indivíduos da mesma idade, também devem ser consideradas as diferenças entre faixas etárias. Em geral, atletas mais velhos tendem a apresentar melhores resultados em testes físicos quando comparados a atletas mais jovens, mesmo quando apresentam níveis semelhantes de potencial esportivo. Dessa forma, parte do desempenho observado pode refletir principalmente o estágio de desenvolvimento físico associado à idade, e não necessariamente diferenças reais de capacidade esportiva.

Diante desse contexto, foi conduzido o Experimento IV com o objetivo de investigar uma abordagem capaz de reduzir a influência dessas diferenças etárias e maturacionais nas previsões do modelo.

#### Configuração do experimento

Foram adicionadas ao conjunto de variáveis utilizadas no treinamento novas variáveis derivadas baseadas em percentis, calculados exclusivamente a partir dos dados do conjunto de treinamento. Este cuidado foi tomado para evitar qualquer vazamento de informação do conjunto de teste. Os percentis foram calculados para os principais testes físicos utilizados no modelo: `Handgrip`, `CMJ`, `Arremesso`, `Vel20m` e `Leger_Distância`.

Para cada uma dessas variáveis, criou-se uma nova coluna representando o percentil do indivíduo em relação à distribuição observada no treinamento. Dessa forma, além dos valores absolutos dos testes físicos, o modelo passou a receber também informações **relativas** ao desempenho do atleta em comparação com os demais participantes da mesma base.

Esta estratégia permite contextualizar os resultados obtidos por cada indivíduo, reduzindo parcialmente o impacto direto das diferenças de idade e maturação nas previsões. Um atleta mais jovem que apresente um desempenho absoluto inferior, mas que esteja em um percentil elevado para sua faixa etária, terá essa informação capturada pela variável de percentil.

Após a inclusão dessas novas variáveis derivadas, os modelos foram novamente treinados utilizando o mesmo protocolo experimental adotado nos experimentos anteriores, incluindo:

- Divisão treino-teste (80/20) com `random_state=42`
- Otimização de hiperparâmetros via PSO com validação cruzada de 10-folds
- Avaliação final no conjunto de teste isolado

#### Síntese dos experimentos

A Tabela 11 resume as configurações de cada experimento realizado.

Tabela 11 – Síntese das configurações experimentais

<b>Experimento</b>	<b>Variáveis subjetivas</b>	<b>Ano_Avaliação</b>	<b>Percentis</b>
Experimento I	Incluídas	Incluído	Não
Experimento II	Removidas	Incluído	Não
Experimento III	Removidas	Removido	Não
Experimento IV (Final)	Removidas	Removido	Sim

Fonte: Elaborado pelo autor.

## 5 RESULTADOS E DISCUSSÃO

Este capítulo apresenta os resultados obtidos a partir do protocolo experimental descrito no Capítulo 4. De forma geral, os modelos tiveram seus hiperparâmetros otimizados por meio do algoritmo PSO, utilizando validação cruzada com 10 *folds* no conjunto de treinamento. Após a definição dos hiperparâmetros ótimos, os modelos foram treinados no conjunto de treinamento e avaliados no conjunto de teste, permitindo a análise de seu desempenho preditivo e de possíveis indícios de *overfitting*.

Os resultados são organizados em quatro experimentos, cada um correspondendo a um conjunto específico de variáveis preditoras, conforme delineamento apresentado no Capítulo 4. Para cada experimento, são apresentadas as métricas de desempenho dos modelos, com destaque para os dois melhores algoritmos, os gráficos de dispersão entre valores reais e preditos, e a análise da capacidade de ordenação dos escolares com melhor *Gold Score* por meio da comparação do Top 20 real versus previsto.

### 5.1 Experimento I: Modelo de Referência

O Experimento I constitui o modelo de referência do estudo, no qual os modelos foram treinados utilizando o conjunto de variáveis resultante do pré-processamento e seleção descritos no Capítulo 4. Este conjunto contempla todas as variáveis que compõem a fórmula original do *Gold Score*, incluindo as avaliações subjetivas do treinador (*PotencialTreinador*, *AltoPotencial\_Treinador* e *Intangíveis*) e as demais medidas físicas, motoras e contextuais que integram o cálculo do índice. A partir desta seção, todas as métricas, gráficos e análises de ranking apresentados referem-se ao desempenho dos modelos no conjunto de teste.

#### 5.1.1 Desempenho Preditivo dos Modelos

Os resultados do Experimento I indicam que, de forma geral, os algoritmos apresentaram elevada capacidade de predição do *Gold Score* quando todas as variáveis disponíveis foram utilizadas como entrada do modelo. As Tabelas 12 e 13 apresentam o ranking de desempenho dos modelos no conjunto de teste para os dados femininos e masculinos, respectivamente.

Observa-se que o modelo CatBoost apresentou o melhor desempenho em ambos os conjuntos de dados. Para os dados femininos, o modelo alcançou um coeficiente de determinação  $R^2$  de 0,9979, com MAE de 0,6089 e RMSE de 0,8002. Resultados semelhantes foram observados nos dados masculinos, nos quais o CatBoost atingiu  $R^2$  de 0,9959, MAE de 0,8809 e RMSE de 1,1661.

De modo geral, os modelos baseados em técnicas de boosting (CatBoost, Gradient Boosting, XGBoost e LightGBM) ocuparam as primeiras posições no ranking de

Tabela 12 – Ranking de performance dos modelos no conjunto de teste - Experimento I (Dados Femininos)

Modelo	$R^2$ Teste	MAE	RMSE
CatBoost	0,9979	0,6089	0,8002
XGBoost	0,9923	1,1667	1,5336
Gradient Boosting	0,9902	1,3517	1,7240
LightGBM	0,9887	1,4120	1,8505
Linear Regression	0,9681	2,4905	3,1109
SVR	0,9681	2,4830	3,1130
Ridge	0,9679	2,4973	3,1196
ElasticNet	0,9672	2,5158	3,1566
Lasso	0,9670	2,5215	3,1634
Extra Tree	0,9491	3,0376	3,9297
Random Forest	0,9472	3,1572	4,0050
Decision Tree	0,8498	5,4066	6,7539
KNN	0,7450	6,9235	8,7995

Fonte: Elaborado pela autora.

**Nota:** Para as métricas MAE e RMSE, valores menores indicam melhor desempenho preditivo ( $\downarrow$  erro). Para o  $R^2$ , valores maiores indicam melhor ajuste do modelo ( $\uparrow$  poder explicativo).

Tabela 13 – Ranking de performance dos modelos no conjunto de teste - Experimento I (Dados Masculinos)

Modelo	$R^2$ Teste	MAE	RMSE
CatBoost	0,9959	0,8809	1,1661
Gradient Boosting	0,9908	1,3590	1,7377
XGBoost	0,9892	1,5015	1,8867
LightGBM	0,9889	1,4885	1,9180
Lasso	0,9522	3,0710	3,9700
ElasticNet	0,9522	3,0794	3,9717
Ridge	0,9521	3,0641	3,9755
Linear Regression	0,9520	3,0606	3,9780
SVR	0,9514	3,0514	4,0026
Random Forest	0,9238	4,0686	5,0127
Extra Tree	0,9118	4,3944	5,3957
Decision Tree	0,7834	6,3289	8,4526
KNN	0,7449	7,2369	9,1739

Fonte: Elaborado pela autora.

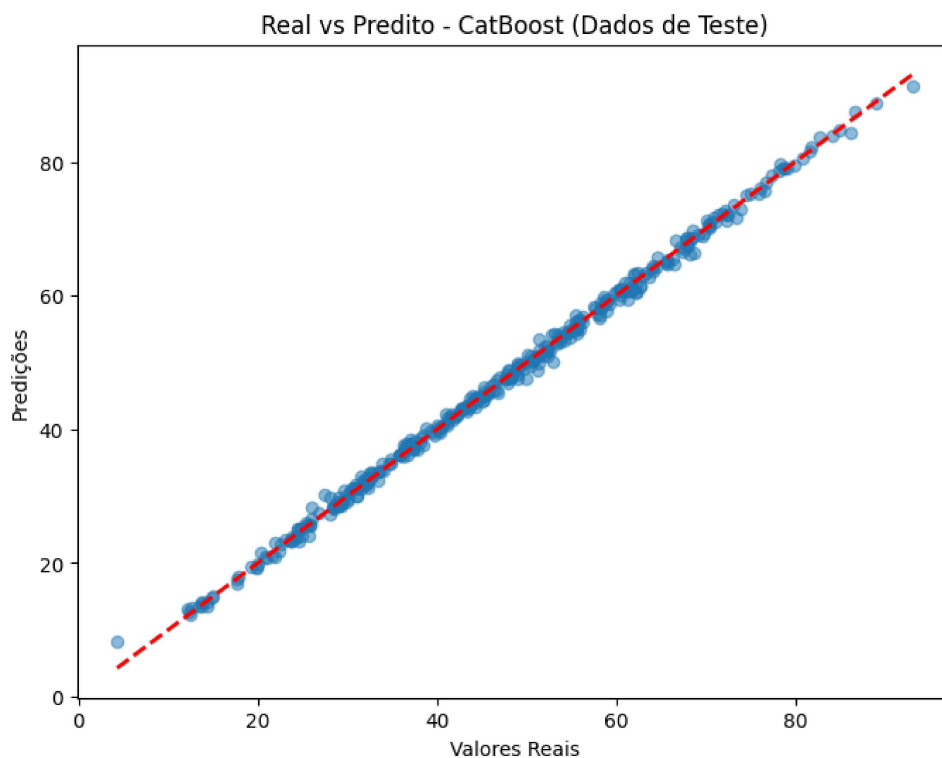
**Nota:** Para as métricas MAE e RMSE, valores menores indicam melhor desempenho preditivo ( $\downarrow$  erro). Para o  $R^2$ , valores maiores indicam melhor ajuste do modelo ( $\uparrow$  poder explicativo).

desempenho em ambos os conjuntos de dados. Esses algoritmos apresentaram valores elevados de  $R^2$  e menores erros de predição quando comparados aos demais métodos. Em contrapartida, modelos mais simples, como regressões lineares regularizadas (Ridge, Lasso e ElasticNet), apresentaram desempenho intermediário, enquanto algoritmos baseados em árvores isoladas ou métodos baseados em distância, como Decision Tree e KNN, apresentaram os piores resultados preditivos.

### 5.1.2 Análise Gráfica das Previsões

As Figuras 20 e 21 apresentam os gráficos de dispersão entre os valores reais e os valores preditos para os dois melhores modelos no conjunto feminino: CatBoost e XGBoost. Para o conjunto masculino, as Figuras 22 e 23 apresentam os resultados para CatBoost e Gradient Boosting, respectivamente.

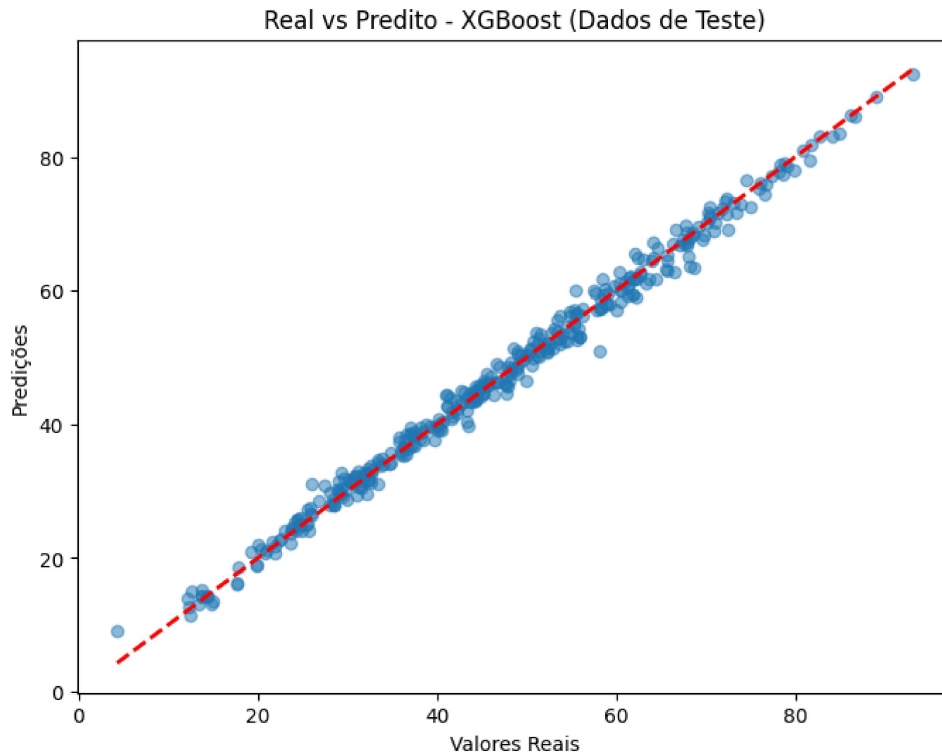
Figura 20 – Gráfico de dispersão entre valores reais e preditos pelo modelo CatBoost - Experimento I (Dados Femininos)



Fonte: Elaborado pela autora.

Em todos os casos, observa-se que os pontos se concentram fortemente próximos à linha de identidade, indicando concordância entre os valores observados e estimados. Esse resultado sugere que os modelos foram capazes de representar adequadamente a relação entre as variáveis e o *Gold Score*, sem indícios claros de subestimação ou superestimação ao longo da escala.

Figura 21 – Gráfico de dispersão entre valores reais e preditos pelo modelo XGBoost - Experimento I (Dados Femininos)



Fonte: Elaborado pela autora.

### 5.1.3 Análise de Importância das Variáveis

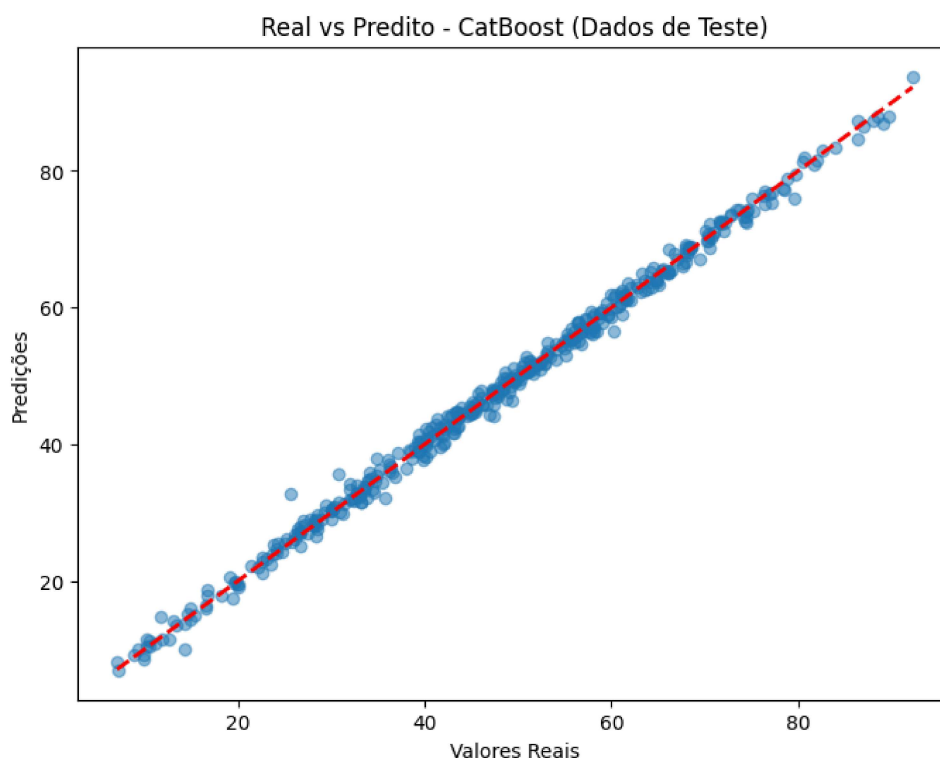
Para compreender melhor o comportamento dos modelos e identificar quais variáveis mais contribuíram para as previsões, foi realizada uma análise de importância das variáveis utilizando os valores SHAP para os dois melhores algoritmos em cada conjunto. As Figuras 24 e 25 apresentam as dez variáveis com maior impacto médio na predição do *Gold Score* para os dados femininos, segundo os modelos CatBoost e XGBoost. Para os dados masculinos, as Figuras 26 e 27 apresentam os resultados para CatBoost e Gradient Boosting.

Observa-se que as variáveis relacionadas à avaliação do treinador, como `Potencial Treinador` e `AltoPotencial_Treinador`, encontram-se entre os fatores de maior influência nos modelos para ambos os conjuntos de dados. Esse resultado indica que tais informações exercem papel relevante na determinação do *Gold Score*, o que era esperado, uma vez que `PotencialTreinador` integra a fórmula original do índice.

### 5.1.4 Capacidade de Ordenação dos Escolares com Melhor *Gold Score*

Para avaliar a capacidade dos modelos em preservar as posições mais altas do ranking, foi realizada uma comparação entre o ranking real e o ranking previsto para os 20 escolares com maiores valores de *Gold Score* no conjunto de teste. As Tabelas 14 e 16

Figura 22 – Gráfico de dispersão entre valores reais e preditos pelo modelo CatBoost - Experimento I (Dados Masculinos)



Fonte: Elaborado pela autora.

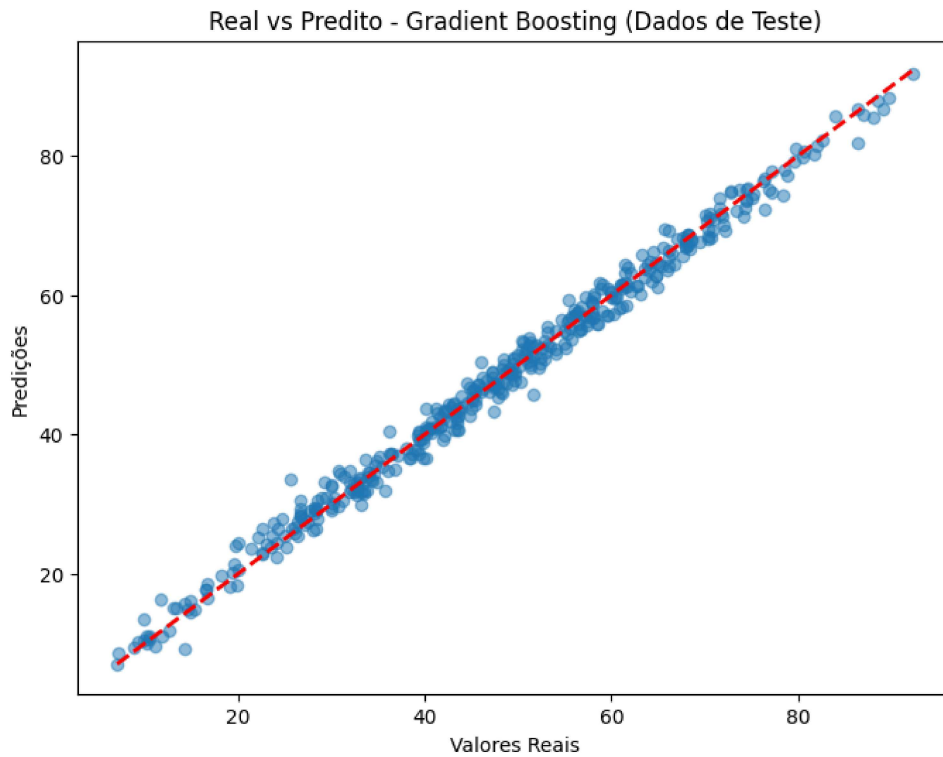
apresentam essa comparação para o modelo CatBoost nos conjuntos feminino e masculino, respectivamente. Complementarmente, as Tabelas 15 e 17 apresentam os resultados para o segundo melhor modelo em cada conjunto: XGBoost no feminino e Gradient Boosting no masculino.

Nessas tabelas, o campo *Índice* corresponde ao identificador do registro no banco de dados, utilizado apenas como referência interna, uma vez que os códigos individuais dos estudantes não podem ser divulgados. A coluna  $\Delta$ Ranking representa a diferença absoluta entre a posição real e a posição prevista pelo modelo.

Para o conjunto feminino, o modelo CatBoost apresentou uma correlação de Spearman de 0,9989 entre o ranking real e o ranking previsto, indicando concordância quase perfeita entre as ordenações. Além disso, verificou-se uma sobreposição de 100% entre os conjuntos, ou seja, todos os escolares presentes no Top 20 real também aparecem no Top 20 previsto pelo modelo. Em relação aos indicadores esportivos, o Top 20 real apresenta sete estudantes selecionadas para os Jogos da Amizade e nenhuma medalhista, sendo que o ranking previsto manteve exatamente a mesma composição.

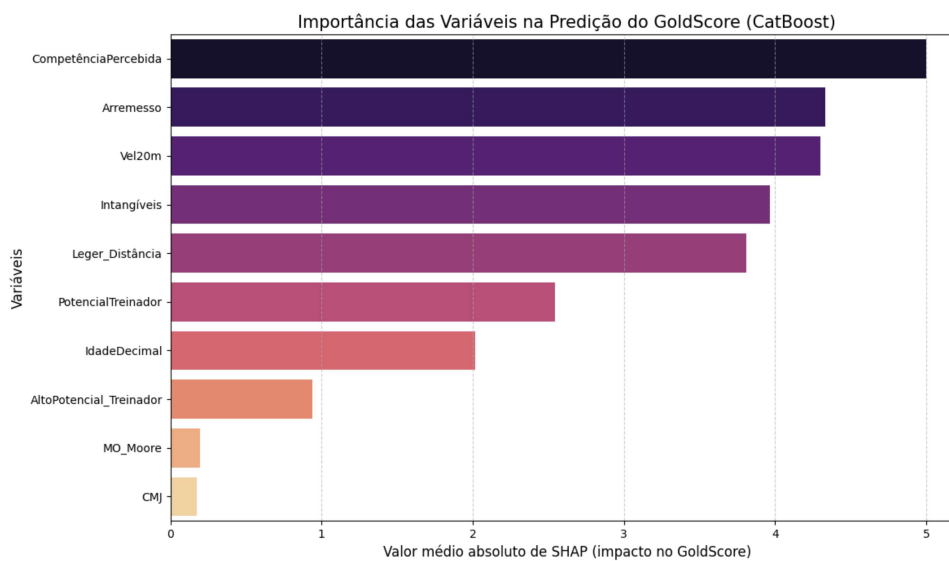
O modelo XGBoost apresentou comportamento similar, com correlação de Spearman de 0,9960 e sobreposição de 95% (19 em comum) no Top 20, mantendo também os sete estudantes selecionados entre as primeiras posições.

Figura 23 – Gráfico de dispersão entre valores reais e preditos pelo modelo Gradient Boosting - Experimento I (Dados Masculinos)



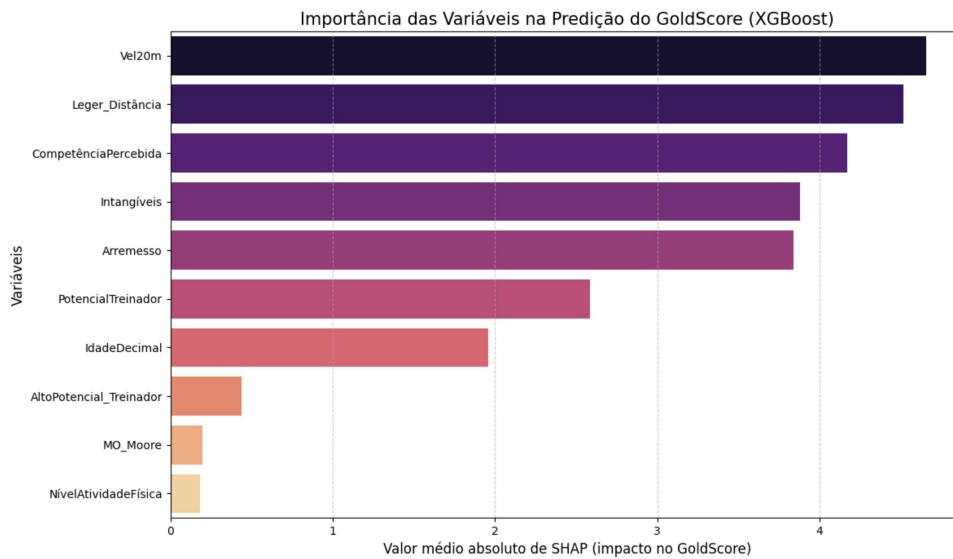
Fonte: Elaborado pela autora.

