

Série dos Seminários de Acompanhamento à Pesquisa

DEI
DEPARTAMENTO
DE ENGENHARIA
INDUSTRIAL

Número 21 | 09 2021

Aplicação de inteligência artificial interpretável para o problema de Churn não-contratual

Autor:

Leonardo Domingues



Série dos Seminários de Acompanhamento à Pesquisa



Número 21 | 09 2021

Aplicação de inteligência artificial interpretável para o problema de Churn não-contratual

Autor:

Leonardo Domingues

Orientador: Davi Michel Valladão



CRÉDITOS:

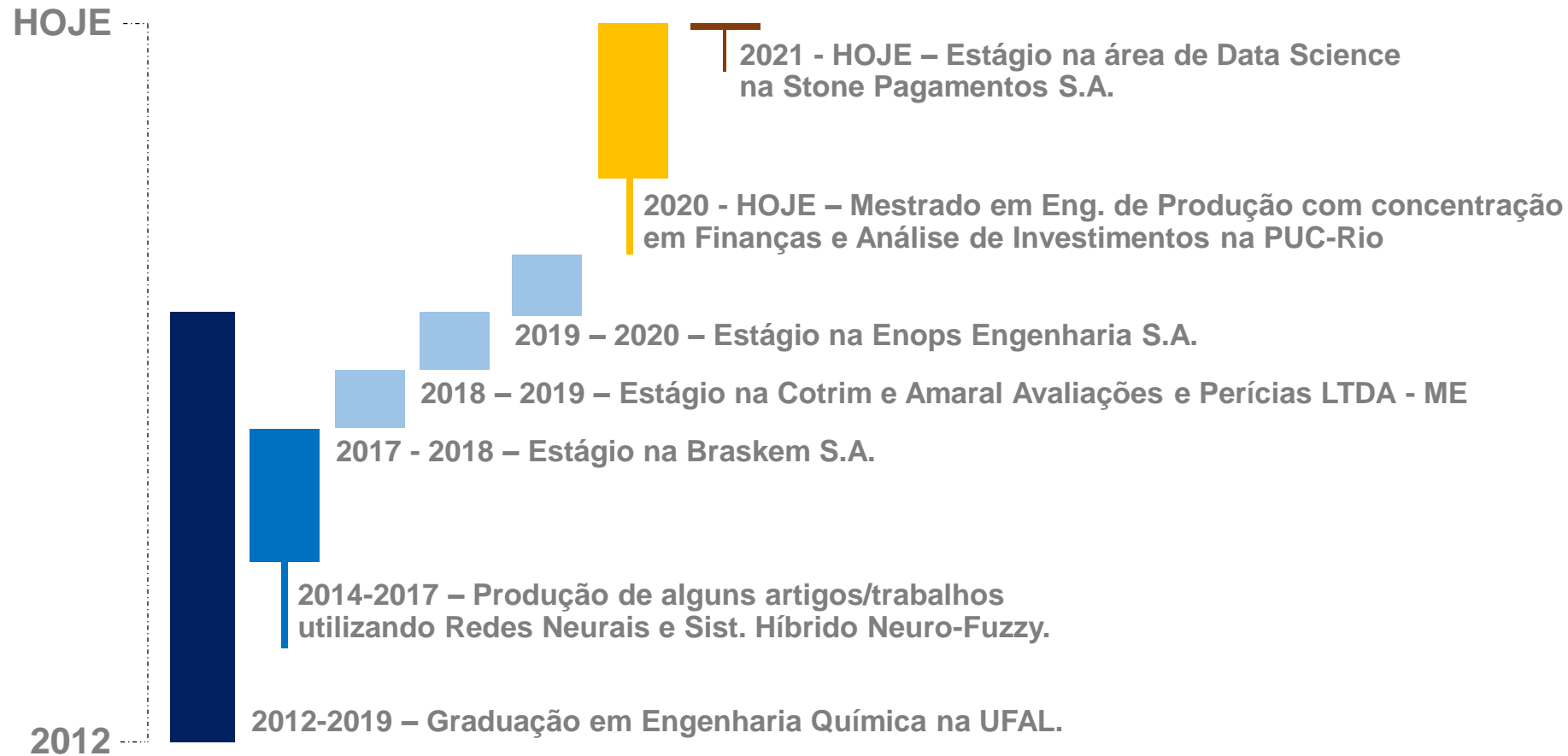
SISTEMA MAXWELL / LAMBDA
<https://www.maxwell.vrac.puc-rio.br/>

Organizadores: Fernanda Baião / Soraida Aguilar

Layout da Capa: Aline Magalhães dos Santos

Apresentação Pessoal

Minha linha do tempo acadêmica/profissional...



