# DA ANÁLISE EXPLORATÓRIA À SÍNTESE DIMENSIONAL: desenvolvimento de um indicador sintético de bem-estar estudantil através de modelagem preditiva[1]

**Luiz Gustavo Barbosa Camargos**[2]
**Ismael Santana Silva**[3]

Submetido em: 19/06/2026 | Aprovado em: 19/06/2026

## RESUMO

A crescente incidência de distúrbios emocionais entre estudantes universitários tem impulsionado o desenvolvimento de soluções baseadas em dados para compreender e monitorar o bem-estar psicológico. Este trabalho propõe uma abordagem para análise de questionários estruturados de saúde mental, combinando análise exploratória, aprendizado de máquina e construção de indicadores sintéticos. Utilizando um conjunto de dados público com informações psicológicas, acadêmicas, fisiológicas e sociais de estudantes, foram realizadas três etapas: (1) análise exploratória das relações entre fatores de estresse e bem-estar; (2) modelagem preditiva para classificação de níveis de estresse e satisfação emocional; e (3) desenvolvimento de um índice de bem-estar inspirado no WHO-5 Well-Being Index. Os resultados apontaram depressão, ansiedade, autoestima e qualidade do sono como os principais fatores associados ao bem-estar, além de validar modelos preditivos explicáveis. O índice proposto, composto por cinco dimensões interpretáveis, apresentou alta consistência interna e capacidade discriminativa em um conjunto de dados independente. Em comparação a um agregado simples das mesmas variáveis, reduziu o erro padrão de medida em aproximadamente 39%, mostrando potencial para aplicações de triagem e análise populacional. A pesquisa contribui para o aprimoramento da análise computacional de questionários psicológicos e para a compreensão dos fatores relacionados ao bem-estar subjetivo no contexto acadêmico.

**Palavras-chave**: bem-estar estudantil; aprendizado de máquina; indicador sintético; análise exploratória; modelagem preditiva.

## ABSTRACT

[1] Artigo científico elaborado para o trabalho de conclusão do curso de Engenharia de Computação do campus Nova Gameleira do Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG).

[2] Graduando em Engenharia de Computação no campus Nova Gameleira do Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG). luiz.880088@gmail.com

[3] Orientador do trabalho. Docente do Departamento de Computação do campus Nova Gameleira do Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG). Av. Amazonas, 7675, Nova Gameleira, Belo Horizonte (MG).

The increasing incidence of emotional disorders among university students has driven the development of data-based solutions to understand and monitor psychological well-being. This work proposes an approach for the analysis of structured mental health questionnaires, combining exploratory analysis, machine learning, and the construction of synthetic indicators. Using a public dataset containing psychological, academic, physiological, and social information about students, three stages were carried out: (1) exploratory analysis of the relationships between stress factors and well-being; (2) predictive modeling for the classification of stress levels and emotional satisfaction; and (3) development of a well-being index inspired by the WHO-5 Well-Being Index. The results pointed to depression, anxiety, self-esteem, and sleep quality as the main factors associated with well-being, in addition to validating explainable predictive models. The proposed index, composed of five interpretable dimensions, showed high internal consistency and discriminative capacity in an independent dataset. Compared to a simple aggregate of the same variables, it reduced the standard error of measurement by approximately 39%, showing potential for screening applications and population-level analysis. This research contributes to the improvement of the computational analysis of psychological questionnaires and to the understanding of the factors related to subjective well-being in the academic context.

**Keywords**: student well-being; machine learning; synthetic indicator; exploratory analysis; predictive modeling.

## 1 INTRODUÇÃO

A saúde mental tem se consolidado como uma das principais preocupações contemporâneas, especialmente no contexto acadêmico (World Health Organization, 2022; Auerbach *et al.*, 2018). Estudos epidemiológicos recentes demonstram que 30–40% dos estudantes universitários relatam sintomas clinicamente significativos de ansiedade e depressão durante a graduação (Lipson *et al.*, 2019; Bruffaerts *et al.*, 2018). O ambiente universitário, embora propício ao desenvolvimento intelectual e social, frequentemente impõe pressões relacionadas ao desempenho, à competitividade e à transição para a vida adulta, contribuindo para o aumento de sintomas de ansiedade, estresse e desmotivação (Pascoe *et al.*, 2020; Baghurst; Kelley, 2014). Diante desse cenário, cresce a necessidade de métodos sistemáticos capazes de auxiliar a identificação precoce de sinais de sofrimento psicológico.

Paralelamente, o avanço da ciência de dados tem ampliado as possibilidades de análise em áreas tradicionalmente qualitativas, como a psicologia e a educação (Dwyer *et al.*, 2018; Shatte *et al.*, 2019). Por exemplo, modelos de aprendizado de máquina e técnicas de mineração de dados vêm sendo aplicadas com sucesso para extrair padrões complexos e prever estados emocionais a partir de dados estruturados (Nemesure *et al.*, 2021; Graham *et al.*, 2019), possibilitando abordagens quantitativas e explicáveis para compreender o bem-estar humano (Wies *et al.*, 2021).

Entre os instrumentos mais utilizados na mensuração do bem-estar subjetivo, destacam-se escalas padronizadas como o WHO-5 Well-Being Index (Topp *et al.*, 2015), o Short Warwick-Edinburgh Mental Well-Being Scale (SWEMWBS) (Stewart-Brown *et al.*, 2009) e o Mental Health Continuum – Short Form (MHC-SF) (Keyes, 2007). Essas ferramentas fornecem medidas consistentes e avaliadas internacionalmente em dimensões como satisfação, vitalidade e equilíbrio emocional (Tennant *et al.*, 2007). No entanto, a aplicação

dessas escalas em contextos de análise de grandes volumes de dados coletados ainda é limitada, especialmente quando se trata de ambientes digitais (Torous *et al.*, 2021; Mohr *et al.*, 2017).

Estudos recentes têm demonstrado que a integração de Aprendizado de Máquina com dados de autorrelato pode alcançar precisões superiores a 90% na classificação de níveis de estresse estudantil (Ovi *et al.*, 2025; Meng *et al.*, 2024). Contudo, persiste a lacuna metodológica na tradução de padrões algorítmicos complexos em indicadores clínicos interpretáveis que possam orientar intervenções práticas (Panicheva *et al.*, 2022; Eder *et al.*, 2025). Instrumentos psicométricos tradicionais, embora validados, são desenvolvidos através de processos puramente teóricos que não capturam plenamente estruturas latentes identificadas por análises dirigidas por dados (Graham *et al.*, 2019).

Nesse contexto, este trabalho propõe uma abordagem computacional para a análise automatizada do bem-estar psicológico de estudantes universitários, utilizando um conjunto de dados público composto por variáveis emocionais, acadêmicas, fisiológicas e sociais. O estudo é estruturado em três etapas complementares. A primeira consiste em uma análise exploratória e comparativa para identificar correlações entre fatores de estresse e bem-estar. A segunda aborda a modelagem preditiva por meio de algoritmos de Aprendizado de Máquina, com o objetivo de classificar níveis de bem-estar e detectar padrões associados ao sofrimento emocional. Por fim, a terceira etapa propõe a construção de um indicador sintético de bem-estar, inspirado em instrumentos psicométricos reconhecidos, visando sintetizar os resultados de forma interpretável e aplicável.

A principal contribuição deste trabalho está na integração entre métodos analíticos e construtos psicológicos, buscando não apenas identificar padrões estatísticos, mas também oferecer interpretações consistentes sobre o bem-estar subjetivo. Espera-se que os resultados obtidos contribuam para o desenvolvimento de ferramentas automatizadas de triagem, capazes de apoiar instituições de ensino e profissionais de saúde mental na formulação de estratégias preventivas e personalizadas.

## 2 TRABALHOS RELACIONADOS

A consolidação de abordagens computacionais aplicadas à avaliação do bem-estar psicológico tem estimulado uma produção crescente de estudos que combinam psicometria e aprendizado de máquina para investigar fatores emocionais, comportamentais e sociais em populações estudantis. Essa convergência entre psicologia e ciência de dados possibilita novas formas de compreender o bem-estar subjetivo por meio da extração de padrões em dados de questionários, registros digitais e contextos ambientais. Nos últimos anos, observa-se um avanço significativo no uso de algoritmos supervisionados, técnicas de seleção de atributos e modelos explicáveis voltados à triagem automatizada de estresse e à modelagem de estados emocionais. Dentro desse panorama, destacam-se pesquisas que buscam não apenas prever níveis de bem-estar, mas também desenvolver estruturas metodológicas integradas e reprodutíveis, capazes de aproximar os resultados computacionais das interpretações psicométricas tradicionalmente empregadas.

É nesse contexto que se insere o estudo de Ovi *et al.* (2025), que representa uma das abordagens mais abrangentes e recentes neste domínio, cuja metodologia e base de dados fundamentam diretamente a proposta deste projeto. Os autores propõem um *framework* de aprendizado de máquina sensível ao contexto para a classificação de estresse em estudantes universitários, integrando múltiplos domínios psicológico, acadêmico, fisi-

ológico, social e ambiental. A pesquisa utiliza dois conjuntos de dados complementares derivados de questionários: o *Student Stress Factors: A Comprehensive Analysis* e o *Stress and Well-being Data of College Students*, ambos coletados de estudantes entre 18 e 21 anos. O *pipeline* desenvolvido inclui etapas de pré-processamento, seleção de atributos com SelectKBest e *Recursive Feature Elimination with Cross-Validation* (RFECV), redução de dimensionalidade com *Principal Component Analysis* (PCA) e treinamento de seis classificadores-base: *Support Vector Machine* (SVM), Random Forest, Gradient Boosting, XGBoost, AdaBoost e Bagging. A combinação desses modelos em estruturas de *ensemble* (*hard voting*, *soft voting*, *weighted voting* e *stacking*) resultou em acurácias superiores a 99%, superando estudos anteriores com os mesmos conjuntos de dados. O trabalho destaca ainda a importância de integrar informações contextuais para aprimorar a precisão de modelos de predição de estresse, aproximando a análise de uma aplicação realista em ambientes educacionais.

Os dois conjuntos de dados empregados por Ovi *et al.* (2025) constituem também a base empírica do presente trabalho. O *Student Stress Factors* (RxNach, 2024) fornece uma base estruturada de variáveis de estresse categorizadas em fatores psicológicos, fisiológicos, acadêmicos e sociais, com ênfase em determinantes contextuais e estruturais. O *Stress and Well-being Data of College Students* (Singh, 2024) prioriza variáveis associadas a sintomas psicossomáticos, qualidade do sono e ansiedade, demonstrando forte correlação com dimensões de bem-estar subjetivo. A complementaridade entre as ênfases dos dois *datasets*, contextual no primeiro e reativa no segundo, fundamenta a estratégia de validação cruzada adotada neste trabalho, cujo detalhamento é apresentado na Seção 4.

A comparação entre abordagens destaca diferenças metodológicas essenciais. Enquanto Ovi *et al.* (2025) adotam um enfoque fortemente orientado a *ensembles* e combinação de classificadores com redução de dimensionalidade e seleção robusta de atributos, outros estudos revelam *trade-offs* frequentes entre complexidade do modelo e interpretabilidade. Meng *et al.* (2024) aplicaram seis modelos supervisionados em uma amostra internacional de 61.585 estudantes, observando que Random Forest apresentou o melhor desempenho preditivo para escores de bem-estar medidos pelo WHO-5; contudo, os autores discutem limitações inerentes a modelos menos interpretáveis quando o objetivo é derivar intervenções clínicas acionáveis. A escolha de Random Forest, neste caso, equilibra capacidade preditiva com alguma interpretabilidade via importância de variáveis, reforçando a tendência da literatura de que classificadores baseados em árvores oferecem desempenho estável com menor exigência amostral em dados tabulares de autorrelato.

Panicheva *et al.* (2022) exploraram uma via distinta ao focar em dados digitais passivos e em linguagem natural extraída de mensagens de usuários de um aplicativo de saúde mental. Em contraste com *pipelines* baseados em questionários, esse trabalho demonstrou que recursos linguísticos e padrões de uso de aplicativos conseguem predizer, de forma moderada, escores WHO-5 (Pearson $r \approx 0{,}37$) e classificar risco depressivo com sensibilidade relevante, embora com limitações de especificidade. O estudo ilustra a utilidade de *features* comportamentais e textuais quando integradas a modelos de regressão e classificação, mas também ressalta problemas de amostragem pequena e ausência de validação externa, que comprometem a generalização dos modelos baseados em traços digitais.

No campo da integração multimodal, Eder *et al.* (2025) apresentaram um exemplo de aplicação clínica em atenção primária, combinando questionários (incluindo WHO-5 e PHQ-9), sinais vitais e biomarcadores com técnicas como Boruta para seleção de atributos

e XGBoost para classificação. A estratégia de empilhamento e de modelos sequenciais resultou em ganho de desempenho em relação a instrumentos isolados, além de possibilitar a identificação de subgrupos clínicos por meio de *clustering* por mistura gaussiana. Esse trabalho destaca a vantagem de combinar fontes heterogêneas: quando múltiplas modalidades são bem integradas, é possível obter classificadores mais robustos e ao mesmo tempo extrair perfis clínicos com significado biológico. As limitações apontadas por Eder *et al.* (2025) envolvem tamanho amostral moderado e necessidade de validação multicêntrica.

Em perspectiva complementar, Graham *et al.* (2019) argumentam que abordagens puramente teóricas de construção de instrumentos psicométricos não capturam plenamente estruturas latentes identificáveis por análises dirigidas por dados. Os autores demonstram que modelos de *machine learning* aplicados a dados de autorrelato conseguem identificar padrões de associação entre variáveis que escapam à formulação clássica de itens por construto teórico, sugerindo que a integração entre métodos preditivos e psicometria pode produzir instrumentos com maior aderência empírica. Essa constatação fundamenta diretamente a estratégia de triangulação adotada no presente trabalho, na qual indicadores derivados de modelos preditivos (*feature importance*) são combinados com métricas estatísticas tradicionais (correlação de Pearson e *Mutual Information*) para a derivação dos pesos dimensionais do indicador proposto.

A literatura também evidencia diferentes estratégias para seleção de atributos e explicabilidade. Métodos como Boruta e RFECV têm sido frequentemente empregados para reduzir ruído e priorizar variáveis com relevância clínica, enquanto técnicas de explicabilidade pós-hoc, como SHAP e LIME, são recomendadas para traduzir decisões de modelos complexos (particularmente redes neurais e *ensembles* de alta dimensionalidade) em conhecimento útil acessível a profissionais de saúde. No entanto, para modelos baseados em árvores de decisão, a importância nativa dos atributos (*mean decrease in impurity*, *gain*) já fornece interpretabilidade diretamente derivada da estrutura do classificador, dispensando camadas adicionais de explicabilidade quando o objetivo é derivar pesos para um indicador sintético, e não explicar predições individuais. O presente trabalho adota essa segunda via, utilizando a importância nativa de seis modelos *tree-based* como uma das três fontes de triangulação para a ponderação dimensional do ISBE, conforme detalhado nas Seções 7.1 e 7.2.

No que se refere aos instrumentos de referência para mensuração de bem-estar e estresse, quatro escalas consolidadas fornecem os *benchmarks* contra os quais indicadores emergentes devem ser avaliados. O WHO-5 Well-Being Index (Topp *et al.*, 2015), com cinco itens e estrutura unidimensional, destaca-se pela brevidade e pela ampla validação transcultural, sendo frequentemente utilizado como desfecho em estudos de intervenção e como critério de triagem para depressão. A Perceived Stress Scale em sua versão de 10 itens (PSS-10) (Cohen; Williamson, 1988; Lee, 2012) mede a percepção subjetiva de estresse em dimensão única, com confiabilidade na faixa de 0,78 a 0,84. A Depression Anxiety Stress Scales de 21 itens (DASS-21) (Lovibond; Lovibond, 1995; Henry; Crawford, 2005) avalia três constructos relacionados em estrutura tridimensional, apresentando alfas entre 0,93 e 0,96. Por fim, a Warwick-Edinburgh Mental Well-being Scale (WEMWBS) (Tennant *et al.*, 2007; Stewart-Brown *et al.*, 2009) mede bem-estar positivo com 14 itens em dimensão única, com alfas na faixa de 0,89 a 0,91. Todos esses instrumentos possuem estabilidade temporal documentada (coeficiente de correlação intraclasse, ICC, entre 0,72 e 0,85) e validação multicêntrica, propriedades que constituem o padrão de maturidade contra o qual o ISBE é posicionado na Tabela 21.

Por fim, a literatura revisada aponta lacunas recorrentes que motivam a presente investigação. Há escassez de estudos que integrem instrumentos de avaliação de bem-estar com *pipelines* preditivos automatizados, traduzindo padrões algorítmicos em indicadores interpretáveis e reprodutíveis; muitos trabalhos empregam *proxies* ou instrumentos diferentes como variáveis-alvo, dificultando comparações diretas. Adicionalmente, a articulação entre medidas de sintomas (como o PHQ-9) e medidas positivas de bem-estar (como o WHO-5 ou o WEMWBS) permanece insuficiente na maioria dos *pipelines* automatizados, o que limita a compreensão do fenômeno a uma única dimensão do espectro saúde-doença. A maioria das investigações também esbarra em amostras restritas, falta de validação externa e insuficiente relato de medidas de proteção de privacidade e anonimização de dados. Essas questões metodológicas e éticas são especialmente relevantes quando se pretende aplicar soluções automatizadas em ambientes educacionais, que exigem transparência, explicabilidade e salvaguardas adequadas.

Em síntese, os trabalhos analisados apresentam avanços complementares: os estudos sobre *ensembles* e seleção de atributos demonstram elevado desempenho preditivo em contextos controlados; as pesquisas com dados digitais passivos e integração multimodal ampliam o escopo de fontes além dos questionários tradicionais; e os instrumentos psicométricos consolidados fornecem o padrão de referência contra o qual indicadores emergentes devem ser avaliados. Ao mesmo tempo, a literatura evidencia lacunas persistentes na tradução de padrões algorítmicos em métricas clinicamente interpretáveis, na validação externa e na articulação entre medidas de sintomas e medidas positivas de bem-estar. O presente trabalho se insere nesse panorama propondo uma sequência metodológica integradora: análise exploratória, modelagem preditiva e construção de um indicador sintético inspirado no WHO-5, com ênfase em explicabilidade e validação, buscando mitigar as lacunas apontadas e contribuir com uma abordagem reprodutível e aplicável a contextos universitários.

## 3 METODOLOGIA

Esta pesquisa adota abordagem quantitativa, aplicada e preditiva, fundamentada em técnicas de aprendizado supervisionado para classificação multiclasse de níveis de estresse estudantil. O diferencial metodológico central reside na construção do Indicador Sintético de Bem-Estar Estudantil (ISBE), métrica composta que agrega múltiplas dimensões de bem-estar em *score* único interpretável na escala 0–100, permitindo triagem institucional, monitoramento longitudinal e identificação de perfis de vulnerabilidade específicos.

As etapas de análise exploratória de dados e treinamento de modelos preditivos seguem o *framework context-aware* proposto por Ovi *et al.* (2025), que demonstrou acurácias superiores a 93% na detecção de estresse estudantil através da integração de fatores psicológicos, acadêmicos, fisiológicos, sociais e ambientais. Este trabalho expande tal *framework* ao desenvolver um indicador com potencial para ser psicometricamente validado que traduz padrões identificados por algoritmos de Aprendizado de Máquina em métrica clinicamente interpretável e comparável a instrumentos estabelecidos.

O fluxo metodológico organiza-se em quatro macrofases. As três primeiras, de natureza preparatória e preditiva, têm seus resultados sintetizados nas Seções 4, 5 e 6, com o detalhamento integral dos procedimentos e das análises no Apêndice A: Preparação e Exploração dos Dados (Macrofase I), Engenharia de *Features* e Análise Exploratória (Macrofase II) e Modelagem Preditiva e *Ensemble* (Macrofase III). A quarta macrofase,

Construção e Validação do ISBE, constitui a contribuição original do trabalho e é desenvolvida nas Seções 7.2 e 7.3. A Figura 1 sintetiza o encadeamento das etapas.

Figura 1 – Fluxo Metodológico Geral do Trabalho


Fonte: Elaborado pelo autor (2026).

O encadeamento metodológico segue uma lógica cumulativa, na qual cada macrofase alimenta a seguinte. A Macrofase I prepara e caracteriza os dados. A Macrofase II realiza a análise exploratória e a seleção de características que sustenta a modelagem preditiva. A Macrofase III treina os classificadores e, como subproduto, produz a *feature importance* agregada dos modelos baseados em árvore. A Macrofase IV constrói o ISBE, integrando três fontes de evidência: a correlação de Pearson (Cover; Thomas, 2006) e a *Mutual Information* (Battiti, 1994) (Macrofase II) e a *feature importance* (Macrofase III). Cabe distinguir, desde já, dois processos de seleção de características independentes e com propósitos distintos: a cascata por modelo descrita na Seção 5, que reduz a dimensionalidade para a modelagem preditiva; e a seleção via *Mutual Information* descrita na Seção 7.2, que define o conjunto de 14 *features* que compõem o ISBE. Os dois processos não se confundem: o ISBE não é construído sobre as *features* selecionadas pela cascata, mas sobre um subconjunto próprio derivado de critérios psicométricos e informacionais. A seguir, é apresentada cada fase do projeto.

# 4 PREPARAÇÃO E EXPLORAÇÃO DOS DADOS (MACROFASE I)

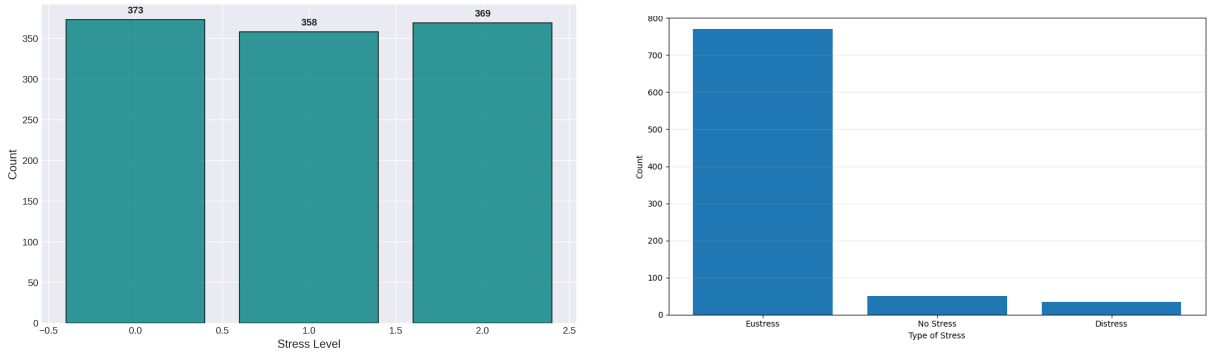
A primeira macrofase preparou e caracterizou os dados que sustentam todas as análises subsequentes. Esta seção reporta os principais achados de coleta, qualidade e exploração; os procedimentos completos e a caracterização gráfica das distribuições estão detalhados no Apêndice A.1.