Figura 24 – Importância das variáveis segundo valores SHAP - CatBoost (Experimento I, Dados Femininos)



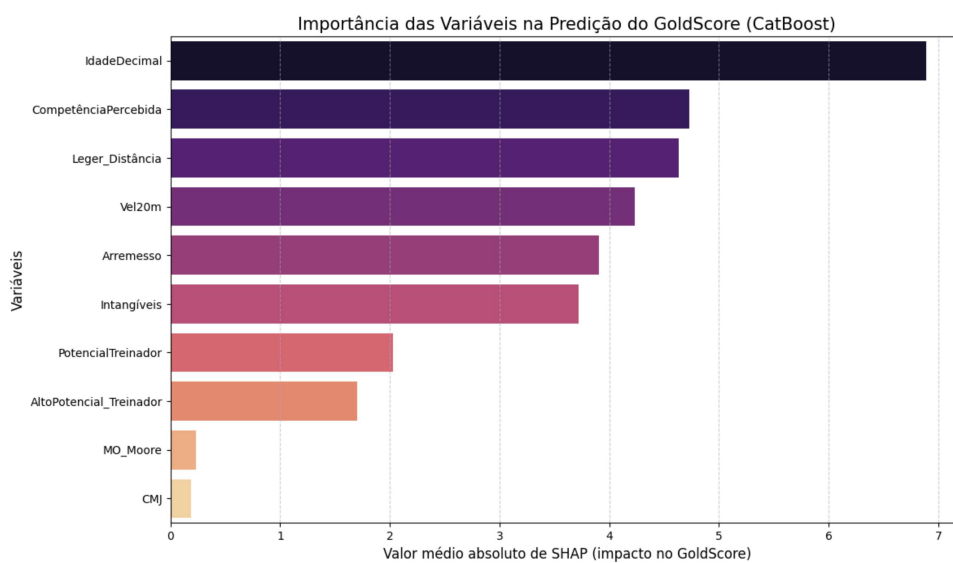
Fonte: Elaborado pela autora.

Figura 25 – Importância das variáveis segundo valores SHAP - XGBoost (Experimento I, Dados Femininos)



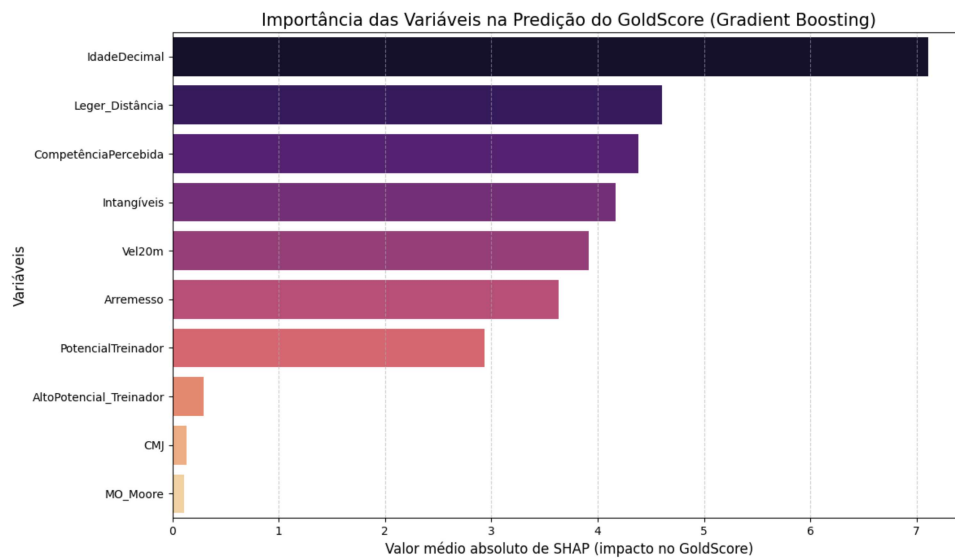
Fonte: Elaborado pela autora.

Figura 26 – Importância das variáveis segundo valores SHAP - CatBoost (Experimento I, Dados Masculinos)



Fonte: Elaborado pela autora.

Figura 27 – Importância das variáveis segundo valores SHAP - Gradient Boosting (Experimento I, Dados Masculinos)



Fonte: Elaborado pela autora.

Tabela 14 – Comparação entre ranking real e ranking previsto pelo CatBoost para os 20 maiores valores de *Gold Score* no conjunto de teste (dados femininos)

Índice	Ranking Real	GS Real	Ranking Prev.	GS Prev.	$\Delta$ Ranking	Selec. Jogos Amizade	Med. Jogos Amizade
1211	1	93,1	1	91,31	0	1	0
427	2	89,1	2	88,86	0	0	0
1201	3	86,7	3	87,53	0	0	0
115	4	86,2	5	84,35	1	1	0
1161	5	85,0	4	84,81	1	0	0
1767	6	84,1	6	84,01	0	0	0
1769	7	82,7	7	83,72	0	0	0
237	8	81,8	8	82,27	0	1	0
1421	9	81,7	9	81,68	0	0	0
495	10	80,8	10	80,48	0	0	0
1476	11	79,9	12	79,50	1	0	0
637	12	79,1	15	78,99	3	0	0
1755	13	78,8	14	79,16	1	0	0
76	14	78,7	13	79,16	1	1	0
1247	15	78,4	11	79,72	4	1	0
1468	16	78,2	16	78,74	0	1	0
398	17	77,4	17	78,00	0	0	0
78	18	76,7	18	76,90	0	0	0
250	19	76,6	20	75,69	1	0	0
261	20	76,2	19	76,02	1	1	0

Fonte: Elaborado pela autora.

Tabela 15 – Comparação entre ranking real e ranking previsto pelo XGBoost para os 20 maiores valores de *Gold Score* no conjunto de teste (dados femininos)

Índice	Ranking Real	GS Real	Ranking Prev.	GS Prev.	$\Delta$ Ranking	Selec. Jogos Amizade	Med. Jogos Amizade
1211	1	93,1	1	92,35	0	1	0
427	2	89,1	2	89,10	0	0	0
1201	3	86,7	4	86,14	1	0	0
115	4	86,2	3	86,37	1	1	0
1161	5	85,0	5	83,58	0	0	0
1767	6	84,1	6	83,21	0	0	0
1769	7	82,7	7	83,17	0	0	0
237	8	81,8	8	81,90	0	1	0
1421	9	81,7	10	79,59	1	0	0
495	10	80,8	9	81,00	1	0	0
1476	11	79,9	14	78,10	3	0	0
637	12	79,1	13	78,69	1	0	0
1755	13	78,8	11	79,11	2	0	0
76	14	78,7	16	77,34	2	1	0
1247	15	78,4	12	78,98	3	1	0
1468	16	78,2	15	77,85	1	1	0
398	17	77,4	17	77,24	0	0	0
78	18	76,7	20	75,90	2	0	0
250	19	76,6	22	74,49	3	0	0
261	20	76,2	19	76,23	1	1	0

Fonte: Elaborado pela autora.

Para o conjunto masculino, o CatBoost apresentou correlação de Spearman de 0,9978 e sobreposição de 95% (19 em comum) entre o Top 20 real e o previsto, mantendo os quatro medalhistas e os nove selecionados.

O Gradient Boosting apresentou correlação de Spearman de 0,9952 e sobreposição de 90% (18 em comum) no Top 20, preservando também os quatro medalhistas e oito dos nove selecionados.

## 5.2 Experimento II: Remoção das Variáveis Subjetivas do Treinador

O Experimento II teve como objetivo avaliar o desempenho dos modelos preditivos após a remoção das variáveis subjetivas do treinador: *AltoPotencial\_Treinador*, *PotencialTreinador* e *Intangíveis*. Este experimento busca responder se é possível construir um modelo preditivo que, mesmo sem essas avaliações especializadas, mantenha capacidade satisfatória de predição do *Gold Score*, utilizando exclusivamente as demais variáveis disponíveis no conjunto de dados.

Tabela 16 – Comparação entre ranking real e ranking previsto pelo CatBoost para os 20 maiores valores de *Gold Score* no conjunto de teste (dados masculinos)

Índice	Ranking Real	GS Real	Ranking Prev.	GS Prev.	$\Delta$ Ranking	Selec. Jogos Amizade	Med. Jogos Amizade
416	1	92,2	1	93,79	0	1	1
1431	2	89,7	3	87,93	1	0	0
952	3	89,1	6	86,93	3	0	0
324	4	88,5	2	88,02	2	0	0
135	5	88,0	4	87,32	1	1	1
1405	6	87,0	7	86,51	1	1	1
810	7	86,3	5	87,25	2	1	0
1000	7	86,3	8	84,70	1	1	0
1708	9	84,0	9	83,35	0	0	0
979	10	82,6	10	82,99	0	0	0
1124	11	81,9	12	81,54	1	0	0
422	12	81,7	14	80,94	2	0	0
581	13	80,6	11	82,05	2	1	0
203	14	80,4	13	81,35	1	0	0
807	15	79,7	15	79,54	0	1	0
772	16	79,6	23	75,89	7	0	0
507	17	78,8	16	78,82	1	1	1
298	18	78,5	18	77,14	0	0	0
817	19	78,3	17	77,49	2	0	0
1647	20	77,1	20	76,71	0	1	0

Fonte: Elaborado pela autora.

### 5.2.1 Desempenho Preditivo dos Modelos

Após a remoção das variáveis subjetivas, os modelos foram novamente treinados e avaliados. As Tabelas 18 e 19 apresentam os resultados para os conjuntos feminino e masculino, respectivamente.

Os resultados apresentados na Tabela 18 indicam que, para os dados femininos, os algoritmos baseados em *gradient boosting* continuam apresentando os melhores desempenhos. O modelo CatBoost obteve o maior coeficiente de determinação ( $R^2 = 0,8724$ ), seguido de XGBoost ( $R^2 = 0,8709$ ) e Gradient Boosting ( $R^2 = 0,8644$ ).

Resultados semelhantes foram observados para os dados masculinos, apresentados na Tabela 19. O modelo CatBoost novamente apresentou o melhor desempenho ( $R^2 = 0,8681$ ), seguido por XGBoost ( $R^2 = 0,8616$ ) e Gradient Boosting ( $R^2 = 0,8504$ ).

Mesmo após a remoção das variáveis subjetivas, os modelos mantiveram elevado poder explicativo, com valores de  $R^2$  próximos de 0,87 em ambos os conjuntos de dados. Esse resultado indica que grande parte da variabilidade do *Gold Score* pode ser explicada a partir das demais variáveis presentes no conjunto de dados.

Tabela 17 – Comparação entre ranking real e ranking previsto pelo Gradient Boosting para os 20 maiores valores de *Gold Score* no conjunto de teste (dados masculinos)

Índice	Ranking Real	GS Real	Ranking Prev.	GS Prev.	$\Delta$ Ranking	Selec. Jogos Amizade	Med. Jogos Amizade
416	1	92,2	1	91,72	0	1	1
1431	2	89,7	2	88,35	0	0	0
952	3	89,1	5	86,68	2	0	0
324	4	88,5	3	87,80	1	0	0
135	5	88,0	8	85,46	3	1	1
1405	6	87,0	6	85,94	0	1	1
810	7	86,3	4	86,73	3	1	0
1000	7	86,3	10	81,84	3	1	0
1708	9	84,0	7	85,60	2	0	0
979	10	82,6	9	82,30	1	0	0
1124	11	81,9	11	81,41	0	0	0
422	12	81,7	14	80,18	2	0	0
581	13	80,6	13	80,62	0	1	0
203	14	80,4	15	79,82	1	0	0
807	15	79,7	12	81,03	3	1	0
772	16	79,6	16	79,26	0	0	0
507	17	78,8	19	77,13	2	1	1
298	18	78,5	17	77,86	1	0	0
817	19	78,3	30	74,39	11	0	0
1647	20	77,1	27	74,77	7	1	0

Fonte: Elaborado pela autora.

Tabela 18 – Ranking de performance dos modelos no conjunto de teste - Experimento II (Dados Femininos)

Modelo	$R^2$ Teste	MAE	RMSE
CatBoost	0,8724	4,9513	6,2250
XGBoost	0,8709	5,0212	6,2598
Gradient Boosting	0,8644	5,0978	6,4165
LightGBM	0,8585	5,2533	6,5535
Linear Regression	0,8451	5,5886	6,8576
Lasso	0,8450	5,6089	6,8604
Ridge	0,8448	5,6013	6,8644
ElasticNet	0,8447	5,6104	6,8671
SVR	0,8444	5,5628	6,8744
Extra Tree	0,8183	5,9042	7,4274
Random Forest	0,8123	6,1973	7,5490
Decision Tree	0,7153	7,5257	9,2966
KNN	0,6560	8,2802	10,2203

Fonte: Elaborado pela autora.

**Nota:** Para as métricas MAE e RMSE, valores menores indicam melhor desempenho preditivo ( $\downarrow$  erro). Para o  $R^2$ , valores maiores indicam melhor ajuste do modelo ( $\uparrow$  poder explicativo).

Tabela 19 – Ranking de performance dos modelos no conjunto de teste - Experimento II (Dados Masculinos)

Modelo	$R^2$ Teste	MAE	RMSE
CatBoost	0,8681	5,2962	6,5962
XGBoost	0,8616	5,3457	6,7576
Gradient Boosting	0,8504	5,5722	7,0265
LightGBM	0,8458	5,6828	7,1333
SVR	0,8239	6,0270	7,6219
Ridge	0,8194	6,1660	7,7190
Linear Regression	0,8192	6,1597	7,7238
ElasticNet	0,8191	6,1770	7,7262
Lasso	0,8188	6,1828	7,7318
Random Forest	0,7977	6,6685	8,1701
Extra Tree	0,7856	6,9077	8,4110
Decision Tree	0,6715	8,3941	10,4100
KNN	0,5995	9,1609	11,4954

Fonte: Elaborado pela autora.

**Nota:** Para as métricas MAE e RMSE, valores menores indicam melhor desempenho preditivo ( $\downarrow$  erro). Para o  $R^2$ , valores maiores indicam melhor ajuste do modelo ( $\uparrow$  poder explicativo).

### 5.2.2 Análise Gráfica das Previsões

Considerando o desempenho superior obtido pelos algoritmos baseados em *boosting*, as Figuras 28 a 31 apresentam os gráficos de dispersão para os dois melhores modelos em cada conjunto.

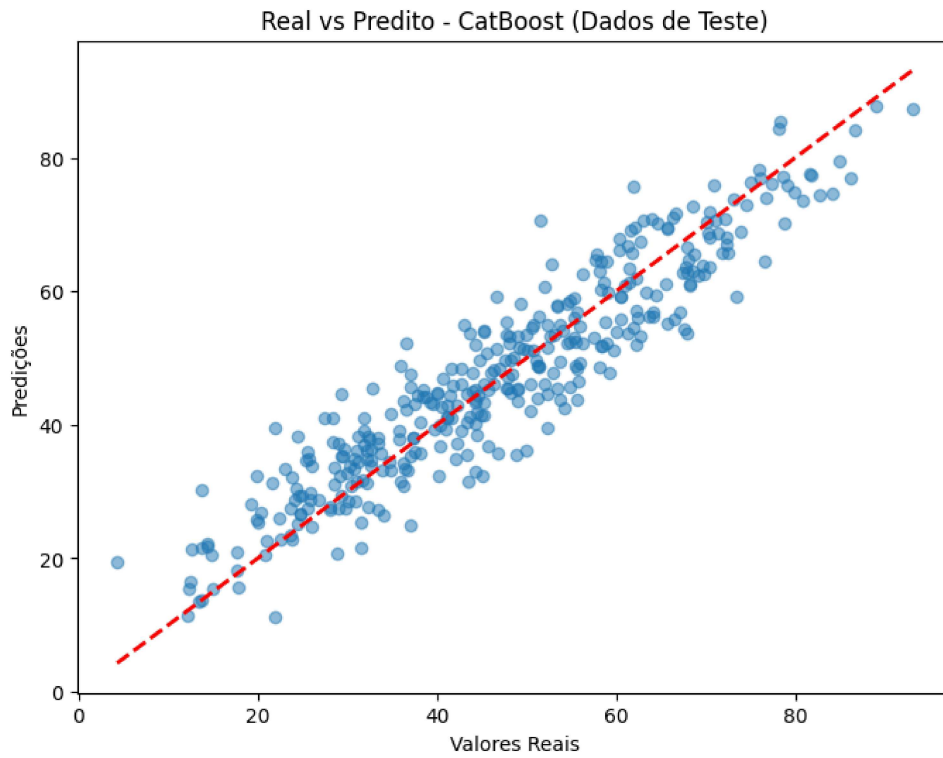
Em todos os casos, observa-se que grande parte dos pontos se concentra próxima à linha de identidade, indicando boa concordância entre valores observados e estimados, embora com dispersão visivelmente maior que no Experimento I. Esse padrão se mantém de forma consistente entre os dois melhores modelos e em ambos os conjuntos de dados, sugerindo que os algoritmos de *boosting* foram capazes de capturar adequadamente as relações presentes nos dados, ainda que com precisão reduzida após a remoção das variáveis subjetivas.

### 5.2.3 Análise de Importância das Variáveis

Para compreender melhor o comportamento dos modelos e identificar quais variáveis passaram a ter maior relevância após a remoção das avaliações subjetivas, foi realizada uma nova análise de importância das variáveis utilizando os valores SHAP para os dois melhores algoritmos em cada conjunto. As Figuras 32 e 33 apresentam os resultados para CatBoost e XGBoost nos dados femininos. Para os dados masculinos, as Figuras 34 e 35 apresentam as importâncias para os mesmos modelos.

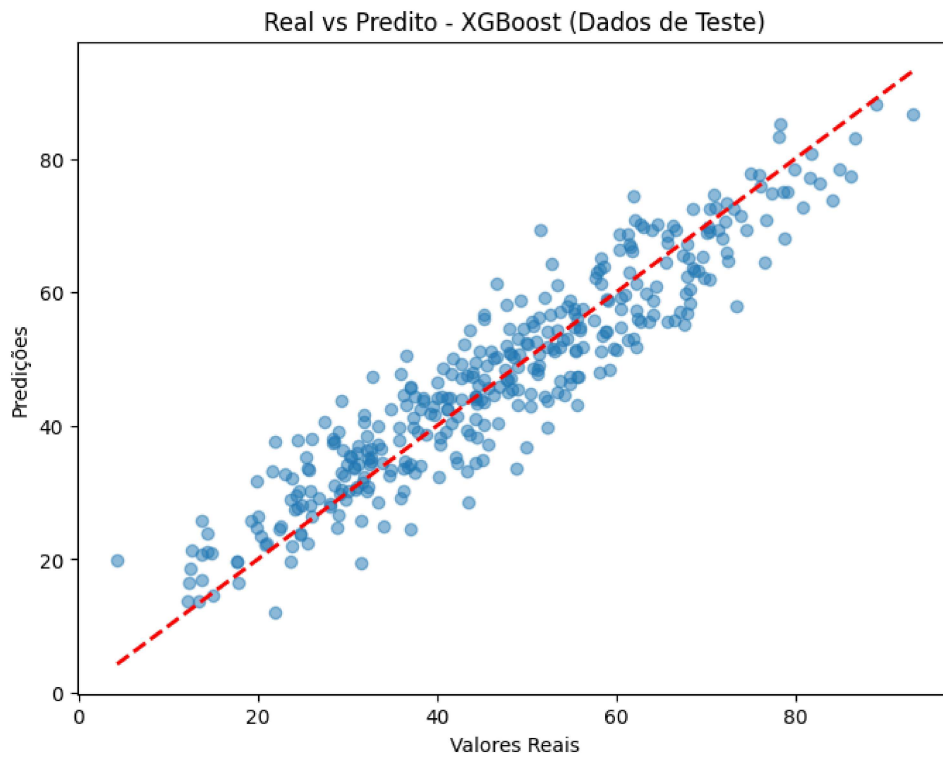
Observa-se que, com a remoção das avaliações subjetivas do treinador, variáveis

Figura 28 – Gráfico de dispersão - CatBoost (Experimento II, Dados Femininos)



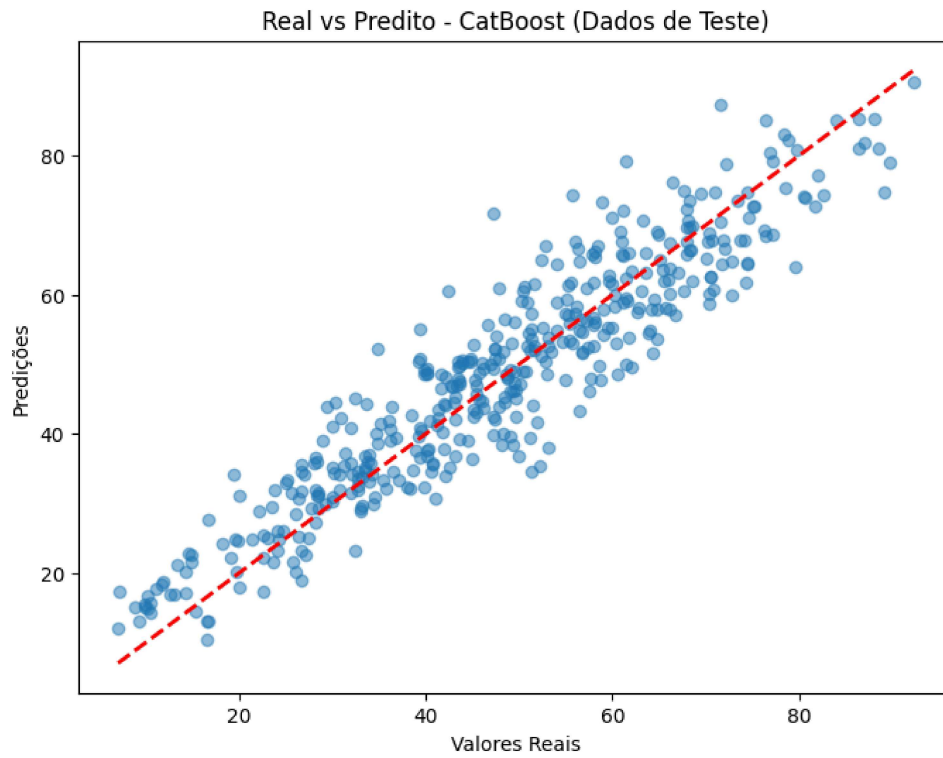
Fonte: Elaborado pela autora.

Figura 29 – Gráfico de dispersão - XGBoost (Experimento II, Dados Femininos)



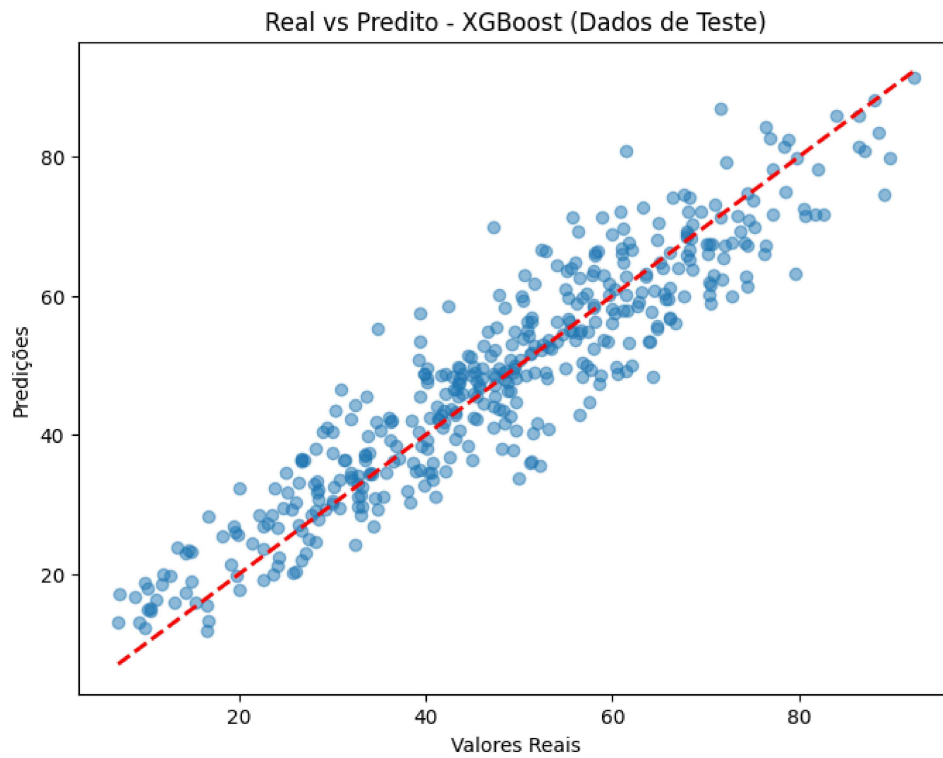
Fonte: Elaborado pela autora.

