UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Redes Bayesianas aplicada à predição de vendas em uma grande rede de fast-food brasileira

Robson Fernandes da Silva

Dissertação de Mestrado do Programa de Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria (MECAI)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Robson Fernandes da Silva

Redes Bayesianas aplicada à predição de vendas em uma grande rede de fast-food brasileira

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestre – Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria.
EXEMPLAR DE DEFESA

Área de Concentração: Matemática, Estatística e Computação

Orientador: Prof. Dr. Alneu de Andrade Lopes

USP – São Carlos
Janeiro de 2019

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

F363r Fernandes, Robson
 Redes Bayesianas aplicada à predição de vendas em
 uma grande rede de fast-food brasileira / Robson
 Fernandes; orientador Alneu Lopes. -- São Carlos,
 2019.
 119 p.

 Dissertação (Mestrado - Programa de Pós-Graduação
 em Mestrado Profissional em Matemática, Estatística
 e Computação Aplicadas à Indústria) -- Instituto de
 Ciências Matemáticas e de Computação, Universidade
 de São Paulo, 2019.

 1. Redes Bayesianas. 2. Relações Causais. 3.
 Séries Temporais. 4. Previsão de Vendas. 5. Fast-
 Food. I. Lopes, Alneu, orient. II. Título.

Robson Fernandes da Silva

**Bayesian Networks applied to the prediction of sales in a
large Brazilian fast food chain**

Master dissertation submitted to the Institute of
Mathematics and Computer Sciences – ICMC-USP,
in partial fulfillment of the requirements for the
degree of the Master – Professional Masters in
Mathematics, Statistics and Computing Applied to
Industry. *EXAMINATION BOARD PRESENTATION
COPY*

Concentration Area: Mathematics, Statistics and
Computing

Advisor: Prof. Dr. Alneu de Andrade Lopes

USP – São Carlos
January 2019

Este trabalho é dedicado primeiramente ao grande arquiteto do universo, pela confiança, equilíbrio e vitórias depositadas ao longo de minha carreira. Agradeço a oportunidade de poder escrever este trabalho com grandes professores, cientistas e profissionais, com o apoio do orientador Prof. Dr. Alneu de Andrade Lopes.

Agradeço a todos os professores, pesquisadores e profissionais do Instituto de Ciências Matemáticas e de Computação (ICMC) da Universidade de São Paulo (USP) pelo apoio e conhecimento compartilhado ao longo do meu período de mestrado.

A minha mãe Fátima Saraiva pelo esforço constante para promover uma boa estrutura emocional e educacional durante toda minha vida. Ao meu pai Altamir Fernandes que embora não esteja neste plano, sei que tem olhado por mim durante todo este trajeto. Ao meu irmão gêmeo David Fernandes pela ajuda e apoio em todos esses anos juntos, sempre me motivando a crescer cada vez mais. As minhas irmãs Paula Fernandes e Cléia Fernandes e meus cunhados Lauro Castilho e Helder Ribeiro que tem um carinho enorme por mim, sempre estendendo as mãos em tudo que precisei. Ao meu sobrinho querido Philipe Fernandes Castilho tão jovem e motivado. Aos meus amigos no âmbito profissional pelo qual tenho um carinho imenso, sempre me apoiando e colaborando nas minhas decisões profissionais e acadêmicas. Aos meus amigos e companheiros de infância e ao meu companheiro Rodrigo de Abreu pelo apoio ao longo de minha vida e carreira. A todos os meus amigos que, direta e indiretamente, me incentivaram no decorrer deste trabalho. Agradeço ao grande arquiteto do universo pela convivência maravilhosa e aprendizado que tenho com todos eles.

Dedico esta obra à minha mãe Fátima Saraiva por todo o apoio ao longo destes anos. Mesmo enfrentando oceanos de batalhas constantes, tendo que atuar como protagonista no papel de mãe e pai para nos fornecer uma estrutura familiar digna, sempre saiu vitoriosa, agindo com honestidade e humildade com todos nós.

Obrigado.

AGRADECIMENTOS

Ao professor Dr. Alneu de Andrade Lopes, pela orientação e oportunidade de crescimento acadêmico.

Ao Dr. Brett Drury pelo apoio e oportunidades de me envolver em atividades científicas nacionais e internacionais.

À empresa Linx, filial Bauru, pelo incentivo e compreensão durante o desenvolvimento deste trabalho.

À minha família pelo incentivo e pelas lições de vida.

*“Eu sei o preço do sucesso: dedicação, trabalho duro,
e uma incessante devoção às coisas que você quer ver acontecer.”
(Frank Lloyd Wright)*

RESUMO

FERNANDES. R. **Redes Bayesianas aplicada à predição de vendas em uma grande rede de fast-food brasileira**. 2019. 119 p. Dissertação (Mestrado – Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2019.

O segmento de *fast-food* tornou-se um mercado muito concorrido e com empresas bem conhecidas, tais como: Subway, McDonald's, Burguer King, Bob's e Habib's. Técnicas de inteligência artificial e ciência de dados podem oferecer inúmeros benefícios para este mercado, como por exemplo, permitir o desenvolvimento de modelos computacionais para tomada de decisões. No contexto de finanças onde envolvam a comercialização de determinados produtos, é muito comum deparar-se com cenários que envolvam incerteza, principalmente quando se deseja realizar projeções financeiras, avaliar riscos e estimativas. O objetivo deste trabalho consiste em desenvolver modelos probabilísticos baseados em Redes Bayesianas (RB) para realizar predições em vendas e análise de causalidade entre variáveis que influenciam no processo de comercialização de determinados grupos de produtos no seguimento de *fast-food*. Nesta análise foram avaliadas Redes Bayesianas com aprendizado de estrutura baseado em restrições, através do algoritmo *Grow Shrink* (GS), e Redes Bayesianas com aprendizado de estrutura baseado em pontuação, através do algoritmo *Hill-Climbing* (HC), posteriormente foram comparadas com um modelo de série temporal baseado em *Generalized Additive Model* (GAM). Os dados para análise foram adquiridos de uma rede de *fast-food* brasileira que possui cerca de 1100 lojas associadas, destas, foram utilizadas lojas que pertencem ao estado de São Paulo, assim como avaliado variáveis de grupos de vendas no período de 2010 à 2017. Os resultados foram avaliados através da métrica *Mean Absolute Percentage Error* (MAPE), que considera valores reais alimentados em modelos e valores ajustados a partir do modelo e calcula a diferença absoluta entre os dois como porcentagem do valor real, com base neste cálculo é possível obter a acurácia de cada modelo. A Rede Bayesiana (RB) com aprendizagem de estrutura baseada em pontuação, utilizando o algoritmo *Hill Climbing* (HC), foi escolhida como o melhor modelo, pois apresentou relações causais mais coerentes entre os vértices que influenciam o processo de venda, bem como combinações de vértices que resultam em combos de produtos, além disso, resultou em 97.60% de acurácia na previsão de vendas das lojas do estado de São Paulo (SP) na amostra de teste avaliada, com base na métrica *Mean Absolute Percentage Error* (MAPE).

Palavras-chave: Redes Bayesianas; Relações Causais; Séries Temporais; Previsão de Vendas; *Fast-Food*.

ABSTRACT

FERNANDES. R. **Bayesian Networks applied to the prediction of sales in a large Brazilian fast food chain**. 2019. 119 p. Dissertação (Mestrado – Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2019.

The *fast-food* segment has become a busy market with well-known companies such as: Subway, McDonald's, Burger King, Bob's and Habib's. Artificial intelligence and data science techniques can offer innumerable benefits to this market, such as allowing the development of computational models for decision making. In the context of finances involving the marketing of certain products, it is very common to come across scenarios where uncertainty is involved, especially when financial projections are desired, to evaluate risks and estimation. The objective of this work is to develop probabilistic models based on Bayesian Networks (BN) to make sales predictions and causality analysis among variables that influence the commercialization process of certain product groups in the *fast-food* segment. In this analysis we evaluated Bayesian networks with learning of structure based on constraints, through the algorithm *Grow Shrink* (GS), and Bayesian Networks with learning of structure based on score, through the algorithm *Hill-Climbing* (HC), later were compared with a model time series based on *Generalized Additive Model* (GAM). The data for analysis were acquired from a Brazilian *fast-food* chain with approximately 1100 associated stores, of which stores were used that belong to the state of São Paulo, as well as evaluated variables of sales groups in the period from 2010 to 2017. The results were evaluated by using the *Mean Absolute Percentage Error* (MAPE), which considers real values fed in models and values adjusted from the model and calculates the absolute difference between the two as a percentage of the real value, based on this calculation it is possible to obtain the accuracy of each model. The Bayesian Network (BN) with scoring based structure learning, using the *Hill Climbing* (HC) algorithm, was chosen as the best model because it presented more coherent causal relationships between vertices that influence the sales process, as well as combinations of vertices that result in product combos, in addition, achieved a 97.60% accuracy in the sales forecast of stores in the state of Sao Paulo (SP) in the test sample evaluated, based on the *Mean Absolute Percentage Error* (MAPE) metric.

Keywords: Bayesian Networks; Causal Relationships; Time Series; Sales Forecast; Fast Food;.

LISTA DE ILUSTRAÇÕES

Figura 1 – Metodologia aplicada neste trabalho - Fluxo das atividades	34
Figura 2 – Rede Bayesiana determinada apenas com o retorno do preço da ação da <i>Toyota Motor Corporation</i>	40
Figura 3 – Rede Bayesiana determinada apenas com o retorno do preço da ação da <i>NIKKEI</i>	40
Figura 4 – Previsão de exibições da página da <i>Wikipedia</i> do jogador de futebol americano <i>Peyton Manning</i>	42
Figura 5 – Componentes da Série Temporal	43
Figura 6 – Exemplos de vários tipos de redes (NEWMAN, 2003).	50
Figura 7 – Conjunto de dados do sistema de monitoramento de alarme (sintético) (BEINLICH <i>et al.</i> , 1989). Rede Bayesiana projetada para fornecer um sistema de mensagens de alarme para monitoramento do paciente.	52
Figura 8 – Conjunto de dados do sistema de monitoramento de alarme (sintético) (BEINLICH <i>et al.</i> , 1989). Rede Bayesiana projetada para fornecer um sistema de mensagens de alarme para monitoramento do paciente. <i>Zoom</i> na Rede Bayesiana para visualizar as relações causais entre os vértices.	52
Figura 9 – Ilustração de um procedimento de busca do algoritmo de aprendizado de estrutura <i>Hill-Climbing</i> (HC)	54
Figura 10 – Uma rede simples com variáveis discretas (<i>Subsídio e Compra</i>) e variáveis contínuas (<i>Colheita e Custo</i>).	58
Figura 11 – Análise de Importância - Relação de variáveis importantes	61
Figura 12 – Análise de tendência em vendas - Grupo de Milkshake	63
Figura 13 – Análise de tendência em vendas - Grupo de Sanduíches	64
Figura 14 – Análise de tendência em vendas - Grupo de Bebidas	65
Figura 15 – Análise de tendência em vendas - Grupo de Acompanhamento	66
Figura 16 – Análise de tendência em vendas - Grupo de Prato	67
Figura 17 – Análise de tendência em vendas - Grupo de Brinde	68
Figura 18 – Análise de tendência em vendas - Grupo de Café	69
Figura 19 – Análise de tendência em vendas - Grupo de Adicional	70
Figura 20 – Análise de tendência em vendas - Grupo de Venda (total)	71
Figura 21 – Análise de densidade das variáveis mais importantes	72
Figura 22 – Análise de densidade das variáveis mais importantes	73
Figura 23 – Modelo de Rede Bayesiana - Aprendizado de Estrutura - HC	75

Figura 24 – Modelo de Rede Bayesiana - Aprendizado de Estrutura - HC. Densidade Condicional sobre as variáveis <i>Venda e Bebida</i>	76
Figura 25 – Modelo de Rede Bayesiana - Aprendizado de Estrutura - HC. Densidade Condicional sobre as variáveis <i>Sanduíche e Acompanhamento</i>	77
Figura 26 – Modelo de Rede Bayesiana - Aprendizado de Estrutura - HC. Densidade Condicional sobre as variáveis <i>Brinde e Café</i>	78
Figura 27 – Modelo de Rede Bayesiana - Aprendizado de Estrutura - HC. Densidade Condicional sobre as variáveis <i>Prato e Adicional</i>	79
Figura 28 – Modelo de Rede Bayesiana - Aprendizado de Estrutura - HC. Densidade Condicional sobre a variável <i>Milkshake</i>	80
Figura 29 – Modelo de Rede Bayesiana - Aprendizado de Estrutura - GS	81
Figura 30 – Modelo de Rede Bayesiana - Aprendizado de Estrutura - GS. Densidade Condicional sobre as variáveis <i>Venda e Acompanhamento</i>	82
Figura 31 – Modelo de Rede Bayesiana - Aprendizado de Estrutura - GS. Densidade Condicional sobre as variáveis <i>Café e Prato</i>	83
Figura 32 – Modelo de Rede Bayesiana - Aprendizado de Estrutura - GS. Densidade Condicional sobre as variáveis <i>Brinde e Adicional</i>	84
Figura 33 – Modelo de Rede Bayesiana - Aprendizado de Estrutura - GS. Densidade Condicional sobre as variáveis <i>Sanduíche e Bebida</i>	85
Figura 34 – Modelo de Rede Bayesiana - Aprendizado de Estrutura - GS. Densidade Condicional sobre a variável <i>Milkshake</i>	86
Figura 35 – Modelo - Rede Bayesiana - <i>Hill-Climbing</i> (HC)	94
Figura 36 – Previsão de Vendas - Rede Bayesiana - <i>Hill-Climbing</i> (HC)	95
Figura 37 – Modelo - Rede Bayesiana - <i>Grow Shrink</i> (GS)	98
Figura 38 – Previsão de Vendas - Rede Bayesiana - <i>Grow Shrink</i> (GS)	99
Figura 39 – Previsão de Vendas - Série Temporal - Generalized Additive Model (GAM)	102
Figura 40 – Comparativo entre Modelos de Previsão de Vendas	103
Figura 41 – Monitoramento de pacientes de atendimento de emergência. Rede Bayesiana Discreta - Alarme, layout com sugiyama.	114
Figura 42 – Monitoramento de pacientes de atendimento de emergência. Rede Bayesiana Discreta - Alarme, layout com sugiyama. <i>Zoom</i> na Rede Bayesiana para visualizar as relações causais entre os vértices.	114
Figura 43 – Monitoramento de pacientes de atendimento de emergência. Rede Bayesiana Discreta - Alarme, layout em grade.	115
Figura 44 – Monitoramento de pacientes de atendimento de emergência. Rede Bayesiana Discreta - Alarme, layout em grade. <i>Zoom</i> na Rede Bayesiana para visualizar as relações causais entre os vértices.	116

Figura 45 – Monitoramento de pacientes de atendimento de emergência. Rede Bayesiana Discreta - Alarme. Avaliação da força das relações de causalidade entre os vértices.	118
Figura 46 – Monitoramento de pacientes de atendimento de emergência. Rede Bayesiana Discreta - Alarme. Avaliação da força das relações de causalidade entre os vértices. <i>Zoom</i> na Rede Bayesiana para visualizar as relações causais entre os vértices.	119

LISTA DE ALGORITMOS

Algoritmo 1 – Algoritmo <i>Hill-Climbing</i> (HC)	55
Algoritmo 2 – Algoritmo <i>Grow-Shrink</i> (GS)	56

LISTA DE TABELAS

Tabela 1 – Conjunto de Dados - <i>Food Storage</i> - Variáveis de Venda	35
Tabela 2 – Conjunto de Dados - <i>Food Storage</i> - Variáveis de Venda	36
Tabela 3 – Grupo de Produtos e Vendas - SP	60
Tabela 4 – Análise de Importância - Relação de variáveis importantes do conjunto de dados referente ao estado de São Paulo (SP)	62
Tabela 5 – Coeficientes associados ao vértice <i>Sanduíche</i> da Rede Bayesiana com aprendizagem de estrutura baseada em pontuação - HC	87
Tabela 6 – Coeficientes associados ao vértice <i>Café</i> da Rede Bayesiana com aprendizagem de estrutura baseada em pontuação - HC	87
Tabela 7 – Coeficientes associados ao vértice <i>Bebida</i> da Rede Bayesiana com aprendizagem de estrutura baseada em pontuação - HC	87
Tabela 8 – Coeficientes associados ao vértice <i>Acompanhamento</i> da Rede Bayesiana com aprendizagem de estrutura baseada em pontuação - HC	88
Tabela 9 – Coeficientes associados ao vértice <i>Adicional</i> da Rede Bayesiana com aprendizagem de estrutura baseada em pontuação - HC	88
Tabela 10 – Coeficientes associados ao vértice <i>Brinde</i> da Rede Bayesiana com aprendizagem de estrutura baseada em pontuação - HC	88
Tabela 11 – Coeficientes associados ao vértice <i>Prato</i> da Rede Bayesiana com aprendizagem de estrutura baseada em pontuação - HC	88
Tabela 12 – Coeficientes associados ao vértice <i>Venda</i> da Rede Bayesiana com aprendizagem de estrutura baseada em pontuação - HC	89
Tabela 13 – Coeficientes associados ao vértice <i>Milkshake</i> da Rede Bayesiana com aprendizagem de estrutura baseada em pontuação - HC	89
Tabela 14 – Coeficientes associados ao vértice <i>Milkshake</i> da Rede Bayesiana com aprendizagem de estrutura baseada em restrições - GS	89
Tabela 15 – Coeficientes associados ao vértice <i>Acompanhamento</i> da Rede Bayesiana com aprendizagem de estrutura baseada em restrições - GS	89
Tabela 16 – Coeficientes associados ao vértice <i>Café</i> da Rede Bayesiana com aprendizagem de estrutura baseada em restrições - GS	90
Tabela 17 – Coeficientes associados ao vértice <i>Prato</i> da Rede Bayesiana com aprendizagem de estrutura baseada em restrições - GS	90
Tabela 18 – Coeficientes associados ao vértice <i>Sanduíche</i> da Rede Bayesiana com aprendizagem de estrutura baseada em restrições - GS	90