Agenda

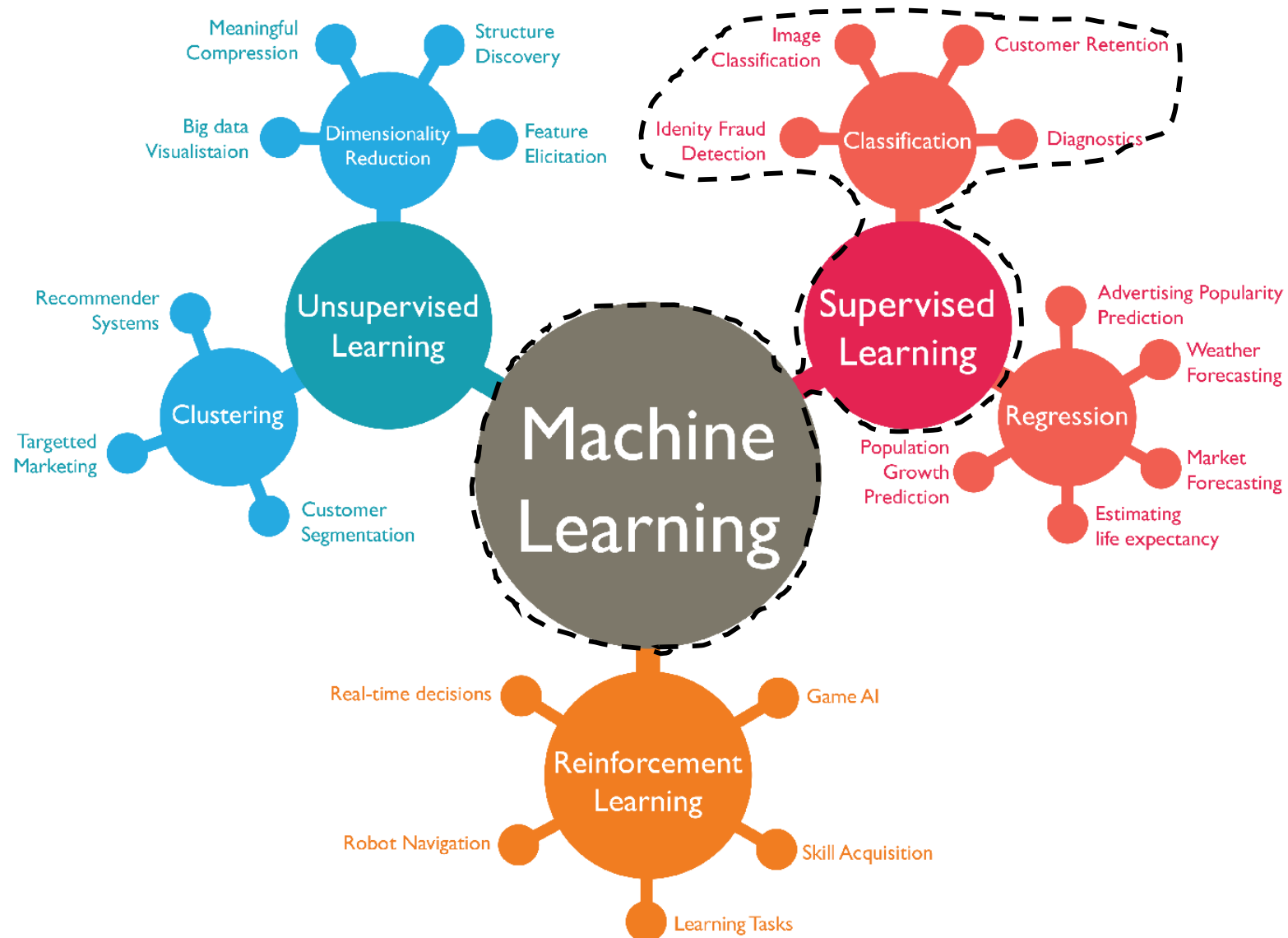
- **Motivação**
- **Técnicas Disponíveis**
- **Árvore de Decisão**
- **IA Interpretável**
- **Aplicações de Churn**
- **Desafios**
- **Próximos Passos**
- **Referências Bibliográficas**

Motivação

- Rotatividade de clientes
- Gerenc. do Relac. com o Cliente (CRM)
- Custo de aquisição >> Retenção de cliente
- Clientes de longa data consomem mais e ainda recomendam os produtos/serviços
- Num mundo competitivo a retenção de clientes é essencial
- São 2 tipos de churn: **contratual e não-contr.**
- **Objetivo:** utilização de machine learning interpretável para identificação do padrão pré-churn para tomada de decisão de retenção.



Motivação

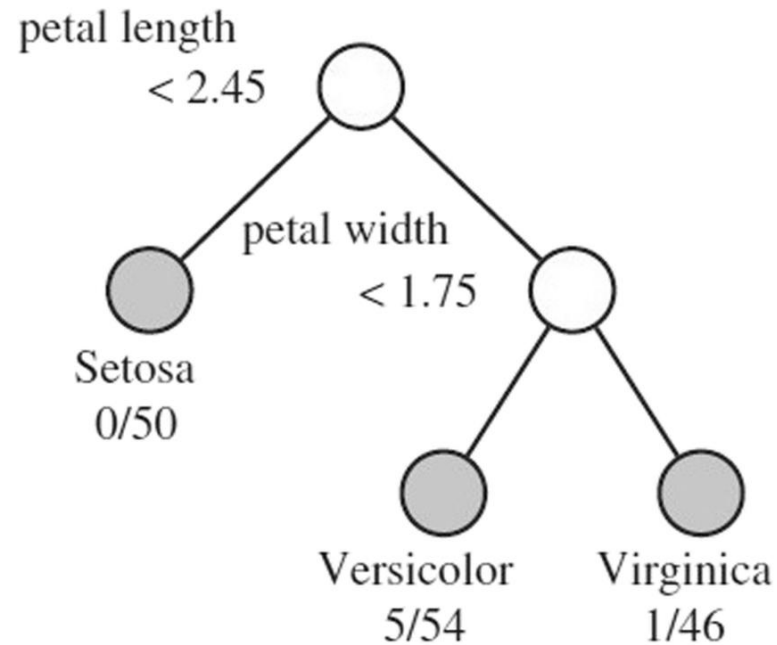


Várias técnicas disponíveis:

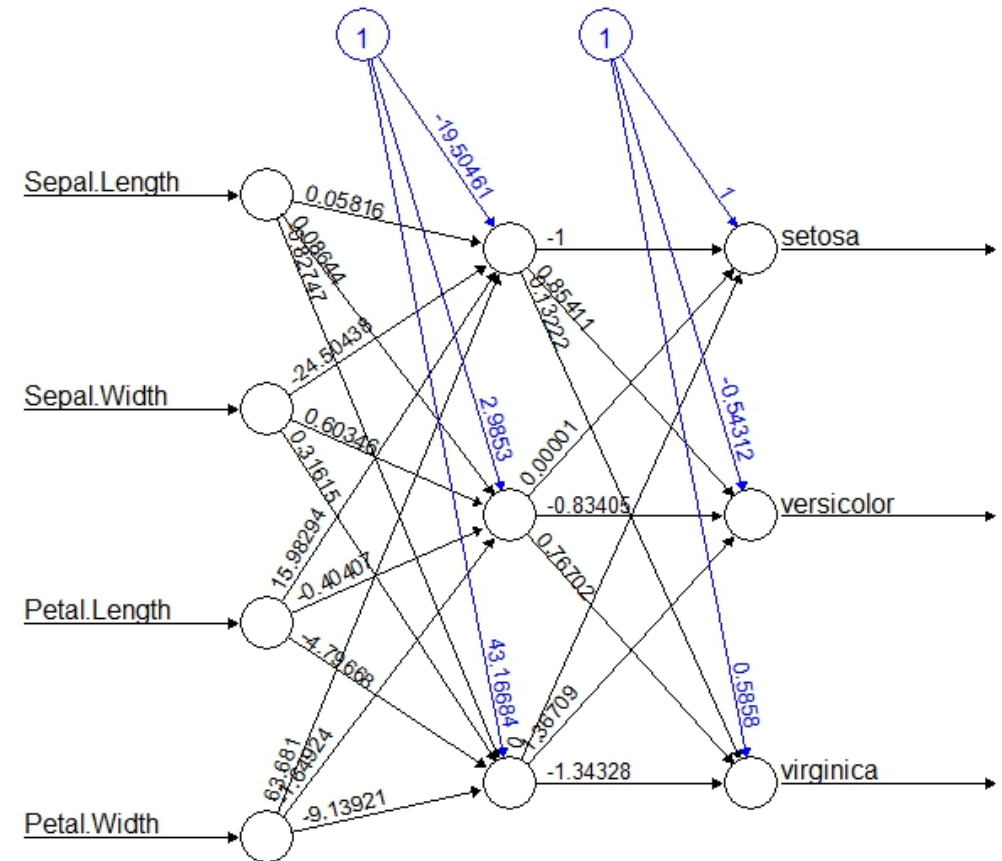
Tradicionais	Ensembles	Neural Networks
<ul style="list-style-type: none">• Classification Trees• KNN• SVM• Bayesian Classification	<ul style="list-style-type: none">• Random Forest• Gradient Boosting• Adaptive Boosting• XGBoost	<ul style="list-style-type: none">• Artificial Neural Networks (ANN)• Convolution Neural Networks (CNN)• Recurrent Neural Networks (RNN)
<ul style="list-style-type: none">• Interpretáveis• Baixa acurácia	<ul style="list-style-type: none">• Menos Interpretáveis (caixa preta)• Maior acurácia	

Output de algumas técnicas para Iris dataset:

CART



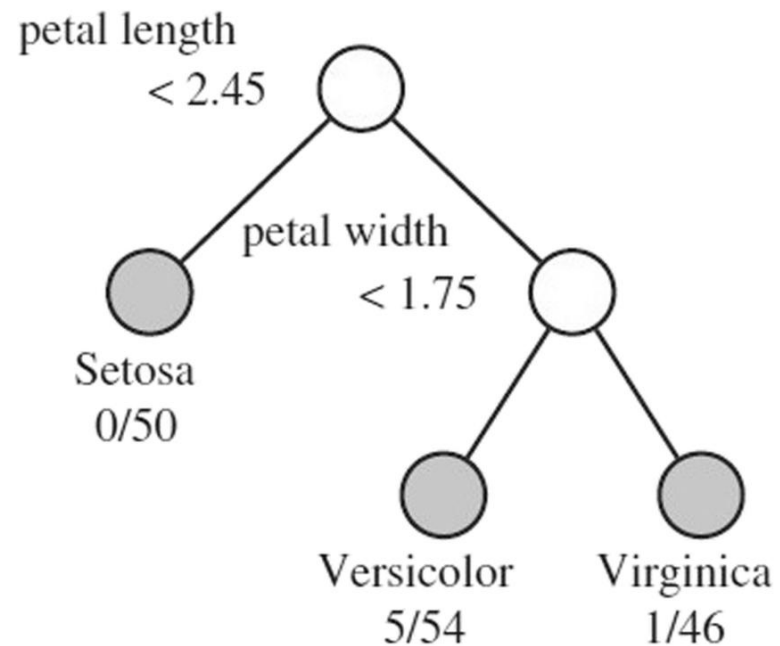
RNN



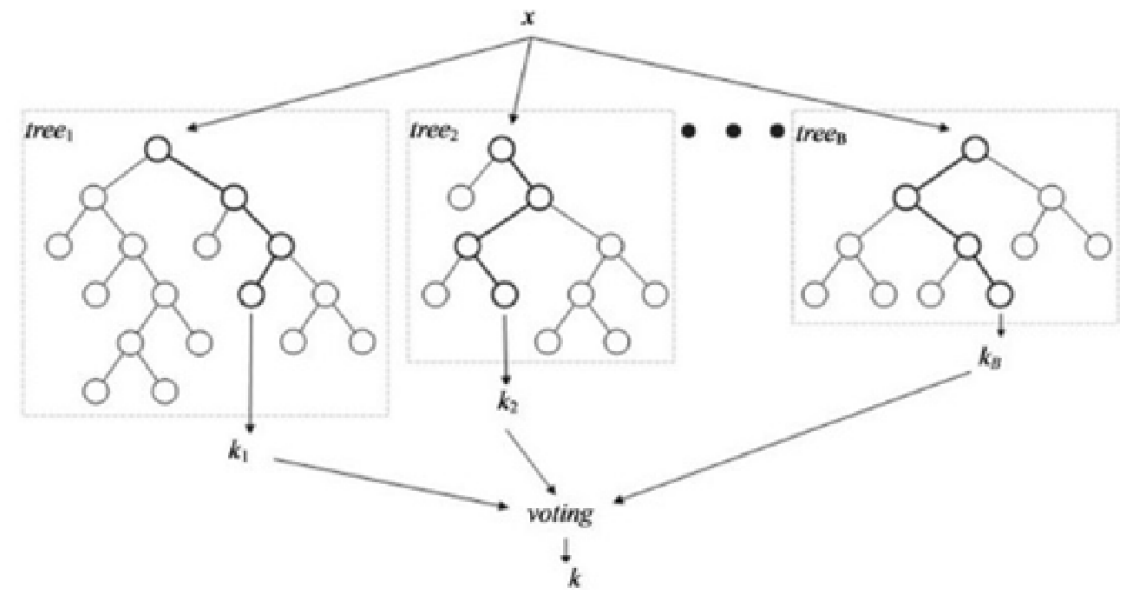
Error: 0.456121 Steps: 12733

Output de algumas técnicas para Iris dataset:

CART



Random Forest



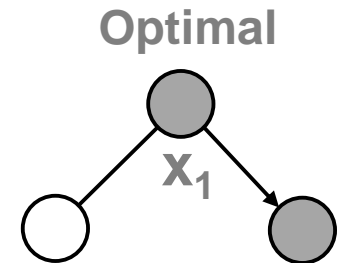
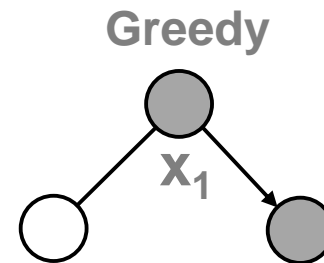
CART - Vantagens / Desvantagens



- Simples de entender, interpretar e visualizar;
- Realizam a triagem de variáveis ou seleção de atributos;
- Pouco esforço para a preparação dos dados;



- Árvores complexas que não generalizam bem (overfitting);
- Baixa acurácia;
- Greedy heuristic, onde a decisão ótima é tomada passo-a-passo;



CART - Vantagens / Desvantagens



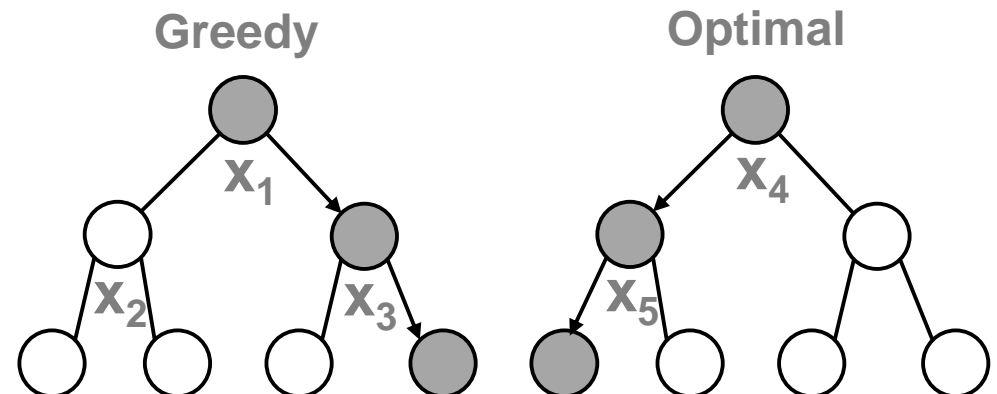
- Simples de entender, interpretar e visualizar;
- Realizam a triagem de variáveis ou seleção de atributos;
- Pouco esforço para a preparação dos dados;