## 4.1 Delineamento ético e coleta de dados

Foram utilizados os mesmos dois *datasets* públicos e anonimizados empregados por Ovi *et al.* (2025). O Dataset 1 (*Student Stress Factors*) (RxNach, 2024) reúne 1.100 registros e 20 *features* preditoras que cobrem dimensões psicológicas, fisiológicas, ambientais,

acadêmicas e sociais. Sua variável-alvo `stress_level` apresentou distribuição equilibrada entre as três classes, baixo (33,9%), médio (32,5%) e alto (33,5%), condição favorável à modelagem multiclasse. O Dataset 2 (*Stress and Well-being Data of College Students*) (Singh, 2024) reúne 843 registros e 25 *features* voltadas a sintomas psicossomáticos e estressores contextuais. Sua variável-alvo, o tipo de estresse, mostrou-se fortemente desbalanceada: 91,1% dos casos concentram-se em *eustress*, contra 5,1% sem estresse e 3,8% em *distress* (Figura 2). Esse desbalanceamento é um achado consequente, pois um classificador trivial que atribuísse *eustress* a todos os casos já alcançaria cerca de 91% de acurácia, o que tornou obrigatório o uso de métricas *macro* na avaliação dos modelos (Macrofase III). A complementaridade entre as ênfases dos dois conjuntos, contextual no primeiro e reativa no segundo, e a avaliação empírica prévia que permite comparação com a literatura justificaram a escolha. Por se tratar de dados secundários públicos, a pesquisa dispensa submissão a Comitê de Ética em Pesquisa, conforme a Resolução nž 510/2016 do Conselho Nacional de Saúde (Brasil, 2016).

Figura 2 – Distribuição das variáveis-alvo dos dois *datasets*




(a) Dataset 1: classes equilibradas

(b) Dataset 2: predomínio de *eustress*

Fonte: Elaborado pelo autor (2026).

## 4.2 Pré-processamento dos dados

A inspeção estrutural não identificou valores ausentes em nenhum dos conjuntos de dados. O Dataset 1 não exigiu correções, enquanto o Dataset 2 continha 27 registros duplicados (3,2%), que foram removidos, resultando em 816 observações válidas. Após a normalização e a divisão estratificada em conjuntos de treino e teste, foi realizada uma análise exploratória dos *outliers*. Como os valores extremos estavam predominantemente associados a características individuais e contextuais dos participantes, optou-se por preservá-los, uma vez que representavam variações legítimas da amostra e não evidências de erros de medição. A única exceção foi a variável idade do Dataset 2, cuja elevada dispersão e ausência de associação consistente com o estresse indicaram baixa relevância preditiva, motivando sua exclusão das etapas subsequentes da análise. O detalhamento desse procedimento é apresentado no Apêndice A.1.

## 4.3 Análise exploratória de dados

A exploração univariada revelou distribuições compatíveis com dados psicossociais reais. No Dataset 1, as variáveis ligadas a sintomas emocionais e fisiológicos (e.g., ansiedade, depressão, dor de cabeça e problemas respiratórios) apresentaram maior assimetria,

refletindo variação substancial na intensidade com que esses fatores são vivenciados. Em contraste, os fatores protetivos (e.g., autoestima, qualidade do sono, desempenho acadêmico e atendimento de necessidades básicas) concentraram-se nas faixas superiores das escalas, sugerindo a presença de recursos de suporte e estabilidade na amostra.

No Dataset 2, removida a variável idade, as demais variáveis exibiram assimetria predominantemente baixa, compatível com sua natureza ordinal e binária, sem distorções que comprometam as análises subsequentes. A análise das relações entre as *features* e o estresse, de natureza bivariada e multivariada, é apresentada na Macrofase II (Seção 5); os gráficos de distribuição completos constam no Apêndice A.1.

# 5 ENGENHARIA DE ATRIBUTOS E ANÁLISE MULTIVARIADA (MACROFASE II)

A segunda macrofase aprofundou a compreensão das relações entre as variáveis e definiu, por algoritmo, o subconjunto de *features* utilizado na modelagem. Esta seção reporta os principais achados. Os gráficos, as tabelas e o detalhamento dos procedimentos constam no Apêndice A.2.

## 5.1 Seleção de características

A seleção de características foi conduzida por uma cascata de *SelectKBest*, RFECV e PCA, reexecutada de forma independente para cada um dos sete algoritmos, de modo que o número de *features* retidas não é fixo, mas determinado pelo desempenho de cada classificador. No bloco do *Support Vector Machine*, por exemplo, o procedimento convergiu para 19 *features*. A etapa de PCA evidenciou a diferença estrutural entre os *datasets*, no Dataset 1, o primeiro componente principal concentra cerca de 60% da variância total e a variância acumulada atinge 95% com aproximadamente 15 componentes, revelando forte redundância informacional que favorece a redução de dimensionalidade; no Dataset 2, a variância distribui-se de forma mais gradual, exigindo cerca de 22 componentes para os mesmos 95%, o que indica baixa redundância e ganhos limitados com a compressão. O detalhamento da cascata e os gráficos de variância explicada constam no Apêndice A.2.

## 5.2 Análise exploratória multivariada

A correlação de Pearson revelou um contraste estrutural entre os *datasets*. No Dataset 1, que mede o estresse de forma contínua, as associações mais fortes ocorreram com *bullying* ($r = 0{,}75$), preocupações com a carreira ($r = 0{,}74$), ansiedade ($r = 0{,}74$), depressão ($r = 0{,}73$) e dor de cabeça ($r = 0{,}71$), caracterizando determinantes crônicos e estruturais. A análise apontou ainda multicolinearidade relevante entre esses preditores psicossociais, como entre ansiedade e preocupações com a carreira ($r = 0{,}72$), o que exige cautela na seleção de variáveis (Tabela 1).

No Dataset 2, focado na experiência recente de estresse, as correlações foram bem menores e concentradas em respostas fisiológicas agudas, com destaque para batimentos cardíacos acelerados ($r = 0{,}44$), seguidos de ansiedade ou tensão recente ($r = 0{,}23$) e conflitos entre atividades acadêmicas e extracurriculares ($r = 0{,}21$); nesse conjunto não foram identificados pares com colinearidade elevada.

A *Mutual Information* reordenou a relevância das variáveis ao capturar dependências não lineares. No Dataset 1, `blood_pressure`, apesar de correlação linear modesta,

Tabela 1 – Correlações de Pearson mais fortes no Dataset 1 ($|r| > 0{,}7$)

| **Variável 1** | **Variável 2** | **Correlação** |
|---|---|---|
| *Correlações com a variável-alvo* | | |
| bullying | stress_level | 0,75 |
| future_career_concerns | stress_level | 0,74 |
| anxiety_level | stress_level | 0,74 |
| depression | stress_level | 0,73 |
| headache | stress_level | 0,71 |
| sleep_quality | stress_level | −0,75 |
| self_esteem | stress_level | −0,76 |
| academic_performance | stress_level | −0,72 |
| basic_needs | stress_level | −0,71 |
| safety | stress_level | −0,71 |
| *Pares entre preditores (multicolinearidade)* | | |
| anxiety_level | future_career_concerns | 0,72 |
| anxiety_level | bullying | 0,71 |
| future_career_concerns | bullying | 0,71 |
| depression | future_career_concerns | 0,71 |
| anxiety_level | sleep_quality | −0,71 |
| self_esteem | future_career_concerns | −0,71 |
| blood_pressure | social_support | −0,75 |

Fonte: Elaborado pelo autor (2026).

obteve a maior pontuação de MI, sinalizando que sua relação com o estresse é predominantemente não linear; qualidade do sono, preocupações com a carreira, depressão e ansiedade também figuraram entre as mais informativas (Figura 3). No Dataset 2, a percepção de batimentos cardíacos acelerados liderou novamente, coerente com a natureza reativa daquele conjunto.

A análise inferencial reforçou o panorama e a análise de variância (ANOVA) apontou diferenças significativas entre os grupos de estresse ($p < 0{,}001$), com os maiores efeitos em ansiedade, depressão e desempenho acadêmico no Dataset 1, e em batimentos cardíacos e tensão recente no Dataset 2. O conjunto desses achados, detalhado no Apêndice A.2, fundamentou tanto a seleção das *features* quanto a posterior ponderação do ISBE.

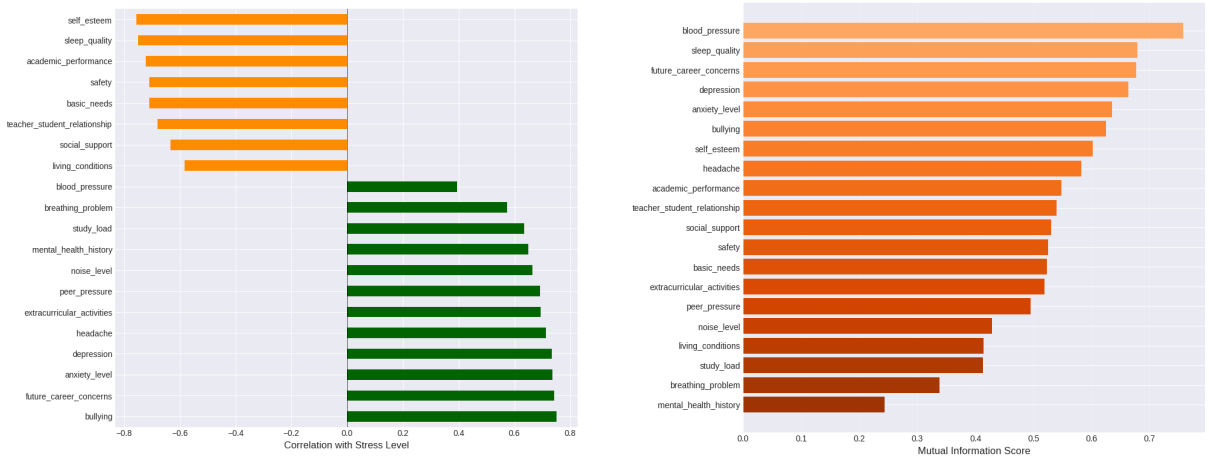
## 6 MODELAGEM PREDITIVA E *ENSEMBLE* (MACROFASE III)

A terceira macrofase avaliou o desempenho preditivo dos algoritmos e, como subproduto, consolidou a importância das *features* que alimenta a construção do ISBE. Os protocolos experimentais completos e as tabelas de desempenho constam no Apêndice A.3.

### 6.1 Modelagem preditiva

Sete algoritmos foram avaliados sob diferentes configurações de pré-processamento e redução de dimensionalidade. O desempenho variou conforme a estrutura de cada *dataset*: no Dataset 1, as melhores configurações associaram-se ao *SelectKBest*, método de seleção de *features* baseado no teste F da ANOVA (*f_classif*), que ranqueia cada variá-

Figura 3 – Relevância das variáveis em relação ao nível de estresse no Dataset 1




(a) Correlação de Pearson com o nível de estresse

(b) Pontuações de *Mutual Information*

Fonte: Elaborado pelo autor (2026).

vel pela sua capacidade de discriminar individualmente as classes de estresse. O número ótimo de *features* (K) não foi fixado manualmente: para cada modelo, o código testou todos os valores de K possíveis (de 1 até o total de variáveis disponíveis) e selecionou aquele que maximizou a acurácia média em validação cruzada estratificada de cinco *folds*. Sobre esse subconjunto inicial, aplicou-se ainda o RFECV (*Recursive Feature Elimination with Cross-Validation*), que eliminou iterativamente as variáveis menos relevantes considerando suas interações, resultando no conjunto final de *features* utilizado por cada modelo. Como esse processo é executado de forma independente para cada algoritmo, o K resultante pode variar entre eles, refletindo a combinação de atributos que melhor se adequou a cada estrutura de aprendizado.

Com esse protocolo, destacaram-se o *Random Forest* (92,4% de acurácia), seguido de AdaBoost (92,0%), XGBoost (91,6%) e *Gradient Boosting* (91,3%), conforme a Tabela 2. No Dataset 2, predominou o PCA, com o *Support Vector Machine* alcançando 99,1%, à frente de AdaBoost (96,7%) e XGBoost (96,2%). Esse padrão indica que a seleção prévia de atributos favoreceu o conjunto mais redundante (Dataset 1), enquanto a transformação dimensional beneficiou o conjunto mais disperso (Dataset 2).

Como subproduto da modelagem, consolidou-se a importância intrínseca das *features* a partir dos seis modelos baseados em árvore (*Random Forest*, *Gradient Boosting*, AdaBoost, XGBoost, LightGBM e *Bagging*). Trata-se da relevância que cada algoritmo atribui automaticamente às variáveis durante o próprio treinamento: modelos baseados em árvore calculam, a cada nó de decisão, o quanto cada *feature* reduz a impureza do conjunto (critério de Gini ou ganho de informação), e a importância final de cada variável corresponde à média ponderada dessas reduções ao longo de todas as árvores construídas. A variável `blood_pressure` despontou como a de maior importância média, muito acima das demais, confirmando que sua relação com o estresse, fraca em termos lineares, é capturada por modelos capazes de representar interações não lineares. Essa importância agregada constitui a terceira fonte da triangulação que pondera o ISBE, desenvolvida na Macrofase IV (Seção 7.2). A tabela completa de importâncias por modelo consta no Apêndice A.3.

Tabela 2 – Desempenho dos melhores modelos individuais e *ensembles* por *dataset*

| **Modelo** | **Config.** | **Acurácia (%)** |
|---|---|---|
| *Dataset 1 (classes equilibradas)* | | |
| Random Forest | SelectKBest | 92,364 |
| AdaBoost | SelectKBest | 92,000 |
| XGBoost | SelectKBest | 91,636 |
| Voting (*hard*) | *ensemble* | 93,091 |
| *Dataset 2 (classes desbalanceadas)* | | |
| SVM | PCA | 99,052 |
| AdaBoost | SelectKBest | 96,683 |
| XGBoost | SelectKBest | 96,209 |
| Stacking | *ensemble* | 99,530 |

Fonte: Elaborado pelo autor (2026).

## 6.2 Estratégias de *ensemble*

As estratégias de *ensemble* igualaram ou superaram os melhores modelos individuais. No Dataset 1, com classes equilibradas, o *Voting Classifier* (*hard*) atingiu 93,1% de acurácia, valor superior ao do melhor modelo individual desse conjunto; no Dataset 2, o *Stacking Classifier* alcançou 99,5%, o melhor resultado geral do estudo. Como antecipado pelo desbalanceamento identificado na Macrofase I, as acurácias do Dataset 2 devem ser lidas com cautela: um classificador trivial já alcançaria cerca de 91%, de modo que as métricas *macro* são mais informativas; as métricas *macro* reportadas no Apêndice A.3 confirmam que o ganho sobre esse piso é real, ainda que menor do que a acurácia bruta sugere.

# 7 CONSTRUÇÃO E AVALIAÇÃO DO ISBE (MACROFASE IV)

A quarta macrofase constitui a contribuição principal deste trabalho e integra, em um único encadeamento, a derivação empírica do indicador e a sua avaliação. Esta seção organiza-se em quatro partes complementares: (1) a fundamentação da importância das *features* por triangulação de três fontes (Seção 7.1); (2) a formulação do ISBE e sua aplicação ao Dataset 1 (Seção 7.2); (3) a avaliação da confiabilidade, validade, robustez, estrutura fatorial, capacidade discriminativa, validação em uma amostra externa utilizando o Dataset 2 e comparação com um agregado parcimonioso (Seção 7.3); e, por fim, (4) os aspectos éticos e as diretrizes de uso (Seção 7.4). Os dados são os mesmos das macrofases anteriores, descritos na Macrofase I, e correspondem aos empregados por Ovi *et al.* (2025), o que assegura comparabilidade com a literatura recente. Salvo indicação contrária, os resultados referem-se à versão ISBE 2.3.

## 7.1 Interpretabilidade e importância de *features*

A interpretabilidade do ISBE foi sustentada por três fontes complementares de relevância preditiva, posteriormente combinadas por triangulação para derivar os pesos do indicador: (1) a correlação absoluta de Pearson com a variável-alvo ($|r|$) capta associações lineares diretas entre cada *feature* e o estresse; (2) a *Mutual Information* (MI) capta

dependências não lineares; e (3) *feature importance* (FI) agregada dos seis modelos baseados em árvore (*Random Forest*, *Gradient Boosting*, AdaBoost, XGBoost, LightGBM e *Bagging*), consolidada na Macrofase III, quantifica a contribuição efetiva de cada variável nas decisões dos classificadores, via redução de impureza ou ganho. O emprego de três fontes reduz a dependência de um único método, explorando suas complementaridades.

Esses indicadores informaram três aplicações metodológicas distintas, que não devem ser confundidas: (1) a seleção das *features* de cada dimensão do ISBE, conduzida exclusivamente por *Mutual Information* na construção da versão 2.2; (2) a derivação dos pesos dimensionais e intradimensionais por triangulação das três fontes na versão 2.3; e (3) a verificação de que *features* teoricamente centrais apresentam importâncias empíricas elevadas. A seleção e a ponderação apoiam-se, portanto, em conjuntos de fontes diferentes e a *feature importance* entra apenas na etapa de ponderação, não na de seleção.

O procedimento de derivação dos pesos seguiu quatro passos. Primeiro, calcularam-se os valores brutos de $|r|$, MI e FI para cada uma das 14 *features*. A FI foi obtida normalizando as importâncias atribuídas às *features* de cada modelo para soma unitária e tomando a média aritmética dos seis modelos. Segundo, cada vetor foi normalizado pelo procedimento min-max dentro do conjunto de 14 *features*, gerando três vetores no intervalo $[0, 1]$. Terceiro, o escore compósito de cada *feature* ($i$) foi calculado como média aritmética simples das três fontes normalizadas:

$$\text{Composite}_i = \frac{|r|_{\text{norm},i} + \text{MI}_{\text{norm},i} + \text{FI}_{\text{norm},i}}{3} \tag{1}$$

Quarto, os pesos intradimensionais foram obtidos normalizando os escores compósitos dentro de cada dimensão (soma = 1), e os pesos dimensionais normalizando a soma dos compósitos por dimensão (soma = 1). Esse procedimento assegura que cada *feature* contribua proporcionalmente à sua relevância consolidada nas três fontes e que cada dimensão contribua proporcionalmente à relevância agregada de seus itens componentes, sem intervenção manual *a posteriori*.

As importâncias utilizadas como terceira fonte da triangulação não foram extraídas dos hiperparâmetros ótimos efetivamente salvos na Macrofase III. Para quatro dos seis modelos baseados em árvore, os `best_params_` originais não foram preservados, de modo que esses modelos precisaram ser retreinados a partir dos *grids* de busca, e não recarregados a partir do estado exato que produziu as métricas de desempenho reportadas. Como o procedimento de ajuste envolve componentes estocásticos (amostragem *bootstrap* em *Random Forest* e *Bagging*, subamostragem de *features* por nó, ordem de avaliação em *boosting*), a reconstrução sem sementes (`random_state`) fixas e idênticas às originais introduz uma variabilidade não mensurada nos valores brutos de FI e, por propagação, nos pesos finais. Em consequência, os pesos derivados da triangulação devem ser lidos como estimativas pontuais sujeitas a uma margem de flutuação, e não como coeficientes determinísticos. Essa incerteza foi caracterizada por duas vias complementares: (i) a análise de sensibilidade de cinco esquemas de ponderação (Seção 7.3, Tabela 9), que mostrou variação máxima de AUC (área sob a curva ROC) de apenas 0,015 entre cenários extremos, incluindo o cenário de pesos invertidos; e (ii) o registro formal dessa limitação na Seção 8.4 (L7). Recomenda-se, em replicações futuras, fixar e versionar as sementes de todos os estimadores e persistir os `best_params_` ao final da otimização, ou, alternativamente, estimar a distribuição dos pesos por reexecução múltipla sob sementes distintas, reportando intervalos em vez de valores pontuais.

As 14 *features* organizam-se em cinco dimensões conceituais, detalhadas na Seção 7.2: Vitalidade Psicoemocional (VPE), Energia e Engajamento (EEE), Capacidade de Enfrentamento Acadêmico (CEA), Suporte Social e Ambiental (SSA) e Autorregulação e Resiliência (ARR). As siglas são utilizadas na Tabela 3 para indicar a dimensão de pertencimento de cada *feature*.

Tabela 3 – Derivação dos pesos por triangulação de três fontes (ISBE 2.3)

| **Feature** | **Dim.** | $\|r\|$ | **MI** | **FI** | $\|r\|_n$ | $\mathrm{MI}_n$ | $\mathrm{FI}_n$ | **Comp.** | **Peso** |
|---|---|---|---|---|---|---|---|---|---|
| depression | VPE | 0,751 | 0,659 | 0,064 | 0,955 | 0,741 | 0,206 | 0,634 | 0,341 |
| anxiety_level | VPE | 0,755 | 0,646 | 0,033 | 0,967 | 0,705 | 0,076 | 0,583 | 0,314 |
| self_esteem | VPE | 0,762 | 0,666 | 0,058 | 0,985 | 0,762 | 0,179 | 0,642 | 0,345 |
| sleep_quality | EEE | 0,767 | 0,681 | 0,076 | 1,000 | 0,803 | 0,254 | 0,686 | 0,370 |
| blood_pressure | EEE | 0,407 | 0,750 | 0,252 | 0,000 | 1,000 | 1,000 | 0,667 | 0,359 |
| headache | EEE | 0,730 | 0,585 | 0,034 | 0,896 | 0,532 | 0,079 | 0,502 | 0,271 |
| academic_perf. | CEA | 0,748 | 0,528 | 0,095 | 0,947 | 0,367 | 0,338 | 0,551 | 0,394 |
| future_career | CEA | 0,751 | 0,699 | 0,029 | 0,955 | 0,857 | 0,059 | 0,623 | 0,446 |
| study_load | CEA | 0,646 | 0,399 | 0,017 | 0,664 | 0,000 | 0,007 | 0,224 | 0,160 |
| social_support | SSA | 0,656 | 0,558 | 0,067 | 0,691 | 0,453 | 0,217 | 0,454 | 0,355 |
| safety | SSA | 0,714 | 0,509 | 0,034 | 0,852 | 0,315 | 0,080 | 0,415 | 0,326 |
| basic_needs | SSA | 0,719 | 0,505 | 0,028 | 0,868 | 0,301 | 0,053 | 0,407 | 0,319 |
| bullying | ARR | 0,757 | 0,675 | 0,030 | 0,973 | 0,786 | 0,061 | 0,607 | 0,706 |
| noise_level | ARR | 0,660 | 0,419 | 0,015 | 0,702 | 0,058 | 0,000 | 0,253 | 0,294 |

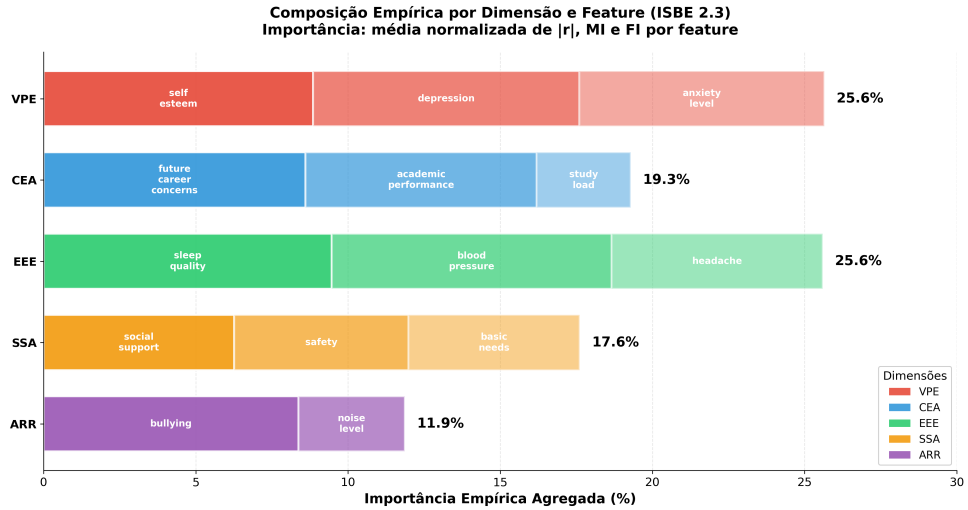
Fonte: Elaborado pelo autor (2026). $|r|$ = correlação absoluta de Pearson; MI = *Mutual Information*; FI = *feature importance* média; $\_n$ = normalizado min-max; Comp. = média das 3 fontes; Peso = peso intradimensional. Pesos derivados da partição de treino do Dataset 1 ($n = 825$); a distribuição do ISBE é avaliada na partição de teste ($n = 275$).

