

# Monitoramento e Análise Preditiva do Estresse em Estudantes Universitários

Ana Clara Guerra - 108205

Thales Barcelos de Castro - 116229

**Abstract**—O estresse em estudantes universitários representa um desafio crescente na saúde pública, impactando negativamente o desempenho acadêmico e o bem-estar. O presente trabalho propõe o desenvolvimento e a análise de modelos de Aprendizado de Máquina para a predição do nível de estresse (classificação multiclasse) em 1100 estudantes. Utilizamos o *Student Stress Monitoring Datasets*, aplicando técnicas de seleção de *features* que se mostraram cruciais para otimizar a performance. O modelo de melhor desempenho foi o MLP (Multilayer Perceptron), que alcançou uma acurácia de aproximadamente 91%. Para decifrar a natureza "caixa-preta" dos modelos preditivos, empregamos técnicas de *eXplainable AI (XAI)*: SHAP (para análise global) e LIME (para análise local). Os resultados de explicabilidade revelaram que a predição do estresse é dominantemente influenciada por fatores fisiológicos (como '*blood\_pressure*' e '*sleep\_quality*') e psicossociais (como '*selfEsteem*' e '*depression*'), validando a integridade do pipeline de classificação. Os achados fornecem *insights* valiosos para a criação de intervenções direcionadas ao manejo do estresse acadêmico.

**Index Terms**—Estresse Universitário, Aprendizado de Máquina, *eXplainable AI (XAI)*, Classificação Multiclasse.

## 1 INTRODUÇÃO

O ambiente universitário, caracterizado por altas demandas acadêmicas, pressão social e incerteza sobre o futuro profissional, configura-se como um cenário de alto risco para o desenvolvimento de níveis elevados de estresse [Lipp 2004]. O estresse não é apenas um problema psicológico; ele se manifesta em uma complexa interconexão de sintomas fisiológicos (como distúrbios do sono e variações na pressão arterial), psicológicos (depressão e ansiedade) e sociais, impactando diretamente o bem-estar e o desempenho acadêmico dos estudantes [Bao et al. 2019]. A incapacidade de lidar com o estresse negativo, ou angústia, pode levar a consequências graves, incluindo o abandono dos estudos e o agravamento de quadros de saúde mental.

Com o avanço da Ciência de Dados, torna-se viável a construção de ferramentas objetivas e preditivas para auxiliar na saúde mental. O *Machine Learning* (ML) oferece a capacidade de processar e modelar essa complexidade de fatores ('*headache*', '*sleep\_quality*', '*anxiety\_level*', etc.) de forma simultânea, permitindo a **classificação do nível de estresse** em categorias bem definidas (0: Sem estresse, 1: Estresse positivo, 2: Estresse negativo). Todas as etapas do pipeline — desde pré-processamento até explicabilidade com SHAP e LIME — foram documentadas e disponibilizadas publicamente no repositório GitHub associado ao projeto [Guerra e Castro 2025].

### 1.1 Problema e Justificativa

A dificuldade central no manejo do estresse se encontra na sua natureza multifatorial e na falta de mecanismos de diagnóstico e monitoramento objetivos em tempo real. Embora algoritmos avançados, como *Multilayer Perceptron* (MLP) demonstre alta capacidade preditiva, ele é considerado "caixa-preta" [Ribeiro, Singh e Guestrin 2016]. Essa falta de transparência impede que especialistas da área da saúde compreendam quais *features* e em que medida elas estão, de fato, direcionando o modelo para uma predição de estresse.

O problema de pesquisa abordado é, portanto, duplo: (i) desenvolver modelos de ML de alta *performance* para a classificação do nível de estresse (multiclasse); e (ii) explicabilidade desses modelos.

A justificativa do trabalho é fornecer uma ferramenta preditiva e interpretável. Ao integrar *eXplainable AI (XAI)* através das técnicas SHAP e LIME, não apenas validamos a robustez das predições, mas também geramos *insights* transparentes sobre a contribuição global e local (para um estudante específico) de cada fator, transformando o modelo preditivo em uma ferramenta de diagnóstico auxiliar.

### 1.2 Motivações

As motivações deste estudo são intrinsecamente ligadas ao impacto prático e científico. Motivamo-nos a: (i) oferecer uma metodologia robusta para o monitoramento da saúde mental, capaz de identificar alunos em estado de angústia antes que o estresse se torne crônico; e (ii) superar a barreira da interpretabilidade inerente aos modelos de alta *performance*. Os *insights* gerados pela análise de explicabilidade são vitais para a comunidade acadêmica e psicológica, pois apontam com clareza quais fatores (fisiológicos, psicológicos, sociais e ambientais) devem ser o foco primário das intervenções e políticas de apoio ao estudante.

### 1.3 Objetivos

Com base nos embasamentos descritos, o objetivo geral é desenvolver e analisar modelos preditivos de Aprendizado de Máquina para a classificação do nível de estresse em estudantes universitários, utilizando técnicas de XAI para interpretar a contribuição global e local dos fatores envolvidos. Especificamente, pretende-se:

- Avaliar o impacto de diferentes técnicas de **seleção de features** (*Mutual Information*, *Recursive Feature Elimination (RFE)* e *Random Forest Importance (RF)*) no desempenho dos modelos preditivos;
- Treinar, refinar e comparar estatisticamente diversos algoritmos de *Machine Learning* (incluindo MLP e *Random Forest*) para identificar o modelo mais robusto;
- Utilizar **SHAP** para identificar a importância global dos fatores (fisiológicos, psicológicos, sociais, ambientais e acadêmicos) na predição do estresse;
- Aplicar **LIME** para fornecer explicações locais, validando a integridade das decisões do modelo "caixa-preta" (MLP) em casos críticos de acerto e erro.

## 2 MATERIAIS E MÉTODOS

### 2.1 Ambiente Computacional

O desenvolvimento dos modelos foi realizado na linguagem Python (v3.8+), utilizando o ambiente Google Colab. As principais bibliotecas empregadas incluíram *Scikit-learn* para modelagem, *Pandas* e *NumPy* para manipulação de dados, e *SHAP* e *LIME* para explicabilidade. A visualização final e a interface do usuário foram construídas

• E-mail: [ana.torres1@ufv.br](mailto:ana.torres1@ufv.br);  
• E-mail: [thales.castro@ufv.br](mailto:thales.castro@ufv.br);

no Microsoft Power BI Desktop, integrando os dados processados via arquivos CSV.

## 2.2 Conjunto de Dados e Pré-processamento

O conjunto de dados utilizado neste estudo é denominado *Student Stress Factors: A Comprehensive Analysis*, um dataset público e colaborativo disponibilizado via Google Forms e posteriormente publicado na plataforma Kaggle. Ele reúne respostas de 1.100 estudantes universitários, com idades entre 18 e 21 anos, referentes a diversos fatores associados ao estresse estudantil. O arquivo é fornecido em formato CSV, com tamanho aproximado de 48 KB, contendo 21 variáveis independentes e uma variável alvo multiclasses denominada *stress\_level*.