Tabela 19 – Coeficientes associados ao vértice <i>Brinde</i> da Rede Bayesiana com aprendizagem de estrutura baseada em restrições - GS	90
Tabela 20 – Coeficientes associados ao vértice <i>Bebida</i> da Rede Bayesiana com aprendizagem de estrutura baseada em restrições - GS	90
Tabela 21 – Coeficientes associados ao vértice <i>Adicional</i> da Rede Bayesiana com aprendizagem de estrutura baseada em restrições - GS	91
Tabela 22 – Coeficientes associados ao vértice <i>Venda</i> da Rede Bayesiana com aprendizagem de estrutura baseada em restrições - GS	91
Tabela 23 – Previsão de Vendas - Rede Bayesiana - <i>Hill-Climbing</i> (HC)	96
Tabela 24 – Previsão de Vendas - Rede Bayesiana - <i>Grow Shrink</i> (GS)	100
Tabela 25 – Previsão de Vendas - Série Temporal - Generalized Additive Model (GAM)	102
Tabela 26 – Comparativo entre Modelos de Previsão de Vendas	104

LISTA DE ABREVIATURAS E SIGLAS

AR	<i>Auto Regressive</i>
ARCH	<i>Auto Regressive Conditionally Heteroscedastic</i>
ARMA	<i>Auto Regressive Moving Average</i>
BIC	<i>Bayesian Information Criterion</i>
CPD	Distribuições de Probabilidade Condicional
CRAN	<i>Comprehensive R Archive Network</i>
DAG	<i>Directed Acyclic Graph</i>
DGU	Distribuição Gaussiana Univariada
GAM	<i>Generalized Additive Model</i>
GS	<i>Grow-Shrink</i>
HC	<i>Hill-Climbing</i>
MA	<i>Moving Average</i>
MAP	<i>Maximum a Posteriori Probability</i>
MAPE	<i>Mean Absolute Percentage Error</i>
MB	<i>Markov Blank</i>
RB	Rede Bayesiana
RBD	Rede Bayesiana Discreta
RBG	Rede Bayesiana Gaussiana
RBGL	Redes Bayesianas Gaussianas Lineares

SUMÁRIO

1	INTRODUÇÃO	29
1.1	Contextualização	29
1.2	Objetivos	31
1.2.1	<i>Geral</i>	31
1.2.2	<i>Específicos</i>	31
1.3	Organização	32
2	METODOLOGIA	33
2.1	Fluxo das Atividades	34
2.2	Conjunto de Dados	35
2.3	Pré-Processamento	37
2.4	Análise Exploratória	37
2.5	Desenvolvimento dos Modelos de Previsão	37
2.6	Avaliação	37
2.7	Comparação	38
3	TRABALHOS RELACIONADOS	39
3.1	Modelos probabilísticos	39
3.1.1	<i>Redes Bayesianas</i>	39
3.2	Modelos Estatísticos	41
3.2.1	<i>Séries Temporais</i>	41
4	FUNDAMENTAÇÃO TEÓRICA	45
4.1	Ciência de Dados	45
4.1.1	<i>Visão Geral</i>	45
4.1.2	<i>Coleta dos Dados</i>	46
4.1.3	<i>Pré-processamento</i>	46
4.1.4	<i>Modelagem</i>	47
4.1.5	<i>Avaliação</i>	47
4.2	Teoremas probabilísticos	48
4.2.1	<i>Probabilidade</i>	48
4.2.2	<i>Probabilidade Condicional</i>	48
4.2.3	<i>Teorema de Bayes</i>	48
4.3	Teoria de Redes	50

4.4	Redes Probabilísticas	50
4.5	Redes Bayesianas	51
4.6	Algoritmos de Aprendizagem de Estrutura de Redes Bayesianas . .	53
4.6.1	<i>Algoritmos baseados em pontuação</i>	53
4.6.2	<i>Algoritmos baseados em restrições</i>	55
4.7	Rede Bayesiana Discreta	56
4.8	Redes Bayesianas Gaussianas	57
4.8.1	<i>Redes Bayesianas Híbridas</i>	57
5	PROPOSTA - REDES BAYESIANAS APLICADA À PREDIÇÃO DE VENDAS EM UMA GRANDE REDE DE FAST-FOOD BRASI- LEIRA	59
5.1	Análise Exploratória de Dados	60
5.1.1	<i>Sumarização - Grupo de Produtos mais vendidos</i>	60
5.1.2	<i>Análise de Variáveis Importantes</i>	61
5.1.3	<i>Análise de Tendência</i>	63
5.1.4	<i>Análise de Densidade - Variáveis Importantes</i>	72
5.1.5	<i>Análise de Densidade - Variáveis Importantes (Transformação Lo- garítmica)</i>	73
5.2	Modelagem	74
5.2.1	<i>Modelagem - Rede Bayesiana</i>	74
5.2.2	<i>Aprendizado de estrutura baseado em Pontuação</i>	75
5.2.3	<i>Aprendizado de estrutura baseado em Restrições</i>	81
5.3	Ajustes de parâmetros	87
5.3.1	<i>Ajustes de parâmetros (Rede Bayesiana - Aprendizado de Estrutura - HC)</i>	87
5.3.2	<i>Ajustes de parâmetros (Rede Bayesiana - Aprendizado de Estrutura - GS)</i>	89
6	RESULTADOS	93
6.1	Rede Bayesiana - Hill Climbing	94
6.1.1	<i>Modelo Matemático da Previsão de Vendas</i>	95
6.1.2	<i>Avaliação da Previsão de Vendas</i>	95
6.1.3	<i>Uso do modelo na prática</i>	96
6.2	Rede Bayesiana - Grow Shrink	98
6.2.1	<i>Modelo Matemático da Previsão de Vendas</i>	98
6.2.2	<i>Avaliação da Previsão de Vendas</i>	99
6.2.3	<i>Uso do modelo na prática</i>	100
6.3	Série Temporal - Generalized Additive Model (GAM)	101
6.3.1	<i>Modelo Matemático da Previsão de Vendas</i>	101

6.3.2	<i>Avaliação da Previsão de Vendas</i>	101
6.4	Comparativo entre Modelos de Previsão de Vendas	103
7	CONCLUSÃO	105
	REFERÊNCIAS	109
APÊNDICE A	BNVIEWER - AN R PACKAGE FOR INTERACTIVE VISUALIZATION OF BAYESIAN NETWORKS	111
A.1	Introdução	111
A.2	Instalação	113
A.3	Como utilizar o bnviewer	113

INTRODUÇÃO

1.1 Contextualização

Fast-food é uma expressão de origem inglesa que significa uma modalidade alimentar. Este segmento demanda agilidade no preparo e consumo, isto é, as refeições devem ser preparadas e vendidas em pouco tempo, por isso a exigência da padronização, mecanização e a rapidez desse sistema. O cardápio desses estabelecimentos geralmente são constituídos de lanches, acompanhados por batatas fritas e outras frituras, além de refrigerantes. Eles são servidos para consumidores que possuem pouco tempo para realizar suas refeições. (Toda Matéria, Fast-Food, 2017)

O sistema de *fast-food* geralmente é associado às grandes redes de alimentação, apesar de também ser um tipo de consumo encontrado em alguns restaurantes mais elaborados. As cadeias de lanchonete mais promissoras ganharam o mundo a partir de 1970 e tornaram-se um dos símbolos do capitalismo durante a Guerra Fria. Como exemplos de maior destaque podemos citar McDonald's, a maior de todas as cadeias de *fast-food*, seguida por Burger King, KFC, Subway e Pizza Hut. (Toda Matéria, Fast-Food, 2017)

No Brasil, as redes de *fast-food* têm início no Rio de Janeiro, a partir de 1951. Na altura Robert Bob Falkenburg, esportista e empresário nova-iorquino, fundou a “Falkenburg Sorveteria Ltda”. Inicialmente, o estabelecimento vendia sorvetes de baunilha. A receita foi trazida dos Estados Unidos e mais tarde viria a ser a primeira e uma das maiores empresas de *fast-food* do Brasil: o Bob's. Atualmente, cerca de 370 empresas atuam nesse empreendimento no país sendo que algumas são de origem brasileira: Habib's, Giraffas, Estupendo, Vivenda do Camarão, Spoleto, Mini Calzone, Au Au, A Pestikeira, Pitts Burg, etc. (Toda Matéria, Fast-Food, 2017)

As redes de *fast-food* ainda estão em crescimento no Brasil, pois o número de estabelecimentos aumentaram em cerca de 11% em 2015¹. Em 2014, o Brasil era 4º maior consumidor de

¹ Geofusion, 2016.

fast-food no mundo e a receita nacional era de cerca de 53 bilhões de reais². Este é um segmento de mercado muito concorrido com empresas bem conhecidas, tais como: Subway, McDonald's, Burger King, Bob's e Habib's. Segundo a Associação Brasileira das Indústrias da Alimentação (Abia), o faturamento da indústria de alimentos no País registrou alta de 85,8% entre os anos de 2010 e 2016³.

De acordo com o SIS/Sebrae, as motivações para esse aumento da alimentação fora de casa são⁴:

- Crescimento demográfico em regiões urbanizadas, o que demanda mais esse tipo de serviço.
- Modificação da estrutura familiar com aumento no número de casais sem filhos, o que ocasiona na redução nas despesas essenciais do orçamento familiar, e consequente investimento em lazer e novas experiências gastronômicas.
- Aumento na renda familiar ou individual que demanda um consumo mais qualificado, impulsionando negócios focados em produtos e serviços diferenciados (artesanais, frescos, orgânicos, cardápios para pessoas com restrições alimentares etc.).
- A crescente participação das mulheres no mercado de trabalho fez com que restasse pouco tempo para as atividades domésticas, como o preparo das refeições, o que também estimulou o aumento da frequência das refeições familiares fora de casa.

O SIS-Sebrae também apresenta números do *food service* brasileiro⁵:

- O mercado de alimentos e bebidas representa 10,1% do PIB nacional e estima-se que o segmento de *food service* abocanhe 2,7% desse montante.
- Em 2016, o segmento de alimentação fora do lar movimentou R\$184 bilhões no Brasil.
- O gasto (tíquete) médio foi de R\$13,40, um acréscimo de 8% em comparação ao ano anterior.
- Em torno de 56% da população consome fora de casa, e tem idade entre 18 e 49 anos.
- A estimativa de crescimento para o mercado de alimentação fora de casa (ainda não fechada), para 2017, deve girar em quase 11%.

² El País, 2016.

³ Food news Oficial, 2018.

⁴ Food news Oficial, 2018.

⁵ Food news Oficial, 2018.

Mediante este cenário, técnicas relacionadas à Ciência de Dados podem oferecer inúmeros benefícios neste nicho de mercado que costumam responder rapidamente as reações do mercado, dado o grau de concorrência. Dentre os diversos suportes que as técnicas relacionadas às Ciências de Dados podem oferecer, as relacionadas à previsão de vendas e da comercialização de determinados produtos estão entre os mais comuns para a realização de tomadas de decisões.

Contudo, não é uma tarefa trivial a realização destas análises, uma vez que os dados necessários para as análises citadas, costumam possuir um grande volume, apresentam ruídos, excesso de categorias e diversos outros problemas, além da dificuldade de seleção do modelo mais adequado para o problema que será analisado.

Apesar de tudo, tais análises podem fortalecer a comercialização de produtos específicos, permitir o melhor gerenciamento de estoques e otimizar o processo de negociação com fornecedores. Para alcançar os benefícios citados acima, é necessária a exploração de um conjunto de técnicas de pré-processamento dos dados coletados, além da modelagem e avaliação das ferramentas de predição.

No decorrer deste trabalho será apresentado uma abordagem baseada em modelos probabilísticos, com foco em Redes Bayesianas, para a inferir as vendas da rede de uma franquia de *fast-food* e analisar as relações de causalidade entre as variáveis que influenciam a comercialização de determinados grupos de produtos.

1.2 Objetivos

1.2.1 Geral

Objetivo geral deste trabalho consiste em desenvolver modelos probabilísticos baseados em Redes Bayesianas para realizar previsões de vendas bem como analisar causalidade entre variáveis de venda que influenciam no processo de comercialização de determinados grupos de produtos de uma rede de *fast-food*.

1.2.2 Específicos

- Analisar a relação de causalidade entre variáveis que influenciam no processo de venda.
- Gerar Redes Bayesianas a partir de algoritmos de aprendizado de estrutura baseados em pontuação e restrições.
- Comparar os modelos de Redes Bayesianas desenvolvidos com o modelo de série temporal do Facebook baseado em *Generalized Additive Model (GAM)*, intitulado "Prophet", e posteriormente avaliá-los.
- Analisar a eficiência entre modelos Bayesianos no contexto de predição de vendas.

- Desenvolver um pacote para linguagem R voltado para visualização interativa de Redes Bayesianas.

1.3 Organização

Este trabalho é organizado em sete capítulos: Introdução, Metodologia, Trabalhos Relacionados, Fundamentação Teórica, Proposta, Resultados e Conclusão. No segundo capítulo será descrito a base de dados utilizada, bem como todos os procedimentos adotados para a construção dos modelos de previsão de vendas, apresentará o fluxo das atividades, contendo todos os passos conduzidos para a elaboração deste trabalho, apresentará detalhes do conjunto de dados utilizado para elaboração dos modelos de previsão, bem como informações sobre a etapa de pré-processamento e transformações de dados aplicados na base de dados. Além disso, descreverá quais técnicas de análise exploratória e modelos de previsão adotados para o desenvolvimento deste trabalho. Ainda neste capítulo será descrito qual método de avaliação adotado para avaliar os modelos de previsão e modelo comparado. No terceiro capítulo serão apresentados trabalhos relacionados a modelos de previsão voltados a técnicas de redes Bayesianas e séries temporais. No quarto capítulo será apresentada a fundamentação teórica de Ciência de Dados, Teoremas Probabilísticos, Redes Probabilísticas, Algoritmos de Aprendizado de Estrutura em Redes Bayesianas baseados em pontuação, por meio do algoritmo *Hill-Climbing* (HC) e baseado em restrições, por meio do algoritmo *Grow Shrink* (GS), e Medidas de Avaliação dos Modelos. No quinto capítulo serão apresentados detalhes sobre a proposta deste trabalho intitulada **Redes Bayesianas aplicada à predição de vendas em uma grande rede de *fast-food* brasileira**. No sexto capítulo serão apresentados os resultados das previsões baseadas em Redes Bayesianas com aprendizado de estrutura, por meio dos algoritmos *Hill Climbing* (HC) e *Grow Shrink* (GS), assim como o comparativo com o modelo de série temporal baseada em *Generalized Additive Model* (GAM). Finalmente, no sétimo capítulo há a conclusão deste estudo e serão propostos trabalhos futuros.

METODOLOGIA

Neste capítulo é descrito a base de dados utilizada, bem como todos os procedimentos adotados para construção dos modelos de previsão de vendas. A Seção 2.1 apresenta o fluxo das atividades, contendo todos os passos que serão conduzidos para elaboração deste trabalho. A Seção 2.2 apresenta detalhes do conjunto de dados utilizado para elaboração dos modelos de previsão. A Seção 2.3 apresenta informações sobre a etapa de pré-processamento e transformações de dados aplicados na base de dados. A Seção 2.4 descreve técnicas de análise exploratória adotadas para o trabalho proposto. A Seção 2.5 apresenta os modelos de previsão adotados para o desenvolvimento deste trabalho. A Seção 2.6 descreve qual método de avaliação adotado para avaliar os modelos de previsão. A Seção 2.7 apresenta qual comparação será realizada nos modelos de previsão.

2.1 Fluxo das Atividades

A Figura 1 resume a metodologia da pesquisa apresentando o fluxo das principais tarefas adotadas para o desenvolvimento dos modelos probabilísticos baseados em Redes Bayesianas.

Neste fluxo, temos respectivamente: (1) A coleta da base de dados bruta (*fast-food*). (2) Pré-processamento, onde será realizado a limpeza e transformação dos dados. (3) Análise exploratória através de estatística descritiva. (4) Análise de características (*features*) mais importantes para os modelos, separação em amostra de treinamento e teste, desenvolvimento de modelos de previsão de vendas, avaliação e comparação de modelos.

Observa-se que embora as tarefas estejam em sequência, essas tarefas, assim como em mineração de dados, eventualmente tem que se retomar etapas anteriores em um processo iterativo e iterativo.

Figura 1 – Metodologia aplicada neste trabalho - Fluxo das atividades

Fonte: Elaborada pelo autor.

2.2 Conjunto de Dados

O conjunto de dados analisado foi adquirido de uma rede de *fast-food* brasileira que possui cerca de 1100 lojas associadas, destas, foram utilizadas lojas que pertencem ao estado de São Paulo, além disso, foram avaliadas variáveis de grupos de vendas no período de 2010 à 2017. No conjunto em questão, considerou-se a aplicação de transformações nos dados, para adequar aos modelos propostos neste trabalho. Nesta etapa, realizou-se o agrupamento das vendas diárias em vendas mensais, bem como a sumarização das vendas de grupos de produtos por estado. Nesta transformação, resultou-se nas variáveis observadas: *Data* ordenada de forma crescente, *Venda*, Grupos de *Milkshake*, *Sanduíche*, *Bebida*, *Acompanhamento*, *Adicional*, *Café*, *Brinde*, *Prato*, *Doce e Torta*, *Calda de Milkshake*, *Chopp*, *Refeição de Colaboradores*, *Molho Festival*, *Itens de Composição*, *Sachet Molho de Picanha*, *Lanche Franqueado*, *Molho*, *Outros*, conforme a Tabela 1 e 2.

