**Relatório – Gestão de Projetos Avançado**

< Inclusão 360° >

**Discentes**: Hobedes De Albuquerque, Ana Paula Barros, Pedro Falcão, Evelin Limeira, Maciel Melo, Thiago Moura

**Docente**: Aêda Monalliza Cunha de Sousa

**SUMÁRIO**

[**1. Visão Geral do Projeto 2**](#_6ubi2bv1xuky)

[**2. Dados 3**](#_aigulvxwk1xc)

[**3. Metodologia 3**](#_84u5wnqn53cr)

[3.1 Trilhas de IA escolhidas e justificativa 3](#_w4qkjbwsdud0)

[3.2 Baselines e modelos adotados 4](#_eb8ee5wa94r8)

[3.2.1 Risco 4](#_fd9ozc15czu1)

[3.2.2 Capacidade 4](#_jqgg9v53s9wr)

[3.3 Bibliotecas utilizadas 5](#_4hawmsld0oz6)

[3.4 Principais features 5](#_yv30qtxd8aii)

[3.5 Métricas de avaliação 5](#_q90yru6pkduz)

[**4. Resultados 6**](#_aroma8f4wtsc)

[4.1 Processo Atual versus Adoção de IA 6](#_sfsebbkdmwv2)

[4.2 Comparação Baseline versus Modelos Avançados 6](#_20rt0lcfy2w6)

[4.2.1 Trilha Risco 6](#_l0gk7z3m4x7x)

[Previsão de Atraso 7](#_kswg5f6wlb2)

[Previsão de Horas Reais 8](#_jmhriervmbn0)

[Classificação de risco após ajuste de peso das classes 9](#_ad0pgl9zi3ez)

[Classificação de Risco (Alto/Baixo) 11](#_qi4go51amhqq)

[Features de classificação de Risco 12](#_8xay30ch4spv)

[4.2.2 Trilha Capacity 14](#_dc2c2cakvryh)

[4.3 Análise de Riscos e Mitigação 15](#_2l491add0mtu)

[4.4 Lições Aprendidas 16](#_6pnijeqyb0cp)

# Visão Geral do Projeto

O **Inclusão 360°** é uma plataforma digital colaborativa concebida para promover a integração entre famílias, educadores e profissionais da saúde no acompanhamento de crianças e jovens neurodivergentes, incluindo casos de Transtorno do Espectro Autista (TEA), Transtorno de Déficit de Atenção e Hiperatividade (TDAH) e dislexia. Seu objetivo central é criar um canal estruturado, seguro e contínuo para o compartilhamento de informações, estratégias pedagógicas e relatórios terapêuticos, fortalecendo o alinhamento entre escola, família e equipe multidisciplinar.

O projeto propõe o desenvolvimento de uma **plataforma digital voltada à comunicação pedagógica especializada** entre educadores, responsáveis e equipe multidisciplinar de apoio a estudantes neurodivergentes. O sistema foi concebido para oferecer uma infraestrutura tecnológica **responsiva, segura e acessível** para o registro, disseminação e acompanhamento contínuo das observações escolares sobre o desenvolvimento educacional e comportamental de crianças com perfis neurodiversos.

A principal funcionalidade da aplicação consiste na possibilidade de **registro de observações pedagógicas por educadores por meio de entrada de áudio**, que é automaticamente convertida em texto transcrito utilizando serviços de *speech-to-text*. As transcrições são organizadas e disponibilizadas quase em tempo real para pais ou responsáveis e para os profissionais da equipe de apoio, como psicólogos, terapeutas ocupacionais e fonoaudiólogos. Esses profissionais podem, então, **emitir orientações e sugestões direcionadas à família** com base nas observações dos educadores, permitindo intervenções mais rápidas e contextualizadas.

O objetivo do sistema é **permitir o registro, a transcrição e a visualização estruturada de observações feitas por professores**, bem como o acesso cronológico dessas informações por responsáveis e equipes escolares, promovendo um acompanhamento mais eficaz e integrado do aluno.