### 2.2.1 Descrição Geral das Variáveis

As variáveis capturam dimensões psicológicas, fisiológicas, sociais, ambientais e acadêmicas relacionadas ao cotidiano do estudante. Abaixo, detalhamos os domínios representados e as respectivas escalas de medição:

- **Psicológicos:** *anxiety\_level* (0-21), *selfEsteem* (0-30), *mentalHealthHistory* (0-1), *depression* (0-27);
- **Fisiológicos:** *sleep\_quality* (0-5), *blood\_pressure* (0-3), *headache* (0-5), *breathing\_problem* (0-5);
- **Ambientais:** *noise\_level* (0-5), *living\_conditions* (0-5), *safety* (0-5), *basic\_needs* (0-5);
- **Sociais:** *peer\_pressure* (0-5), *social\_support* (0-3), *bullying* (0-5), *extracurricular\_activities* (0-5);
- **Acadêmicos:** *study\_load* (0-5), *academic\_performance* (0-5), *teacher\_student\_relationship* (0-5), *future\_career\_concerns* (0-5).

A variável alvo *stress\_level* apresenta três categorias:

- 0: Sem/baixo estresse;
- 1: Estresse positivo;
- 2: Angústia (estresse negativo).

O dataset não apresenta valores ausentes nem registros duplicados, e sua distribuição de classes é equilibrada, o que elimina a necessidade de técnicas de balanceamento supervisionado.

### 2.2.2 Pré-processamento dos Dados

As etapas de pré-processamento foram conduzidas conforme documentado nos notebooks disponibilizados no repositório GitHub do projeto [Guerra e Castro 2025], abrangendo desde a inspeção inicial até a preparação final para os modelos de aprendizado supervisionado.

### 2.2.3 Análise Inicial e Limpeza

Uma primeira inspeção foi realizada com o objetivo de verificar valores ausentes (nenhum identificado), registros duplicados (não encontrados), tipos das variáveis e a consistência dos valores nas escalas *Likert* e binárias. Para conferir rigorosamente se todas as variáveis estavam dentro do intervalo esperado e para verificar a presença de *outliers* que pudessem distorcer a modelagem, foi gerado um *boxplot* para todas as *features* numéricas.

A análise visual da distribuição, conforme ilustrada na Figura 1, confirmou que as variáveis estavam corretamente tipadas e **não foi detectado nenhum outlier** ou valor inconsistente em relação às escalas de medição definidas, o que eliminou a necessidade de correções adicionais nessa fase.

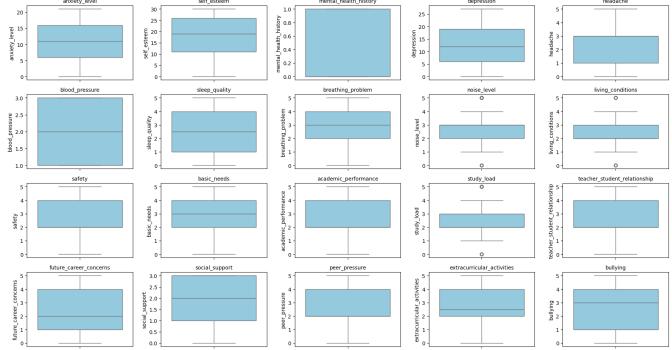


Fig. 1. Boxplot das variáveis do *dataset* para identificação de *outliers* e verificação da consistência das escalas de medição.

### 2.2.4 Normalização e Codificação

Como os modelos avaliados incluíam tanto algoritmos sensíveis à escala (como regressão logística e k-NN) quanto métodos baseados em árvores (como Random Forest e Gradient Boosting), adotou-se a seguinte estratégia:

- As features numéricas foram normalizadas utilizando *StandardScaler* e salvas para usar posteriormente como *StressLevelDataset\_Scaled.csv*;
- A variável alvo foi mantida como inteiro, respeitando a natureza categórica multiclasse.

A normalização foi aplicada exclusivamente aos dados de treino, prevenindo *data leakage*.

### 2.2.5 Divisão Treino-Teste

A base foi dividida em 70% para treino e 30% para teste, utilizando estratificação da variável alvo, o que garantiu a preservação das proporções entre as classes em ambos os subconjuntos.

### 2.2.6 Tratamento das Features

Embora o conjunto de dados seja composto majoritariamente por variáveis numéricas discretas, foi necessário identificar quais delas contribuíram de forma significativa para a predição do nível de estresse. Para isso, aplicamos diferentes abordagens de seleção de *features*, de modo a comparar metodologias e reduzir dimensionalidade sem perda de desempenho.

**Métodos Aplicados** Foram aplicadas três abordagens complementares:

- **Filtro — Mutual Information (MI):** Mede a dependência estatística entre cada variável independente e a variável alvo, avaliando relevância individual.
- **Wrapper — Recursive Feature Elimination (RFE) com Regressão Logística:** Remove iterativamente as variáveis menos relevantes otimizando o modelo ao longo do processo.
- **Embedded — Importância das Features via Random Forest:** Atribui importância com base na redução de impureza em árvores de decisão ao longo do treinamento.

O resultado comparativo desses métodos está sintetizado no heatmap da Figura 2, que mostra o ranqueamento relativo das variáveis.

**Conclusões da Seleção de Features** Os três métodos revelaram padrões consistentes no comportamento das variáveis:

- **Preditores fisiológicos foram dominantes**, especialmente *blood\_pressure*, *sleep\_quality* e *headache*.
- **Variáveis psicológicas como *self\_esteem*, *depression* e *anxiety\_level*** apresentaram relevância moderada a alta.
- **Variáveis sociais como *bullying* e *social\_support*** também surgiram entre os fatores mais determinantes em alguns métodos.
- **Variáveis com menor relevância:** *mental\_health\_history*, *study\_load*, *living\_conditions* foram classificadas consistentemente como pouco informativas.

Com base nos resultados obtidos por cada técnica de seleção de variáveis, foram gerados subconjuntos distintos para experimentação comparativa dos modelos. Ao invés de utilizar apenas um ranking unificado, optou-se por preservar os resultados específicos de cada abordagem para permitir análises mais rigorosas sobre seu impacto no desempenho do modelo.

