

**MBA
USP
ESALQ**

Gestão de demanda aérea: um estudo com árvores de decisão e modelos de “ensemble”

Rafael Carneiro da Silveira
Ana Julia Righetto

Objetivo do Trabalho

- O objetivo do trabalho é avaliar como modelos “ensemble” podem auxiliar na gestão de demanda em companhias aéreas ao automatizar o processo de escolha das classes de tarifa.

Introdução – Revenue Management

Capacidade Fixa

- Limitação de assentos disponíveis em cada voo, exigindo otimização para maximizar a receita.

Mercado Competitivo

- Características de oligopólio, com poucas empresas dominando o setor aéreo, o que exige estratégias para manter competitividade.

Precificação Dinâmica

- Ajustes de preços em tempo real para responder rapidamente às variações de demanda.

Introdução – Cenário

Decisão Condicional “SE... ENTÃO...”

- Analistas definem classes de tarifa com base em regras condicionais, adaptando-se às variáveis disponíveis.

Exemplos de Critérios Condicionais:

- **SE** a ocupação do voo é baixa, **ENTÃO** aplicar classe menor para aumentar a resposta da demanda.
- **SE** a concorrência ajusta tarifas, **ENTÃO** igualar ou reduzir levemente a classe para manter competitividade.

Introdução – Variáveis do Estudo

Gestão de demanda

Tabela 7. Grupos de "mix"

Classe	Classificação
0	"Mix High"
1	
2	
3	
4	"Mix Mid"
5	
6	
7	
8	"Mix Low"
9	
10	
11	
12	

Fonte: Base de dados originais

Gestão de "pricing"

	ADVP			
	0	7	14	21
0	1500			
1	1200			
2	1000			
3		900		
4			800	
5			700	
6			650	
7			500	
8				450
9				375
10				300
11				290
12				250

Material e Métodos – Variáveis do Estudo

Variável	Descrição
CLASSE	Variável dependente - Classe a ser aplicada para tarifação
DTD	Quantidade de dias até a decolagem
ADVP	Agrupamento da antecedência do voo
CAP	Capacidade da aeronave
Season	Flag de temporada: alta ou baixa
DOW	Dia da semana
CLUSTER	Cluster de tarifação inicial
BKD	Booked Load Factor - Ocupação de assentos em %
EXP	Expectativa - Cálculo de projeção de demanda em %
VAR_EXP	Variação de Expectativa versus dia anterior
Revenue	Receita
RASK	Indicador de receita do negócio
TM	Tarifa média
FX	Projeção de expectativa

Variável	Descrição
DM1	Ocupação de assentos d-1 em pontos percentuais
DM2	Ocupação de assentos d-2 em pontos percentuais
DM3	Ocupação de assentos d-3 em pontos percentuais
LW	Ocupação de assentos versus semana anterior em pontos percentuais
TLW	Variação das transações de passagem versus semana anterior em p.p.
CNX_percent	Abertura para venda de conexões em %
CL	Classe mínima para abertura de conexão
CM1	Classe aplicada em d-1
CM2	Classe aplicada em d-2
CM3	Classe aplicada em d-3
CMT_0	Último indicador de competitividade – variável binária
Match_min	Classe em que se precisa chegar para ficar igual ao competidor em tarifa
CMT	Indicador de competitividade atual da classe – variável binária
Fluxo	Flag de fluxo: fluxo, contrafluxo ou indiferente

Material e Métodos – Base de dados

Base 1

Representa um domínio específico de mercado com 11.368 registros:

- Ambiente mais controlado;
- Avalia a capacidade de especialização e precisão em cenário homogêneo;
- Referência de análise como padrão estável.

Base 2

Composta por um mix de dois domínios (50% de cada), totalizando 6.218 registros:

- Ambiente mais diversificado;
- Avalia a capacidade de generalização em múltiplos domínios de mercado;
- Teste de robustez do modelo em cenários variados.

