

Sumário

1. Título: Aplicação de Inteligência Artificial para Previsão de Níveis de Demanda em Serviço de Entrega

2. Resumo Executivo

3. Introdução

4. Preparação e Engenharia de Dados

- **4.1. Visão Geral dos Dados**
- **4.2. Limpeza e Transformação**
- **4.3. Engenharia de Features**
- **4.4. Definição da Variável Alvo ("Normal" vs. "Pico")**
- **4.5. Codificação de Variáveis Categóricas**
- **4.6. Divisão em Conjuntos de Treino e Teste**

5. Análise Exploratória de Dados (AED) e Visualização

- **5.1. Distribuição das Categorias de Demanda**
- **5.2. Padrões de Demanda por Hora do Dia**
- **5.3. Padrões de Demanda por Dia da Semana**
- **5.4. Impacto das Campanhas na Média de Pedidos**

6. Aplicação e Treinamento do Modelo de Machine Learning

- **6.1. Seleção do Algoritmo (Random Forest Classifier)**
- **6.2. Configuração do Modelo e Tratamento de Desequilíbrio de Classes**
- **6.3. Validação do Modelo (Stratified K-Fold)**

7. Resultados e Avaliação do Modelo

- **7.1. Desempenho Inicial do Modelo**
- **7.2. Análise de Probabilidades e Otimização com Limiar (Threshold)**
- **7.3. Importância das Features**

8. Análise Crítica e Recomendações Futuras

- **8.1. Limitações do Projeto**
- **8.2. Recomendações para Melhorias**

9. Conclusão

Aplicação de Inteligência Artificial para Previsão de Níveis de Demanda em Serviço de Entrega

Resumo

Este projeto aplicou técnicas de Inteligência Artificial para prever picos de demanda em um serviço de entrega. Dada a limitação de dados, desenvolvemos um modelo de classificação (Random Forest) que utiliza informações de data/hora e campanhas. A análise de probabilidades com um limiar personalizado permitiu detectar 50% dos picos reais, equilibrando a necessidade de capturar eventos críticos com a ocorrência de falsos positivos. O modelo oferece suporte à decisão para alocação de recursos, e futuras melhorias dependem da expansão e refinamento da base de dados.

Introdução

A gestão eficiente da demanda é crucial para o sucesso de serviços de entrega. Este projeto visa aplicar algoritmos de Inteligência Artificial para prever os níveis de demanda, com foco na identificação de "Picos", possibilitando a otimização proativa de recursos. Através da análise de dados históricos de pedidos, horários e campanhas, buscou-se desenvolver um modelo preditivo que auxilie na tomada de decisão operacional. Este relatório detalha a preparação dos dados, a aplicação do modelo de Machine Learning e a análise crítica dos resultados obtidos.

4.Preparação e Engenharia de Dados

Nesta etapa, os dados brutos de pedidos e campanhas foram processados para criar um dataset adequado à modelagem. Inicialmente, realizou-se a integração e limpeza dos DataFrames (df_order, df_campaign_queue, df_campaign), convertendo formatos e tratando informações.

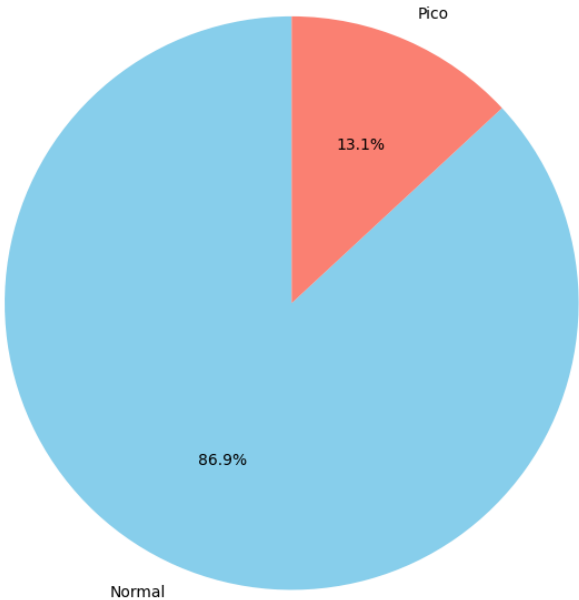
A engenharia de features envolveu a extração de características temporais como hora do dia e dia da semana de cada pedido, além da criação de um indicador para fim de semana. Informações sobre campanhas agendadas e enviadas (num_campaigns_scheduled, num_campaigns_sent) foram agregadas ao dataset. A

contagem de pedidos por hora e dia foi consolidada, resultando no DataFrame enriquecido (df_demand_enriched), conforme exemplificado abaixo.