A Tabela 3 evidencia convergência parcial entre as três fontes para a maioria das *features*. Variáveis como `sleep_quality` (compósito = 0,686) e `self_esteem` (0,642) mantêm posições elevadas em $|r|$ e MI, embora com FI relativamente modesta. A principal discrepância envolve `blood_pressure`, que ocupa a última posição em correlação linear ($|r| = 0{,}407$) mas a primeira em MI (0,750) e em FI (0,252). A inclusão da FI como terceira fonte eleva seu compósito para 0,667, valor próximo ao de `sleep_quality` (0,686), reequilibrando sua contribuição intradimensional na dimensão EEE. Esse resultado confirma, de forma quantitativa, o achado da Macrofase III de que a relação de `blood_pressure` com o estresse é melhor capturada por modelos capazes de representar interações não lineares.

No extremo oposto, `study_load` apresenta o menor compósito (0,224), refletindo posições inferiores nas três fontes simultaneamente. Embora integre conceitualmente a dimensão CEA, sua contribuição empírica é substancialmente menor que a de `future_career_concerns` (0,623) e `academic_performance` (0,551), resultado que se traduz em peso intradimensional proporcionalmente reduzido (0,160). A Figura 4 sintetiza a composição empírica das cinco dimensões e das *features* componentes.

Cabe explicitar que os três indicadores de importância ($|r|$, MI e FI) foram derivados do mesmo *dataset* utilizado para treinar os modelos preditivos. Adicionalmente, a FI foi calculada em etapa posterior ao treinamento original. Para quatro dos seis modelos, os melhores hiperparâmetros encontrados pelo *GridSearchCV* não haviam sido preservados, de modo que esses modelos precisaram ser retreinados repetindo o processo de busca a partir dos *grids* originais. Como esse retreinamento pode convergir para configurações ligeiramente diferentes das obtidas no treinamento original, os valores de FI desses quatro modelos carregam uma incerteza adicional. A coerência observada entre as três

Figura 4 – Composição empírica por dimensão e *feature* componente, ISBE 2.3




Fonte: Elaborado pelo autor (2026). Importância calculada como média normalizada de $|r|$ (correlação de Pearson), *Mutual Information* e *Feature Importance* por *feature*, agregada por dimensão.

fontes deve ser lida, portanto, como consistência interna do procedimento de derivação, e não como validação independente. A confirmação de que essa estrutura captura padrões generalizáveis requer avaliação em dados externos, conduzida na Seção 7.3.

## 7.2 Construção do ISBE

O Indicador Sintético de Bem-Estar Estudantil transforma padrões complexos identificados por modelos de *machine learning* em uma métrica na escala de 0 a 100. Sua construção resultou de um refinamento progressivo, ou seja: (1) a versão inicial (1.0) definiu a arquitetura dimensional e a seleção *data-driven* de *features* via *Mutual Information* e (2) as versões intermediárias (2.0, 2.1 e 2.2) corrigiram a inversão de polaridade (que, na versão 2.0, produzia alfas negativos em três dimensões), substituíram métricas de confiabilidade inadequadas e acrescentaram análise fatorial exploratória (EFA) e análise de sensibilidade dos pesos. A versão aqui reportada, ISBE 2.3, incorpora como avanço central a *feature importance* agregada dos modelos da Macrofase III como terceira fonte de triangulação na derivação dos pesos (Seção 7.1), completando a integração entre as etapas preditiva e psicométrica.

O ISBE agrega cinco dimensões conceituais: (1) Vitalidade Psicoemocional (VPE), que captura ansiedade, depressão e autoestima; (2) Energia e Engajamento (EEE), que integra qualidade do sono, ativação fisiológica e sintomas somáticos; (3) Capacidade de Enfrentamento Acadêmico (CEA), que reúne desempenho, carga de estudos e preocupações com a carreira; (4) Suporte Social e Ambiental (SSA), que operacionaliza relações interpessoais e adequação ambiental; e (5) Autorregulação e Resiliência (ARR), que captura a exposição a estressores interpessoais e ambientais. O *score* é calculado em três etapas sequenciais que transformam os valores brutos das *features* em valor interpretável na escala 0–100.

### 7.2.1 Normalização

Todas as *features* são normalizadas pelo procedimento min-max para o intervalo $[0, 1]$, padronizando escalas heterogêneas. Para *features* em que valores elevados indicam piores condições de bem-estar (e.g., `anxiety_level`, `depression`, `bullying`, `study_load`), aplica-se inversão de polaridade pela transformação $(1 - \text{valor normalizado})$, assegurando que *scores* mais altos sempre representem melhor bem-estar.

O conjunto de 14 *features* que compõe o ISBE resultou da redução da versão inicial do indicador (1.0), que reunia 17 *features* distribuídas pelas cinco dimensões. Por meio de seleção via *Mutual Information*, seis *features* de baixa contribuição informacional foram removidas e três de maior relevância foram incorporadas, resultando nas 14 atuais. Na dimensão EEE, `breathing_problem` e `extracurricular_activities` deram lugar a `blood_pressure` e `headache`; na dimensão SSA, `teacher_student_relationship` e `living_conditions` foram removidas; na dimensão ARR, `mental_health_history` e `peer_pressure` foram substituídas por `noise_level`. A versão 2.3 mantém esse mesmo conjunto de 14 *features*, alterando apenas os pesos pela incorporação da *feature importance* como terceira fonte de triangulação.

### 7.2.2 Agregação Dimensional

As *features* normalizadas são agregadas em cinco dimensões por médias ponderadas intradimensionais, com pesos derivados da triangulação descrita na Seção 7.1:

- Vitalidade Psicoemocional (VPE) integra `depression` (peso 0,341; invertido), `anxiety_level` (0,314; invertido) e `self_esteem` (0,345). Peso dimensional: 0,256.
- Energia e Engajamento (EEE) combina `sleep_quality` (0,370), `blood_pressure` (0,359) e `headache` (0,271; invertido). Peso dimensional: 0,256.
- Capacidade de Enfrentamento Acadêmico (CEA) reúne `academic_performance` (0,394), `future_career_concerns` (0,446; invertido) e `study_load` (0,160; invertido). Peso dimensional: 0,193.
- Suporte Social e Ambiental (SSA) agrega `social_support` (0,355), `safety` (0,326) e `basic_needs` (0,319). Peso dimensional: 0,176.
- Autorregulação e Resiliência (ARR) combina `bullying` (0,706; invertido) e `noise_level` (0,294; invertido). Peso dimensional: 0,119.

### 7.2.3 *Score* Final

O *score* ISBE é obtido pela soma ponderada das cinco dimensões, multiplicada por 100:

$$\text{ISBE} = 100 \times (0{,}256 \times \text{VPE} + 0{,}256 \times \text{EEE} + 0{,}193 \times \text{CEA} + 0{,}176 \times \text{SSA} + 0{,}119 \times \text{ARR}) \quad (2)$$

O resultado situa-se no intervalo $[0, 100]$, onde 0 indica pior e 100 indica melhor bem-estar. Para orientar interpretação preliminar, foram definidas cinco categorias provisórias ancoradas pelo *threshold* do WHO-5 (52%): Crítico (0–39), Reduzido (40–51), Moderado (52–64), Bom (65–79) e Excelente (80–100), conforme a Tabela 4. Esses *cutoffs* são aproximações teóricas, não limiares clinicamente validados.

Tabela 4 – Categorização Clínica do ISBE

| **Categoria** | **Faixa** | **Interpretação** | **Ação Recomendada** |
|---|---|---|---|
| Crítico | 0–39 | Distress severo, múltiplos estressores | Encaminhamento imediato |
| Reduzido | 40–51 | Indicadores significativos de distress | Avaliação psicológica detalhada |
| Moderado | 52–64 | Funcionamento adequado com vulnerabilidades | Intervenções preventivas |
| Bom | 65–79 | Recursos adaptativos ativos | Manutenção |
| Excelente | 80–100 | Florescimento psicológico | Modelo de adaptação positiva |

Fonte: Elaborado pelo autor (2026).

### 7.2.4 Distribuição empírica no Dataset 1

O Dataset 1 constitui o cenário de aplicação ideal porque a estrutura dimensional do ISBE foi construída a partir de suas *features*, eliminando ambiguidades interpretativas. A aplicação ao conjunto de teste ($n = 275$) produziu distribuição com média de 51,00 pontos (DP = 23,46), mediana de 55,45 e amplitude de 11,89 a 87,62 (Tabela 5). A assimetria de $-0{,}09$ indica distribuição aproximadamente simétrica com leve deslocamento negativo, e a curtose de $-1{,}27$ sugere distribuição platicúrtica, consistente com população heterogênea.

Tabela 5 – Estatísticas Descritivas do ISBE – Dataset 1

| **Métrica** | **Valor** |
|---|---|
| $n$ | 275 |
| Média ($\mu$) | 51,00 |
| Mediana | 55,45 |
| Desvio Padrão ($\sigma$) | 23,46 |
| Mínimo | 11,89 |
| Máximo | 87,62 |
| Assimetria | $-0{,}09$ |
| Curtose | $-1{,}27$ |

Fonte: Elaborado pelo autor (2026).

A prevalência por categoria foi: Crítico 31,6% ($n = 87$), Reduzido 11,3% ($n = 31$), Moderado 32,0% ($n = 88$), Bom 2,2% ($n = 6$) e Excelente 22,9% ($n = 63$), como mostra a Tabela 6. A distribuição concentra-se em três faixas (Crítico, Moderado e Excelente), com representação reduzida nas categorias intermediárias. Esse padrão trimodal sugere que o ISBE tende a agrupar estudantes em perfis distintos de bem-estar, com transições abruptas entre categorias. A sub-representação nas faixas Reduzido e Bom pode refletir descontinuidades nas *features* originais ou inadequação dos *cutoffs* para esta população, aspecto que requer investigação em amostras independentes.

A análise dimensional revelou padrões informativos: EEE apresentou a média mais baixa (47,31; DP = 26,84), seguida por ARR (47,67; DP = 26,36). SSA obteve o *score* mais elevado (56,84; DP = 26,11), indicando que suporte social e ambiental é o recurso

Tabela 6 – Prevalência por Categoria – Dataset 1

| **Categoria** | **Faixa** | **n** | **%** |
|---|---|---|---|
| Crítico | 0–39 | 87 | 31,6% |
| Reduzido | 40–51 | 31 | 11,3% |
| Moderado | 52–64 | 88 | 32,0% |
| Bom | 65–79 | 6 | 2,2% |
| Excelente | 80–100 | 63 | 22,9% |

Fonte: Elaborado pelo autor (2026).

mais preservado na amostra. VPE (53,28; DP = 24,89) e CEA (49,56; DP = 25,05) posicionaram-se próximos à média global.

### 7.2.5 Capacidade discriminativa

A comparação entre grupos de estresse conhecidos revelou gradiente consistente: estudantes com baixo estresse ($n = 93$) apresentaram ISBE médio de 72,50 (DP = 16,91); estresse médio ($n = 90$), 55,75 (DP = 7,53); alto estresse ($n = 92$), 24,61 (DP = 10,62), conforme a Tabela 7. A diferença de 47,89 pontos entre grupos extremos indica separação ampla. A ANOVA *one-way* confirmou diferenças significativas ($F = 356{,}42$; $p < 0{,}001$; $\eta^2 = 0{,}724$), com 72,4% da variância no ISBE explicada pelo nível de estresse. O Cohen's $d$ entre grupos extremos foi de 3,387.

Tabela 7 – ISBE por Grupo de Estresse – Dataset 1

| **Grupo** | **n** | **ISBE Médio** | **Desvio Padrão** |
|---|---|---|---|
| Baixo Estresse | 93 | 72,50 | 16,91 |
| Médio Estresse | 90 | 55,75 | 7,53 |
| Alto Estresse | 92 | 24,61 | 10,62 |

Fonte: Elaborado pelo autor (2026).

A variável-critério (`stress_level`) foi derivada das mesmas *features* utilizadas no cálculo do ISBE. Consequentemente, esta etapa não constitui avaliação de validade discriminante, mas estritamente uma verificação de coerência construtiva: os resultados de ANOVA e Cohen's $d$ demonstram apenas que a ponderação dimensional preserva a separação entre grupos já codificada nos dados originais, e não capacidade de discriminar contra critério externo independente. A validade discriminante no sentido estrito requer comparação com diagnóstico clínico ou instrumentos como a PSS-14, o WHO-5 ou o *Maslach Burnout Inventory-Student Survey* (MBI-SS), e é parcialmente endereçada pela avaliação externa utilizando o Dataset 2 (Seção 7.3).

## 7.3 Avaliação de Desempenho do ISBE

A avaliação de desempenho consiste no conjunto de procedimentos estatísticos empregados para examinar a qualidade de um instrumento de medida, investigando propriedades como confiabilidade, validade, sensibilidade e capacidade de generalização. Seu

objetivo é verificar se o instrumento produz medidas consistentes e representa adequadamente o construto que se propõe a avaliar.

A avaliação de desempenho do ISBE seguiu diretrizes para indicadores compostos, estruturando-se em seis componentes complementares: avaliação de confiabilidade, análise fatorial exploratória, análise de sensibilidade, capacidade discriminativa, validação em uma amostra externa em um *dataset* independente e comparação com um agregado parcimonioso. Cada componente apoia-se em indicadores consolidados na literatura, definidos a seguir no ponto em que são introduzidos.

A consistência interna foi avaliada pelo alfa de Cronbach ($\alpha$) (Cronbach, 1951), coeficiente que estima a confiabilidade de um instrumento a partir da intercorrelação média entre seus itens e do número de itens, variando de 0 a 1, valores mais altos indicam que os itens medem de forma coerente um mesmo constructo latente. Formalmente,

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum_{i=1}^{k}\sigma_i^2}{\sigma_t^2}\right), \tag{3}$$

em que $k$ é o número de itens, $\sigma_i^2$ é a variância do item $i$ e $\sigma_t^2$ é a variância do escore total. Adotaram-se os limiares convencionais, $\alpha \geq 0{,}90$ (excelente), $\geq 0{,}80$ (bom) e $\geq 0{,}70$ (aceitável), calculando-se o coeficiente para o indicador global e para cada uma das cinco dimensões.

Como o $\alpha$ pressupõe contribuições iguais entre itens (tau-equivalência) e tende a subestimar a confiabilidade em dimensões com poucos itens, calculou-se também a Confiabilidade Composta (*Composite Reliability*, CR), que pondera cada item por sua carga fatorial e é mais robusta quando os itens contribuem de forma heterogênea. A confiabilidade por metades (*split-half*) foi estimada dividindo o instrumento em duas metades e correlacionando seus escores; como a correlação bruta subestima a confiabilidade do teste completo, aplicou-se a correção de Spearman-Brown, $r_{SB} = 2r/(1+r)$, que projeta a confiabilidade para o comprimento integral do instrumento. Por fim, o Erro Padrão de Medida (*Standard Error of Measurement*, SEM) quantifica a imprecisão esperada de um escore individual, derivando do desvio padrão e da confiabilidade global por SEM $= \sigma\sqrt{1-\alpha}$; quanto menor o SEM, mais preciso o escore atribuído a cada estudante. O SEM foi reportado como percentual da escala 0–100 para facilitar a interpretação.

Antes da extração de fatores, verificou-se se os dados eram fatoráveis. O índice Kaiser-Meyer-Olkin (KMO) mede a proporção de variância compartilhada entre as variáveis, variando de 0 a 1, com valores acima de 0,80 indicando que as correlações são suficientemente difusas para sustentar a análise fatorial. O teste de esfericidade de Bartlett avalia a hipótese de que a matriz de correlação é uma identidade; sua rejeição ($p < 0{,}001$) confirma a existência de estrutura correlacional explorável. O número de fatores a reter foi determinado convergindo dois critérios: o critério de Kaiser, que retém fatores com autovalor maior que 1, e a análise paralela, que retém apenas os fatores cujos autovalores observados superam os obtidos de dados aleatórios.

A capacidade de triagem foi avaliada pela área sob a curva ROC (*Receiver Operating Characteristic*), a AUC-ROC. Como a curva ROC é definida para problemas binários, as três classes de estresse foram dicotomizadas de forma alinhada ao propósito de triagem: no Dataset 1, contrastou-se a ausência de estresse contra a presença de estresse médio ou alto; no Dataset 2, o estado de *distress* contra as demais condições. Em ambos os casos, utilizou-se como escore de risco o ISBE invertido ($100 - \text{ISBE}$), uma vez que

valores elevados do indicador representam melhor bem-estar. A AUC resume num único valor entre 0,5 (acaso) e 1,0 (separação perfeita) a relação entre sensibilidade e especificidade; valores acima de 0,80 indicam excelente capacidade de triagem. O ponto de corte ótimo foi determinado pelo índice de Youden ($J$ = sensibilidade + especificidade − 1). Para avaliar a dependência dos resultados em relação ao esquema de ponderação, cinco cenários foram comparados: pesos da versão 2.3 (triangulação de três fontes), pesos da versão 2.2 (triangulação de duas fontes), pesos dimensionais iguais (1/5), pesos uniformes para todas as *features* e dimensões, e pesos invertidos (*stress test*). Por fim, a aplicação ao Dataset 2 constituiu teste de generalização contra critério externo, e a comparação com um agregado parcimonioso, construído sobre o mesmo conjunto de 14 *features* sem agregação dimensional nem ponderação, isolou a contribuição específica da arquitetura do ISBE.

### 7.3.1 Confiabilidade e estrutura fatorial (Dataset 1)

A avaliação de confiabilidade do ISBE 2.3 (14 itens, 5 dimensões) empregou um conjunto de métricas complementares aplicadas ao Dataset 1 ($n = 275$). O alfa de Cronbach global foi de 0,941 (intervalo de confiança, IC 95%: 0,930–0,951), classificado como excelente (Tabela 8). Três dimensões situam-se acima do limiar de 0,70 (VPE, CEA, SSA), enquanto duas ficam abaixo: ARR é penalizada pelo número reduzido de itens ($k = 2$), configuração que sabidamente subestima $\alpha$; EEE apresenta heterogeneidade entre seus componentes (`sleep_quality`, `blood_pressure`, `headache`), cujas correlações interitem são mais fracas. A *Composite Reliability* (CR) convergiu com os valores de $\alpha$ em todas as dimensões (diferenças $\leq 0{,}039$), confirmando tau-equivalência aproximada. A CR de EEE (0,683) superou o $\alpha$ correspondente (0,644), sugerindo que a ponderação por cargas fatoriais compensa parcialmente a heterogeneidade dos itens nessa dimensão.

Tabela 8 – Confiabilidade por dimensão ISBE

| **Dimensão** | **k** | **$\alpha$ de Cronbach** | **IC 95%** | **CR** | **Classif.** |
|---|---|---|---|---|---|
| VPE | 3 | 0,842 | 0,807–0,872 | 0,843 | Boa |
| CEA | 3 | 0,793 | 0,747–0,832 | 0,794 | Aceitável |
| SSA | 3 | 0,795 | 0,749–0,834 | 0,804 | Aceitável |
| ARR | 2 | 0,672 | 0,584–0,741 | 0,676 | Questionável[†] |
| EEE | 3 | 0,644 | 0,564–0,711 | 0,683 | Questionável |
| **Global** | **14** | **0,941** | **0,930–0,951** | N/D | **Excelente** |

Fonte: Elaborado pelo autor (2026). $n = 275$. $k$ = número de itens; CR = *Composite Reliability*.
[†]Estimativa instável devido a $k = 2$.

A correlação entre metades aleatórias foi de 0,924, e o coeficiente de Spearman-Brown corrigido foi de 0,960 (IC 95%: 0,950–0,969), indicando que as duas metades do instrumento medem de forma consistente o mesmo constructo latente. O SEM global foi de 5,70 pontos (5,7% da escala 0–100), com IC 95% de ±11,17 pontos. Diferenças individuais superiores a ±11 pontos podem ser interpretadas como mudanças clinicamente significativas, embora essa estimativa deva ser confirmada por estudo de responsividade específico.

A adequação dos dados para análise fatorial foi confirmada pelo índice KMO de 0,951 (classificação: meritório) e pelo teste de esfericidade de Bartlett ($\chi^2 = 12.116{,}6$;

$p < 0{,}001$). O critério de Kaiser sugeriu dois fatores (autovalor $> 1$), enquanto a análise paralela reteve um único fator dominante. Optou-se pela extração de dois fatores, que explicaram conjuntamente 64,8% da variância. Três das cinco dimensões teóricas (VPE, CEA e ARR) convergiram para o fator principal, enquanto EEE e SSA apresentaram dispersão entre fatores. Esse resultado sugere que a estrutura empírica observada é mais parcimoniosa do que o modelo pentadimensional proposto.

Os resultados indicam que as cinco dimensões devem ser interpretadas como uma estrutura conceitual fundamentada na literatura e na modelagem preditiva. Ou seja, não como dimensões empiricamente confirmadas pela estrutura fatorial observada nesta amostra. Contudo, essa divergência não invalida necessariamente a formulação multidimensional. Instrumentos amplamente utilizados, como o DASS-21, também apresentam um fator geral dominante coexistindo com subescalas teoricamente justificadas. Os achados sugerem, entretanto, que futuras revisões do ISBE podem considerar alternativas de reorganização dimensional. A confirmação da estrutura proposta requer análise fatorial confirmatória (CFA) em amostra independente.

A análise de sensibilidade dos pesos comparou cinco cenários de ponderação. A variação de AUC entre eles foi de apenas 0,015, e a variação de Cohen's $d$, de 0,241, indicando que o desempenho discriminativo do ISBE não depende criticamente do esquema de ponderação específico. A comparação entre os cenários v2.3 e v2.2 é particularmente informativa: a recalibração dos pesos pela inclusão da FI produziu variação de AUC de apenas $-0{,}002$, com aumento do Cohen's $d$ de $+0{,}097$, confirmando que a incorporação da terceira fonte refina a ponderação sem alterar substancialmente o desempenho. A correlação entre os *scores* individuais das versões 2.2 e 2.3 foi de $r = 0{,}997$, com concordância categórica de 94,5% e diferença média de $+0{,}01$ ponto, evidenciando que a recalibração preservou a estrutura geral do instrumento.