Material e Métodos – Escolha dos Algoritmos

Árvore de Decisão

(Quinlan, 1986)

- Modelo mais simples;
- Ponto de referência inicial no trabalho e modelo base de comparação.

Random Forest

(Breiman, 1996)

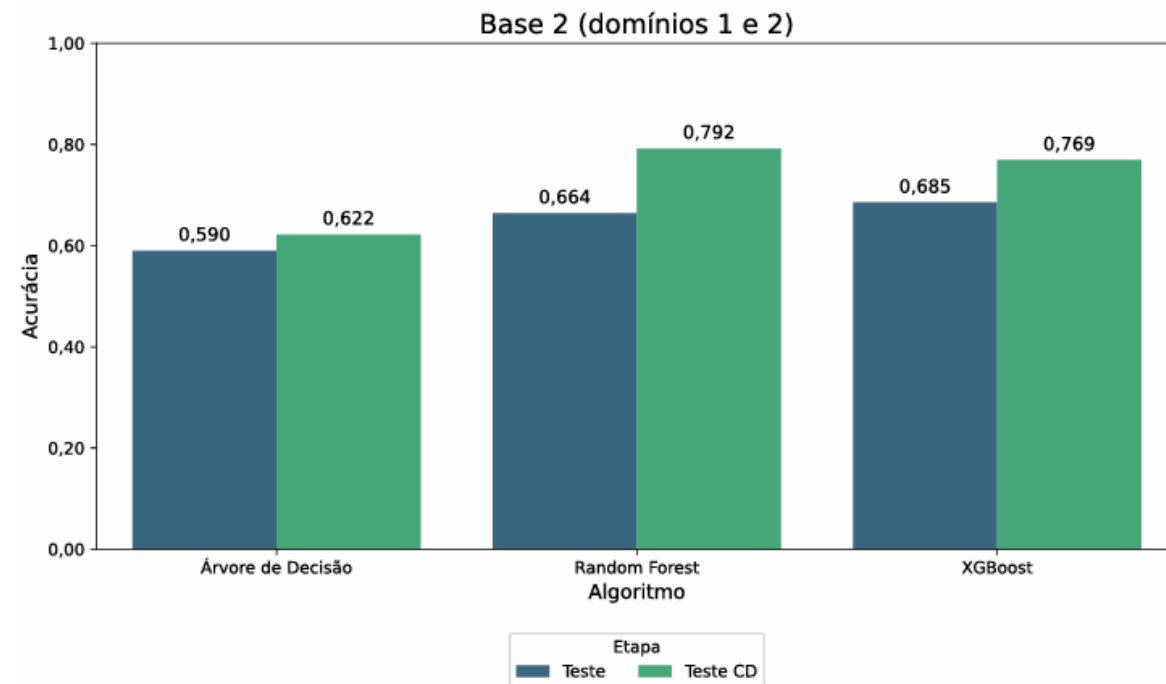
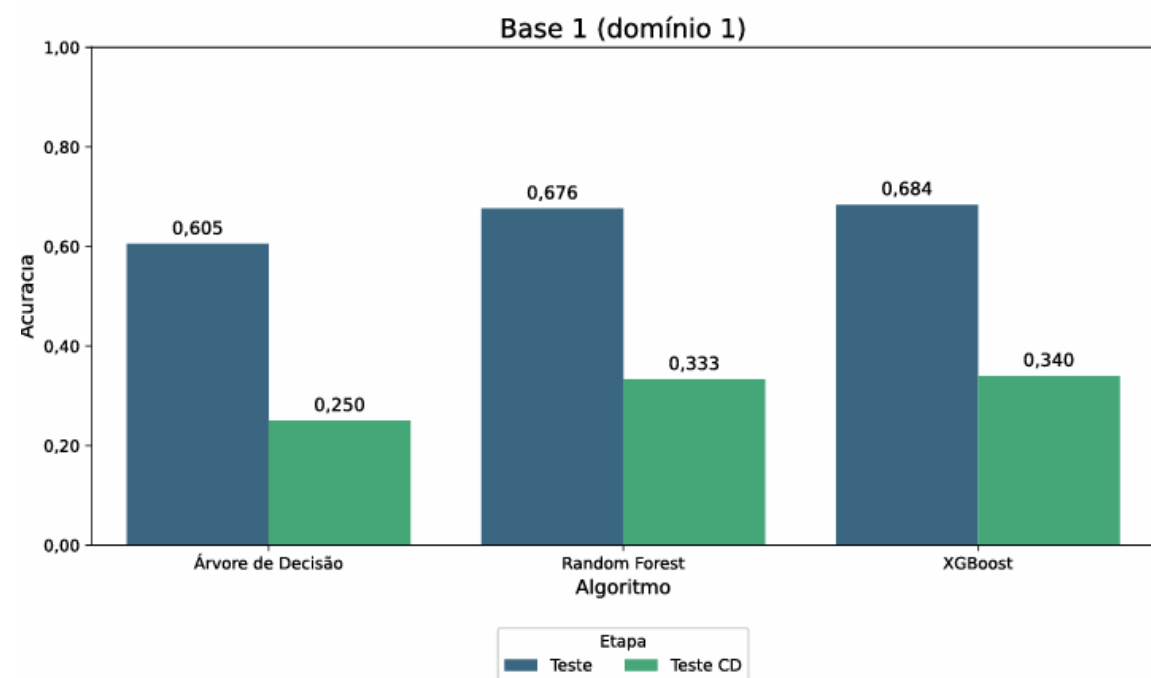
- Combinação de múltiplas árvores de decisão;
- Redução da variância e aumento de robustez.