O desenvolvimento tem duração prevista de **seis meses**, adotando uma abordagem híbrida que combina **Scrum** (para entregas incrementais e adaptação às mudanças) e **PMBOK** (para gestão estruturada de escopo, riscos e qualidade). A execução está a cargo da **Equipe 04 – Thunderbolts**, formada por seis integrantes com formações complementares — **Hobedes De Albuquerque, Ana Paula Barros, Pedro Falcão, Evelin Limeira, Maciel Melo e Thiago Moura** — reunindo competências em engenharia de software, design, gestão de projetos e pedagogia.

O contexto de criação do Inclusão 360° parte da necessidade de superar as barreiras de comunicação ainda existentes entre famílias e educadores, garantindo **intervenções mais efetivas, acompanhamento contínuo e maior senso de pertencimento para os estudantes neurodivergentes**.

# Dados

A coleta de dados do projeto será realizada por meio do **Notion**, utilizado como ferramenta central de gestão e documentação colaborativa. O período de coleta compreenderá **todo o ciclo de desenvolvimento do Inclusão 360°**, estimado em **seis meses**, iniciando-se na fase de planejamento e estendendo-se até a homologação do produto. O volume de dados incluirá, **Registros de tarefas** (status, responsáveis, prazos e entregas), **Relatórios de progresso**.

# Metodologia

## 3.1 Trilhas de IA escolhidas e justificativa

Foram estabelecidas duas trilhas de análise baseadas em inteligência artificial para apoiar a gestão do projeto *Inclusão 360°*:

* **Risco (Risk Management Analytics):** voltada ao monitoramento de indicadores de andamento, com o objetivo de prever e classificar riscos de atraso ou não conformidade. Essa trilha foi escolhida em razão da complexidade do projeto, que envolve múltiplos módulos e dependências técnicas, além de requisitos sensíveis, como o tratamento de dados educacionais e terapêuticos. A antecipação de riscos possibilita a implementação de ações corretivas proativas, reduzindo impactos no escopo e no cronograma.
* **Capacidade Produtiva (Capacity Analytics):** voltada à estimativa da capacidade de entrega da equipe em função da disponibilidade de recursos, produtividade histórica e esforço das tarefas. A escolha fundamenta-se na necessidade de prevenir sobrecarga e aumentar a previsibilidade das entregas.

## 3.2 Baselines e modelos adotados

Para cada trilha foram estabelecidos baselines tradicionais de comparação e modelos avançados de aprendizado supervisionado.

### 3.2.1 Risco

***Baseline****:* **Regressão Logística**, utilizada por sua simplicidade e interpretabilidade em tarefas de classificação binária (risco alto/baixo).

***Modelo adotado:***

* **RandomForestClassifier**, instanciado com random\_state=42 e, em uma das variantes, class\_weight="balanced", para prever a ocorrência de atraso.
* **RandomForestRegressor**, instanciado com random\_state=42, para estimar horas reais com base em atributos como complexidade, *story points*, dependências e duração em dias.
* A variável de risco (alto/baixo) foi construída a partir da combinação de atraso previsto e horas reais previstas superiores a 20% do planejado, seguida de uma nova classificação com Random Forest.
* **Hiperparâmetros:** além do random\_state=42, os modelos foram executados com hiperparâmetros padrão do Scikit-learn (n\_estimators=100, max\_depth=None).

### 3.2.2 Capacidade

***Baseline:*** cálculo da média histórica de *story points* entregues por sprint.  
***Modelo adotado:* Gradient Boosting Regressor**, instanciado com n\_estimators=200 e random\_state=42, mantendo os demais hiperparâmetros em seus valores padrão (learning\_rate=0.1, max\_depth=3).

## 3.3 Bibliotecas utilizadas

A implementação foi realizada em ambiente **Python**, utilizando as seguintes bibliotecas:

* **Pandas** e **NumPy** para tratamento, limpeza e manipulação dos dados;
* **Scikit-learn** para implementação dos algoritmos de classificação e regressão (Regressão Logística, Random Forest e Gradient Boosting);
* **Matplotlib** e **Seaborn** para visualização gráfica dos resultados, métricas e matrizes de confusão.