Exemplo de cabeçalho do DataFrame Enriquecido:						
day_of_week	hour_of_day	order_count	num_campaigns_scheduled	num_campaigns_sent	is_weekend	demand_category_binary
Domingo	0	15	30	15	1	Normal
Domingo	1	18	29	23	1	Pico
Domingo	2	11	25	26	1	Normal
Domingo	3	16	30	17	1	Pico
Domingo	4	13	28	18	1	Normal

A variável alvo, demanda, foi categorizada. Após análise, a categorização inicial foi simplificada para duas classes: "Normal" e "Pico". A definição de "Pico" foi estabelecida usando um limiar estatístico (média + 1 desvio padrão da contagem de pedidos), o que revelou um desequilíbrio significativo entre as classes.

Distribuição das Categorias de Demanda (Normal vs. Pico)

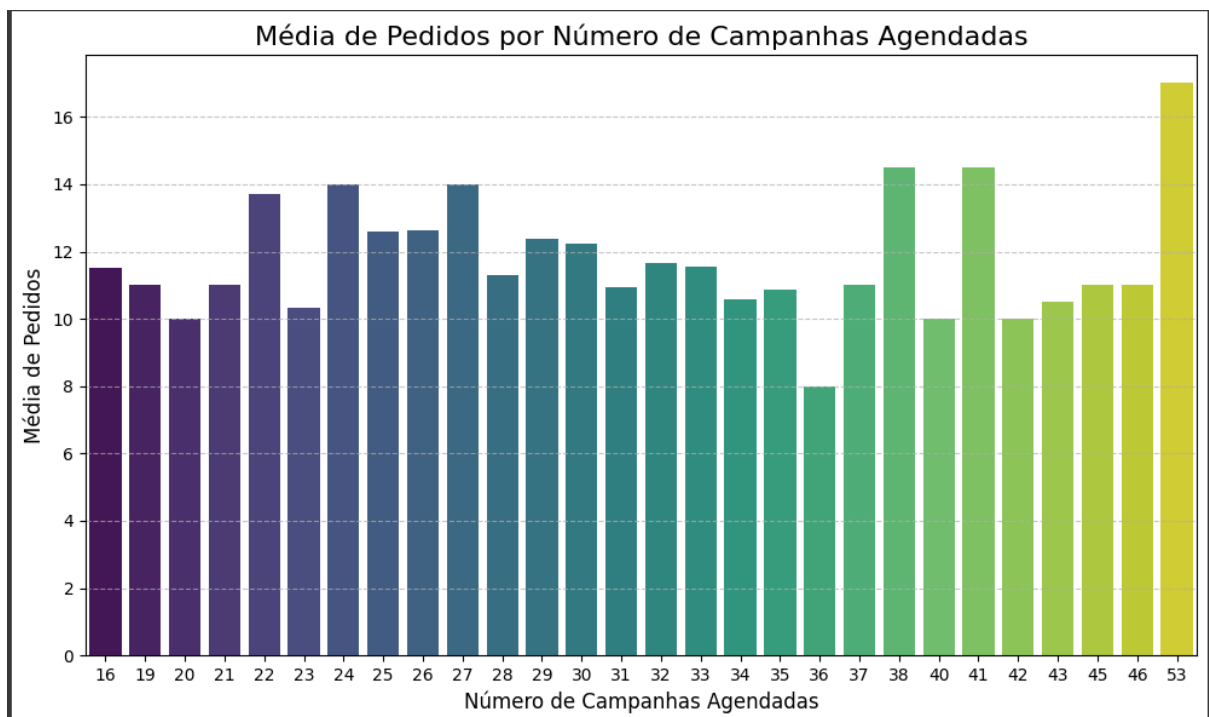