Assim, para cada método (*Mutual Information*, *Recursive Feature Elimination* — *RFE*, e *Random Forest Importance*), foram criados dois subconjuntos reduzidos contendo as **Top 10** e **Top 5** variáveis mais relevantes. Além disso, foi mantido o conjunto original com todas as variáveis. No total, os seguintes arquivos foram gerados e armazenados para uso nos experimentos posteriores:

- **Dataset completo:** *X\_all.csv*
- **Mutual Information:** *X\_mi\_top10.csv*, *X\_mi\_top5.csv*
- **Recursive Feature Elimination (RFE):** *X\_rfe\_top10.csv*, *X\_rfe\_top5.csv*
- **Random Forest Importance:** *X\_rf\_top10.csv*, *X\_rf\_top5.csv*

Essa organização permitiu avaliar não apenas o impacto da redução dimensional, mas também comparar a influência metodológica de cada técnica na performance dos modelos treinados nas etapas seguintes.

**Análise Exploratória com PCA e LDA** Para avaliar a separabilidade entre as classes após a seleção das variáveis, técnicas de redução de dimensionalidade foram aplicadas:

- **PCA (Principal Component Analysis):** utilizado com finalidade exploratória para observar a variância explicada e possíveis agrupamentos naturais.
- **LDA (Linear Discriminant Analysis):** utilizado para visualizar a separabilidade supervisionada entre as classes da variável alvo.

Os resultados estão apresentados nas Figuras 3 e 4.

**Síntese dos Achados** Os resultados indicaram que:

- Existe uma **separabilidade parcial** entre as classes de estresse, especialmente entre as classes 0 (sem estresse) e 2 (estresse negativo).
- Modelos lineares tendem a capturar apenas parte da estrutura dos dados, enquanto modelos não lineares são mais adequados.
- A interação entre fatores fisiológicos, psicológicos e sociais sugere um **mecanismo multifatorial** que exige classificadores capazes de lidar com relações não lineares — justificando a escolha posterior por algoritmos como *Random Forest*, *Gradient Boosting* e *MLP*.

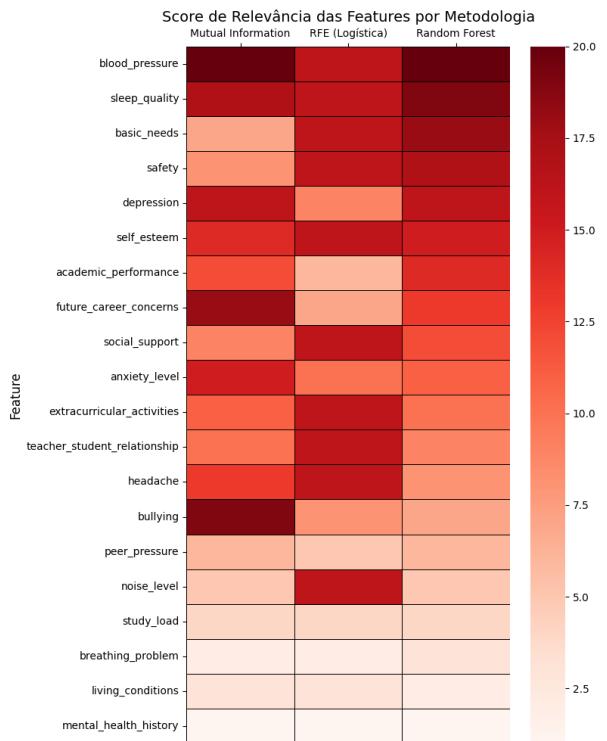


Fig. 2. Heatmap comparativo dos métodos de seleção de features.

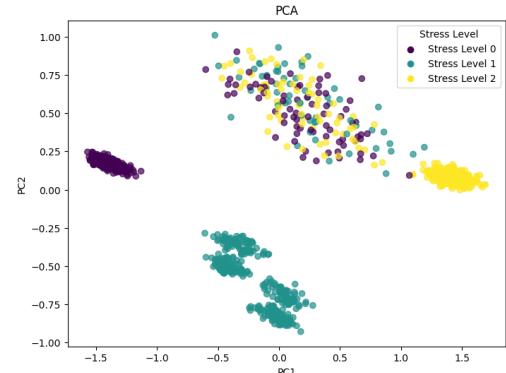


Fig. 3. Visualização PCA das classes do nível de estresse.

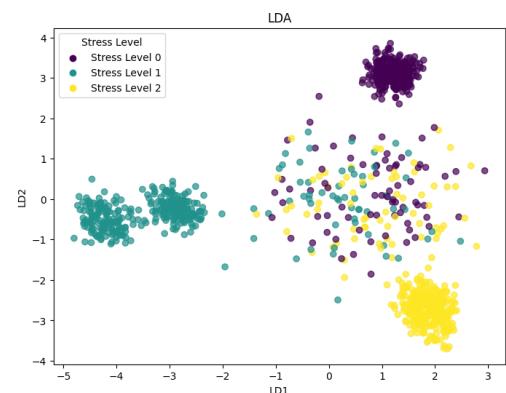


Fig. 4. Visualização LDA evidenciando separabilidade parcial entre classes.

### 2.2.7 Preparação Final e Pipeline

Após a definição dos subconjuntos de *features* selecionados pelos métodos Mutual Information, RFE e Random Forest (`X_mi_top10.csv`, `X_mi_top5.csv`, `X_rfe_top10.csv`, `X_rfe_top5.csv`, `X_rf_top10.csv`, `X_rf_top5.csv` e `X_all.csv`), foi estruturado um *pipeline* completo para garantir rigor metodológico e reprodutibilidade em todas as etapas do aprendizado supervisionado.

Esse *pipeline*, documentado no notebook `1_Preparacao.ipynb`, incluiu:

- separação estratificada da base em treino e teste (80% / 20%), garantindo preservação proporcional das classes;
- normalização dos atributos numéricos utilizando *StandardScaler*, aplicada exclusivamente ao conjunto de treino para prevenir *data leakage*;
- padronização do tipo de dados e organização das variáveis conforme o subconjunto analisado;
- exportação dos conjuntos finais para arquivos CSV nomeados de forma sistemática, a fim de possibilitar a comparação justa entre diferentes experimentos de modelagem;
- integração direta com os algoritmos de classificação testados, permitindo mudanças estruturadas e automáticas nas entradas dos modelos.

Além disso, o pipeline foi projetado para possibilitar a substituição de qualquer subconjunto de *features* com mínimo retrabalho computacional, permitindo avaliar com clareza o impacto da seleção de atributos no desempenho dos modelos preditivos.

## 2.3 Algoritmos de Aprendizado de Máquina

Após a etapa de seleção de *features* (Seção ??), o processo de modelagem consistiu no treinamento de diversos algoritmos de *Machine Learning* para a classificação multiclasse do nível de estresse (0, 1 e 2). O objetivo foi identificar o modelo com a melhor *performance* em cada um dos subconjuntos de *features* gerados (*Todas*, *Top 10* e *Top 5*).