Tabela 1 – Conjunto de Dados - *Food Storage* - Variáveis de Venda

Variável	Tipo de Variável	Descrição
Data (MM/YYYY)	Categórica Ordinal	Mês de referência da venda
Venda	Contínua	Total de Vendas no mês de referência
Milkshake	Contínua	Vendas do grupo de Milkshake no mês de referência
Sanduíche	Contínua	Vendas do grupo de Sanduíche no mês de referência
Bebida	Contínua	Vendas do grupo de Bebida no mês de referência
Acompanhamento	Contínua	Vendas do grupo de Acompanhamento no mês de referência
Adicional	Contínua	Vendas do grupo de Adicional no mês de referência
Café	Contínua	Vendas do grupo de Café no mês de referência
Brinde	Contínua	Vendas do grupo de Brinde no mês de referência
Prato	Contínua	Vendas do grupo de Prato no mês de referência
Doce e Torta	Contínua	Vendas do grupo de Doce e Torta no mês de referência

Tabela 2 – Conjunto de Dados - *Food Storage* - Variáveis de Venda

Variável	Tipo de Variável	Descrição
Data (MM/YYYY)	Categórica Ordinal	Mês de referência da venda
Venda	Contínua	Total de Vendas no mês de referência
Calda de Milkshake	Contínua	Vendas do grupo de Calda de Milkshake no mês de referência
Chopp	Contínua	Vendas do grupo de Chopp no mês de referência
Refeição de Colaboradores	Contínua	Vendas do grupo de Refeição de Colaboradores no mês de referência
Molho Festival	Contínua	Vendas do grupo de Molho Festival no mês de referência
Itens de Composição	Contínua	Vendas do grupo de Itens de Composição no mês de referência
Sachet Molho de Picanha	Contínua	Vendas do grupo de Sachet Molho de Picanha no mês de referência
Lanche Franqueado	Contínua	Vendas do grupo de Lanche Franqueado no mês de referência
Molho	Contínua	Vendas do grupo de Molho no mês de referência
Outros	Contínua	Vendas do grupo de Outros no mês de referência

2.3 Pré-Processamento

Na etapa de pré-processamento, foi considerado a limpeza de *outliers* que possam interferir no desenvolvimento dos modelos de previsão, bem como, a transformação dos dados em questão, de modo a auxiliar no ajuste e na convergência dos modelos propostos neste trabalho. Mais detalhes sobre a transformação de dados propostos neste trabalho são descritos na Seção [5.1.5](#)

2.4 Análise Exploratória

Na etapa de análise exploratória de dados, foram consideradas técnicas de estatística descritiva para obter conhecimento prévio do comportamento do conjunto de dados. As técnicas utilizadas neste trabalho foram: sumarização das variáveis analisadas, análise das variáveis mais importantes, análise de tendências para verificar comportamento de crescimento ou decréscimo das variáveis de vendas e análise de densidade. Mais detalhes sobre a análise exploratória de dados propostas neste trabalho são descritos na Seção [5.1](#).

2.5 Desenvolvimento dos Modelos de Previsão

Na etapa de desenvolvimento dos modelos de previsão, foram adotados técnicas de Redes Bayesianas com diferentes abordagens de aprendizado de estrutura, sendo: Redes Bayesianas com aprendizado de estrutura baseado em restrições, através do algoritmo *Grow Shrink* (GS), e aprendizado de estrutura baseado em pontuação, através do algoritmo *Hill-Climbing* (HC). Mais detalhes sobre o desenvolvimento dos modelos de previsão propostos neste trabalho são descritos na Seção [5.2](#).

2.6 Avaliação

Os modelos de previsão propostos neste trabalho, foram avaliados por meio da métrica *Mean Absolute Percentage Error* (MAPE) e por meio de validação cruzada, utilizando o método *holdout*. O método *holdout* consiste em dividir o conjunto total de dados em dois subconjuntos mutuamente exclusivos, um para treinamento (estimação dos parâmetros) e outro para teste (validação), desta forma, com base no conjunto de dados analisado, considerando o período de 2010 à 2017, será utilizado seis anos para treinamento, que contempla o período de 2010 à 2016, e, um ano para teste, utilizando a amostra do ano de 2017. Mais detalhes sobre a avaliação dos modelos de previsão propostos neste trabalho estão descritos na Seção [4.1.5](#).

2.7 Comparação

Os modelos probabilísticos baseados em Redes Bayesianas, propostos neste trabalho, foram comparados com um modelo estatístico baseado em séries temporais. O Capítulo 3 abordará maiores detalhes sobre os modelos utilizados nesta comparação.

TRABALHOS RELACIONADOS

Neste capítulo, são levantados trabalhos baseados em paradigmas probabilísticos e estatísticos, relacionados a predição. Os trabalhos relacionados estão divididos em duas categorias: (i) Modelos probabilísticos de predição, baseados em Redes Bayesianas, (ii) Modelos estatísticos de predição, baseados em Séries Temporais. Esta divisão tem como objetivo, facilitar o entendimento sobre as diversas abordagens existentes, para cada um dos aspectos de implementação dos modelos de predição.

3.1 Modelos probabilísticos

3.1.1 *Redes Bayesianas*

No artigo "*Application of Bayesian Network to stock price prediction*", [Eisuke, Harada e Mizuno \(2012\)](#), apresentam um algoritmo de predição do preço das ações da *Toyota Motor Corporation* e *NIKKEI*, usando Redes Bayesianas, e comparam com modelos de séries temporais, *Auto Regressive* (AR), *Moving Average* (MA), *Auto Regressive Moving Average* (ARMA) e modelos *Auto Regressive Conditionally Heteroscedastic* (ARCH). Primeiro, a rede é determinada a partir do preço diário das ações, e em seguida, é aplicada para prever o preço diário das ações que já foi observada. Os erros de previsão são estimados pela diferença entre os preços das ações reais e previstos. No caso das predições da *NIKKEI*, a média e máxima de erros do modelo proposto em Redes Bayesianas, ficou entre 6% e 30%, abaixo dos algoritmos de previsão baseados em séries temporais. Com relação às predições da *Toyota Motor Corporation*, a média e máxima de erros do modelo proposto em Redes Bayesianas, ficou entre 2% e 20%, abaixo dos algoritmos de previsão baseados em séries temporais. Desta forma, é evidente que os modelos baseados em Redes Bayesianas, tiveram desempenho superior às séries temporais, no aspecto da predição do preço das ações da *Toyota Motor Corporation* e *NIKKEI*.

A Figura 2, apresenta a rede gerada apenas pelo retorno do preço das ações da *Toyota*

Motor Corporation. Observe que o retorno $R(t)$ depende do retorno anterior de 1 dia $R(t-1)$, 4 dias antes do retorno $R(t-4)$, 6 dias antes do retorno $R(t-6)$, 8 dias antes do retorno $R(t-8)$ e 9 dias retorno anterior $R(t-9)$.

Figura 2 – Rede Bayesiana determinada apenas com o retorno do preço da ação da *Toyota Motor Corporation*

Fonte: Artigo "*Application of Bayesian Network to stock price prediction*", (EISUKE; HARADA; MIZUNO, 2012). Alterado graficamente pelo autor.

A Figura 3, apresenta a rede gerada apenas pelo retorno do preço das ações da *NIKKEI*. Observe que o retorno $R(t)$ depende do retorno anterior de 5 dias $R(t-5)$, 7 dias antes do retorno $R(t-7)$, 9 dias antes do retorno $R(t-9)$, 10 dias antes do retorno $R(t-10)$.

Figura 3 – Rede Bayesiana determinada apenas com o retorno do preço da ação da *NIKKEI*

Fonte: Artigo "*Application of Bayesian Network to stock price prediction*", (EISUKE; HARADA; MIZUNO, 2012). Alterado graficamente pelo autor.

No artigo "*A survey of the applications of bayesian networks in agriculture*", para Drury *et al.* (2017), a aplicação de Redes Bayesianas tem sido utilizadas em diversos campos de

pesquisa, principalmente em domínios onde envolvam incerteza. Essa atenção aumentou uma série de abordagens que usam diferentes técnicas de aprendizagem de máquina. É discutível que, as Redes Bayesianas sejam particularmente adequadas à pesquisa agrícola, devido à sua capacidade de fundamentar com informações incompletas e incorporar novas informações.

3.2 Modelos Estatísticos

3.2.1 Séries Temporais

No artigo "Forecasting at scale", [Taylor e Letham \(2017\)](#), apontam que a previsão é uma tarefa comum em ciência dos dados, que ajuda as organizações com planejamento de capacidade, definição de metas e detecção de anomalias. Apesar de sua importância, há grandes desafios associados à produção de previsões confiáveis e de alta qualidade.

Produzir previsões de alta qualidade, não é um problema fácil para máquinas ou para a maioria dos analistas. Assim, [Taylor e Letham \(2017\)](#), observaram dois temas principais, na prática de criar uma variedade de previsões de negócios:

1. As técnicas de previsão totalmente automáticas, podem ser frágeis e muitas vezes são inflexíveis demais, para incorporar suposições ou heurísticas úteis.
2. Os analistas que podem produzir previsões de alta qualidade, são bastante raros, porque a previsão, é uma habilidade de ciência de dados especializada que requer uma experiência substancial.

Desta forma, [Taylor e Letham \(2017\)](#), ambos pesquisadores do *Facebook*, abordam a ferramenta de previsão chamada *Prophet*, baseada em séries temporais, disponível em Python e R, que se mostra robusta em tarefas de previsão, e em situações quando há dados faltantes, mudanças na tendência e grande quantidade *outliers*.

Prophet, possui em seu núcleo, um Modelo Aditivo Generalizado com ajustes Bayesianos, utilizando linguagem probabilística de programação Stan¹. *Stan* executa a otimização do *Maximum a Posteriori Probability* (MAP) para parâmetros extremamente rápidos (<1 segundo), que permite estimar a incerteza dos parâmetros usando o algoritmo Monte Carlo Hamiltoniano. O modelo é resistente aos efeitos de *outliers*, e suporta dados coletados em uma escala de tempo irregular (presença de dados em falta), sem a necessidade de interpolação.

Na Figura 4, temos um exemplo de previsão de exibições da página da *Wikipedia* do jogador de futebol americano *Peyton Manning*, em escala logarítmica, utilizando o *Prophet*. No exemplo, trata-se como *Peyton Manning* é um jogador de futebol americano, é possível verificar

¹ [Stan - Inferência Bayesiana](#)

que a sazonalidade anual desempenha um papel importante, enquanto a periodicidade semanal também é evidenciada.

Figura 4 – Previsão de exibições da página da *Wikipedia* do jogador de futebol americano *Peyton Manning*.

Fonte: Artigo "*Forecasting at scale*", (TAYLOR; LETHAM, 2017).

A Figura 5, descreve os componentes da série temporal. Ela mostra mais claramente a sazonalidade anual associada à navegação na página de *Peyton Manning* (temporada de futebol e os *playoffs*), bem como a sazonalidade semanal: mais visitas no dia e após os jogos (domingos e segundas-feiras). Também é possível observar o ajuste descendente para o componente de tendência, desde que ele se aposentou recentemente.

A idéia importante no *Prophet*, é que, ao fazer um trabalho melhor de ajustar o componente de tendência, de forma muito flexível, modelamos de forma mais precisa a sazonalidade, e conseqüentemente, o resultado é uma previsão mais precisa. O *Prophet* usa um modelo de regressão muito flexível (um pouco como ajuste de curva), em vez de um modelo tradicional de séries temporais para essa tarefa, porque permite a flexibilidade de modelagem, facilita a adaptação ao modelo e manipula dados ausentes ou *outliers*.

Figura 5 – Componentes da Série Temporal

Fonte: Artigo "*Forecasting at scale*", ([TAYLOR; LETHAM, 2017](#)).

FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, são descritos os principais conceitos de toda a teoria envolvida para o desenvolvimento deste trabalho. A Seção 4.1, apresentada a fundamentação teórica sobre ciência de dados. A Seção 4.2, aborda fundamentos de teoremas probabilísticos. A Seção 4.4, aborda as redes probabilísticas. A Seção 4.5, aborda as Redes Bayesianas. A Seção 4.6, aborda algoritmos de aprendizado de estrutura em Redes Bayesianas baseados em pontuação, por meio do algoritmo *Hill-Climbing* (HC) e baseado em restrições, por meio do algoritmo *Grow Shrink* (GS). A Seção 4.7, aborda Redes Bayesianas Discretas. A Seção 4.8, aborda Redes Bayesianas Gaussianas e a Seção 4.8.1, aborda Redes Bayesianas Híbridas.

4.1 Ciência de Dados

4.1.1 Visão Geral

Com o advento da era da informação, cada vez mais pessoas utilizam a Internet para realizar as mais variadas atividades, desde entretenimento à produção profissional, e cada uma dessas atividades geram uma quantidade massiva de dados. Segundo Witten *et al.* (2016), a quantidade de dados armazenada nos bancos de dados do mundo, dobra a cada 20 meses. Nunca foram gerados tantos dados na história mundial. Para Provost e Fawcett (2013), tais dados, se explorados adequadamente, podem representar vantagens competitivas.

Logo, a Ciência de dados é a área que realiza a exploração e análise de todos os dados disponíveis, sejam eles estruturados ou não, com o objetivo de desenvolver compreensão, extrair conhecimento e formular ações que gerem resultados, tendo como mecanismo básico os sistemas computacionais. Após a extração e interpretação dos dados é importante apresentá-los através de uma linguagem simples e não técnica para os usuários finais e tomadores de decisão.

Os projetos em Ciência de Dados, geralmente possuem três etapas básicas e sequenciais: coleta de dados, decisões e ações executadas alinhadas às decisões obtidas. A falha em qualquer

etapa impacta no produto final, remetendo a perda de recursos ou em uma baixa qualidade do produto final.

É interessante ressaltar que a Ciência de Dados possui forte ligação com Big Data e Mineração de Dados. Big Data, representam as tecnologias que permitem coletar, gerenciar e analisar conjuntos de dados muito grandes, para os sistemas de banco de dados convencionais (TAMBE, 2014). A ligação de Ciências de dados com Big Data se dá no processo de coleta de dados, em que os dados fornecidos pelo Big Data poderão ser úteis no processo de tomada de decisão (PROVOST; FAWCETT, 2013).

Os conceitos de Mineração de Dados e Ciência de Dados, podem gerar algumas confusões, dada a semelhança de ambos. A Mineração de Dados é a extração de conhecimento a partir de dados, através de tecnologias que incorporam esses princípios, de acordo com Provost e Fawcett (2013), e estruturada nas fases: 1) Análise do Problema, 2) Pré-processamento 3) Extração de padrões, 4) Pós-processamento, 5) Uso do conhecimento extraído. E elementarmente, a Mineração de Dados pode ser considerada como um subconjunto da Ciência de Dados, ou seja, ela pertence a Ciência de Dados. A Ciência de Dados inclui as seguintes etapas adicionais: Planejamento dos Experimentos e Avaliação.

Entre a coleta de dados e as ações, existe um caminho a ser percorrido. Neste caminho, está o planejamento dos experimentos, que permitirá entender o problema que deverá ser resolvido, além das etapas de pré-processamento, modelagem, avaliação, metas e o tempo de cada etapa.

4.1.2 Coleta dos Dados

O primeiro passo no processo de Ciências de Dados, é a coleta dos dados. Nesta fase deve-se realizar uma busca minuciosa, para identificar quais dados relacionados ao problema estão disponíveis, e coletá-los. Esses dados serão usados na criação de modelos capazes de extrair conhecimento. Não se deve negligenciar nenhum dado disponível. Dados podem vir de diferentes fontes, como: sinais de localização de *smartphones*, logs de servidores, sensores de dados, câmeras de monitoramento de trânsito e segurança, blogs, e-mails, redes sociais e etc.

A incorporação de dados externos podem aumentar o valor comercial de um sistema de previsão, no entanto, é essencial avaliar a viabilidade da aquisição dos dados externos. Para ser viável, os dados externos devem aumentar a discernibilidade do classificador. Weiss e Provost (2003), discutiram o impacto da distribuição de classes nos dados externos sobre o desempenho do classificador, respondendo à questão da qualidade dos dados externos.

4.1.3 Pré-processamento

Esta fase se inicia logo após a coleta dos dados, e a organização dos mesmos ocorre geralmente na forma de um conjunto. Um dos objetivos nesta fase, é de solucionar problemas

nos dados, podendo ser eles valores desconhecidos, atributos irrelevantes ou identificar e tratar dados corrompidos.

As ferramentas analíticas estão cada vez mais robustas, porém ainda se faz necessária a realização da preparação de dados, antes de sua utilização. O seu objetivo, é o de viabilizar e organizar os dados para a sua utilização. Os requisitos desta etapa são variados, e estão fortemente ligados às características dos dados que serão utilizados, bem como às ferramentas que serão utilizadas e a análise que se pretende realizar (PROVOST; FAWCETT, 2013).

Algumas ferramentas possuem limitações diretas aos tipos de dados que serão utilizados, algumas atuam apenas com dados numéricos e outras apenas com dados categóricos ou simbólicos, sendo muitas vezes necessárias conversões. Podem ser necessárias também, a normalização, redimensionamento dos dados, preenchimento de dados ausentes, substituição ou remoção de dados suspeitos e ordenação ou tabulação dos dados, de acordo com Provost e Fawcett (2013) e Faceli *et al.* (2011).

4.1.4 Modelagem

A fase da modelagem busca fornecer algum tipo de modelo ou padrão, que capta regularidades nos dados. No momento da implementação da modelagem, é importante considerar os diversos tipos de técnicas e algoritmos para a mineração de dados, uma vez que a real capacidade dos sistemas oferecerem decisões, são objetos desta etapa, de acordo com Provost e Fawcett (2013).