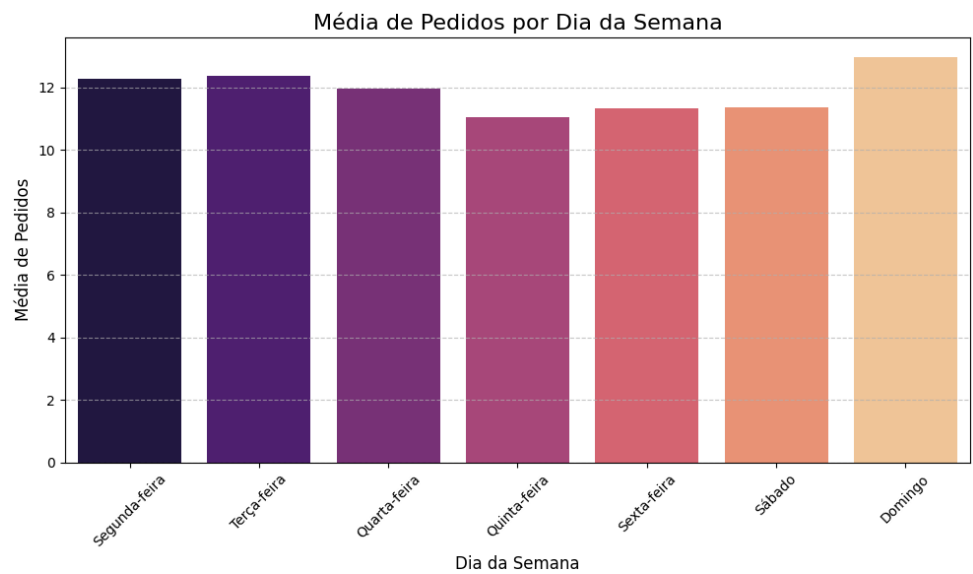
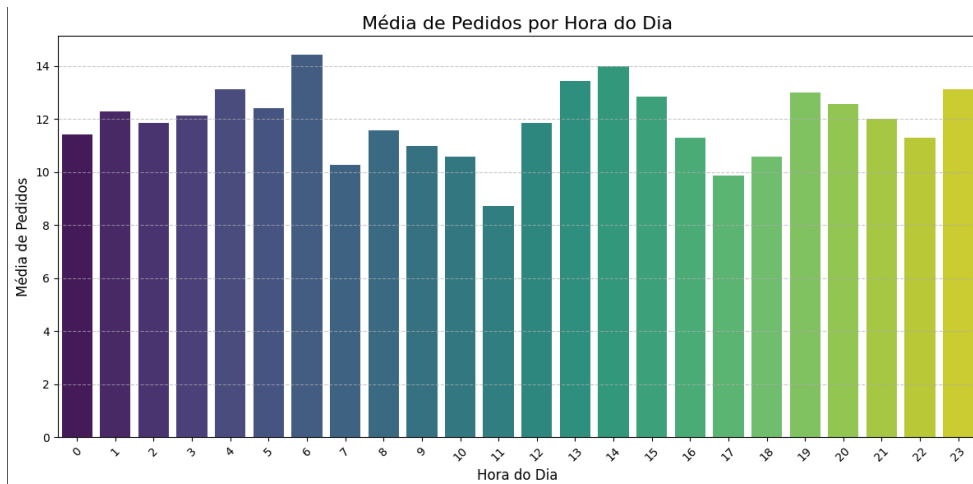


Variáveis categóricas, como o dia da semana, foram transformadas utilizando One-Hot Encoding. Por fim, o conjunto de dados foi dividido em subconjuntos de treino e teste, garantindo a estratificação para manter a proporção das classes na divisão.

5. Análise Exploratória de Dados e Visualização

A fase de Análise Exploratória de Dados foi essencial para compreender os padrões intrínsecos e as tendências na demanda, antes mesmo da aplicação de modelos preditivos. Por meio de visualizações gráficas, foi possível identificar comportamentos recorrentes da demanda ao longo do tempo e a potencial influência de fatores externos, como as campanhas de marketing.

