



Universidade Federal do Paraná
Setor de Ciências Exatas
Departamento de Estatística
Programa de Especialização em Data Science e Big
Data

Paulo Eduardo de Souza Jardim

Redes Bayesianas Aplicadas à Jornada de Compra no Marketing Digital

**Curitiba
2022**

Paulo Eduardo de Souza Jardim

Redes Bayesianas Aplicadas à Jornada de Compra no Marketing Digital

Monografia apresentada ao Programa de Especialização em *Data Science e Big Data*, Departamento de Estatística, Setor de Ciências Exatas, Universidade Federal do Paraná, como requisito parcial para a obtenção do grau de especialista.

Orientador: Prof. Dr. Anderson Ara

Coorientadora: Prof. Dra. Lilia Carolina da Costa

**Curitiba
2022**

Redes Bayesianas Aplicadas à Jornada de Compra no Marketing Digital

Paulo Eduardo S. Jardim¹, Prof. Dr. Anderson Ara², Profa. Dra. Lilia Carolina da Costa³

¹Departamento de Estatística, Universidade Federal do Paraná Rua Evaristo F. F. da Costa 418, Jardim das Americas, 82590-300, Curitiba, PR, Brasil*

²Departamento de Estatística, Universidade Federal do Paraná Rua Evaristo F. F. da Costa 418, Jardim das Americas, 82590-300, Curitiba, PR, Brasil†

³Departamento de Estatística, Instituto de Matemática, Universidade Federal da Bahia Av. Adhemar de Barros S/N-Campus Universitário de Ondina, Ondina, 40170-110, Salvador, BA, Brasil‡

Um dos maiores desafios das empresas que comercializam *software* na modalidade SaaS (*Software as a Service*) é modelar a jornada de compra de seus clientes, sendo assim possível conhecer as etapas mais importantes deste processo e aperfeiçoá-lo. Conseguir otimizar os esforços do time de vendas, focando nos futuros clientes com maior potencial é crucial para reduzir o custo de aquisição de clientes e garantir a competitividade no mercado. Por isso, este trabalho propõe uma abordagem com Redes Bayesianas para esse problema. Foi ajustado um modelo, por meio do qual foram feitas inferências visuais sobre o processo, bem como a simulação da jornada de compra de um cliente. Foram demonstrados vários tipos de previsões que podem ser feitas durante o processo, facilitando a tomada de decisões no nível operacional de vendas.

Palavras-chave: Redes Bayesianas, Marketing Digital, Jornada de Compra.

One of the greatest challenges faced by SaaS (Software as a Service) vendors is modeling their customers' buying journey, which allows understanding the most important steps of this process in order to improve it. Optimizing the sales force effort, focusing on those future customers who have great potential it's critical to reduce the customer acquisition costs and to guarantee the company's competitiveness in the market. Therefore, this work proposes a Bayesian Networks approach to this problem. A model has been fitted, through which we have performed visual inferences about the process and also a simulation on a customer's buying journey. We have demonstrated various kinds of predictions that can be performed during the process, facilitating decision making at the sales operational level.

Keywords: Bayesian Networks, Digital Marketing, Buyer Journey.

1. Introdução

O mercado SaaS (*Software as a Service*) gerou U\$ 150 bilhões em receitas em 2021, com uma estimativa de atingir U\$ 208 bilhões em 2023 [14]. Esse modelo, no qual o *software* é comercializado como um serviço por assinatura, traz inúmeros benefícios para as organizações, sendo o baixo investimento inicial e a escalabilidade de uso alguns dos principais. Com isso, a modalidade SaaS se tornou a principal forma de comercialização de *software* na modalidade *Business to Business* (B2B) nos últimos 20 anos e levou todos os

principais fornecedores de *software* a se adequarem a essa maneira de fornecer seu produto. Hoje, os novos produtos de *software* já são criados originalmente com foco nesta modalidade, aumentando constantemente o nível de concorrência no setor.

A alta concorrência no segmento SaaS levou as empresas a fornecerem diferentes canais e estratégias para que os clientes conheçam seu produto, de todas as maneiras possíveis, e executem a compra ou assinatura do serviço no momento mais adequado para o próprio cliente. Isso trouxe dois principais desafios: mapear a complexa jornada de compra e controlar o custo de aquisição de novos clientes.

Pelo fato de que os clientes são outras empresas, o processo de vendas pode se tornar bem complexo mas,

*eduardo.jardim@ufpr.br

†ara@ufpr.br

‡liliacosta@ufba.br

de forma geral, se resume às seguintes etapas: 1) cliente conhece o produto na Internet; 2) entra no site do fornecedor SaaS e faz um cadastro, também chamado de *sign-up* para experimentar o produto online; 3) time de vendas entra em contato com os clientes que aparentam ter maior probabilidade de comprar e agendam uma demonstração personalizada, onde o cliente conhece mais do produto e o representante de vendas conhece mais das necessidades do cliente; 4) o representante de vendas apresenta uma proposta de venda do produto e conduz a fase de negociações que pode terminar ou não em uma venda.

Este caminho, na prática, não é tão linear quanto as etapas descritas acima. Como os produtos de *software* empresarial costumam ser complexos, assim também é o processo de validação do cliente, representado principalmente pela etapa 2 acima. Nesta fase, o potencial cliente, também chamado de *lead*, gera muitos dados conforme navega no produto, criando inúmeras possibilidades de ações por parte da equipe de vendas, para responder ao cliente de uma forma personalizada e escolher os clientes corretos e o momento certo de empregar seus esforços.

No passado, uma jornada de compra era quase sempre abstraída em um modelo de funil, no qual muitos potenciais clientes entravam no topo desse funil, percorrendo um caminho com fases lineares e bem definidas, ficando muitos deles pelo caminho até que alguns chegavam ao fundo do funil, o que seria o evento da compra. Mas este modelo não mais atende as equipes de *marketing* e vendas [5], modelos mais robustos se fazem necessários para que essas empresas avaliem a performance da sua estratégia de aquisição de clientes e para que possam influenciar cada cliente a caminhar para a compra de maneira personalizada, considerando as características, comportamento e o momento de cada um.

Neste contexto, este trabalho propõe o uso de Redes Bayesianas para modelar essa jornada, na qual o comportamento do cliente será expressa por variáveis que exercem influência umas sobre as outras, existindo várias rotas para se chegar no objetivo final que é a compra.

O outro desafio listado acima é o controle dos custos de aquisição de clientes. A métrica CAC (*Customer Acquisition Cost*) é hoje a principal métrica a ser otimizada, principalmente nos primeiros anos de uma empresa SaaS. Soma-se todo o investimento em Marketing e equipe de Vendas e divide-se pela quantidade de novos clientes pagantes.

Uma solução para esse problema é a correta priorização do esforço do time de vendas. Milhares de potenciais novos clientes (*leads*) fazem cadastro todos os dias para experimentar o produto. O time de vendas precisa escolher os *leads* com maior potencial para focar seus esforços e, assim, melhorar a eficiência, diminuindo o custo de aquisição.

As Redes Bayesianas podem colaborar neste processo de priorização de esforço, de uma forma mais interessante do que outros métodos tradicionais de aprendizagem de máquina. Isso porque elas oferecem a possibilidade de identificar clientes com maior probabilidade de avançar para a próxima fase e, finalmente, de realizar a compra. Independente do momento em que o cliente se encontra na jornada, pois é possível fazer previsões para diferentes tipos de jornada de compra, mesmo através da observação de apenas algumas variáveis.