Figura 30 – Gráfico de dispersão - CatBoost (Experimento II, Dados Masculinos)



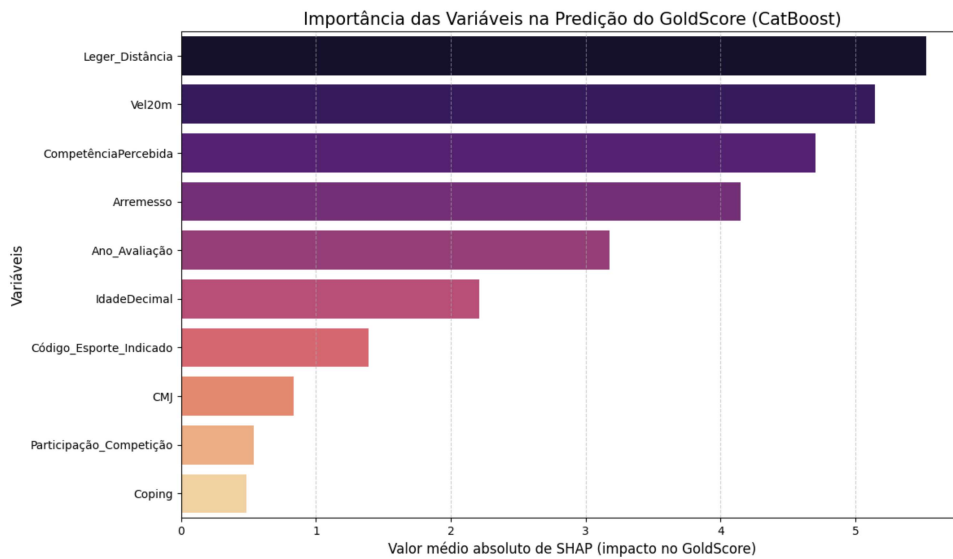
Fonte: Elaborado pela autora.

Figura 31 – Gráfico de dispersão - XGBoost (Experimento II, Dados Masculinos)



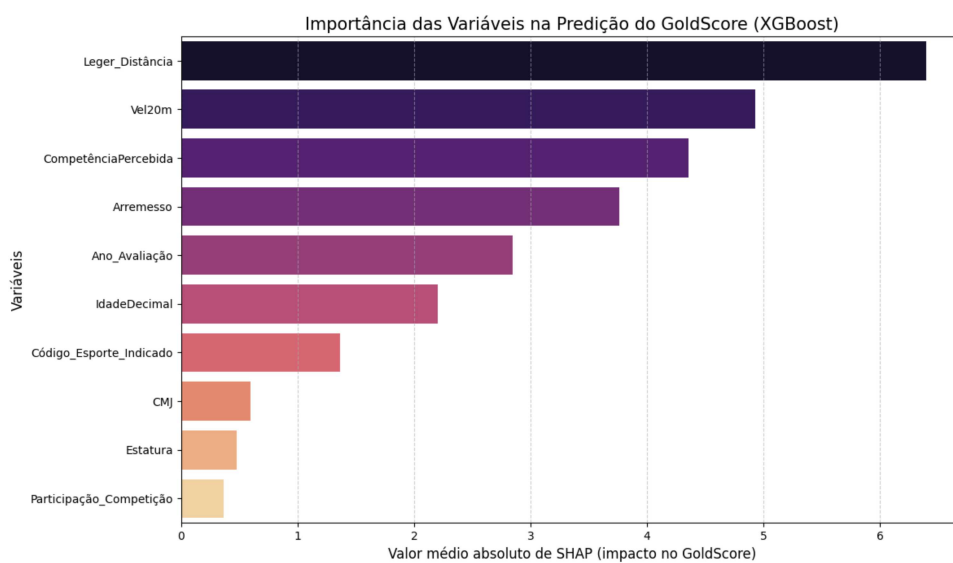
Fonte: Elaborado pela autora.

Figura 32 – Importância das variáveis segundo valores SHAP - CatBoost (Experimento II, Dados Femininos)



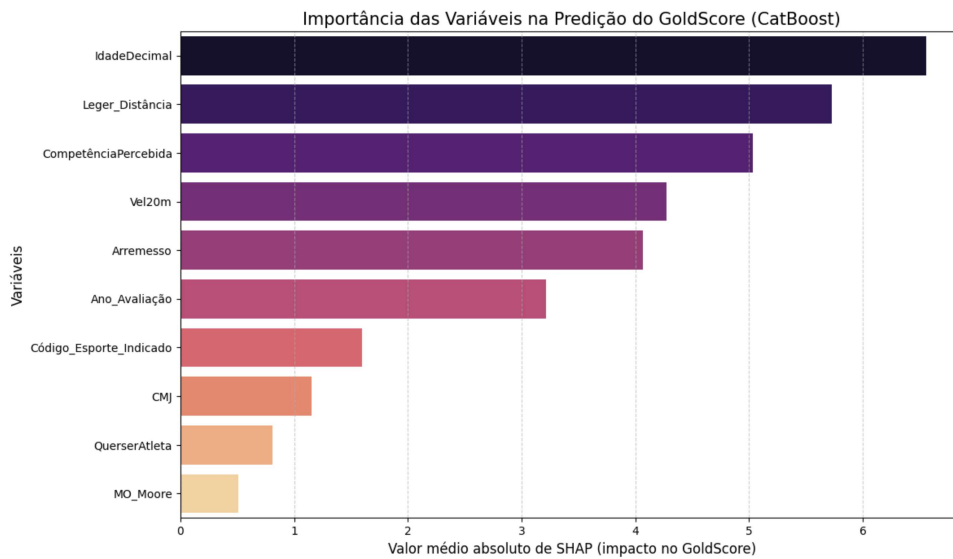
Fonte: Elaborado pela autora.

Figura 33 – Importância das variáveis segundo valores SHAP - XGBoost (Experimento II, Dados Femininos)



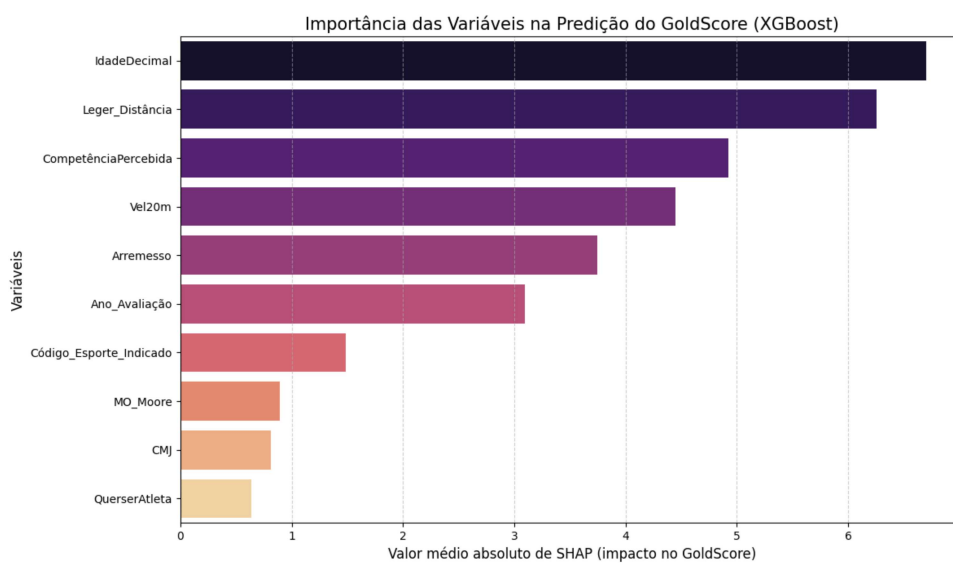
Fonte: Elaborado pela autora.

Figura 34 – Importância das variáveis segundo valores SHAP - CatBoost (Experimento II, Dados Masculinos)



Fonte: Elaborado pela autora.

Figura 35 – Importância das variáveis segundo valores SHAP - XGBoost (Experimento II, Dados Masculinos)



Fonte: Elaborado pela autora.

relacionadas ao desempenho físico objetivo passam a figurar entre as mais importantes em ambos os modelos e conjuntos de dados. Destacam-se a distância percorrida no teste de Léger (*Leger\_Distância*), a velocidade em 20 metros (*Vel20m*) e o arremesso de medicine ball (*Arremesso*), que aparecem consistentemente nas primeiras posições.

Entretanto, um aspecto relevante observado em todas as análises foi a elevada importância atribuída à variável *Ano\_Avaliação*. Esta variável não representa uma característica individual do estudante, mas sim um aspecto relacionado ao contexto temporal em que os dados foram coletados. Sua posição de destaque na importância SHAP — figurando entre as cinco mais relevantes em todos os modelos — indica que os algoritmos podem estar capturando padrões associados ao período de avaliação, e não exclusivamente às características esportivas dos participantes.

A Tabela 20 apresenta a distribuição etária e dos valores de *Gold Score* dos participantes do sexo masculino ao longo dos anos de coleta, enquanto a Tabela 21 traz os dados para o sexo feminino. Essas diferenças na composição etária e nos escores da amostra justificam a relevância da variável *Ano\_Avaliação* nos modelos preditivos.

Tabela 20 – Distribuição da idade e do *Gold Score* dos participantes (masculino) por ano de avaliação

<b>Ano</b>	<b>n</b>	<b>Idade (média ± DP)</b>	<b>Gold Score (média ± DP)</b>
2015	212	14,52 ± 1,94	56,03 ± 19,84
2016	285	13,96 ± 1,81	51,19 ± 19,02
2017	327	14,03 ± 1,93	53,28 ± 19,67
2018	359	14,05 ± 1,99	47,57 ± 18,93
2019	372	13,93 ± 1,99	51,45 ± 20,12
2022	369	14,32 ± 2,10	43,12 ± 17,86
2023	212	13,07 ± 1,41	40,30 ± 16,74

Fonte: Elaborado pela autora.

Tabela 21 – Distribuição da idade e do *Gold Score* dos participantes (feminino) por ano de avaliação

<b>Ano</b>	<b>n</b>	<b>Idade (média ± DP)</b>	<b>Gold Score (média ± DP)</b>
2015	172	13,96 ± 2,07	54,41 ± 18,76
2016	223	13,44 ± 1,64	52,32 ± 18,43
2017	249	13,47 ± 1,76	54,13 ± 18,91
2018	311	13,70 ± 1,95	45,93 ± 17,02
2019	342	13,80 ± 1,87	46,58 ± 17,34
2022	344	14,11 ± 2,04	41,60 ± 15,98
2023	191	13,05 ± 1,28	41,07 ± 16,23

Fonte: Elaborado pela autora.

Observa-se uma variação significativa na composição etária e nos valores de *Gold Score* ao longo dos anos em ambos os sexos:

- **Ano de 2023:** apresenta as menores médias de idade (13,07 anos no masculino; 13,05 no feminino) e os menores desvios padrão (1,41 e 1,28, respectivamente), indicando amostras mais jovens e homogêneas. Coincidentemente, também apresenta as menores médias de *Gold Score* (40,30 no masculino; 41,07 no feminino).
- **Anos de 2015-2017:** concentram as maiores médias de *Gold Score* em ambos os sexos, associadas a amostras com idades médias mais elevadas e maior variabilidade etária.
- **Tendência decrescente:** observa-se uma queda progressiva nos valores médios de *Gold Score* a partir de 2018 em ambos os sexos, com os menores valores concentrados nos anos mais recentes (2022 e 2023).
- **Padrão consistente:** a relação entre ano de coleta e *Gold Score* se mantém similar entre os sexos, sugerindo que fatores contextuais ou mudanças no perfil da amostra ao longo do tempo afetam ambos os grupos de forma semelhante.

Essa heterogeneidade na distribuição etária e nos escores ao longo do tempo ajuda a explicar a elevada importância da variável *Ano\_Avaliação* nos modelos preditivos. No entanto, é importante ressaltar que essa variável pode estar associada a uma combinação de fatores que transcendem a simples diferença etária, tais como:

- **Mudanças no perfil socioeconômico e cultural** dos estudantes ao longo dos anos;
- **Variações no processo de avaliação** ou na calibragem dos instrumentos de coleta;
- **Diferenças nas condições de treinamento e preparação** oferecidas pela instituição;
- **Flutuações no nível de exigência ou nos critérios de avaliação** do *Gold Score*;
- **Características não mensuradas das coortes**, como motivação, engajamento ou exposição prévia ao esporte.

A consistência desse padrão entre os sexos reforça a necessidade de considerar esses fatores contextuais na modelagem, uma vez que a variável temporal pode estar incorporando efeitos que não são estritamente relacionados às características individuais dos escolares.

Diante dessa constatação, tornou-se necessário redelinear os experimentos subsequentes com o objetivo de reduzir possíveis vieses associados a variáveis contextuais e garantir que as previsões sejam fundamentadas principalmente nas características físicas, motoras e psicossociais dos escolares, minimizando a influência de fatores externos.

### 5.2.4 Capacidade de Ordenação dos Escolares com Melhor *Gold Score*

Para avaliar a capacidade dos modelos em preservar as posições mais altas do ranking após a remoção das variáveis subjetivas, foi realizada uma comparação entre o ranking real e o ranking previsto para os 20 escolares com maiores valores de *Gold Score* no conjunto de teste. As tabelas completas com a comparação detalhada para os dois melhores modelos (CatBoost e XGBoost) em ambos os sexos encontram-se no Apêndice B, evitando a quebra do fluxo de discussão dos resultados principais.

Para o conjunto feminino, o modelo CatBoost apresentou uma correlação de Spearman de 0,9311 entre o ranking real e o previsto, com sobreposição de 80% (16 em comum) no Top 20. Em relação aos indicadores esportivos, o Top 20 real apresenta sete estudantes selecionadas para os Jogos da Amizade, sendo que o ranking previsto pelo CatBoost manteve oito dessas atletas, indicando uma leve superestimação. O XGBoost apresentou comportamento similar, com correlação de 0,9322 e também 80% de sobreposição, mantendo oito selecionadas no Top 20 previsto.

Para o conjunto masculino, o CatBoost alcançou correlação de Spearman de 0,9286 e sobreposição de 70% (14 em comum) no Top 20, preservando todos os quatro medalhistas e oito dos nove selecionados. O XGBoost apresentou resultados equivalentes, com correlação de 0,9242 e também 70% de sobreposição, mantendo os quatro medalhistas e oito selecionados.

Esses resultados indicam que, mesmo com a redução no poder preditivo absoluto (refletido nos valores de  $R^2$ ), os modelos mantiveram boa capacidade de identificar os escolares de maior potencial, preservando a maior parte dos medalhistas e selecionados entre as primeiras posições do ranking.

## 5.3 Experimento III: Remoção da Variável Temporal

No Experimento III foi realizada a remoção da variável **Ano\_Avaliação** do conjunto de variáveis utilizadas no treinamento dos modelos. Conforme discutido na seção anterior, a presença dessa variável poderia introduzir um viés temporal no processo de modelagem, permitindo que o algoritmo aprendesse padrões específicos associados ao período de coleta dos dados em vez de capturar relações estruturais relacionadas às características individuais dos escolares.

### 5.3.1 Desempenho Preditivo dos Modelos

Após a remoção da variável temporal, os modelos foram novamente treinados e avaliados. As Tabelas 22 e 23 apresentam os resultados para os conjuntos masculino e feminino.

Tabela 22 – Ranking de performance dos modelos no conjunto de teste - Experimento III (Dados Masculinos)

Modelo	$R^2$ Teste	MAE	RMSE
CatBoost	0,8422	5,8182	7,2151
XGBoost	0,8348	5,9628	7,3829
Gradient Boosting	0,8309	5,9451	7,4693
LightGBM	0,8265	6,0670	7,5661
SVR	0,8206	6,0856	7,6937
Lasso	0,8063	6,4656	7,9943
ElasticNet	0,8058	6,4590	8,0043
Ridge	0,8054	6,4608	8,0134
Linear Regression	0,8050	6,4557	8,0201
Extra Tree	0,7982	6,6948	8,1593
Random Forest	0,7959	6,6766	8,2067
Decision Tree	0,6850	8,1541	10,1946
KNN	0,6583	8,5270	10,6182

Fonte: Elaborado pela autora.

**Nota:** Para as métricas MAE e RMSE, valores menores indicam melhor desempenho preditivo ( $\downarrow$  erro). Para o  $R^2$ , valores maiores indicam melhor ajuste do modelo ( $\uparrow$  poder explicativo).

Para o conjunto masculino, o modelo CatBoost apresentou o melhor desempenho, com  $R^2 = 0,8422$ ,  $MAE = 5,82$  e  $RMSE = 7,22$ . Observa-se uma redução no poder preditivo em relação ao Experimento II (de 0,8681 para 0,8422), indicando que parte da variabilidade explicada anteriormente estava associada à informação temporal.

Para o conjunto feminino, o CatBoost também apresentou o melhor resultado, com  $R^2 = 0,8449$ ,  $MAE = 5,54$  e  $RMSE = 6,86$ . Similarmente ao observado no masculino, houve uma redução no  $R^2$  em relação ao Experimento II (de 0,8724 para 0,8449).

Observa-se que os algoritmos baseados em *boosting* (CatBoost, XGBoost e Gradient Boosting) continuam apresentando desempenho consistentemente superior aos demais métodos avaliados, mantendo-se como as melhores opções para modelagem mesmo após a remoção da variável temporal.

### 5.3.2 Análise Gráfica das Previsões

Considerando o desempenho superior obtido pelos algoritmos baseados em *boosting*, as Figuras 36 a 39 apresentam os gráficos de dispersão para os dois melhores modelos em cada conjunto.

Em todos os casos, observa-se que os pontos se concentram ao redor da linha de identidade, embora com dispersão ligeiramente maior que nos experimentos anteriores, o que é consistente com a redução nos valores de  $R^2$ . O padrão se mantém de forma consistente entre os dois melhores modelos e em ambos os conjuntos de dados.

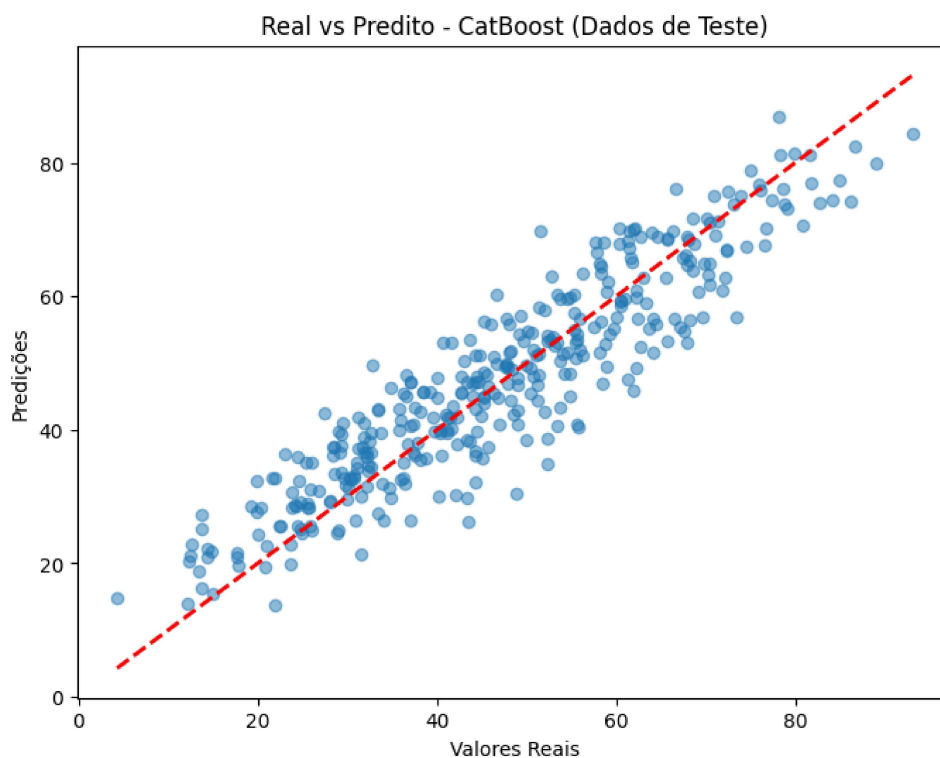
Tabela 23 – Ranking de performance dos modelos no conjunto de teste - Experimento III (Dados Femininos)

Modelo	$R^2$ Teste	MAE	RMSE
CatBoost	0,8449	5,5427	6,8620
XGBoost	0,8431	5,6348	6,9019
Gradient Boosting	0,8347	5,7690	7,0851
SVR	0,8347	5,7635	7,0854
Linear Regression	0,8332	5,7957	7,1170
Lasso	0,8324	5,8337	7,1325
Ridge	0,8323	5,8223	7,1356
ElasticNet	0,8318	5,8522	7,1473
LightGBM	0,8293	5,8099	7,1993
Extra Tree	0,8025	6,2773	7,7429
Random Forest	0,7996	6,4361	7,7998
Decision Tree	0,7117	7,6105	9,3558
KNN	0,6545	8,1522	10,2427

Fonte: Elaborado pela autora.

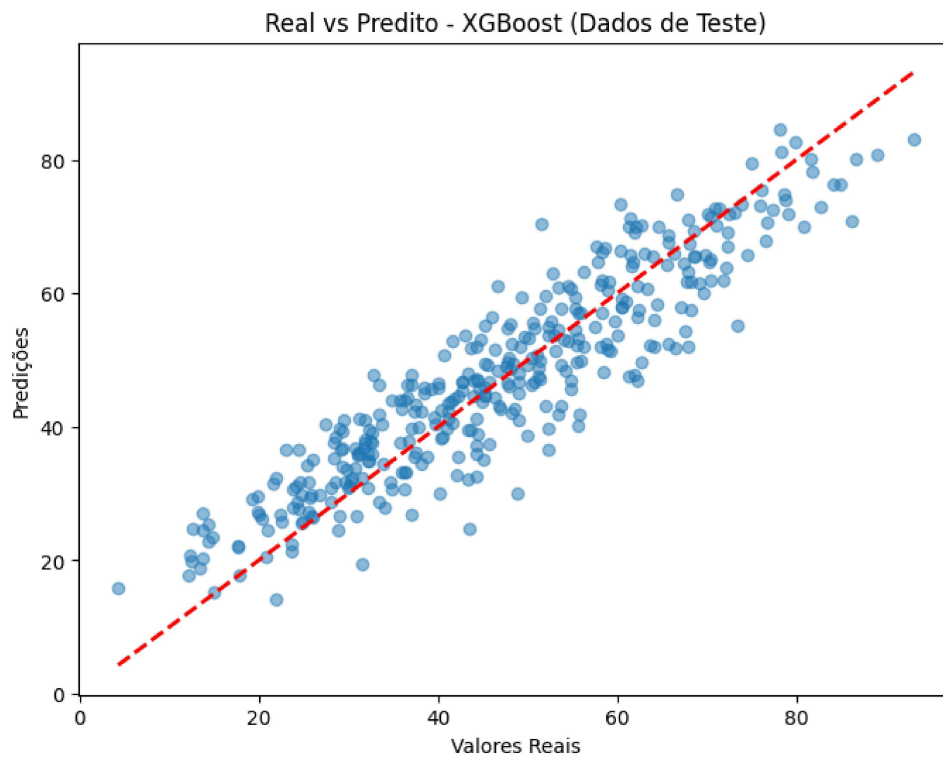
**Nota:** Para as métricas MAE e RMSE, valores menores indicam melhor desempenho preditivo ( $\downarrow$  erro). Para o  $R^2$ , valores maiores indicam melhor ajuste do modelo ( $\uparrow$  poder explicativo).