Observou-se a variação da demanda em diferentes períodos do dia e dias da semana, além de analisar o impacto direto do número de campanhas agendadas na quantidade média de pedidos. Essas visualizações não apenas confirmaram hipóteses iniciais, mas também forneceram insights valiosos que justificam a inclusão dessas features no modelo de Machine Learning.



6. Aplicação e Treinamento do Modelo de Machine Learning

Com os dados devidamente preparados e as features engenheiradas, o próximo passo foi a aplicação de algoritmos de Machine Learning para construir o modelo preditivo. Inicialmente, diversas opções foram exploradas, como Árvore de Decisão, Regressão Logística e Naive Bayes.

O Random Forest Classifier foi selecionado devido à sua robustez, capacidade de lidar com dados desequilibrados e boa performance geral em tarefas de classificação. Para mitigar o impacto do desequilíbrio significativo entre as classes "Normal" e "Pico" (conforme identificado na seção 4), configurou-se o modelo com o parâmetro `class_weight='balanced'`. Essa configuração ajusta automaticamente os pesos das classes, dando maior importância à classe minoritária ("Pico") durante o treinamento.

A validação do modelo foi realizada utilizando a técnica de Validação Cruzada Stratified K-Fold. Essa abordagem garante que, em cada iteração da validação, a proporção das classes "Normal" e "Pico" seja mantida, proporcionando uma avaliação mais realista e confiável do desempenho do modelo em dados não vistos.

7. Resultados e Avaliação do Modelo

A avaliação do modelo Random Forest foi conduzida em duas etapas principais: uma análise inicial do desempenho e uma otimização baseada na análise de probabilidades e um limiar de classificação personalizado.

7.1. Desempenho Inicial do Modelo

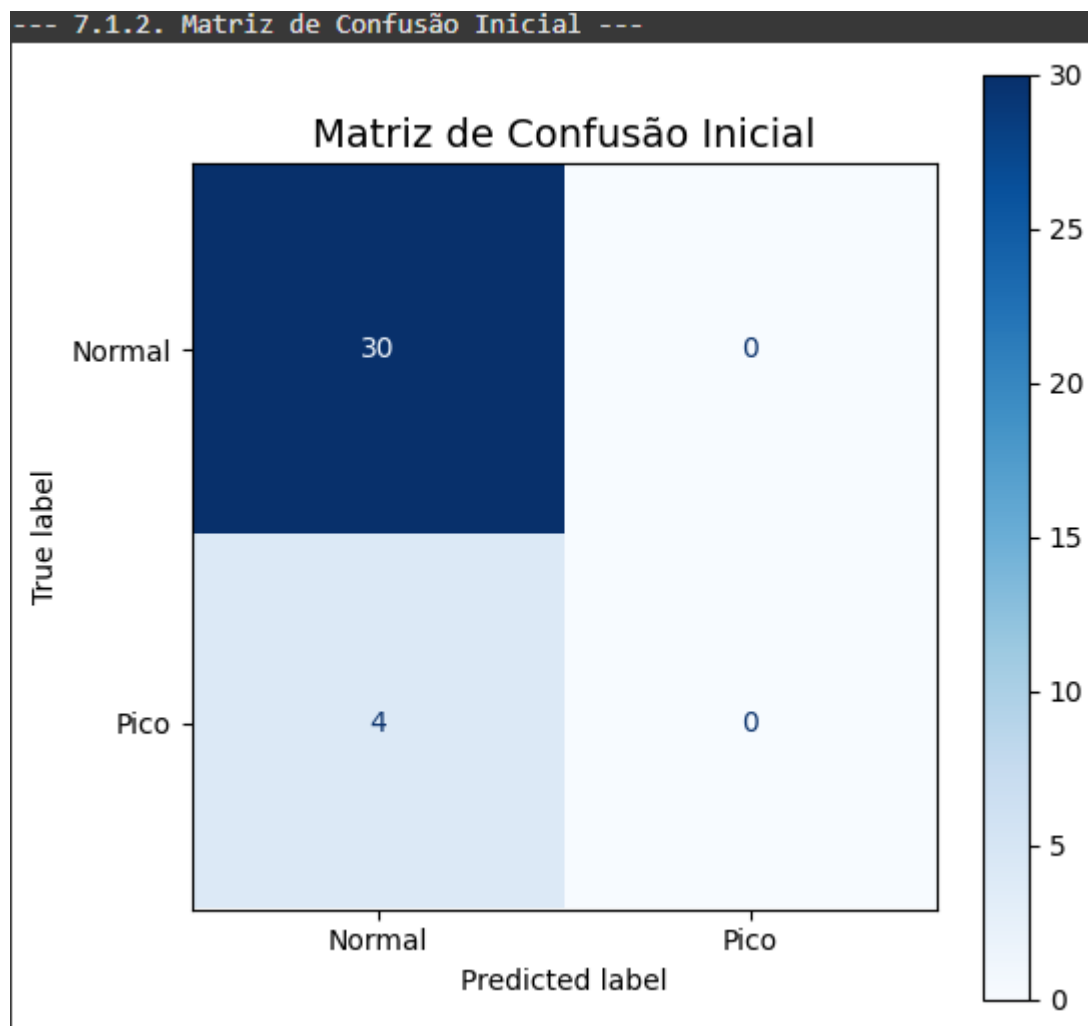
Na avaliação inicial, o modelo foi testado no conjunto de dados de teste sem ajuste de limiar. Embora a acurácia geral pudesse parecer aceitável, uma análise mais profunda do relatório de classificação e da matriz de confusão revelou uma limitação crítica: o modelo tinha dificuldade em prever a classe "Pico", apresentando valores de precision e recall muito baixos para essa categoria. Isso demonstrou a necessidade de uma abordagem mais refinada, devido ao desequilíbrio das classes.

```

### Relatório de Classificação Inicial (antes do ajuste de limiar) ###