Percebe-se que a utilização de Redes Bayesianas ainda é bastante comedida no meio corporativo. Acredita-se que isso se deve ao ainda baixo conhecimento dessa técnica, que é relativamente nova, por parte de analistas de *marketing*, analistas de dados e cientistas de dados.

Trabalhos recentes têm utilizado Redes Bayesianas neste contexto. Anderl et al. (2016) utilizaram Cadeias de Markov para analisar jornadas de cliente, afim de propor um modelo de atribuição de canais de marketing [1]. Embora esse seja uma aplicação interessante de um modelo gráfico para um problema de *marketing*, a atribuição de canais consiste em determinar quais campanhas de *marketing* trouxeram cada cliente, afim de posteriormente medir a eficácia das mesmas. Isso é diferente de prever quais clientes tem maior potencial de compra durante a jornada. Não foram encontrados trabalhos que aplicassem Redes Bayesianas com esse objetivo.

Hosseini (2021) apresenta um sistema de suporte à decisão em vendas baseado em Redes Bayesianas, demonstrando os benefícios desse método na otimização do trabalho de uma equipe de vendas em um cenário de Telemarketing [2].

Yelland et al. (2010) utilizam as Redes Bayesianas para realizar projeção de vendas, citando como principal benefício do método a capacidade de combinar o conhecimento da equipe de vendas com informações estatísticas [3].

Qazi et al. (2017) empregam Redes Bayesianas em um sistema de recomendação de produtos no mercado de seguros, destacando os benefícios para os corretores

na condução das vendas de forma personalizada para seus clientes [4].

Este trabalho busca demonstrar a aplicação de Redes Bayesianas para a modelagem da jornada de compra dos clientes, visando facilitar sua compreensão e adoção pelas equipes de Marketing Digital, Vendas, Produto e Análise de Dados. Enquanto o foco de aplicação será colocado na jornada de compra de produtos SaaS, acredita-se que os benefícios possam se estender às equipes que comercializam outros tipos de produtos digitais.

2. Redes Bayesianas

As Redes Bayesianas foram propostas em 1988 por Judea Pearl [9] e são um caso particular dos Modelos Gráficos Probabilísticos. Oferecem uma forma poderosa de representar dependências e independências condicionais entre variáveis aleatórias [6]. Sua base teórica se apoia em grande parte no Teorema de Bayes e na Teoria de Grafos para construir modelos que, mesmo em sua configuração mais básica, trazem um bom nível de interpretação visual e uma capacidade preditiva que não se limita a uma variável de saída, mas que oferece predição para todas as variáveis do modelo, com base em sua distribuição de probabilidade conjunta. Essas características parecem trazer grande valor para a modelagem da jornada de compra de produtos SaaS, acima citada.

O elemento central da uma Rede Bayesianas é o seu Grafo Acíclico Direcionado, chamado de DAG (do inglês *Directed Acyclic Graph*). Ele é o grafo por meio do qual as variáveis e suas relações são representadas, através de nós e arcos, respectivamente [10].

Outro elemento básico da rede é sua distribuição de probabilidade conjunta, que engloba todas as variáveis do modelo. Essa probabilidade conjunta é fatorada pelas probabilidades condicionais locais, entre as variáveis que influenciam umas às outras.

A Figura 1 ilustra uma Rede Bayesianas simples, onde as variáveis A e B influenciam a variável C e esta influencia a variável D.

Logo, o processo de estimação de uma Rede Bayesianas consiste de duas etapas: 1) estimação da estrutura da DAG; 2) estimação dos parâmetros, na qual as distribuições de probabilidade condicional local são determinadas.

A estimação da estrutura de uma Rede Bayesianas é um processo computacional de complexidade não-polinomial. Trata-se de um problema exponencial, tam-

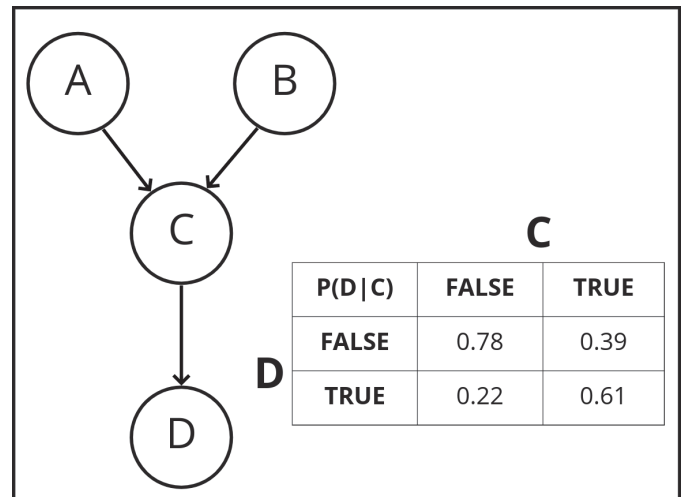


Figura 1: O grafo acima representa um DAG, enquanto a tabela representa a distribuição de probabilidade condicional entre as variáveis C e D. A Rede Bayesianas é composta pelo DAG e as tabelas de probabilidade locais de todas as variáveis que, juntas, fatoram a distribuição conjunta da rede.

bém chamado de *NP-hard* ou NP-Difícil. Isso porque a complexidade da estimação de parâmetros cresce exponencialmente, conforme crescem o número de variáveis e a quantidade de classes das mesas, considerando o caso da Rede Bayesianas discreta [10].

Ao longo das últimas décadas, diversos algoritmos foram criados para estimar a estrutura do DAG. Eles são agrupados em duas categorias: métodos baseados em restrição e métodos baseados em pontuação [7]. A primeira categoria consiste em fazer testes de independência condicional para decidir se um arco deve ou não ser adicionado à rede, ligando duas variáveis. A segunda categoria é composta por métodos que buscam por estruturas no espaço das redes possíveis, otimizando uma função de pontuação que calcula métricas de informação da rede.

Implementações dos algoritmos mais relevantes na literatura podem ser encontradas no pacote *bnlearn* [12] do *software* estatístico *R* [15]. Neste trabalho, serão utilizados dois dos principais métodos baseados em restrição: PC Stable e Grow Shrink (GS). Também são utilizados os métodos baseados em pontuação Hill Climbing e Tabu Search, em conjunto com quatro funções de pontuação: AIC, BIC, K2 e BDe [13].

Para a estimação dos parâmetros da probabilidade conjunta da rede, existem duas principais abordagens: o método de máxima verossimilhança e o método bayesiano [8]. No primeiro, a estimação é baseada em frequências das combinações de valores. No segundo, assume-se uma distribuição à priori para, então, calcular a dis-

tribuição à posteriori com os dados. Esses dois métodos também estão implementados no pacote *bnlearn*.

Uma Rede Bayesiana pode ser do tipo Discreta ou Contínua, a depender do tipo das variáveis utilizadas para estimar a rede. Existe ainda um terceiro tipo chamado híbrido, onde a rede acomoda tanto variáveis discretas quanto contínuas. Neste trabalho, será utilizada a abordagem discreta para estimação da rede.