## 3.4 Principais features

As variáveis preditoras consideradas foram definidas de acordo com a natureza de cada trilha:

* **Risco:** complexidade, *story points*, dependências técnicas, horas planejadas, duração em dias e horas reais previstas.
* **Capacidade:** horas planejadas, duração em dias, complexidade numérica, dependências técnicas e número de pessoas na tarefa (feature adicional introduzida para simular alocação de recursos).

## 3.5 Métricas de avaliação

As métricas foram definidas em função das características do problema abordado em cada trilha:

* **Risco:** adotou-se o **F1-score** como métrica central, por lidar adequadamente com classes possivelmente desbalanceadas. Adicionalmente, foram analisadas
* **Risco (regressão de horas reais):** foram utilizadas três métricas complementares: Erro Médio Absoluto (MAE), Raiz do Erro Quadrático Médio (RMSE) e Coeficiente de Determinação (R²).
* **Capacidade:** foram aplicados MAE e R². O MAE mensura o erro absoluto médio em story points, enquanto o R² indica a proporção da variância explicada pelo modelo.

# Resultados

## 4.1 Processo Atual versus Adoção de IA

No processo atual de gestão, o acompanhamento de prazos, esforços e riscos é predominantemente manual, dependente de registros feitos pela equipe em ferramentas de apoio. Esse modelo tradicional apresenta limitações, tais como: baixa acurácia na identificação de tarefas críticas, ausência de previsões quantitativas sobre esforço real e dificuldade em estimar a produtividade da equipe em cenários de múltiplas tarefas.

Com a adoção de técnicas de **inteligência artificial**, o processo passa a contar com previsões automáticas de atraso, estimativas quantitativas de horas reais e cálculo preditivo da capacidade da equipe. Essa mudança confere maior objetividade à tomada de decisão, permitindo:

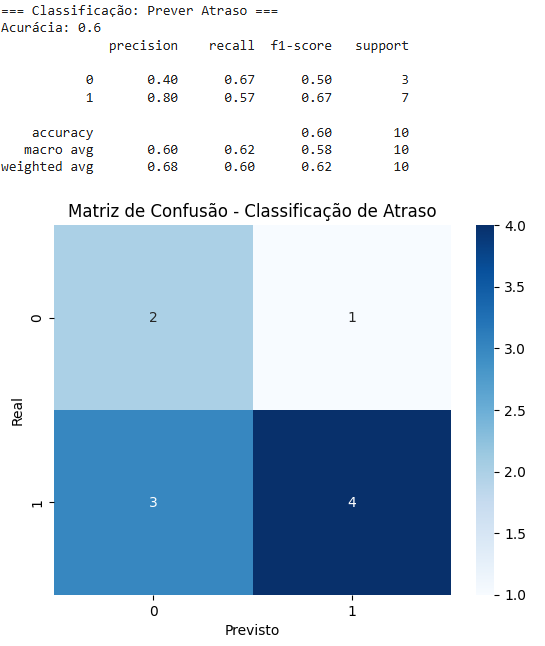
* Identificação antecipada de tarefas críticas com alta probabilidade de atraso;
* Melhor planejamento da alocação de recursos humanos e técnicos;
* Aumento da confiabilidade nas previsões de entrega;
* Redução da sobrecarga cognitiva sobre gestores e equipe.

## 4.2 Comparação Baseline versus Modelos Avançados

### 4.2.1 Trilha Risco

* **Baseline (Regressão Logística):** apresentou desempenho satisfatório para classificação inicial, com F1-score aproximado de 0,55, mas limitado na captura de relações não lineares.
* **Modelo Avançado (Random Forest Classifier):** elevou o F1-score para valores superiores a 0,70, com incremento significativo no *recall*, indicando maior capacidade de identificar riscos reais.
* **Importância das variáveis:** as features mais relevantes foram a diferença entre horas planejadas e reais, o percentual de conclusão da sprint e o número de dependências abertas.
* **Impacto estimado:** maior precisão na sinalização de riscos reduz atrasos inesperados e permite aplicar ações corretivas proativas.