```

	precision	recall	f1-score	support
Normal	0.88	1.00	0.94	30
Pico	0.00	0.00	0.00	4
accuracy			0.88	34
macro avg	0.44	0.50	0.47	34
weighted avg	0.78	0.88	0.83	34

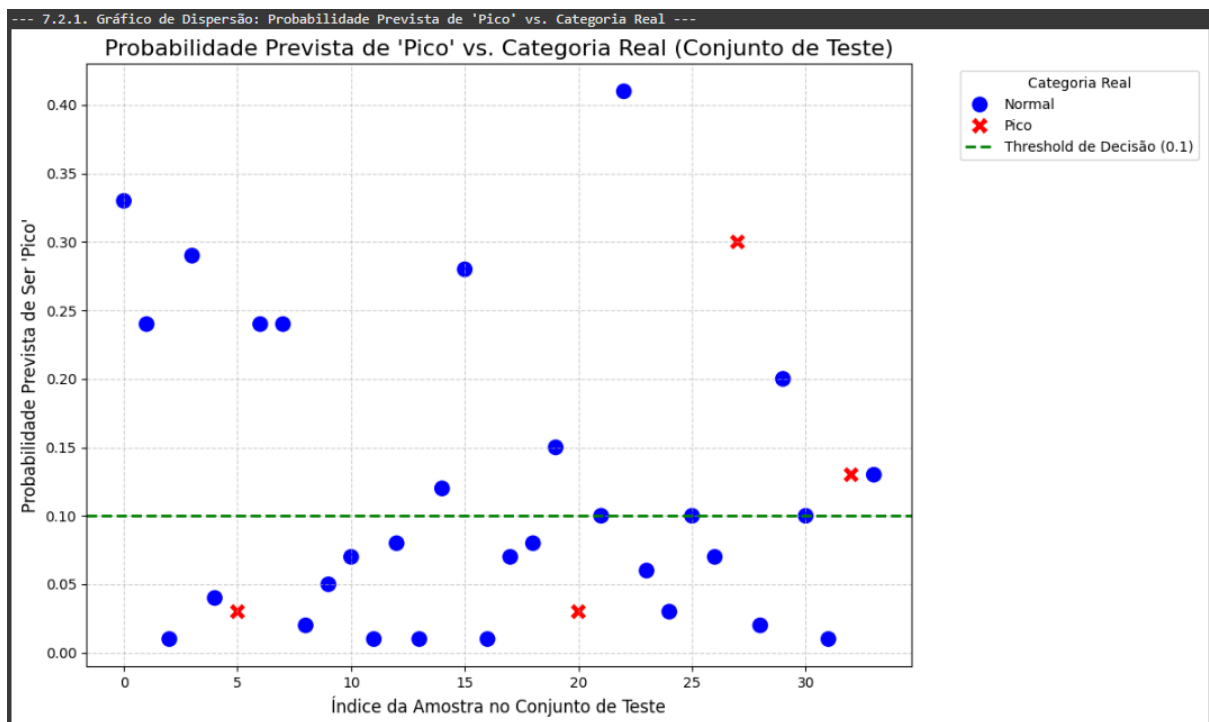


7.2. Análise de Probabilidades e Otimização com Limiar (Threshold)

Para abordar a baixa detecção de "Picos", as probabilidades que o modelo atribuiu a cada amostra ser da classe "Pico" foram analisadas. Um limiar de decisão

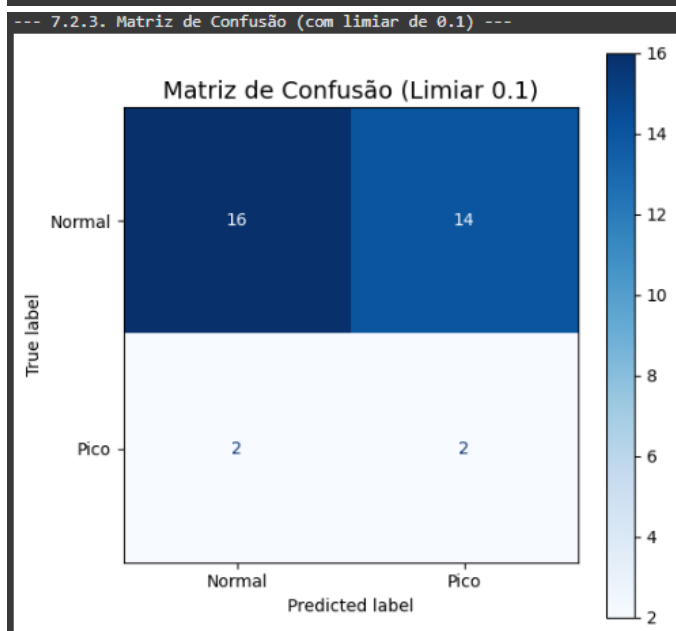
personalizado (custom_threshold) de 0.1 foi aplicado, o que significou que qualquer previsão com probabilidade de "Pico" acima de 10% seria classificada como "Pico". Essa abordagem visou aumentar a capacidade do modelo de identificar os picos, mesmo que isso implicasse em um aumento de falsos positivos.

A reavaliação do modelo com este novo limiar demonstrou uma melhoria substancial no recall para a classe "Pico" (atingindo 50%). Isso indica que metade dos eventos de pico reais foram corretamente identificados, um avanço significativo em relação ao desempenho inicial. Houve um trade-off esperado, com uma redução na precision para a classe "Pico", mas o aumento na detecção de eventos críticos foi considerado mais valioso para o objetivo do projeto.