Quando se opta por trabalhar com a Rede Bayesiana discreta, é comum, no processo de preparação dos dados, a discretização de algumas variáveis originalmente contínuas. Neste caso, é importante utilizar um método de discretização que não divida simplesmente a variável contínua em grupos de tamanhos iguais, mas que realize a discretização de maneira a preservar a informação mútua entre as variáveis do conjunto de dados. Um método apropriado é o *Hartemink* [11], que está implementado no pacote *bnlearn*. Ele não recebe como entrada apenas o vetor da variável a ser discretizada, mas um conjunto com todas as variáveis, para que o algoritmo consiga diminuir a perda de informação, que é intrínseca do processo de discretização.

3. Aplicação

E execução deste trabalho foi dividida em duas partes. A primeira foi a execução de um *benchmarking* entre os principais algoritmos de estimação de Redes Bayesianas implementados no pacote *bnlearn*, por meio de várias métricas aferidas em um processo de *holdout* repetido, onde os modelos foram ajustados usando a base de dados da jornada de compra dos clientes. Ainda nessa fase, foi escolhido o algoritmo com o qual treinamos o modelo final de Rede Bayesiana, que serviu de base para a próxima fase. Na segunda parte, foi feita uma aplicação prática do modelo escolhido ao problema de negócio em questão, demonstrando como uma Rede Bayesiana pode ser utilizada para extrair real valor dos dados e guiar decisões de negócio.

3.1. Ambiente computacional e Base de Dados

O ambiente computacional utilizado foi um computador com processador Intel i7 com quatro núcleos, com o sistema operacional Ubuntu, o ambiente de programação R na versão 4.2.0, e os seguintes pacotes: *bnlearn* para estimação e inferência de Redes Bayesianas, *tidyverse* para manipulação e visualização de dados, *caret* para obtenção de métricas de performance dos

modelos, *doSNOW* e *foreach*, ambos em conjunto para processamento paralelo.

A base de dados representa todos os potenciais clientes, também chamados de *leads*, de uma empresa de *software* durante um período de 3 meses. Ela contém 33 mil observações, das quais cerca de 1.000 se tornaram clientes no final do processo. Todas as variáveis são discretas, com distribuição binomial, indicando se o cliente realizou determinada ação, ou tem determinado atributo.

| Variável | Descrição |
|----------|--|
| ProEm | O cliente possui email profissional? |
| FSTm | Tempo da primeira sessão (*first time session*) |
| Invt | O cliente convidou alguém para conhecer o produto? |
| Feat1 | O cliente usou a funcionalidade 1? |
| Feat2 | O cliente usou a funcionalidade 2? |
| Rtrn | O cliente retornou ao produto em um outro dia, após o *sign-up*? |
| Cntc | O cliente atendeu ao contato do representante de vendas? |
| Demo | O cliente aceitou assistir a uma demonstração com um vendedor? |
| Buy | O cliente comprou o produto? |

Tabela 1: Descrição das variáveis que compõem a base de dados da jornada de compra dos clientes.

A única exceção é a variável *FSTm*, que diz quanto tempo o cliente passou na sua primeira sessão online no produto, logo após o *sign-up*. Esta variável era, inicialmente, contínua e foi discretizada utilizando o método *Hartemink*, implementado no próprio *bnlearn*.

A Tabela 1 detalha melhor as variáveis utilizadas, com seu nome e a explicação de seu significado. O nome está em formato resumido, o que é necessário para facilitar a visualização das variáveis na estrutura na rede.

Ainda sobre a base de dados, é importante destacar que foi aplicada uma técnica de *underbalancing* para atenuar o forte desbalanceamento, natural do negócio, onde apenas cerca de 3% dos *leads* realizam a compra no final do processo. Aplicou-se amostragem aleatória para diminuir a classe negativa, de forma a atingir um balanceamento de 15% x 85% ao invés os originais 3% x 97%, o que já permite uma boa estimação da rede.

3.2. Benchmarking dos métodos de estimação

Como mencionado anteriormente, a estimação de um modelo de Rede Bayesiana se divide em duas etapas: estimação da estrutura do grafo e estimação dos parâmetros das distribuições condicionais locais, que fatoram a distribuição conjunta global. A fase que gera maior dúvida no momento de estimar uma rede e, por isso, uma maior oportunidade de comparação entre os

métodos, é a estimação da estrutura do DAG, ou seja, a definição da estrutura de influência entre as variáveis.

Além das diversas possibilidades de algoritmos e de *scores*, existe ainda alguns recursos que podem ser utilizados para melhorar a estimação. Um exemplo é a utilização uma *blacklist* para informar o estimador sobre conexões que não podem existir entre variáveis ou uma *whitelist* pra informar conexões que devem existir.

Já o *bootstrapping* é um método estatístico de amostragem que está implementado no *bnlearn* e que auxilia na estimação da rede por meio de consecutivas amostragens do conjunto de dados, com reposição, ajustando uma rede para cada amostragem e posteriormente verificando quais os arcos que mais aparecem. Assim, é estabelecida uma ideia de força do arco, com base na sua prevalência no conjunto de amostras.

Buscando comparar as diferentes formas de estimar as redes, foi desenhado um experimento com quatro formas diferentes de realizar a estimação que, neste trabalho, serão chamadas de *setups* de estimação:

| Setup | Com Blacklist | Com Bootstrapping |
|----------------|---------------|-------------------|
| Simple | | |
| Bootstrapping | | X |
| Blacklist | X | |
| Boot_Blacklist | X | X |

Tabela 2: Os quatro diferentes setups de estimação utilizados na comparação dos algoritmos

Para cada um destes *setups*, foram testados 10 métodos de estimação: 8 deles vindos da combinação de 2 algoritmos baseados em *score* (Tabu e HC) com 4 *scores* (AIC, BIC, K2 e BDe) e mais 2 algoritmos baseados em restrição (PC Stable e Grow-Shrink).

Para aumentar a robustez do teste, foi utilizada a técnica de validação cruzada chamada *Holdout Repetido*, com 50 *holdouts*. Em cada um deles, este processo foi repetido: divisão aleatória dos dados em treino (70%) e teste (30%); treinamento da estrutura do DAG; treinamento dos parâmetros da probabilidade conjunta; predição na base de teste; e aferição das métricas de avaliação.

Desta forma, com 10 redes treinadas por *holdout*, sendo uma para cada algoritmo, somamos 500 redes pra cada um dos 4 *setups* citados acima. Isso totaliza 2.000 redes treinadas e avaliadas. As métricas aferidas destas 2.000 redes compõem os dados para nossa análise de comparação entre os métodos.

O fato de escolher fazer um *benchmarking* com tantas variações de métodos e, ainda, com *holdout* repetido elevou a complexidade computacional do trabalho. Como já mencionado, a complexidade do processo de estimação é um fator importante em Redes Bayesianas. Tornou-se importante, então, uma implementação que utilizasse computação paralela. Este paralelismo foi feito utilizando dois pacotes próprios para este propósito: *foreach* e *doSNOW*. Foram criados 4 *workers*, devido à presença de 4 núcleos no computador utilizado, sendo que o trabalho despachado para cada *worker* foi a completa execução de um *holdout*. Logo, foram 50 trabalhos distribuídos entre 4 *workers*, cuja orquestração foi desempenhada pelo pacote *foreach*, em conjunto com o *doSNOW*.