### 2.3.1 Justificativa dos Hiperparâmetros baseada em PCA e LDA

As análises de PCA e LDA (Seção 2.4) demonstraram que, embora haja uma separabilidade parcial entre os níveis de estresse, existe uma sobreposição significativa entre as classes no espaço de *features* reduzido. Essa conclusão foi fundamental para a escolha e ajuste dos hiperparâmetros, pois:

- **Regra:** Modelos simples e lineares poderiam apresentar baixo desempenho;
- **Consequência:** A necessidade de modelos **não-lineares** (ex: SVM-RBF, Árvores de Decisão) e **robustos** (*Random Forest*, *Gradient Boosting*, *MLP*);
- **Ajuste:** Hiperparâmetros foram ajustados para evitar o *overfitting* (ex: limitando a profundidade de árvores), já que a sobreposição indica que a separação é complexa, mas o modelo não deve memorizar o ruído.

### 2.3.2 Modelos de Comparação e Hiperparâmetros

Os modelos de comparação e suas configurações definidas foram:

- **Regressão Logística:** Definido com `max_iter = 1000` (para garantir a convergência) e `random_state = 42`. Este modelo, de natureza linear, foi incluído para avaliar a eficácia da separação linear do problema. A sobreposição significativa identificada no PCA e LDA sugeriu que, embora a Regressão Logística possa ser útil, modelos não-lineares seriam necessários para capturar todas as relações.

• **k-Nearest Neighbors (kNN):** Configurado com `n_neighbors = 7` e `weights = 'distance'`. O *kNN* é um modelo baseado em distância. A escolha de um número pequeno de vizinhos ( $k = 7$ ) e o peso 'distance' foram definidos para que o modelo priorizasse a separação local, o que era visível nos agrupamentos densos sugeridos pelo LDA.

• **Support Vector Machine (SVM):** `kernel = "rbf"`, `C = 1.0` e `gamma = 'scale'`. Para lidar com a não-linearidade e a sobreposição das classes observadas no PCA e LDA, o *kernel* de Base Radial (*RBF*) foi escolhido. Este *kernel* mapeia os dados para um espaço de maior dimensão, facilitando a separação não-linear.  $C = 1.0$  e `gamma = 'scale'` fornecem um balanço inicial entre margem de separação e penalidade de erro.

• **Modelos Baseados em Árvores:** Devido à capacidade natural de lidar com relações não-lineares nos *features* (confirmada pelo PCA/LDA), modelos baseados em árvores foram considerados cruciais.

– *Decision Tree*: `max_depth = 6`. A profundidade limitada a 6 foi uma escolha moderada, garantindo que o modelo capture as relações complexas sem se tornar excessivamente específico e propenso ao *overfitting*.

– *Random Forest* (RF): `n_estimators = 200` e `max_depth = 6`. Utilizou-se um grande número de estimadores (200) para aumentar a robustez e a profundidade limitada (6) para promover a generalização.

– *Gradient Boosting* (GB): `n_estimators = 250`, `learning_rate = 0.08` e `max_depth = 4`. Para o GB, a estratégia é aprender de forma gradual. Foi adotado um *learning\_rate* baixo (0.08) e um alto número de estimadores (250) para um ajuste lento e preciso, minimizando o risco de *overfitting* no contexto de classes sobrepostas.

• **Multilayer Perceptron (MLP):** A configuração inicial foi definida com `hidden_layer_sizes=(50,)`, `activation='relu'`, `solver='adam'`, e `max_iter=1000`. Dada a necessidade de modelos não-lineares, o MLP foi um candidato ideal. Foi utilizada uma estrutura inicial simples, com uma camada intermediária moderada (50 neurônios), que serviu como ponto de partida para o refinamento exaustivo da topologia (Seção 2.5.3).

## 3 RESULTADOS E DISCUSSÃO

### 3.1 Desempenho dos Modelos

Após o treinamento supervisionado utilizando os diferentes subconjuntos de *features* (*Todas*, *Top 10* e *Top 5*), cada algoritmo foi avaliado no conjunto de teste com base em duas métricas principais: Acurácia e F1-score Macro. Essas métricas foram selecionadas por fornecerem uma avaliação robusta em cenários multiclasse e balanceada.

A Figura 5 apresenta uma comparação visual consolidada das métricas de desempenho para todos os modelos treinados, conforme documentado no notebook `2_Modelagem.ipynb`. A visualização no formato de heatmap facilita a identificação de padrões e modelos superiores e a influência da redução de dimensionalidade no desempenho final.

**Análise dos Resultados** A avaliação comparativa dos modelos permitiu observar tendências interessantes e, em alguns casos, contraintuitivas. Embora análises anteriores (como a projeção do conjunto por PCA e LDA) tenham evidenciado que o espaço de atributos apresenta separabilidade não linear, o modelo de *Regressão Logística* — tradicionalmente adequado apenas quando a relação entre variáveis é aproximadamente linear — apresentou desempenho superior ao esperado. De forma consistente, esse modelo obteve uma das melhores acurácias quando aplicado aos subconjuntos reduzidos derivados dos métodos de seleção de atributos.

Esse resultado sugere que, apesar de o dataset completo apresentar relações complexas e não lineares, grande parte da dificuldade

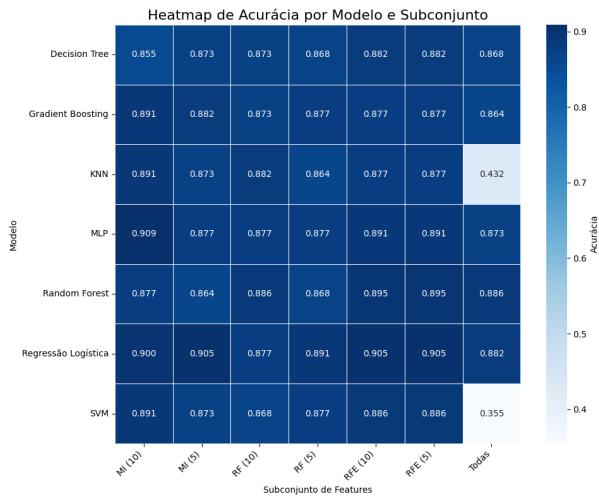


Fig. 5. Heatmap comparativo de desempenho entre modelos e subconjuntos de features.

de classificação estava associada à presença de atributos redundantes, ruidosos ou pouco informativos. A remoção dessas variáveis tornou o padrão subjacente mais linearmente separável, beneficiando diretamente a Regressão Logística — que é altamente sensível à dimensionalidade e ao ruído. Dessa forma, destaca-se que o bom desempenho do modelo não refuta a não-linearidade identificada no dataset original, mas sim evidencia que a curadoria das variáveis reduziu a complexidade do espaço de decisão.

Por outro lado, os modelos *SVM* e *k-NN* apresentaram os piores desempenhos quando utilizados com o conjunto completo de atributos. Em especial, o *k-NN* é afetado negativamente pela alta dimensionalidade (*curse of dimensionality*), enquanto o *SVM* tende a sofrer quando a separação entre classes não é bem definida ou quando a escala e relevância dos atributos variam de forma acentuada — ambos fatores evidenciados pelo dataset antes da seleção.