XGBoost

(Chen & Guestrin, 2016)

- Modelo com correção de erros sequenciais em várias etapas;
- Captura de padrões mais complexos.

Resultados e Discussão



Resultados e Discussão

Tabela 5. Parâmetros de Avaliação - Base 1 (domínio 1)

Modelo	Precisão	Sensibilidade	F1-Score
Árvore de Decisão	0,593	0,593	0,593
"Random Forest"	0,673	0,669	0,666
"XGBoost"	0,682	0,677	0,678

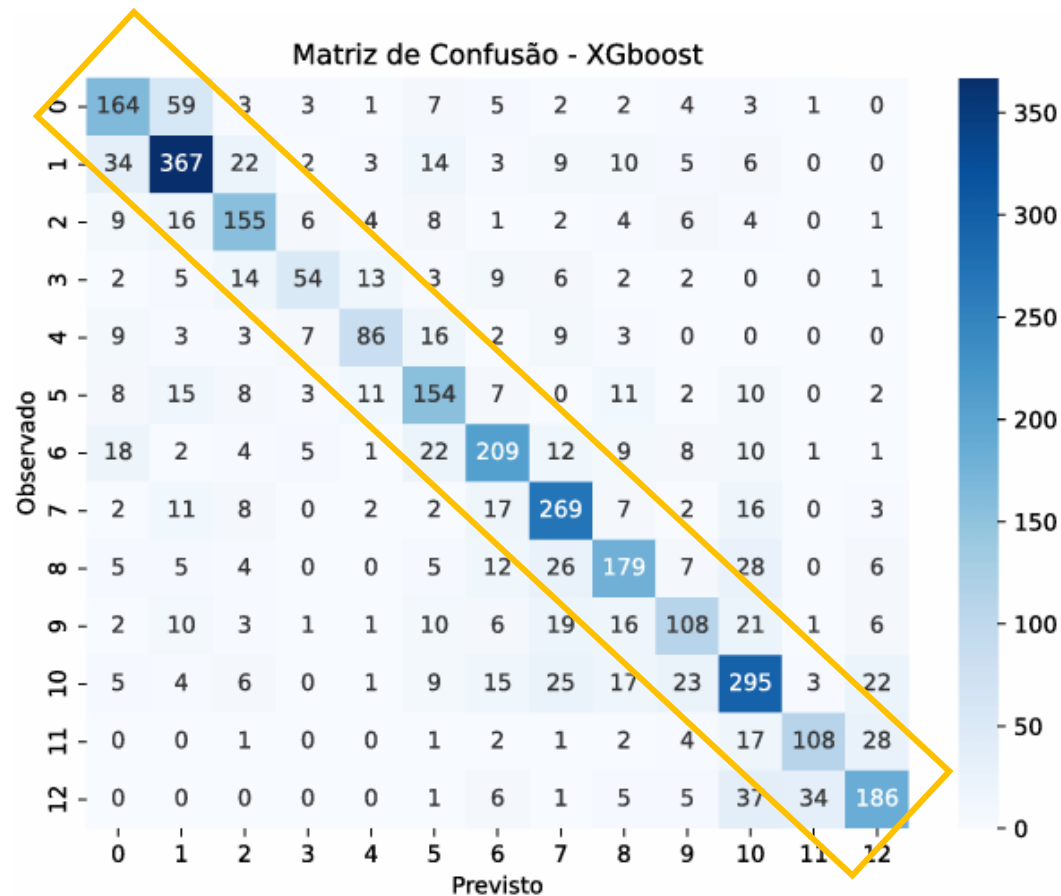
Fonte: resultados originais da pesquisa

Tabela 6. Parâmetros de Avaliação - Base 2 (domínios 1 e 2)

Modelo	Precisão	Sensibilidade	F1-Score
Árvore de Decisão	0,587	0,583	0,584
"Random Forest"	0,657	0,634	0,642
"XGBoost"	0,674	0,643	0,653

Fonte: resultados originais da pesquisa

Resultados e Discussão – Exemplo base 1



Destaques

- Tendência a erros adjacentes
- Especialização em algumas classes
- Classes altas x classes baixas

Análise Complementar – Base 1 (domínio 1)