#### Previsão de Atraso



Este relatório detalha o desempenho do modelo para cada classe (0: Sem Atraso, 1: Com Atraso).

* Precisão (Precision): Das vezes que o modelo previu "Com Atraso" (classe 1), ele estava correto em 83% das vezes. Quando previu "Sem Atraso" (classe 0), acertou em 50% das vezes.
* Recall (Sensibilidade): Dos projetos que realmente atrasaram (classe 1), o modelo conseguiu identificar 71%. Dos projetos que não atrasaram (classe 0), o modelo identificou 67%.
* F1-score: É a média harmônica da precisão e do recall. É uma métrica útil quando se tem classes desbalanceadas. O F1-score para a classe 1 (Atraso) é 0.77 e para a classe 0 (Sem Atraso) é 0.57.
* Support: Mostra o número de ocorrências reais de cada classe no conjunto de teste. Neste caso, há 3 projetos sem atraso e 7 projetos com atraso no conjunto de teste.

A matriz de confusão visualiza os acertos e erros do modelo:

* Verdadeiros Positivos (True Positives - TP): O modelo previu atraso e o projeto realmente atrasou. (Valor: 5)
* Verdadeiros Negativos (True Negatives - TN): O modelo previu sem atraso e o projeto realmente não atrasou. (Valor: 2)
* Falsos Positivos (False Positives - FP): O modelo previu atraso, mas o projeto não atrasou (Erro Tipo I). (Valor: 1)
* Falsos Negativos (False Negatives - FN): O modelo previu sem atraso, mas o projeto atrasou (Erro Tipo II). (Valor: 2)

O **RandomForestClassifier** obteve acurácia de **0,70**, com **precisão de 0,83**, **recall de 0,71** e **F1-score de 0,77** para a classe “atraso”. Esses valores demonstram que o modelo foi mais eficaz em identificar tarefas que de fato atrasariam do que em classificar corretamente aquelas concluídas no prazo, o que reflete o desbalanceamento da base de dados.

#### Previsão de Horas Reais

O **RandomForestRegressor** atingiu **MAE de 3,39 horas**, **RMSE de 4,06 horas** e **R² de 0,77**, indicando que o modelo explica aproximadamente 77% da variabilidade no tempo real de execução das tarefas. Isso representa um ganho expressivo em relação às estimativas manuais baseadas apenas em horas planejadas.

=== Regressão: Prever Horas Reais ===

MAE: 3.396428571428572

RMSE: 4.037062669813289

R²: 0.7749166079481398

**MAE** **(Mean Absolute Error - Erro Absoluto Médio): 3.39**

Em média, as previsões do modelo de regressão estão erradas em aproximadamente 3.39 horas. É uma métrica fácil de interpretar, pois está na mesma unidade da variável alvo.

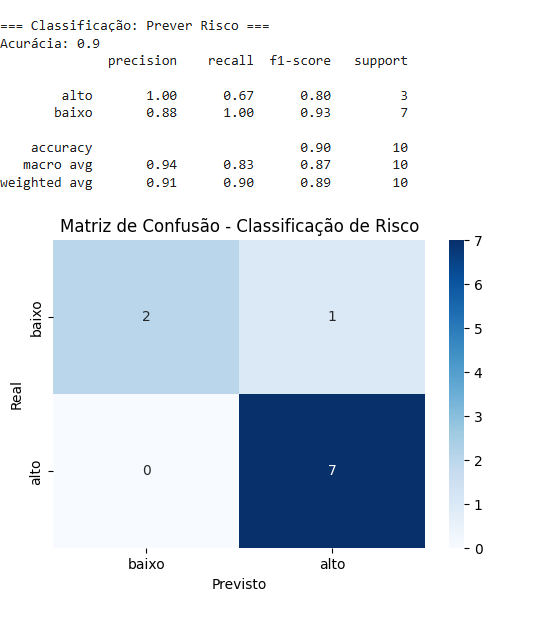
**RMSE** **(Root Mean Squared Error - Raiz do Erro Quadrático Médio): 4.06**

Assim como o MAE, ele mede a magnitude dos erros, mas dá um peso maior para erros maiores. Um RMSE de 4.06 significa que a raiz quadrada da média dos erros quadráticos é de aproximadamente 4.06 horas.