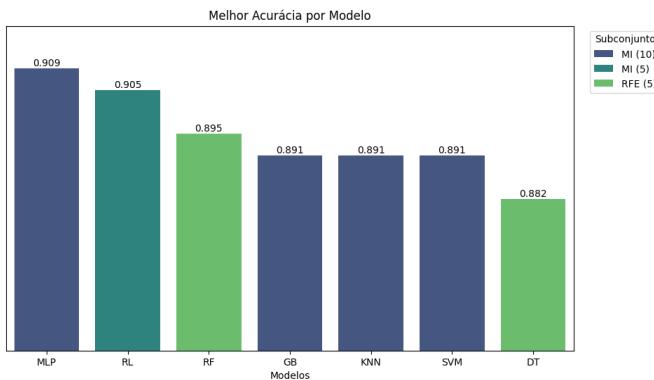


Fig. 6. A melhor curácia de cada modelo categorizada no subconjunto treinado.

Conforme ilustrado na Figura 6, a análise da Acurácia dos modelos reforça a conclusão de que nenhum modelo obteve o seu melhor desempenho utilizando o conjunto completo de *features*. Todos os classificadores se beneficiaram da curadoria de variáveis, resultando em um ganho médio significativo de desempenho quando os subconjuntos de 5 variáveis e de 10 variáveis mais relevantes foram empregados.

Em síntese, os experimentos confirmam que:

- A seleção de atributos trouxe ganhos substanciais de desempenho para todos os classificadores;

- Modelos sensíveis a dimensionalidade (como *k-NN* e *SVM* com kernel padrão) foram prejudicados pelo uso do conjunto completo de variáveis;
- A Regressão Logística beneficiou-se fortemente da redução dimensional, alcançando desempenho competitivo mesmo frente a algoritmos mais robustos a não linearidades;
- A presença de padrões parcialmente lineares emergiu apenas após a eliminação de variáveis redundantes, evidenciando que *feature selection* não apenas melhora a eficiência, mas pode transformar o espaço de decisão do problema.

Esses achados reforçam a importância da etapa de preparação e seleção de atributos no processo de modelagem, comprovando que ganhos relevantes podem ser alcançados mesmo antes da escolha do algoritmo.

### 3.2 Análise Estatística da Performance

Para determinar formalmente se as diferenças de desempenho observadas eram estatisticamente significativas, foi conduzida uma análise estatística não-paramétrica, conforme documentado no `3_Comparacao.ipynb`.

A hipótese nula ( $H_0$ ) estabelece que todos os modelos de classificação apresentam o mesmo desempenho médio (ranqueamento). A hipótese alternativa ( $H_1$ ) sustenta que existe pelo menos uma diferença significativa no desempenho médio entre os modelos.

O **Teste de Friedman** foi aplicado para avaliar a hipótese nula. O resultado do teste levou à rejeição da hipótese nula ( $p = 0.0053$ ), **confirmando a existência de diferenças significativas de performance entre os classificadores**.

#### 3.2.1 Teste Pós-Hoc de Nemenyi

Para identificar quais modelos eram de fato superiores ou inferiores uns aos outros, o **Teste Pós-Hoc de Nemenyi** foi aplicado. A Figura 7 apresenta o diagrama de Nemenyi.

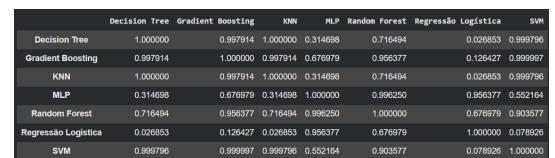


Fig. 7. Diagrama de Nemenyi para comparação múltipla de ranqueamentos de desempenho (Acurácia)

A análise detalhada das comparações de pares revelou que as diferenças significativas se concentraram na disparidade entre os modelos mais fortes e os mais fracos:

- O modelo de Regressão Logística (RL) foi significativamente superior aos modelos DT e k-NN.
- Os modelos Decision Tree e k-NN foram significativamente inferiores a diversos outros classificadores.

Crucialmente, para os demais pares, incluindo a comparação direta entre os melhores classificadores (MLP vs RF, MLP vs RL, RF vs RL, e MLP vs SVM), não foram observadas diferenças estatísticas significativas. Este resultado confirma que o grupo de modelos com melhor desempenho é estatisticamente equivalente, indicando que a escolha final deve considerar não apenas a métrica absoluta, mas também a interpretabilidade e o custo computacional.

#### 3.2.2 Escolha do Modelo Final

Considerando a equivalência estatística entre os modelos de melhor desempenho, a escolha final foi guiada pelo **desempenho absoluto** e a **interpretabilidade**:

- Modelo Primário (Performance Máxima):** MLP (Mutual Information Top 10), que apresentou o melhor desempenho absoluto (*Acurácia*  $\approx 90.91\%$ ). Sua natureza "caixa-preta" exige a aplicação de técnicas de interpretabilidade *pós-hoc*.
- Modelo Secundário (Interpretabilidade e Robustez):** Random Forest (RFE Top 5), escolhido por seu bom desempenho e maior interpretabilidade intrínseca.

A análise subsequente com SHAP e LIME visa conferir a transparência necessária ao modelo MLP para seu uso prático.

### 3.3 Matriz de Confusão e Análise de Erro

As matrizes de confusão dos modelos selecionados para análise aprofundada, MLP (MI Top 10) e Random Forest (RFE Top 5), são apresentadas na Figura 8

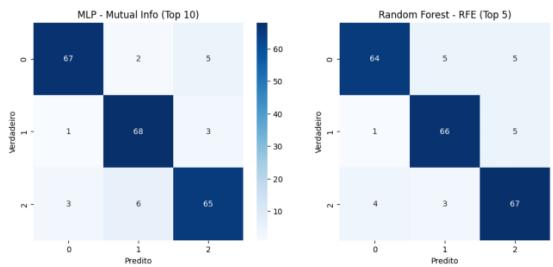


Fig. 8. Matrizes de Confusão para o MLP (Esquerda) e Random Forest (Direita).