Figura 36 – Gráfico de dispersão - CatBoost (Experimento III, Dados Femininos)



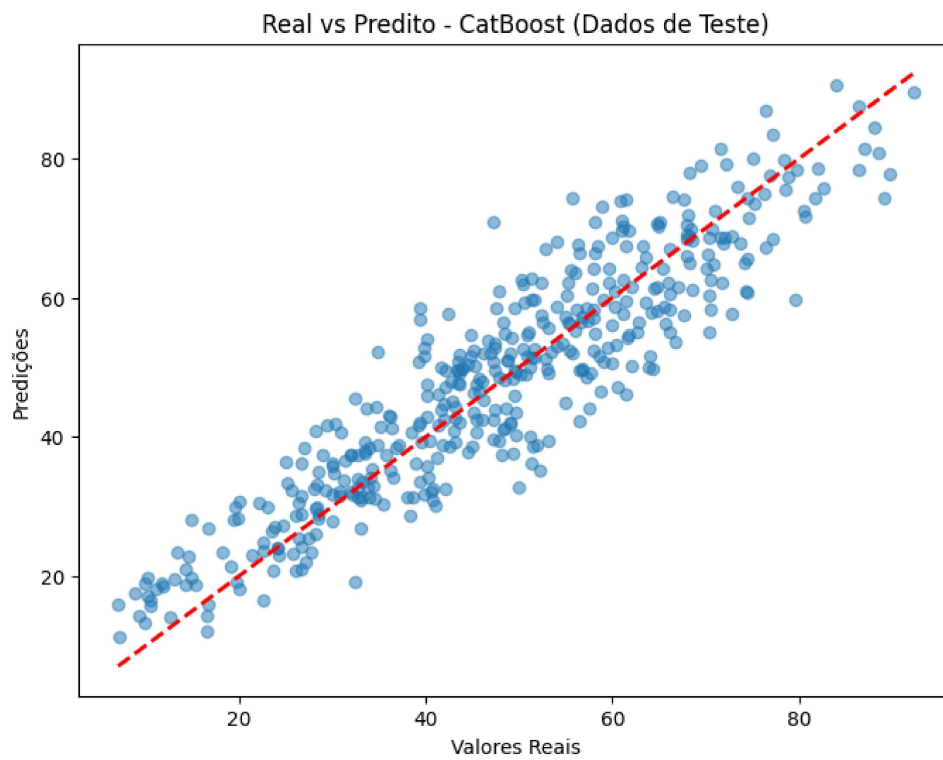
Fonte: Elaborado pela autora.

Figura 37 – Gráfico de dispersão - XGBoost (Experimento III, Dados Femininos)



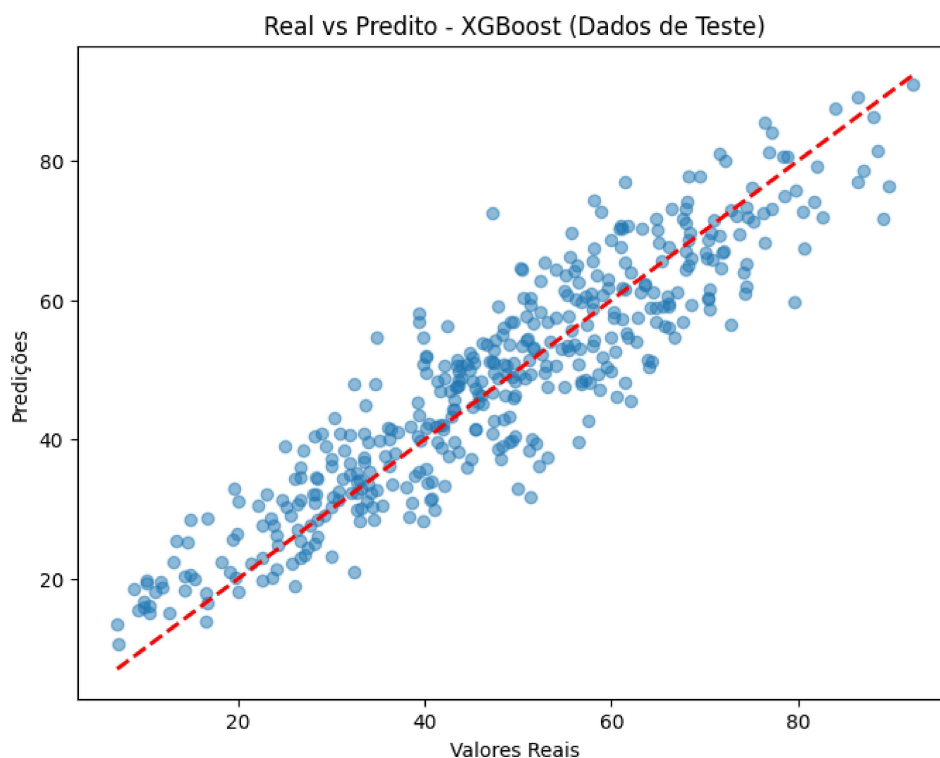
Fonte: Elaborado pela autora.

Figura 38 – Gráfico de dispersão - CatBoost (Experimento III, Dados Masculinos)



Fonte: Elaborado pela autora.

Figura 39 – Gráfico de dispersão - XGBoost (Experimento III, Dados Masculinos)



Fonte: Elaborado pela autora.

### 5.3.3 Análise de Importância das Variáveis

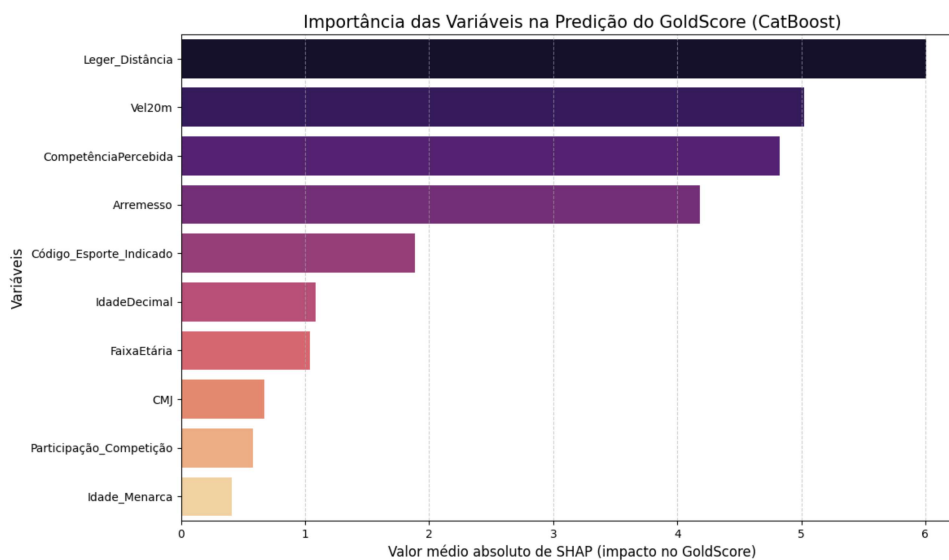
Para compreender melhor o comportamento dos modelos após a remoção da variável `Ano_Avaliação` e identificar quais variáveis passaram a ter maior relevância, foi realizada uma nova análise de importância das variáveis utilizando os valores SHAP para os dois melhores algoritmos em cada conjunto. As Figuras 40 e 41 apresentam os resultados para CatBoost e XGBoost nos dados femininos. Para os dados masculinos, as Figuras 42 e 43 apresentam as importâncias para os mesmos modelos.

Observa-se que, com a remoção da variável `Ano_Avaliação`, as medidas de desempenho físico objetivo assumem o protagonismo na importância preditiva. Destacam-se a distância percorrida no teste de Léger (`Leger_Distância`), a velocidade em 20 metros (`Vel20m`) e a competência percebida (`CompetênciaPercebida`), que aparecem consistentemente nas primeiras posições em todos os modelos e conjuntos de dados.

No entanto, chama a atenção o surgimento ou fortalecimento de variáveis como `FaixaEtária` e `IdadeDecimal`, que passam a figurar entre as mais importantes, especialmente no masculino. Isso sugere que, na ausência da informação temporal explícita (`Ano_Avaliação`), os modelos passam a se apoiar mais fortemente em variáveis relacionadas à idade e à maturação dos escolares para realizar as previsões.

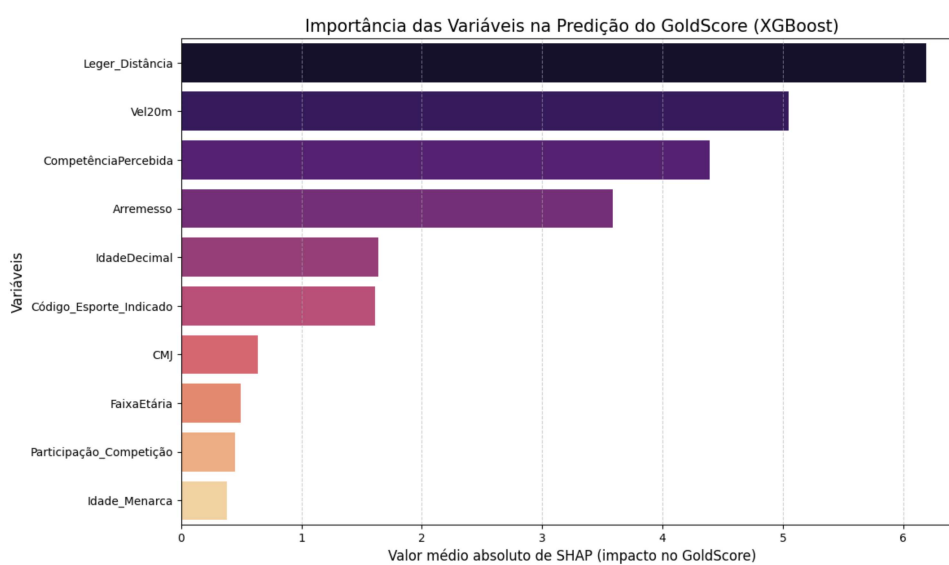
A consistência desse padrão entre os dois melhores modelos e em ambos os sexos

Figura 40 – Importância das variáveis segundo valores SHAP - CatBoost (Experimento III, Dados Femininos)



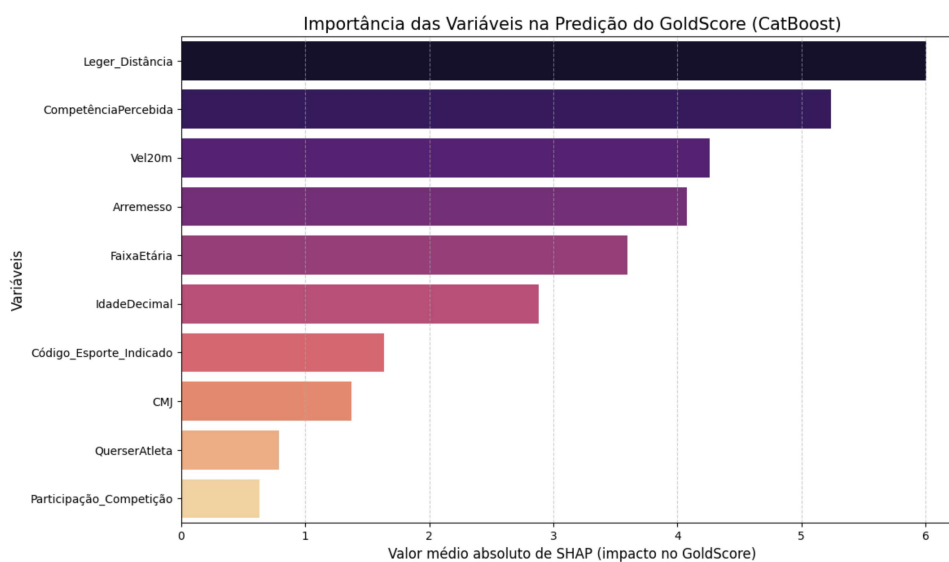
Fonte: Elaborado pela autora.

Figura 41 – Importância das variáveis segundo valores SHAP - XGBoost (Experimento III, Dados Femininos)



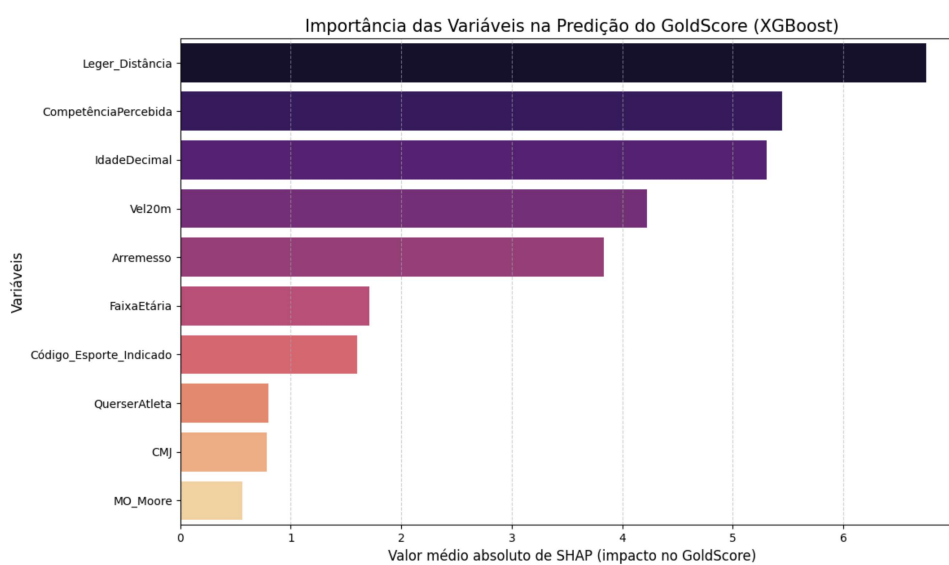
Fonte: Elaborado pela autora.

Figura 42 – Importância das variáveis segundo valores SHAP - CatBoost (Experimento III, Dados Masculinos)



Fonte: Elaborado pela autora.

Figura 43 – Importância das variáveis segundo valores SHAP - XGBoost (Experimento III, Dados Masculinos)



Fonte: Elaborado pela autora.

reforça a importância das características individuais — especialmente as físicas e motoras — na determinação do *Gold Score*, ao mesmo tempo que evidencia a necessidade de considerar adequadamente os efeitos etários na modelagem.

#### 5.3.4 Capacidade de Ordenação dos Escolares com Melhor *Gold Score*

Para avaliar a capacidade dos modelos em preservar as posições mais altas do ranking após a remoção da variável temporal, foi realizada uma comparação entre o ranking real e o ranking previsto para os 20 escolares com maiores valores de *Gold Score* no conjunto de teste. As Tabelas 24 a 27 apresentam essa comparação para os dois melhores modelos em cada conjunto.

Tabela 24 – Comparação Top 20 - CatBoost (Experimento III, Dados Femininos)

Índice	Ranking Real	GS Real	Ranking Prev.	GS Prev.	$\Delta$ Ranking	Selec. Jogos Amizade	Med. Jogos Amizade
1211	1	93,1	2	84,38	1	1	0
427	2	89,1	7	80,03	5	0	0
1201	3	86,7	3	82,38	0	0	0
115	4	86,2	20	74,32	16	1	0
1161	5	85,0	9	77,44	4	0	0
1767	6	84,1	18	74,51	12	0	0
1769	7	82,7	21	74,01	14	0	0
237	8	81,8	10	77,07	2	1	0
1421	9	81,7	6	81,16	3	0	0
495	10	80,8	29	70,59	19	0	0
1476	11	79,9	4	81,37	7	0	0
637	12	79,1	24	73,27	12	0	0
1755	13	78,8	23	73,78	10	0	0
76	14	78,7	13	76,08	1	1	0
1247	15	78,4	5	81,29	10	1	0
1468	16	78,2	1	86,96	15	1	0
398	17	77,4	19	74,36	2	0	0
78	18	76,7	32	70,16	14	0	0
250	19	76,6	50	67,59	31	0	0
261	20	76,2	14	75,84	6	1	0

Fonte: Elaborado pela autora.

Para o conjunto feminino, o modelo CatBoost apresentou correlação de Spearman de 0,9168 entre o ranking real e o previsto, com sobreposição de 70% (14 em comum) no Top 20. Em relação aos indicadores esportivos, o Top 20 real apresenta sete estudantes selecionadas para os Jogos da Amizade, sendo que o ranking previsto pelo CatBoost manteve oito dessas atletas, indicando uma leve superestimação. O XGBoost apresentou comportamento similar, com correlação de 0,9185 e também 70% de sobreposição, porém mantendo exatamente as sete selecionadas no Top 20 previsto.

Tabela 25 – Comparação Top 20 - XGBoost (Experimento III, Dados Femininos)

Índice	Ranking Real	GS Real	Ranking Prev.	GS Prev.	$\Delta$ Ranking	Selec. Jogos Amizade	Med. Jogos Amizade
1211	1	93,1	2	83,01	1	1	0
427	2	89,1	5	80,71	3	0	0
1201	3	86,7	6	80,17	3	0	0
115	4	86,2	30	70,78	26	1	0
1161	5	85,0	10	76,36	5	0	0
1767	6	84,1	11	76,28	5	0	0
1769	7	82,7	19	73,05	12	0	0
237	8	81,8	9	78,21	1	1	0
1421	9	81,7	7	80,08	2	0	0
495	10	80,8	35	70,02	25	0	0
1476	11	79,9	3	82,61	8	0	0
637	12	79,1	26	71,83	14	0	0
1755	13	78,8	15	73,93	2	0	0
76	14	78,7	14	74,90	0	1	0
1247	15	78,4	4	81,12	11	1	0
1468	16	78,2	1	84,64	15	1	0
398	17	77,4	22	72,62	5	0	0
78	18	76,7	31	70,69	13	0	0
250	19	76,6	43	67,80	24	0	0
261	20	76,2	12	75,59	8	1	0

Fonte: Elaborado pela autora.

Para o conjunto masculino, o CatBoost alcançou correlação de Spearman de 0,9123 e sobreposição de 65% (13 em comum) no Top 20, preservando todos os quatro medalhistas e todos os nove selecionados. O XGBoost apresentou correlação de 0,9076 e sobreposição de 60% (12 em comum), mantendo os quatro medalhistas e oito dos nove selecionados.

Esses resultados indicam que, mesmo com a redução no poder preditivo absoluto decorrente da remoção da variável temporal, os modelos mantiveram boa capacidade de identificar os escolares de maior potencial, com destaque para a preservação dos medalhistas em ambos os sexos.

### 5.3.5 Síntese do Experimento III

Os resultados obtidos no Experimento III mostram que a remoção da variável `Ano_Avaliação` resultou em uma redução adicional no poder preditivo dos modelos, com quedas de aproximadamente 3% no  $R^2$  em relação ao Experimento II. Essa redução confirma que parte do desempenho anterior estava associada a fatores correlacionados ao ano de coleta, incluindo diferenças na idade cronológica e no estágio de desenvolvimento dos escolares avaliados.

Apesar da queda no desempenho absoluto, os modelos mantiveram boa capacidade

Tabela 26 – Comparação Top 20 - CatBoost (Experimento III, Dados Masculinos)

Índice	Ranking Real	GS Real	Ranking Prev.	GS Prev.	$\Delta$ Ranking	Selec. Jogos Amizade	Med. Jogos Amizade
416	1	92,2	2	89,42	1	1	1
1431	2	89,7	18	77,66	16	0	0
952	3	89,1	28	74,28	25	0	0
324	4	88,5	9	80,87	5	0	0
135	5	88,0	5	84,37	0	1	1
1405	6	87,0	7	81,40	1	1	1
810	7	86,3	15	78,40	8	1	0
1000	7	86,3	3	87,45	4	1	0
1708	9	84,0	1	90,43	8	0	0
979	10	82,6	22	75,78	12	0	0
1124	11	81,9	14	78,48	3	0	0
422	12	81,7	29	74,28	17	0	0
581	13	80,6	38	71,73	25	1	0
203	14	80,4	35	72,50	21	0	0
807	15	79,7	16	78,39	1	1	0
772	16	79,6	112	59,75	96	0	0
507	17	78,8	20	77,36	3	1	1
298	18	78,5	23	75,59	5	0	0
817	19	78,3	11	79,72	8	0	0
1647	20	77,1	6	83,45	14	1	0

Fonte: Elaborado pela autora.

de ordenação, com correlações de Spearman superiores a 0,90 e preservação da maioria dos medalhistas e selecionados no Top 20. A análise de importância SHAP revelou que, na ausência da variável temporal, as medidas de desempenho físico e a competência percebida assumem o papel central nas previsões, com a idade cronológica ganhando relevância como fator complementar.

Esses resultados motivaram a realização do Experimento IV, no qual foram investigadas estratégias para reduzir a influência de diferenças etárias nas previsões, por meio da incorporação de medidas relativas de desempenho (percentis).

#### 5.4 Experimento IV: Inclusão de Percentis

O Experimento IV teve como objetivo investigar se a incorporação de medidas relativas de desempenho, na forma de percentis calculados por faixa etária, poderia melhorar a capacidade preditiva dos modelos e reduzir a dependência da idade cronológica. Para isso, foram adicionadas ao conjunto de variáveis os percentis correspondentes a cada teste físico, conforme descrito no Capítulo 4.

Tabela 27 – Comparação Top 20 - XGBoost (Experimento III, Dados Masculinos)

Índice	Ranking Real	GS Real	Ranking Prev.	GS Prev.	$\Delta$ Ranking	Selec. Jogos Amizade	Med. Jogos Amizade
416	1	92,2	1	90,91	0	1	1
1431	2	89,7	19	76,27	17	0	0
952	3	89,1	40	71,60	37	0	0
324	4	88,5	7	81,31	3	0	0
135	5	88,0	4	86,24	1	1	1
1405	6	87,0	14	78,52	8	1	1
810	7	86,3	18	76,85	11	1	0
1000	7	86,3	2	89,15	5	1	0
1708	9	84,0	3	87,54	6	0	0
979	10	82,6	37	71,87	27	0	0
1124	11	81,9	13	79,12	2	0	0
422	12	81,7	25	74,12	13	0	0
581	13	80,6	63	67,34	50	1	0
203	14	80,4	32	72,64	18	0	0
807	15	79,7	21	75,77	6	1	0
772	16	79,6	115	59,68	99	0	0
507	17	78,8	11	80,65	6	1	1
298	18	78,5	22	74,91	4	0	0
817	19	78,3	10	80,67	9	0	0
1647	20	77,1	6	84,01	14	1	0

Fonte: Elaborado pela autora.

#### 5.4.1 Desempenho Preditivo dos Modelos

Após a inclusão dos percentis, os modelos foram novamente treinados e avaliados. As Tabelas 28 e 29 apresentam os resultados para os conjuntos feminino e masculino.

Para o conjunto feminino, o modelo XGBoost apresentou o melhor desempenho, com  $R^2 = 0,8474$ ,  $MAE = 5,46$  e  $RMSE = 6,80$ , superando ligeiramente o CatBoost ( $R^2 = 0,8444$ ). Observa-se uma pequena recuperação no poder preditivo em relação ao Experimento III (de 0,8449 para 0,8474).

Para o conjunto masculino, o XGBoost também apresentou o melhor resultado, com  $R^2 = 0,8544$ ,  $MAE = 5,63$  e  $RMSE = 6,93$ , superando o CatBoost ( $R^2 = 0,8518$ ). Observa-se uma recuperação mais expressiva no poder preditivo em relação ao Experimento III (de 0,8422 para 0,8544), representando um ganho de aproximadamente 1,2%.

Destaca-se que, pela primeira vez, o XGBoost superou o CatBoost como melhor modelo em ambos os conjuntos de dados, sugerindo que a inclusão dos percentis pode ter beneficiado diferentemente os algoritmos.

Tabela 28 – Ranking de performance dos modelos no conjunto de teste - Experimento IV (Dados Femininos)

Modelo	$R^2$ Teste	MAE	RMSE
XGBoost	0,8474	5,4596	6,8035
CatBoost	0,8444	5,5932	6,8743
Gradient Boosting	0,8425	5,5897	6,9207
LightGBM	0,8347	5,7123	7,0853
SVR	0,8334	5,8093	7,1135
Ridge	0,8329	5,8161	7,1256
Linear Regression	0,8329	5,8225	7,1257
Lasso	0,8324	5,8349	7,1332
ElasticNet	0,8321	5,8415	7,1401
Random Forest	0,8131	6,2228	7,5329
Extra Tree	0,8086	6,1896	7,6192
Decision Tree	0,7042	7,7022	9,4747
KNN	0,6572	8,2254	10,1983

Fonte: Elaborado pela autora.