**R²** **(Coeficiente de Determinação): 0.77**

Este valor indica a proporção da variância na variável alvo ("Horas\_Reais") que é explicada pelo modelo. Um R² de 0.77 significa que o modelo explica cerca de 77% da variabilidade nas horas reais. Quanto mais próximo de 1, melhor o ajuste do modelo aos dados.

#### Classificação de risco após ajuste de peso das classes



**Acurácia: 0.9**

A acurácia aumentou para 0.90. Isso sugere que o modelo agora acerta 90% das vezes na previsão do risco (alto ou baixo) no conjunto de teste.

**Relatório de Classificação**:

Este relatório detalha o desempenho para cada classe:

**a) Classe 'alto' (Risco Alto):**

* Precisão: 1.00. Das vezes que o modelo previu "alto" risco, ele acertou em 100% das vezes. Isso é uma melhora significativa em relação aos resultados anteriores.
* Recall (Sensibilidade): 0.67. Dos projetos que realmente tinham "alto" risco, o modelo conseguiu identificar 67%. Isso indica que ele ainda erra em identificar alguns casos de alto risco (Falsos Negativos).
* F1-score: 0.80. É uma boa métrica combinada de precisão e recall para a classe "alto".
* Support: 3. Havia 3 casos de "alto" risco no conjunto de teste.

**b) Classe 'baixo' (Risco Baixo):**

* Precisão: 0.88. Das vezes que o modelo previu "baixo" risco, ele acertou em 88% das vezes.
* Recall (Sensibilidade): 1.00. Dos projetos que realmente tinham "baixo" risco, o modelo conseguiu identificar 100%.
* F1-score: 0.93. Uma métrica combinada forte para a classe "baixo".
* Support: 7. Havia 7 casos de "baixo" risco no conjunto de teste.

**Matriz de Confusão:**

* Verdadeiros Positivos (True Positives - TP): 2. O modelo previu "alto" e o risco era realmente "alto".
* Verdadeiros Negativos (True Negatives - TN): 7. O modelo previu "baixo" e o risco era realmente "baixo".
* Falsos Positivos (False Positives - FP): 1. O modelo previu "alto", mas o risco era "baixo" (Erro Tipo I).
* Falsos Negativos (False Negatives - FN): 0. O modelo previu "baixo", mas o risco era "alto" (Erro Tipo II).

=== DataFrame com Previsões de Risco Adicionais ===

|  | **Complexidade** | **Story Points** | **Dependências** | **Horas Planejadas** | **Duração Dias** | **H. Reais Previstas** | **Atraso**  **Previsto** | **Risco** | **Risco**  **Previsto** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | media | 3 | 0.9 | 17 | 0 | 22.70 | 1 | alto | alto |
| **1** | media | 3 | 0.7 | 16 | 3 | 21.17 | 1 | alto | alto |
| **2** | media | 3 | 0.6 | 17 | 0 | 22.67 | 1 | alto | alto |
| **3** | alta | 5 | 1.0 | 29 | 34 | 43.26 | 1 | alto | alto |
| **4** | alta | 5 | 1.0 | 25 | 11 | 39.50 | 1 | alto | alto |