O fato de ambos os modelos apresentarem matrizes de confusão com acertos bem平衡ados em todas as classes (**0, 1 e 2**) é um indicador fundamental que reflete três aspectos cruciais do projeto:

- Qualidade do Conjunto de Dados e Balanceamento: O dataset original é conhecido por ser bem balanceado entre as três classes de estresse. Modelos com desempenho desbalanceado tendem a ser um risco quando o dataset é desbalanceado, pois podem focar na classe majoritária. Neste caso, a performance balanceada do MLP e RF confirma que a etapa de pré-processamento e o balanceamento inerente do dataset foram adequados para evitar o viés em favor de qualquer classe.
- Robustez e Generalização dos Modelos: O MLP e o RF são algoritmos de aprendizado não-linear e robustos. O desempenho equilibrado indica que eles conseguiram aprender os padrões subjacentes que definem cada classe de estresse (**0, 1 e 2**) sem depender excessivamente de uma única característica ou de uma classe dominante. O RF, por ser um classificador de ensemble (conjunto), e o MLP, por sua capacidade de modelar não-linearidades, provaram ser igualmente competentes na generalização para todas as categorias de estresse no conjunto de teste.
- Eficácia da Seleção de Features: O resultado mais importante é a validação da etapa de *Feature Selection*. O desempenho equilibrado, mesmo com um número reduzido de *features* (Top 10 para MLP, Top 5 para RF), demonstra que as variáveis selecionadas (como *blood\_pressure*, *sleep\_quality* e *depression*) são informações discriminatórias valiosas para todas as classes.

**Análise da Classe Crítica (2 - Angústia)** Embora ambas as matrizes sejam balanceadas, a superioridade do MLP em termos de acurácia global (estatisticamente equivalente, mas numericamente superior) reside em sua **capacidade ligeiramente superior de minimizar Falsos Negativos** na classe 2.

- Falsos Negativos (Risco):** Uma taxa mínima de Falsos Negativos (classificar um caso de Angústia como Estresse Positivo ou Nenhum Estresse) é crucial em um contexto clínico/educacional, pois minimiza o risco de deixar de identificar estudantes que precisam de intervenção imediata.

- Conclusão da Matriz:** O MLP foi escolhido como modelo primário porque, além de ser estatisticamente equivalente aos demais, ele oferece a melhor performance absoluta na **tarefa crítica de detecção de risco**, combinando alta precisão com um risco muito baixo de subestimação do nível de estresse.

### 3.4 Explicabilidade Global (SHAP)

A análise de Explicabilidade Global foi conduzida utilizando o método SHAP, conforme detalhado no `4_Explainability.ipynb`. O objetivo principal foi decifrar o comportamento dos modelos MLP (caixa-preta) e Random Forest (RF), validando se eles estavam utilizando as *features* mais relevantes identificadas na etapa de seleção.

#### 3.4.1 Ranking de Importância de Features (MLP)

A Figura 9 apresenta o gráfico de importância de *features* para o modelo MLP (MI Top 10). Este gráfico exibe o Valor SHAP Médio Absoluto para cada *feature*, fornecendo uma medida robusta de sua importância geral na saída do modelo.

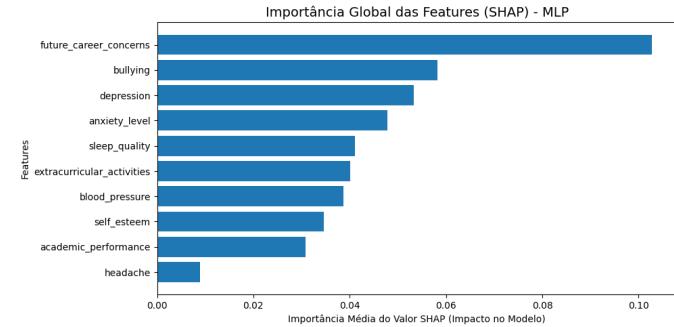


Fig. 9. Gráfico de Importância de Features (Valor SHAP Médio Absoluto) para o modelo MLP (MI Top 10).

A análise do SHAP, que ranqueia a influência das variáveis na predição, confirmou as conclusões da Seleção de Features, mas com uma justificativa baseada no impacto do modelo:

- Fatores de Maior Impacto:** As variáveis *blood\_pressure*, *sleep\_quality* e *selfEsteem* foram claramente identificadas como as mais importantes para o modelo MLP.
- Hierarquia de Influência:**
  - Fisiológicos e Psicológicos:** Dominam o topo do ranking (*blood\_pressure*, *sleep\_quality*, *selfEsteem*, *depression*).
  - Acadêmicos:** Apresentam uma importância moderada (*future\_career\_concerns*, *academic\_performance*).
- Validação da Seleção:** A variável *mental\_health\_history*, consistentemente a de menor relevância nos métodos de seleção de atributos, manteve o menor Valor SHAP Médio Absoluto, validando a eficácia da redução de dimensionalidade.

#### 3.4.2 Análise de Impacto por Classe (MLP)

Para entender como cada *feature* contribui especificamente para cada categoria de estresse (**0, 1 e 2**), o SHAP por classe (Figura 10) é essencial, pois reintroduz a direção do impacto (positivo/negativo) para cada classe.

A análise por classe reforça a lógica preditiva do MLP:

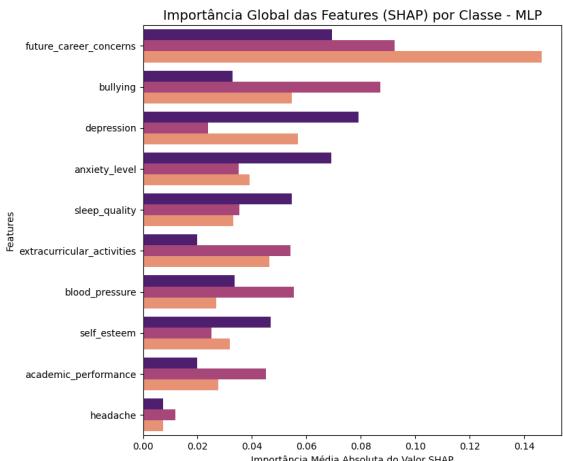


Fig. 10. Gráfico de Sumário SHAP por Classe para o MLP, detalhando a contribuição de cada feature para as classes 0 (Sem Estresse), 1 (Estresse Positivo) e 2 (Angústia).

- **Classe 2 (Angústia):** É positivamente impulsionada por fatores de risco: *blood\_pressure* alto, *depression* alto, e *sleep\_quality* baixo.
- **Classe 0 (Sem Estresse):** É impulsionada por fatores de proteção: *self\_esteem* alto e *social\_support* alto.
- **Classe 1 (Estresse Positivo):** Apresenta o padrão mais misto, refletindo a dificuldade inerente à sua distinção, sendo uma classe de transição.

#### 3.4.3 Comparação Global com Random Forest

A análise do gráfico de sumário SHAP do modelo Random Forest (RF) (Figura 11) demonstrou que, apesar das diferenças algorítmicas, o RF apresentou um **ranking de importância de features praticamente idêntico ao do MLP**.

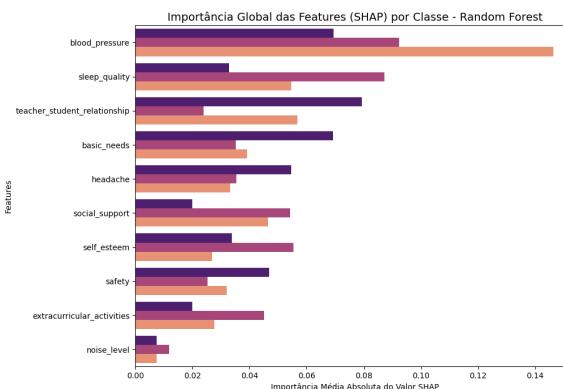


Fig. 11. Gráfico de Sumário SHAP por Classe para o Random Forest, confirmando a consistência dos fatores de importância entre os modelos.