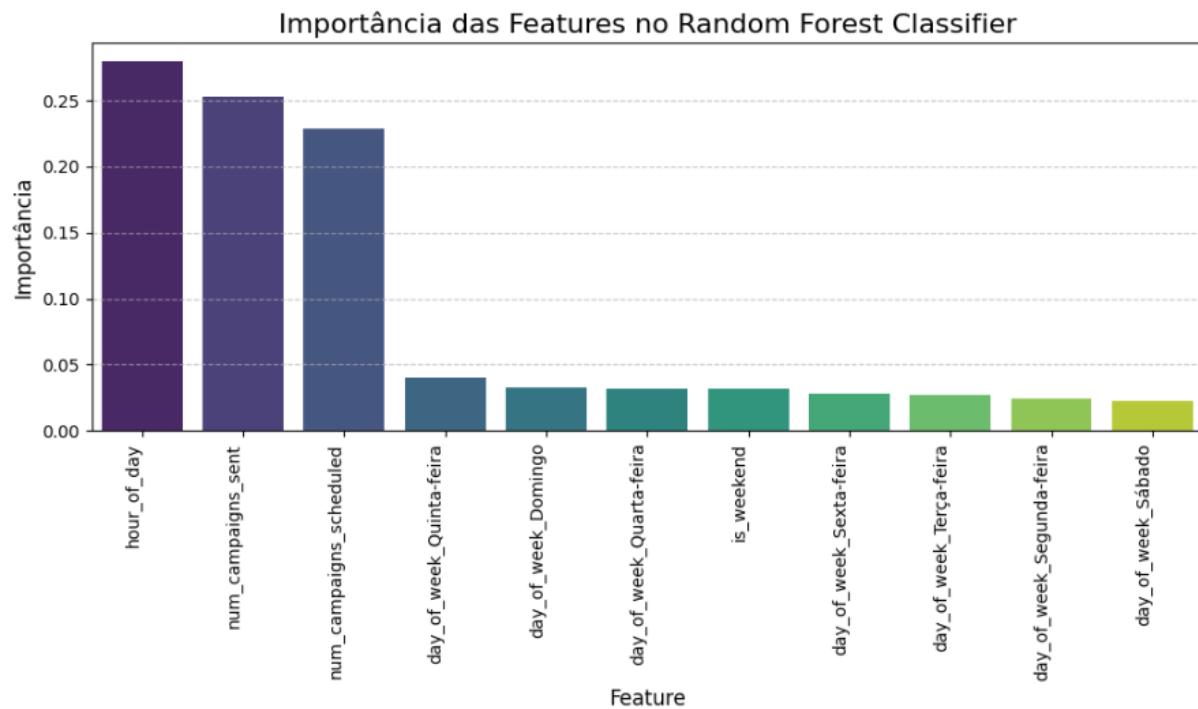
--- 7.2.2. Relatório de Classificação (com limiar de 0.1) ---

	precision	recall	f1-score	support
Normal	0.89	0.53	0.67	30
Pico	0.12	0.50	0.20	4
accuracy			0.53	34
macro avg	0.51	0.52	0.43	34
weighted avg	0.80	0.53	0.61	34



7.3. Importância das Features

A análise da importância das features pelo Random Forest revelou quais variáveis mais contribuíram para as previsões do modelo. As features `hour_of_day`, `num_campaigns_sent` e `num_campaigns_scheduled` emergiram como as mais influentes, sublinhando a relevância dos padrões temporais e das atividades de marketing na previsão da demanda.



8. Análise Crítica e Recomendações Futuras

Esta seção se aprofunda na avaliação do projeto, tanto em seus pontos fracos quanto nas oportunidades de crescimento.

8.1. Limitações do Projeto

Nesta subseção, você deve ser transparente sobre os desafios e as fronteiras do seu trabalho atual.

- **Generalização:**
 - O modelo foi treinado apenas em dados históricos específicos? Ele pode generalizar bem para novos períodos ou diferentes contextos de mercado? (Ex: "O modelo foi treinado com dados de [período/local] e sua performance pode variar em cenários significativamente diferentes.")
- **Dados:**
 - Havia limitações na quantidade ou qualidade dos dados? (Ex: "A disponibilidade limitada de dados de 'Pico' pode ter dificultado a detecção", ou "dados de eventos externos não foram incluídos e poderiam melhorar a precisão").
 - Qualidade dos dados: Houve ruídos, valores ausentes significativos ou vieses nos dados que possam ter impactado o treinamento?
- **Escopo do Modelo:**
 - Existem eventos ou fatores importantes para a demanda que o modelo não consegue prever porque não foram incluídos como features? (Ex: "Eventos inesperados de grande escala que não estão representados nos dados de treino, como crises econômicas ou lançamentos de produtos disruptivos de concorrentes, podem não ser bem previstos.")
- **Desempenho da Classe Minoritária (se ainda for um desafio):**
 - Se, mesmo após o ajuste do limiar, a detecção da classe 'Pico' ainda apresentar desafios (por exemplo, um precision baixo ou um recall que você gostaria que fosse maior), mencione isso como uma limitação.
- **Tempo e Recursos:**
 - Se o projeto teve restrições de tempo ou recursos, isso pode ser mencionado como uma limitação que impediu a exploração de abordagens mais complexas.