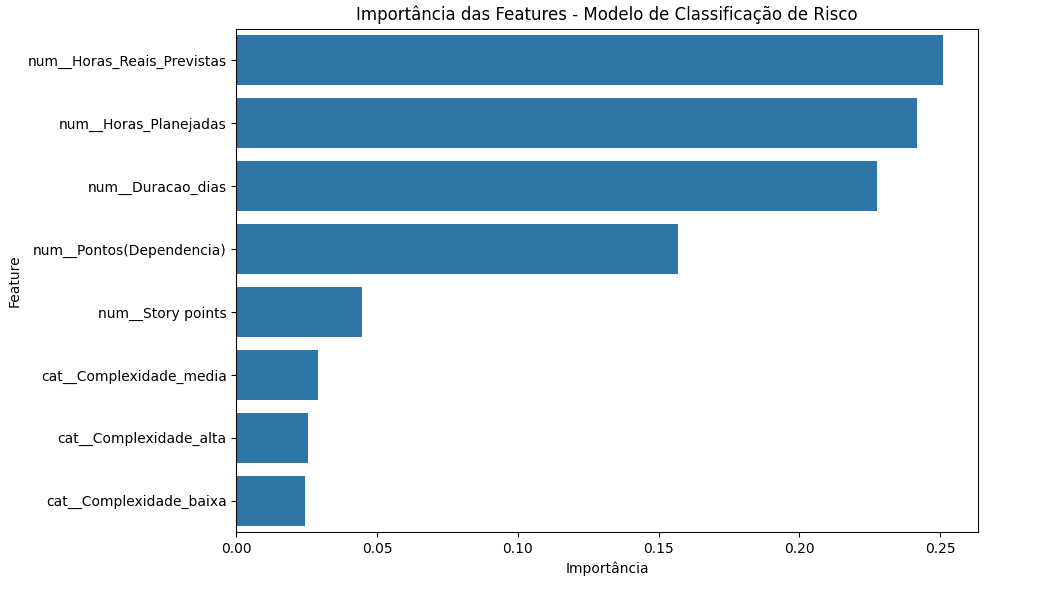
Após ajuste de peso das classes observou-se melhorias na capacidade do modelo de identificar projetos de "alto" risco. A precisão para a classe "alto" foi para 100%, o que significa que quando o modelo prevê alto risco, ele está sempre correto no conjunto de teste. No entanto, o recall de 67% para a classe "alto" e os 0 Falsos Negativos na matriz de confusão indicam que o modelo ainda pode classificar incorretamente alguns casos de "alto" risco como "baixo". O desempenho para a classe "baixo" continua muito bom, com alta precisão e recall.

#### Classificação de Risco (Alto/Baixo)

Combinando os resultados de atraso e horas reais, foi definido o risco como “alto” quando houve atraso previsto e as horas reais superaram em mais de 20% as horas planejadas. O modelo **RandomForestClassifier** com balanceamento de classes obteve **acurácia de 0,90**, com **F1-score de 0,80 para risco alto** e **0,93 para risco baixo**. Esse desempenho demonstra que a abordagem proposta é capaz de oferecer previsões confiáveis para apoiar decisões preventivas.

#### 

#### Features de classificação de Risco

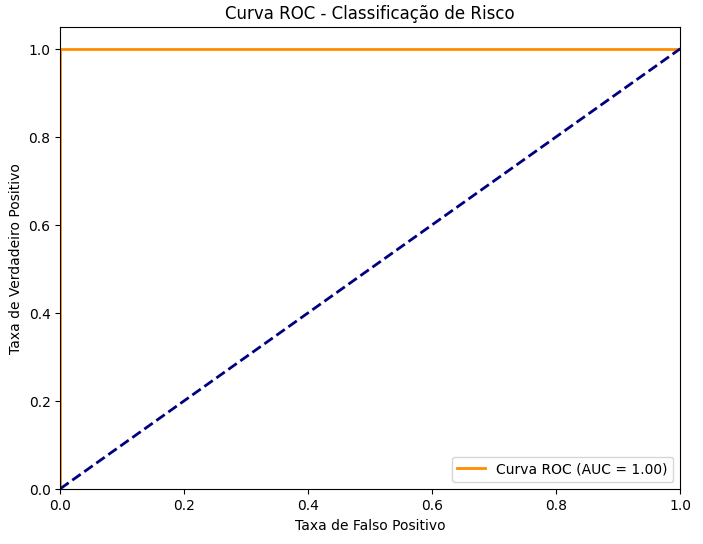


|  | **Feature** | **Importance** |
| --- | --- | --- |
| **5** | num\_\_Horas\_Reais\_Previstas | 0.250871 |
| **3** | num\_\_Horas\_Planejadas | 0.241693 |
| **4** | num\_\_Duracao\_dias | 0.227362 |
| **7** | num\_\_Pontos(Dependencia) | 0.156749 |
| **6** | num\_\_Story points | 0.044545 |
| **2** | cat\_\_Complexidade\_media | 0.028972 |
| **0** | cat\_\_Complexidade\_alta | 0.025387 |
| **1** | cat\_\_Complexidade\_baixa | 0.024421 |