4.1.5 Avaliação

A fase de avaliação, é a que fornece indícios de quão confiável um modelo é na captura de padrões nos dados. Devido o conjunto de dados possuir variáveis contínuas e o trabalho proposto ser direcionado a previsão de vendas no seguimento de *fast-food*, que está relacionado a problemas de aprendizado supervisionado com foco em regressão, será adotado a métrica *Mean Absolute Percentage Error* (MAPE). O MAPE considera valores reais alimentados em modelos e valores ajustados a partir do modelo, e calcula a diferença absoluta entre os dois como porcentagem do valor real, com base neste cálculo, é possível obter a acurácia de cada modelo.

Segundo Barros (2015), o MAPE pode ser definido pela equação 4.1:

$$MAPE = \frac{\sum_{i=1}^N \frac{Y_{(t)} - \hat{Y}_{(t)}}{Y_{(t)}} \cdot 100}{N} \quad (4.1)$$

Onde $Y_{(t)}$ é o valor real no período, e $\hat{Y}_{(t)}$ é o valor da previsão para o período t com N totais de observações.

4.2 Teoremas probabilísticos

4.2.1 Probabilidade

Definição 1. Uma medida de probabilidade P , é uma função que atribui para cada evento A , um número denotado por $P(A)$, chamada de probabilidade de A . Se $S = \{A_1, A_2, \dots, A_n\}$, é o conjunto de todas possibilidades que um experimento pode fornecer, S , é chamado de espaço amostral deste experimento e cada um de seus elementos de ponto amostral ou evento elementar. Tal função satisfaz as seguintes condições:

1. (P1) P é não negativa; $P(A) \geq 0$ para todo evento A .
2. (P2) P é normalizada; $P(S) = 1$.
3. (P3) P é enumeravelmente aditiva, descrito também como, σ -aditiva (medida da união enumerável disjunta é igual a soma das medidas); $P(\bigcup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} P(A_j)$ para toda coleção de eventos mutuamente disjuntos $A_j, j = 1, 2, \dots$.

4.2.2 Probabilidade Condicional

Suponha que sabemos que um evento B ocorreu, e que queremos saber a probabilidade de outro evento A acontecer, mas levando em conta nosso conhecimento sobre a ocorrência de B . Sendo assim, o que realmente queremos é atualizar nosso conhecimento sobre A , baseando-se na ocorrência de outro evento. Essa atualização pode ser calculada através da probabilidade condicional de A , dado que B ocorreu.

Definição 2. Seja B um evento tal que $P(B) > 0$. Então a probabilidade condicional, dado B , é a função escrita como $P(.|B)$, e definida para todo evento A , da seguinte forma:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (4.2)$$

A função $P(.|B)$ satisfaz os axiomas de Kolmogorov, portanto, é uma medida de probabilidade, de acordo com [DeGroot e Schervish \(2012\)](#).

4.2.3 Teorema de Bayes

Dado que nosso interesse seja calcular a probabilidade de cada um dos eventos disjuntos $B_1, \dots, B_n$ acontecer, e que temos a possibilidade de observar a ocorrência de outro evento A . Se tivermos $P(A|B_j)$, para cada j podemos calcular, usando o Teorema de Bayes, as probabilidades $P(B_j|A)$, também para cada j .

Seja $\{B_j, j = 1, 2, \dots, n\}$, é uma partição de S , se $P(B_j) > 0 \forall j = 1, 2, 3, \dots, n$, e se existe um evento A com $P(A) > 0$, então:

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{P(A)} \quad (4.3)$$

Desta forma, o teorema de Bayes nos permite computar a probabilidade condicional de cada evento numa partição, dada a observação de um evento A , conforme [DeGroot e Schervish \(2012\)](#).

4.3 Teoria de Redes

Rede, ou grafo, é um par ordenado $G = (V, E)$, onde V , é o conjunto de vértices e $E \subset V \times V$, é o conjunto de arestas que conectam pares de vértices. Em grafos simples, então uma aresta é um par não-ordenado $e = \{u, v\}$, que conecta, ou incide em, dois vértices distintos $u, v \in V$. Vértices conectados por uma aresta são chamados adjacentes ou vizinhos. Um grafo é chamado simples, se não for ponderado, não for direcionado e não tiver várias arestas ou laços próprios. Em outras palavras, as arestas do grafo não têm pesos e são bidirecionais, não há duas arestas conectando o mesmo par de vértices e nenhuma aresta conecta um vértice a si mesmo, conforme [Gibbons \(1985\)](#).

Na Figura 6, temos tipos de redes: (a) uma rede não direcionada com apenas um único tipo de vértice e um único tipo de aresta; (b) uma rede com um número de tipos discretos de vértice e de aresta; (c) uma rede com pesos variados de vértice e aresta; (d) uma rede direcionada na qual cada aresta tem uma direção, conforme [Newman \(2003\)](#).

Figura 6 – Exemplos de vários tipos de redes ([NEWMAN, 2003](#)).

4.4 Redes Probabilísticas

Redes probabilísticas são modelos gráficos que representam interações entre variáveis aleatórias, podendo tais relações serem vistas como simples conjuntos de dependências ou como associações de causa-efeito, dependendo da construção e interpretação de cada modelo.

Em geral, a construção de redes probabilísticas faz uso da chamada teoria de grafos,

na qual as variáveis aleatórias são representadas como vértices e as interações entre elas como arestas. Distribuições de probabilidades conjuntas podem ser representadas naturalmente através desses modelos, onde as presenças e ausências de arestas representam, respectivamente, as relações de dependência ou independência entre as variáveis, de acordo com [Faria \(2014\)](#).

4.5 Redes Bayesianas

Uma Rede Bayesiana é uma estrutura gráfica que nos permite representar e raciocinar sobre um domínio incerto. Os vértices de uma Rede Bayesiana representam um conjunto de variáveis aleatórias do domínio. Um conjunto de arestas direcionadas conectam pares de vértices, representando as dependências diretas entre variáveis, conforme [Korb e Nicholson \(2010\)](#).

Segundo a [ISBA - International Society for Bayesian Analysis \(1992\)](#), as Redes Bayesianas são um modelo de representação do conhecimento que trabalham com o conhecimento incerto e incompleto através da teoria da probabilidade bayesiana, publicada pelo matemático Thomas Bayes em 1763. A estrutura dada por, $G = (V, E)$ de uma Rede Bayesiana, é um *Directed Acyclic Graph* (DAG), em que V é o conjunto de vértices e E é o conjunto de arestas. O DAG define uma fatoração da distribuição de probabilidade conjunta de $V = X_1, X_2, \dots, X_v$, frequentemente chamado de distribuição de probabilidade global, em um conjunto de distribuições de probabilidade locais, uma para cada variável. A forma da fatoração é dada pela propriedade Markov das Redes Bayesianas, conforme [Korb e Nicholson \(2010\)](#), que afirma que toda variável aleatória X_i depende diretamente apenas de seus pais $\prod_{X_i}$:

$$P(X_1, \dots, X_v) = \prod_{i=1}^v P(X_i | \prod_{X_i}) \quad (\text{para variáveis discretas}) \quad (4.4)$$

$$f(X_1, \dots, X_v) = \prod_{i=1}^v f(X_i | \prod_{X_i}) \quad (\text{para variáveis contínuas}) \quad (4.5)$$

A correspondência entre independência condicional (das variáveis aleatórias) e separação gráfica (dos nós correspondentes do grafo), foi estendida a um trio arbitrário de subconjuntos disjuntos de V definidos por [Pearl \(2014\)](#), com a *d-separation* (da separação dependente da direção). Portanto, os algoritmos de seleção de modelos tentam primeiro aprender a estrutura da Rede Bayesiana (daí o nome dos algoritmos de aprendizado de estrutura) e então estimar os parâmetros das funções de distribuição local, condicionais à estrutura aprendida. Essa abordagem em duas etapas tem a vantagem de considerar uma função de distribuição local por vez, e não exige modelar explicitamente a função de distribuição global. Outra vantagem, é que os algoritmos de aprendizagem são capazes de escalar, para se adaptarem a modelos de alta dimensão, sem incorrer na chamada maldição de dimensionalidade, conforme [Scutari \(2009\)](#).

Para ilustrar a representação gráfica das Redes Bayesianas, na Figura 41, temos a vi-

sualização interativa da Rede Bayesiana Discreta - Alarme com layout Sugiyama. Na Figura 42, ampliamos a Rede Bayesiana para visualizar as relações causais entre os vértices. Ambos desenvolvidos através do pacote *bnviewer*, para visualização interativa de Redes Bayesianas, idealizado por [Fernandes \(2018\)](#).

Figura 7 – Conjunto de dados do sistema de monitoramento de alarme (sintético) ([BEINLICH et al., 1989](#)). Rede Bayesiana projetada para fornecer um sistema de mensagens de alarme para monitoramento do paciente.

Figura 8 – Conjunto de dados do sistema de monitoramento de alarme (sintético) ([BEINLICH et al., 1989](#)). Rede Bayesiana projetada para fornecer um sistema de mensagens de alarme para monitoramento do paciente. *Zoom* na Rede Bayesiana para visualizar as relações causais entre os vértices.

4.6 Algoritmos de Aprendizagem de Estrutura de Redes Bayesianas

4.6.1 Algoritmos baseados em pontuação

Esses algoritmos atribuem uma pontuação a cada Rede Bayesiana candidata, e tentam maximizá-la com algum algoritmo de busca heurística. Os algoritmos de busca gulosos (busca escalada na montanha (*Hill-Climbing*) ou busca tabu (*Tabu Search*)), são uma escolha comum, mas quase qualquer tipo de procedimento de busca pode ser usado.

A abordagem baseada em pontuação é dos métodos mais populares para induzir Redes Bayesianas a partir de dados, especialmente para fins de estimação por funções de distribuição de probabilidade. O processo atribui uma pontuação a cada Rede Bayesiana candidata, e mede quão bem essa Rede Bayesiana descreve o conjunto de dados D . Assumindo uma estrutura G , sua pontuação é dada pela Equação 4.6,

$$Score(G, D) = Pr(G|D), \quad (4.6)$$

em outras palavras, a probabilidade posterior de G , dado o conjunto de dados. Um algoritmo baseado em pontuação tenta maximizar essa pontuação. A Equação 4.6, pode ser convertida em uma definição mais conveniente usando a lei de Bayes conforme a Equação 4.7.

$$Score(G, D) = Pr(G|D) = \frac{Pr(G|D)Pr(G)}{Pr(D)} \quad (4.7)$$

Para calcular $Pr(G|D)Pr(G)$, a abordagem bayesiana calcula a média de todos os parâmetros possíveis, pesando cada um deles pela probabilidade posterior:

A Figura 9, ilustra o procedimento de busca do algoritmo de aprendizado de estrutura *Hill-Climbing* (HC). O algoritmo HC, busca a melhor estrutura da rede Bayesiana, adicionando e removendo arestas, além de ajustar arestas reversas (verifica o sentido das arestas). A melhor estrutura é aquela que maximiza (ou minimiza) as medidas de *score*. O algoritmo HC, utiliza a métrica *Bayesian Information Criterion* (BIC) como método de pontuação.

Figura 9 – Ilustração de um procedimento de busca do algoritmo de aprendizado de estrutura *Hill-Climbing* (HC)

Fonte: Artigo "*Learning Bayesian Network Model Structure from Data*", (MARGARITIS, 2003).

No Algoritmo 1, a busca é iniciada a partir de uma rede vazia ou possivelmente aleatória, embora, se houver conhecimento prévio, possa ser usado para propagar a rede inicial. O procedimento *ProbabilityTables()*, estima os parâmetros das funções de distribuição de probabilidade locais. Geralmente, esta é uma estimativa de máxima verossimilhança das entradas de probabilidade do conjunto de dados, que para funções de distribuição de probabilidade locais multinomiais, consiste na contagem do número de tuplas que se enquadram em cada entrada da tabela de probabilidade multinomial, na Rede Bayesiana. O *loop* principal do algoritmo, verifica tentativas de adição, remoção ou reversão de uma única aresta da rede, para posteriormente avaliá-la, e assim, adotar a rede que tiver a maior pontuação. O processo é interrompido quando não há uma mudança de aresta única que aumente a pontuação.

Algoritmo 1 – Algoritmo *Hill-Climbing* (HC)

```

1: procedimento BICHILLCLIMBING( $D$ )
2:    $E \leftarrow 0$ 
3:    $T \leftarrow \text{ProbabilityTables}(E, D)$ 
4:    $B \leftarrow \langle U, E, T \rangle$ 
5:    $score \leftarrow -\infty$ 
6:   enquanto  $score > maxscore$  faça
7:      $maxscore \leftarrow score$ 
8:     para cada attribute pair  $(X, Y)$  faca
9:       para cada  $E' \in \{E \cup \{X \rightarrow Y\},$ 
            $\{E - \{X \rightarrow Y\},$ 
            $\{E - \{X \rightarrow Y\} \cup \{Y \rightarrow X\}\}$  faca
10:         $T' \leftarrow \langle U, E', T' \rangle$ 
11:         $newscore \leftarrow \text{BICscore}(B', D)$ 
12:        se  $newscore > score$  então
13:           $B \leftarrow B'$ 
14:           $score \leftarrow newscore$ 
15:        fim se
16:      fim para cada
17:    fim para cada
18:  fim enquanto
19:  retorna  $B$ 
20: fim procedimento

```

4.6.2 Algoritmos baseados em restrições

A abordagem de algoritmos baseados em restrições, aprendem a estrutura da rede analisando as relações probabilísticas decorrentes da propriedade de *Markov* de Redes Bayesianas, com testes de independência condicional, e em seguida, construindo um grafo que satisfaça as instruções de *d-separation*. Os modelos resultantes, são muitas vezes interpretados como modelos causais, mesmo quando aprendidos com dados observados (PEARL, 2014).

A propriedade do *Markov Blank* (MB), de um vértice v , é a união dos pais de v , seus filhos e os pais de seus filhos. Uma propriedade útil do *Markov Blank* é que cada conjunto de nós é condicionalmente independente em v quando condicionado no MB de v .

Para qualquer variável X , a propriedade *Markov Blank*(X), é qualquer variável definida, tal que, para qualquer $Y \in V \setminus \{BL(X), X\}, X \perp\!\!\!\perp Y \mid BL(X)$.

O Algoritmo 2, visa recuperar a propriedade *Markov Blank* de X baseada em testes de independência pareados. Desta forma, o algoritmo necessita de duas fases: Uma crescente (*growing*) e outra encolhida (*shrink*), daí o seu nome. Partindo de um conjunto vazio S , a fase de crescimento adiciona variáveis a S , desde que sejam dependentes X , dado o conteúdo atual de S . A idéia por trás é que: contanto que a propriedade *Markov Blank* de X seja violada (ou seja, existe uma variável em U que depende de X , dado o S atual), nós a adicionamos ao atual conjunto S ,

até que não haja mais essas variáveis. Nesse processo, no entanto, pode haver algumas variáveis que foram adicionadas a S que estavam realmente fora do *Blanket*. Tais variáveis, são aquelas que se tornaram independentes de X , em um ponto posterior quando os vértices intermediários, *d-separations* da Rede Bayesiana subjacente, foram adicionados a S . Esta observação motiva a fase de encolhimento (*shrink*), que identifica e remove estas variáveis.

Algoritmo 2 – Algoritmo *Grow-Shrink* (GS)

```

1: procedimento GROWSHRINK( $D$ )
2:    $S \leftarrow \emptyset$ 
3:   enquanto  $\exists Y \in U - \{X\}$  de tal modo que  $Y \not\perp X | S$ , faça
4:      $S \leftarrow S \cup \{Y\}$ 
5:   fim enquanto
6:   enquanto  $\exists Y \in S$  de tal modo que  $Y \perp X | S - \{Y\}$ , faça
7:      $S \leftarrow S - \{Y\}$ 
8:   fim enquanto
9:   retorna  $B(X) \leftarrow S$ 
10: fim procedimento
  
```

4.7 Rede Bayesiana Discreta

De acordo com [Kjaerulff e Madsen \(2008\)](#), uma Rede Bayesiana Discreta (RBD), $N = (X, G, P)$, sobre variáveis X , consiste em um grafo acíclico direcionado $G = (V, E)$ e um conjunto de distribuições de probabilidade condicional P . Cada vértice v , em G , corresponde um a um, a uma variável aleatória discreta $X_v \in X$, com um conjunto finito de estados mutuamente exclusivos. As arestas direcionadas em $E \subseteq V \times V$ de G , especificam suposições de dependência condicional, e independência entre variáveis aleatórias, de acordo com o critério de *d-separation*.

Há uma distribuição de probabilidade condicional, $P(X_v | X_{pa(v)}) \in P$, para cada variável $X_v \in X$. O conjunto de variáveis representadas pelos pais, $pa(v)$, de $v \in V$ em $G = (V, E)$, às vezes são chamados de variáveis de condicionamento de X_v .

Definição 3. Para [Nielsen e JENSEN \(2013\)](#), uma Rede Bayesiana Discreta (RBD) $N = (X, G, P)$ consiste de

1. Um DAG, $G = (V, E)$, com vértices, $V = \{v_1, \dots, v_n\}$, e arestas direcionadas, E .
2. Um conjunto de variáveis aleatórias discretas, X , representadas pelos vértices de G .
3. Um conjunto de distribuições de probabilidade condicional, P , contendo uma distribuição, $P(X_v | X_{pa(v)})$, para cada variável aleatória, $X_v \in X$.

4.8 Redes Bayesianas Gaussianas

Uma Rede Bayesianas Gaussianas (RBG), é definido como uma Rede Bayesianas (RB) onde todas as variáveis são contínuas, e onde todas as Distribuições de Probabilidade Condicional (CPD) são gaussianas lineares.