**Nota:** Para as métricas MAE e RMSE, valores menores indicam melhor desempenho preditivo ( $\downarrow$  erro). Para o  $R^2$ , valores maiores indicam melhor ajuste do modelo ( $\uparrow$  poder explicativo).

Tabela 29 – Ranking de performance dos modelos no conjunto de teste - Experimento IV (Dados Masculinos)

Modelo	$R^2$ Teste	MAE	RMSE
XGBoost	0,8544	5,6318	6,9287
CatBoost	0,8518	5,6288	6,9908
Gradient Boosting	0,8451	5,7556	7,1529
LightGBM	0,8386	5,8818	7,2941
SVR	0,8258	6,0015	7,5821
Ridge	0,8221	6,0675	7,6601
Lasso	0,8216	6,1083	7,6740
ElasticNet	0,8214	6,1115	7,6773
Linear Regression	0,8212	6,0780	7,6806
Random Forest	0,8092	6,3816	7,9324
Extra Tree	0,7957	6,7202	8,2079
Decision Tree	0,6919	8,0561	10,0829
KNN	0,6605	8,4722	10,5811

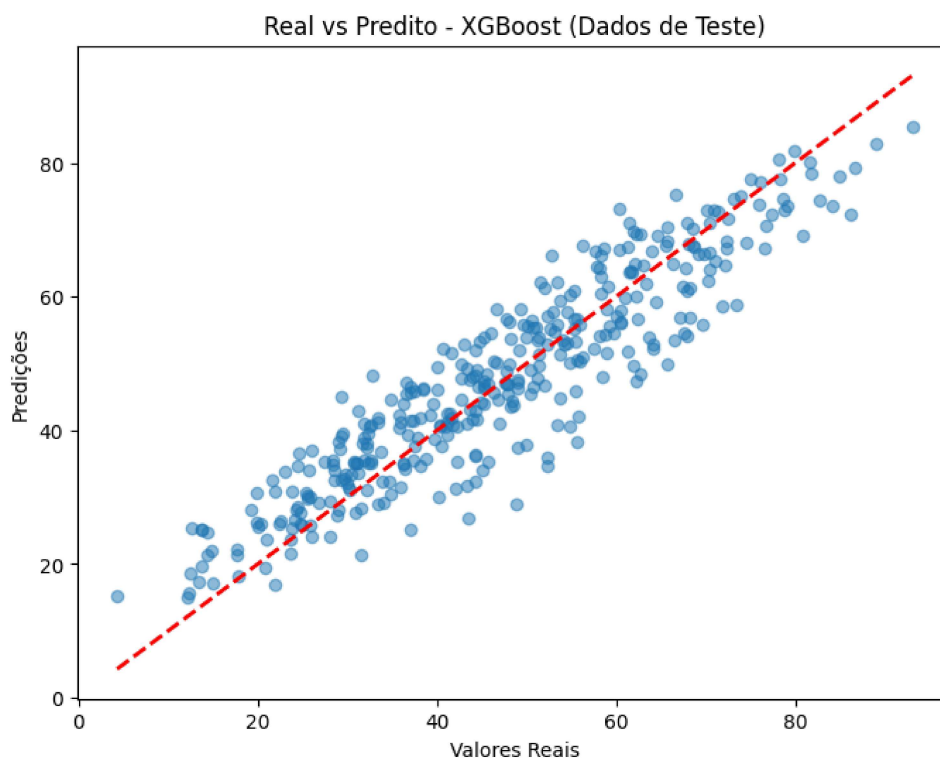
Fonte: Elaborado pela autora.

**Nota:** Para as métricas MAE e RMSE, valores menores indicam melhor desempenho preditivo ( $\downarrow$  erro). Para o  $R^2$ , valores maiores indicam melhor ajuste do modelo ( $\uparrow$  poder explicativo).

### 5.4.2 Análise Gráfica das Previsões

Considerando o desempenho superior obtido pelo XGBoost em ambos os conjuntos, as Figuras 44 e 46 apresentam os gráficos de dispersão para este modelo. Complementarmente, as Figuras 45 e 47 apresentam os resultados para o CatBoost, segundo melhor modelo.

Figura 44 – Gráfico de dispersão - XGBoost (Experimento IV, Dados Femininos)



Fonte: Elaborado pela autora.

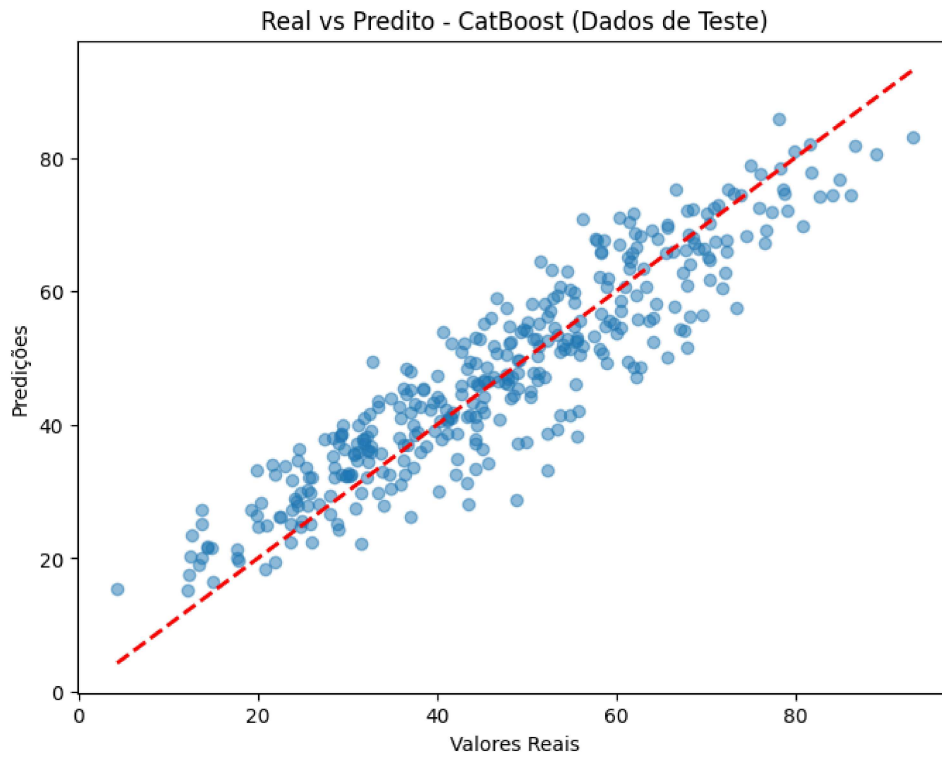
Observa-se que os pontos se concentram ao redor da linha de identidade, com padrão de dispersão semelhante ao observado no Experimento III, porém com ligeira melhora na precisão, consistente com o aumento nos valores de  $R^2$ .

### 5.4.3 Análise de Importância das Variáveis

Para compreender o impacto da inclusão dos percentis na importância das variáveis, foi realizada uma nova análise SHAP para os dois melhores algoritmos em cada conjunto. As Figuras 48 e 49 apresentam os resultados para os dados femininos, enquanto as Figuras 50 e 51 apresentam os resultados para os dados masculinos.

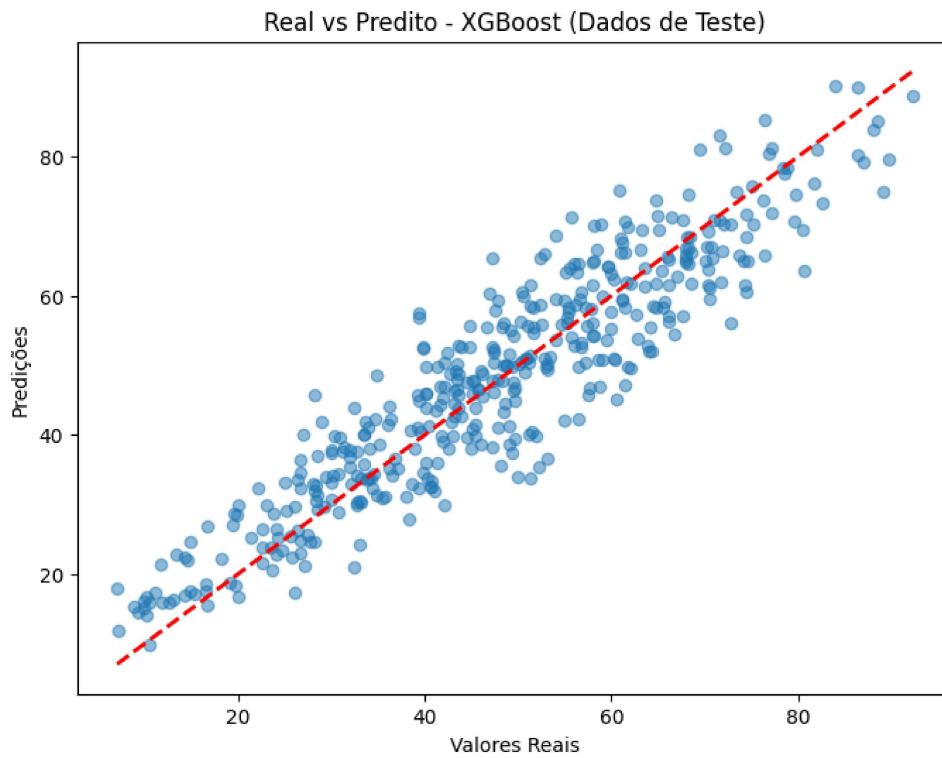
Observa-se que, com a inclusão dos percentis, as variáveis percentílicas passam a dominar as primeiras posições na importância SHAP em todos os modelos e conjuntos de dados. Destacam-se:

Figura 45 – Gráfico de dispersão - CatBoost (Experimento IV, Dados Femininos)



Fonte: Elaborado pela autora.

Figura 46 – Gráfico de dispersão - XGBoost (Experimento IV, Dados Masculinos)



Fonte: Elaborado pela autora.

Figura 47 – Gráfico de dispersão - CatBoost (Experimento IV, Dados Masculinos)

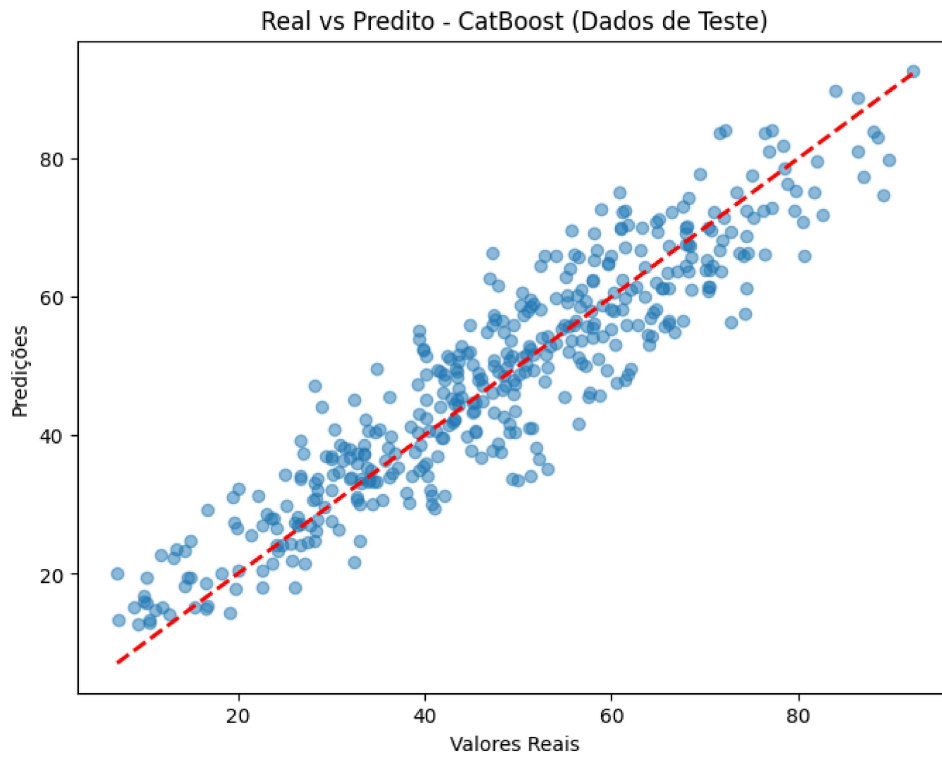


Figura 48 – Importância das variáveis segundo valores SHAP - XGBoost (Experimento IV, Dados Femininos)

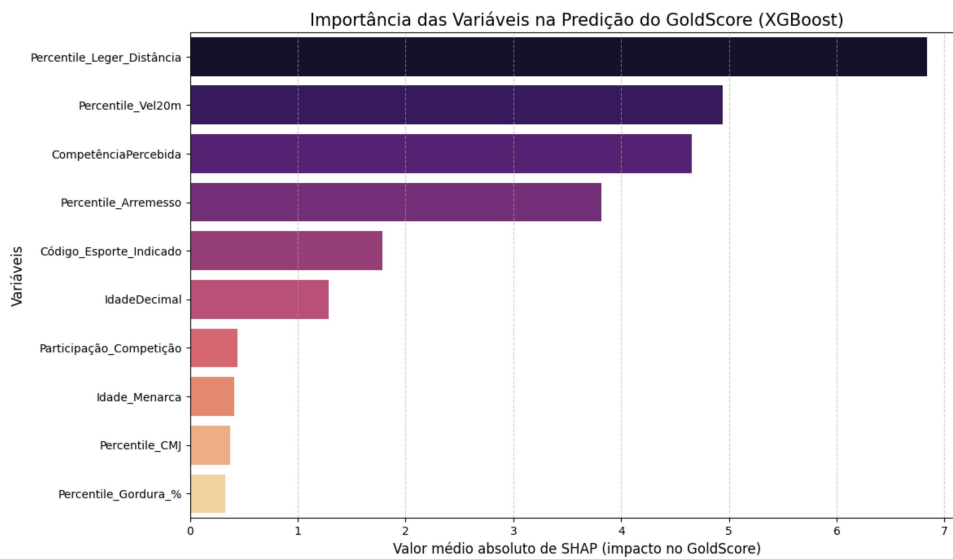
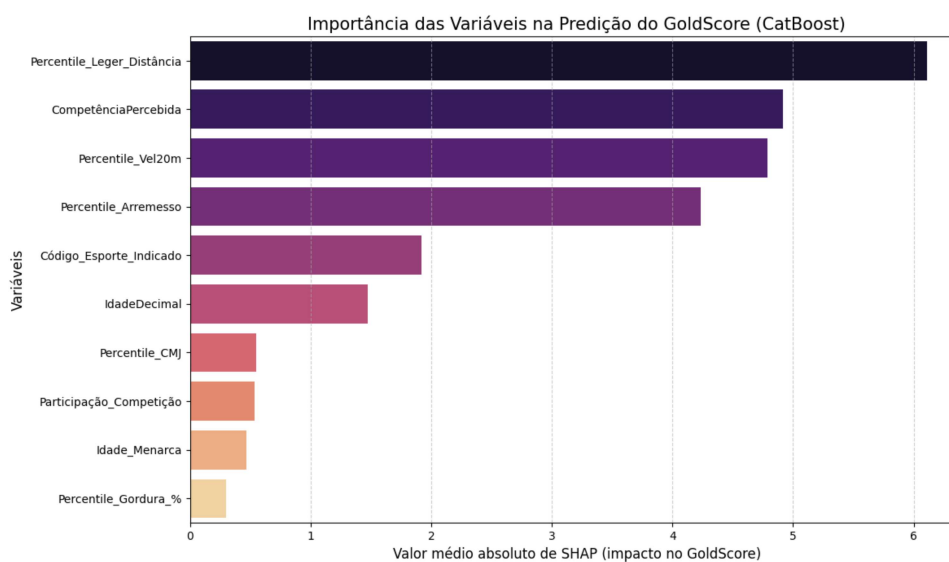
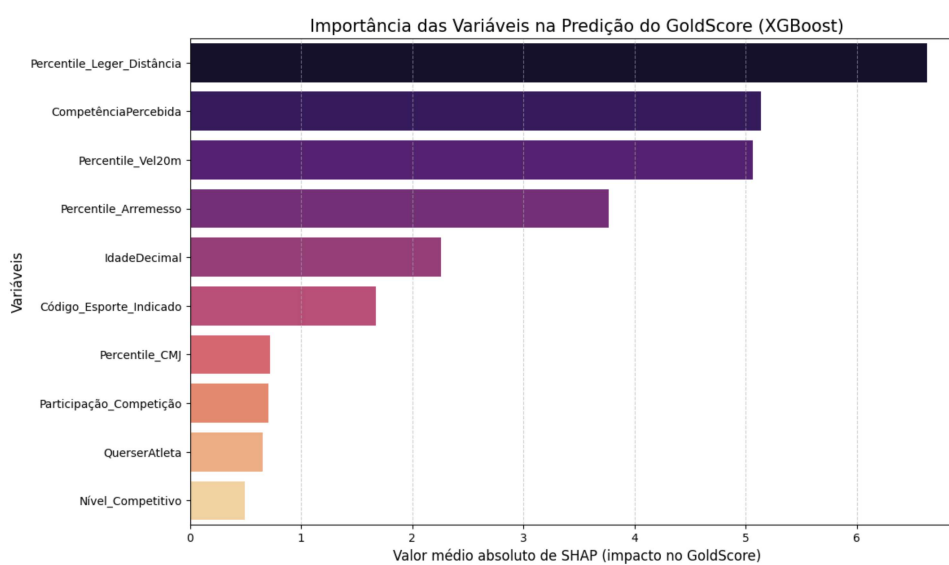


Figura 49 – Importância das variáveis segundo valores SHAP - CatBoost (Experimento IV, Dados Femininos)



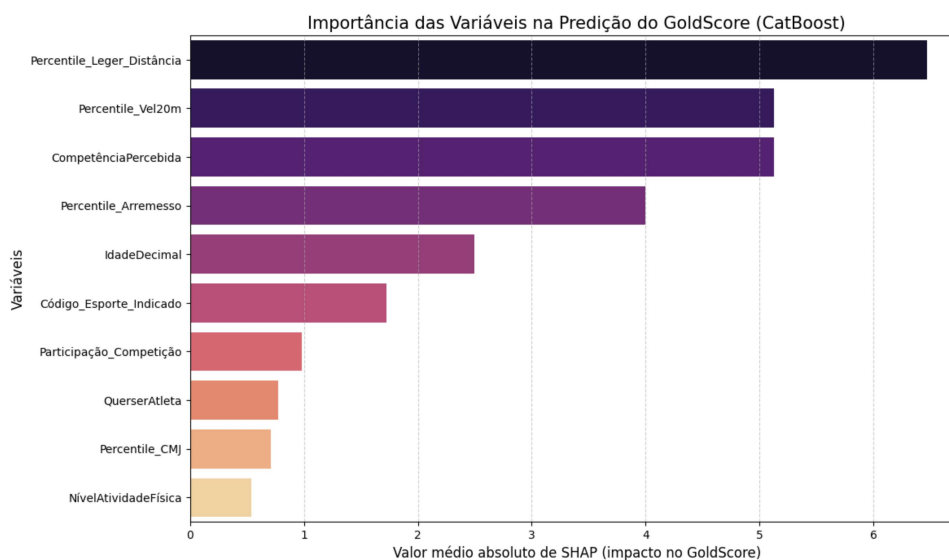
Fonte: Elaborado pela autora.

Figura 50 – Importância das variáveis segundo valores SHAP - XGBoost (Experimento IV, Dados Masculinos)



Fonte: Elaborado pela autora.

Figura 51 – Importância das variáveis segundo valores SHAP - CatBoost (Experimento IV, Dados Masculinos)



Fonte: Elaborado pela autora.

- **Leger\_Distância\_Percentil:** assume a primeira posição em três dos quatro modelos analisados, confirmando a importância da aptidão cardiorrespiratória na predição do *Gold Score*.
- **Vel20m\_Percentil e Arremesso\_Percentil:** aparecem consistentemente entre as cinco variáveis mais importantes em ambos os sexos.
- **CompetênciaPercebida:** mantém-se como uma variável relevante, especialmente no feminino, sugerindo que aspectos psicológicos continuam sendo importantes mesmo após a inclusão de medidas físicas relativas.
- **IdadeDecimal e FaixaEtária:** perdem posições relativas na importância, indicando que a inclusão dos percentis cumpriu seu objetivo de reduzir a dependência da idade cronológica nas predições.

A consistência desse padrão entre os modelos e sexos confirma que a transformação das medidas absolutas em percentis por faixa etária foi eficaz para destacar o desempenho relativo dos escolares, reduzindo a influência de diferenças maturacionais nas predições.

#### 5.4.4 Capacidade de Ordenação dos Escolares com Melhor *Gold Score*

Para avaliar a capacidade dos modelos em preservar as posições mais altas do ranking após a inclusão dos percentis, foi realizada uma comparação entre o ranking real e o ranking previsto para os 20 escolares com maiores valores de *Gold Score* no conjunto de teste. As Tabelas 30 a 33 apresentam essa comparação para os dois melhores modelos em cada conjunto.

Tabela 30 – Comparação Top 20 - CatBoost (Experimento IV, Dados Femininos)

Índice	Ranking Real	GS Real	Ranking Prev.	GS Prev.	$\Delta$ Ranking	Selec. Jogos Amizade	Med. Jogos Amizade
1211	1	93,1	2	83,10	1	1	0
427	2	89,1	6	80,54	4	0	0
1201	3	86,7	4	81,83	1	0	0
115	4	86,2	17	74,53	13	1	0
1161	5	85,0	11	76,87	6	0	0
1767	6	84,1	18	74,48	12	0	0
1769	7	82,7	20	74,22	13	0	0
237	8	81,8	9	77,88	1	1	0
1421	9	81,7	3	82,09	6	0	0
495	10	80,8	35	69,84	25	0	0
1476	11	79,9	5	81,09	6	0	0
637	12	79,1	25	72,20	13	0	0
1755	13	78,8	16	74,59	3	0	0
76	14	78,7	14	75,19	0	1	0
1247	15	78,4	8	78,50	7	1	0
1468	16	78,2	1	85,82	15	1	0
398	17	77,4	27	71,97	10	0	0
78	18	76,7	38	69,10	20	0	0
250	19	76,6	50	67,35	31	0	0
261	20	76,2	10	77,70	10	1	0

Fonte: Elaborado pela autora.

Para o conjunto feminino, o modelo CatBoost apresentou correlação de Spearman de 0,9164 e sobreposição de 75% (15 em comum) no Top 20, mantendo exatamente as sete estudantes selecionadas para os Jogos da Amizade. O XGBoost apresentou correlação de 0,9176 e sobreposição de 70% (14 em comum), também preservando as sete selecionadas.

Para o conjunto masculino, o CatBoost alcançou correlação de Spearman de 0,9177 e sobreposição de 70% (14 em comum) no Top 20, preservando todos os quatro medalhistas e todos os nove selecionados. O XGBoost apresentou correlação de 0,9242 e sobreposição de 70% (14 em comum), mantendo os quatro medalhistas e oito dos nove selecionados.

Esses resultados indicam que, com a inclusão dos percentis, os modelos mantiveram boa capacidade de identificar os escolares de maior potencial, com destaque para a preservação dos medalhistas em ambos os sexos e a recuperação parcial da sobreposição no Top 20 em relação ao Experimento III.

## 5.5 Síntese Comparativa dos Experimentos

A Tabela 34 apresenta uma visão consolidada do desempenho dos melhores modelos ao longo dos quatro experimentos, para ambos os sexos, considerando as principais métricas de avaliação.