- Árvores complexas que não generalizam bem (overfitting);
- Baixa acurácia;
- Greedy heuristic, onde a decisão ótima é tomada passo-a-passo;

D = 1

D = 2



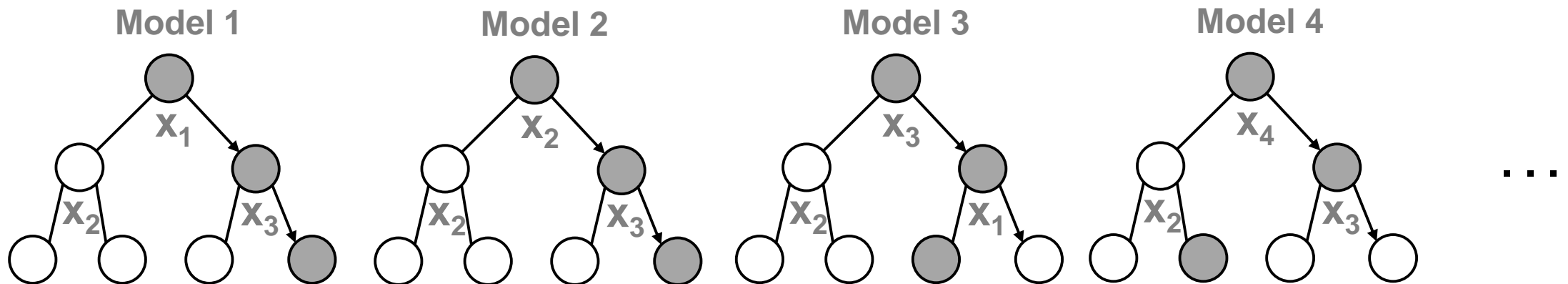
CART - Vantagens / Desvantagens



- Simples de entender, interpretar e visualizar;
- Realizam a triagem de variáveis ou seleção de atributos;
- Pouco esforço para a preparação dos dados;



- Árvores complexas que não generalizam bem (overfitting);
- Baixa acurácia;
- Greedy heuristic, onde a decisão ótima é tomada passo-a-passo;



CART - Vantagens / Desvantagens



- Simples de entender, interpretar e visualizar;
- Realizam a triagem de variáveis ou seleção de atributos;
- Pouco esforço para a preparação dos dados;



- Árvores complexas que não generalizam bem (overfitting);
- Baixa acurácia;
- Greedy heuristic, onde a decisão ótima é tomada passo-a-passo;

Será que conseguimos aumentar a interpretabilidade e a acurácia?

Sim! Através da Árvore de Decisão Ótima:

Mach Learn (2017) 106:1039–1082
DOI 10.1007/s10994-017-5633-9

Optimal classification trees

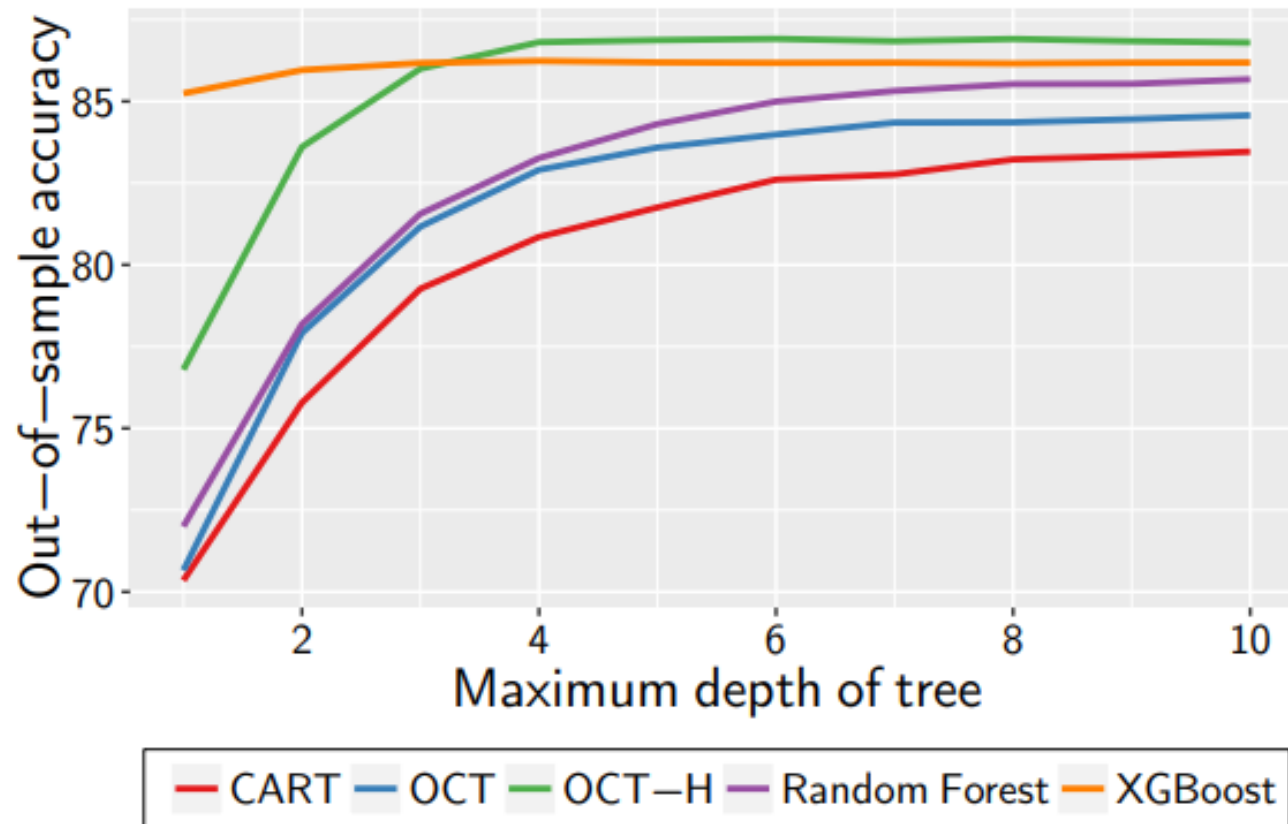
Dimitris Bertsimas¹ · Jack Dunn²

- Deixa de ser heurístico e se torna exato (MILP)
- A otimização é feita globalmente e não passo-a-passo
- Apresenta melhorias significativas em relação aos métodos heurísticos.

Received: 17 September 2015 / Accepted: 3 March 2017 / Published online: 3 April 2017
© The Author(s) 2017

Desempenho de árvores de classificação ótima

Acurácia média out-of-sample em 60 conjuntos de dados do mundo real:



Fonte: Bertsimas, D.
Interpretable AI. Acesso em :
<https://orfe.princeton.edu/pdo/sites/orfe.princeton.edu.pdo/files/Bertsimas_PDO.pdf>

Aplicação: Churn

Contratual

- **Características**
 - Serviço atrelado a um contrato
 - Churn é observado e é total (para cada produto)
 - Definição clara de churn (fim da contratação do serviço)
- **Aplicações**
 - Telefone pós-paga
 - Conta Corrente / Cartão de crédito
 - TV por assinatura
 - Subscrições online
 - Seguros



Não-contratual

- **Características**
 - Custo do cliente trocar o provedor do serviço / fornecedor é baixo ou nulo
 - Churn não é observado (cliente não avisa)
 - Um contrato não cancelado pode significar churn
 - Definição de churn é subjetiva
 - Os analistas levam em conta o que o churn representa na área de estudo.
- **Aplicações**
 - Telefonia pré-paga
 - Supermercado
 - Serviços financeiros



Tipos de Churn: • Ativo • Passivo • Forçado • Esperado

Aplicação de Churn contratual na prática (e direta)



Prever se um cliente mudará de telefônica

O dataset de treinamento contém 4250 (80%T / 20%V) amostras. Cada amostra contém 19 atributos e uma label "churn" que indica a classe da amostra.

Para a resolução deste problema utilizou-se o PyCaret e o pacote da empresa Interpretable AI (www.interpretable.ai) para a Árvore de Decisão Ótima

Resultados

	Modelo	Accuracy	AUC	Recall	Prec.	F1
octmio	Opt. Classification Tree	0,9660	0,8782	0,7560	1,0000	0,8608
catboost	CatBoost Classifier	0,9547	0,9102	0,7388	0,9363	0,8259
lightgbm	Light Gradient Boosting machine	0,9529	0,8949	0,7265	0,9350	0,8177
xgboost	Extreme Gradient Boosting	0,9512	0,8977	0,7245	0,9225	0,8116
gbc	Gradient Boosting Classifier	0,9447	0,9100	0,6939	0,9011	0,7840
dt	Decision Tree Classifier	0,9188	0,8338	0,7143	0,7210	0,7176
rf	Random Forest Classifier	0,9179	0,9058	0,4735	0,9164	0,7914

“De todos identificados como churners, quantos realmente eram churners”