Tabela 9 – Análise de sensibilidade dos pesos dimensionais do ISBE

| **Cenário** | **AUC** | **Cohen's $d$ (B×A)** |
|---|---|---|
| Pesos v2.3 (triangulação 3 fontes) | 0,884 | 3,387 |
| Pesos v2.2 (triangulação 2 fontes) | 0,887 | 3,290 |
| Dimensões iguais (1/5) | 0,888 | 3,471 |
| Tudo igual (dimensões + *features*) | 0,890 | 3,530 |
| Pesos invertidos (*stress test*) | 0,899 | 3,486 |
| **Variação máxima** | **0,015** | **0,241** |

Fonte: Elaborado pelo autor (2026).

Cabe registrar que as versões 2.0 a 2.3 foram calibradas no mesmo conjunto de 275 observações, o que configura risco de capitalização sobre o acaso (Limitação L4, detalhada na Discussão). A avaliação em uma amostra externa, utilizando o Dataset 2, apresentada adiante, fornece evidência parcial de generalização ao demonstrar validade discriminativa contra critério externo, ainda que com degradação de confiabilidade atribuível a diferenças de instrumentação.

### 7.3.2 Concordância preditiva e coerência construtiva (Dataset 1)

A concordância entre os *scores* ISBE e as predições dos modelos de Aprendizado de Máquina foi avaliada sob protocolo de *holdout*, sem validação cruzada: os quatro mo-

delos foram treinados exclusivamente na partição de treino ($n = 825$) e suas predições foram geradas na partição de teste ($n = 275$), a mesma utilizada para o cálculo do ISBE, assegurando que nenhum modelo tivesse contato com os dados de avaliação durante o treinamento. A validação cruzada não foi empregada nesta métrica porque o ISBE da partição de teste é fixo, calculado com pesos derivados da partição de treino, de modo que a concordância deve ser medida sobre um conjunto de avaliação único e independente. Como escore contínuo de cada modelo, adotou-se a esperança das probabilidades preditas para as três classes de estresse; o ISBE foi reescalonado para a mesma direção e amplitude, e a incerteza das correlações foi estimada por intervalos de confiança de 95% via transformação $z$ de Fisher. A correlação média entre os quatro modelos foi de $r = 0{,}926$ (Tabela 10).

Tabela 10 – Concordância preditiva por modelo de *machine learning*

| **Modelo** | $r$ | $r^2$ |
|---|---|---|
| Random Forest | 0,961 | 0,923 |
| Voting Ensemble | 0,937 | 0,878 |
| XGBoost | 0,915 | 0,838 |
| Gradient Boosting | 0,891 | 0,794 |
| **Média** | **0,926** | **0,858** |

Fonte: Elaborado pelo autor (2026).

Essa métrica não constitui validade convergente no sentido psicométrico clássico, dado que os modelos foram treinados nas mesmas *features* utilizadas para calcular o ISBE. O resultado indica que a ponderação dimensional reproduz com fidelidade a estrutura preditiva dos modelos, o que é coerente com o procedimento de derivação, mas não equivale a correlação com critério externo independente. Validade convergente *stricto sensu* requer correlação com instrumentos validados de forma independente (e.g., PSS-14, WHO-5, MBI-SS).

Quanto à triagem, o ISBE 2.3 obteve AUC-ROC de 0,884. No *cutoff* ótimo de 63,8 pontos (determinado pelo índice de Youden), a sensibilidade foi de 1,000 (100%) e a especificidade de 0,742 (74,2%). A sensibilidade unitária indica que todos os casos de alto estresse foram corretamente identificados nesse *cutoff*, ao custo de uma taxa de falsos positivos de 25,8%. O Cohen's $d$ entre grupos extremos de estresse (baixo *vs.* alto) foi de 3,387; a separação entre grupos adjacentes também foi substancial: baixo *vs.* médio ($d = 1{,}273$) e médio *vs.* alto ($d = 3{,}376$).

### 7.3.3 Avaliação Externa em Amostra Independente (Dataset 2)

A aplicação do ISBE ao Dataset 2 (Singh, 2024) constitui validação externa em amostra independente, teste essencial para avaliar a generalização do indicador para além do *dataset* de derivação. Diferentemente de uma abordagem de cobertura parcial, o mapeamento conceitual revisado permitiu identificar correspondências defensáveis para todas as 14 *features* do ISBE nas 25 variáveis do Dataset 2, assegurando cobertura completa das cinco dimensões (Tabela 11). Uma distinção metodológica fundamental diferencia esta análise da realizada no Dataset 1 é que nesse experimento a variável-critério (tipo de estresse: sem estresse, *eustress*, *distress*) é **externa** às *features* utilizadas no cálculo do ISBE, eliminando a circularidade parcial (limitação L2) que afeta as métricas discri-

minativas do Dataset 1. Os resultados discriminativos reportados nesta parte constituem, portanto, evidência de validade contra critério independente.

Tabela 11 – Mapeamento Semântico e Alinhamento de Polaridade entre os *Datasets*

| **Variável no Dataset 1** | **Pergunta correspondente no Dataset 2** | **Polar.** |
|---|---|---|
| depression | Have you been feeling sadness or low mood? | Direta |
| anxiety_level | Have you been dealing with anxiety or tension recently? | Direta |
| self_esteem | Do you lack confidence in your academic performance? | Invert. |
| sleep_quality | Do you face any sleep problems or difficulties falling asleep? | Invert. |
| blood_pressure | Have you noticed a rapid heartbeat or palpitations? | Direta |
| headache | Have you been getting headaches more often than usual? | Direta |
| academic_perf. | Do you have trouble concentrating on your academic tasks? | Invert. |
| future_career | Do you lack confidence in your choice of academic subject? | Direta |
| study_load | Do you feel overwhelmed with your academic workload? | Direta |
| social_support | Do you often feel lonely or isolated? | Invert. |
| safety | Are you facing difficulties with your professors/environment? | Invert. |
| basic_needs | Is your hostel or home environment causing you difficulties? | Invert. |
| bullying | Are you in competition with your peers, and does it affect you? | Direta |
| noise_level | Is your working environment unpleasant or stressful? | Direta |

Fonte: Elaborado pelo autor (2026).

A aplicação ao conjunto completo ($n = 816$) produziu distribuição com média de 60,22 pontos (DP = 12,63), mediana de 61,42 e amplitude de 0,00 a 96,47 (Tabela 12). A assimetria de $-0{,}53$ indica leve assimetria negativa, e a curtose de 1,03 sugere distribuição levemente leptocúrtica. Comparada ao Dataset 1 (média = 51,00; DP = 23,46), observa-se deslocamento da média em +9,22 pontos e compressão do desvio padrão ($-10{,}83$), refletindo as diferenças de instrumentação e a distribuição assimétrica da variável *target* do Dataset 2 (91,3% *eustress*).

Tabela 12 – Estatísticas Descritivas do ISBE – Dataset 2

| **Métrica** | **Valor** |
|---|---|
| $n$ | 816 |
| Média ($\mu$) | 60,22 |
| Mediana | 61,42 |
| Desvio Padrão ($\sigma$) | 12,63 |
| Mínimo | 0,00 |
| Máximo | 96,47 |
| Assimetria | $-0{,}53$ |
| Curtose | 1,03 |

Fonte: Elaborado pelo autor (2026).

A prevalência por categoria concentrou-se nas faixas intermediárias: Moderado 40,2% ($n = 328$) e Bom 49,6% ($n = 405$), somando 89,8% da amostra (Tabela 13). As categorias extremas foram sub-representadas: Crítico 0,4% ($n = 3$), Reduzido 5,9% ($n = 48$) e Excelente 3,9% ($n = 32$). Essa concentração central é coerente com o desbalanceamento da variável *target* (91,3% *eustress*) e indica que o ISBE atribuiu *scores* intermediários à vasta maioria dos estudantes classificados como *eustress*, diferenciando-os adequadamente dos extremos.

Tabela 13 – Prevalência por Categoria – Dataset 2

| **Categoria** | **Faixa** | **n** | **%** |
|---|---|---|---|
| Crítico | 0–39 | 3 | 0,4% |
| Reduzido | 40–51 | 48 | 5,9% |
| Moderado | 52–64 | 328 | 40,2% |
| Bom | 65–79 | 405 | 49,6% |
| Excelente | 80–100 | 32 | 3,9% |

Fonte: Elaborado pelo autor (2026).

A análise dimensional revelou padrões coerentes com a instrumentação do Dataset 2. Isso porque, ARR apresentou a média mais elevada (63,26; DP = 24,71), seguida por SSA (63,63; DP = 19,72). Além disso, a EEE obteve a média mais baixa (56,58; DP = 19,63). VPE (60,82; DP = 19,06) e CEA (59,26; DP = 21,39) posicionaram-se próximos à média global.

A comparação entre os grupos de estresse originais do Dataset 2 revelou gradiente robusto e monotônico. Isso é, sem estresse ($n = 42$) apresentaram ISBE médio de 81,18 (DP = 6,00); *eustress* ($n = 745$), 60,26 (DP = 10,29); *distress* ($n = 29$), 28,86 (DP = 10,16), conforme a Tabela 14. A diferença de 52,32 pontos entre grupos extremos superou a separação obtida no Dataset 1 (47,89 pontos) e, crucialmente, opera com critério externo independente.

Tabela 14 – ISBE por Grupo de Estresse – Dataset 2 (critério externo)

| **Grupo** | **n** | **ISBE Médio** | **Desvio Padrão** |
|---|---|---|---|
| Sem estresse | 42 | 81,18 | 6,00 |
| Eustress | 745 | 60,26 | 10,29 |
| Distress | 29 | 28,86 | 10,16 |

Fonte: Elaborado pelo autor (2026).

A ANOVA *one-way* confirmou diferenças significativas ($F = 229{,}81$; $p < 0{,}001$; $\eta^2 = 0{,}361$), com 36,1% da variância no ISBE explicada pelo tipo de estresse externo. O Cohen's $d$ entre grupos extremos foi de 6,576, efeito muito grande. A separação entre grupos adjacentes também foi substancial, uma vez que, sem estresse *vs. eustress* ($d = 2{,}070$) e *eustress vs. distress* ($d = 3{,}054$). A AUC-ROC atingiu 0,990. Esse valor, contudo, deve ser interpretado com cautela e não como "capacidade discriminativa quase perfeita". A tarefa de triagem no Dataset 2 é fortemente desbalanceada: a classe *distress* representa apenas 3,8% da amostra e, mesmo na dicotomização adotada, a classe positiva permanece minoritária. Sob essa prevalência, a AUC-ROC tende a ser otimista, porque a taxa de falsos positivos que compõe seu eixo é calculada sobre a classe majoritária, muito numerosa, o que torna a especificidade alta e estável mesmo quando o número absoluto de falsos positivos não é desprezível. Métricas dependentes da prevalência, como a própria AUC-ROC global e a acurácia, são, portanto, insuficientes para caracterizar o desempenho. Para uma leitura honesta, é necessário acompanhar a AUC-ROC de indicadores insensíveis ao desbalanceamento, em particular a área sob a curva de Precisão-Recall (AUC-PR) e o *score* F1-Macro, que ponderam igualmente as classes minoritária e majoritária. A curva

de Precisão-Recall é especialmente informativa neste cenário, pois explicita o *trade-off* entre sensibilidade e valor preditivo positivo justamente na faixa de operação de uma ferramenta de triagem, na qual o custo de falsos positivos sobre uma base majoritariamente saudável é elevado. As métricas *macro* dos classificadores no Dataset 2, reportadas no Apêndice A.3 (Tabela 24), evidenciam essa diferença: enquanto a acurácia se mantém acima de 94% para todos os modelos, o F1-Macro cai de forma acentuada (por exemplo, 68,8% no *Bagging* e 72,3% no *Random Forest*), confirmando que o ganho real sobre o piso trivial de 91,1% é substancialmente menor do que a acurácia e a AUC-ROC isoladamente sugerem.

A consistência interna global no Dataset 2 foi de $\alpha = 0{,}606$ (IC 95%: 0,565–0,645), classificação questionável. Todas as dimensões individuais apresentaram alfas inaceitáveis (VPE = 0,172; EEE = 0,303; CEA = 0,248; SSA = 0,270; ARR = 0,342), com *Composite Reliability* convergente (diferenças $\leq 0{,}005$). A confiabilidade por metades produziu $r = 0{,}453$ e Spearman-Brown corrigido de 0,623 (IC 95%: 0,568–0,672). O SEM global foi de 7,93 pontos (7,9% da escala). A queda substancial em relação ao Dataset 1 ($\Delta\alpha = -0{,}335$) é esperada e interpretável, uma vez que o Dataset 2 emprega instrumentação diferente (variáveis focadas em sintomas psicossomáticos e estressores contextuais, em escalas e formulações distintas das do Dataset 1), de modo que o mapeamento entre variáveis, embora conceitualmente defensável, não assegura equivalência de mensuração. As correlações interitem são necessariamente mais fracas porque os itens não foram desenhados como escalas coerentes dentro das dimensões do ISBE. A Tabela 15 sintetiza a comparação das métricas entre os dois *datasets*.

Tabela 15 – Comparação de Métricas entre Datasets 1 e 2

| **Métrica** | **Dataset 1** | **Dataset 2** | **Δ** | **Tipo de evidência** |
|---|---|---|---|---|
| $\alpha$ Global | 0,941 | 0,606 | −0,335 | Interna |
| Split-half (SB) | 0,960 | 0,623 | −0,337 | Interna |
| SEM (% escala) | 5,7% | 7,9% | +2,2 pp | Interna |
| Cohen's $d$ (extr.) | 3,387 | 6,576 | +3,189 | D2: critério externo |
| AUC-ROC | 0,884 | 0,990 | +0,106 | D2: critério externo |

Fonte: Elaborado pelo autor (2026).

A análise fatorial exploratória com extração de quatro fatores revelou estrutura parcialmente alinhada ao modelo pentadimensional. O Fator 1 agregou variáveis de ambiente e suporte (`study_load`, `noise_level`, `bullying`, `safety`, `social_support`), alinhando-se às dimensões SSA e parcialmente a CEA. O Fator 2 concentrou `self_esteem`, `future_career_concer` e `basic_needs`. O Fator 3 agrupou os atributos`depression` e `academic_performance`. O Fator 4 associou `anxiety_level` e `sleep_quality`. A dispersão entre fatores é coerente com a queda de consistência interna: o mapeamento conceitual preserva a estrutura teórica, mas a instrumentação diferente fragmenta a coerência fatorial.

A análise no Dataset 2 estabelece três achados complementares. Primeiro, o ISBE demonstra validade discriminativa contra critério externo: a capacidade de separar grupos definidos independentemente ($d = 6{,}576$; AUC = 0,990) constitui evidência que transcende a coerência construtiva do Dataset 1, mitigando parcialmente a limitação L2. Segundo, a consistência interna degrada-se substancialmente ($\alpha = 0{,}606$) quando o ins-

trumento é operacionalizado com variáveis de instrumentação diferente, confirmando que o ISBE não é agnóstico ao instrumento de coleta; aplicações prospectivas devem replicar formulações equivalentes às *features* originais. Terceiro, apesar da queda de confiabilidade, o ISBE retém capacidade discriminativa robusta no nível populacional, sugerindo que a estrutura dimensional e o esquema de ponderação capturam relações substantivas entre os constructos, não artefatos do Dataset 1.

### 7.3.4 Comparação com agregado parcimonioso

Para avaliar a contribuição específica da arquitetura dimensional, conduziu-se análise comparativa contra um agregado construído sobre o mesmo conjunto de 14 *features* selecionadas, com polaridade alinhada e normalização Min-Max, porém sem agregação dimensional ou ponderação por triangulação. Esse *baseline* responde à pergunta: dado o mesmo conjunto de itens, qual o ganho líquido da arquitetura pentadimensional do ISBE em relação a um agregado simples?

Tabela 16 – Comparação do ISBE com agregados de referência no Dataset 1 ($n = 275$)

| **Indicador** | **ISBE 2.3** | **Agr. 20 itens** | **Agr. 14 itens** |
|---|---|---|---|
| $\alpha$ de Cronbach | **0,941** | 0,861 | 0,900 |
| Split-half (SB) | **0,960** | 0,906 | 0,914 |
| SEM (% escala) | **5,7%** | 10,8% | 9,4% |
| Cohen's $d$ (extremos) | 3,387 | 3,235 | **3,522** |
| AUC-ROC | 0,884 | **0,946** | 0,931 |

Fonte: Elaborado pelo autor (2026). Em negrito, o melhor valor de cada linha.

A decomposição do ganho de consistência interna (Tabela 17) evidencia que a maior parcela decorre de pré-processamento elementar. Aproximadamente 95% do ganho total de $\alpha$ decorre do alinhamento de polaridade, operação exigida por qualquer prática psicométrica responsável. O ganho atribuível à arquitetura pentadimensional do ISBE é de +0,041 em $\alpha$ e, principalmente, da redução do SEM de 9,4% para 5,7% da escala (queda de 39,4%). Em contrapartida, observa-se perda de 3,8% em Cohen's $d$ e 5,0% em AUC em relação ao agregado equivalente. A avaliação em uma amostra externa no Dataset 2 reforça esse padrão (Tabela 18): naquele *dataset*, os itens são Likert uniformes 1–5, eliminando o problema de escalas heterogêneas, e o agregado parcimonioso iguala ou supera o ISBE em todas as métricas.

Tabela 17 – Decomposição do ganho de $\alpha$ no Dataset 1

| **Componente** | $\Delta\alpha$ | **Atribuição** |
|---|---|---|
| Alinhamento de polaridade | +1,522 | Pré-processamento elementar |
| Seleção de 14 itens via MI | +0,039 | Parcimônia *data-driven* |
| Agregação dimensional e ponderação | +0,041 | Arquitetura do ISBE |
| **Ganho total** | **+1,602** | Total |

Fonte: Elaborado pelo autor (2026).

Tabela 18 – Comparação do ISBE com agregado equivalente no Dataset 2

| **Indicador** | **ISBE 2.3** | **Agr. 14 itens mapeados** |
|---|---|---|
| $\alpha$ de Cronbach | 0,606 | 0,617 |
| Split-half (SB) | 0,623 | 0,668 |
| SEM (% escala) | 7,9% | 8,3% |
| Cohen's $d$ (No Stress *vs* Distress) | 6,576 | 7,292 |
| AUC-ROC | 0,990 | 1,000 |

Fonte: Elaborado pelo autor (2026).

A leitura conjunta dessas tabelas sustenta três conclusões. Primeira, o ISBE não supera um agregado parcimonioso bem construído em métricas de discriminação populacional. Segunda, entrega ganho substancial e específico em redução do erro padrão de medida individual quando a escala possui itens em amplitudes heterogêneas. Terceira, sua contribuição distintiva não reside na maximização de métricas psicométricas clássicas, mas na transformação de um agregado em *score* interpretável e decomposto em cinco dimensões clinicamente acionáveis, propriedade não captada por nenhuma das métricas comparadas. Essa interpretação fundamenta o reenquadramento de uso discutido na Discussão.

### 7.3.5 Consolidação dos resultados

O conjunto de evidências apresentado ao longo desta seção constitui a base empírica sobre a qual se apoiam a interpretação e o posicionamento do ISBE, desenvolvidos na Discussão. A Tabela 19 reúne as propriedades métricas do ISBE 2.3 nos dois *datasets*, e a Tabela 20 sintetiza as diferenças de resolução entre os critérios de estresse originais e o indicador proposto.

Tabela 19 – Consolidação das propriedades métricas do ISBE 2.3

| **Propriedade** | **Dataset 1** ($n = 275$) | **Dataset 2** ($n = 816$) |
|---|---|---|
| $\alpha$ de Cronbach global | 0,941 [0,930–0,951] | 0,606 [0,565–0,645] |
| Split-half (Spearman-Brown) | 0,960 [0,950–0,969] | 0,623 [0,568–0,672] |
| SEM (% da escala) | 5,70% | 7,93% |
| Concordância preditiva ($r$ médio) | 0,926 | N/D |
| Cohen's $d$ (extremos) | 3,387 | 6,576 |
| ANOVA ($\eta^2$) | 0,724 | 0,361 |
| AUC-ROC | 0,884 | 0,990 |
| Sensibilidade / Especificidade | 1,000 / 0,742 | 1,000 / 0,914 |
| KMO | 0,951 | 0,666 |
| Variância explicada (EFA) | 64,8% | 47,9% |
| Sensibilidade dos pesos ($\Delta$AUC) | 0,015 | N/D |

Fonte: Elaborado pelo autor (2026). As métricas discriminativas do D1 operam sob circularidade parcial; as do D2 utilizam critério externo independente. A AUC-ROC do D2 deve ser lida com cautela: sob o forte desbalanceamento da variável-alvo (3,8% *distress*), tende a superestimar o desempenho, devendo ser complementada pela AUC-PR e pelo F1-Macro (Apêndice A.3).

Tabela 20 – Comparação de resolução informacional entre critérios originais e ISBE 2.3

| **Aspecto** | **D1: stress_level** | **D2: Likert** | **ISBE 2.3** |
|---|---|---|---|
| Tipo de escala | Categórica (3 níveis) | Categórica (3 classes) | Contínua (0–100) |
| Dimensionalidade | Unidimensional | Unidimensional | 5 dimensões + global |
| Derivação | Mesmas features | Autorreportada (externa) | Triangulação empírica |
| Info. por indivíduo | 1 rótulo (0, 1, 2) | 1 rótulo | 6 scores |
| Variância intra-grupo | Descartada | Descartada | Preservada |
| Acionabilidade | Nenhuma | Mínima | Perfil dimensional |

Fonte: Elaborado pelo autor (2026).

Por fim, a Tabela 21 posiciona as métricas do ISBE em relação a instrumentos estabelecidos de avaliação de bem-estar e estresse. A coluna 'Tipo de evidência' explicita a natureza dos valores: todas as métricas do ISBE derivam de validação interna no Dataset 1 (parte delas adicionalmente afetada pela circularidade parcial L2), ou da avaliação em uma amostra externa no Dataset 2; os valores dos instrumentos de referência provêm de estudos com validação externa multicêntrica. A comparação deve ser interpretada como contextualização, não como equivalência metodológica.