Definição 4. Para [Gómez et al. \(2010\)](#), uma Rede Bayesianas Gaussianas (RBG), é uma Rede Bayesianas (RB) onde a densidade de probabilidade conjunta associada às variáveis $X = \{X_1, \dots, X_n\}$, segue uma distribuição gaussiana multivariada $N(\mu, \Sigma)$, dada por:

$$f(x) = (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp -\frac{1}{2} (x - \mu)' |\Sigma|^{-1} (x - \mu) \quad (4.8)$$

μ é um vetor médio de dimensão n e Σ é a matriz de co-variância definida positiva $n \times n$

A densidade de probabilidade condicional para $X_i (i = 1, \dots, n)$, é uma Distribuição Gaussiana Univariada (DGU) dada por:

$$X_i | pa(X_i) \sim N(\mu_i + \sum_{j=1}^{i-1} \beta_{ji} (x_j - \mu_j), v_i) \quad (4.9)$$

Onde μ_i é a média de X_i , β_{ji} é o coeficiente de regressão quando X_i é regredido em seus pais $X_j \in Pa(X_i)$, e v_i é a variância condicional de X_i dada aos pais. Pode-se ressaltar que $\beta_{ji} = 0$ se e somente se não houver link de X_j para X_i .

4.8.1 Redes Bayesianas Híbridas

Uma rede com variáveis discretas e contínuas, é chamada de Rede Bayesianas Híbrida. Para especificar uma rede híbrida, temos de especificar dois novos tipos de distribuições: a distribuição condicional, para uma variável contínua, dado pais discretos ou contínuos, e a distribuição condicional para uma variável discreta, dado pais contínuos. Considere o exemplo simples na Figura 10, em que um cliente compra alguma fruta dependendo de seu custo, que por sua vez, depende do volume da colheita e do fato de o esquema de subsídios do governo estar em vigor. A variável *Custo* é contínua, e tem pais contínuos e discretos; a variável *Compra* é discreta e tem pai contínuo.

Figura 10 – Uma rede simples com variáveis discretas (*Subsídio* e *Compra*) e variáveis contínuas (*Colheita* e *Custo*).

Para a variável *Custo*, precisamos especificar $(Custo | Colheita, Subsídio)$. O pai discreto é manipulado por enumeração explícita, ou seja, pela especificação de $(Custo | Colheita, Subsídio)$ e de $(Custo | Colheita, \neg Subsídio)$. Para tratar *Colheita*, especificamos como a distribuição sobre o custo c depende do valor contínuo h de *Colheita*. Em outras palavras, especificamos os parâmetros da distribuição de custo como uma função de h . A escolha mais comum é a distribuição gaussiana linear, na qual o filho tem uma distribuição gaussiana cuja média μ varia linearmente com o valor do pai, e cujo desvio-padrão σ é fixo. Precisamos de duas distribuições, uma para *Subsídio* e uma para $\neg Subsídio$, com parâmetros diferentes:

$$P(c|h, Subsídio) = N(a_t h + b_t, \sigma_t^2)(c) = \frac{1}{\sigma_t \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{c - (a_t h + b_t)}{\sigma_t} \right)^2} \quad (4.10)$$

$$P(c|h, \neg Subsídio) = N(a_f h + b_f, \sigma_f^2)(c) = \frac{1}{\sigma_f \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{c - (a_f h + b_f)}{\sigma_f} \right)^2} \quad (4.11)$$

Para esse exemplo, a distribuição condicional para *Custo* é especificada pela nomenclatura da distribuição gaussiana linear, fornecendo-se os parâmetros $a_t, b_t, \Sigma_t, a_f, b_f$ e Σ_f , conforme [Norvig e Russell \(2015\)](#)

PROPOSTA - REDES BAYESIANAS APLICADA À PREDIÇÃO DE VENDAS EM UMA GRANDE REDE DE FAST-FOOD BRASILEIRA

Neste capítulo, são apresentados detalhes sobre a proposta deste trabalho, intitulada **Redes Bayesianas aplicada à predição de vendas em uma grande rede de *fast-food* brasileira**. A Seção 5.1, descreve a análise exploratória do conjunto de dados e contempla: a Seção 5.1.1, apresenta a sumarização dos grupos de produtos mais vendidos. A Seção 5.1.2, apresenta a análise das variáveis mais importantes do conjunto de dados. A Seção 5.1.3, apresenta os gráficos de análise de tendência das variáveis que influenciam o processo de venda. A Seção 5.1.4, apresenta a análise de densidade das variáveis mais importantes. A Seção 5.1.5, apresenta a análise de densidade das variáveis mais importantes após transformação logarítima. A Seção 5.2, apresenta a modelagem das Redes Bayesianas com aprendizado de estrutura baseado em restrições, através do algoritmo *Grow Shrink* (GS), e, aprendizado de estrutura baseado em pontuação, através do algoritmo *Hill-Climbing* (HC). A Seção 5.3, apresenta a abordagem de ajuste de parâmetros das Redes Bayesianas.

5.1 Análise Exploratória de Dados

Nesta Seção, são apresentadas análises exploratórias no conjunto de dados, considerando o período de 2010 à 2017, no estado de São Paulo - SP.

5.1.1 Sumarização - Grupo de Produtos mais vendidos

Na Tabela 3, são apresentados os grupos de produtos com maior fluxo de vendas. Pode-se observar que o grupo **Milkshake**, foi o grupo em que houve o maior volume em vendas de produtos.

Tabela 3 – Grupo de Produtos e Vendas - SP

Grupo de Produtos	Vendas R\$
Milkshake	538.285.471,70
Sanduiche	324.145.965,94
Bebida	154.870.264,64
Acompanhamento	85.991.336,46
Adicional	13.354.495,45
Prato	4.403.061,03
Café	4.361.322,26
Brinde	1.660.749,93
Chopp	793.814,05
Doce e Torta	531.526,32
Itens de Composição	117.948,82
Sachet Molho Picanha	3.328,15
Molho Festival	2.161,75
Calda Milkshake	226,63
Lanche Franqueado	0,00
Linha Original	0,00
Molho	0,00
Refeição Colaboradores	0,00
Outros	0,00

5.1.2 Análise de Variáveis Importantes

A Figura 11 apresenta a análise das variáveis importantes do conjunto de dados, referente o estado de São Paulo (SP). Para análise, foi utilizado o pacote *caret* para linguagem R (KUHN, 2008). A análise considerou 19 variáveis (grupos de produtos), que buscam explicar a variável venda. Na avaliação, foi realizado um *cross-validation* com $k=10$ folds e 3 repetições, mediante um modelo de regressão linear, visando estimar a contribuição de cada variável para o modelo por meio do valor absoluto da *t-statistic* para cada parâmetro, bem como também, a métrica *Root Mean Square Error* (RMSE), como métrica para os modelos avaliados. Em estatística, o *t-statistic*, é a razão entre a saída do valor estimado de um parâmetro e seu valor hipotético para seu erro padrão. É usado no teste de hipóteses, por meio do teste *t* de Student. Após análise, verificou-se que 15 variáveis são consideradas importantes para explicar a variável venda.

Figura 11 – Análise de Importância - Relação de variáveis importantes

Fonte: Elaborada pelo autor.

Na Tabela 4, são apresentadas as variáveis importantes do conjunto de dados, referente o estado de São Paulo (SP), com seus respectivos *scores*.

Dentre as variáveis analisadas, 8 possuem percentual de importância acima de 1%, sendo que as variáveis **Milkshake**, **Sanduiche**, **Bebida e Acompanhamento** apresentaram respectivamente 100%, 24%, 11% e 9% de importância para o modelo em questão. As demais variáveis apresentaram importância abaixo de 5%.

Tabela 4 – Análise de Importância - Relação de variáveis importantes do conjunto de dados referente ao estado de São Paulo (SP)

Variável	Score
Milkshake	100,00
Sanduiche	24,53
Bebida	11,30
Acompanhamento	9,15
Adicional	2,23
Café	1,79
Brinde	1,73
Prato	1,40
Doce e Torta	0,19
Calda de Milkshake	0,13
Chopp	0,09
Refeição Colaboradores	0,05
Molho Festival	0,01
Itens de Composição	0,01
Sachet Molho Picanha	0,00

5.1.3 Análise de Tendência

A análise de tendência está associada ao estudo de séries temporais, uma vez que se trata de uma das componentes de uma série temporal. As componentes de tendência são frequentemente, aquelas que produzem mudanças graduais em longo prazo. São normalmente provocadas, por exemplo, pelo crescimento constante na população, no produto interno bruto, no efeito da competição, ou por outros fatores que falham na tentativa de produzir mudanças repentinas, mas produzem variações graduais e regulares ao longo do tempo.

Na Figura 12, temos a análise de tendência em vendas, no grupo de Milkshake, indicado pela cor verde, no período de 2010 à 2017.

Figura 12 – Análise de tendência em vendas - Grupo de Milkshake

Fonte: Elaborada pelo autor.

Na Figura 13, temos a análise de tendência em vendas, no grupo de Sanduíches, indicado pela cor laranja, no período de 2010 à 2017.

Figura 13 – Análise de tendência em vendas - Grupo de Sanduíches

Fonte: Elaborada pelo autor.

Na Figura 14, temos a análise de tendência em vendas, no grupo de Bebidas, indicado pela cor magenta, no período de 2010 à 2017.

Figura 14 – Análise de tendência em vendas - Grupo de Bebidas

Fonte: Elaborada pelo autor.

Na Figura 15, temos a análise de tendência em vendas, no grupo de Acompanhamento, indicado pela cor vermelha, no período de 2010 à 2017.

Figura 15 – Análise de tendência em vendas - Grupo de Acompanhamento

Fonte: Elaborada pelo autor.

Na Figura 16, temos a análise de tendência em vendas, no grupo de Prato, indicado pela cor amarela, no período de 2010 à 2017.

Figura 16 – Análise de tendência em vendas - Grupo de Prato

Fonte: Elaborada pelo autor.

Na Figura 17, temos a análise de tendência em vendas, no grupo de Brinde, indicado pela cor azul, no período de 2010 à 2017. No ano de 2010, 2011 e 2012, é possível observar que não houve vendas de brindes, isto ocorreu pois itens relacionados a este grupo de venda não haviam sido lançados no cardápio da rede de *fast-food* neste período.

Figura 17 – Análise de tendência em vendas - Grupo de Brinde

Fonte: Elaborada pelo autor.

Na Figura 18, temos a análise de tendência em vendas, no grupo de Café, indicado pela cor vermelho escuro, no período de 2010 à 2017.

Figura 18 – Análise de tendência em vendas - Grupo de Café

Fonte: Elaborada pelo autor.

Na Figura 19, temos a análise de tendência em vendas, no grupo de Adicional, indicado pela cor cinza, no período de 2010 à 2017.

Figura 19 – Análise de tendência em vendas - Grupo de Adicional

Fonte: Elaborada pelo autor.

Na Figura 20, temos a análise de tendência em vendas, no grupo de Venda (total), indicado pela cor preta, no período de 2010 à 2017.

Figura 20 – Análise de tendência em vendas - Grupo de Venda (total)

Fonte: Elaborada pelo autor.

5.1.4 Análise de Densidade - Variáveis Importantes

Nas Figuras, 21a e 21b, temos respectivamente o histograma e diagrama de densidade das variáveis de grupos de vendas mais importantes. Pode-se observar que, a distribuição das variáveis estão distorcidas, isto, pode dificultar a convergência das Redes Bayesianas, bem como, as relações não lineares nos modelos de regressão. Portanto, será aplicado uma transformação logarítmica nas variáveis analisadas, com a finalidade melhorar tais situações. Transformações logarítmicas são um meio conveniente de transformar uma variável altamente distorcida, em uma que é mais aproximadamente normal, ou seja, há uma distribuição log-normal, definida como uma distribuição cujo logaritmo é normalmente distribuído, mas cuja escala não transformada é distorcida.

Figura 21 – Análise de densidade das variáveis mais importantes

Fonte: Elaborada pelo autor.

5.1.5 Análise de Densidade - Variáveis Importantes (Transformação Logarítmica)

Na Figuras, 22a e 22b, temos respectivamente o histograma e diagrama de densidade das variáveis de grupos de vendas mais importantes com transformação logarítmica. Pode-se observar, que as variáveis seguem uma distribuição mais próxima da normal. Transformações logarítmicas em variáveis são técnicas muito comuns em lidar com situações onde existe uma relação não linear entre as variáveis independentes e dependentes, desta forma, usar o logaritmo de uma ou mais variáveis, torna o relacionamento efetivo não-linear, preservando o modelo linear.

Figura 22 – Análise de densidade das variáveis mais importantes

Fonte: Elaborada pelo autor.

5.2 Modelagem

5.2.1 Modelagem - Rede Bayesiana

Os modelos de aprendizado de estrutura de Redes Bayesianas foram desenvolvidos utilizando o pacote *bnlearn* para linguagem R, concebido por [Scutari, Scutari e MMPC \(2017\)](#). O pacote possui algoritmos de aprendizagem de estrutura de Redes Bayesianas com suporte a variáveis discretas e contínuas, aprendizagem de parâmetros e inferência. Ele implementa algoritmos de aprendizagem de estrutura baseados em restrições (*GS*, *IAMB*, *Inter-IAMB*, *Fast-IAMB*, *MMPC*, *HitonPC*), em pares (*ARACNE* e *Chow-Liu*), baseados em pontuação (*Hill-Climbing* e *Tabu Search*) e híbridos (*MMHC* e *RSMAX2*) para Redes Bayesianas Discretas, Gaussianas e Gaussianas Condicionais, juntamente com várias funções de *scores* e testes de independência condicional, conforme [Scutari \(2009\)](#).

As visualizações dos modelos de Redes Bayesianas foram desenvolvidos utilizando o pacote de *bnviewer* para linguagem R, idealizado por [Fernandes \(2018\)](#). O *bnviewer* é um pacote R para visualização interativa de Redes Bayesianas, baseado nos pacotes *bnlearn* e *visNetwork*. Este pacote foi desenvolvido no contexto desta pesquisa e disponibilizado publicamente em (<http://robsonfernandes.net/bnviewer/>). O *bnviewer* interpreta vários algoritmos de aprendizado de estrutura, fornecidos pelo pacote *bnlearn*, e permite visualizá-los de forma interativa por meio de layouts customizados, bem como permite realizar interações com operações de *drag and drop*, *zoom* e *click* nos vértices e arestas da rede.

Além disso, o pacote *bnviewer* permite interpretar a força das relações de causalidade entre os vértices da Rede Bayesiana analisada. Para visualizar a força das arestas entre vértices da Rede Bayesiana, é necessário utilizar a função *boot.strength* do pacote *bnlearn*. Esta função utiliza *bootstrap* não paramétrico para avaliar a força e a direções das arestas. O método *bootstrap*, definido por [Efron e Tibshirani \(1994\)](#), é um método de amostragem baseado na construção de sub-amostras a partir de uma amostra inicial. O método *bootstrap* pode ser construído com abordagem paramétrica ou não-paramétrico. No *bootstrap* paramétrico, a distribuição é realizada por meio de suposições, e assim, posteriormente reamostramos observações da distribuição postulada, mas usando os valores das estimativas dos parâmetros no processo de geração de pseudo-amostras. No *bootstrap* não-paramétrico, o processo de amostragem se dá a partir da função de distribuição empírica dos dados (ou dos resíduos quando há uma estrutura de regressão).

5.2.2 Aprendizagem de estrutura baseado em Pontuação

Na Figura 23, temos a representação gráfica da Rede Bayesiana com aprendizagem de estrutura baseada em pontuação, no qual foi utilizado o algoritmo *Hill-Climbing* (HC), através de um processo de aprendizagem de estrutura automático. O grafo direcionado formado pela Figura 23 é composto por 9 vértices e 12 arestas direcionadas. Os vértices indicam as variáveis de grupos de vendas, as arestas direcionadas indicam a influência que cada variável tem sobre a outra. Todas as arestas possuem pesos que vão de 0 a 1, e indicam a força da aresta, sendo que as arestas nas cores magenta indicam forças iguais ou superiores a 0.95, e as arestas em tons de azul indicam forças inferiores a 0.95.

Figura 23 – Modelo de Rede Bayesiana - Aprendizagem de Estrutura - HC

Fonte: Elaborada pelo autor.

A Figura 24a apresenta a relação de variáveis que influenciam a variável *Venda*, dado pela densidade condicional: $f(Venda | Milkshake + Bebida + Adicional)$. A Figura 24b apresenta a relação de variáveis que influenciam a variável *Bebida*, dado pela densidade condicional: $f(Bebida | Sanduíche)$. Os vértices indicam as variáveis de grupos de vendas, as arestas direcionadas indicam a influência que cada variável tem sobre a outra. Todas as arestas possuem pesos que vão de 0 a 1, e indicam a força da aresta, sendo que as arestas nas cores magenta indicam forças iguais ou superiores a 0.95, e as arestas em tons de azul indicam forças inferiores a 0.95. Os vértices são apresentados na cor azul, e ao realizar o evento *click* sobre um vértice, ele se apresenta na cor magenta, dando ênfase ao mesmo, posteriormente os demais vértices são inativados visualmente, indicados pela cor cinza.

Figura 24 – Modelo de Rede Bayesiana - Aprendizado de Estrutura - HC. Densidade Condicional sobre as variáveis *Venda* e *Bebida*

(a) Rede Bayesiana - Aprendizado de Estrutura - HC.
Densidade Condicional:
 $f(Venda | Milkshake + Bebida + Adicional)$

(b) Rede Bayesiana - Aprendizado de Estrutura - HC.
Densidade Condicional:
 $f(Bebida | Sanduíche)$

Fonte: Elaborada pelo autor.