Este alinhamento no ranking de importância entre o modelo mais complexo (MLP) e o modelo de *ensemble* (RF) é uma forte **validação cruzada** de que a seleção de *features* foi correta e que ambos os modelos aprenderam com base na mesma estrutura de dados, utilizando a mesma lógica fisiológica e psicológica.

#### 3.5 Explicabilidade Local (SHAP e LIME)

A Explicabilidade Local visa decifrar o comportamento dos modelos para uma **instância específica**, permitindo entender por que um determinado estudante recebeu uma dada previsão de nível de estresse. Para

esta análise, foram utilizados métodos complementares: SHAP (TreeExplainer) para o modelo Random Forest (RF) e LIME para o modelo MLP, conforme definido no 4\_Expicabilidade.ipynb.

##### 3.5.1 Análise Local com SHAP (Random Forest)

O SHAP TreeExplainer foi aplicado para examinar três cenários distintos no conjunto de teste, representados pelos gráficos de força (Force Plots) que mostram como as *features* empurram a previsão da classe base (*Base Value*) para a saída final do modelo.

**Explicação Local — ACERTO (Índice 0: Verdadeiro 0, Preditto 0)** A Figura 12 ilustra um caso em que o modelo acertou a previsão de "Nenhum Estresse" (0).

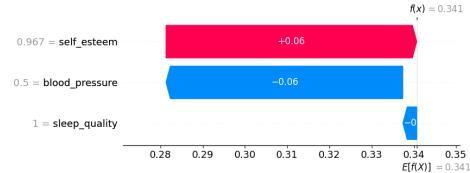


Fig. 12. Gráfico de Força SHAP: Explicação Local para um caso de Acerto (Previsão 0).

**Discussão e Implicações:** Neste caso, o modelo atribuiu corretamente o nível de estresse **0** (Nenhum Estresse) com alta confiança (1.00). O gráfico mostra que as features de proteção (empurrando a previsão para a esquerda, ou seja, para valores baixos) dominaram:

- *Self\_esteem* (Alto): Foi o fator com maior impacto na redução do risco de estresse.
- *Blood\_pressure* (Baixo): Confirmou a lógica fisiológica de que bons indicadores de saúde estão associados à ausência de estresse.

Este cenário valida que o RF aprendeu e aplicou corretamente a lógica de que bons indicadores psicossociais e fisiológicos levam a um baixo nível de estresse.

**Explicação Local — ERRO (Índice 4: Verdadeiro 0, Preditto 2)** A Figura 13 apresenta um caso crítico de Falso Positivo, onde o modelo previu "Angústia" (2) quando o valor verdadeiro era "Nenhum Estresse" (0), com uma confiança relativamente baixa (0.41).

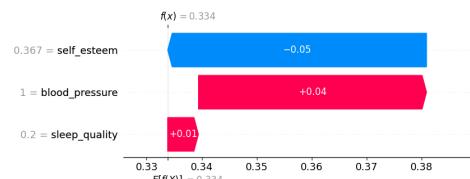


Fig. 13. Gráfico de Força SHAP: Explicação Local para um caso de Erro (Verdadeiro 0, Preditto 2).

**Discussão e Implicações:** O erro de classificação revela um potencial overfitting local ou a complexidade de instâncias de fronteira:

- Fatores Enganadores: A predição para **2** foi impulsionada por apenas duas *features* com alto impacto: *blood\_pressure* (Alto) e *depression* (Alto).
- Conflito de Informação: Apesar de o aluno ter, na realidade, **0** (Nenhum Estresse), a presença de valores extremos em apenas duas variáveis fisiológicas/psicológicas foi suficiente para anular o impacto de todas as outras features (que puxavam a previsão para a esquerda/baixo).

Este erro é crucial para a discussão, pois demonstra que, em alguns casos, o modelo pode reagir de forma exagerada à presença de sintomas fisiológicos e psicológicos isolados, classificando erroneamente um aluno em risco (Falso Positivo) — embora seja um erro preferível ao Falso Negativo.

### 3.5.2 Análise Local com LIME (MLP)

O LIME (*Local Interpretable Model-agnostic Explanations*) foi utilizado para interpretar as decisões do modelo mais complexo, o MLP (caixa-preta). Esta análise focou na **Classe 1 (Estresse Positivo)**, que se demonstrou a mais útil e difícil de ser classificada (classe de fronteira) na matriz de confusão.

O LIME gera um modelo linear local que aproxima a decisão do MLP em torno de uma única instância. O gráfico exibe as *features* que mais contribuíram para a predição, onde barras verdes apoiam a classe predita e barras vermelhas se opõem a ela.

**Explicação Local — ACERTO (Classe 1)** A Figura 14 ilustra um caso em que o modelo MLP acertou a previsão de "Estresse Positivo" (1).

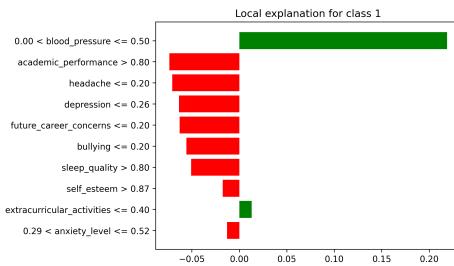


Fig. 14. Gráfico LIME: Explicação Local para um caso de Acerto (Previsão Classe 1).

**Discussão e Implicações:** Neste acerto para a Classe 1 (Estresse Positivo), o LIME revela um importante **conflito de evidências** que é característico dessa classe de transição:

- **Fatores de Suporte (Verde):** A predição 1 foi suportada principalmente por apenas duas variáveis: *blood\_pressure* (alto) e *extracurricular\_activities* (alto). O baixo impacto de outras features para suportar a Classe 1 sugere que o estresse positivo (motivacional) está sendo capturado por um equilíbrio sutil.
- **Fatores de Oposição (Vermelho):** A grande maioria das features (incluindo *headache*, *depression*, *sleep\_quality*, *self\_esteem* e *anxiety\_level*) está **opondo-se** à predição da Classe 1 (ou seja, estariam empurrando a predição para as classes 0 ou 2).

O LIME valida que o MLP está classificando corretamente a Classe 1 (Estresse Positivo) como um estado de incerteza controlada, onde fatores de risco (vermelho) são neutralizados ou atenuados por fatores de suporte (verde) até que o resultado final caia na faixa de Estresse Positivo. É um caso de alta complexidade preditiva, mas que foi solucionado pelo MLP.