Tabela 21 – Posicionamento métrico do ISBE em relação a instrumentos estabelecidos

| **Métrica** | **ISBE D1** | **ISBE D2** | **WHO-5** | **PSS-10** | **DASS-21** | **WEMWBS** |
|---|---|---|---|---|---|---|
| $\alpha$ (Cronbach) | 0,941 | 0,606 | 0,82–0,87 | 0,78–0,84 | 0,93–0,96 | 0,89–0,91 |
| Split-half (SB) | 0,960 | 0,623 | N/D | N/D | N/D | N/D |
| SEM (% escala) | 5,7% | 7,9% | 8–9% | ∼8% | 5–6% | ∼6% |
| AUC-ROC | 0,884 | 0,990 | 0,84 | 0,86 | 0,87 | 0,82 |
| Cohen's $d$ (B×A) | 3,387 | 6,576 | 2,31 | 2,45 | 2,60 | 2,10 |
| Sensibilidade | 100% | 100% | 78,9% | 81,2% | 82,3% | 76% |
| Especificidade | 74,2% | 91,4% | 76,4% | 77,1% | 80,1% | 74% |
| ICC (reteste) | N/D | N/D | 0,83 | 0,72–0,85 | 0,84 | 0,83 |
| Nº itens | 14 | 14 | 5 | 10 | 21 | 14 |
| Dimensões | 5 | 5 | 1 | 1 | 3 | 1 |

Fonte: Elaborado pelo autor (2026). Fontes: WHO-5 (Topp *et al.*, 2015); PSS-10 (Lee, 2012; Cohen; Williamson, 1988); DASS-21 (Lovibond; Lovibond, 1995; Henry; Crawford, 2005); WEMWBS (Tennant *et al.*, 2007; Stewart-Brown *et al.*, 2009).

## 7.4 Aspectos éticos e limitações

A pesquisa observou princípios éticos de transparência, anonimização e uso apropriado de dados públicos. Reconhecem-se limitações inerentes à abordagem adotada, entre as quais se destacam: a utilização de dados secundários cujas *features* não foram desenhadas como itens do ISBE; a circularidade parcial entre as *features* do indicador e a variável-critério do Dataset 1; a impossibilidade de inferências causais dada a natureza transversal dos dados; e o refinamento iterativo do indicador em amostra única. A análise completa das limitações metodológicas é desenvolvida na Discussão.

O ISBE destina-se a triagem e monitoramento, não substituindo avaliação diagnóstica por profissionais de saúde mental. Implementações institucionais devem observar proteção de privacidade conforme a legislação vigente e conduzir análises de equidade para prevenir vieses algorítmicos contra subgrupos vulneráveis.

# 8 DISCUSSÃO

Este trabalho percorreu quatro macrofases, da exploração dos dados à construção e validação de um indicador sintético, com o objetivo de investigar se técnicas de aprendizado de máquina podem fundamentar empiricamente a construção de instrumentos de avaliação de bem-estar estudantil. As seções seguintes contextualizam os achados em relação à literatura, examinam as implicações metodológicas e delimitam o alcance das evidências produzidas.

## 8.1 Desempenho preditivo e comparação com a literatura

Os resultados da Macrofase III situam-se dentro da faixa reportada pela literatura recente sobre classificação de estresse estudantil. No Dataset 1, o melhor *ensemble* (*Voting Classifier hard*) atingiu 93,1% de acurácia com distribuição equilibrada entre três classes. No Dataset 2, o *Stacking Classifier* alcançou 99,5%. Esses valores são consistentes com os reportados por Ovi *et al.* (2025), cuja abordagem *context-aware* com os mesmos *datasets* produziu acurácias superiores a 93% no Dataset 1 e a 99% no Dataset 2 com estratégias de *stacking*.

Dois aspectos merecem ponderação. Primeiro, as acurácias elevadas no Dataset 2 devem ser interpretadas com cautela em razão do desbalanceamento da variável-alvo: 91,3% dos registros pertencem a uma única classe (*Eustress*), de modo que um classificador trivial que atribuísse essa classe a todos os casos alcançaria acurácia de 91,3%. Segundo, os resultados no Dataset 1, com classes equilibradas, são mais informativos sobre a capacidade real dos modelos. Cabe, ainda, uma ressalva interpretativa quanto ao Dataset 1: os preditores e a variável-alvo `stress_level` medem facetas de um mesmo construto, de modo que a elevada acurácia é, em parte, esperada pela sobreposição conceitual entre as variáveis, e não decorre de circularidade computacional, visto que `stress_level` não constitui função determinística das *features* (o teto de acurácia em torno de 92%, distante de 100%, confirma a existência de variância não explicada pelos preditores). Esse aspecto reforça que os resultados do Dataset 1 atestam a consistência informacional da base, mas não constituem, isoladamente, evidência forte de capacidade preditiva sobre uma tarefa difícil.

A comparação com Meng *et al.* (2024), que aplicaram *Random Forest* a uma amostra internacional de 61.585 estudantes para predizer *scores* WHO-5, revela uma diferença de abordagem relevante: enquanto aquele estudo tratou o bem-estar como variável contínua, o presente trabalho abordou o estresse como classificação multiclasse. Ambos convergem na identificação de métodos de *ensemble* como abordagens com desempenho consistente para dados tabulares de autorrelato. A diferença de escala amostral (275–816 *vs.* 61.585) é uma limitação relevante deste trabalho.

## 8.2 A triangulação como ponte entre modelagem preditiva e psicometria

A contribuição central deste trabalho reside não nos resultados preditivos *per se*, mas na proposição de um *pipeline* que transforma evidências de *machine learning* em um indicador psicometricamente avaliável. O procedimento de triangulação, combinando correlação de Pearson, *Mutual Information* e *feature importance* agregada, representa uma tentativa de superar a lacuna apontada por Panicheva *et al.* (2022) e Eder *et al.* (2025) entre a capacidade preditiva dos modelos e a produção de métricas clinicamente interpretáveis.

O caso de `blood_pressure` ilustra o valor da abordagem. Essa *feature* ocupa a última posição em correlação linear ($|r| = 0{,}407$) mas a primeira em MI (0,750) e FI (0,252). Um procedimento baseado exclusivamente em correlação a subponderaria; um baseado apenas em FI a sobreponderaria. A média das três fontes normalizadas (compósito = 0,667) produziu peso intermediário, coerente com a interpretação de que a relação dessa variável com o estresse é predominantemente não linear. Esse resultado confirma, por via distinta, o achado de Graham *et al.* (2019) de que abordagens puramente teóricas não capturam plenamente estruturas latentes identificáveis por análises *data-driven*.

Entretanto, é necessário distinguir entre o que o *pipeline* demonstra e o que ele não demonstra. O *pipeline* demonstra que é possível derivar pesos dimensionais empiricamente sobre dados de autorrelato e produzir um indicador com consistência interna elevada, capacidade de separação entre grupos e perfil dimensional clinicamente interpretável. A comparação adicional contra um agregado parcimonioso sobre as mesmas 14 *features*, reportada na Seção 7.3, qualifica esse achado: o ganho do ISBE em consistência interna sobre o *baseline* equivalente é marginal ($\Delta\alpha \approx +0{,}041$), mas a redução do erro padrão de medida é substancial (de 9,4% para 5,7% da escala, redução de 39%). Em contrapartida, o ISBE apresenta pequena perda em métricas de discriminação populacional (Cohen's $d$ e AUC), evidenciando um *trade-off* explícito entre precisão individual e separação de grupos. O *pipeline* não demonstra que o ISBE mede bem-estar estudantil de forma válida no sentido psicométrico clássico, para isso, seria necessária validação convergente com instrumentos estabelecidos, estabilidade temporal e aplicação prospectiva com itens dedicados. Essa distinção decorre de um conjunto de limitações metodológicas discutidas na Seção 8.4.

## 8.3 Confiabilidade e validade discriminativa: o que o Dataset 2 revelou

A avaliação na amostra externa utilizando o Dataset 2 produziu um resultado inesperado e metodologicamente informativo: a confiabilidade degradou-se substancialmente ($\alpha$ de 0,941 para 0,606), mas a capacidade discriminativa melhorou (AUC de 0,884 para 0,990; Cohen's $d$ de 3,387 para 6,576). Essa dissociação admite interpretação coerente.

O alfa de Cronbach mede a covariância entre itens. No Dataset 1, as 14 *features* apresentam correlações interitem elevadas (KMO = 0,951), refletindo a estrutura específica daquele *survey*. No Dataset 2, perguntas diferentes medem construtos semanticamente análogos mas não idênticos, introduzindo heterogeneidade que reduz as correlações interitem e, por consequência, o alfa. Isso não significa que o ISBE funciona pior no Dataset 2; significa que a consistência interna, uma propriedade da relação entre itens, é sensível à instrumentação de forma que a capacidade discriminativa, uma propriedade da relação entre o índice e um critério externo, não é.

O ganho epistemológico central do Dataset 2 reside no critério utilizado. No Dataset 1, a variável `stress_level` é derivada das mesmas *features* que compõem o ISBE, de modo que as métricas ali obtidas configuram uma verificação de coerência construtiva, e não evidência de validade discriminante: demonstram apenas que a ponderação dimensional preserva a separação já codificada nos dados. No Dataset 2, o critério (tipo de estresse autorreportado: sem estresse, *eustress*, *distress*) é coletado independentemente das *features* mapeadas. A AUC de 0,990 e o Cohen's $d$ de 6,576 constituem evidência de que o ISBE discrimina grupos cuja definição não depende do índice, algo que os resultados do Dataset 1 não podiam demonstrar.

Esse resultado ganha contorno adicional quando comparado com a resolução informacional dos critérios originais de cada *dataset*. No Dataset 1, a variável `stress_level` condensa 20 *features* em três categorias discretas (0, 1, 2), descartando toda a variância intracategórica: dois estudantes classificados como `stress_level = 1` podem apresentar *scores* ISBE de 48 e 62, com perfis dimensionais distintos, por exemplo, VPE comprometido com CEA preservado *versus* o padrão inverso. Essa diferença não é acessível pelo rótulo original. O ISBE, ao operar como *score* contínuo (0–100) com cinco dimensões, preserva a variância que ambos os critérios descartam e a decompõe em fontes interpretáveis. Assim, a circularidade parcial (L2) é uma limitação da estratégia de validação, não do instrumento em si.

A validação no Dataset 2 reforça essa interpretação. Naquele *dataset*, o critério externo constitui medida igualmente discreta e unidimensional. O ISBE, operando como *score* contínuo com cinco dimensões, obteve AUC-ROC de 0,990 contra esse critério e Cohen's $d$ de 6,576 entre os grupos extremos. Dentro do grupo *eustress* ($n = 745$), que o critério original trata como homogêneo, o ISBE produziu DP de 10,29, revelando heterogeneidade substantiva que a classificação categórica não captura. Essa diferença de resolução, de três rótulos discretos para um *score* contínuo com perfil dimensional, constitui o valor agregado mensurável do indicador proposto.

Contudo, a mitigação é parcial por três razões. Primeiro, o critério do Dataset 2 é uma pergunta categórica autorreportada, não um diagnóstico clínico ou um escore de instrumento validado. Segundo, o mapeamento entre *surveys* foi realizado por correspondência semântica *post-hoc*, com duas correspondências conceitualmente imperfeitas (competição entre pares $\neq$ *bullying*; ambiente de trabalho $\neq$ `noise_level`). Terceiro, o forte desbalanceamento do Dataset 2 (91,1% *eustress*; apenas 3,8% *distress*) infla métricas calculadas sobre a classe majoritária, como a especificidade e a própria AUC-ROC de 0,990, de modo que o desempenho aparente quase perfeito não se sustenta sob indicadores insensíveis à prevalência. Como discutido na Seção 7.3, a leitura correta exige acompanhar a AUC-ROC da AUC-PR e do F1-Macro; os valores *macro* dos classificadores (Apêndice A.3) confirmam que o ganho real sobre o piso trivial de 91,1% é bem inferior ao sugerido pela acurácia e pela AUC-ROC isoladas. A circularidade está endereçada, não resolvida; a resolução completa requer aplicação simultânea do ISBE e de um instrumento validado à mesma amostra.

A dissociação entre confiabilidade interna e capacidade discriminativa observada na transição Dataset 1 $\rightarrow$ Dataset 2 admite ainda uma leitura adicional, que se conecta à comparação com o *baseline* reportado na Seção 7.3. O agregado parcimonioso favorece a discriminação populacional ao preservar a variância bruta dos itens originais, enquanto o ISBE favorece a precisão individual ao homogeneizar contribuições via normalização Min-Max e ponderação dimensional. Essas duas propriedades, capacidade de separar populações e precisão da medida individual, não são equivalentes nem necessariamente correlacionadas. Instrumentos psicométricos têm sido tradicionalmente avaliados pela primeira, mas a triagem individual em contexto institucional depende criticamente da segunda. Assim, a perda do ISBE em Cohen's $d$ e AUC contra o agregado equivalente não constitui inferioridade global, mas especialização funcional, conforme discutido na Seção 8.5.

## 8.4 Limitações metodológicas

As evidências produzidas neste trabalho são condicionadas por oito limitações que delimitam o alcance das conclusões.

A primeira e mais estrutural é a utilização de dados secundários (L1). As *features* do Dataset 1 não foram coletadas como itens de um instrumento ISBE; o mapeamento entre variáveis do *dataset* e dimensões teóricas foi realizado *post-hoc*, introduzindo graus de liberdade interpretativos que só podem ser eliminados por aplicação prospectiva do questionário a uma amostra coletada para esse fim. A segunda é a circularidade parcial (L2), discutida em detalhe na Seção 8.3. As métricas discriminativas do Dataset 1 utilizam como critério a variável `stress_level`, derivada das mesmas *features* do ISBE. A validação no Dataset 2 endereçou parcialmente essa limitação, mas sem eliminá-la.

A terceira é a ausência de estabilidade temporal (L3). Sem dados de teste-reteste, a confiabilidade temporal do ISBE permanece desconhecida, lacuna que se torna evidente na comparação com instrumentos de referência, todos com ICC documentado na faixa de 0,72–0,85. A quarta decorre do refinamento iterativo em amostra única (L4). As versões 2.0 a 2.3 foram todas calibradas no mesmo conjunto de 275 observações, configurando risco de capitalização sobre o acaso. A validação no Dataset 2 demonstrou capacidade discriminativa em amostra independente, mas a generalização das propriedades de confiabilidade permanece não demonstrada.

A quinta refere-se à dimensão ARR, composta por apenas dois itens (L5), configuração que torna a estimativa de alfa instável (0,672) e que requer expansão em versões futuras. A sexta diz respeito à análise fatorial exploratória (L6), conduzida nos mesmos dados usados para construir o índice. A análise paralela reteve um fator dominante com convergência de três das cinco dimensões, sugerindo que a estrutura empírica pode ser mais parcimoniosa que o modelo pentadimensional proposto. Análise fatorial confirmatória em amostra independente é necessária para dirimir essa questão.

A sétima envolve a extração retroativa da *feature importance* (L7). A FI, terceira fonte da triangulação, foi obtida retroativamente dos modelos treinados na Macrofase III, com hiperparâmetros de quatro dos seis modelos reconstruídos a partir dos *grids* originais, não dos `best_params_` preservados. A análise de sensibilidade (variação de AUC = 0,015) sugere impacto limitado no desempenho final, mas a incerteza adicional nos pesos FI deve ser declarada. A oitava limitação refere-se à comparação com o *baseline* (Seção 7.3), conduzida *post-hoc* após a finalização da arquitetura do ISBE. Embora o *baseline* tenha sido construído sobre o mesmo conjunto de *features* selecionadas e com pré-processamento equivalente, a definição operacional de "agregado parcimonioso" admite variantes. Particularmente, não foram testadas variantes alternativas como agregados sobre Z-*scores* ou *rankings*, nem agregados parciais por dimensão. As diferenças observadas em $\alpha$ ($\approx +0{,}04$), Split-half ($\approx +0{,}05$) e SEM ($-39\%$) devem ser interpretadas como ordens de grandeza, não como estimativas de precisão calibrada. Recomenda-se que estudos futuros reportem intervalos de confiança das diferenças entre o ISBE e múltiplos *baselines* via *bootstrap* pareado.

## 8.5 Posicionamento frente a instrumentos estabelecidos e implicações de uso

A Tabela 21 posiciona as métricas do ISBE em relação ao WHO-5, PSS-10, DASS-21 e WEMWBS. A comparação contextualiza, mas não equivale. Os valores de referência provêm de décadas de validação multicêntrica com critério externo, normas populacionais e estabilidade temporal documentada; os valores do ISBE provêm de validação interna em amostra única, complementada por avaliação em uma amostra externa com mapeamento *post-hoc*.

A consistência interna preliminar do ISBE no Dataset 1 ($\alpha = 0{,}941$) atinge magnitudes análogas às faixas reportadas para WHO-5 (0,82–0,87), PSS-10 (0,78–0,84) e WEMWBS (0,89–0,91), aproximando-se do DASS-21 (0,93–0,96). Cabe ressaltar, contudo, que esse valor reflete a coerência estrutural da amostra de calibração e não é diretamente comparável às métricas populacionais *out-of-sample* dos instrumentos de referência, cuja estabilização depende de validações externas futuras. O SEM (5,7%) é competitivo. No entanto, a confiabilidade não se replicou no Dataset 2, e a lacuna mais evidente, a ausência de ICC, está presente em todos os instrumentos de referência. Essas assimetrias impedem qualquer afirmação de equivalência; o que se pode afirmar é que as propriedades preliminares são compatíveis com a faixa inicial observada em instrumentos que, no início de seu desenvolvimento, também partiam de amostras únicas e validação interna.

Do ponto de vista de uso, dois objetivos legítimos e complementares devem ser distinguidos para situar o ISBE: a triagem individual e a classificação populacional. A triagem individual requer alta precisão na medida atribuída a cada estudante (operacionalizada pelo erro padrão de medida); a classificação populacional requer alta capacidade de separar grupos com diferentes perfis (operacionalizada por Cohen's $d$ e AUC). A comparação reportada na Tabela 16 indica que o ISBE serve melhor à primeira finalidade. Em relação a um agregado parcimonioso construído sobre as mesmas 14 *features* com polaridade alinhada, o ISBE apresenta SEM substancialmente menor (5,7% contra 9,4% da escala no Dataset 1), reduzindo o intervalo de confiança de 95% de cada escore individual de aproximadamente $\pm 18{,}4$ para $\pm 11{,}2$ pontos, diferença prática relevante para uma ferramenta de triagem orientada a casos. Em contrapartida, o agregado equivalente supera o ISBE em separação populacional (Cohen's $d$ de 3,52 contra 3,39; AUC de 0,931 contra 0,884). Ainda assim, o SEM de 5,7 pontos implica que diferenças individuais menores que $\pm 11$ pontos podem ser atribuídas a erro de medida, o que limita a aplicabilidade para decisões clínicas isoladas e reforça a necessidade de interpretação conjunta com o perfil dimensional e com avaliação profissional. Em síntese, o ISBE situa-se como ferramenta complementar de rastreamento que combina precisão individual aceitável e perfilagem dimensional acionável, propriedades não simultaneamente oferecidas por um agregado simples ou por critérios categóricos como os disponíveis nos *datasets* originais.

Do ponto de vista aplicado, o principal diferencial do ISBE em relação aos critérios disponíveis nos *datasets* originais, e em certa medida, em relação a instrumentos unidimensionais como WHO-5 e PSS-10, é a decomposição dimensional do *score* global. Um rótulo de `stress_level = 2` ou uma resposta Likert de nível 4 informa a intensidade do estresse, mas não sua composição. O ISBE, ao decompor o *score* em cinco dimensões, permite identificar se o comprometimento é predominantemente psicoemocional (VPE), acadêmico (CEA), energético (EEE), social (SSA) ou associado a estressores ambientais (ARR). Essa decomposição é condição necessária para intervenções institucionais direcionadas, por exemplo, encaminhamento psicológico para estudantes com VPE comprometido *versus* reestruturação de carga horária para estudantes com CEA comprometido. Nenhuma das variáveis-alvo disponíveis nos *datasets* originais oferece esse nível de especificidade. A confirmação dessa vantagem aplicada requer, naturalmente, estudos de campo que avaliem se as recomendações derivadas dos perfis dimensionais produzem melhores desfechos do que intervenções baseadas em *scores* globais.

# 9 CONSIDERAÇÕES FINAIS

Este trabalho propôs e avaliou uma abordagem computacional para a construção de um indicador de bem-estar estudantil a partir de dados de autorrelato, integrando análise exploratória, modelagem preditiva com aprendizado de máquina e validação psicométrica. O percurso de quatro macrofases produziu três resultados centrais.

Na etapa de modelagem preditiva, modelos de *machine learning*, particularmente *Random Forest*, XGBoost e estratégias de *ensemble*, classificaram com acurácias entre 89,5% e 99,5% os níveis de estresse de estudantes universitários a partir de variáveis psicológicas, acadêmicas, fisiológicas e sociais, confirmando o potencial preditivo dos dados de autorrelato reportado por Ovi *et al.* (2025) e Meng *et al.* (2024). A análise de *feature importance* revelou que variáveis como `blood_pressure`, cujas relações com o estresse são predominantemente não lineares, ocupam posições de destaque apenas quando avaliadas por métricas que capturam esse tipo de dependência, achado que fundamentou a escolha de uma estratégia de triangulação multimétodo para derivação dos pesos dimensionais.

Na etapa de construção do indicador, o *pipeline* proposto demonstrou a viabilidade de derivar pesos dimensionais por triangulação de correlação linear, *mutual information* e *feature importance* de modelos preditivos, sem intervenções manuais na ponderação. O indicador resultante (ISBE 2.3, 14 itens, 5 dimensões, escala 0–100) apresentou, no Dataset 1, consistência interna elevada ($\alpha = 0{,}941$; SB = 0,960), erro de medida controlado (SEM = 5,7%), concordância com as predições de *machine learning* ($r = 0{,}926$) e separação ampla entre grupos de estresse ($d = 3{,}387$). A comparação contra um agregado parcimonioso construído sobre as mesmas 14 *features* com polaridade alinhada (Seção 7.3) qualificou esses números: o ISBE preserva consistência interna comparável ao *baseline* equivalente, com vantagem específica e substancial em precisão individual, redução do SEM de 9,4% para 5,7% da escala (queda de 39%), ao custo de pequena perda em discriminação populacional. Diferentemente dos critérios originais dos *datasets*, uma variável categórica de três níveis no Dataset 1 e uma pergunta Likert de cinco níveis no Dataset 2, e diferentemente também do agregado parcimonioso, que oferece apenas um valor global, o ISBE transforma as mesmas informações em *score* contínuo decomposto em cinco dimensões interpretáveis, preservando variância intracategórica que os critérios originais descartam e oferecendo perfil dimensional acionável para intervenções direcionadas.

Na etapa de validação cruzada, a aplicação do ISBE ao Dataset 2 ($n = 816$) produziu evidência complementar de natureza distinta: a capacidade discriminativa foi confirmada contra critério externo independente (AUC = 0,990; $d = 6{,}576$), parcialmente eliminando a circularidade que limita as métricas do Dataset 1. Ao mesmo tempo, a confiabilidade degradou-se ($\alpha = 0{,}606$), revelando que a consistência interna é sensível à instrumentação e que a generalização requer aplicação prospectiva com questionário padronizado.

A contribuição principal deste trabalho é metodológica. O *pipeline* proposto, da seleção *data-driven* de *features* à triangulação de importância e à avaliação das evidências de validade e desempenho, constituem abordagem reprodutível para a construção de indicadores sintéticos em contextos onde dados de autorrelato estão disponíveis mas instrumentos psicométricos dedicados são inexistentes. A integração explícita entre as etapas preditiva e psicométrica, com declaração transparente das limitações de cada tipo de evidência, representa contribuição ao debate sobre como traduzir padrões algorítmicos em métricas interpretáveis, lacuna apontada pela literatura recente. Como contribuição secundária, o trabalho produziu uma análise comparativa de 210 modelos de *machine learning*

em dois *datasets* de estresse estudantil, documentando o efeito de diferentes estratégias de pré-processamento, seleção de atributos e *ensemble* sobre o desempenho preditivo.