Tabela 31 – Comparação Top 20 - XGBoost (Experimento IV, Dados Femininos)

Índice	Ranking Real	GS Real	Ranking Prev.	GS Prev.	$\Delta$ Ranking	Selec. Jogos Amizade	Med. Jogos Amizade
1211	1	93,1	1	85,50	0	1	0
427	2	89,1	2	82,97	0	0	0
1201	3	86,7	6	79,28	3	0	0
115	4	86,2	26	72,23	22	1	0
1161	5	85,0	8	77,94	3	0	0
1767	6	84,1	18	73,64	12	0	0
1769	7	82,7	16	74,51	9	0	0
237	8	81,8	7	78,49	1	1	0
1421	9	81,7	5	80,19	4	0	0
495	10	80,8	37	69,22	27	0	0
1476	11	79,9	3	81,89	8	0	0
637	12	79,1	19	73,61	7	0	0
1755	13	78,8	22	73,00	9	0	0
76	14	78,7	14	74,70	0	1	0
1247	15	78,4	9	77,59	6	1	0
1468	16	78,2	4	80,63	12	1	0
398	17	77,4	25	72,27	8	0	0
78	18	76,7	31	70,58	13	0	0
250	19	76,6	50	67,15	31	0	0
261	20	76,2	11	77,27	9	1	0

Fonte: Elaborado pela autora.

A Figura 52 apresenta a evolução do  $R^2$  do melhor modelo ao longo dos experimentos, permitindo visualizar o impacto dos experimentos.

Observa-se uma queda acentuada no desempenho entre o Experimento I e o Experimento II, decorrente da remoção das variáveis subjetivas do treinador. Essa redução era esperada, uma vez que tais variáveis apresentavam forte correlação com o *Gold Score* e foram deliberadamente suprimidas para testar a viabilidade de um modelo que não dependesse dessas avaliações especializadas.

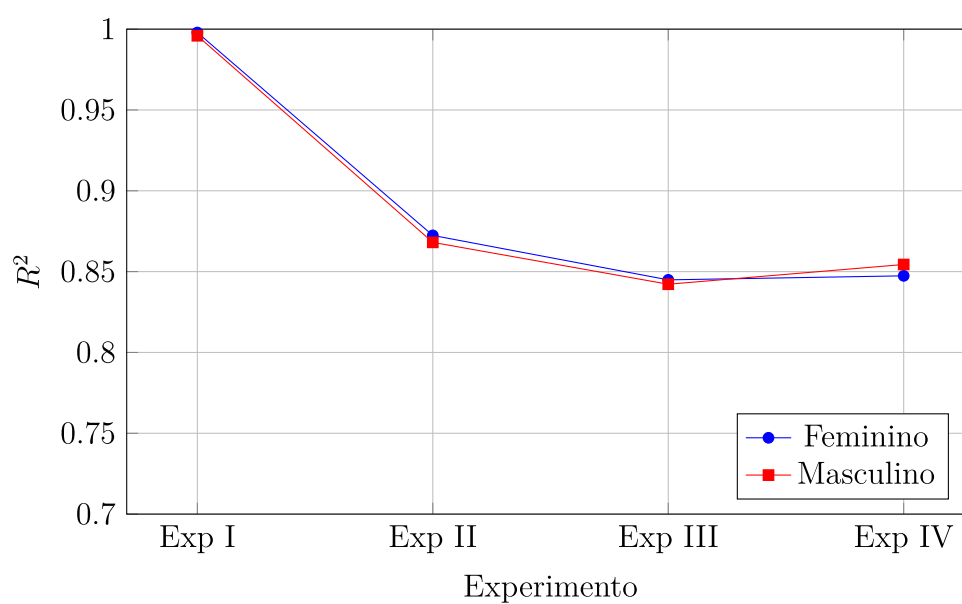
Entre os Experimentos II e III, observa-se uma redução adicional no  $R^2$  (aproximadamente 3%), associada à remoção da variável *Ano\_Avaliação*. Esta queda confirma que parte do desempenho anterior estava associada a fatores temporais e contextuais, e não exclusivamente às características individuais dos escolares.

No Experimento IV, observa-se uma recuperação parcial do desempenho para ambos os sexos, mais acentuada no masculino (de 0,8422 para 0,8544) e modesta no feminino (de 0,8449 para 0,8474). A inclusão dos percentis permitiu que o modelo capturasse informações relativas de desempenho, compensando parcialmente a perda de informação associada à remoção de variáveis temporais e subjetivas.

Tabela 32 – Comparação Top 20 - CatBoost (Experimento IV, Dados Masculinos)

Índice	Ranking Real	GS Real	Ranking Prev.	GS Prev.	$\Delta$ Ranking	Selec. Jogos Amizade	Med. Jogos Amizade
416	1	92,2	1	92,69	0	1	1
1431	2	89,7	13	79,87	11	0	0
952	3	89,1	24	74,61	21	0	0
324	4	88,5	9	83,00	5	0	0
135	5	88,0	6	83,79	1	1	1
1405	6	87,0	18	77,33	12	1	1
810	7	86,3	12	81,03	5	1	0
1000	7	86,3	3	88,77	4	1	0
1708	9	84,0	2	89,75	7	0	0
979	10	82,6	36	71,83	26	0	0
1124	11	81,9	14	79,67	3	0	0
422	12	81,7	23	75,05	11	0	0
581	13	80,6	70	66,01	57	1	0
203	14	80,4	40	70,91	26	0	0
807	15	79,7	20	75,33	5	1	0
772	16	79,6	30	72,53	14	0	0
507	17	78,8	19	76,34	2	1	1
298	18	78,5	15	78,66	3	0	0
817	19	78,3	10	81,81	9	0	0
1647	20	77,1	4	84,12	16	1	0

Fonte: Elaborado pela autora.

Figura 52 – Comparação do  $R^2$  do melhor modelo ao longo dos experimentos

Fonte: Elaborado pela autora.

Tabela 33 – Comparação Top 20 - XGBoost (Experimento IV, Dados Masculinos)

Índice	Ranking Real	GS Real	Ranking Prev.	GS Prev.	$\Delta$ Ranking	Selec. Jogos Amizade	Med. Jogos Amizade
416	1	92,2	3	88,72	2	1	1
1431	2	89,7	14	79,60	12	0	0
952	3	89,1	22	74,87	19	0	0
324	4	88,5	5	84,96	1	0	0
135	5	88,0	6	83,75	1	1	1
1405	6	87,0	15	79,14	9	1	1
810	7	86,3	13	80,10	6	1	0
1000	7	86,3	2	89,85	5	1	0
1708	9	84,0	1	90,12	8	0	0
979	10	82,6	28	73,36	18	0	0
1124	11	81,9	10	80,95	1	0	0
422	12	81,7	19	76,15	7	0	0
581	13	80,6	90	63,53	77	1	0
203	14	80,4	46	69,42	32	0	0
807	15	79,7	24	74,52	9	1	0
772	16	79,6	38	70,65	22	0	0
507	17	78,8	16	78,38	1	1	1
298	18	78,5	18	77,61	0	0	0
817	19	78,3	17	78,32	2	0	0
1647	20	77,1	9	81,16	11	1	0

Fonte: Elaborado pela autora.

Tabela 34 – Síntese comparativa dos experimentos - Melhores modelos

Experimento	Sexo	Melhor Modelo	$R^2$	Spearman	Sobreposição Top 20
Exp I	Feminino	CatBoost	0,9979	0,9989	100%
	Masculino	CatBoost	0,9959	0,9978	95%
Exp II	Feminino	CatBoost	0,8724	0,9311	80%
	Masculino	CatBoost	0,8681	0,9286	70%
Exp III	Feminino	CatBoost	0,8449	0,9168	70%
	Masculino	CatBoost	0,8422	0,9123	65%
Exp IV	Feminino	XGBoost	0,8474	0,9176	75%
	Masculino	XGBoost	0,8544	0,9242	70%

Fonte: Elaborado pela autora.

### 5.5.1 Análise da Capacidade de Ordenação

A Tabela 35 detalha a evolução da capacidade de ordenação dos dois melhores modelos ao longo dos experimentos, incluindo a informação sobre a preservação de medalhistas e selecionados.

Tabela 35 – Evolução da capacidade de ordenação dos dois melhores modelos ao longo dos experimentos

Exp.	Sexo	Modelo	Spearman	Sobreposição	Medalhistas	Selecionados
I	Feminino	CatBoost	0,9989	100%	0/0	7/7
		XGBoost	0,9960	95%	0/0	7/7
	Masculino	CatBoost	0,9978	95%	4/4	9/9
		Gradient Boosting	0,9952	90%	4/4	8/9
II	Feminino	CatBoost	0,9311	80%	0/0	8/7
		XGBoost	0,9322	80%	0/0	8/7
	Masculino	CatBoost	0,9286	70%	4/4	8/9
		XGBoost	0,9242	70%	4/4	8/9
III	Feminino	CatBoost	0,9168	70%	0/0	8/7
		XGBoost	0,9185	70%	0/0	7/7
	Masculino	CatBoost	0,9123	65%	4/4	9/9
		XGBoost	0,9076	60%	4/4	8/9
IV	Feminino	XGBoost	0,9176	70%	0/0	7/7
		CatBoost	0,9164	75%	0/0	7/7
	Masculino	XGBoost	0,9242	70%	4/4	8/9
		CatBoost	0,9177	70%	4/4	9/9

Fonte: Elaborado pela autora.

Observa-se que a capacidade de ordenação acompanha a tendência observada nas métricas de erro. No Experimento I, a ordenação é praticamente perfeita para o CatBoost em ambos os sexos, com os segundos colocados também apresentando desempenho excelente (Spearman  $> 0,99$  e sobreposição  $> 90\%$ ).

Após as remoções sucessivas de variáveis, há uma redução gradual na concordância dos rankings, embora os valores de correlação de Spearman permaneçam elevados (acima de 0,90 em todos os casos). Destacam-se os seguintes padrões:

- **Medalhistas:** Todos os modelos, em todos os experimentos, preservaram integralmente os quatro medalhistas do sexo masculino no Top 20 previsto, demonstrando robustez na identificação dos escolares de mais alto rendimento.
- **Selecionados femininos:** A preservação das sete selecionadas variou entre os modelos, com o XGBoost apresentando melhor desempenho nos Experimentos III e IV (mantendo exatamente as sete), enquanto o CatBoost tendeu a superestimar levemente (prevendo oito) nos Experimentos II e III.

- **Selecionados masculinos:** O CatBoost destacou-se por preservar todos os nove selecionados nos Experimentos I, III e IV, enquanto o XGBoost manteve oito dos nove na maioria dos cenários.
- **Experimento IV:** A inclusão dos percentis beneficiou especialmente o XGBoost no feminino (mantendo as sete selecionadas) e o CatBoost no masculino (mantendo os nove selecionados).

Destaca-se que, mesmo no cenário mais restritivo (Experimento III), os modelos são capazes de identificar 12 a 14 dos 20 escolares com maior potencial (60-70% de sobreposição), além de preservar a detecção de atletas com reconhecimento competitivo (medalhistas e selecionados para os Jogos da Amizade).

### 5.5.2 Análise de *overfitting*

Para avaliar a ocorrência de sobreajuste, foi calculada a diferença entre o  $R^2$  obtido no conjunto de treinamento (durante a validação cruzada) e o  $R^2$  no conjunto de teste ( $\Delta R^2 = R^2_{treino} - R^2_{teste}$ ). Valores próximos de zero indicam boa capacidade de generalização, enquanto diferenças elevadas sugerem que o modelo se ajustou excessivamente aos padrões específicos dos dados de treinamento.

A Tabela 36 apresenta essa comparação para os dois melhores modelos de cada experimento: CatBoost (melhor na maioria) e o respectivo segundo colocado (XGBoost ou Gradient Boosting).

Observa-se que todos os modelos apresentam diferenças reduzidas entre o desempenho em treino e teste ( $|\Delta R^2| < 0,025$  em todos os casos), indicando baixo nível de sobreajuste. No Experimento I, os valores negativos de  $\Delta R^2$  indicam desempenho ligeiramente superior no conjunto de teste, o que pode ocorrer devido à variabilidade amostral ou à natureza conservadora da validação cruzada.

Nos experimentos subsequentes, observa-se um padrão consistente de  $\Delta R^2$  positivos, porém de pequena magnitude, sugerindo que os modelos mantêm boa capacidade de generalização mesmo após a remoção de variáveis preditoras. Destaca-se que, no Experimento IV, o XGBoost apresentou os menores valores de  $\Delta R^2$  (0,0096 no feminino e 0,0036 no masculino), indicando excelente capacidade de generalização.

A consistência desses resultados entre os dois melhores modelos e em ambos os sexos reforça a robustez das modelagens realizadas, indicando que as estratégias de otimização de hiperparâmetros e validação cruzada foram eficazes para evitar *overfitting*.

Tabela 36 – Comparação do desempenho treino vs teste para os dois melhores modelos

Experimento	Sexo	Modelo	$R^2$ Treino (CV)	$R^2$ Teste	$\Delta R^2$ (Treino - Teste)
Experimento I	Feminino	CatBoost	0,9975	0,9979	-0,0004
		XGBoost	0,9922	0,9923	-0,0001
	Masculino	CatBoost	0,9945	0,9959	-0,0014
		Gradient Boosting	0,9885	0,9908	-0,0024
Experimento II	Feminino	CatBoost	0,8902	0,8724	+0,0178
		XGBoost	0,8848	0,8709	+0,0139
	Masculino	CatBoost	0,8788	0,8681	+0,0107
		XGBoost	0,8767	0,8616	+0,0151
Experimento III	Feminino	CatBoost	0,8649	0,8449	+0,0200
		XGBoost	0,8607	0,8431	+0,0176
	Masculino	CatBoost	0,8602	0,8422	+0,0180
		XGBoost	0,8583	0,8348	+0,0235
Experimento IV	Feminino	XGBoost	0,8570	0,8474	+0,0096
		CatBoost	0,8619	0,8444	+0,0175
	Masculino	XGBoost	0,8580	0,8544	+0,0036
		CatBoost	0,8596	0,8518	+0,0078

Fonte: Elaborado pela autora.

**Nota:**  $\Delta R^2$  positivo (+) indica desempenho superior no treino;  $\Delta R^2$  negativo (-) indica desempenho superior no teste. Todos os valores são inferiores a 0,025.

## 5.6 Validação Externa dos Escores Preditos

Além da avaliação do desempenho, foi conduzida uma análise para verificar se os escores gerados pelos modelos se comportam de forma semelhante ao *Gold Score* real. Seguindo o mesmo método do estudo original, comparamos os escores preditos com os escores reais em três aspectos:

1. **Estatísticas descritivas:** médias, desvios padrão e variabilidade dos escores;
2. **Comparação entre selecionados e não selecionados:** verificar se os alunos convocados para os Jogos da Amizade apresentam escores mais altos (no estudo original, esta análise é denominada *validade de construto*);
3. **Comparação entre medalhistas e não medalhistas:** verificar se os alunos que ganharam medalhas apresentam escores ainda mais altos (no estudo original, esta análise é denominada *validade de critério*).

### 5.6.1 Comparação das Estatísticas Descritivas

A Tabela 37 apresenta a comparação entre as estatísticas descritivas dos escores reais e preditos para os quatro experimentos, em ambos os sexos.

Tabela 37 – Comparação das estatísticas descritivas entre escores reais e preditos

Sexo	Exp.	Real		Predito	
		Média ± DP	CV (%)	Média ± DP	CV (%)
Feminino	I	46,98 ± 17,45	37,1	46,97 ± 17,34	36,9
	II	46,98 ± 17,45	37,1	47,31 ± 15,83	33,5
	III	46,98 ± 17,45	37,1	47,31 ± 15,83	33,5
	IV	46,98 ± 17,45	37,1	47,49 ± 15,50	32,6
Masculino	I	47,75 ± 18,18	38,1	47,75 ± 18,06	37,8
	II	47,75 ± 18,18	38,1	48,35 ± 16,95	35,1
	III	47,75 ± 18,18	38,1	48,32 ± 17,00	35,2
	IV	47,75 ± 18,18	38,1	48,13 ± 16,91	35,1

Fonte: Elaborado pela autora.

**Nota:** CV = Coeficiente de Variação (DP/Média × 100). Quanto menor o CV, mais homogêneos são os dados.

Observa-se que, para ambos os sexos, as médias dos escores preditos são muito próximas às médias reais, com diferenças inferiores a 0,5 pontos. O coeficiente de variação (CV) também se mantém consistente entre real e predito, indicando que o modelo preserva a variabilidade original dos dados. Para o sexo feminino, o CV real foi de 37,1%, enquanto os valores preditos variaram entre 32,6% e 36,9%. Para o sexo masculino, o CV real foi de 38,1%, com os preditos variando entre 35,1% e 37,8%.

### 5.6.2 Diferenciação de Seleccionados e Medalhistas

Para responder a esta questão, foram realizadas duas análises: (1) comparação entre os escores de alunos seleccionados e não seleccionados para os Jogos da Amizade; (2) comparação entre os escores de alunos medalhistas e não medalhistas. Em ambas, calculou-se o tamanho do efeito (Cohen's  $d$ ), adotando-se como referência:  $d = 0,20$  (efeito pequeno),  $d = 0,50$  (efeito médio),  $d = 0,80$  (efeito grande) e  $d > 1,00$  (efeito muito grande).

**Resultados para o sexo feminino:** Para o conjunto feminino, os resultados devem ser interpretados com cautela devido à presença de apenas uma atleta medalhista na amostra do conjunto de teste, o que inviabilizou a análise de comparação entre medalhistas e não medalhistas. A Tabela 38 apresenta os resultados da comparação entre seleccionadas e não seleccionadas para os quatro experimentos.

Observa-se que, em todos os experimentos, os escores preditos das alunas seleccionadas foram significativamente superiores aos das não seleccionadas ( $p < 0,001$ ), com tamanhos de efeito grandes ( $d$  entre 0,88 e 0,96), muito próximos ao observado para os escores reais ( $d = 0,95$ ). Este resultado indica que o modelo é capaz de diferenciar adequadamente as escolares seleccionadas para os Jogos da Amizade. A comparação entre

Tabela 38 – Comparação entre selecionadas e não selecionadas (dados femininos)

Exp.	Modelo	Selecionadas vs. Não Selecionadas			Medalhistas vs. Não Medalhistas		
		Selecionadas (média ± DP)	Não Sel. (média ± DP)	d	Medalhistas (média ± DP)	Não Med. (média ± DP)	d
I	Real	61,52 ± 16,35	45,55 ± 16,91	0,95	<i>Apenas 1 medalhista</i>		
	Predito	61,50 ± 16,30	45,54 ± 16,79	0,95	— análise inviável —		
II	Predito	60,68 ± 15,76	45,99 ± 15,24	0,96	— análise inviável —		
III	Predito	60,68 ± 15,76	45,99 ± 15,24	0,96	— análise inviável —		
IV	Predito	59,56 ± 13,89	46,30 ± 15,15	0,88	— análise inviável —		

Fonte: Elaborado pela autora.

**Nota:** d = tamanho do efeito (Cohen). Valores acima de 0,80 indicam efeito grande.

medalhistas e não medalhistas não pôde ser avaliada devido ao número insuficiente de medalhistas na amostra feminina ( $n = 1$ ).

**Resultados para o sexo masculino:** Para o conjunto masculino, a amostra continha número suficiente de medalhistas ( $n = 15$ ), permitindo a avaliação completa de ambas as comparações. A Tabela 39 apresenta os resultados.

Tabela 39 – Comparação entre selecionados e não selecionados, e entre medalhistas e não medalhistas (dados masculinos)

Exp.	Modelo	Selecionados vs. Não Selecionados			Medalhistas vs. Não Medalhistas		
		Selecionados (média ± DP)	Não Sel. (média ± DP)	d	Medalhistas (média ± DP)	Não Med. (média ± DP)	d
I	Real	62,86 ± 15,44	46,02 ± 17,68	0,96	67,35 ± 15,03	47,04 ± 17,90	1,14
	Predito	62,85 ± 15,44	46,02 ± 17,54	0,97	67,46 ± 14,87	47,03 ± 17,77	1,16
II	Predito	62,45 ± 13,72	46,74 ± 16,54	0,97	66,49 ± 12,98	47,69 ± 16,72	1,13
III	Predito	62,20 ± 14,12	46,73 ± 16,59	0,95	65,33 ± 12,87	47,70 ± 16,82	1,05
IV	Predito	61,43 ± 13,65	46,60 ± 16,59	0,91	64,49 ± 13,19	47,53 ± 16,74	1,02

Fonte: Elaborado pela autora.

**Nota:** d = tamanho do efeito (Cohen). Valores acima de 0,80 = efeito grande; acima de 1,00 = efeito muito grande.

Os escores preditos mantiveram elevada capacidade de diferenciar selecionados de não selecionados em todos os experimentos, com tamanhos de efeito variando entre  $d = 0,91$  e  $d = 0,97$  (real:  $d = 0,96$ ), todos classificados como *grandes*. Para a comparação entre medalhistas e não medalhistas, os escores preditos também foram capazes de diferenciar os grupos de forma consistente, com tamanhos de efeito variando entre  $d = 1,02$  e  $d = 1,16$  (real:  $d = 1,14$ ), todos classificados como *muito grandes*.

### 5.6.3 Síntese dos Principais Achados

Os resultados obtidos ao longo dos quatro experimentos permitem extrair as seguintes conclusões principais:

- **Superioridade dos métodos de boosting:** Algoritmos como CatBoost, XGBoost e Gradient Boosting apresentaram consistentemente os melhores resultados em todas as configurações experimentais, confirmando sua adequação para problemas com dados tabulares estruturados.
- **Papel central das variáveis subjetivas:** A remoção das avaliações do treinador (Experimento II) causou a maior queda no desempenho, evidenciando a importância dessas informações na predição do *Gold Score*. No entanto, os modelos mantiveram-se capazes de explicar cerca de 87% da variância mesmo sem essas variáveis.
- **Influência de fatores temporais:** A variável *Ano\_Avaliação* mostrou-se altamente relevante, atuando como proxy para diferenças etárias, maturacionais e contextuais entre as coortes. Sua remoção (Experimento III) resultou em queda adicional de aproximadamente 3% no  $R^2$ .
- **Contribuição dos percentis:** A inclusão de medidas relativas de desempenho (Experimento IV) permitiu recuperar parcialmente o poder preditivo, com ganhos mais expressivos no masculino. A análise SHAP confirmou que as variáveis percentílicas passaram a dominar as importâncias, reduzindo o peso da idade cronológica.
- **Capacidade de identificação de talentos:** Mesmo nos cenários mais restritivos, os modelos preservaram a capacidade de identificar a maioria dos escolares de alto potencial, com destaque para a manutenção de todos os medalhistas nos Top 20 previstos ao longo de todos os experimentos.
- **Robustez e generalização:** Os baixos valores de  $\Delta R^2$  entre treino e teste indicam que os modelos são robustos e generalizáveis, sem evidências de *overfitting* significativo.

## 6 CONCLUSÃO

Este estudo investigou a aplicação de técnicas de Aprendizado de Máquina para a predição do *Gold Score*, um índice multidimensional de potencial esportivo, utilizando dados de escolares coletados entre 2015 e 2023. A pesquisa foi estruturada em quatro experimentos progressivos, que permitiram avaliar o impacto da remoção de variáveis subjetivas, do ajuste de vieses temporais e da incorporação de medidas relativas de desempenho. **Com isso, todos os objetivos específicos foram alcançados:** (i) o pré-processamento e a análise exploratória dos dados foram realizados (Capítulo 4); (ii) modelos foram construídos e treinados em quatro cenários distintos; (iii) as estimativas foram comparadas aos valores reais e ao ranking Top 20; (iv) a análise SHAP identificou as variáveis mais importantes; (v) a ordenação foi avaliada por correlação de Spearman e sobreposição; e (vi) os resultados apontam viabilidade para uma ferramenta computacional de triagem de talentos.

### 6.1 Contribuições do Estudo

Este trabalho oferece contribuições teóricas e práticas:

- **Metodológica:** Propõe e valida um protocolo experimental para a predição de índices multidimensionais de potencial esportivo, incluindo estratégias para tratamento de dados ausentes (cópias), otimização de hiperparâmetros (PSO) e avaliação robusta (validação cruzada 10-fold com análise de overfitting). A abordagem em quatro experimentos progressivos, com remoção controlada de variáveis, constitui um modelo replicável para estudos semelhantes.
- **Analítica:** Demonstra a importância da investigação de vieses contextuais (como *Ano\_Avaliação*) e quantifica o impacto das variáveis subjetividade nas predições, evidenciando que modelos baseados apenas em métricas objetivas podem alcançar desempenho competitivo. A análise SHAP apresenta evidências interpretáveis sobre quais características físicas, motoras e psicossociais são mais determinantes na composição do índice.
- **Aplicada:** Fornece evidências de que modelos de AM podem operar de forma autônoma, sem depender de avaliações subjetivas, mantendo capacidade satisfatória de identificação dos atletas de maior potencial. Isso abre caminho para a implementação de sistemas escaláveis de triagem inicial de talentos em contextos escolares com acesso limitado a profissionais especializados.
- **Quantitativa:** Demonstra que a normalização por percentis etários é uma estratégia eficaz para reduzir o viés maturacional em modelos preditivos aplicados a populações

jovens em desenvolvimento, representando uma contribuição metodológica extensível a outros índices de avaliação esportiva juvenil.