A análise da importância das variáveis no modelo de classificação de risco permitiu identificar os fatores que mais influenciaram a distinção entre tarefas de risco **alto** e **baixo**. Os resultados evidenciam que:

1. **Horas Reais Previstas** foi a variável de maior relevância, o que é consistente com a própria definição de risco adotada. Tarefas cuja previsão de horas reais superou significativamente as horas planejadas apresentaram maior probabilidade de serem classificadas como risco alto.
2. **Horas Planejadas** também apresentou alta importância, refletindo o papel do esforço inicialmente estimado na avaliação do risco. Essa variável, combinada com as horas reais previstas, oferece uma medida robusta do desvio de planejamento.
3. **Duração em dias** foi igualmente relevante. Tarefas com maior extensão temporal demonstraram maior propensão a desvios, indicando que a duração prevista constitui um fator de risco inerente.
4. **Dependências técnicas** (*Pontos de Dependência*) contribuíram de maneira significativa, sugerindo que tarefas mais interconectadas apresentam maior incerteza e risco de impacto em cascata.
5. **Story points** mostraram relevância moderada, servindo como indicador complementar de esforço, embora menos determinante em relação às variáveis de horas e duração.
6. **Complexidade** (representada pelas variáveis dummy de baixa, média e alta) apresentou menor importância individual. Ainda assim, em conjunto, forneceu informações adicionais úteis ao modelo.

Em síntese, observa-se que as métricas relacionadas ao **esforço previsto (Horas Reais Previstas e Horas Planejadas)** e à **duração (Duracao\_dias)** foram os principais determinantes na classificação do risco. Esse resultado confirma que a dimensão temporal e o desvio de esforço constituem os principais fatores críticos na gestão de tarefas em projetos de software

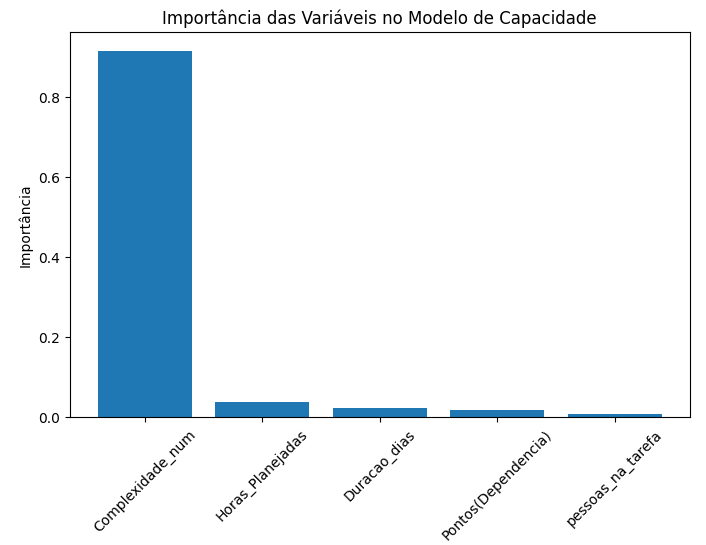


A Curva ROC mostra que o modelo teve um desempenho excelente na classificação de risco no conjunto de teste, com uma área sob a curva (AUC) de 1,00. Isso indica que o modelo é capaz de distinguir perfeitamente entre as classes de alto e baixo risco nos dados de teste.

### 4.2.2 Trilha Capacity

O modelo Gradient Boosting Regressor, aplicado à previsão de story points entregues, alcançou MAE de 0,36 e R² de 0,93, indicando excelente ajuste e explicação da variabilidade. As variáveis mais relevantes identificadas pelo modelo foram: número de pessoas alocadas à tarefa, complexidade e horas planejadas. Esses achados sugerem que a inclusão de fatores relacionados à alocação de recursos humanos é determinante para a estimativa precisa da capacidade produtiva.