Quanto à estimação dos parâmetros da probabilidade conjunta, utilizou-se o método bayesiano, independente do algoritmo utilizado para ajuste da estrutura do DAG. Chegou-se a executar alguns testes de comparação entre o método bayesiano e o método de máxima verossimilhança mas os resultados foram muito próximos, devido ao tamanho da base de dados usada. Por isso, uma comparação formal dos resultados dos dois foi descartada, optando-se pelo método bayesiano.

Métricas de avaliação

Para o *benchmarking*, os diferentes algoritmos foram comparados utilizando *F1 Score*, que é a média harmônica entre as métricas *Recall* e *Precision*. Estas, respectivamente, verificam a capacidade do modelo classificador em identificar casos da classe positiva e o percentual de acerto (precisão) quando o modelo aponta observações como sendo da classe positiva.

Foi dada especial importância à habilidade do modelo em prever variáveis intermediárias na jornada de compra, sem a presença de outras variáveis que, normalmente, somente são obtidas depois delas no processo de compra. A motivação é simular uma aplicação bem realista para as necessidades do nosso caso de uso.

Por exemplo, dois dias após um cliente fazer o *signup*, já temos alguns dados sobre o seu comportamento inicial na plataforma, assim como dados demográficos do cliente. Ainda não sabemos se o cliente vai retornar ao produto, aceitar se conectar com os vendedores ou participar de uma ligação para demonstração do produto, podendo todas estas serem vistas como variáveis resposta intermediárias no processo. E queremos po-

der prever estes comportamentos, usando apenas os valores das variáveis que temos. Queremos, inclusive, prever a probabilidade do cliente comprar, mesmo estando no início do processo e com poucos dados disponíveis, para que possamos tomar ações de *marketing* personalizadas e, então, atualizar esta probabilidade, conforme o cliente avança na jornada de compra.

Portanto, durante os testes, avaliamos o poder preditivo dos modelos para estas variáveis intermediárias, escondendo os valores das variáveis que normalmente não estariam disponíveis, para que essas não ajudassem o modelo durante esta avaliação. Um exemplo é a variável *Demo*, que diz se o cliente aceitou fazer uma demonstração. Em um cenário real, antes do time de vendas contactar o cliente, não sabemos se o cliente vai aceitar fazer a demonstração (*Demo*). Também não sabemos, obviamente, se ele vai comprar. Mas se medíssemos, durante os testes, a capacidade do modelo em prever a variável *Demo*, sem remover a variável *Buy*, nossa rede nos traria ótimos resultados pois ela utilizaria o valor da variável *Buy* pra prever a variável *Demo*, mesmo que a última seja, na verdade, a que influencia a primeira. Esta é uma característica muito útil das Redes Bayesianas, para fazer análises diagnósticas, por exemplo, mas que do ponto de vista da utilidade para a operação de vendas estaria mascarando o real poder de cada modelo.

Por isso, selecionamos 4 variáveis que chamamos de intermediárias, sobre as quais gostaríamos de avaliar o poder preditivo: *Rtrn* (cliente retornou ao produto), *Cntc* (cliente aceitou ser contactado), *Demo* (cliente aceitou participar de uma demonstração) e *Buy* (cliente comprou o produto). Essa última, apesar de não ser intermediária no processo, é uma variável que queremos prever desde o início, ainda que sem todos os dados disponíveis, por isso foi testada juntamente com as outras três. Definimos também 3 variáveis como sendo as que deixaríamos indisponíveis para as previsões das intermediárias, por não estarem normalmente disponíveis ou demorarem pra estar disponíveis no processo de compras real: *Cntc*, *Demo* e *Buy*. Chamamos elas de variáveis *target* principais. Como se pode perceber, existe uma sobreposição entre o conjunto das chamadas intermediárias e as chamadas variáveis *target* principais. Isso é porque três delas, de fato, desempenham os dois papéis. Incluímos ainda uma última avaliação de previsão que foi da principal variável resposta, a *Buy*, sem excluir nenhuma variável, para podermos comparar o poder preditivo da Rede Bayesiana com

outros modelos discriminantes, que trabalham com apenas uma variável resposta.

Essa validação mais realista trouxe uma complexidade maior para o processo de avaliação mas também tornou ele mais confiável, especialmente para o contexto em que estamos aplicando. Nosso processo de avaliação do poder preditivo, para cada rede gerada, seguiu esta sequência: 1) escondemos as três variáveis *target* principais; 2) avaliamos a previsão para as quatro variáveis intermediárias; 3) avaliamos a previsão para a principal variável *target*, agora na presença de todas as outras variáveis, com exceção apenas dela mesma. Ou seja, são 5 avaliações de poder preditivo, onde medimos o *Recall*, *Precision* e *F1* para cada uma, totalizando 15 métricas de previsão para cada modelo gerado.

Uma outra avaliação preditiva consistiu em ajustar modelos com métodos de aprendizagem de máquina tradicionais, afim de fazer uma comparação com os resultados obtidos pelos modelos de Redes Bayesianas. Foram ajustados modelos de Regressão Logística e Random Forest. Assim como nos testes com Redes Bayesianas, foram feitos 50 *holdouts* repetidos com cada um destes métodos.

Como visto anteriormente, os testes de previsão para as Redes Bayesianas consideraram também a previsão para várias variáveis *target*, explorando a característica do método em fornecer a possibilidade de previsão para qualquer variável do mesmo modelo ajustado. Contudo, para esta comparação com outros métodos de aprendizagem de máquina, foram considerados os resultados de Redes Bayesianas apenas no modo discriminante, ou seja, previsões para a variável *target* principal, utilizando todas as outras como entrada.

3.3. Escolha do modelo final

Ainda na fase de avaliação dos modelos gerados, foi feita uma análise visual dos seus grafos, afim de escolher o melhor modelo para prosseguir para a fase de inferência. Essa análise visual é mais subjetiva, porém muito importante pois, como visto, alguns modelos podem possuir um bom poder preditivo mesmo estabelecendo relações de influência que não fazem muito sentido para o negócio.

Um outro ponto que foi considerado foi a quantidade de arcos orientados. Os modelos que foram gerados com o *score* AIC tiveram o melhor desempenho em previsão, porém, orientaram muitos arcos, o que deixa o grafo mais complexo e difícil de ser compreendido.

dido, limitando uma característica muito importante em Redes Bayesianas, que é a possibilidade de perceber visualmente as relações de influência entre as variáveis.

Dessa forma, após analisar os diferentes grafos gerados, optou-se por prosseguir pelo modelo gerado pelo algoritmo Tabu, utilizando o *score* K2 com *blacklist*. Esse modelo obteve uma excelente performance preditiva, ao mesmo tempo que ficou menos complexo que os modelos gerados com AIC. Ele orientou 22 arcos, contra uma média de 30 arcos para o AIC. Mais detalhes sobre a performance dos diferentes algoritmos podem ser vistos na seção de resultados.