O amadurecimento do ISBE requer cinco passos, consistentes com o percurso de validação de instrumentos psicométricos: (a) aplicação prospectiva a amostra independente ($n > 300$), com questionário desenhado para replicar a estrutura dimensional do ISBE; (b) validação convergente simultânea com instrumentos estabelecidos (PSS-10, WHO-5 ou DASS-21); (c) estudo de teste-reteste com intervalo de 2 a 4 semanas; (d) análise fatorial confirmatória em amostra independente; e (e) expansão da dimensão ARR para ao menos três itens. Adicionalmente, a replicação do *pipeline* com outros constructos, por exemplo, engajamento acadêmico ou satisfação institucional, permitiria avaliar sua generalização para além do caso específico do bem-estar estudantil.

## REFERÊNCIAS


Auerbach, R. *et al.* WHO World Mental Health Surveys International College Student Project: Prevalence and distribution of mental disorders. **Journal of Abnormal Psychology**, v. 127, n. 7, p. 623–638, 2018. Disponível em: <https://doi.org/10.1037/abn0000362>. Acesso em: 10 jul. 2026. Citado na página 1.

Baghurst, T.; Kelley, B. An examination of stress in college students over the course of a semester. **Health Promotion Practice**, v. 15, n. 3, p. 438–447, 2014. Disponível em: <https://doi.org/10.1177/1524839913510316>. Acesso em: 10 jul. 2026. Citado na página 1.

Battiti, R. Using mutual information for selecting features in supervised neural net learning. **IEEE Transactions on Neural Networks**, v. 5, n. 4, p. 537–550, 1994. Disponível em: <https://doi.org/10.1109/72.298224>. Acesso em: 10 jul. 2026. Citado na página 6.

Brasil. **Resolução nž 510, de 7 de abril de 2016**. 2016. Conselho Nacional de Saúde. Diário Oficial da União: Brasília, DF, 24 maio 2016. Dispõe sobre as normas aplicáveis a pesquisas em Ciências Humanas e Sociais. Disponível em: <https://conselho.saude.gov.br/resolucoes/2016/Reso510.pdf>. Acesso em: 10 jul. 2026. Citado na página 7.

Bruffaerts, R. *et al.* Mental health problems in college freshmen: Prevalence and academic functioning. **Journal of Affective Disorders**, v. 225, p. 97–103, 2018. Disponível em: <https://doi.org/10.1016/j.jad.2017.07.044>. Acesso em: 10 jul. 2026. Citado na página 1.

Cohen, S.; Williamson, G. Perceived stress in a probability sample of the United States. In: SPACAPAN, S.; OSKAMP, S. (Ed.). **The Social Psychology of Health**. Newbury Park, CA: Sage, 1988. p. 31–67. Citado 2 vezes nas páginas 4 e 27.

Cover, T.; Thomas, J. **Elements of Information Theory**. 2. ed. Hoboken, NJ: Wiley, 2006. Citado na página 6.

Cronbach, L. Coefficient alpha and the internal structure of tests. **Psychometrika**, v. 16, n. 3, p. 297–334, 1951. Disponível em: <https://doi.org/10.1007/BF02310555>. Acesso em: 10 jul. 2026. Citado na página 18.

Dwyer, D. *et al.* Machine learning approaches for clinical psychology and psychiatry. **Annual Review of Clinical Psychology**, v. 14, p. 91–118, 2018. Disponível em: <https://doi.org/10.1146/annurev-clinpsy-032816-045037>. Acesso em: 10 jul. 2026. Citado na página 1.

Eder, J. *et al.* A multimodal approach to depression diagnosis: insights from machine learning algorithm development in primary care. **European Archives of Psychiatry and Clinical Neuroscience**, 2025. Disponível em: <https://doi.org/10.1007/s00406-025-01990-5>. Acesso em: 10 jul. 2026. Citado 4 vezes nas páginas 2, 3, 4 e 28.

Graham, S. *et al.* Artificial intelligence for mental health and mental illnesses: an overview. **Current Psychiatry Reports**, v. 21, n. 11, p. 116, 2019. Disponível em: <https://doi.org/10.1007/s11920-019-1094-0>. Acesso em: 10 jul. 2026. Citado 4 vezes nas páginas 1, 2, 4 e 29.

Henry, J.; Crawford, J. The short-form version of the Depression Anxiety Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample. **British Journal of Clinical Psychology**, v. 44, n. 2, p. 227–239, 2005. Disponível em: <https://doi.org/10.1348/014466505X29657>. Acesso em: 10 jul. 2026. Citado 2 vezes nas páginas 4 e 27.

Keyes, C. Promoting and protecting mental health as flourishing: A complementary strategy for improving national mental health. **American Psychologist**, v. 62, n. 2, p. 95–108, 2007. Disponível em: <https://doi.org/10.1037/0003-066X.62.2.95>. Acesso em: 10 jul. 2026. Citado na página 1.

Lee, E. Review of the psychometric evidence of the Perceived Stress Scale. **Asian Nursing Research**, v. 6, n. 4, p. 121–127, 2012. Disponível em: <https://doi.org/10.1016/j.anr.2012.08.004>. Acesso em: 10 jul. 2026. Citado 2 vezes nas páginas 4 e 27.

Lipson, S.; Lattie, E.; Eisenberg, D. Increased rates of mental health service utilization by U.S. college students: 10-year population-level trends (2007–2017). **Psychiatric Services**, v. 70, n. 1, p. 60–63, 2019. Disponível em: <https://doi.org/10.1176/appi.ps.201800332>. Acesso em: 10 jul. 2026. Citado na página 1.

Lovibond, P.; Lovibond, S. The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. **Behaviour Research and Therapy**, v. 33, n. 3, p. 335–343, 1995. Disponível em: <https://doi.org/10.1016/0005-7967(94)00075-U>. Acesso em: 10 jul. 2026. Citado 2 vezes nas páginas 4 e 27.

Meng, X. *et al.* Applying machine learning to understand the role of social–emotional skills on subjective well-being and physical health. **Applied Psychology: Health and Well-Being**, 2024. Disponível em: <https://doi.org/10.1111/aphw.12624>. Acesso em: 10 jul. 2026. Citado 4 vezes nas páginas 2, 3, 28 e 33.

Mohr, D. *et al.* Personal sensing: Understanding mental health using ubiquitous sensors and machine learning. **Annual Review of Clinical Psychology**, v. 13, p. 23–47, 2017. Disponível em: <https://doi.org/10.1146/annurev-clinpsy-032816-044949>. Acesso em: 10 jul. 2026. Citado na página 2.

Nemesure, M. *et al.* Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence. **Scientific Reports**, v. 11, p. 1980, 2021. Disponível em: <https://doi.org/10.1038/s41598-021-81368-4>. Acesso em: 10 jul. 2026. Citado na página 1.

Ovi, M. *et al.* **Protecting Student Mental Health with a Context-Aware Machine Learning Framework for Stress Monitoring**. 2025. Pré-print. arXiv:2508.01105. Disponível em: <https://arxiv.org/abs/2508.01105>. Acesso em: 10 jul. 2026. Citado 9 vezes nas páginas 2, 3, 5, 6, 11, 28, 33, 38 e 40.

Panicheva, P. *et al.* Predicting subjective well-being in a high-risk sample of russian mental health app users. **EPJ Data Science**, v. 11, n. 1, p. 21, 2022. Disponível em: <https://doi.org/10.1140/epjds/s13688-022-00333-x>. Acesso em: 10 jul. 2026. Citado 3 vezes nas páginas 2, 3 e 28.

Pascoe, M. *et al.* The impact of stress on students in secondary school and higher education. **International Journal of Adolescence and Youth**, v. 25, n. 1, p. 104–112, 2020. Disponível em: <https://doi.org/10.1080/02673843.2019.1596823>. Acesso em: 10 jul. 2026. Citado na página 1.

RxNach. **Student Stress Factors: A Comprehensive Analysis**. 2024. Conjunto de dados. Kaggle. Disponível em: <https://www.kaggle.com/datasets/rxnach/student-stress-factors-a-comprehensive-analysis>. Acesso em: 10 jul. 2026. Citado 2 vezes nas páginas 3 e 6.

Shatte, A. *et al.* Machine learning in mental health: a scoping review of methods and applications. **Psychological Medicine**, v. 49, n. 9, p. 1426–1448, 2019. Disponível em: <https://doi.org/10.1017/S0033291719000151>. Acesso em: 10 jul. 2026. Citado na página 1.

Singh, A. **Stress and Well-being Data of College Students**. 2024. Conjunto de dados. Kaggle. Disponível em: <https://www.kaggle.com/datasets/ashutoshsingh22102/stress-and-well-being-data-of-college-students>. Acesso em: 10 jul. 2026. Citado 3 vezes nas páginas 3, 7 e 21.

Stewart-Brown, S. *et al.* Internal construct validity of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS): a Rasch analysis using data from the Scottish Health Education Population Survey. **Health and Quality of Life Outcomes**, v. 7, p. 15, 2009. Disponível em: <https://doi.org/10.1186/1477-7525-7-15>. Acesso em: 10 jul. 2026. Citado 3 vezes nas páginas 1, 4 e 27.

Tennant, R. *et al.* The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. **Health and Quality of Life Outcomes**, v. 5, p. 63, 2007. Disponível em: <https://doi.org/10.1186/1477-7525-5-63>. Acesso em: 10 jul. 2026. Citado 3 vezes nas páginas 1, 4 e 27.

Topp, C. *et al.* The WHO-5 Well-Being Index: A systematic review of the literature. **Psychotherapy and Psychosomatics**, v. 84, n. 3, p. 167–176, 2015. Disponível em: <https://doi.org/10.1159/000376585>. Acesso em: 10 jul. 2026. Citado 3 vezes nas páginas 1, 4 e 27.

Torous, J. *et al.* The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. **World Psychiatry**, v. 20, n. 3, p. 318–335, 2021. Disponível em: <https://doi.org/10.1002/wps.20883>. Acesso em: 10 jul. 2026. Citado na página 2.

Wies, B.; Landers, C.; Ienca, M. Digital mental health for young people: A scoping review of ethical promises and challenges. **Frontiers in Digital Health**, v. 3, p. 697072, 2021. Disponível em: <https://doi.org/10.3389/fdgth.2021.697072>. Acesso em: 10 jul. 2026. Citado na página 1.

World Health Organization. **World mental health report: transforming mental health for all**. 2022. Geneva: World Health Organization. Disponível em: <https://www.who.int/publications/i/item/9789240049338>. Acesso em: 10 jul. 2026. Citado na página 1.

# APÊNDICE A – DETALHAMENTO DA ANÁLISE EXPLORATÓRIA, SELEÇÃO DE ATRIBUTOS E AVALIAÇÃO DOS MODELOS

Este apêndice reúne os procedimentos e os resultados completos das três primeiras macrofases do trabalho: preparação e exploração dos dados (Macrofase I), engenharia de *features* e análise exploratória (Macrofase II) e modelagem preditiva e *ensemble* (Macrofase III). Todos os procedimentos seguem o *framework* de Ovi *et al.* (2025) e foram implementados em Python 3.10 com as bibliotecas *pandas*, NumPy e *scikit-learn*. As figuras e tabelas aqui apresentadas fundamentam as decisões metodológicas que sustentam a construção do ISBE.

## A.1 MACROFASE I: PREPARAÇÃO E EXPLORAÇÃO DOS DADOS

A primeira macrofase abrangeu a inspeção de qualidade, a limpeza e o pré-processamento dos dois *datasets*. A inspeção estrutural verificou valores ausentes, duplicatas e inconsistências de tipagem. Todas as *features* foram normalizadas ao intervalo $[0, 1]$ pelo procedimento Min-Max; a variável-alvo categórica do Dataset 2 foi codificada numericamente via *LabelEncoder*; e a divisão treino-teste empregou amostragem estratificada (75% e 25%) com semente fixa (*seed* = 42), assegurando a proporção das classes em ambos os conjuntos. A detecção de *outliers* pelo método *Interquartile Range* (IQR) foi conduzida apenas para caracterização dos dados; optou-se por preservar os valores extremos, dado que representam fenômenos psicológicos legítimos e que os algoritmos baseados em árvores são robustos a eles.

### A.1.1 Origem dos dados, coleta e delineamento ético

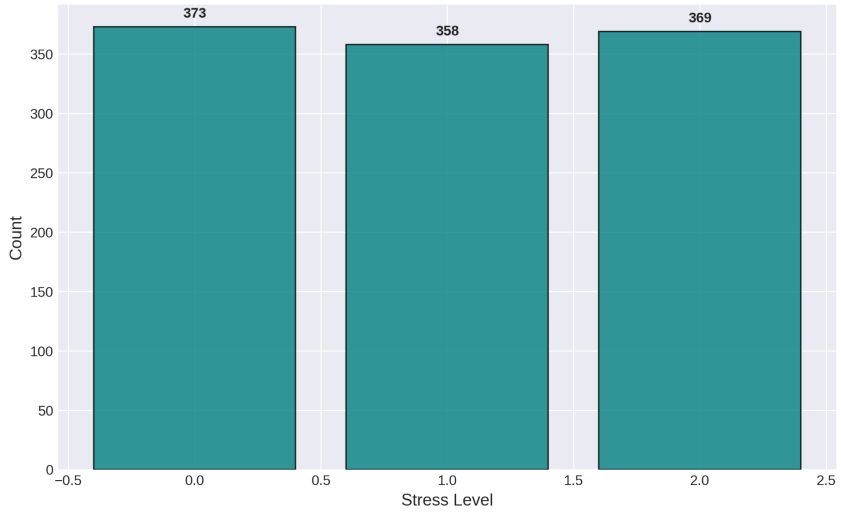
Os dois *datasets* foram submetidos à inspeção inicial de qualidade, com os resultados reportados a seguir.

No Dataset 1 (*Student Stress Factors: A Comprehensive Analysis*), a inspeção não identificou valores ausentes nem duplicatas. A variável *target* confirmou distribuição equilibrada: baixo ($n$ = 373; 33,9%), médio ($n$ = 358; 32,5%) e alto ($n$ = 369; 33,5%), conforme a Figura 5.

No Dataset 2 (*Stress and Well-being Data of College Students*), a inspeção identificou 27 registros duplicados (3,2%), removidos conforme o pré-processamento descrito na Seção A.1, resultando em 816 registros válidos. Não foram detectados valores ausentes.

A variável alvo *"Have you recently experienced stress in your life?"*, mensurada em escala ordinal de cinco pontos, apresentou concentração predominante nos níveis intermediários, especialmente no nível 3 ($n$ = 263; 31,2%), seguido pelos níveis 2 ($n$ = 219; 26,0%) e 4 ($n$ = 193; 22,9%). Os extremos da escala foram menos frequentes, com menor incidência de indivíduos que relataram ausência quase total de estresse (nível 1: $n$ = 78; 9,3%) ou níveis muito elevados (nível 5: $n$ = 90; 10,7%). Esse padrão indica que a maior parte dos respondentes percebe o estresse como uma experiência recorrente, porém moderada, sugerindo variabilidade suficiente para análises preditivas, embora com leve concentração central (Figura 6).

Figura 5 – Distribuição da variável *target* do Dataset 1




Fonte: Elaborado pelo autor (2026).

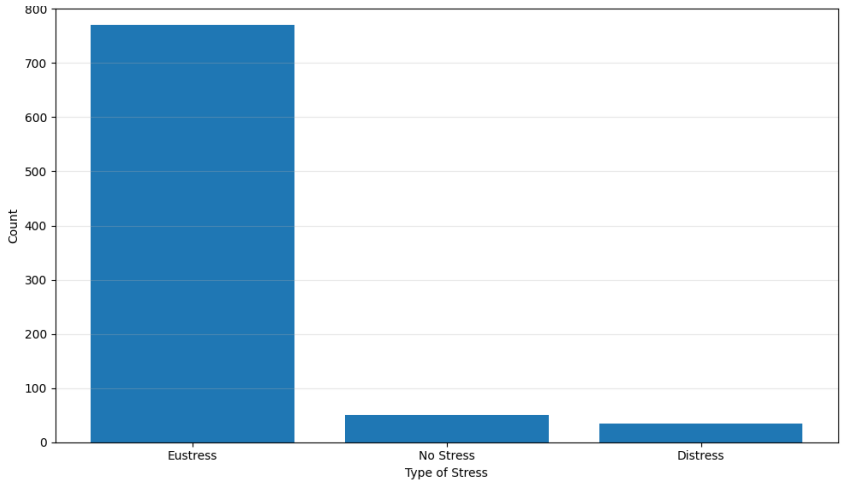
Figura 6 – Distribuição da variável categórica do Dataset 2


Fonte: Elaborado pelo autor (2026).

Em complemento, a variável categórica *"Which type of stress do you primarily experience?"* revelou distribuição fortemente assimétrica, com ampla predominância de *eustress* ($n = 744$; 91,1%), seguida por sem estresse ($n = 41$; 5,1%) e *distress* ($n = 31$; 3,8%), conforme a Figura 7. Tal desbalanceamento severo configura um desafio metodológico relevante, uma vez que modelos de classificação tendem a apresentar viés em favor da classe majoritária, podendo atingir acurácia aparentemente elevada (≈91%) mesmo ao predizer exclusivamente *eustress*. Dessa forma, a análise conjunta das variáveis evidencia a necessidade de cuidados adicionais na modelagem, como métricas além da acurácia e técnicas de balanceamento, a fim de garantir interpretações mais robustas e representativas do fenômeno estudado.

Ambos os conjuntos de dados foram disponibilizados de forma anonimizada, sem in-

Figura 7 – Distribuição da variável *target* do Dataset 2




Fonte: Elaborado pelo autor (2026).

formações pessoais identificáveis, e foram submetidos a avaliações de qualidade, incluindo análise de valores ausentes e duplicados. Essas práticas garantem que a coleta esteja em conformidade com os padrões éticos de pesquisa, conforme as diretrizes estabelecidas em Ovi *et al.* (2025). A análise ética e de privacidade é particularmente relevante, considerando que os dados de saúde mental envolvem informações sensíveis.

### A.1.2 Pré-processamento dos dados: limpeza, preparação e heterogeneidade

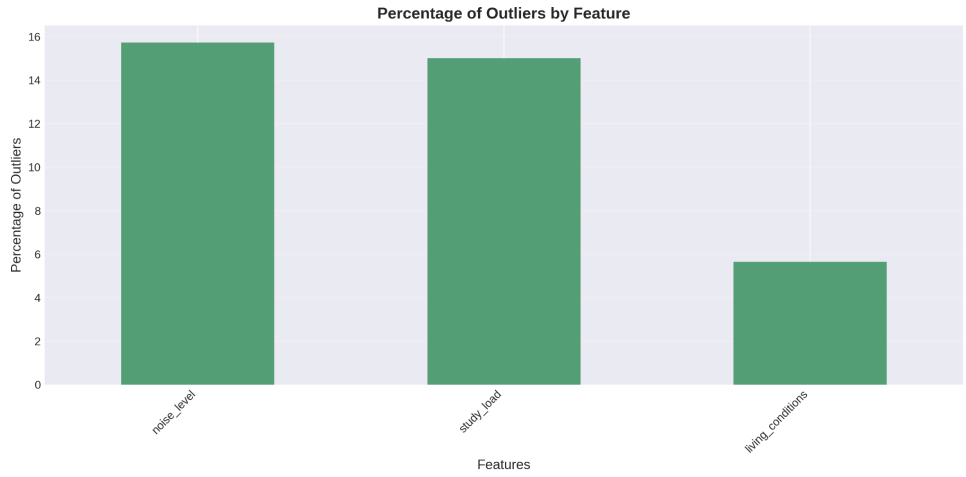
Executado o *pipeline* de pré-processamento descrito na Seção A.1, a análise de *outliers* concentrou-se principalmente em variáveis associadas à heterogeneidade individual e ao contexto. No Dataset 1, observaram-se percentuais moderados a elevados em variáveis como nível de ruído (∼16%), carga de estudos (∼15%) e, em menor grau, condições de moradia (∼6%). Esses valores indicam a existência de grupos específicos expostos a condições extremas, potencialmente relevantes do ponto de vista substantivo. Assim, tais *outliers* não configuram necessariamente ruído estatístico, mas sim manifestações legítimas de contextos adversos, devendo ser tratados com cautela para não comprometer a representatividade do modelo (Figura 8).

No Dataset 2, a variável idade apresenta elevada dispersão e presença de valores extremos, sem um padrão claro de associação com o estresse, sugerindo baixa relevância preditiva e justificando sua exclusão em etapas posteriores. A detecção de *outliers* aponta maior incidência em idade (∼11%), tristeza ou humor deprimido (∼9%), competição com pares (∼8%) e ansiedade ou tensão (∼7%). Assim como no Dataset 1, esses *outliers* parecem refletir experiências individuais extremas, mais associadas à diversidade da amostra do que a erros de medição (Figura 9).

### A.1.3 Distribuição das variáveis explicativas e padrões observados

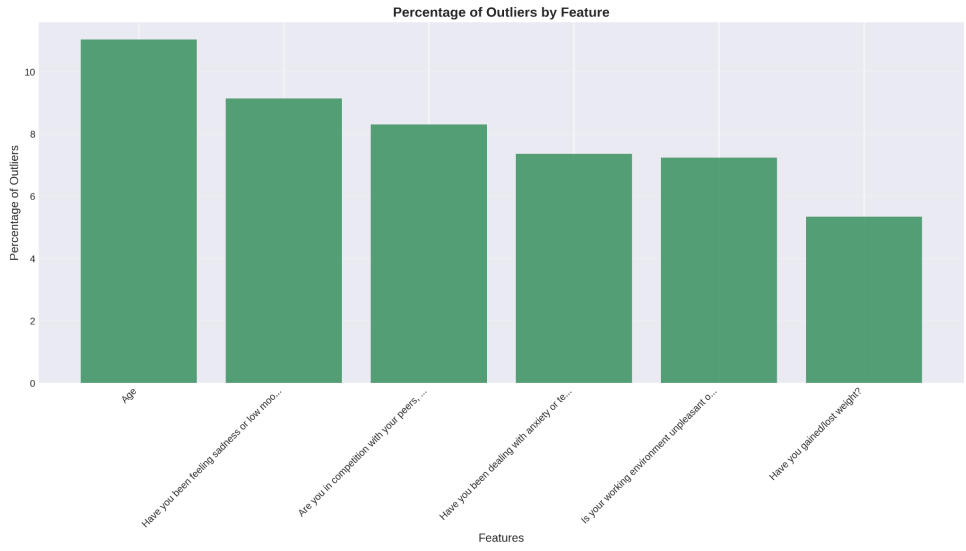
As distribuições das variáveis explicativas, apresentadas nas Figuras 10 e 11, mostram que as variáveis possuem distribuição compatível com dados psicossociais reais. As variáveis do Dataset 1 associadas a sintomas emocionais e fisiológicos, como `anxiety_level`,

Figura 8 – Gráfico de *outliers* no Dataset 1




Fonte: Elaborado pelo autor (2026).