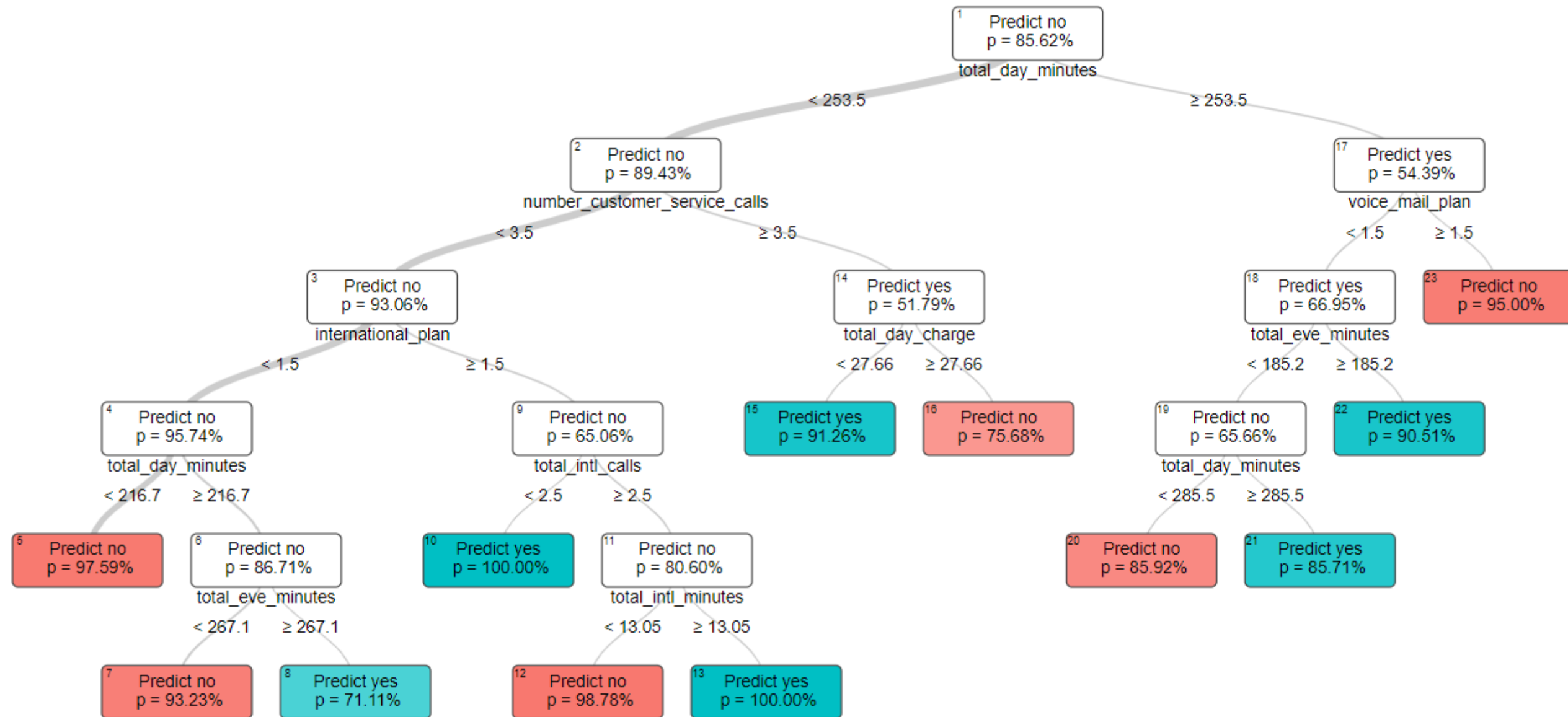
“De todos que realmente eram churners, quantos foram identificados como tal”

Real		
	Churner	Não-Churner
Previsto	Churner	FP
	Não-Churner	TN

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

Optimal Decision Tree (MIO)



Desafios do churn não-contratual:

- Tratar série temporais dentro de Machine Learning;
- Qual janela os dados serão agregados ? 1 hora? 1 dia? 1 semana? . . .
- Definição de Churn: quantos dias sem transacionar? 7 dias? 14 dias? 28 dias? . . .
- O cliente que transaciona um valor pequeno todo dia deu churn?
- **Seleção de amostras:** considerar clientes esporádicos? Clientes com padrão irregular? Clientes recentes? Clientes com ticket alto? Clientes com nº de transações extremamente alto? . . .
- Definir e escolher as variáveis RFM a serem implantadas no modelo
- Considerar no modelo variáveis estáticas? Nº de dias na base, tipo de cliente (CPF, CNPJ ou MEI), segmento . . .

Churn não contratual na prática para uma adquirente:

Recency	Frequency	Monetary
<i>days from first:</i> Tempo decorrido desde a primeira transação até hoje	<i>average trx per day (all):</i> Frequência média - número de transações / número de dias corridos.	<i>tpv monthly average:</i> Média do TPV mensal
<i>days from first to last:</i> Tempo decorrido desde a primeira transação até a última.	<i>average trx per day with trx:</i> Frequência média em dias ativos - número de transações / número de dias com transação.	<i>tpv average trx:</i> Ticket médio por transação
	<i>trx days / days from first to last:</i> Proporção dos dias corridos que tiveram transação	<i>average trx per day with trx:</i> Frequência média em dias ativos - número de transações / número de dias com transação.
	<i>standard deviation IPT*:</i> Desvio padrão dos intervalos entre transações.	<i>trx days / days from first to last:</i> Proporção dos dias corridos que tiveram transação

Próximos passos . . .



- **Na literatura, em relação à configuração não-contratual:**
 - Como está sendo abordada? Quais são os preditores mais utilizados?
 - Utilizam as variáveis RFM somente? Utilizam clusterização? Utilizam Customer Lifetime Value (CLV)?
 - Como lidam com as séries temporais?
 - Quais as técnicas de machine learning mais utilizadas?
 - Existem estudos para uma adquirente?
- **Definir a feature engineering para tratar o problema**

Referências Bibliográficas

- [1] ALBADVI, A.; VARASTEY, S. Customers churn behavior modeling using decision trees (a case study in non-contractual setting). *Advances in Industrial Engineering, University of Tehran*, v. 44, n. 2, p. 127–139, 2010.
- [2] ALI, Ö. G.; ARITÜRK, U. Dynamic churn prediction framework with more effective use of rare event data: The case of private banking. *Expert Systems with Applications, Elsevier*, v. 41, n. 17, p. 7889–7903, 2014.
- [3] BABKIN, A.; GOLDBERG, I. Incorporating time-dependent covariates into bg-nbd model for churn prediction in non-contractual settings. Available at SSRN 2905307, 2017.
- [4] BENGIO, Y.; LODI, A.; PROUVOST, A. Machine learning for combinatorial optimization: a methodological tour d’horizon. *European Journal of Operational Research, Elsevier*, 2020.
- [5] BERTSIMAS, D.; DUNN, J. Optimal classification trees. *Machine Learning, Springer*, v. 106, n. 7, p. 1039–1082, 2017.
- [6] BIN, L.; PEIJI, S.; JUAN, L. Customer churn prediction based on the decision tree in personal handy phone system service. In: *IEEE.2007 International Conference on Service Systems and Service Management*. [S.l.], 2007. p. 1–5.