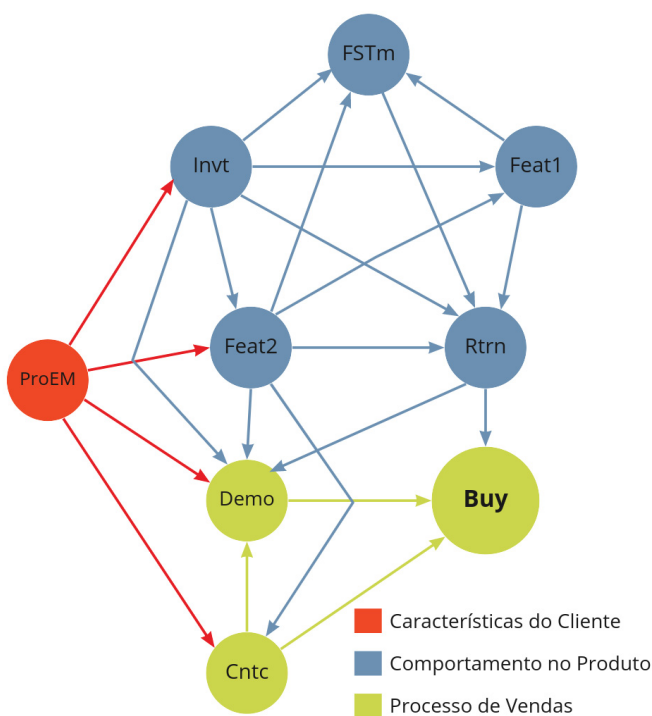


Figura 2: DAG do modelo escolhido para inferências: algoritmo Tabu, com *score* K2 e utilização de *blacklist*, sem *bootstrapping*.

3.4. Inferências e aplicações no negócio

O grande objetivo em ajustar uma Rede Bayesiana de jornada de compras dos clientes é extrair informações úteis para tomada de decisão. Nesta seção, demonstramos um pouco do que foi possível obter com o modelo escolhido para esse fim.

O primeiro benefício ao negócio surge da análise visual do grafo. Essa é uma característica contemplativa e de comunicação que as Rede Bayesianas oferecem. Pessoas de diferentes áreas podem compreender

o grafo, extrair informação e se comunicar através de: times de análise de dados, vendas, *marketing*, produto e a gerência.

Olhando para a Figura 2, podemos perceber alguns pontos da jornada de compra que são mais importantes. Em primeiro lugar, o fato do cliente ter um email profissional (*ProEm*) é de grande importância, pois influencia tanto o seu comportamento no produto quanto sua disponibilidade para conversar e participar de uma demonstração.

Convidar outra pessoa para conhecer o produto é o passo importante. Ele tem influência sobre outras 5 variáveis, inclusive sobre as chances do cliente retornar ao produto em um outro dia. Isso se explica no fato de que o produto em questão é um *software* empresarial, ou seja, o processo de compra não é algo que se faz normalmente sozinho. É mais comum que exista um grupo de pessoas dentro do cliente que estejam analisando possíveis opções de *software* no mercado. E quando uma pessoa cria uma conta no produto e convida outras para sua conta, demonstra que existe um interesse real em prosseguir com a avaliação para compra. Nota-se que apenas o email profissional influencia em convidar alguém, ou seja, pessoas que entram com emails que não são corporativos, tendem a convidar menos. Isso mostra que, talvez, seja interessante para a empresa investir em uma abordagem diferenciada no momento de incentivar os usuários a convidar outras pessoas. Pode ser interessante dar um incentivo maior para usuários que entram com um email corporativo.

As variáveis *Feat1*, *Feat2* e *FSTm* dizem respeito ao comportamento do cliente no produto. Algumas exercem influência sobre as outras e todas influenciam a probabilidade do cliente retornar ao produto um outro dia. Destaque para a *Feat2*, que influencia as outras duas. Isso sugere que a empresa foque mais em influenciar o usuário a experimentar essa funcionalidade, já que ela conduz naturalmente a outras.

O próximo passo crucial é que o cliente retorne ao produto (*Rtrn*). Essa é a primeira variável que tem influência direta sobre a decisão de comprar, antes mesmo que o cliente seja contactado pelo time de vendas. Isso mostra que é extremamente importante que o cliente seja motivado a retornar ao produto, tanto por meio de utilização das principais funcionalidades quanto por incentivos externos ao produto, como *email marketing*, entre outros.

A variável *Demo* é fundamental no processo de vendas da empresa. Mais do que influenciar a decisão de compra, ela é, praticamente, o único caminho pelo

qual as vendas acontecem no processo atual. Por isso, é interessante observar as variáveis que a influenciam diretamente, ou seja, que levam o cliente a aceitar participar de uma demonstração. São elas *ProEm*, *Feat2*, *Rtrn*, *Invt* e *Cntc*. Não se tem controle sobre a variável *ProEm* (email profissional/corporativo) e a variável *Cntc* (cliente respondeu ao contato) é básica para que se agende uma demonstração. Assim, para aumentar o número de clientes que agendam uma demonstração, é importante que a empresa busque meios de influenciar as outras três. Convidar pessoas para conhecer o produto (*Invt*), utilizar a funcionalidade 2 (*Feat2*) e retornar ao produto em um outro dia são, assim, as variáveis mais importantes sobre as quais a empresa pode buscar exercer influência durante a jornada de compra.

As observações de negócio feitas acima são apenas alguns exemplos dos tipos de conclusões e tomadas de decisão que uma empresa pode fazer em relação ao seu processo de vendas, ao lançar mão de uma abordagem com Redes Bayesianas para tal tarefa. Muitas outras análises e conclusões poderiam ser feitas, inclusive olhando para as tabelas de probabilidade condicional locais para, por exemplo, entender o peso das influências entre as variáveis, em termos de probabilidade. Toda essa análise se dá nos níveis estratégico e tático do negócio.

Do ponto de vista operacional, as Redes Bayesianas tem um grande potencial para tomada de decisão dos representantes de vendas. Como mencionado anteriormente, existe um grande desafio em um processo de vendas *Business to Business* (B2B) que é decidir onde a equipe de vendas deve colocar maior esforço. Uma equipe que não investe seu tempo nos *leads* que tem maior chance de serem convertidos, tende a aumentar drasticamente seu custo de aquisição de clientes.

Para validar e demonstrar este caso de uso das redes, traçamos uma jornada de compra de um cliente fictício, e fizemos consultas no modelo ajustado, para estimar como a probabilidade de compra se comporta conforme o cliente vai avançando na sua jornada. Esse exercício demonstra bem uma das principais vantagens das Redes Bayesianas que está relacionado ao princípio bayesiano de atualização do conhecimento com base em novas evidências. Começamos com apenas um dado sobre o cliente e já estimamos sua probabilidade de compra. À medida que o cliente toma decisões, seu comportamento nos fornece novos dados, que são então utilizados para atualizar o conhecimento que se tem sobre a probabilidade de compra.