- **Interpretabilidade sem fórmula explícita:** Os modelos de AM não produzem uma fórmula fechada como o *Gold Score* original. No entanto, as análises de importância das variáveis (valores SHAP) permitem identificar quais características mais influenciam o resultado — por exemplo, que o teste de Léger e a velocidade em 20m são mais determinantes do que outras variáveis. Assim, mesmo sem conhecer a “conta” interna do modelo, é possível compreender sua lógica de funcionamento, tornando-o transparente e auditável para profissionais da educação física.

## 6.2 Limitações do Estudo

Apesar dos resultados promissores, algumas limitações devem ser consideradas na interpretação dos resultados obtidos:

1. **Contexto institucional restrito:** Todos os dados foram coletados em uma única instituição (CMJF), com características socioeconômicas, culturais e de infraestrutura esportiva particulares. Isso traz uma limitação na generalização dos modelos para populações escolares em outros contextos.
2. **Ausência de validação prospectiva:** Os modelos foram avaliados com base em dados históricos de participação nos Jogos da Amizade. Não foi possível acompanhar prospectivamente a trajetória esportiva dos escolares identificados, o que seria necessário para confirmar a validade preditiva de longo prazo dos modelos.
3. **Natureza *cross-sectional* da avaliação:** A opção metodológica de tratar cada avaliação anual como observação independente, embora justificada pelo desenvolvimento maturacional, não considera a estrutura longitudinal dos dados. Modelos que considerem trajetórias individuais poderiam capturar dinâmicas de desenvolvimento não acessíveis nesta abordagem.
4. **Percentis calculados intra-amostra:** Os percentis utilizados no Experimento IV foram calculados com base na própria amostra do estudo, o que pode introduzir dependência estatística entre preditores e alvo. Tabelas normativas externas e validadas poderia trazer maior robustez ao trabalho.
5. **Lacuna entre 2019 e 2022:** A ausência de coletas nos anos de 2020 e 2021 devido à pandemia de COVID-19 pode ter introduzido descontinuidades não controladas no perfil das amostras dos anos subsequentes, potencialmente afetando a representatividade da série temporal.

### 6.3 Recomendações para Trabalhos Futuros

Com base nos resultados e limitações identificadas, sugerem-se as seguintes direções para pesquisas futuras:

1. **Expansão da base de dados:** Incluir dados de outras instituições e contextos socioeconômicos para avaliar a generalização dos modelos para populações mais diversas.
2. **Modelagem longitudinal:** Investigar abordagens que tratem adequadamente a estrutura longitudinal dos dados, como modelos mistos ou redes neurais recorrentes, que possam capturar trajetórias individuais de desenvolvimento.
3. **Validação externa com critérios esportivos:** Realizar estudos prospectivos que acompanhem a trajetória esportiva dos atletas identificados pelo modelo, verificando se as previsões se correlacionam com o sucesso esportivo futuro (por exemplo, medalhas, convocações e profissionalização).
4. **Análise aprofundada dos erros de classificação:** Investigar sistematicamente os casos de discordância entre o Top 20 real e o Top 20 estimado, buscando compreender:
  - Quais atletas presentes no ranking real não foram identificados pelo modelo (falsos negativos) e quais características os distinguem?
  - Quais atletas foram erroneamente classificados como Top 20 pelo modelo (falsos positivos) e que variáveis contribuíram para essa superestimação?

Esta análise pode revelar dimensões do potencial esportivo não capturadas adequadamente pelos preditores atuais.

5. **Modelagem da subjetividade do treinador com aprendizado de máquina:** Desenvolver estudos que utilizem algoritmos de ML para compreender e quantificar os fatores implícitos considerados pelos avaliadores na atribuição das notas subjetivas. A integração destes fatores aos modelos preditivos poderia reduzir a lacuna entre a estimativa algorítmica e o olhar especializado.
6. **Redução de ruído e refinamento temporal:** Desenvolver estratégias para mitigar o ruído introduzido por flutuações anuais nas avaliações e condições de teste.
7. **Percentis baseados em normas externas:** Desenvolver tabelas normativas para uma população maior, permitindo que os percentis sejam calculados com base em referências populacionais validadas, em vez de intra-amostra.
8. **Interpretabilidade:** Explorar técnicas adicionais de explicabilidade, como gráficos de dependência de SHAP e análise de interações entre variáveis, para compreender como cada preditor influencia as estimativas ao longo de toda a escala do *Gold Score*.

9. **Análise do efeito da pandemia:** Investigar se a descontinuidade introduzida pela ausência de coletas em 2020–2021 afetou o perfil das coortes subsequentes e, em caso positivo, desenvolver estratégias de ajuste para garantir a comparabilidade temporal das avaliações.
10. **Prática-operacional:** Os modelos desenvolvidos podem ser implementados como um programa computacional de fácil utilização. Nesse sistema, o usuário insere as variáveis de entrada (testes físicos, idade, competência percebida, etc.) e o programa retorna um único número — o *Gold Score* estimado.

#### 6.4 Considerações

Por fim, ressalta-se que ferramentas de AM aplicadas à identificação de talentos esportivos devem ser utilizadas como instrumentos de apoio à decisão, e não como substitutos da avaliação humana especializada. A combinação entre a eficiência computacional dos modelos e a sensibilidade contextual dos profissionais tende a produzir os melhores resultados, garantindo que aspectos éticos, sociais e individuais sejam adequadamente considerados no processo.

O desenvolvimento contínuo destas ferramentas deve ser acompanhado de reflexão crítica sobre seus limites e vieses, assegurando que contribuam para a democratização do acesso ao esporte e para a identificação equitativa de talentos em diferentes contextos socioeconômicos. Nesse sentido, a incorporação de critérios de equidade na modelagem constituem direções prioritárias para a evolução desta linha de pesquisa.

A presente dissertação representa uma contribuição para a interseção entre ciência do esporte e inteligência artificial, demonstrando que abordagens computacionais rigorosas podem gerar conhecimento aplicável à promoção do esporte escolar e à descoberta de novos talentos no Brasil.

## REFERÊNCIAS

- ABBOTT, A.; COLLINS, D. Eliminating the dichotomy between theory and practice in talent identification and development: Considering the role of psychology. **Journal of Sports Sciences**, Taylor & Francis, v. 20, n. 2, p. 157–178, 2002.
- ALPAYDIN, E. **Introduction to Machine Learning**. 2nd. ed. [S.l.]: MIT Press, 2010.
- BAILEY, R.; MORLEY, D. Towards a model of talent development in physical education. **Sport, Education and Society**, Routledge, v. 11, n. 3, p. 211–230, 2006.
- BAKER, J.; COBLEY, S.; SCHORER, J.; WATTIE, N. **Routledge handbook of talent identification and development in sport**. London: Routledge, 2017.
- BERGSTRA, J.; BENGIO, Y. Random search for hyper-parameter optimization. **Journal of Machine Learning Research**, v. 13, n. 2, p. 281–305, 2012.
- BOURDIEU, P. **O senso prático**. Petrópolis: Vozes, 1998.
- BREIMAN, L. Bagging predictors. **Machine learning**, v. 24, n. 2, p. 123–140, 1996.
- BREIMAN, L. Random forests. **Machine Learning**, Springer, v. 45, n. 1, p. 5–32, 2001.
- BREIMAN, L.; FRIEDMAN, J.; STONE, C. J.; OLSHEN, R. A. **Classification and Regression Trees**. [S.l.]: CRC press, 1984.
- BROWN, J. **Sports Talent: How to Identify and Develop Outstanding Athletes**. Champaign, IL: Human Kinetics, 2001. 312 p. ISBN 978-0736033909.
- BÖHME, M. T. S. O talento esportivo e o processo de treinamento a longo prazo. **Revista Brasileira de Ciência e Movimento**, 2000. Verifique o ano e título corretos desta autora para o seu contexto.
- CHAPELLE, O.; SCHÖLKOPF, B.; ZIEN, A. **Semi-Supervised Learning**. [S.l.]: MIT Press, 2006.
- CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. In: **Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining**. [S.l.: s.n.], 2016. p. 785–794.
- CHICCO, D.; WARRENS, M. J.; JURMAN, G. The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. **PeerJ Computer Science**, PeerJ Inc., v. 7, p. e623, 2021.
- CHRISTENSEN, M. K. “an eye for talent”: Talent identification and the “practical sense” of top-level soccer coaches. **Sociology of Sport Journal**, v. 26, n. 3, p. 365–382, 2009.
- CIACCIO, A. D.; GIORGI, G. Deep learning for supervised classification. **Rivista italiana di economia, demografia e statistica**, Vol. LXVIV, 07 2016.
- COBLEY, S.; BAKER, J.; SCHORER, J. Talent identification and development in sport: An introduction to a field of expanding research and practice. In: BAKER, J.; COBLEY, S.; SCHORER, J. (Ed.). **Talent Identification and Development in Sport: International Perspectives**. 2nd. ed. London: Routledge, 2020. p. 1–10. ISBN 9780367501983.

COBLEY, S.; SCHORER, J.; BAKER, J. Identification and development of sport talent: A brief introduction to a growing field of research and practice. In: BAKER, J.; COBLEY, S.; SCHORER, J. (Ed.). **Talent identification and development in sport: International perspectives**. 1st. ed. Abingdon: Routledge, 2012. p. 1–10. ISBN 9780415581615.

COHEN, J. **Statistical Power Analysis for the Behavioral Sciences**. 2. ed. Hillsdale, NJ: Lawrence Erlbaum Associates, 1992.

COVER, T. M.; HART, P. E. Nearest neighbor pattern classification. **IEEE Transactions on Information Theory**, IEEE, v. 13, n. 1, p. 21–27, 1967.

DOROGUSH, A. V.; ERSHOV, V.; GULIN, A. Catboost: gradient boosting with categorical features support. **arXiv preprint arXiv:1810.11363**, 2018. Disponível em: <https://arxiv.org/abs/1810.11363>. Acesso em: 04 dez. 2025.

DRUCKER, H.; BURGESS, C. J.; KAUFMAN, L.; SMOLA, A. J.; VAPNIK, V. N. Support vector regression machines. In: **Advances in Neural Information Processing Systems 9 (NIPS 1996)**. [S.l.]: MIT Press, 1996.

EFRON, B.; TIBSHIRANI, R. J. **An Introduction to the Bootstrap**. New York: Chapman and Hall/CRC, 1994.

ENGELBRECHT, A. P. **Computational intelligence: an introduction**. 2nd. ed. Chichester, England; Hoboken, NJ: John Wiley & Sons, 2007. ISBN 978-0-470-03561-0.

FABER, I. R.; SLOOT, L.; HOOGEVEEN, L.; ELFERINK-GEMSER, M. T.; SCHORER, J. Western approaches for the identification and development of talent in schools and sports contexts from 2009 to 2019—a literature review. **High Ability Studies**, Routledge, v. 33, n. 2, p. 135–168, 2022.

FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. C. P. L. F. d. **Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina**. 2. ed. Rio de Janeiro: LTC, 2021.

FIELD, A.; MILES, J.; FIELD, Z. **Discovering Statistics Using R**. London: Sage Publications, 2012.

FRIEDMAN, J. H. Greedy function approximation: a gradient boosting machine. **Annals of statistics**, JSTOR, p. 1189–1232, 2001.

FRIEDMAN, J. H. Stochastic gradient boosting. **Computational statistics & data analysis**, Elsevier, v. 38, n. 4, p. 367–378, 2002.

FURLEY, P.; WOOD, G. Working memory, attentional control, and expertise in sports: A review of current literature. **Journal of Applied Sport Psychology**, v. 22, n. 2, p. 207–223, 2010.

GARAY, A. L. de; LEVINE, L.; CARTER, J. E. L. Anthropometric and constitutional characteristics of top athletes. In: **Proceedings of the Olympic Scientific Congress**. Munich: International Council of Sport Science and Physical Education, 1974.

GAYA, A. **Projeto Esporte Brasil: manual de aplicação de medidas e testes, normas e critérios de avaliação**. Porto Alegre, 2002.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. Cambridge: MIT press, 2016.

HARTIGH, R. J. R. D.; NIESSEN, A. S. M.; FRENCKEN, W. G. P.; MEIJER, R. R. How psychological and behavioral characteristics influence talent development in sports. **Journal of Sports Sciences**, v. 36, n. 17, p. 2002–2011, 2018.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. 2nd. ed. [S.l.]: Springer Science & Business Media, 2009.

HOARE, D. G. Predicting success in junior elite basketball players—the contribution of anthropometric and physiological attributes. **Journal of Science and Medicine in Sport**, v. 1, n. 1, p. 1–12, 1995.

HOWELL, D. C. **Statistical Methods for Psychology**. 7. ed. Belmont: Wadsworth Cengage Learning, 2010.

ISSURIN, V. B. Evidence-based prerequisites and precursors of athletic talent: A review. **Sports Med**, v. 47, n. 10, p. 1993–2010, 2017.

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. **An Introduction to Statistical Learning**. [S.l.]: Springer, 2013.

JÚNIOR, D. B. R.; VIANNA, J. M.; OLIVEIRA, H. Z.; SILVA, R. C. P.; WERNECK, F. Z. Gold score basketball: a hybrid scientific talent identification model for male basketball. **Motricidade**, v. 17, n. 4, p. 346–358, 2021. Disponível em: <https://doi.org/10.6063/motricidade.22607>.

KE, G.; MENG, Q.; FINLEY, T.; WANG, T.; CHEN, W.; MA, W.; YE, Q.; LIU, T.-Y. Lightgbm: A highly efficient gradient boosting decision tree. In: **Advances in neural information processing systems**. [S.l.: s.n.], 2017. p. 3146–3154.

KENNEDY, J.; EBERHART, R. Particle swarm optimization. In: IEEE. **Proceedings of ICNN'95 - International Conference on Neural Networks**. [S.l.], 1995. v. 4, p. 1942–1948.

KEULEN, G. E. van; WERNECK, F. Z.; COELHO, E. F.; AGUIAR, C. M.; MIRANDA, L.; LIMA, J. R. D. P. Gold score athletics: Talent detection model for track and field. **Revista Brasileira de Medicina do Esporte**, SciELO Brasil, v. 30, n. 1, p. e2022\_0147, 2024.

KHAMIS, H. J.; ROCHE, A. F. Predicting adult stature without using skeletal age: The khamis-roche method. **Pediatrics**, American Academy of Pediatrics, v. 94, n. 4, p. 504–507, 1994. ISSN 0031-4005.

KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: **Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI)**. [S.l.]: Morgan Kaufmann, 1995. v. 2, p. 1137–1143.

LARKIN, P.; MESAGNO, C.; BERRY, J. Planning for the future: The role of talent identification in youth soccer. **International Journal of Sports Science & Coaching**, v. 14, p. xx–xx, 2019.

LATH, F.; KOOPMANN, T.; FABER, I. R.; BAKER, J. O.; SCHORER, J. Focusing on the coach's eye; towards a working model of coach decision-making in talent selection.

**Psychology of Sport and Exercise**, v. 56, p. 102011, 2021. Disponível em:

<https://doi.org/10.1016/j.psychsport.2021.102011>.

LLOYD, R. S.; OLIVER, J. L. The youth physical development model: A new approach to long-term athletic development. **Strength and Conditioning Journal**, v. 34, n. 3, p. 61–72, 2012.

LUGER, G. F. **Artificial Intelligence: Structures and Strategies for Complex Problem Solving**. 6th. ed. Boston: Addison-Wesley, 2008. ISBN 978-0321545893.

LUNDBERG, S. M.; LEE, S.-I. A unified approach to interpreting model predictions. **Advances in Neural Information Processing Systems (NIPS)**, Curran Associates, Inc., v. 30, p. 4765–4774, 2017. Disponível em: <https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf>.

MALINA, R. M.; BOUCHARD, C.; BAR-OR, O. **Growth, Maturation, and Physical Activity**. 2. ed. Champaign, IL: Human Kinetics, 2004.

MATSUDO, V. K. R. Talento esportivo e prescrição de exercício. **Revista Brasileira de Ciência e Movimento**, v. 10, n. 2, p. 7–18, 1996.

MELLO, J. B.; NAGORNY, G. A. K.; HAIACHI, M. d. C.; GAYA, A. R.; GAYA, A. C. A. Projeto esporte brasil: perfil da aptidão física relacionada ao desempenho esportivo de crianças e adolescentes. **Revista Brasileira de Cineantropometria e Desempenho Humano**, v. 18, n. 6, p. 658–666, nov 2016.

MIRANDA, L.; WERNECK, F. Z.; COELHO, E. F. et al. Talento motor e maturação biológica em escolares de um colégio militar. **Revista Brasileira de Medicina do Esporte**, v. 25, p. 372–378, 2019.

MIRANDA, L.; WERNECK, F. Z.; COELHO, E. F.; PRIETO-AYUSO, A.; LIMA, J. R. P. de. Valores normativos para avaliação do potencial esportivo de escolares de educação física. **Cuadernos de Educación y Desarrollo - QUALIS A4**, v. 16, n. 1, p. 3609–3639, 2024. Acesso em: 28 out. 2025. Disponível em: <https://ojs.cuadernoseducacion.com/ojs/index.php/ced/article/view/3277>.

MIRANDA, L. J. V. Pyswarms: a research toolkit for particle swarm optimization in python. **Journal of Open Source Software**, v. 3, n. 21, p. 433, 2018. Disponível em: <https://doi.org/10.21105/joss.00433>.

MITCHELL, T. M. **Machine Learning**. New York: McGraw-Hill, 1997. ISBN 978-0070428072.

MORETTIN, P. A.; BUSSAB, W. O. **Estatística Básica**. 9. ed. São Paulo: Saraiva, 2017.

NAEEM, M. S.; HASAN, M. M.; KHAN, M. A. W. N.; IQBAL, K. M. S. S.; UDDIN, M. S. An Unsupervised Machine Learning Algorithms: Comprehensive Review. **International Journal of Innovative Technology and Exploring Engineering (IJITEE)**, Blue Eyes Intelligence Engineering and Sciences Publication, v. 8, n. 9S3, p. 20–25, jul 2019.

PAPIĆ, V.; ROGULJ, N.; PLEŠTINA, V. Expert system for sports talent identification and orientation. In: **2nd International Conference on Computers and Devices for Communication**. [S.l.: s.n.], 2009.

PEARSON, K. Mathematical contributions to the theory of evolution. iii. regression, heredity and panmixia. **Philosophical Transactions of the Royal Society of London. Series A**, v. 187, p. 253–318, 1895.

PION, J. The flemish sports compass: From sports orientation to elite performance prediction. **Ghent University (PhD Thesis)**, Ghent, Belgium, 2015.

POLIKAR, R. Ensemble based systems in decision making. **IEEE circuits and systems magazine**, v. 6, n. 3, p. 21–45, 2006.

POOLE, D.; MACKWORTH, A.; GOEBEL, R. **Computational Intelligence: A Logical Approach**. New York: Oxford University Press, 1998. ISBN 9780195102703.

PRIETO-AYUSO, A.; PASTOR-VICEDO, J. C.; CONTRERAS-JORDÁN, O. R. Talent identification and development in soccer: a systematic review. **International Journal of Sports Science & Coaching**, v. 15, n. 2, p. 245–246, 2020.

QUINLAN, J. R. Induction of decision trees. **Machine Learning**, Springer, v. 1, n. 1, p. 81–106, 1986.

RIBEIRO, A. A. d. S. **Gold Score Judo: modelo de detecção de talentos para o judô**. Dissertação (Dissertação de Mestrado) — Universidade Federal de Juiz de Fora (UFJF), Juiz de Fora, 2022. Também publicado como artigo em 2024 no Cuadernos de Educación y Desarrollo.

ROBERTS, A. H.; GREENWOOD, D.; HUMBERSTONE, C.; RAYNOR, A. J. Pilot study on the reliability of the coach's eye: Identifying talent throughout a 4-day cadet judo camp. **Frontiers in Sports and Active Living**, Frontiers Media SA, v. 2, p. 596369, 2020.

ROBERTS, A. H.; GREENWOOD, D.; STANLEY, M.; HUMBERSTONE, C.; IREDALE, F.; RAYNOR, A. Coach knowledge in talent identification: A systematic review and meta-synthesis. **Sports Medicine**, v. 49, p. 1163–1183, 2019.

RUSSELL, S.; NORVIG, P. **Artificial Intelligence: A Modern Approach**. 4th. ed. [S.l.]: Pearson, 2020. ISBN 978-0134610993.

SHAPLEY, L. S. A value for n-person games. In: KUHN, H. W.; TUCKER, A. W. (Ed.). **Contributions to the Theory of Games II**. Princeton: Princeton University Press, 1953, (Annals of Mathematics Studies, v. 28). p. 307–317.

SHELDON, J. P.; ECCLES, J. S. Physical and psychological predictors of perceived ability in adult male and female tennis players. **Journal of Applied Sport Psychology**, Taylor & Francis, v. 17, n. 1, p. 48–63, 2005.

SHI, Y.; EBERHART, R. C. A modified particle swarm optimizer. In: **Proceedings of the IEEE International Conference on Evolutionary Computation**. [S.l.]: IEEE, 1998. p. 69–73.

- SKLAR, A. Fonctions de répartition à n dimensions et leurs marges. **Publications de l'Institut de Statistique de l'Université de Paris**, Université de Paris, v. 8, p. 229–231, 1959.
- SLAUGHTER, M. H.; LOHMAN, T. G.; BOILEAU, R. A.; HORSWILL, C. A.; STILLMAN, R. J.; LOAN, M. D. V.; BEMBEN, D. A. Skinfold equations for estimation of body fatness in children and youth. **Human Biology**, Wayne State University Press, v. 60, n. 5, p. 709–723, 1988.
- SMOLA, A. J.; SCHÖLKOPF, B. A tutorial on support vector regression. **Statistics and Computing**, Springer, v. 14, n. 3, p. 199–222, 2004.
- SPEARMAN, C. The proof and measurement of association between two things. **The American Journal of Psychology**, v. 15, n. 1, p. 72–101, 1904.
- SUTTON, R. S.; BARTO, A. G. **Reinforcement Learning: An Introduction**. 2nd ed. [S.l.]: MIT Press, 2018.
- TANG, Q.; WEI, X.; TAN, B. The role of machine learning in talent identification for team sports: A systematic review. **Journal of Sports Science and Medicine**, JSSM, v. 25, n. 1, p. 58–83, 2026.
- VAPNIK, V. N. **The Nature of Statistical Learning Theory**. [S.l.]: Springer-Verlag, 1995.
- WERNECK, F.; COELHO, E.; MIRANDA, L. Projeto atletas de ouro®: Uma inovação na detecção de talentos esportivos. **E-Legis - Revista Eletrônica do Programa de Pós-Graduação da Câmara dos Deputados**, v. 15, p. 1–16, 03 2022.
- WERNECK, F. Z.; MIRANDA, L.; COELHO, E. F.; LIMA, J. R. P. d. Índice de potencial esportivo: uma abordagem científica para detecção de talentos na escola. **Revista Educação Especial**, v. 37, n. 1, p. e19/1–26, 2024. Acesso em: 28 out. 2025. Disponível em: <https://periodicos.ufsm.br/educacaoespecial/article/view/85079>.
- WILLIAMS, A. M.; REILLY, T. Talent identification and development in soccer. **Journal of sports sciences**, Taylor & Francis, v. 18, n. 9, p. 703–715, 2000.
- XIANG, C.; KAMALDEN, T. F. T.; LIU, H.; ISMAIL, N. Exploring the multidisciplinary factors affecting sports talent identification. **Frontiers in Psychology**, v. 13, p. 948121, 2022.
- ZHU, X. **Semi-supervised learning literature survey**. [S.l.], 2005.