A Figura 25a apresenta a relação de variáveis que influenciam a variável *Sanduíche*, observa-se que a variável *Sanduíche* não é influenciada pelas demais variáveis, dado pela densidade condicional: $f(\text{Sanduíche})$. A Figura 25b apresenta a relação de variáveis que influenciam a variável *Acompanhamento*, dado pela densidade condicional: $f(\text{Acompanhamento} | \text{Sanduíche} + \text{Brinde})$. Os vértices indicam as variáveis de grupos de vendas, as arestas direcionadas indicam a influência que cada variável tem sobre a outra. Todas as arestas possuem pesos que vão de 0 a 1, e indicam a força da aresta, sendo que as arestas nas cores magenta indicam forças iguais ou superiores a 0.95, e as arestas em tons de azul indicam forças inferiores a 0.95. Os vértices são apresentados na cor azul, e ao realizar o evento *click* sobre um vértice, ele se apresenta na cor magenta, dando ênfase ao mesmo, posteriormente os demais vértices são inativados visualmente, indicados pela cor cinza.

Figura 25 – Modelo de Rede Bayesiana - Aprendizado de Estrutura - HC. Densidade Condicional sobre as variáveis *Sanduíche* e *Acompanhamento*

(a) Rede Bayesiana - Aprendizado de Estrutura - HC.
Densidade Condicional:
 $f(\text{Sanduíche})$

(b) Rede Bayesiana - Aprendizado de Estrutura - HC.
Densidade Condicional:
 $f(\text{Acompanhamento} | \text{Sanduíche} + \text{Brinde})$

Fonte: Elaborada pelo autor.

A Figura 26a apresenta a relação de variáveis que influenciam a variável *Brinde*, observa-se que a variável *Brinde* não é influenciada pelas demais variáveis, dado pela densidade condicional: $f(\text{Brinde})$. A Figura 26b apresenta a relação de variáveis que influenciam a variável *Café*, observa-se que a variável *Café* não é influenciada pelas demais variáveis, dado pela densidade condicional: $f(\text{Café})$. Os vértices indicam as variáveis de grupos de vendas, as arestas direcionadas indicam a influência que cada variável tem sobre a outra. Todas as arestas possuem pesos que vão de 0 a 1, e indicam a força da aresta, sendo que as arestas nas cores magenta indicam forças iguais ou superiores a 0.95, e as arestas em tons de azul indicam forças inferiores a 0.95. Os vértices são apresentados na cor azul, e ao realizar o evento *click* sobre um vértice, ele se apresenta na cor magenta, dando ênfase ao mesmo, posteriormente os demais vértices são inativados visualmente, indicados pela cor cinza.

Figura 26 – Modelo de Rede Bayesiana - Aprendizado de Estrutura - HC. Densidade Condicional sobre as variáveis *Brinde* e *Café*

(a) Rede Bayesiana - Aprendizado de Estrutura - HC.
Densidade Condicional:
 $f(\text{Brinde})$

(b) Rede Bayesiana - Aprendizado de Estrutura - HC.
Densidade Condicional:
 $f(\text{Café})$

Fonte: Elaborada pelo autor.

A Figura 27a apresenta a relação de variáveis que influenciam a variável *Prato*, dado pela densidade condicional: $f(\text{Prato} | \text{Bebida} + \text{Brinde})$. A Figura 27b apresenta a relação de variáveis que influenciam a variável *Adicional*, dado pela densidade condicional: $f(\text{Adicional} | \text{Café})$. Os vértices indicam as variáveis de grupos de vendas, as arestas direcionadas indicam a influência que cada variável tem sobre a outra. Todas as arestas possuem pesos que vão de 0 a 1, e indicam a força da aresta, sendo que as arestas nas cores magenta indicam forças iguais ou superiores a 0.95, e as arestas em tons de azul indicam forças inferiores a 0.95. Os vértices são apresentados na cor azul, e ao realizar o evento *click* sobre um vértice, ele se apresenta na cor magenta, dando ênfase ao mesmo, posteriormente os demais vértices são inativados visualmente, indicados pela cor cinza.

Figura 27 – Modelo de Rede Bayesiana - Aprendizado de Estrutura - HC. Densidade Condicional sobre as variáveis *Prato* e *Adicional*

(a) Rede Bayesiana - Aprendizado de Estrutura - HC.
Densidade Condicional:
 $f(\text{Prato} | \text{Bebida} + \text{Brinde})$

(b) Rede Bayesiana - Aprendizado de Estrutura - HC.
Densidade Condicional:
 $f(\text{Adicional} | \text{Café})$

Fonte: Elaborada pelo autor.

Na Figura 28a apresenta a relação de variáveis que influenciam a variável *Milkshake*, dado pela densidade condicional: $f(\text{Milkshake} | \text{Sanduíche} + \text{Bebida} + \text{Acompanhamento})$. Os vértices indicam as variáveis de grupos de vendas, as arestas direcionadas indicam a influência que cada variável tem sobre a outra. Todas as arestas possuem pesos que vão de 0 a 1, e indicam a força da aresta, sendo que as arestas nas cores magenta indicam forças iguais ou superiores a 0.95, e as arestas em tons de azul indicam forças inferiores a 0.95. Os vértices são apresentados na cor azul, e ao realizar o evento *click* sobre um vértice, ele se apresenta na cor magenta, dando ênfase ao mesmo, posteriormente os demais vértices são inativados visualmente, indicados pela cor cinza.

Figura 28 – Modelo de Rede Bayesiana - Aprendizado de Estrutura - HC. Densidade Condicional sobre a variável *Milkshake*

(a) Rede Bayesiana - Aprendizado de Estrutura - HC.

Densidade Condicional:

$$f(\text{Milkshake} | \text{Sanduíche} + \text{Bebida} + \text{Acompanhamento})$$

Fonte: Elaborada pelo autor.

5.2.3 Aprendizagem de estrutura baseado em Restrições

Na Figura 29, temos a representação gráfica da Rede Bayesiana com aprendizagem de estrutura baseada em restrições, no qual foi utilizado o algoritmo *Grow-Shrink* (GS), através de um processo de aprendizagem de estrutura automático. O grafo direcionado formado pela Figura 29 é composto por 9 vértices e 5 arestas direcionadas. Os vértices indicam as variáveis de grupos de vendas, as arestas direcionadas indicam a influência que cada variável tem sobre a outra. Todas as arestas possuem pesos que vão de 0 a 1, e indicam a força da aresta, sendo que as arestas nas cores magenta indicam forças iguais ou superiores a 0.95, e as arestas em tons de azul indicam forças inferiores a 0.95.

Figura 29 – Modelo de Rede Bayesiana - Aprendizagem de Estrutura - GS

Fonte: Elaborada pelo autor.

A Figura 30a apresenta a relação de variáveis que influenciam a variável *Venda*, dado pela densidade condicional: $f(Venda | Milkshake + Adicional)$. A Figura 30b apresenta a relação de variáveis que influenciam a variável *Acompanhamento*, observa-se que a variável *Acompanhamento* não é influenciada pelas demais variáveis, dado pela densidade condicional: $f(Acompanhamento | Sanduíche)$. Os vértices indicam as variáveis de grupos de vendas, as arestas direcionadas indicam a influência que cada variável tem sobre a outra. Todas as arestas possuem pesos que vão de 0 a 1, e indicam a força da aresta, sendo que as arestas nas cores magenta indicam forças iguais ou superiores a 0.95, e as arestas em tons de azul indicam forças inferiores a 0.95. Os vértices são apresentados na cor azul, e ao realizar o evento *click* sobre um vértice, ele se apresenta na cor magenta, dando ênfase ao mesmo, posteriormente os demais vértices são inativados visualmente, indicados pela cor cinza.

Figura 30 – Modelo de Rede Bayesiana - Aprendizado de Estrutura - GS. Densidade Condicional sobre as variáveis *Venda* e *Acompanhamento*

Fonte: Elaborada pelo autor.

A Figura 31a apresenta a relação de variáveis que influenciam a variável *Café*, observa-se que a variável *Café* não influencia bem como também não é influenciada pelas demais variáveis, dado pela densidade condicional: $f(\text{Café})$. A Figura 31b apresenta a relação de variáveis que influenciam a variável *Prato*, observa-se que a variável *Prato* não influencia bem como também não é influenciada pelas demais variáveis, dado pela densidade condicional: $f(\text{Prato})$. Os vértices indicam as variáveis de grupos de vendas, as arestas direcionadas indicam a influência que cada variável tem sobre a outra. Os vértices são apresentados na cor azul, e ao realizar o evento *click* sobre um vértice, ele se apresenta na cor magenta, dando ênfase ao mesmo, posteriormente os demais vértices são inativados visualmente, indicados pela cor cinza.

Figura 31 – Modelo de Rede Bayesiana - Aprendizado de Estrutura - GS. Densidade Condicional sobre as variáveis *Café* e *Prato*

(a) Rede Bayesiana - Aprendizado de Estrutura - GS.
Densidade Condicional:
 $f(\text{Café})$

(b) Rede Bayesiana - Aprendizado de Estrutura - GS.
Densidade Condicional:
 $f(\text{Prato})$

Fonte: Elaborada pelo autor.

A Figura 32a apresenta a relação de variáveis que influenciam a variável *Brinde*, observa-se que a variável *Brinde* não influencia bem como também não é influenciada pelas demais variáveis, dado pela densidade condicional: $f(\text{Brinde})$. A Figura 32b apresenta a relação de variáveis que influenciam a variável *Adicional*, observa-se que a variável *Adicional* não é influenciada pelas demais variáveis, dado pela densidade condicional: $f(\text{Adicional})$. Os vértices indicam as variáveis de grupos de vendas, as arestas direcionadas indicam a influência que cada variável tem sobre a outra. Todas as arestas possuem pesos que vão de 0 a 1, e indicam a força da aresta, sendo que as arestas nas cores magenta indicam forças iguais ou superiores a 0.95, e as arestas em tons de azul indicam forças inferiores a 0.95. Os vértices são apresentados na cor azul, e ao realizar o evento *click* sobre um vértice, ele se apresenta na cor magenta, dando ênfase ao mesmo, posteriormente os demais vértices são inativados visualmente, indicados pela cor cinza.

Figura 32 – Modelo de Rede Bayesiana - Aprendizado de Estrutura - GS. Densidade Condicional sobre as variáveis *Brinde* e *Adicional*

Fonte: Elaborada pelo autor.

A Figura 33a apresenta a relação de variáveis que influenciam a variável *Sanduíche*, dado pela densidade condicional: $f(\text{Sanduíche})$. A Figura 33b apresenta a relação de variáveis que influenciam a variável *Bebida*, dado pela densidade condicional: $f(\text{Bebida} | \text{Sanduíche})$. Os vértices indicam as variáveis de grupos de vendas, as arestas direcionadas indicam a influência que cada variável tem sobre a outra. Todas as arestas possuem pesos que vão de 0 a 1, e indicam a força da aresta, sendo que as arestas nas cores magenta indicam forças iguais ou superiores a 0.95, e as arestas em tons de azul indicam forças inferiores a 0.95. Os vértices são apresentados na cor azul, e ao realizar o evento *click* sobre um vértice, ele se apresenta na cor magenta, dando ênfase ao mesmo, posteriormente os demais vértices são inativados visualmente, indicados pela cor cinza.

Figura 33 – Modelo de Rede Bayesiana - Aprendizado de Estrutura - GS. Densidade Condicional sobre as variáveis *Sanduíche* e *Bebida*

Fonte: Elaborada pelo autor.

Na Figura 34a apresenta a relação de variáveis que influenciam a variável *Milkshake*, dado pela densidade condicional: $f(\text{Milkshake} | \text{Bebida})$. Os vértices indicam as variáveis de grupos de vendas, as arestas direcionadas indicam a influência que cada variável tem sobre a outra. Todas as arestas possuem pesos que vão de 0 a 1, e indicam a força da aresta, sendo que as arestas nas cores magenta indicam forças iguais ou superiores a 0.95, e as arestas em tons de azul indicam forças inferiores a 0.95. Os vértices são apresentados na cor azul, e ao realizar o evento *click* sobre um vértice, ele se apresenta na cor magenta, dando ênfase ao mesmo, posteriormente os demais vértices são inativados visualmente, indicados pela cor cinza.

Figura 34 – Modelo de Rede Bayesiana - Aprendizado de Estrutura - GS. Densidade Condicional sobre a variável *Milkshake*

Fonte: Elaborada pelo autor.

5.3 Ajustes de parâmetros

Em Redes Bayesianas, onde todos os vértices são discretos, os ajustes de parâmetros são tabulados, no que geralmente é referido como Tabela de Probabilidade Condicional (TPC).

No modelo em questão, trata-se de uma Rede Bayesiana, onde todos os vértices são contínuos. Desta forma, os ajustes de parâmetros são realizados por coeficientes de regressão, para cada variável em relação aos seus pais, bem como também, encontram o intercepto (parâmetro β_0 é chamado intercepto ou coeficiente linear e representa o ponto em que a reta regressora corta o eixo dos y 's, quando $x = 0$).

Para obter os coeficientes das Redes Bayesianas, foi utilizado o pacote *rbmn* para linguagem R, concebido por Denis (2013). O pacote permite a criação, manipulação e simulação de Redes Bayesianas Gaussianas Lineares (RBGL).

5.3.1 Ajustes de parâmetros (Rede Bayesiana - Aprendizado de Estrutura - HC)

Tabela 5 – Coeficientes associados ao vértice *Sanduiche* da Rede Bayesiana com aprendizagem de estrutura baseada em pontuação - HC

Parâmetros do vértice Sanduíche (Distribuição Gaussiana)		
Densidade Condicional		
$f(\text{Sanduíche})$		
Coeficientes	(Intercepto)	14.82548
Desvio padrão resíduos		0.8265545

Tabela 6 – Coeficientes associados ao vértice *Café* da Rede Bayesiana com aprendizagem de estrutura baseada em pontuação - HC

Parâmetros do vértice Café (Distribuição Gaussiana)		
Densidade Condicional		
$f(\text{Café})$		
Coeficientes	(Intercepto)	10.34651
Desvio padrão resíduos		1.05642

Tabela 7 – Coeficientes associados ao vértice *Bebida* da Rede Bayesiana com aprendizagem de estrutura baseada em pontuação - HC

Parâmetros do vértice Bebida (Distribuição Gaussiana)		
Densidade Condicional		
$f(\text{Bebida} \mid \text{Sanduíche})$		
Coeficientes	(Intercepto)	-0.6811503
	Sanduíche	0.9963611
Desvio padrão resíduos		0.06057377

Tabela 8 – Coeficientes associados ao vértice *Acompanhamento* da Rede Bayesiana com aprendizagem de estrutura baseada em pontuação - HC

Parâmetros do vértice Acompanhamento (Distribuição Gaussiana)		
Densidade Condicional		
$f(\text{Acompanhamento} \mid \text{Sanduíche} + \text{Brinde})$		
Coeficientes	(Intercepto)	-2.20242689
	Sanduíche	0.99610525
	Brinde	0.09056828
Desvio padrão resíduos		0.08403771

Tabela 9 – Coeficientes associados ao vértice *Adicional* da Rede Bayesiana com aprendizagem de estrutura baseada em pontuação - HC

Parâmetros do vértice Adicional (Distribuição Gaussiana)		
Densidade Condicional		
$f(\text{Adicional} \mid \text{Café})$		
Coeficientes	(Intercepto)	10.91523505
	Café	0.07322855
Desvio padrão resíduos		0.7583874

Tabela 10 – Coeficientes associados ao vértice *Brinde* da Rede Bayesiana com aprendizagem de estrutura baseada em pontuação - HC

Parâmetros do vértice Brinde (Distribuição Gaussiana)		
Densidade Condicional		
$f(\text{Brinde})$		
Coeficientes	(Intercepto)	9.949161
Desvio padrão resíduos		0.4794455

Tabela 11 – Coeficientes associados ao vértice *Prato* da Rede Bayesiana com aprendizagem de estrutura baseada em pontuação - HC

Parâmetros do vértice Prato (Distribuição Gaussiana)		
Densidade Condicional		
$f(\text{Prato} \mid \text{Bebida} + \text{Brinde})$		
Coeficientes	(Intercepto)	5.5560801
	Bebida	0.5598613
	Brinde	-0.2738339
Desvio padrão resíduos		0.2143587

Tabela 12 – Coeficientes associados ao vértice *Venda* da Rede Bayesiana com aprendizagem de estrutura baseada em pontuação - HC

Parâmetros do vértice <i>Venda</i> (Distribuição Gaussiana)		
Densidade Condicional		
$f(Venda \mid Milkshake + Bebida + Adicional)$		
Coeficientes	(Intercepto)	1.58330852
	Milkshake	0.45548406
	Bebida	0.47455019
	Adicional	0.07053049
Desvio padrão resíduos		0.02270405

Tabela 13 – Coeficientes associados ao vértice *Milkshake* da Rede Bayesiana com aprendizagem de estrutura baseada em pontuação - HC

Parâmetros do vértice <i>Milkshake</i> (Distribuição Gaussiana)		
Densidade Condicional		
$f(Milkshake \mid Sanduíche + Bebida + Acompanhamento)$		
Coeficientes	(Intercepto)	0.9125164
	Sanduíche	-0.2091924
	Bebida	1.0045829
	Acompanhamento	0.2477580
Desvio padrão resíduos		0.1473436

5.3.2 Ajustes de parâmetros (Rede Bayesiana - Aprendizado de Estrutura - GS)

Tabela 14 – Coeficientes associados ao vértice *Milkshake* da Rede Bayesiana com aprendizagem de estrutura baseada em restrições - GS

Parâmetros do vértice <i>Milkshake</i> (Distribuição Gaussiana)		
Densidade Condicional		
$f(Milkshake \mid Bebida)$		
Coeficientes	(Intercepto)	0.5773822
	Bebida	1.0450475
Desvio padrão resíduos		0.1473457