Figura 9 – Gráfico de *outliers* no Dataset 2



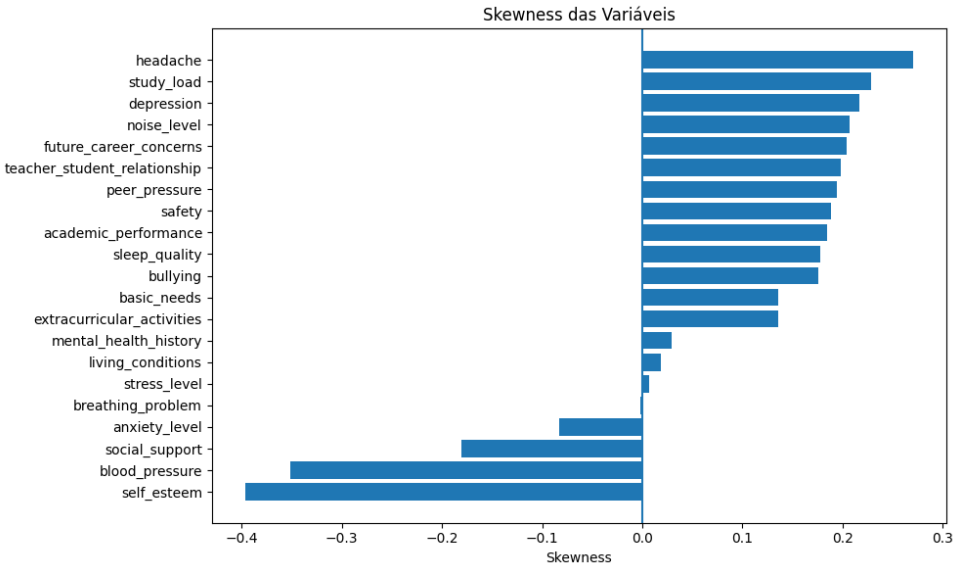

Fonte: Elaborado pelo autor (2026).

`depression`, `headache` e `breathing_problem`, apresentam maior dispersão, refletindo variações substanciais na intensidade com que esses fatores são vivenciados pelos estudantes.

Por outro lado, variáveis relacionadas a fatores protetivos, como `self_esteem`, `sleep_quality`, `academic_performance` e `basic_needs`, mostram-se concentradas em faixas superiores das escalas, sugerindo que a amostra apresenta níveis significativos de suporte e estabilidade. Esse comportamento está em linha com os achados empíricos apresentados no artigo original, que discute a presença de fatores mitigadores do estresse em ambientes acadêmicos.

Após a exclusão da variável *Age* do gráfico do Dataset 2, devido à assimetria positiva extrema confirmada na análise de *outliers*, observa-se que as demais variáveis numéricas apresentam distorção (*skewness*) predominantemente baixa. Esse comportamento indica distribuições levemente assimétricas, compatíveis com a natureza ordinal e biná-

Figura 10 – Distorção (*skewness*) das variáveis no Dataset 1




Fonte: Elaborado pelo autor (2026).

ria dos dados coletados. De modo geral, não são observadas distorções significativas que comprometam a análise estatística subsequente.

## A.2 MACROFASE II: ENGENHARIA DE *FEATURES* E ANÁLISE EXPLORATÓRIA

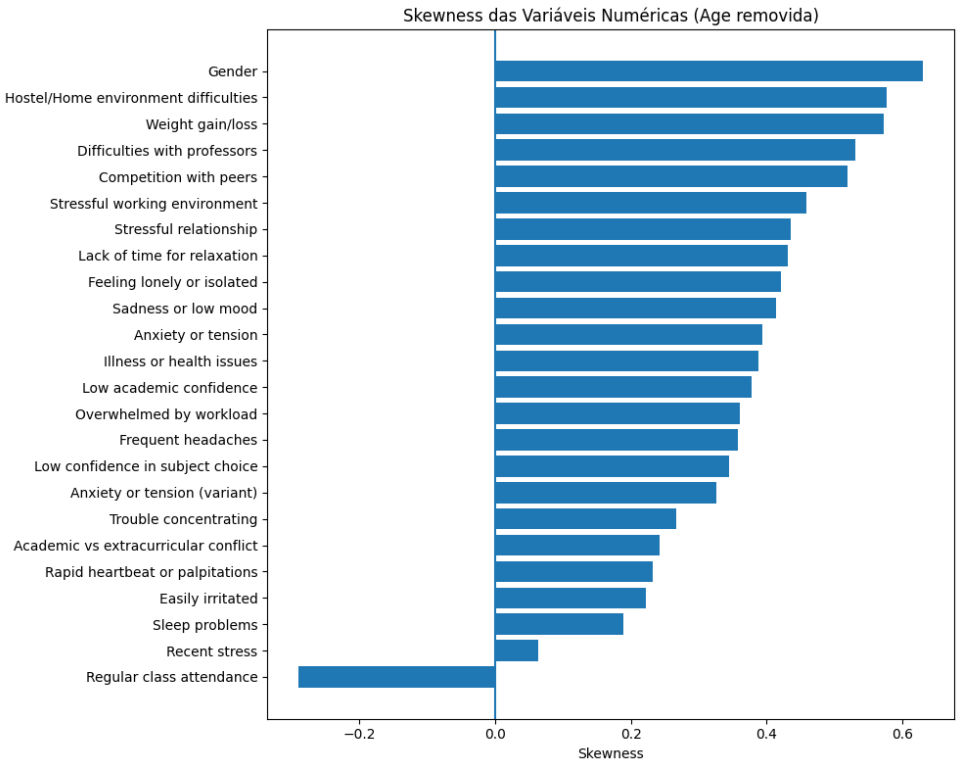
A segunda macrofase aprofundou a análise exploratória e produziu as evidências que orientam a seleção de atributos. Empregaram-se a matriz de correlação de Pearson, com identificação de pares de *features* com $|r| > 0{,}70$ para diagnóstico de multicolinearidade; a *Mutual Information* (MI), para capturar dependências não lineares; a *Principal Component Analysis* (PCA), para avaliar a redundância dimensional de cada *dataset*; e testes inferenciais (normalidade e ANOVA), para verificar diferenças entre os grupos de estresse. A cascata de seleção propriamente dita (*SelectKBest*, RFECV e PCA), reexecutada de forma independente em cada algoritmo, e seus efeitos sobre o desempenho preditivo constam na Macrofase III deste apêndice.

### A.2.1 Padrões de associação com níveis de estresse e detecção de multicolinearidade

A análise de correlação de Pearson revelou associações consistentes entre múltiplas dimensões psicossociais, emocionais e fisiológicas e os níveis de estresse reportados, com diferenças claras na intensidade e na natureza dessas relações entre os dois conjuntos de dados, detalhadas a seguir.

No Dataset 1, que mensura o nível de estresse de forma contínua e abrangente, observam-se correlações elevadas, sobretudo com variáveis associadas a sofrimento psicológico persistente e contextos sociais adversos. As associações mais fortes ocorrem com *bullying* ($r = 0{,}75$), preocupações com a carreira futura ($r = 0{,}74$), nível de ansiedade ($r = 0{,}74$) e depressão ($r = 0{,}73$), indicando que níveis mais altos de estresse estão

Figura 11 – Distorção (*skewness*) das variáveis no Dataset 2 (*Age* removido)




Fonte: Elaborado pelo autor (2026).

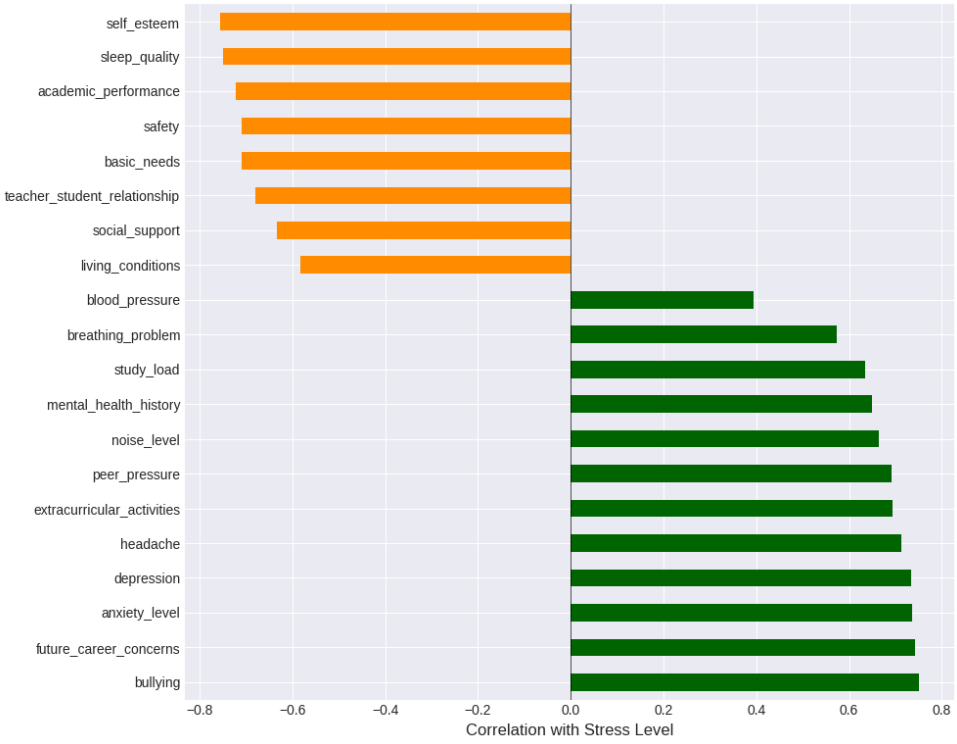
fortemente relacionados a pressões emocionais crônicas, insegurança quanto ao futuro e experiências sociais negativas. Além disso, a variável dor de cabeça ($r = 0{,}71$) apresenta correlação expressiva, sugerindo processos de somatização do estresse quando este se manifesta de forma mais intensa e prolongada (Figura 12).

Em contraste, o Dataset 2, focado na experiência recente de estresse, apresenta correlações de menor magnitude, concentradas principalmente em respostas fisiológicas imediatas e manifestações agudas de ansiedade. Destaca-se a associação com batimentos cardíacos acelerados ou palpitações ($r = 0{,}44$), indicando que sinais corporais agudos são marcadores relevantes da vivência recente de estresse. Observam-se ainda correlações moderadas com ansiedade ou tensão recente ($r = 0{,}23$) e com conflitos entre atividades acadêmicas e extracurriculares ($r = 0{,}21$), apontando para a influência combinada de fatores emocionais e pressões relacionadas à organização do tempo. Alterações de peso corporal ($r = 0{,}16$) e problemas de sono ($r = 0{,}12$) também se associam positivamente ao estresse, embora com menor intensidade (Figura 13).

De forma integrada, os resultados indicam que o Dataset 1 captura determinantes estruturais e crônicos do estresse, enquanto o Dataset 2 reflete manifestações mais reativas e situacionais, reforçando a ideia de que diferentes operacionalizações do estresse revelam camadas complementares do fenômeno, aspecto fundamental para a interpretação dos resultados e para as etapas subsequentes de modelagem preditiva.

A análise de correlação também identificou a presença de multicolinearidade, aspecto relevante para a estabilidade e a interpretabilidade dos modelos preditivos. A Tabela 22 reúne as correlações mais fortes do Dataset 1, separando as associações com a

Figura 12 – Correlação das variáveis com o nível de estresse no Dataset 1




Fonte: Elaborado pelo autor (2026).

variável-alvo dos pares entre preditores, estes últimos relevantes para o diagnóstico de multicolinearidade.

No Dataset 1, observam-se correlações fortes entre o nível de estresse e variáveis de saúde mental, experiências sociais adversas e sintomas somáticos, além de multicolinearidade relevante entre preditores psicossociais (ansiedade, *bullying* e preocupações com a carreira). Isso indica redundância informacional, exigindo cuidados na seleção de variáveis ou o uso de modelos robustos à colinearidade. No Dataset 2 não foram identificados pares altamente correlacionados, indicando baixo risco de multicolinearidade e maior independência entre os preditores. Em síntese, o Dataset 1 demanda estratégias de seleção ou regularização para mitigar a multicolinearidade, enquanto o Dataset 2 permite modelagem mais direta sem perda de estabilidade.

### A.2.2 Avaliação de dependências não lineares via *Mutual Information* (MI)

A utilização da *Mutual Information* (MI) permitiu identificar o grau de dependência estatística entre as variáveis explicativas e o estresse, capturando tanto relações lineares quanto não lineares, o que amplia a compreensão dos fatores mais informativos para a predição.

No Dataset 1, observa-se que as maiores pontuações de MI concentram-se em variáveis associadas a sintomas físicos, saúde mental e condições psicossociais estruturais. A variável `blood_pressure` apresenta o maior valor de MI, indicando elevada capacidade informativa para explicar variações no nível de estresse. Em seguida, destacam-se qualidade do sono, preocupações com a carreira futura, depressão e nível de ansiedade, reforçando a forte interdependência entre estresse, saúde mental e alterações fisiológicas (Figura 14).

Figura 13 – Correlação das variáveis com o nível de estresse no Dataset 2


Fonte: Elaborado pelo autor (2026).

Variáveis como *bullying*, autoestima, dor de cabeça e desempenho acadêmico também exibem alta relevância informativa, evidenciando que experiências sociais negativas e prejuízos no funcionamento acadêmico são componentes centrais na caracterização de níveis elevados de estresse. Já fatores contextuais, como apoio social, segurança, necessidades básicas, atividades extracurriculares, pressão dos pares e nível de ruído, apresentam MI moderada, sugerindo influência indireta ou mediada por outras dimensões psicológicas. Por fim, variáveis como histórico de saúde mental e problemas respiratórios apresentam menor MI, indicando contribuição informativa mais limitada no contexto global do modelo, embora não necessariamente irrelevante do ponto de vista clínico ou social.

No Dataset 2, o *ranking* de MI evidencia um padrão distinto, coerente com a natureza aguda e situacional da variável alvo. A variável percepção de batimentos cardíacos acelerados ou palpitações apresenta a maior MI, destacando-se como o principal indicador informativo da experiência recente de estresse. Em seguida, observa-se a relevância de an-

Tabela 22 – Correlações de Pearson mais fortes no Dataset 1 ($|r| > 0{,}7$)

| **Variável 1** | **Variável 2** | **Correlação** |
|---|---|---|
| *Correlações com a variável-alvo* | | |
| bullying | stress_level | 0,75 |
| future_career_concerns | stress_level | 0,74 |
| anxiety_level | stress_level | 0,74 |
| depression | stress_level | 0,73 |
| headache | stress_level | 0,71 |
| sleep_quality | stress_level | −0,75 |
| self_esteem | stress_level | −0,76 |
| academic_performance | stress_level | −0,72 |
| basic_needs | stress_level | −0,71 |
| safety | stress_level | −0,71 |
| *Pares entre preditores (multicolinearidade)* | | |
| anxiety_level | future_career_concerns | 0,72 |
| anxiety_level | bullying | 0,71 |
| future_career_concerns | bullying | 0,71 |
| depression | future_career_concerns | 0,71 |
| anxiety_level | sleep_quality | −0,71 |
| self_esteem | future_career_concerns | −0,71 |
| blood_pressure | social_support | −0,75 |

Fonte: Elaborado pelo autor (2026).

siedade ou tensão recente, conflitos entre atividades acadêmicas e extracurriculares e falta de confiança na escolha acadêmica, sugerindo que o estresse recente é fortemente influenciado por respostas emocionais imediatas e pressões relacionadas ao contexto educacional (Figura 15).

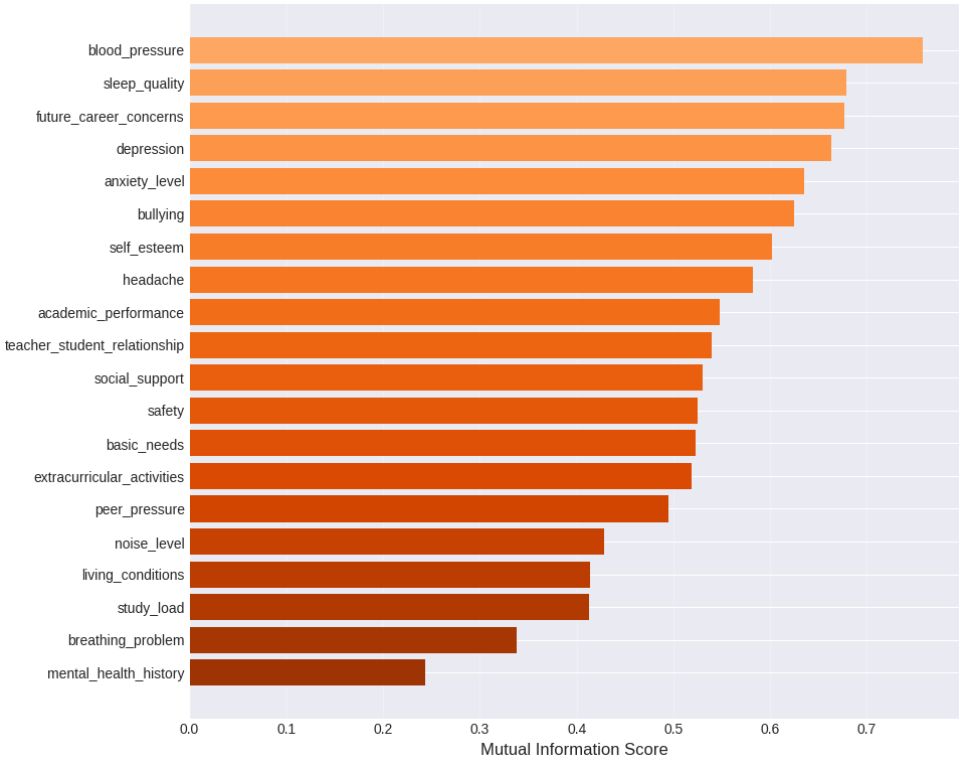
Variáveis demográficas como idade e gênero exibem baixa MI, indicando fraca dependência direta com a experiência recente de estresse. De modo semelhante, fatores como frequência às aulas, irritabilidade, problemas de concentração, tristeza persistente e problemas de saúde apresentam valores próximos de zero, sugerindo baixo poder informativo quando considerados isoladamente.

De maneira consoante com a análise de correlação (Seção A.2.1), os resultados de MI reafirmam a natureza estrutural e cumulativa do Dataset 1 frente ao caráter reativo e imediato do Dataset 2, justificando a adoção de estratégias diferenciadas de seleção de características para cada conjunto de dados.

### A.2.3 Redução de dimensionalidade (PCA)

A visualização do PCA no Dataset 1 indica forte concentração de variância nos primeiros componentes principais. O primeiro componente explica aproximadamente 60% da variância total, evidenciando a presença de uma dimensão latente dominante, associada principalmente a fatores psicossociais e de saúde mental. Observa-se um ponto de inflexão claro após o segundo componente, sugerindo que grande parte da informação relevante está concentrada nos primeiros eixos (Figura 16).

Figura 14 – Pontuações de *Mutual Information* no Dataset 1




Fonte: Elaborado pelo autor (2026).

A curva de variância explicada acumulada atinge cerca de 95% com aproximadamente 15 componentes, indicando que é possível reduzir substancialmente a dimensionalidade sem perda significativa de informação. Esse resultado é consistente com a multicolinearidade previamente identificada, reforçando a adequação do PCA como estratégia de compactação e estabilização do espaço de atributos.

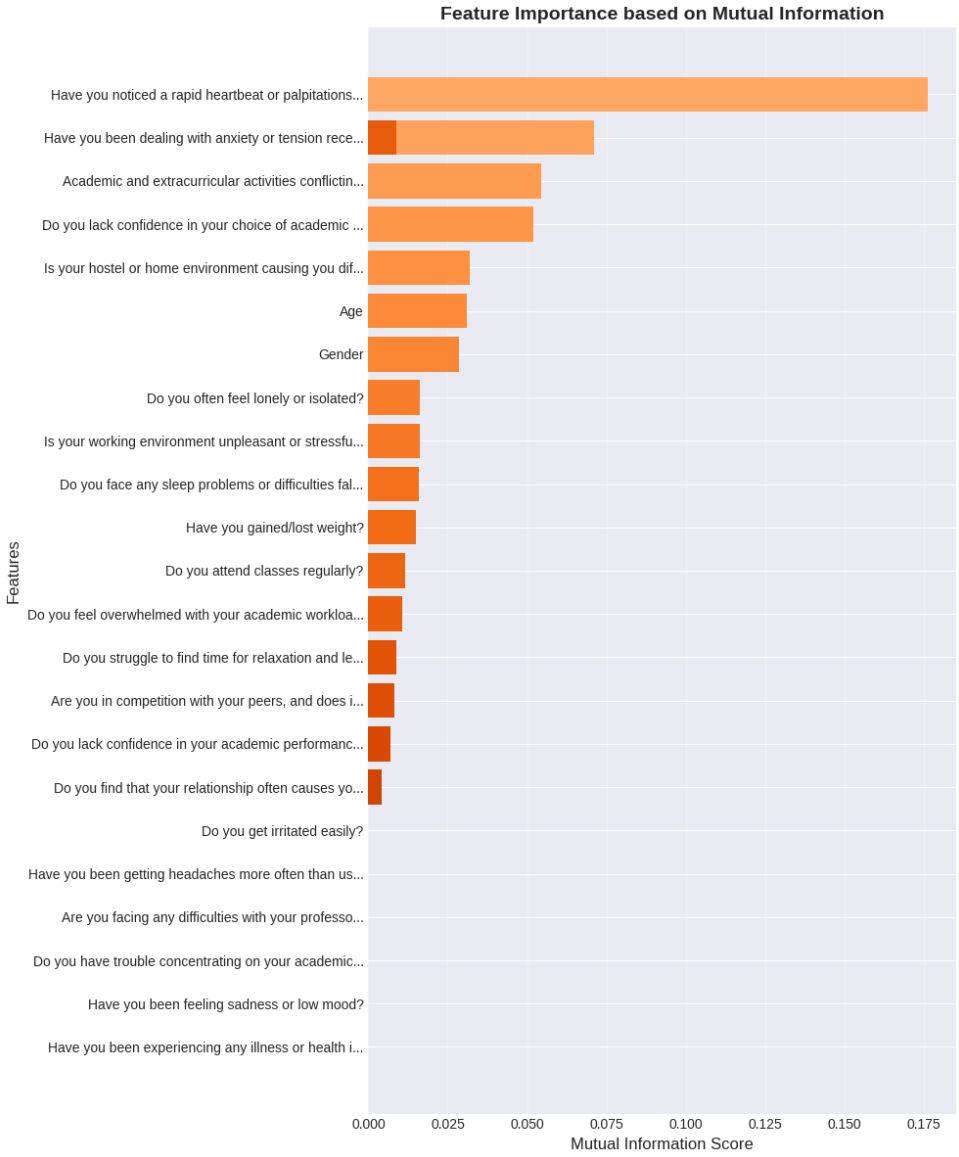
No Dataset 2, o PCA apresenta uma distribuição mais gradual da variância entre os componentes principais. O primeiro componente explica cerca de 13% da variância, sem a presença de um componente claramente dominante. A variância acumulada cresce de forma mais linear, atingindo 95% apenas após aproximadamente 22 componentes, o que indica menor redundância entre as variáveis (Figura 17).