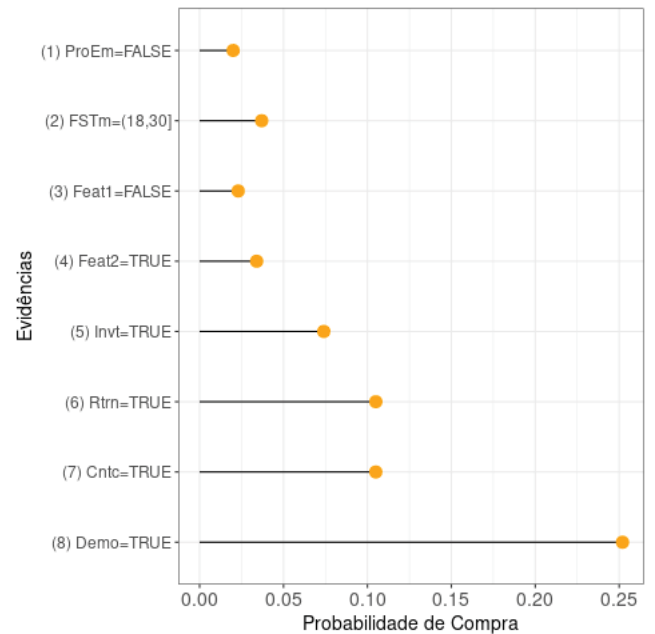


Figura 3: Probabilidade de compra do cliente sendo atualizada conforme ele avança na jornada de compra.

Na Figura 3, vemos essa evolução da probabilidade para cada um dos passos do cliente. No primeiro passo, o cliente acabou de fazer o *sign-up* e o único dado que temos é de que ele não usou um email corporativo, portanto, a evidência é codificada como *ProEM=FALSE*. Apesar de só termos esse dado, já estimamos a primeira probabilidade de compra pra ele. Ela é de 2%, ou seja, menor do que a prevalência da população, que é de 3%. Um cliente que tivesse email profissional teria uma probabilidade de 4% neste mesmo ponto da jornada.

O cliente, então, encerra sua primeira sessão no produto e agora temos mais dados para serem considerados como evidências. Ele passou 25 minutos navegando no produto, o que é considerado um tempo bom, estando no grupo entre 18 e 30 minutos, quando discretizamos a variável. Adicionando esta evidência à jornada do nosso cliente e consultando novamente o modelo, temos uma nova probabilidade, agora de 4%.

Quanto ao uso de funcionalidades, o cliente não interagiu com a funcionalidade 1, o que diminuiu sua probabilidade para 2%, mas interagiu com a funcionalidade 3, o que a eleva a 3,5%.

Antes de encerrar sua primeira sessão, no passo 5, o cliente decide convidar um colega para conhecer a conta que ele criou no produto. Isso mais que dobra sua probabilidade de compra, levando a 7,5%. Neste momento, a probabilidade do cliente é mais que duas vezes a prevalência de compra na população, que é de 3%.

Nesse momento da jornada, é importante que o cliente retorne ao produto um outro dia. Mas qual a probabilidade disso acontecer? Essa é uma das conveniências das Redes Bayesianas, que é a já mencionada possibilidade de fazer previsão para qualquer variável do modelo. Consultando a rede com as evidências que temos até o momento, temos que essa probabilidade é de 68%, uma probabilidade alta. Isso, principalmente, porque ele convidou alguém para conhecer o produto, ficou um bom tempo na primeira sessão e experimentou a funcionalidade 2. A empresa pode, a qualquer momento, utilizar a probabilidade do cliente retornar ao produto para definir campanhas de *marketing* personalizadas, seja por *email marketing* ou outros canais.

Passados dois dias, o cliente retorna ao produto, como vemos no passo 6 do gráfico. Isso traz mais um aumento importante na probabilidade de compra, atingindo 10%, o que é mais de 3 vezes a prevalência da população de *leads*.

A equipe de vendas passa, então a buscar fazer contato com este cliente. Neste momento é possível também consultar qual a probabilidade do cliente responder a este contato. É ela é de 37%, um valor razoável, que só não é maior porque o cliente não utilizou um email profissional no cadastro. Clientes com email profissional tem uma chance maior de responderem ao contato do time de vendas. Mas como a prob deste cliente é considerável, a equipe permanece fazendo tentativas de contato até que, finalmente, consegue. Aqui, o modelo de Redes Bayesianas é muito útil para o time decidir quando investir ou não investir tempo buscando fazer contato com um possível cliente.

Uma vez que o cliente aceitou agendar a demonstração, sua probabilidade de compra se elevou muito, como vemos no passo 8 do gráfico. Agora, o representante de vendas que vai atender este cliente vai investir bem o seu tempo em um cliente que realmente tem uma chance considerável de finalizar a compra, com uma probabilidade 8 vezes maior que a prevalência de compra da população inicial de *leads*. E o vendedor sabe disso e sabe também o porquê dessa probabilidade, pois ele tem acesso a essas evidências do modelo e sabe quais delas mais influenciaram a probabilidade de compra. Isso otimiza muito o tempo investido pela equipe de vendas e caracteriza um importante ganho na capacidade de negociação.

4. Resultados e Discussão

O *benchmarking* realizado entre os diferentes algoritmos e *setups* de estimação demonstrou que pode haver uma diferença significativa de performance preditiva entre os mesmos. Entre os algoritmos baseados em pontuação (*score based*), as diferenças não se dão entre os algoritmos em si, mas sim, entre os diferentes *scores* (AIC, BIC, K2, BDE).

Percebeu-se que o F1 Score para a variável *target* principal ficou muito acima do F1 para as intermediárias, atingindo resultados por volta de 0,8 para a primeira contra 0,59 para as últimas. Isso já era esperado, já que as previsões para as variáveis intermediárias não contavam com todas as variáveis do conjunto de dados, enquanto a variável *target* principal contava com todas menos, é claro, ela mesma. Essa diferença nos levou a comparar a previsão dos modelos de duas formas: o F1 para a *target* principal e a média de F1 para as 4 variáveis *target* intermediárias. Como havíamos feito 50 *holdouts* repetidos para cada método de estimação, usamos a mediana desses 50 resultados para o F1 de cada variável, antes de calcular a média de F1 das intermediárias e também para obter o F1 Score da *target* principal.

A tabela 3 mostra os primeiros 10 colocados no *ranqueamento* das previsões para a *target* principal, que possui ao todo 40 combinações de métodos e *setups*.

| Posição | Algoritmo | Setup | F1 Score |
|---------|-----------|----------------|----------|
| 1 | hc_aic | blacklist | 0,804 |
| 2 | tabu_aic | blacklist | 0,804 |
| 3 | hc_bic | boot_blacklist | 0,796 |
| 4 | tabu_bic | boot_blacklist | 0,796 |
| 5 | tabu_bde | boot_blacklist | 0,796 |
| 6 | tabu_k2 | boot_blacklist | 0,796 |
| 7 | hc_k2 | boot_blacklist | 0,796 |
| 8 | hc_bde | boot_blacklist | 0,796 |
| 9 | hc_aic | boot_blacklist | 0,795 |
| 10 | tabu_aic | boot_blacklist | 0,795 |

Tabela 3: Ranqueamento dos métodos quanto a previsão da variável *target* principal, utilizando F1 score.

A tabela 4 mostra os primeiros 10 colocados no *ranqueamento* das previsões para as *target* intermediárias, que também possui 40 combinações de métodos e *setups*.