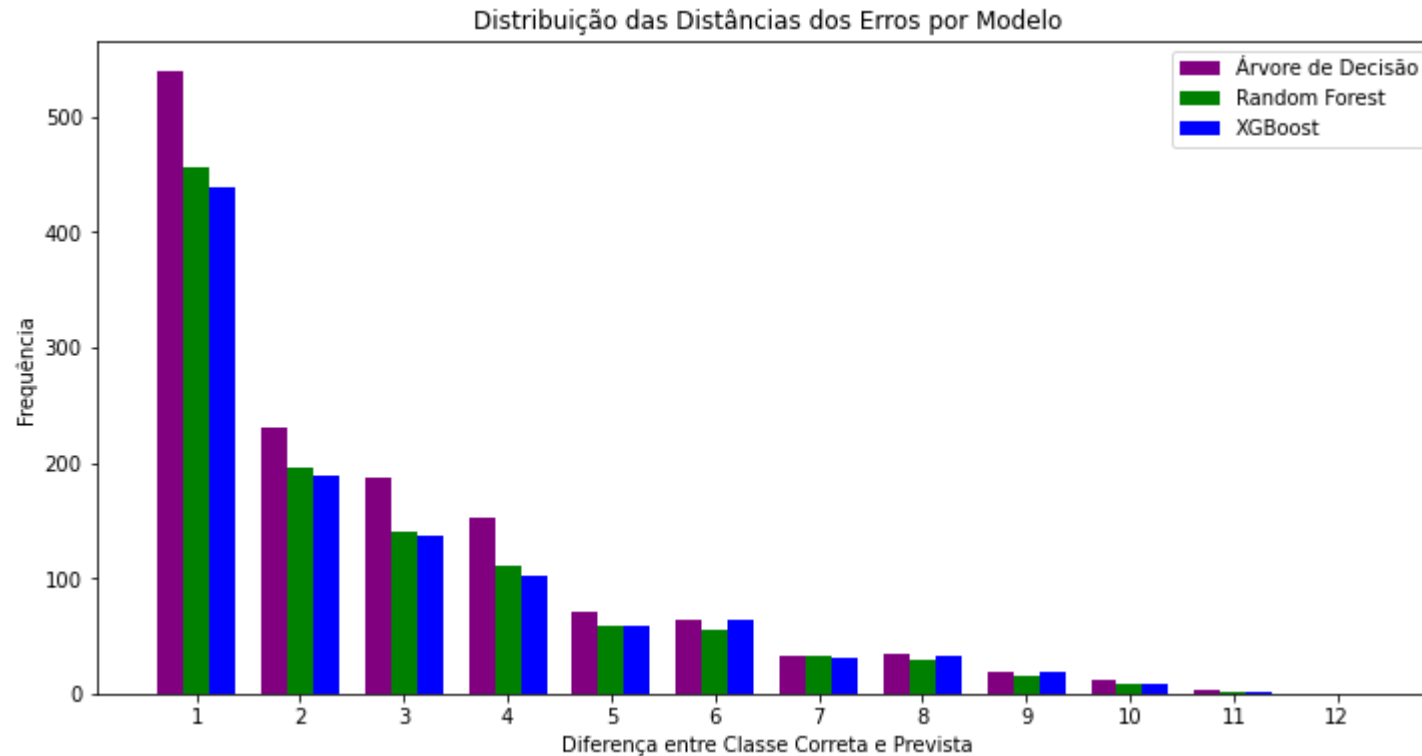


Tabela 11. Detalhamento de Acurácia e Erros

Modelo	Acurácia		Erros		
	Geral	Top-3 (K=3)	Quantidade	Média da Distância	Desvio-padrão
Árvore de Decisão	0,605		1347	2,783	2,146
"Random Forest"	0,676	0,977	1105	2,755	2,154
"XGBoost"	0,684	0,983	1077	2,822	2,210

Fonte: elaboração própria

Análise Complementar – Base 2 (domínios 1 e 2)