Esse comportamento é coerente com a análise de correlação, que apontou baixa multicolinearidade no Dataset 2. Assim, a redução dimensional neste conjunto tende a gerar ganhos mais limitados, sugerindo que a preservação de um maior número de variáveis pode ser necessária para manter o poder explicativo. De forma comparativa, o Dataset 1 apresenta alta redundância informacional, favorecendo técnicas de redução de dimensionalidade como o PCA, enquanto o Dataset 2 exibe estrutura mais dispersa, na qual a redução agressiva pode resultar em perda relevante de informação.

### A.2.4 Análise estatística inferencial

Inicialmente, registra-se uma ressalva metodológica sobre o teste de normalidade. Tanto a variável-alvo do Dataset 1 (nível de estresse) quanto a do Dataset 2 (experiência recente de estresse) são intrinsecamente ordinais e assumem um número finito e pequeno

Figura 15 – Pontuações de *Mutual Information* no Dataset 2

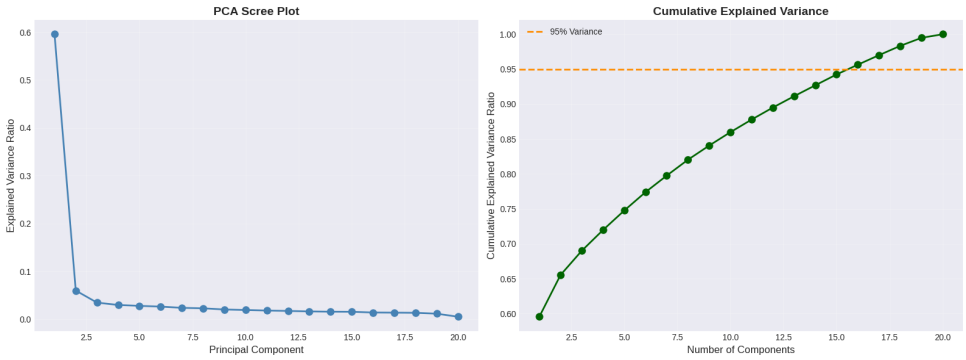



Fonte: Elaborado pelo autor (2026).

de valores discretos. Para variáveis dessa natureza, qualquer teste formal de normalidade (Shapiro-Wilk, D'Agostino ou Kolmogorov-Smirnov) rejeita a hipótese nula de forma praticamente automática, sobretudo em amostras de centenas de observações, nas quais a sensibilidade do teste é alta. A rejeição observada (Dataset 1: estatística = 7361,72; Dataset 2: estatística = 61,04; ambos $p < 0{,}001$) é, portanto, esperada por construção e tem valor informativo limitado: ela não decorre de uma propriedade substantiva dos dados, mas da própria discretude da escala. Por essa razão, este teste não é utilizado aqui como fundamento das escolhas analíticas, sendo reportado apenas para fins de completude.

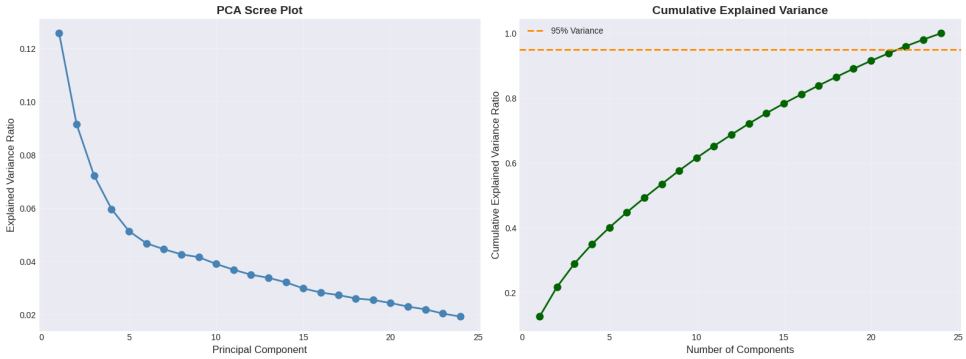
A justificativa para o emprego de métodos que capturam relações não lineares apoia-se, em vez disso, em dois elementos diretamente pertinentes à estrutura dos dados: a matriz de correlação (interpretável, no caso de itens ordinais, como aproximação de correlações policóricas entre variáveis latentes contínuas subjacentes) e, principalmente, as pontuações de *Mutual Information* apresentadas na Seção A.2.1. O contraste entre a

Figura 16 – Autovalores do PCA e variância explicada acumulada do Dataset 1




Fonte: Elaborado pelo autor (2026).

Figura 17 – Autovalores do PCA e variância explicada acumulada do Dataset 2




Fonte: Elaborado pelo autor (2026).

baixa correlação linear e a elevada MI de variáveis como `blood_pressure` evidencia, de forma muito mais direta do que qualquer teste de normalidade, a presença de dependências não lineares relevantes e, com isso, sustenta tanto a adoção da MI na seleção de *features* quanto a opção por modelos baseados em árvore na Macrofase III. Quanto à ANOVA reportada a seguir, cabe lembrar que o teste F é robusto a desvios de normalidade em amostras de tamanho moderado a grande, de modo que sua aplicação permanece adequada apesar da natureza ordinal das variáveis.

No Dataset 1, os testes ANOVA revelaram diferenças estatisticamente significativas entre os níveis de estresse para variáveis centrais de saúde mental e desempenho acadêmico. Observou-se efeito significativo para nível de ansiedade ($F = 655{,}45$; $p < 0{,}001$), depressão ($F = 652{,}63$; $p < 0{,}001$) e desempenho acadêmico ($F = 639{,}22$; $p < 0{,}001$), indicando que essas variáveis variam sistematicamente conforme o aumento do estresse. Esses achados reforçam a forte associação entre estresse elevado, sofrimento psicológico e prejuízos no funcionamento acadêmico.

No Dataset 2, a ANOVA também apontou diferenças significativas entre os níveis de experiência recente de estresse para variáveis relacionadas a respostas fisiológicas e emocionais imediatas. Destacam-se os efeitos observados para batimentos cardíacos acelerados ou palpitações ($F = 51{,}94$; $p < 0{,}001$), ansiedade ou tensão recente ($F = 12{,}15$; $p < 0{,}001$) e conflitos entre atividades acadêmicas e extracurriculares ($F = 12{,}37$; $p < 0{,}001$). Esses resultados indicam que a intensificação do estresse recente está associada a manifestações

agudas tanto fisiológicas quanto contextuais.

De forma integrada, a análise estatística confirma que há diferenças expressivas e consistentes entre os grupos de estresse em ambos os *datasets*, em consonância com o padrão estrutural *versus* reativo já descrito na Seção A.2.1. A natureza ordinal e discreta das variáveis-alvo, como discutido acima, não compromete essa conclusão, dado o uso de testes robustos a desvios de normalidade e o fato de as decisões de modelagem se ancorarem na *Mutual Information* e na estrutura de correlação, não na hipótese de normalidade.

## A.3 MACROFASE III: MODELAGEM PREDITIVA E *ENSEMBLE*

A terceira macrofase teve como objetivo avaliar o desempenho de diferentes algoritmos de aprendizado de máquina na classificação multiclasse dos níveis de estresse estudantil, considerando os dois conjuntos de dados adotados neste trabalho. Foram testados diferentes classificadores sob combinações específicas de pré-processamento, seleção de atributos e redução de dimensionalidade, com o propósito de identificar as configurações mais adequadas para cada base analisada.

Essa etapa foi relevante não apenas por permitir a comparação entre modelos preditivos, mas também por fornecer evidências empíricas sobre a estrutura dos dados e sobre os atributos mais importantes para a discriminação dos níveis de estresse. Assim, a modelagem preditiva foi utilizada tanto como instrumento de validação do potencial classificatório dos *datasets* quanto como suporte para a etapa posterior de construção do Indicador Sintético de Bem-Estar Estudantil (ISBE).

### A.3.1 Desempenho dos modelos individuais

A análise dos modelos individuais demonstrou que o desempenho dos classificadores variou conforme a estrutura de cada *dataset* e a técnica de transformação aplicada aos dados. No Dataset 1, observou-se melhor desempenho das configurações associadas ao *SelectKBest*, indicando que a seleção prévia dos atributos mais relevantes favoreceu a capacidade de generalização dos modelos. Nesse conjunto, o melhor resultado foi obtido pelo *Random Forest* com *SelectKBest* (92,364% de acurácia), seguido por AdaBoost (92,000%), XGBoost (91,636%) e *Gradient Boosting* (91,273%).

No Dataset 2, por sua vez, verificou-se predominância da configuração com PCA, especialmente para o *Support Vector Machine*, que alcançou 99,052% de acurácia, superando os demais modelos individuais. Esse comportamento indica que, nesse conjunto de dados, a redução de dimensionalidade contribuiu significativamente para tornar a separação entre as classes mais eficiente, favorecendo o desempenho de um classificador baseado em margens máximas. Outros algoritmos também apresentaram resultados elevados, como AdaBoost (96,683%), XGBoost (96,209%), *Gradient Boosting* (95,735%), *Random Forest* (94,787%) e *Bagging* (94,313%).

De forma geral, os resultados evidenciam que o desempenho preditivo depende não apenas do algoritmo empregado, mas também da natureza dos dados e da estratégia de transformação adotada. Isso reforça a importância de uma abordagem experimental abrangente, capaz de explorar diferentes combinações metodológicas antes da definição do modelo mais adequado.

Tabela 23 – Desempenho dos melhores modelos individuais no Dataset 1

| **Modelo** | **Config.** | **Acur. (%)** | **F1 (%)** | **Recall (%)** | **Prec. (%)** |
|---|---|---|---|---|---|
| AdaBoost | SelectKBest | 92,000 | 92,005 | 92,083 | 92,121 |
| Bagging | SelectKBest | 89,455 | 89,493 | 89,508 | 89,716 |
| Gradient Boost. | SelectKBest | 91,273 | 91,279 | 91,404 | 91,498 |
| Random Forest | SelectKBest | 92,364 | 92,365 | 92,356 | 92,378 |
| SVM | PCA | 90,546 | 90,557 | 90,551 | 90,567 |
| XGBoost | SelectKBest | 91,636 | 91,636 | 91,776 | 92,014 |

Fonte: Elaborado pelo autor (2026).

Os resultados apresentados na Tabela 23 indicam que o Dataset 1 apresentou desempenho elevado para diferentes classificadores, com pequena variação entre os melhores modelos. Esse comportamento sugere que a base possui sinais informativos consistentes, embora distribuídos de maneira relativamente heterogênea entre as variáveis, o que torna a seleção de atributos uma etapa particularmente importante para o ganho de desempenho.

Tabela 24 – Desempenho dos melhores modelos individuais no Dataset 2

| **Modelo** | **Config.** | **Acur. (%)** | **F1 (%)** | **Recall (%)** | **Prec. (%)** |
|---|---|---|---|---|---|
| AdaBoost | SelectKBest | 96,683 | 85,866 | 79,372 | 94,818 |
| Bagging | SelectKBest | 94,313 | 68,831 | 59,091 | 98,039 |
| Gradient Boost. | SelectKBest | 95,735 | 79,044 | 69,318 | 98,508 |
| Random Forest | SelectKBest | 94,787 | 72,332 | 63,258 | 98,194 |
| SVM | PCA | 99,052 | 96,494 | 93,939 | 99,656 |
| XGBoost | SelectKBest | 96,209 | 82,058 | 73,485 | 98,667 |

Fonte: Elaborado pelo autor (2026).

Conforme observado na Tabela 24, o Dataset 2 apresentou resultados superiores aos do Dataset 1, com destaque para o desempenho do *Support Vector Machine* associado ao PCA. Esse resultado sugere que a estrutura desse conjunto de dados favorece fortemente abordagens baseadas em transformação dimensional, permitindo uma separação entre classes mais eficiente e, consequentemente, melhor capacidade preditiva.

### A.3.2 Estratégias de *ensemble*

Após a análise dos modelos individuais, foram avaliadas estratégias de *ensemble* com o objetivo de verificar se a combinação de classificadores poderia produzir ganhos adicionais de desempenho e robustez. Para isso, foram consideradas arquiteturas de votação e empilhamento, contemplando *hard voting*, *soft voting*, *weighted hard voting*, *weighted soft voting* e *stacking classifier*.

Os resultados mostraram que os *ensembles*, em geral, superaram ou igualaram os melhores modelos individuais. No Dataset 1, os melhores desempenhos foram observados para o *Voting Classifier* (*hard*) e o *Voting Classifier* (*weighted hard*), ambos com 93,091% de acurácia, valor superior ao melhor resultado individual desse conjunto. No Dataset 2,

o destaque foi o *Stacking Classifier*, com 99,530% de acurácia, superando inclusive o desempenho do SVM com PCA, que já havia obtido resultado bastante elevado (Tabela 25).

Tabela 25 – Desempenho consolidado das estratégias de *ensemble*

| **Estratégia de *Ensemble*** | **D1 – Acur. (%)** | **D2 – Acur. (%)** |
|---|---|---|
| Voting Classifier (hard) | 93,091 | 97,156 |
| Voting Classifier (soft) | 91,636 | 97,156 |
| Voting Classifier (weighted hard) | 93,091 | 97,156 |
| Voting Classifier (weighted soft) | 91,636 | 97,156 |
| Stacking Classifier | 92,727 | 99,530 |

Fonte: Elaborado pelo autor (2026).

### A.3.3 Síntese dos achados preditivos

De modo geral, a Macrofase III confirmou que os *datasets* analisados possuem alto potencial preditivo para classificação de estresse estudantil. Os modelos individuais já apresentaram desempenho robusto, enquanto os *ensembles* mostraram capacidade de elevar ainda mais os resultados, sobretudo no Dataset 2. Esses achados reforçam a consistência da abordagem metodológica adotada e demonstram que a combinação entre pré-processamento, transformação de atributos e modelagem preditiva é adequada para o problema investigado.

Mais do que identificar o classificador de melhor desempenho, esta macrofase possibilitou compreender quais atributos exerceram maior influência nas decisões dos modelos. Esse aspecto é especialmente importante porque o propósito final deste trabalho não se limita à previsão dos níveis de estresse, mas busca também subsidiar a construção de um indicador sintético interpretável, capaz de refletir, de forma estruturada, os fatores mais relevantes associados ao fenômeno.

### A.3.4 Importância das *features* como transição para a Macrofase IV

Como desdobramento da etapa de modelagem, foi realizada a consolidação das importâncias das *features* nos modelos baseados em árvore reproduzidos na configuração de dados originais, sem PCA e sem seleção prévia, considerando seis algoritmos com importância nativa de atributos: *Random Forest*, *Gradient Boosting*, AdaBoost, XGBoost, LightGBM e *Bagging*. As importâncias foram normalizadas de modo que a soma em cada modelo fosse igual a 1, o que permitiu comparar diretamente os pesos atribuídos por cada algoritmo e calcular uma média consolidada por variável.

Os seis algoritmos calculam importância por mecanismos distintos, todos derivados da estrutura interna das árvores treinadas. *Random Forest* e *Bagging* baseiam-se na *Mean Decrease in Impurity* (MDI): para cada divisão de nó em uma árvore, registra-se a redução de impureza de Gini ponderada pela proporção de amostras que atinge o nó; a importância de uma *feature* é a soma dessas reduções ao longo de todas as árvores do *ensemble*. *Gradient Boosting* segue o mesmo princípio, com a diferença de que cada árvore é treinada para corrigir os resíduos da anterior, e a importância acumula as reduções de impureza ponderadas pela taxa de aprendizado de cada estágio. AdaBoost combina a MDI das árvores-base com o peso ($\alpha$) atribuído a cada uma na votação final, refletindo

sua confiabilidade preditiva. XGBoost reporta importância como *gain*, a melhora média na função objetivo (com regularização) produzida pelas divisões que utilizam a *feature*. LightGBM oferece duas variantes, *split* (contagem de divisões) e *gain* (redução acumulada na função objetivo); neste trabalho adotou-se a segunda, por ser conceitualmente equivalente à do XGBoost e por refletir a contribuição efetiva, não a mera frequência de uso, dado que *split* tende a inflar artificialmente *features* com muitos valores possíveis. Para o *Bagging*, cuja API expõe apenas as importâncias dos estimadores individuais sobre subconjuntos aleatórios de *features*, a importância global foi reconstruída agregando os vetores de cada estimador às posições das *features* que ele utilizou, com posterior média sobre o número total de estimadores.

Embora todos os modelos retornem vetores não negativos, suas escalas brutas diferem: importâncias de XGBoost e LightGBM em modo *gain* possuem magnitudes absolutas distintas das demais, e nem todas as implementações garantem soma unitária. Para tornar os pesos comparáveis e permitir uma média aritmética entre os seis modelos, cada vetor foi normalizado para soma igual a 1 antes da consolidação. Essa transformação afim preserva integralmente a ordem e a razão entre as importâncias dentro de cada modelo, mas iguala a contribuição total dos seis vetores na média final, evitando que algoritmos com magnitudes brutas maiores dominem o resultado consolidado.

Tabela 26 – Importância normalizada das *features* por modelo e média consolidada

| **Feature** | **RF** | **GB** | **Ada** | **XGB** | **LGB** | **Bag** | **Média** | **Rank** |
|---|---|---|---|---|---|---|---|---|
| blood_pressure | 0,128 | 0,092 | 0,387 | 0,362 | 0,290 | 0,254 | 0,252 | 1 |
| academic_perf. | 0,068 | 0,069 | 0,026 | 0,103 | 0,189 | 0,118 | 0,095 | 2 |
| sleep_quality | 0,081 | 0,059 | 0,022 | 0,088 | 0,138 | 0,066 | 0,076 | 3 |
| teacher_stud. | 0,072 | 0,073 | 0,048 | 0,098 | 0,045 | 0,090 | 0,071 | 4 |
| social_support | 0,058 | 0,061 | 0,137 | 0,055 | 0,030 | 0,059 | 0,067 | 5 |
| depression | 0,087 | 0,076 | 0,043 | 0,050 | 0,040 | 0,088 | 0,064 | 6 |
| self_esteem | 0,040 | 0,043 | 0,096 | 0,026 | 0,083 | 0,058 | 0,058 | 7 |
| safety | 0,066 | 0,058 | 0,004 | 0,024 | 0,011 | 0,043 | 0,034 | 8 |
| headache | 0,061 | 0,019 | 0,027 | 0,031 | 0,032 | 0,035 | 0,034 | 9 |
| anxiety_level | 0,059 | 0,037 | 0,036 | 0,014 | 0,017 | 0,037 | 0,033 | 10 |
| extracurr. | 0,030 | 0,052 | 0,038 | 0,032 | 0,021 | 0,024 | 0,033 | 11 |
| peer_pressure | 0,045 | 0,078 | 0,021 | 0,013 | 0,014 | 0,021 | 0,032 | 12 |
| bullying | 0,051 | 0,083 | 0,009 | 0,009 | 0,007 | 0,021 | 0,030 | 13 |
| future_career | 0,058 | 0,071 | 0,009 | 0,013 | 0,009 | 0,016 | 0,029 | 14 |
| basic_needs | 0,034 | 0,069 | 0,004 | 0,017 | 0,018 | 0,026 | 0,028 | 15 |
| study_load | 0,014 | 0,015 | 0,027 | 0,016 | 0,019 | 0,009 | 0,017 | 16 |
| noise_level | 0,016 | 0,014 | 0,025 | 0,014 | 0,011 | 0,012 | 0,015 | 17 |
| breathing_prob. | 0,012 | 0,015 | 0,027 | 0,013 | 0,012 | 0,009 | 0,015 | 18 |
| living_cond. | 0,013 | 0,013 | 0,008 | 0,012 | 0,016 | 0,012 | 0,012 | 19 |
| mental_health | 0,007 | 0,004 | 0,006 | 0,009 | 0,002 | 0,003 | 0,005 | 20 |

Fonte: Elaborado pelo autor (2026).

A Tabela 26 mostra que as variáveis mais influentes para os modelos não coincidem necessariamente com aquelas que se destacam apenas em análises lineares. O caso de `blood_pressure` é particularmente relevante, pois sua elevada importância média indica que sua relação com o estresse é melhor capturada por modelos capazes de representar interações e não linearidades. Esse resultado fornece uma base empírica importante para a etapa seguinte do trabalho.

Dessa forma, a Macrofase III não se encerra apenas com a identificação dos classifi-

cadores de melhor desempenho, mas também com a produção de evidências interpretativas sobre os fatores que mais sustentam a classificação do estresse estudantil. É justamente essa evidência que fundamenta a transição para a Macrofase IV, na qual a importância das *features* é incorporada à triangulação com a correlação de Pearson e a *Mutual Information* para orientar a construção e a validação do ISBE.

**MINISTÉRIO DA EDUCAÇÃO**
**CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS**
**DEPARTAMENTO DE COMPUTAÇÃO - NG**

**FOLHA DE APROVAÇÃO DE TCC Nº 20 / 2026 - DECOM (11.56.03)**

**Nº do Protocolo: 23062.032963/2026-91**

**Belo Horizonte-MG, 25 de junho de 2026.**

**LUIZ GUSTAVO BARBOSA CAMARGOS**

# DA ANÁLISE EXPLORATÓRIA À SÍNTESE DIMENSIONAL: DESENVOLVIMENTO DE UM INDICADOR SINTÉTICO DE BEM-ESTAR ESTUDANTIL ATRAVÉS DE MODELAGEM PREDITIVA

Trabalho de Conclusão de Curso apresentado ao Curso de Engenharia de Computação do Centro Federal de Educação Tecnológica de Minas Gerais, do Campus Nova Gameleira, como parte do requisito para obtenção do grau de Engenheiro(a) de Computação.

Aprovado em 19 de junho de 2026.

Banca Examinadora:

Prof. D.Sc. Ismael Santana Silva - Orientador
Centro Federal de Educação Tecnológica de Minas Gerais

Prof. D.Sc. Bruno André Santos
Centro Federal de Educação Tecnológica de Minas Gerais

Prof. D.Sc. Fábio Rocha da Silva
Centro Federal de Educação Tecnológica de Minas Gerais

**Belo Horizonte**
**2026**

***(Assinado digitalmente em 25/06/2026 16:14 )***
BRUNO ANDRE SANTOS
*PROFESSOR DO MAGISTERIO SUPERIOR*
*DECOM (11.56.03)*
*Matrícula: 1459458*

***(Assinado digitalmente em 25/06/2026 19:29 )***
FABIO ROCHA DA SILVA
*PROFESSOR ENS BASICO TECN TECNOLOGICO*
*DECOM (11.56.03)*
*Matrícula: 1703246*

***(Assinado digitalmente em 25/06/2026 15:39 )***
ISMAEL SANTANA SILVA
*PROFESSOR ENS BASICO TECN TECNOLOGICO*
*DECOM (11.56.03)*
*Matrícula: 1028758*

Visualize o documento original em **https://sig.cefetmg.br/public/documentos/index.jsp** informando seu número: **20**, ano: **2026**, tipo: **FOLHA DE APROVAÇÃO DE TCC**, data de emissão: **25/06/2026** e o código de verificação: **5ac6e776f6**