Referências Bibliográficas

- [7] BUCKINX, W.; POEL, D. Van den. Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual fmcg retail setting. European journal of operational research, Elsevier, v. 164, n. 1, p. 252–268, 2005.
- [8] BULT, J. R.; WANSBEEK, T. Optimal selection for direct mail. Marketing Science, INFORMS, v. 14, n. 4, p. 378–394, 1995.
- [9] CALCIU, M. Deterministic and stochastic customer lifetime value models. Evaluating the impact of ignored heterogeneity in non-contractual contexts. Journal of Targeting, Measurement and Analysis for Marketing, Springer, v. 17, n. 4, p. 257–271, 2009.
- [10] DAWES, J.; SWAILES, S. Retention sans frontieres: issues for financial service tailors. International Journal of Bank Marketing, MCB UP Ltd, 1999.
- [11] FADER, P. S.; HARDIE, B. G. Probability models for customer-base analysis. Journal of interactive marketing, Elsevier, v. 23, n. 1, p. 61–69, 2009.
- [12] GLADY, N.; BAESENS, B.; CROUX, C. Modeling churn using customer lifetime value. European Journal of Operational Research, Elsevier, v. 197, n. 1, p. 402–411, 2009.

Referências Bibliográficas

- [13] HADIJI, F.; SIFA, R.; DRACHEN, A.; THU-RAU, C.; KERSTING, K.; BAUCKHAGE, C. Predicting player churn in the wild. In: IEEE. 2014 IEEE Conference on Computational Intelligence and Games.[S.l.], 2014. p. 1–8.
- [14] HOPMANN, J.; THEDE, A. Applicability of customer churn forecasts in a non-contractual setting. In: Innovations in classification, data science, and information systems.[S.l.]: Springer, 2005. p. 330–337.
- [15] JAHROMI, A. T.; SEPEHRI, M. M.; TEIMOURPOUR, B.; CHOEBDAR, S. Modeling customer churn in a non-contractual setting: the case of telecommunications service providers. Journal of Strategic Marketing, Taylor & Francis, v. 18, n. 7, p.587–598, 2010.
- [16] JAHROMI, A. T.; STAKHOVYCH, S.; EWING, M. Managing b2b customer churn, retention and profitability. Industrial Marketing Management, Elsevier, v. 43, n. 7, p. 1258–1268, 2014
- [17] KERAMATI, A.; ARDABILI, S. M. Churn analysis for an Iranian mobile operator. Telecommunications Policy, Elsevier, v. 35, n. 4, p. 344–356, 2011.
- [18] KERAMATI, A.; GHANEEI, H.; MIR-MOHAMMADI, S. M. Developing a prediction model for customer churn from electronic banking services using data mining. Financial Innovation, Springer, v. 2, n. 1, p. 1–13, 2016.

Referências Bibliográficas

- [19] MANDIĆ, M.; KRALJEVIĆ, G.; BOBAN, I. Performance comparison of six data mining models for soft churn customer prediction in telecom. IJEEC-INTERNATIONAL JOURNAL OF ELECTRICAL ENGINEERING AND COMPUTING, v. 2,n. 1, 2018.
- [20] MARTÍNEZ, A.; SCHMUCK, C.; JR, S. P.;PIRKER, C.; HALTMEIER, M. A machine learning framework for customer purchase prediction in the non-contractual setting. European Journal of Operational Research, Elsevier, v. 281, n. 3, p. 588–596,2020.
- [21] MIN, S.; ZHANG, X.; KIM, N.; SRIVAS-TAVA, R. K. Customer acquisition and retention spending: An analytical model and empirical investigation in wireless telecommunications markets. Journal ofmarketing research, SAGE Publications Sage CA: Los Angeles, CA, v. 53, n. 5, p.728–744, 2016.
- [22] POEL, D. Van den; LARIVIERE, B. Customer attrition analysis for financial services using proportional hazard models. European journal of operational research, Elsevier, v. 157, n. 1, p. 196–217,2004.
- [23] PRASAD, U. D.; MADHAVI, S. Prediction of churn behaviour of bank customers using data mining tools. Indian Journal of Marketing, v. 42, n. 9, p. 25–30, 2012.

Referências Bibliográficas

- [24] VAFEIADIS, T.; DIAMANTARAS, K. I.; SARIGIANNIDIS, G.; CHATZISAVVAS, K. C. A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, Elsevier, v. 55, p. 1–9, 2015.
- [25] WANG, Q. Design the churn analysis on games: a review on techniques for churn analysis. Tese (Doutorado) —Northeastern University, 2018.
- [26] WEERAHANDI, S.; MOITRA, S. Using survey data to predict adoption and switching for services. *Journal of Marketing Research*, SAGE Publications Sage CA: Los Angeles, CA, v. 32, n. 1, p. 85–96, 1995.
- [27] YOON, S.; KOEHLER, J.; GHOBARAH, A. Prediction of advertiser churn for google adwords. 2010.