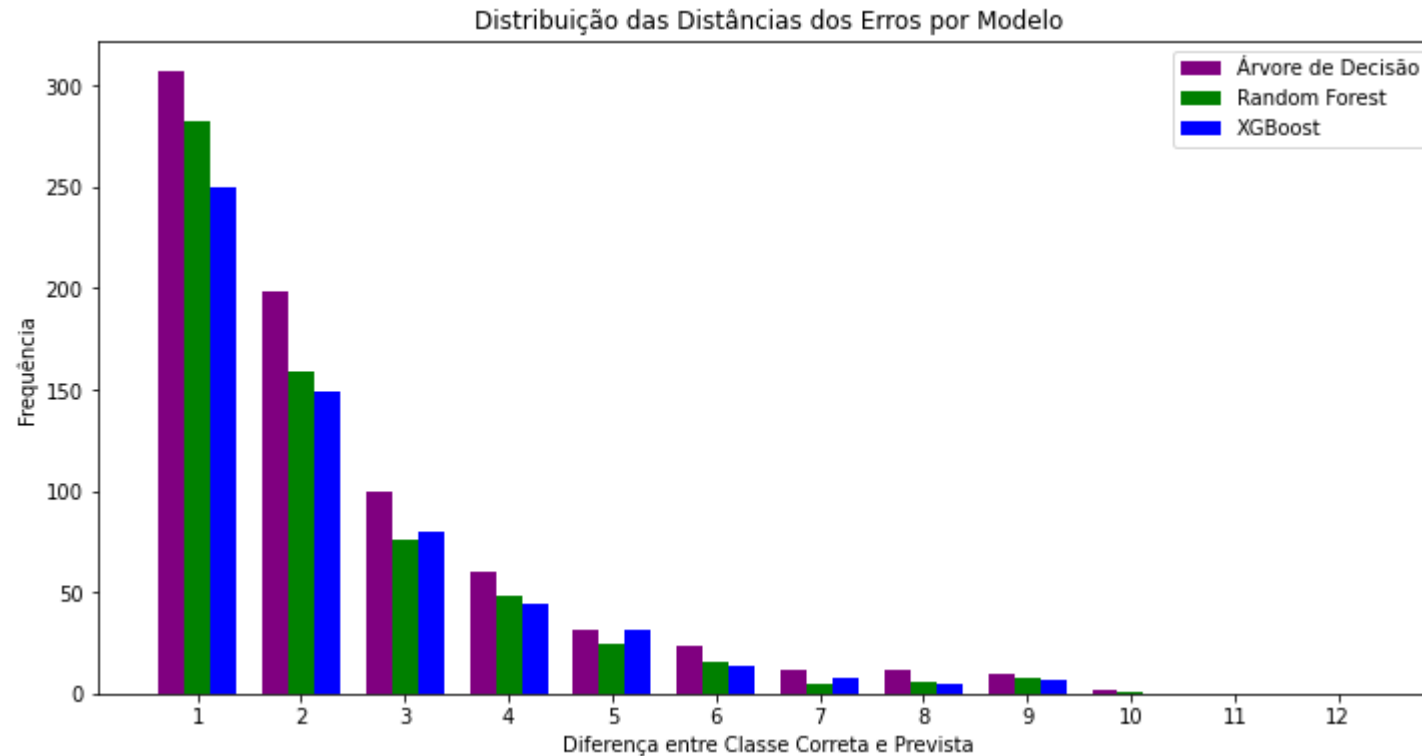


Tabela 12. Detalhamento de Acurácia e Erros

Modelo	Acurácia		Erros		
	Geral	Top-3 (K=3)	Quantidade	Média da Distância	Desvio-padrão
Árvore de Decisão	0,591		756	2,423	1,827
"Random Forest"	0,665	0,935	626	2,246	1,691
"XGBoost"	0,685	0,951	588	2,316	1,687

Fonte: elaboração própria

Resultados e Discussão

Tabela 10. Índice de Competitividade - Geral

Referência	Índice original	Modelos		
		"XGBoost"	"Random Forest"	Árvore de Decisão
Base 1 (domínio 1)	82,0%	82,3%	82,2%	61,2%
Base 2 (domínios 1 e 2)	79,3%	79,6%	79,6%	55,6%

Fonte: resultados originais da pesquisa

Conclusão

Avaliação Geral dos Modelos

Modelo	Acurácia Geral	Consistência dos Erros	Competitividade	Observações
XGBoost	Alta	Boa	Alta	Destacou-se como o mais robusto entre os modelos testados.
Random Forest	Boa	Alta	Alta	Alternativa confiável ao XGBoost com erros mais consistentes.
Árvore de Decisão	Baixa ¹	Baixa	Baixa	Limitações na precisão e competitividade; alta incidência de erros.

¹Acurácia mais baixa entre os modelos apresentados; o índice se refere a termos relativos.

Conclusão

Aplicação de Diferentes Bases

Base 1 (domínio 1):

- Tendência ao overfitting, adaptando-se às particularidades do ambiente controlado;
- Acurácia mais alta, mas com características previsíveis e menor variabilidade.

Base 2 (domínios 1 e 2):

- Redução do overfitting e maior capacidade de adaptação.
- Modelo mais generalista, embora com leve sacrifício de precisão.

Qual abordagem melhor atende aos objetivos do negócio:

Modelo especialista que maximiza precisão em um ambiente específico?

Modelo generalista que traz flexibilidade para um ambiente multissegmentado?

Conclusão

Contribuições do estudo:

Inovação em pricing: Viabilidade na automação do processo de gestão de demanda em companhias aéreas;

Relevância da análise de erros: Visão importante sobre aderência e tolerância do modelo ao negócio.

Perspectivas futuras:

Aprimoramento dos modelos: Deep learning, ajustes finos, redução de erros mais graves;

Aplicação de novas métricas: Aprofundar as análises dos erros, observar impactos de erros por classe ou grupo de mix e melhorar a precisão das previsões.

Bibliografia – Principais Literaturas

- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794
- Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81–106.
- Philips LR. (2011) *Pricing and Revenue Optimization*. Stanford, California: Stanford University Press.
- Ganin, Y. et al. (2016). Domain-Adversarial Training of Neural Networks. *The Journal of Machine Learning Research*.
- Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. New York: Springer.