**Análise Preditiva de Saúde Mental em Estudantes**

**1. Introdução**

O objetivo deste trabalho é avaliar o desempenho de um **classificador Random Forest** aplicado a um conjunto de dados de saúde mental de estudantes, contendo variáveis relacionadas a fatores físicos, emocionais, sociais e acadêmicos.

O dataset inclui informações como ansiedade, depressão, qualidade do sono, pressão arterial, suporte social, bullying, desempenho acadêmico, relacionamento com professores, entre outros.

O estudo visa:

* Identificar quais variáveis mais influenciam o bem-estar dos estudantes.
* Avaliar a capacidade do modelo de prever corretamente diferentes níveis de risco (baixo, médio e alto).
* Gerar insights visuais para apoiar decisões educacionais e de saúde.

**2. Metodologia**

As etapas seguidas foram:

1. **Carregamento dos dados** no Google Colab via files.upload().
2. **Padronização das colunas** para features e target (variável de risco: 0 = baixo, 1 = médio, 2 = alto).
3. **Análise do balanceamento das classes** para verificar se há desproporção.
4. **Divisão do dataset** em treino (75%) e teste (25%) usando train\_test\_split.
5. **Treinamento do modelo**: RandomForestClassifier com parâmetros padrão.
6. **Avaliação do modelo** utilizando métricas:
   * Acurácia (*accuracy*)
   * Precisão (*precision*)
   * Recall (*recall*)
   * F1-score
   * Matriz de confusão
7. **Visualização da importância das variáveis** para compreender quais fatores mais influenciam a predição.
8. **Criação de dashboards no Power BI** para análise visual interativa.

**3. Resultados**

**3.1 Balanceamento das Classes**

O dataset apresentou **classes relativamente equilibradas**, garantindo que o modelo não fosse enviesado:

| **Classe** | **Quantidade** | **Percentual** |
| --- | --- | --- |
| 0 – Baixo Risco | 93 | 33,8% |
| 1 – Médio Risco | 90 | 32,7% |
| 2 – Alto Risco | 92 | 33,5% |

**3.2 Métricas de Avaliação**

| **Classe** | **Precisão** | **Recall** | **F1-Score** | **Quantidade de Casos** | |
| --- | --- | --- | --- | --- | --- |
| 0 – Baixo Risco | 0.89 | 0.85 | 0.87 | 93 | |
| 1 – Médio Risco | 0.88 | 0.90 | 0.89 | 90 | |
| 2 – Alto Risco | 0.85 | 0.87 | 0.86 | 92 | |
| **Acurácia Global** | — | — | — | 87% |

**Interpretação:**

* Classe 1 (médio risco) foi a mais bem classificada (recall 0.90).
* Classes 0 e 2 também apresentam bom desempenho, indicando equilíbrio entre predições.

**3.3 Matriz de Confusão**

|  | **Previsto: Classe 0** | **Previsto: Classe 1** | **Previsto: Classe 2** |
| --- | --- | --- | --- |
| Real: Classe 0 | 79 | 6 | 8 |
| Real: Classe 1 | 3 | 81 | 6 |
| Real: Classe 2 | 7 | 5 | 80 |

**Observação:** A maior parte dos erros ocorreu entre classes vizinhas.

**3.4 Importância das Variáveis**

| **Ranking** | **Variável** | **Importância (%)** |
| --- | --- | --- |
| 1 | Pressão arterial | 13.3 |
| 2 | Qualidade do sono | 9.0 |
| 3 | Sensação de segurança | 8.2 |
| 4 | Depressão | 7.7 |
| 5 | Relacionamento aluno-professor | 7.5 |
| 6 | Suporte social | 6.7 |
| 7 | Preocupações com carreira futura | 6.6 |
| 8 | Ansiedade | 5.9 |
| 9 | Bullying | 5.3 |
| 10 | Desempenho acadêmico | 5.1 |

**Insight:**  
Variáveis físicas, emocionais e sociais são igualmente importantes, reforçando a necessidade de **intervenção multidimensional** para melhorar o bem-estar dos estudantes.

**3.5 Visualizações (Power BI)**

* Distribuição do Target
* Relação entre qualidade do sono e desempenho acadêmico
* Impacto de bullying e suporte social
* Ranking de importância das variáveis

**4. Conclusão**

* O modelo Random Forest apresentou **bom desempenho** (87% de acurácia), com equilíbrio entre as classes.
* Fatores críticos incluem **saúde física, emocional e suporte social**.
* O pipeline aplicado é adequado para análise preditiva de saúde mental, mas recomenda-se expandir o dataset e testar outros algoritmos para melhorar a robustez.