Tabela 15 – Coeficientes associados ao vértice *Acompanhamento* da Rede Bayesiana com aprendizagem de estrutura baseada em restrições - GS

Parâmetros do vértice <i>Milkshake</i> (Distribuição Gaussiana)		
Densidade Condicional		
$f(Acompanhamento \mid Sanduíche)$		
Coeficientes	(Intercepto)	-1.459186
	Sanduíche	1.006752
Desvio padrão resíduos		0.09384212

Tabela 16 – Coeficientes associados ao vértice *Café* da Rede Bayesiana com aprendizagem de estrutura baseada em restrições - GS

Parâmetros do vértice Café (Distribuição Gaussiana)		
Densidade Condicional		
$f(\text{Café})$		
Coeficientes	(Intercepto)	10.34651
Desvio padrão resíduos		1.05642

Tabela 17 – Coeficientes associados ao vértice *Prato* da Rede Bayesiana com aprendizagem de estrutura baseada em restrições - GS

Parâmetros do vértice Prato (Distribuição Gaussiana)		
Densidade Condicional		
$f(\text{Prato})$		
Coeficientes	(Intercepto)	10.72032
Desvio padrão resíduos		0.4995754

Tabela 18 – Coeficientes associados ao vértice *Sanduiche* da Rede Bayesiana com aprendizagem de estrutura baseada em restrições - GS

Parâmetros do vértice Sanduíche (Distribuição Gaussiana)		
Densidade Condicional		
$f(\text{Sanduíche})$		
Coeficientes	(Intercepto)	14.82548
Desvio padrão resíduos		0.8265545

Tabela 19 – Coeficientes associados ao vértice *Brinde* da Rede Bayesiana com aprendizagem de estrutura baseada em restrições - GS

Parâmetros do vértice Brinde (Distribuição Gaussiana)		
Densidade Condicional		
$f(\text{Brinde})$		
Coeficientes	(Intercepto)	9.949161
Desvio padrão resíduos		0.4794455

Tabela 20 – Coeficientes associados ao vértice *Bebida* da Rede Bayesiana com aprendizagem de estrutura baseada em restrições - GS

Parâmetros do vértice Bebida (Distribuição Gaussiana)		
Densidade Condicional		
$f(\text{Bebida} \mid \text{Sanduíche})$		
Coeficientes	(Intercepto)	-0.6811503
	Sanduíche	0.9963611
Desvio padrão resíduos		0.06057377

Tabela 21 – Coeficientes associados ao vértice *Adicional* da Rede Bayesiana com aprendizagem de estrutura baseada em restrições - GS

Parâmetros do vértice Adicional (Distribuição Gaussiana)		
Densidade Condicional		
$f(\textit{Adicional})$		
Coeficientes	(Intercepto)	11.6729
Desvio padrão resíduos	0.7577641	

Tabela 22 – Coeficientes associados ao vértice *Venda* da Rede Bayesiana com aprendizagem de estrutura baseada em restrições - GS

Parâmetros do vértice Venda (Distribuição Gaussiana)		
Densidade Condicional		
$f(\textit{Venda} \mid \textit{Milkshake} + \textit{Adicional})$		
Coeficientes	(Intercepto)	1.8638251
	Milkshake	0.6742545
	Adicional	0.3325343
Desvio padrão resíduos	0.05853934	

RESULTADOS

Neste capítulo, são apresentados os resultados das previsões baseadas em Redes Bayesianas com aprendizado de estrutura, por meio dos algoritmos *Hill Climbing* (HC) e *Grow Shrink* (GS), assim como o comparativo com o modelo de série temporal baseada em *Generalized Additive Model* (GAM). Na Seção 6.1, são descritos os elementos que compõe a Rede Bayesiana com aprendizado de estrutura baseado no algoritmo *Hill Climbing* (HC) e suas relações de causalidade. Na Seção 6.1.1, são descritos os modelos matemáticos da previsão de vendas, gerados através da Rede Bayesiana com aprendizado de estrutura baseado no algoritmo *Hill Climbing* (HC). Na Seção 6.1.2, são apresentados os resultados da avaliação da previsão de vendas, gerado através da Rede Bayesiana com aprendizado de estrutura baseado no algoritmo *Hill Climbing* (HC) com base no conjunto de dados de teste. Na Seção 6.1.3, são descritos os procedimentos para o uso dos modelos matemáticos da previsão de vendas, gerados através da Rede Bayesiana com aprendizado de estrutura baseado no algoritmo *Hill Climbing* (HC) na prática. Na Seção 6.2, são descritos os elementos que compõe a Rede Bayesiana com aprendizado de estrutura baseado no algoritmo *Grow Shrink* (GS) e suas relações de causalidade. Na Seção 6.2.1, são descritos os modelos matemáticos da previsão de vendas, gerados através da Rede Bayesiana com aprendizado de estrutura baseado no algoritmo *Grow Shrink* (GS). Na Seção 6.2.2, são apresentados os resultados da avaliação da previsão de vendas, gerado através da Rede Bayesiana com aprendizado de estrutura baseado no algoritmo *Grow Shrink* (GS) com base no conjunto de dados de teste. Na Seção 6.2.3, são descritos os procedimentos para o uso dos modelos matemáticos da previsão de vendas, gerados através da Rede Bayesiana com aprendizado de estrutura baseado no algoritmo *Grow Shrink* (GS) na prática. Na Seção 6.3, são descritos os elementos que compõe a série temporal baseada em *Generalized Additive Model* (GAM). Na Seção 6.3.1, é descrito a formulação do modelo matemático, proposto para série temporal baseada em *Generalized Additive Model* (GAM). Na Seção 6.3.2, são apresentados os resultados da avaliação da previsão de vendas, gerado através da série temporal baseada em *Generalized Additive Model* (GAM).

6.1 Rede Bayesiana - Hill Climbing

Na Figura 35, temos a representação gráfica da Rede Bayesiana com aprendizagem de estrutura baseada em pontuação, por meio do algoritmo *Hill Climbing* (HC). A rede é formada por um grafo composto por 9 vértices e 12 arestas direcionadas. As arestas nas cores magenta, indicam as relações com maior força entre os vértices, enquanto as arestas nas cores azuis, indicam a menor força. Podemos observar que a variável *Sanduíche* influencia diretamente as variáveis *Milkshake* e *Bebida*, que geralmente estão relacionadas a composição de um combo de produtos, e consecutivamente estas variáveis influenciam a variável *Venda*. Todas as arestas possuem pesos que vão de 0 a 1, e indicam a força da aresta, sendo que as arestas nas cores magenta, indicam forças iguais ou superiores a 0.95, e as arestas em tons de azul, indicam forças inferiores a 0.95.

Figura 35 – Modelo - Rede Bayesiana - *Hill-Climbing* (HC)

Fonte: Elaborada pelo autor.

6.1.1 Modelo Matemático da Previsão de Vendas

Na Equação 6.1, temos o conjunto de modelos matemáticos que descrevem a previsão de vendas, mediante o ajuste da Rede Bayesiana com aprendizado de estrutura baseada em pontuação através do algoritmo *Hill-Climbing* (HC).

$$f_{\text{bebida}}(\text{sanduche}) = -0.681 + (0.996 * \ln(\text{sanduche})) + 0.061 \quad (6.1a)$$

$$f_{\text{acompanhamento}}(\text{sanduche}, \text{brinde}) = -2.202 + (0.996 * \text{sanduche}) + (0.091 * \ln(\text{brinde})) + 0.084 \quad (6.1b)$$

$$f_{\text{milkshake}}(\text{sanduche}, \text{bebida}, \text{acompanhamento}) = 0.913 + (-0.209 * \ln(\text{sanduche})) + (1.005 * \text{bebida}) + (0.248 * \text{acompanhamento}) + 0.147 \quad (6.1c)$$

$$f_{\text{venda}}(\text{milkshake}, \text{bebida}, \text{adicional}) = e^{1.583 + (0.455 * \text{milkshake}) + (0.475 * \text{bebida}) + (0.071 * \ln(\text{adicional}))} + 0.023 \quad (6.1d)$$

6.1.2 Avaliação da Previsão de Vendas

Na Figura 36, temos o resultado do modelo de Rede Bayesiana com aprendizado de estrutura baseado em pontuação através do algoritmo *Hill-Climbing* (HC). O gráfico apresenta valores reais observados durante o período de 2017 que foram utilizados para teste, com relação aos valores previstos pelo modelo proposto. Os valores das variáveis a priori *Sanduíche*, *Brinde* e *Adicional* foram coletados do conjunto de teste para estimar os valores das variáveis *Bebida*, *Acompanhamento*, *Milkshake* e consecutivamente a variável *Venda*.

Figura 36 – Previsão de Vendas - Rede Bayesiana - *Hill-Climbing* (HC)

Fonte: Elaborada pelo autor.

Na Tabela 23, é possível observar que alguns meses tiveram uma diferença maior entre os valores reais e previstos. Com base na rede de *fast-food* considerada na análise, existem fatores externos que possam ter impactado no modelo de previsão proposto, como por exemplo, o fechamento de algumas lojas, devido períodos de crise, bem como outros fatores. Desta forma, o modelo não conseguiu detectar cenários como esse, dado que esta variável não foi considerada na análise inicialmente.

Tabela 23 – Previsão de Vendas - Rede Bayesiana - *Hill-Climbing* (HC)

Vendas - Ano de 2017				
Mês	Real	Previsto	Dif. R\$	Dif. %
Janeiro	14.341.265,56	13.979.833,58	361.431,98	2,52
Fevereiro	11.878.234,94	10.587.232,12	1.291.002,82	10,87
Março	12.319.539,18	11.569.200,44	750.338,74	6,09
Abril	13.047.168,61	11.245.714,49	1.801.454,12	13,81
Maio	11.480.624,06	11.063.481,90	417.142,16	3,63
Junho	11.497.298,16	10.053.069,01	1.444.229,15	12,56
Julho	13.829.977,54	12.683.844,43	1.146.133,11	8,29
Agosto	10.637.173,93	9.988.251,36	648.922,57	6,10
Setembro	11.977.163,98	10.307.788,76	1.669.375,22	13,94
Outubro	12.951.892,49	11.657.647,18	1.294.245,31	9,99
Novembro	13.742.491,53	12.739.365,86	1.003.125,67	7,30
Dezembro	18.616.569,93	17.988.675,31	627.894,62	3,37

6.1.3 Uso do modelo na prática

Na prática, o uso do modelo matemático de previsão de vendas proposto na Seção 6.1.1, necessita estimar o valor das variáveis *Sanduíche*, *Brinde* e *Adicional*, para que o modelo possa encontrar os valores das demais variáveis e assim formular a previsão de vendas.

Desta forma, considerando as estimativas para variáveis *Sanduíche* = 100000, *Brinde* = 25000 e *Adicional* = 5000, temos:

$$f_{\text{bebida}}(\text{sanduiche} = 100000) = -0.681 + (0.996 * \ln(\text{sanduiche})) + 0.061 \quad (6.2a)$$

$$\begin{aligned} f_{\text{acompanhamento}}(\text{sanduiche} = 100000, \text{brinde} = 25000) = & -2.202 + (0.996 * \text{sanduiche}) \\ & + (0.091 * \ln(\text{brinde})) + 0.084 \end{aligned} \quad (6.2b)$$

$$\begin{aligned} f_{\text{milkshake}}(\text{sanduiche} = 100000, \text{bebida} = f_{\text{bebida}}(\text{args}), \\ \text{acompanhamento} = f_{\text{acompanhamento}}(\text{args})) = & 0.913 + (-0.209 * \ln(\text{sanduiche})) \\ & + (1.005 * \text{bebida}) + (0.248 * \text{acompanhamento}) + 0.147 \end{aligned} \quad (6.2c)$$

$$\begin{aligned} f_{\text{venda}}(\text{milkshake} = f_{\text{milkshake}}(\text{args}), \text{bebida} = f_{\text{bebida}}(\text{args}), \text{adicional} = 5000) = \\ e^{1.583 + (0.455 * \text{milkshake}) + (0.475 * \text{bebida}) + (0.071 * \ln(\text{adicional})) + 0.023} \end{aligned} \quad (6.2d)$$

Logo, temos que o valor da venda, considerando as estimativas para variáveis *Sanduiche* = 100000, *Brinde* = 25000 e *Adicional* = 5000, é aproximadamente R\$ 388.240,36 reais.

6.2 Rede Bayesiana - Grow Shrink

Na Figura 37, temos a representação gráfica da Rede Bayesiana com aprendizagem de estrutura baseado em restrições, por meio do algoritmo *Grow Shrink* (GS). A rede é formada por um grafo composto por 9 vértices e 5 arestas direcionadas. As arestas nas cores magenta, indicam as relações com maior força entre os vértices, enquanto as arestas nas cores azuis, indicam a menor força. Podemos observar que a variável *Sanduíche* influencia diretamente as variáveis *Bebida*, *Acompanhamento* e a variável *Bebida* por sua vez, influencia a variável *Milkshake*, sendo estas relacionadas a composição de um combo de produtos, e consecutivamente estas variáveis influenciam a variável *Venda*. Todas as arestas possuem pesos que vão de 0 a 1, e indicam a força da aresta, sendo que as arestas nas cores magenta, indicam forças iguais ou superiores a 0.95, e as arestas em tons de azul, indicam forças inferiores a 0.95.

Figura 37 – Modelo - Rede Bayesiana - *Grow Shrink* (GS)

Fonte: Elaborada pelo autor.

6.2.1 Modelo Matemático da Previsão de Vendas

Na Equação 6.3, temos o conjunto de modelos matemáticos que descrevem a previsão de vendas, mediante o ajuste da Rede Bayesiana com aprendizado de estrutura baseada em

restrições através do algoritmo *Grow Shrink* (GS).

$$f_{\text{bebida}}(\text{sanduche}) = -0.681 + (0.996 * \ln(\text{sanduche})) + 0.061 \quad (6.3a)$$

$$f_{\text{milkshake}}(\text{bebida}) = 0.577 + (1.045 * \text{bebida}) + 0.147 \quad (6.3b)$$

$$f_{\text{venda}}(\text{milkshake}, \text{adicional}) = e^{1.864 + (0.674 * \text{milkshake}) + (0.333 * \text{adicional}) + 0.059} \quad (6.3c)$$

6.2.2 Avaliação da Previsão de Vendas

Na Figura 38, temos o resultado do modelo de Rede Bayesiana com aprendizado de estrutura baseado em restrições através do algoritmo *Grow Shrink* (GS). O gráfico apresenta valores reais observados durante o período de 2017 que foram utilizados para teste, com relação aos valores previstos pelo modelo proposto. Os valores das variáveis *a priori* *Sanduíche* e *Adicional* foram coletados do conjunto de teste para estimar os valores das variáveis *Bebida*, *Milkshake* e consecutivamente a variável *Venda*.

Figura 38 – Previsão de Vendas - Rede Bayesiana - *Grow Shrink* (GS)

Fonte: Elaborada pelo autor.

Na Tabela 24, é possível observar que alguns meses tiveram uma diferença maior entre os valores reais e previstos. Com base na rede de *fast-food* considerada na análise, existem fatores externos que possam ter impactado no modelo de previsão proposto, como por exemplo, o fechamento de algumas lojas, devido períodos de crise, bem como outros fatores. Desta forma, o modelo não conseguiu detectar cenários como esse, dado que esta variável não foi considerada na análise inicialmente.

Tabela 24 – Previsão de Vendas - Rede Bayesiana - *Grow Shrink* (GS)

Venda - Ano de 2017				
Mês	Real	Previsto	Dif. R\$	Dif. %
Janeiro	14.341.265,56	13.891.668,86	449.596,69	3,13
Fevereiro	11.878.234,94	11.880.653,62	-2.418,68	-0,02
Março	12.319.539,18	11.554.235,75	765.303,43	6,21
Abril	13.047.168,61	11.796.796,61	1.250.372,00	9,58
Mai	11.480.624,06	10.768.974,54	711.649,51	6,20
Junho	11.497.298,16	10.812.350,50	684.947,65	5,96
Julho	13.829.977,54	13.454.580,34	375.397,20	2,71
Agosto	10.637.173,93	10.679.539,93	-42.366,00	-0,40
Setembro	11.977.163,98	12.874.399,12	-897.235,13	-7,49
Outubro	12.951.892,49	13.828.691,91	-876.799,41	-6,77
Novembro	13.742.491,53	14.539.150,50	-796.658,96	-5,80
Dezembro	18.616.569,93	19.753.000,06	-1.136.430,12	-6,10

6.2.3 Uso do modelo na prática

Na prática, o uso do modelo matemático de previsão de vendas proposto na Seção 6.2.1, necessita estimar o valor das variáveis *Sanduíche* e *Adicional*, para que o modelo possa encontrar os valores das demais variáveis e assim formular a previsão de vendas.

Desta forma, considerando as estimativas para variáveis *Sanduíche* = 100000 e *Adicional* = 5000, temos:

$$f_{\text{bebida}}(\text{sandui che} = 100000) = -0.681 + (0.996 * \ln(\text{sandui che})) + 0.061 \quad (6.4a)$$

$$f_{\text{milkshake}}(\text{bebida} = f_{\text{bebida}}(\text{args})) = 0.577 + (1.045 * \text{bebida}) + 0.147 \quad (6.4b)$$

$$f_{\text{venda}}(\text{milkshake} = f_{\text{milkshake}}(\text{args}), \text{adicional} = 5000) = e^{1.864 + (0.674 * \text{milkshake}) + (0.333 * \text{adicional}) + 0.059} \quad (6.4c)$$

Logo, temos que o valor da venda, considerando as estimativas para variáveis *Sanduíche* = 100000 e *Adicional* = 5000, é aproximadamente R\$ 395.133,88 reais.