## APÊNDICE A – DICIONÁRIO DE VARIÁVEIS DO ESTUDO

Tabela 40 – Dicionário de Variáveis do Projeto Atletas de Ouro

Variável	Tipo de Dado	Descrição
<b>Geral e Demográfica</b>		
Codigo	Numérico	Código único de identificação do escolar no projeto.
Ano_Avaliação	Numérico	Ano em que a avaliação do escolar foi realizada.
DataNascimento	Data	Data de nascimento do escolar.
Sexo	Categórica	Sexo do escolar (Masculino/Feminino).
QuartilNascimento	Categórica	Quartil do ano de nascimento.
SemestreNascimento	Categórica	Semestre do ano de nascimento (1º ou 2º).
IdadeDecimal	Numérico	Idade do escolar em anos com precisão decimal.
FaixaEtária	Categórica	Faixa etária categórica.
Repetência	Binária	Indica se o escolar já foi reprovado.
Dominância_Mãos	Categórica	Mão dominante.
Dominância_Pés	Categórica	Pé dominante.
<b>Prática Esportiva e Motivação</b>		
NívelAtividadeFísica	Numérico	Nível de atividade física geral.
P_NívelAtvFisica	Numérico	Percentil do Nível de Atividade Física.
Treinamento	Numérico	Envolvimento em treinamento esportivo formal.
P_Treinamento	Numérico	Percentil do Nível de Treinamento.
Idade_Inicio	Numérico	Idade de início da prática esportiva organizada.
TempoPrática_Idade	Numérico	Tempo total de prática esportiva em anos.
Esporte_Preferido	Categórica	Esporte de preferência.
Participação_Competição	Binária	Participação em competições.
Nível_Competitivo	Numérico	Nível de competição.
P_NívelCompetitivo	Numérico	Percentil do Nível Competitivo.
QuerserAtleta	Binária	Desejo de se tornar atleta profissional.
PossuiAtleta_Familia	Binária	Atleta na família.
ApoioFamilia	Numérico	Nível de apoio familiar.

*Continua na próxima página*

Tabela 40 – continuação

Variável	Tipo de Dado	Descrição
P_ApoioFamilia	Numérico	Percentil do Apoio Familiar.
<b>Antropometria e Composição Corporal</b>		
MassaCorporal	Numérico	Peso corporal (kg).
P_MassaCorporal	Numérico	Percentil de Massa Corporal.
Estatura	Numérico	Altura (cm).
P_Estatura	Numérico	Percentil de Estatura.
IMC	Numérico	Índice de Massa Corporal (kg/m <sup>2</sup> ).
AlturaSentado	Numérico	Altura sentado (cm).
P_AlturaSentado	Numérico	Percentil de Altura Sentado.
Envergadura	Numérico	Envergadura (cm).
P_Envergadura	Numérico	Percentil de Envergadura.
ComprimentoMMII	Numérico	Comprimento dos Membros Inferiores (cm).
P_ComprimentoMMII	Numérico	Percentil de Comprimento dos Membros Inferiores.
SomaDC	Numérico	Soma das dobras cutâneas.
Gordura_%	Numérico	Percentual de Gordura Corporal.
P_Gordura	Numérico	Percentil de Gordura Corporal.
<b>Potencial Esportivo e Indicação</b>		
PotencialTreinador	Numérico	Avaliação do potencial pelo treinador.
AltoPotencial_Treinador	Binária	Classificação binária de Alto Potencial.
P_PotencialTreinador	Numérico	Percentil de Potencial avaliado pelo Treinador.
Esporte_Indicado	Catagórica	Modalidade esportiva indicada.
<b>Habilidades Psicológicas e Intangíveis</b>		
Intangíveis	Numérico	Atributos psicológicos gerais.
P_Intangiveis	Numérico	Percentil de Intangíveis.
AutoAvaliacaoPotenciais	Numérico	Autoavaliação do potencial esportivo.
CompetênciaPercebida	Numérico	Percepção da própria competência.
P_CompetenciaPercebida	Numérico	Percentil de Competência Percebida.
Competitivo	Numérico	Traço de competitividade.
P_Competitivo	Numérico	Percentil de Competitividade.
Vencedor	Numérico	Mentalidade de vencedor.

*Continua na próxima página*

Tabela 40 – continuação

Variável	Tipo de Dado	Descrição
P_Vencedor	Numérico	Percentil da Escala Vencedor.
Determinado	Numérico	Nível de determinação.
P_Determinado	Numérico	Percentil da Escala Determinado.
LidarAdversidades	Numérico	Capacidade de lidar com adversidades.
P_LidarAdversidades	Numérico	Percentil de Lidar com Adversidades.
DesempenhoPressão	Numérico	Desempenho sob pressão.
P_DesempenhoPressao	Numérico	Percentil de Desempenho sob Pressão.
MetasPreparoMental	Numérico	Capacidade de definir metas.
P_MetasPreparoMental	Numérico	Percentil de Metas e Preparo Mental.
Concentração	Numérico	Nível de concentração.
P_Concentracao	Numérico	Percentil de Concentração.
LivrePreocupação	Numérico	Capacidade de se manter livre de preocupações.
P_LivrepreReocupacao	Numérico	Percentil de Livre de Preocupação.
ConfiançaMotivação	Numérico	Confiança e motivação intrínseca.
P_ConfiancaMotivacao	Numérico	Percentil de Confiança e Motivação.
Treinabilidade	Numérico	Capacidade de assimilação do treinamento.
P_Treinabilidade	Numérico	Percentil de Treinabilidade.
Coping	Numérico	Estratégias de enfrentamento do estresse.
P_Coping	Numérico	Percentil de Coping.
<b>Aptidão Física e Desempenho Motor</b>		
Flexibilidade	Numérico	Teste de flexibilidade (cm).
P_Flexibilidade	Numérico	Percentil de Flexibilidade.
Handgrip	Numérico	Força de prensão manual (kg).
P_Handgrip	Numérico	Percentil de Força de Prensão Manual.
CMJ	Numérico	Altura do Salto Vertical (cm).
P_CMJ	Numérico	Percentil de Salto Vertical.
Arremesso	Numérico	Distância do Arremesso (m).
P_Arremesso	Numérico	Percentil de Arremesso.
Vel10m	Numérico	Tempo de corrida em 10m (s).
P_Vel10m	Numérico	Percentil de Velocidade (10m).
Vel20m	Numérico	Tempo de corrida em 20m (s).
P_Vel20m	Numérico	Percentil de Velocidade (20m).
Leger_Distância	Numérico	Distância no Teste de Léger (m).

*Continua na próxima página*

Tabela 40 – continuação

Variável	Tipo de Dado	Descrição
P_LegerDistancia	Numérico	Percentil da Distância no Teste de Léger.
<b>Maturação e Projeção de Crescimento</b>		
EstaturaAdultaPrevista	Numérico	Projeção da estatura final adulta (cm).
EAP	Numérico	Estatura Adulta Prevista.
EAP_atingida	Numérico	Estatura Adulta Atingida.
EscoreZ_EAPatingida	Numérico	Escore Z da Estatura Adulta Atingida.
Classif_StatusMaturação	Catagórica	Estágio de maturação biológica.
MO_Moore	Numérico	Offset Maturacional (Método de Moore).
IdadePVC_Moore	Numérico	Idade do Pico de Velocidade de Crescimento.
Menarca	Binária	Indica se a menina já teve menarca.
Idade_Menarca	Numérico	Idade da menarca.
<b>Índices de Potencial Esportivo (Gold Score)</b>		
GoldScore_GS	Numérico	Índice de Potencial Esportivo Geral.
GS_Velocistas	Numérico	Índice para modalidades de Velocidade.
GS_Arremessadores	Numérico	Índice para modalidades de Força e Arremesso.
GS_Fundistas	Numérico	Índice para modalidades de Resistência.
GS_Judo	Numérico	Índice para a modalidade Judô.
<b>Resultados Competitivos</b>		
SeleçãoJogosdaAmizade	Binária	Selecionado para os Jogos da Amizade.
MedalhistaJogosAmizade	Binária	Medalhista nos Jogos da Amizade.
Nível_VITÓRIA	Numérico	Nível de sucesso em competições.

**Fonte:** Elaborado pela autora

## APÊNDICE B – Top 20 - Experimento II

### B.1 Resultados para o Sexo Feminino

Tabela 41 – Comparação entre ranking real e ranking previsto pelo CatBoost para os 20 maiores valores de Gold Score no conjunto de teste (Experimento II, Dados Femininos)

Índice	Ranking Real	GS Real	Ranking Prev.	GS Prev.	$\Delta$ Ranking	Selecionado Jogos da Amizade	Medalhista Jogos da Amizade
1211	1	93.1	2	87.27	1	1	0
427	2	89.1	1	87.82	1	0	0
1201	3	86.7	5	84.21	2	0	0
115	4	86.2	11	77.07	7	1	0
1161	5	85.0	6	79.60	1	0	0
1767	6	84.1	19	74.72	13	0	0
1769	7	82.7	20	74.44	13	0	0
237	8	81.8	9	77.29	1	1	0
1421	9	81.7	8	77.51	1	0	0
495	10	80.8	23	73.59	13	0	0
1476	11	79.9	18	74.76	7	0	0
637	12	79.1	15	75.97	3	0	0
1755	13	78.8	36	70.14	23	0	0
76	14	78.7	10	77.09	4	1	0
1247	15	78.4	3	85.39	12	1	0
1468	16	78.2	4	84.37	12	1	0
398	17	77.4	14	76.15	3	0	0
78	18	76.7	21	73.96	3	0	0
250	19	76.6	61	64.46	42	0	0
261	20	76.2	12	76.87	8	1	0

Fonte: Elaborado pela autora

### B.2 Resultados para o Sexo Masculino

Tabela 42 – Comparação entre ranking real e ranking previsto pelo XGBoost para os 20 maiores valores de Gold Score no conjunto de teste (Experimento II, Dados Femininos)

Índice	Ranking Real	GS Real	Ranking Prev.	GS Prev.	$\Delta$ Ranking	Selecionado Jogos da Amizade	Medalhista Jogos da Amizade
1211	1	93.1	2	86.62	1	1	0
427	2	89.1	1	88.14	1	0	0
1201	3	86.7	5	83.18	2	0	0
115	4	86.2	11	77.36	7	1	0
1161	5	85.0	7	78.54	2	0	0
1767	6	84.1	20	73.74	14	0	0
1769	7	82.7	13	76.43	6	0	0
237	8	81.8	6	80.71	2	1	0
1421	9	81.7	12	77.25	3	0	0
495	10	80.8	22	72.82	12	0	0
1476	11	79.9	8	78.35	3	0	0
637	12	79.1	15	75.14	3	0	0
1755	13	78.8	46	68.03	33	0	0
76	14	78.7	16	75.05	2	1	0
1247	15	78.4	3	85.32	12	1	0
1468	16	78.2	4	83.26	12	1	0
398	17	77.4	17	74.84	0	0	0
78	18	76.7	29	70.82	11	0	0
250	19	76.6	60	64.55	41	0	0
261	20	76.2	14	75.96	6	1	0

Fonte: Elaborado pela autora

Tabela 43 – Comparação entre ranking real e ranking previsto pelo CatBoost para os 20 maiores valores de Gold Score no conjunto de teste (Experimento II, Dados Masculinos)

Índice	Ranking Real	GS Real	Ranking Prev.	GS Prev.	$\Delta$ Ranking	Selecionado Jogos da Amizade	Medalhista Jogos da Amizade
416	1	92.2	1	90.45	0	1	1
1431	2	89.7	16	79.06	14	0	0
952	3	89.1	22	74.77	19	0	0
324	4	88.5	10	81.02	6	0	0
135	5	88.0	4	85.29	1	1	1
1405	6	87.0	9	81.86	3	1	1
810	7	86.3	11	81.02	4	1	0
1000	7	86.3	3	85.33	4	1	0
1708	9	84.0	6	85.10	3	0	0
979	10	82.6	26	74.30	16	0	0
1124	11	81.9	18	77.06	7	0	0
422	12	81.7	33	72.70	21	0	0
581	13	80.6	29	73.85	16	1	0
203	14	80.4	28	74.04	14	0	0
807	15	79.7	12	80.73	3	1	0
772	16	79.6	86	63.91	70	0	0
507	17	78.8	8	82.14	9	1	1
298	18	78.5	20	75.29	2	0	0
817	19	78.3	7	83.08	12	0	0
1647	20	77.1	14	79.14	6	1	0

Fonte: Elaborado pela autora

Tabela 44 – Comparação entre ranking real e ranking previsto pelo XGBoost para os 20 maiores valores de Gold Score no conjunto de teste (Experimento II, Dados Masculinos)

Índice	Ranking Real	GS Real	Ranking Prev.	GS Prev.	$\Delta$ Rank	Selecionado Jogos da Amizade	Medalhista Jogos da Amizade
416	1	92.2	1	91.36	0	1	1
1431	2	89.7	14	79.77	12	0	0
952	3	89.1	22	74.47	19	0	0
324	4	88.5	7	83.49	3	0	0
135	5	88.0	2	88.13	3	1	1
1405	6	87.0	12	80.86	6	1	1
810	7	86.3	11	81.34	4	1	0
1000	7	86.3	4	85.93	3	1	0
1708	9	84.0	5	85.91	4	0	0
979	10	82.6	33	71.71	23	0	0
1124	11	81.9	18	78.06	7	0	0
422	12	81.7	32	71.75	20	0	0
581	13	80.6	35	71.58	22	1	0
203	14	80.4	28	72.39	14	0	0
807	15	79.7	15	79.73	0	1	0
772	16	79.6	88	63.24	72	0	0
507	17	78.8	9	82.33	8	1	1
298	18	78.5	19	75.00	1	0	0
817	19	78.3	10	81.39	9	0	0
1647	20	77.1	17	78.15	3	1	0

Fonte: Elaborado pela autora

## APÊNDICE C – RESULTADOS COMPLETOS DOS EXPERIMENTOS

Tabela 45 – Resultados completos de todos os modelos por experimento e sexo

Exp.	Sexo	Modelo	$R^2$ Treino	$R^2$ Teste	$\Delta R^2$	MAE	RMSE
I	Fem.	CatBoost	0,9975	0,9979	-0,0004	0,6089	0,8002
I	Fem.	XGBoost	0,9922	0,9923	-0,0001	1,1667	1,5336
I	Fem.	Gradient Boosting	0,9885	0,9902	-0,0017	1,3517	1,7240
I	Fem.	LightGBM	0,9866	0,9887	-0,0021	1,4120	1,8505
I	Fem.	Linear Regression	0,9637	0,9681	-0,0044	2,4905	3,1109
I	Fem.	SVR	0,9629	0,9681	-0,0051	2,4830	3,1130
I	Fem.	Ridge	0,9638	0,9679	-0,0042	2,4973	3,1196
I	Fem.	ElasticNet	0,9639	0,9672	-0,0033	2,5158	3,1566
I	Fem.	Lasso	0,9640	0,9670	-0,0030	2,5215	3,1634
I	Fem.	Extra Tree	0,9511	0,9491	+0,0020	3,0376	3,9297
I	Fem.	Random Forest	0,9447	0,9472	-0,0025	3,1572	4,0050
I	Fem.	Decision Tree	0,8583	0,8498	+0,0085	5,4066	6,7539
I	Fem.	KNN	0,7706	0,7450	+0,0256	6,9235	8,7995
I	Masc.	CatBoost	0,9945	0,9959	-0,0014	0,8809	1,1661
I	Masc.	Gradient Boosting	0,9885	0,9908	-0,0024	1,3590	1,7377
I	Masc.	XGBoost	0,9884	0,9892	-0,0008	1,5015	1,8867
I	Masc.	LightGBM	0,9851	0,9889	-0,0038	1,4885	1,9180
I	Masc.	Lasso	0,9522	0,9522	-0,0001	3,0710	3,9700
I	Masc.	ElasticNet	0,9520	0,9522	-0,0002	3,0794	3,9717
I	Masc.	Ridge	0,9518	0,9521	-0,0003	3,0641	3,9755
I	Masc.	Linear Regression	0,9517	0,9520	-0,0003	3,0606	3,9780
I	Masc.	SVR	0,9511	0,9514	-0,0004	3,0514	4,0026
I	Masc.	Random Forest	0,9119	0,9238	-0,0120	4,0686	5,0127
I	Masc.	Extra Tree	0,8921	0,9118	-0,0196	4,3944	5,3957
I	Masc.	Decision Tree	0,8147	0,7834	+0,0313	6,3289	8,4526
I	Masc.	KNN	0,7344	0,7449	-0,0105	7,2369	9,1739
II	Fem.	CatBoost	0,8902	0,8724	+0,0178	4,9513	6,2250
II	Fem.	XGBoost	0,8848	0,8709	+0,0139	5,0212	6,2598
II	Fem.	Gradient Boosting	0,8791	0,8644	+0,0147	5,0978	6,4165
II	Fem.	LightGBM	0,8792	0,8585	+0,0206	5,2533	6,5535
II	Fem.	Linear Regression	0,8635	0,8451	+0,0184	5,5886	6,8576
II	Fem.	Lasso	0,8638	0,8450	+0,0188	5,6089	6,8604
II	Fem.	Ridge	0,8636	0,8448	+0,0188	5,6013	6,8644
II	Fem.	ElasticNet	0,8637	0,8447	+0,0190	5,6104	6,8671
II	Fem.	SVR	0,8610	0,8444	+0,0166	5,5628	6,8744

Continua na próxima página

Tabela 45 – continuação

Exp.	Sexo	Modelo	$R^2$ Treino	$R^2$ Teste	$\Delta R^2$	MAE	RMSE
II	Fem.	Extra Tree	0,8447	0,8183	+0,0264	5,9042	7,4274
II	Fem.	Random Forest	0,8337	0,8123	+0,0214	6,1973	7,5490
II	Fem.	Decision Tree	0,7234	0,7153	+0,0081	7,5257	9,2966
II	Fem.	KNN	0,6838	0,6560	+0,0278	8,2802	10,2203
II	Masc.	CatBoost	0,8788	0,8681	+0,0107	5,2962	6,5962
II	Masc.	XGBoost	0,8767	0,8616	+0,0151	5,3457	6,7576
II	Masc.	Gradient Boosting	0,8719	0,8504	+0,0216	5,5722	7,0265
II	Masc.	LightGBM	0,8647	0,8458	+0,0189	5,6828	7,1333
II	Masc.	SVR	0,8302	0,8239	+0,0063	6,0270	7,6219
II	Masc.	Ridge	0,8431	0,8194	+0,0237	6,1660	7,7190
II	Masc.	Linear Regression	0,8420	0,8192	+0,0228	6,1597	7,7238
II	Masc.	ElasticNet	0,8434	0,8191	+0,0243	6,1770	7,7262
II	Masc.	Lasso	0,8434	0,8188	+0,0246	6,1828	7,7318
II	Masc.	Random Forest	0,8027	0,7977	+0,0050	6,6685	8,1701
II	Masc.	Extra Tree	0,7952	0,7856	+0,0097	6,9077	8,4110
II	Masc.	Decision Tree	0,6585	0,6715	-0,0130	8,3941	10,4100
II	Masc.	KNN	0,6218	0,5995	+0,0223	9,1609	11,4954
III	Fem.	CatBoost	0,8649	0,8449	+0,0200	5,5427	6,8620
III	Fem.	XGBoost	0,8607	0,8431	+0,0176	5,6348	6,9019
III	Fem.	Gradient Boosting	0,8527	0,8347	+0,0181	5,7690	7,0851
III	Fem.	SVR	0,8346	0,8347	-0,0000	5,7635	7,0854
III	Fem.	Linear Regression	0,8371	0,8332	+0,0039	5,7957	7,1170
III	Fem.	Lasso	0,8375	0,8324	+0,0050	5,8337	7,1325
III	Fem.	Ridge	0,8372	0,8323	+0,0049	5,8223	7,1356
III	Fem.	ElasticNet	0,8375	0,8318	+0,0057	5,8522	7,1473
III	Fem.	LightGBM	0,8521	0,8293	+0,0228	5,8099	7,1993
III	Fem.	Extra Tree	0,8264	0,8025	+0,0239	6,2773	7,7429
III	Fem.	Random Forest	0,8236	0,7996	+0,0240	6,4361	7,7998
III	Fem.	Decision Tree	0,7228	0,7117	+0,0111	7,6105	9,3558
III	Fem.	KNN	0,6865	0,6545	+0,0320	8,1522	10,2427
III	Masc.	CatBoost	0,8602	0,8422	+0,0180	5,8182	7,2151
III	Masc.	XGBoost	0,8583	0,8348	+0,0235	5,9628	7,3829
III	Masc.	Gradient Boosting	0,8513	0,8309	+0,0204	5,9451	7,4693
III	Masc.	LightGBM	0,8459	0,8265	+0,0194	6,0670	7,5661
III	Masc.	SVR	0,8123	0,8206	-0,0083	6,0856	7,6937
III	Masc.	Lasso	0,8271	0,8063	+0,0208	6,4656	7,9943
III	Masc.	ElasticNet	0,8267	0,8058	+0,0209	6,4590	8,0043

Continua na próxima página

Tabela 45 – continuação

Exp.	Sexo	Modelo	$R^2$ Treino	$R^2$ Teste	$\Delta R^2$	MAE	RMSE
III	Masc.	Ridge	0,8265	0,8054	+0,0211	6,4608	8,0134
III	Masc.	Linear Regression	0,8262	0,8050	+0,0212	6,4557	8,0201
III	Masc.	Extra Tree	0,8077	0,7982	+0,0095	6,6948	8,1593
III	Masc.	Random Forest	0,8000	0,7959	+0,0042	6,6766	8,2067
III	Masc.	Decision Tree	0,6597	0,6850	-0,0253	8,1541	10,1946
III	Masc.	KNN	0,6525	0,6583	-0,0058	8,5270	10,6182
IV	Fem.	XGBoost	0,8570	0,8474	+0,0096	5,4978	6,8064
IV	Fem.	CatBoost	0,8619	0,8444	+0,0175	5,5745	6,8733
IV	Fem.	Gradient Boosting	0,8525	0,8426	+0,0099	5,5892	6,9135
IV	Fem.	Linear Regression	0,8467	0,8415	+0,0052	5,6031	6,9371
IV	Fem.	Lasso	0,8470	0,8411	+0,0058	5,6120	6,9451
IV	Fem.	Ridge	0,8468	0,8410	+0,0058	5,6109	6,9477
IV	Fem.	LightGBM	0,8478	0,8406	+0,0072	5,6201	6,9563
IV	Fem.	ElasticNet	0,8470	0,8402	+0,0068	5,6259	6,9663
IV	Fem.	SVR	0,8448	0,8388	+0,0059	5,5691	6,9956
IV	Fem.	Extra Tree	0,8322	0,8240	+0,0082	5,9152	7,3097
IV	Fem.	Random Forest	0,8233	0,8204	+0,0029	6,0870	7,3837
IV	Fem.	Decision Tree	0,7295	0,7259	+0,0036	7,4338	9,1230
IV	Fem.	KNN	0,7040	0,6938	+0,0102	7,7647	9,6419
IV	Masc.	XGBoost	0,8580	0,8544	+0,0036	5,5620	6,9320
IV	Masc.	CatBoost	0,8596	0,8518	+0,0078	5,6163	6,9923
IV	Masc.	LightGBM	0,8482	0,8480	+0,0002	5,6639	7,0824
IV	Masc.	Gradient Boosting	0,8533	0,8474	+0,0059	5,7282	7,0962
IV	Masc.	Extra Tree	0,8112	0,8146	-0,0033	6,3117	7,8219
IV	Masc.	SVR	0,8123	0,8139	-0,0016	6,3063	7,8364
IV	Masc.	Linear Regression	0,8127	0,8125	+0,0002	6,3282	7,8657
IV	Masc.	Ridge	0,8129	0,8125	+0,0004	6,3284	7,8659
IV	Masc.	Lasso	0,8135	0,8116	+0,0019	6,3436	7,8833
IV	Masc.	ElasticNet	0,8135	0,8110	+0,0025	6,3588	7,8965
IV	Masc.	Random Forest	0,8088	0,8109	-0,0020	6,4144	7,8989
IV	Masc.	Decision Tree	0,7305	0,7238	+0,0067	7,7787	9,5460
IV	Masc.	KNN	0,7128	0,7104	+0,0024	7,8581	9,7745

**Fonte:** Elaborado pela autora

**Nota:**  $\Delta R^2$  positivo (+) indica desempenho superior no treino;  $\Delta R^2$  negativo (-) indica desempenho superior no teste.