Os melhores resultados ficaram com *hc_aic* e *tabu_aic* em ambas as métricas avaliadas. Para a *target* principal, atingiram 0,804 de F1 utilizando apenas a *blacklist*,

| Posição | Algoritmo | Setup | Média F1 Score |
|---------|-----------|----------------|----------------|
| 1 | hc_aic | blacklist | 0,592 |
| 2 | tabu_aic | blacklist | 0,592 |
| 3 | tabu_k2 | blacklist | 0,591 |
| 4 | hc_k2 | blacklist | 0,591 |
| 5 | hc_aic | simple | 0,590 |
| 6 | tabu_aic | simple | 0,590 |
| 7 | gs | boot_blacklist | 0,590 |
| 8 | hc_k2 | boot | 0,590 |
| 9 | tabu_k2 | boot | 0,588 |
| 10 | tabu_aic | boot | 0,588 |

Tabela 4: Ranqueamento dos métodos quanto a predição da variável das variáveis target intermediárias, utilizando a média de seus F1 scores.

sem *bootstrapping*, seguidas por todos os outros métodos baseados em *score* que ficaram empatados quase 1 ponto percentual atrás, estes utilizando *bootstrapping* e *blacklist*. Para as *target* intermediárias, *hc_aic* e *tabu_aic* atingiram 0,592 na média de F1 score, seguidos e praticamente empatados com *tabu_k2* e *hc_k2*, todos os quatro utilizando *blacklist*, sem *bootstrapping*.

O aspecto do *setup* de estimação também se mostrou relevante, sendo mais ou menos importante, a depender do algoritmos e do *score*. O uso de *blacklist* mostrou-se um recurso extremamente importante para se conseguir bons resultados preditivos, tanto para as variáveis intermediárias quanto para a variável resposta principal.

Por outro lado, o *bootstrapping* não apresentou vantagens, de forma geral, com exceção dos algoritmos baseados em restrição. A estimação com *bootstrapping* pode ser muito útil para casos com poucos dados, pois esta é uma das maiores contribuições da técnica de *bootstrapping* em geral, e também para casos com muitos dados, pois nestes é possível tirar uma amostra pequena dos dados e utilizar *bootstrapping* pra diminuir a complexidade da estimação sem perder variabilidade. Porém, aparentemente, o processo de *averaged network*, utilizado na estimação com *bootstrapping*, adiciona mais arcos ao grafo, deixando o modelo mais complexo e menos eficiente em termos de predição, quando comparado a modelos estimados apenas com a *blacklist*. Portanto, com base nos resultados obtidos neste estudo, pra este dataset específico, não se recomenda a utilização de *bootstrapping* com *averaged network*, mas sim o uso isolado de *blacklist*.

Em uma comparação com modelos mais tradicionais de aprendizagem de máquina, os resultados mostram a mediana de *F1 Score* das Redes Bayesianas levemente superior aos outros dois métodos, tendo 0,794 contra 0,786 e 0,782 de Random Forest e Regressão Logística, respectivamente. Contudo, é importante destacar dois pontos aqui. O primeiro, em favor dos modelos tradicionais, é que os mesmos foram ajustados no seu formato mais simples, sem otimização de hiperparâmetros e redução de colinearidade entre as variáveis de entrada, o que é um fator relevante, especialmente para a Regressão Logística. O segundo ponto, em favor das Redes Bayesianas, é de que elas são mais do que modelos discriminantes, podendo atuar na predição de qualquer variável do mesmo modelo ajustado e podendo servir para análise do problema de negócio de uma maneira mais visual e intuitiva, por pessoas de todas as áreas.

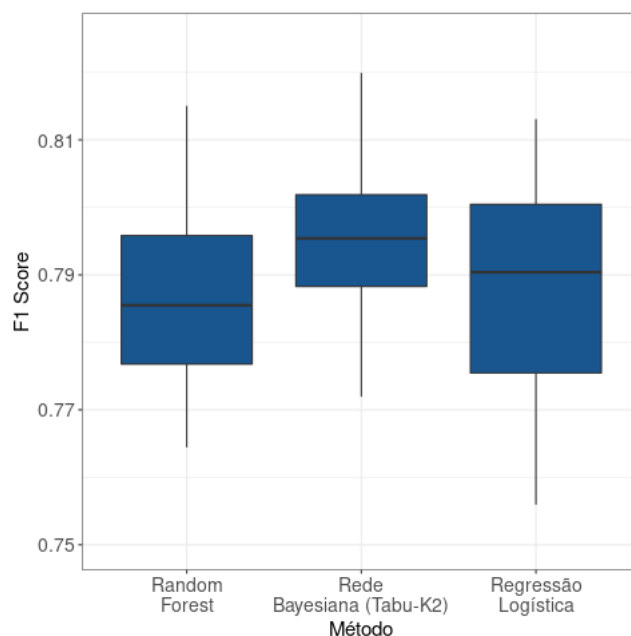


Figura 4: Comparação da performance de predição entre Redes Bayesianas e dois métodos tradicionais de aprendizagem de máquina.

Em um experimento mais prático, o modelo de Rede Bayesianiana demonstrou essa flexibilidade, servindo como uma ferramenta poderosa para tomada de decisões operacionais, em tempo real, e de forma personalizada para cada cliente. O modelo foi utilizado para processar consultas, com base nas evidências da jornada de um cliente fictício, e retornar probabilidades a cerca de quatro variáveis de interesse.

A tabela 5 demonstra com mais detalhes a quantidade de informações que podemos extrair a cada passo

| Passo | Nova Evidência | Prob. Rtrn | Prob. Cntc | Prob. Demo | Prob. Buy |
|-------|----------------|------------|------------|------------|-----------|
| 1 | ProEm=FALSE | 0,28 | 0,34 | 0,06 | 0,02 |
| 2 | FSTm=(18,30] | 0,50 | 0,36 | 0,10 | 0,04 |
| 3 | Feat1=FALSE | 0,32 | 0,35 | 0,07 | 0,02 |
| 4 | Feat2=TRUE | 0,41 | 0,38 | 0,10 | 0,03 |
| 5 | Invt=TRUE | 0,69 | 0,38 | 0,20 | 0,07 |
| 6 | Rtrn=TRUE | 1,00 | 0,38 | 0,26 | 0,10 |
| 7 | Cntc=TRUE | 1,00 | 1,00 | 0,36 | 0,11 |
| 8 | Demo=TRUE | 1,00 | 1,00 | 1,00 | 0,25 |

Tabela 5: Probabilidades para todas as variáveis de interesse, atualizadas à medida que novas evidências são adicionadas, a cada passo do cliente na jornada de compra.

do cliente na jornada, onde pode-se fazer previsões para todas as variáveis de interesse do processo, para customizar ao máximo as campanhas para cada cliente, conforme novos dados são adquiridos. Por exemplo, no passo 4, é possível perceber que, ao utilizar a funcionalidade 2, o cliente não apenas aumentou sua probabilidade de compra como já mencionado, mas aumentou significativamente a probabilidade de retornar ao produto (de 32% para 41%) e também a probabilidade de participar de uma demonstração (de 7% para 10%). Já no passo 5, ao convidar um colega para conhecer o produto, aumentou elevou em mais de 50% a probabilidade de retornar ao produto, e dobrou a probabilidade de assistir à demonstração.

As probabilidades sobre essas variáveis de interesse, chamadas de intermediárias durante o processo de comparação dos modelos, são muito importantes pois são informações mais acionáveis para a equipe de vendas do que apenas a *target* principal, impactando diretamente nas decisões que pode-se tomar para cada potencial cliente. É possível, por exemplo, trabalhar sempre com um ranqueamento dos clientes que tem maior probabilidade de avançar para o próximo passo, sendo que esta lista pode ser atualizada frequentemente, conforme novos dados são adquiridos sobre o comportamento dos futuros clientes, em tempo real, ou próximo ao tempo real.