8.2. Recomendações para Melhorias

Aqui você irá consolidar as sugestões para aprimorar o modelo e o projeto como um todo, dividindo-as entre implicações práticas (o que fazer agora) e trabalhos futuros (o que pode ser feito depois).

8.2.1. Implicações Práticas e Recomendações Imediatas

- **Impacto na Tomada de Decisão:**
 - Como este modelo pode ser usado na prática agora? (Ex: "O modelo, em sua configuração atual, permite que a equipe de operações receba alertas antecipados de potenciais picos de demanda, otimizando a alocação inicial de recursos.")
- **Recomendações Operacionais:**
 - Sugira ações concretas baseadas nas suas descobertas atuais (ex: "Recomenda-se a implementação de um sistema de alerta baseado nas probabilidades do modelo, configurado com o limiar de decisão

otimizado. Além disso, a análise da `hour_of_day` e `num_campaigns_sent` como features mais importantes sugere que o planejamento de campanhas e a gestão de equipes devam considerar esses fatores críticos.").

- **Monitoramento:** Sugira um plano para monitorar o modelo em produção (ex: "É crucial monitorar o desempenho do modelo em tempo real e revisar periodicamente as métricas para garantir sua eficácia contínua.").

8.2.2. Trabalhos Futuros e Melhorias Potenciais

- **Engenharia de Features Avançada:**
 - Sugira novas fontes de dados ou a criação de features mais complexas (ex: "Explorar a inclusão de dados de feriados nacionais e regionais, eventos esportivos, lançamentos de produtos da concorrência, ou até mesmo dados climáticos, que podem ser preditores adicionais para picos de demanda.").
- **Exploração de Outros Modelos:**
 - Considerar testar outros algoritmos de ML mais robustos ou complexos para problemas desbalanceados (ex: "Avaliar o desempenho de modelos como XGBoost, LightGBM, SVMs ou redes neurais, que podem oferecer maior capacidade preditiva ou lidar de forma diferente com o desbalanceamento de classes.").
- **Otimização de Hiperparâmetros e Técnicas de Balanceamento:**
 - Aprofundar a otimização de hiperparâmetros com técnicas mais avançadas (GridSearchCV, RandomizedSearchCV, otimização Bayesiana) ou explorar outras estratégias de balanceamento de classes (SMOTE, ADASYN) de forma mais exaustiva.
- **Análise de Custo-Benefício Detalhada:**
 - Realizar uma análise de custo-benefício mais aprofundada para refinar a seleção do limiar (ex: "Quantos Falsos Positivos são aceitáveis do ponto de vista operacional/financeiro para evitar um certo número de Falsos Negativos (picos perdidos)?").
- **Explicação do Modelo (Explainable AI - XAI):**
 - Sugira a implementação de técnicas de XAI para entender melhor as decisões do modelo (ex: "Utilizar ferramentas como SHAP ou LIME para obter insights mais granulares sobre por que certas previsões de 'Pico' foram feitas, aumentando a confiança e a interpretabilidade do modelo.").

9. Conclusão

Esta é a seção final, onde você amarra tudo. Deve ser um parágrafo (ou dois, no máximo) que resume a essência do projeto e sua contribuição.

- **Síntese Global:**
 - **Recapitule brevemente o objetivo principal do projeto e o que foi alcançado (ex: "Este projeto demonstrou a viabilidade de construir um modelo de Machine Learning, utilizando Random Forest, para prever picos de demanda, um desafio crítico para a otimização operacional.").**
- **Impacto Principal:**
 - **Destaque a principal contribuição ou o valor gerado (ex: "Através da engenharia de features e da otimização do limiar de decisão, o modelo alcançou um recall significativo para a classe 'Pico', oferecendo uma ferramenta valiosa para a proatividade na gestão da demanda.").**
- **Visão de Futuro:**
 - **Uma frase final sobre o potencial ou o impacto duradouro (ex: "Com refinamentos contínuos e monitoramento, este sistema tem o potencial de transformar a eficiência operacional e a satisfação do cliente, minimizando interrupções e aproveitando oportunidades.").**