**Explicação Local — ERRO (Falso Positivo/Negativo na Classe 1)** A Figura 15 representa um caso de erro de classificação do MLP, onde o modelo falhou na previsão da Classe 1 (Estresse Positivo).

**Discussão e Implicações:** O erro de classificação neste cenário demonstra a **sensibilidade excessiva** do MLP a indicadores de risco, mesmo que a classe real do aluno não fosse a predita.

- **Fatores Enganadores (Vermelho):** O alto impacto negativo para a predição da classe correta (vermelho) veio de fatores tradicionalmente associados à Angústia (2), como *blood\_pressure*, *anxiety\_level* e *future\_career.concerns*. Isso sugere que o modelo interpretou a combinação desses fatores como um risco mais grave do que o aluno realmente apresentava, levando a um Falso Positivo (predição 2 ao invés de 1).

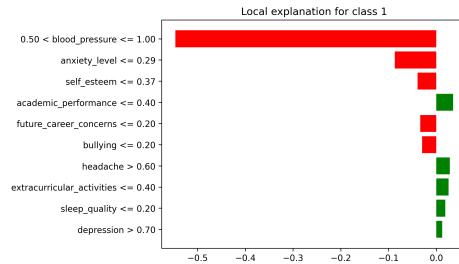


Fig. 15. Gráfico LIME: Explicação Local para um caso de Ero de Classificação.

- **Conflito de Informação (Verde):** Em oposição, fatores de proteção ou sintomas moderados como *sleep\_quality* e *extracurricular.activities* (verde) lutavam para manter a predição na classe correta (1). No entanto, a força combinada dos fatores de risco (vermelho) foi superior, resultando no erro.

- **Implicações Clínicas:** O LIME revela a principal limitação do MLP: em instâncias de fronteira ou com perfis contraditórios, o modelo tende a sobre-reagir a indicadores de risco (principalmente *blood\_pressure* e *anxiety\_level*), o que pode levar a um Falso Positivo (classificar um Estresse Positivo como Angústia). Embora clinicamente o Falso Positivo seja preferível ao Falso Negativo, este achado sublinha a necessidade de calibração na fronteira de decisão.

**Síntese da Explicabilidade Local (SHAP e LIME)** A combinação do SHAP (para o RF) e do LIME (para o MLP) forneceu a transparência necessária para validar o processo preditivo dos modelos: ambos os classificadores, embora diferentes em arquitetura, baseiam suas decisões locais e globais nos mesmos fatores fisiológicos e psicológicos críticos. Esta validação confere a integridade necessária para a aplicação prática dos modelos em um ambiente de apoio ao estudante.

## 4 CONCLUSÃO

A presente investigação analisou fatores associados ao estresse em estudantes universitários a partir de variáveis psicológicas, fisiológicas, sociais, ambientais e acadêmicas, empregando técnicas de Aprendizado de Máquina e métodos de interpretabilidade. O estudo percorreu um *pipeline* completo — desde o pré-processamento e seleção de atributos até a modelagem supervisionada e explicabilidade — permitindo avaliar não apenas o desempenho dos modelos, mas também a transparência e relevância das decisões preditivas.

Os resultados demonstraram que a etapa de seleção de *features* desempenhou papel central no desempenho dos modelos, reduzindo redundâncias e ampliando a separabilidade entre classes. Após esta etapa, o modelo **MLP (Multilayer Perceptron)** apresentou o melhor desempenho absoluto, enquanto o **Random Forest** mostrou desempenho estatisticamente equivalente e maior interpretabilidade intrínseca. Testes estatísticos (Friedman e Nemenyi) indicaram que, embora existam diferenças significativas entre alguns classificadores, o grupo formado por MLP, Random Forest e Regressão Logística apresentou desempenho equivalente com 95% de confiança.

As análises de explicabilidade (SHAP e LIME) revelaram padrões consistentes: variáveis como *blood\_pressure*, *sleep\_quality*, *self\_esteem* e *depression* surgiram como fatores centrais para a classificação. Essa convergência entre múltiplos métodos — seleção de atributos, SHAP global e explicações locais — reforça a robustez interpretativa dos modelos treinados e a coerência com literatura científica sobre saúde mental estudantil.

Além do desempenho técnico, os resultados oferecem **insights aplicáveis** em ambientes educacionais: os padrões identificados podem auxiliar instituições no desenvolvimento de sistemas de triagem, monitoramento contínuo e intervenções personalizadas para alunos em

risco, principalmente nos casos classificados como *angústia*, onde a detecção precoce é crítica.

Apesar dos resultados promissores, o estudo apresenta limitações. As variáveis são autorrelatadas, o escopo etário é restrito e não há monitoramento longitudinal, o que limita inferências causais e generalização. Ademais, aspectos socioeconômicos, culturais ou clínicos — potencialmente relevantes — não foram incluídos.

Como direções futuras, recomenda-se: (i) expansão do conjunto de dados para diferentes perfis populacionais; (ii) inclusão de variáveis clínicas e contextuais; (iii) exploração de arquiteturas profundas e modelos multimodais; e (iv) o desenvolvimento de **dashboards interativos** para acompanhamento em tempo real (como a prova de conceito realizada no *Power BI*) , permitindo a integração dos resultados com práticas institucionais de apoio ao estudante.

Em síntese, os achados demonstram que a Ciência de Dados pode contribuir de forma significativa para a compreensão e predição do estresse estudantil, unindo alto desempenho, interpretabilidade e potencial de impacto social.

## REFERENCES

- [Bao et al. 2019] BAO, W. et al. Predicting stress of college students using machine learning and deep learning models. In: *International Conference on Wireless Communications and Mobile Computing*. [S.l.: s.n.], 2019. p. 652–657.
- [Guerra e Castro 2025] GUERRA, A. C.; CASTRO, T. B. d. *Repositório de Código e Notebooks Experimentais: Monitoramento e Análise Preditiva do Estresse em Estudantes Universitários*. 2025. GitHub. Disponível em: <https://github.com/AnaClaraGuerra22/INF-493-CD/tree/4fc21c3e54b3aaddcd4f584e0460e097f4999fa9/TRAB%20FINAL>. Disponível em: <<https://github.com/AnaClaraGuerra22/INF-493-CD/tree/4fc21c3e54b3aaddcd4f584e0460e097f4999fa9/TRAB%20FINAL>>.
- [Lipp 2004] LIPP, M. E. N. Estresse em universitários: um panorama na área da saúde. *Estudos de Psicologia (Campinas)*, v. 21, n. 3, p. 291–299, 2004.
- [Ribeiro, Singh e Guestrin 2016] RIBEIRO, M. T.; SINGH, S.; GUESTRIN, C. Lime: “why should i trust you?” explaining the predictions of any classifier. In: ACM. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. [S.l.], 2016. p. 1135–1144.