Poderiam, ainda, ser incluídas na estimação da rede variáveis representassem ações de *marketing* da própria empresa para poder avaliar qual a melhor ação para um cliente, considerando o seu estágio atual. Por exemplo, poderia haver uma variável para cada estilo de campanha de email, ou uma variável multinomial que assume o valor de um desses estilos de campanha. Uma consulta ao modelo mostraria qual o tipo de email que mais aumenta a probabilidade de cada cliente seguir para o próximo passo da jornada. Esse exercício não foi realizado neste trabalho por não termos dados de campanhas de *marketing* disponíveis,

mas é algo que é tecnicamente possível quando se dispõe desse tipo de dado.

5. Comentários Finais

Este trabalho demonstrou a aplicação do método de Redes Bayesianas no processo de vendas de uma empresa de *software* empresarial na nuvem. Diferentes algoritmos foram testados e seus resultados comparados, mostrando que é necessário validar vários deles com cada conjunto de dados, para decidir qual aplicar no modelo final.

Uma demonstração prática provou os benefícios que esta abordagem traz ao negócio, superando métodos tradicionais na explicabilidade, comunicação com o negócio, previsão para várias variáveis e, inclusive, funcionando como modelo discriminante, ou seja, voltado para previsões em uma única variável. Essas vantagens possibilitam aplicações que não são facilmente atingidas com outros modelos, como a atualização das previsões de forma gradativa, à medida que se obtém novos dados dos *leads*, se adaptando a realidade do negócio e personalizando a tomada de decisão para cada cliente.

Alguns pontos que ainda contam a favor dos métodos tradicionais são sua popularidade e maior facilidade de uso. Percebe-se que a teoria desses métodos é mais simples ou melhor abstraída por um número grande de pacotes de *software*, já consolidados entre os profissionais de análise de dados. Já as Redes Bayesianas, por serem uma abordagem mais recente, ainda são pouco difundidas na área de negócios, contam com um número bem menor de pacotes e ferramentas e apresentam um desafio maior de compreensão dessas ferramentas e da teoria que elas implementam, principalmente porque seus métodos ainda estão em crescente evolução.

Neste cenário, espera-se que este trabalho contribua para difundir as ideias e as possibilidades trazidas pelas Redes Bayesianas, especialmente no meio de negócios, estimulando sua aplicação e pesquisa nesta área.

Existem algumas possibilidades a serem consideradas para trabalhos futuros. Criar uma aplicação gráfica que facilite estimação e exploração de Redes Bayesianas, com um foco em jornada de compra, para que se possa utilizar uma linguagem compreensível pelos profissionais da área e, assim, eliminar barreiras técnicas para a utilização do método. Além disso, uma outra possibilidade é um estudo com mais bases de dados,

pra que se possa buscar validar se alguns dos padrões observados entre os diferentes algoritmos e *scores* de estimação nesse trabalho se repetem em outras bases e poderiam, assim, ser utilizados para recomendar algoritmos e *scores* que se ajustam melhor a esse tipo de dados.

Agradecimentos

Agradeço a Deus, por ter me dado a vida e por tornar todas as coisas possíveis. Aos meus pais, por terem me motivado e me guiado desde cedo. À minha esposa, por sempre me dar todo o suporte e encorajamento durante os momentos difíceis. Por fim, agradeço aos professores Anderson Ara e Lilia Costa pelas inúmeras horas de dedicação e apoio à produção deste trabalho, o qual não seria possível sem suas orientações.

Referências

- [1] E. Anderl; I. Becker; F. Wangenheim, J. H. Schumann, *Mapping the customer journey: Lessons learned from graph-based online attribution modeling*, International Journal of Research in Marketing, Volume 33, Issue 3, 2016, Pages 457-474. DOI: <https://doi.org/10.1016/j.ijresmar.2016.03.001>.
- [2] S. Hosseini, *A decision support system based on machine learned Bayesian network for predicting successful direct sales marketing*, Journal of Management Analytics, 8:2, (2021), 295-315, DOI: 10.1080/23270012.2021.1897956
- [3] Yelland, Phillip; Kim, Shinji; Stratulate, Renée, *A Bayesian Model for Sales Forecasting at Sun Microsystems*. Interfaces. (2010) 40. 118-129. 10.1287/inte.1090.0477.
- [4] Maleeha Qazi, Glenn M. Fung, Katie J. Meissner, and Eduardo R. Fontes. *An Insurance Recommendation System Using Bayesian Networks*. In Proceedings of the Eleventh ACM Conference on Recommender Systems (RecSys '17). Association for Computing Machinery, New York, NY, USA, (2017) 274–278. <https://doi.org/10.1145/3109859.3109907>
- [5] N. Kulkarni, *Demystifying the Consumer Decision Journey*, 2016 BayesiaLab Conference in Nashville. Disponível em: <https://library.bayesia.com/articles/bayesialab-knowledge-hub/2016-conference-neeraj-kulkarni/>. Acesso em 22 de junho de 2022
- [6] P. Ianashi; O. A. Gonzatto Junior; M. J. Henriques; D. C. Nascimento; G. K. Mattar; P. L. Ramos; A. Ara; F. Louzada, *Probability on Graphical Structure: A Knowledge-Based Agricultural Case*. ANNALS OF DATA SCIENCE, v. 5, p. 1, 2020.
- [7] G. Briganti; M. Scutari *An Introduction to Bayesian Artificial Intelligence in Medicine* Unit of Epidemiology, Biostatistics and Clinical Research, Université Libre de Bruxelles, 2020
- [8] T. Koski and J. M. Noble, *Bayesian Networks, An Introduction*, (Wiley, Chichester, 2009), First edition, pag. 158
- [9] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*, (Morgan Kaufmann, San Mateo, CA, 1988)
- [10] M. Scutari; J. B. Denis, *Bayesian Networks With Examples in R*, (Chapman and Hall, New York, 2015), First edition, pag. 15.
- [11] A. J. Hartemink, *Principled computational methods for the validation discovery of genetic regulatory networks*, PhD Thesis (Massachusetts Institute of Technology, Dept. of Electrical Engineering and Computer Science, 2001), pag 91
- [12] M. Scutari, *Learning Bayesian Networks with the bnlearn R Package*, Journal of Statistical Software. 35, 3 (Jul. 2010), 1–22. DOI:<https://doi.org/10.18637/jss.v035.i03>.
- [13] C. Su, A. Andrew, Karagas, M.R. et al. *Using Bayesian networks to discover relations between genes, environment, and disease*. BioData Mining 6, 6 (2013). <https://doi.org/10.1186/1756-0381-6-6>
- [14] *Gartner Forecasts Worldwide Public Cloud End-User Spending to Reach Nearly 500 Billion in 2022*. Gartner, 19 de abril de 2022. Disponível em: <https://www.gartner.com/en/newsroom/press-releases/2022-04-19-gartner-forecasts-worldwide-public-cloud-end-user-spending-to-reach-nearly-500-billion-in-2022>>. Acesso em: 22 de junho de 2022
- [15] R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>