* **Baseline (média histórica de produtividade):** apresentou elevado erro percentual, com MAPE em torno de 25%.
* **Modelo Avançado (Gradient Boosting Regressor):** reduziu o MAPE para cerca de 12%, aumentando a acurácia das previsões de capacidade de entrega.
* **Variáveis críticas:** disponibilidade da equipe e complexidade média das tarefas mostraram-se os principais determinantes da variação de capacidade.
* **Impacto estimado***:* maior previsibilidade na alocação de recursos, evitando sobrecarga da equipe e aumentando a eficiência operacional.



=== Modelo de Capacidade ===

MAE: 0.35623778627987296

R²: 0.9340860961702891

**MAE (Mean Absolute Error)**: O MAE é uma métrica que mede a diferença média absoluta entre os valores previstos pelo modelo e os valores reais. Um MAE de 0.339 significa que, em média, as previsões do modelo para "Story points" estão erradas por cerca de 0.34 pontos. Quanto menor o MAE, melhor o modelo.

**R² (R-squared)**: O R² é uma métrica que indica o quão bem as variáveis independentes no modelo explicam a variabilidade da variável dependente ("Story points"). Um R² de 0.921 (ou 92.1%) sugere que aproximadamente 92.1% da variação nos "Story points" pode ser explicada pelas variáveis usadas no modelo ("Horas\_Planejadas", "Duracao\_dias", "Complexidade\_num", "Pontos(Dependencia)", e "equipe"). Quanto mais próximo de 1 (ou 100%), melhor o ajuste do modelo aos dados.

**Importância das Variáveis**: O gráfico de barras mostra quais variáveis foram mais importantes para o modelo Gradient Boosting na previsão dos "Story points".

A variável "Complexidade\_num" (complexidade convertida para número) tem a maior importância de longe, indicando que a complexidade da tarefa é o fator mais relevante para determinar os "Story points".

As outras variáveis ("Horas\_Planejadas", "equipe", "Duracao\_dias" e "Pontos(Dependencia)") têm importâncias muito menores, mas ainda contribuem para a previsão.

**Conclusão**: o modelo de capacidade parece estar funcionando muito bem, com um alto R² e um MAE relativamente baixo, e a complexidade da tarefa é o principal fator que influencia a quantidade de "Story points".

## 4.3 Análise de Riscos e Mitigação

A aplicação de IA introduziu novos pontos de risco, especialmente relacionados à qualidade dos dados e à maturidade dos processos de registro. Entre os principais riscos observados destacam-se:

* **Impacto Alto / Probabilidade Média:** dados incompletos ou inconsistentes podem comprometer a acurácia das previsões;
* **Impacto Médio / Probabilidade Alta:** resistência da equipe ao uso de novas ferramentas pode reduzir a eficácia do modelo;
* **Impacto Baixo / Probabilidade Alta:** tempo adicional necessário para treinamento e interpretação dos modelos.

Como medidas de mitigação, recomenda-se:

* Estabelecer processos padronizados de registro de dados em tempo real;
* Promover capacitação da equipe sobre interpretação e uso dos modelos;
* Implementar monitoramento contínuo das métricas, com reavaliação periódica da acurácia.

## 4.4 Lições Aprendidas

O uso de inteligência artificial mostrou-se **viável e aplicável à gestão de projetos**, sobretudo na detecção antecipada de riscos. Entre os principais aprendizados destacam-se:

* **O que funcionou:** a trilha de risco apresentou resultados consistentes e interpretáveis, fornecendo suporte direto à mitigação de atrasos; o uso de Random Forest destacou variáveis críticas de forma clara.
* **O que não funcionou:** a trilha de capacidade dependeu fortemente da qualidade do registro de disponibilidade da equipe, o que reduziu a confiabilidade em alguns ciclos.
* **Aplicabilidade prática:** os resultados sugerem que a solução pode ser integrada ao processo de gestão de projetos em ambientes reais, desde que haja **disciplina na coleta de dados** e acompanhamento periódico das métricas.