6.3 Série Temporal - Generalized Additive Model (GAM)

O modelo de série temporal baseada em *Generalized Additive Model* (GAM) com ajustes Bayesianos, proposto para o comparativo, identifica e soma várias funções lineares e não lineares que resultam em uma linha de tendência que melhor se adapta aos dados, desta forma, estes ajustes são realizados de forma iterativa para reduzir o erro de previsão. O modelo considera uma abordagem de série temporal univariada e utiliza para análise a variável *Venda*. Além disso, são considerados fatores de sazonalidade mensais e anuais.

6.3.1 Modelo Matemático da Previsão de Vendas

O modelo proposto pode ser decomposto em três componentes principais: tendência, sazonalidade e feriados. Eles são combinados conforme a Equação 6.5:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \quad (6.5)$$

1. $g(t)$: curva de crescimento linear ou logística para modelar mudanças não periódicas em séries temporais.
2. $s(t)$: mudanças periódicas (por exemplo, sazonalidade semanal / mensal / anual).
3. $h(t)$: feriados (fornecido pelo usuário).
4. ε_t erro associado a alterações incomuns não ajustadas pelo modelo.

6.3.2 Avaliação da Previsão de Vendas

Na Figura 39, temos o resultado do modelo de série temporal baseada em Generalized Additive Model (GAM) com ajustes Bayesianos, desenvolvida pelos pesquisadores do *Facebook*. O gráfico apresenta valores reais observados durante o período de 2017 que foram utilizados para teste, com relação aos valores previstos pelo modelo proposto.

Figura 39 – Previsão de Vendas - Série Temporal - Generalized Additive Model (GAM)

Fonte: Elaborada pelo autor.

Na Tabela 25, é possível observar que alguns meses tiveram uma diferença maior entre os valores reais e previstos. Com base na rede de *fast-food* considerada na análise, existem fatores externos que possam ter impactado no modelo de previsão proposto, como por exemplo, o fechamento de algumas lojas, devido períodos de crise, bem como outros fatores. Desta forma, o modelo não conseguiu detectar cenários como esse, dado que o modelo proposto é baseado em uma série temporal univariada.

Tabela 25 – Previsão de Vendas - Série Temporal - Generalized Additive Model (GAM)

Venda - Ano de 2017				
Mês	Real	Previsto	Dif. R\$	Dif. %
Janeiro	14.341.265,56	12.557.854,32	1.783.411,24	12,44
Fevereiro	11.878.234,94	10.587.232,12	1.291.002,82	10,87
Março	12.319.539,18	11.569.200,44	750.338,74	6,09
Abril	13.047.168,61	11.245.714,49	1.801.454,12	13,81
Maio	11.480.624,06	11.063.481,90	417.142,16	3,63
Junho	11.497.298,16	10.053.069,01	1.444.229,15	12,56
Julho	13.829.977,54	12.683.844,43	1.146.133,11	8,29
Agosto	10.637.173,93	9.988.251,36	648.922,57	6,10
Setembro	11.977.163,98	10.307.788,76	1.669.375,22	13,94
Outubro	12.951.892,49	11.657.647,18	1.294.245,31	9,99
Novembro	13.742.491,53	12.739.365,86	1.003.125,67	7,30
Dezembro	18.616.569,93	17.988.675,31	627.894,62	3,37

6.4 Comparativo entre Modelos de Previsão de Vendas

Na Figura 40, temos o comparativo entre os modelos *Rede Bayesiana - HC*, *Rede Bayesiana - GS* e *Série Temporal - GAM* para previsão de vendas. O gráfico apresenta valores reais observados durante o período de 2017 que foram utilizados para teste, com relação aos valores previstos pelo modelo proposto.

Figura 40 – Comparativo entre Modelos de Previsão de Vendas

Fonte: Elaborada pelo autor.

Na Tabela 26, temos o comparativo entre os modelos *Rede Bayesiana - HC*, *Rede Bayesiana - GS* e *Série Temporal - GAM* para previsão de vendas. A Tabela apresenta valores reais observados durante o período de 2017 que foram utilizados para teste, com relação aos valores previstos pelo modelo proposto. Observa-se que o modelo de Rede Bayesiana com aprendizado de estrutura baseado em pontuação, através do algoritmo *Hill-Climbing* (HC), obteve os melhores resultados entre a diferença dos valores reais e previstos, sendo que seu *Mean Absolute Percentage Error* (MAPE), resultou em 2.40% e acurácia de 97.60% em relação ao conjunto de teste avaliado.

Tabela 26 – Comparativo entre Modelos de Previsão de Vendas

Venda - Ano de 2017							
Mês	Real	Previsto HC	Previsto GS	Previsto GAM	Dif. % HC	Dif. % GS	Dif. % GAM
Janeiro	14.341.265,56	13.979.833,58	13.891.668,86	12.557.854,32	2,52	3,13	12,44
Fevereiro	11.878.234,94	11.794.083,86	11.880.653,62	10.587.232,12	0,71	-0,02	10,87
Março	12.319.539,18	12.030.629,97	11.554.235,75	11.569.200,44	2,35	6,21	6,09
Abril	13.047.168,61	12.541.901,79	11.796.796,61	11.245.714,49	3,87	9,58	13,81
Maiο	11.480.624,06	10.803.431,12	10.768.974,54	11.063.481,90	5,90	6,20	3,63
Junho	11.497.298,16	10.937.465,58	10.812.350,50	10.053.069,01	4,87	5,96	12,56
Julho	13.829.977,54	13.485.842,97	13.454.580,34	12.683.844,43	2,49	2,71	8,29
Agosto	10.637.173,93	10.476.783,11	10.679.539,93	9.988.251,36	1,51	-0,40	6,10
Setembro	11.977.163,98	11.787.679,22	12.874.399,12	10.307.788,76	1,58	-7,49	13,94
Outubro	12.951.892,49	12.684.768,79	13.828.691,91	11.657.647,18	2,06	-6,77	9,99
Novembro	13.742.491,53	13.675.684,43	14.539.150,50	12.739.365,86	0,49	-5,80	7,30
Dezembro	18.616.569,93	18.542.943,48	19.753.000,06	17.988.675,31	0,40	-6,10	3,37

CONCLUSÃO

Neste trabalho foram analisados comparativamente aplicações de Redes Bayesianas e séries temporais aplicados a previsão de vendas em um rede de *fast-food*. Foram avaliadas Redes Bayesianas com aprendizado de estrutura baseado em restrições, através do algoritmo *Grow Shrink* (GS) e baseado em pontuação, através do algoritmo *Hill-Climbing* (HC), posteriormente foram comparadas com o modelo de série temporal baseada em *Generalized Additive Model* (GAM).

Nos resultados observou-se que os modelos de Redes Bayesianas com aprendizado de estrutura baseado em restrições através do algoritmo *Grow Shrink* (GS) e baseado em pontuação através do algoritmo *Hill-Climbing* (HC), apresentaram desempenho superior ao modelo de série temporal baseada em *Generalized Additive Model* (GAM), tendo erros menores com relação aos valores previstos.

A Rede Bayesiana com aprendizado de estrutura baseado em restrições, através do algoritmo *Grow Shrink* (GS) encontrou poucas relações entre as variáveis que fazem parte do conjunto de dados da rede de *fast-food*. No entanto, este modelo conseguiu detectar relações muito fortes entre a variável *Milkshake* influenciando o processo de venda, bem como a forte influência da variável *Sanduíche* com relação a variável *Bebida*, o que indica que na aquisição de sanduíche um indivíduo geralmente realiza a compra de uma bebida em conjunto. A Rede Bayesiana com aprendizado de estrutura baseado em restrições, através do algoritmo *Grow Shrink* (GS) obteve um *Mean Absolute Percentage Error* (MAPE) de 5.04% e acurácia de 94.96% em relação ao conjunto de teste avaliado.

A Rede Bayesiana com aprendizado de estrutura baseado em pontuação através do algoritmo *Hill-Climbing* (HC) encontrou as melhores relações entre as variáveis que fazem parte do conjunto de dados da rede de *fast-food*. O modelo conseguiu observar que a variável *Sanduíche* influencia diretamente as variáveis *Milkshake* e *Bebida* que geralmente estão relacionadas a composição de um combo de produtos, e consecutivamente estas variáveis influenciam a variável

Venda, bem como detectou uma relação muito forte entre a variável *Milkshake* influenciando o processo de venda. Outro aspecto interessante diz respeito a influência entre a variável *Sanduíche* e *Acompanhamento*, pois na aquisição de um sanduíche geralmente é sugerido o acompanhamento de batata frita ou porção. A Rede Bayesiana com aprendizado de estrutura baseado em pontuação através do algoritmo *Hill-Climbing* (HC) obteve um *Mean Absolute Percentage Error* (MAPE) de 2.40% e acurácia de 97.60% em relação ao conjunto de teste avaliado.

O modelo de série temporal baseada em *Generalized Additive Model* (GAM), embora tenha apresentado desempenho inferior em relação aos valores previstos comparado com as Redes Bayesianas, o modelo conseguiu realizar estimativas acima de 85% de acurácia mesmo se tratando de um modelo de série temporal univariada. Tal técnica poderia ser utilizada para estimar variáveis *a priori* das Redes Bayesianas de modo a ter um modelo totalmente independente de intervenção de especialistas da área para realizar estimativas de vendas. A série temporal baseada em *Generalized Additive Model* (GAM) obteve um *Mean Absolute Percentage Error* (MAPE) de 9.04% e acurácia de 90.96% em relação ao conjunto de teste avaliado.

Desta forma, a Rede Bayesiana com aprendizado de estrutura baseado em pontuação através do algoritmo *Hill-Climbing* (HC) foi escolhido como o melhor modelo, pois apresentou o melhor desempenho nas previsões de vendas bem como encontrou as melhores relações de causalidade entre variáveis observadas.

Uma contribuição deste trabalho está na análise comparativa do uso dos algoritmos de aprendizado de estrutura em Redes Bayesianas baseado em pontuação através do algoritmo *Hill-Climbing* (HC) e baseado em restrições através do algoritmo *Grow Shrink* (GS) para encontrar as melhores relações de variáveis contínuas para auxiliar na concepção de modelos de previsão de vendas, bem como a comparação com o modelo de série temporal baseada em *Generalized Additive Model* (GAM) desenvolvido pela equipe de pesquisadores do Facebook, intitulado "Prophet".

Outra contribuição deste trabalho está no desenvolvido do pacote para linguagem R intitulado *bnviewer*, para visualização interativa de Redes Bayesianas. Uma vez que este pacote permite a visualização interativa de Redes Bayesianas por meio de layouts customizados bem como permite realizar interações com operações de *drag and drop*, *zoom* e *click* nos vértices e arestas da rede. Além disso, o pacote permite interpretar a força das relações de causalidade entre os vértices da Rede Bayesiana analisada.

Outra contribuição deste trabalho está na produção do artigo intitulado "BrAgriNews: Um Corpus Temporal-Causal (Português-Brasileiro) para a Agricultura", por [Drury, Fernandes e Lopes \(2017\)](#). No qual faz uso de Redes Bayesianas aplicado a análise de relações causais em mineração de textos relacionados a agricultura.

Em trabalhos futuros é sugerida uma análise mais detalhada da variabilidade das partições da amostra de validação por meio de *Cross-Validation* com *K-fold* partições, a fim de se ter

uma análise da estabilidade dos modelos de Redes Bayesianas para previsão de vendas, além da análise e possibilidade de inclusão de variáveis externas que possam influenciar os modelos. Bem como também, a implementação de novas melhorias no pacote *bnviewer*, não só relacionados a Redes Bayesianas (grafos direcionados), mas também a campos aleatórios de Markov (grafos não-direcionados), a fim de contribuir com a comunidade científica que tenham interesse no estudo de modelos gráficos probabilísticos no campo de aprendizado de máquina.

REFERÊNCIAS

- BARROS, M. F. de. **Análise e Previsão de Séries Temporais Utilizando Amortecimento Exponencial com Múltiplos Ciclos e Técnicas de Simulação na Produção de Energia Eólica**. Tese (Doutorado) — PUC-Rio, 2015. Citado na página 47.
- BEINLICH, I. A.; SUERMONDT, H. J.; CHAVEZ, R. M.; COOPER, G. F. The alarm monitoring system: A case study with two probabilistic inference techniques for belief networks. In: **AIME** 89. [S.l.]: Springer, 1989. p. 247–256. Citado nas páginas 15, 52 e 118.
- DEGROOT, M.; SCHERVISH, M. **Probability and Statistics**. Addison-Wesley, 2012. ISBN 9780321500465. Disponível em: <<https://books.google.com.br/books?id=4TIEPgAACAAJ>>. Citado nas páginas 48 e 49.
- DENIS, M. J.-B. Package ‘rbmn’. 2013. Citado na página 87.
- DRURY, B.; FERNANDES, R.; LOPES, A. d. A. Bragrinews: Um corpus temporal-causal (português-brasileiro) para a agricultura. **Linguamática**, v. 9, n. 1, p. 41–54, 2017. Citado na página 106.
- DRURY, B.; VALVERDE-REBAZA, J.; MOURA, M.-F.; LOPES, A. de A. A survey of the applications of bayesian networks in agriculture. **Engineering Applications of Artificial Intelligence**, Elsevier, v. 65, p. 29–42, 2017. Citado na página 40.
- EFRON, B.; TIBSHIRANI, R. J. **An introduction to the bootstrap**. [S.l.]: CRC press, 1994. Citado nas páginas 74 e 116.
- EISUKE, K.; HARADA, M.; MIZUNO, T. Application of bayesian network to stock price prediction. **Artificial Intelligence Research**, v. 1, n. 2, p. 171, 2012. Citado nas páginas 39 e 40.
- FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. Inteligência artificial: Uma abordagem de aprendizado de máquina. **Rio de Janeiro: LTC**, v. 2, p. 192, 2011. Citado na página 47.
- FARIA, R. C. **Redes probabilísticas: aprendendo estruturas e atualizando probabilidades**. Tese (Doutorado) — Universidade de São Paulo, 2014. Citado na página 51.
- FERNANDES, R. **bnviewer: Interactive Visualization of Bayesian Networks**. [S.l.], 2018. R package version 0.1.0. Disponível em: <<https://CRAN.R-project.org/package=bnviewer>>. Citado nas páginas 52 e 74.
- GIBBONS, A. **Algorithmic graph theory**. [S.l.]: Cambridge university press, 1985. Citado na página 50.
- GÓMEZ, M. A.; VILLEGASA, P. M.; NAVARRO, H.; SUSIA, R. Dealing with uncertainty in gaussian bayesian networks from a regression perspective. **on Probabilistic Graphical Models**, p. 145, 2010. Citado na página 57.

- ISBA - International Society for Bayesian Analysis. **ISBA - International Society for Bayesian Analysis**. 1992. <<http://www.bayesian.org>>. Accessed: 2017-01-19. Citado na página 51.
- KJAERULFF, U. B.; MADSEN, A. L. Bayesian networks and influence diagrams. **Springer Science+ Business Media**, Springer, v. 200, p. 114, 2008. Citado na página 56.
- KORB, K. B.; NICHOLSON, A. E. **Bayesian artificial intelligence**. [S.l.]: CRC press, 2010. Citado na página 51.
- KUHN, M. Caret package. **Journal of Statistical Software**, v. 28, n. 5, p. 1–26, 2008. Citado na página 61.
- MARGARITIS, D. **Learning Bayesian network model structure from data**. [S.l.], 2003. Citado na página 54.
- NEWMAN, M. E. The structure and function of complex networks. **SIAM review**, SIAM, v. 45, n. 2, p. 167–256, 2003. Citado nas páginas 15 e 50.
- NIELSEN, T.; JENSEN, F. **Bayesian Networks and Decision Graphs**. Springer New York, 2013. (Information Science and Statistics). ISBN 9781475735024. Disponível em: <<https://books.google.com.br/books?id=cWLaBwAAQBAJ>>. Citado na página 56.
- NORVIG, P.; RUSSELL, S. **Inteligência Artificial: Tradução da 3a Edição**. Elsevier Brasil, 2015. ISBN 9788535251418. Disponível em: <<https://books.google.com.br/books?id=BsNeAwAAQBAJ>>. Citado na página 58.
- PEARL, J. **Probabilistic reasoning in intelligent systems: networks of plausible inference**. [S.l.]: Morgan Kaufmann, 2014. Citado nas páginas 51 e 55.
- PROVOST, F.; FAWCETT, T. Data science and its relationship to big data and data-driven decision making. **Big Data**, Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA, v. 1, n. 1, p. 51–59, 2013. Citado nas páginas 45, 46 e 47.
- SCUTARI, M. Learning bayesian networks with the bnlearn r package. **arXiv preprint arXiv:0908.3817**, 2009. Citado nas páginas 51 e 74.
- SCUTARI, M.; SCUTARI, M. M.; MMPC, H.-P. **Package ‘bnlearn’**. 2017. Citado na página 74.
- TAMBE, P. Big data investment, skills, and firm value. **Management Science**, INFORMS, v. 60, n. 6, p. 1452–1469, 2014. Citado na página 46.
- TAYLOR, S. J.; LETHAM, B. Forecasting at scale. **The American Statistician**, Taylor & Francis, n. just-accepted, 2017. Citado nas páginas 41, 42 e 43.
- Toda Matéria, Fast-Food. **Fast-Food**. 2017. <<https://www.todamateria.com.br/fast-food/>>. Accessed: 2018-01-19. Citado na página 29.
- WEISS, G. M.; PROVOST, F. Learning when training data are costly: The effect of class distribution on tree induction. **Journal of Artificial Intelligence Research**, v. 19, p. 315–354, 2003. Citado na página 46.
- WITTEN, I. H.; FRANK, E.; HALL, M. A.; PAL, C. J. **Data Mining: Practical machine learning tools and techniques**. [S.l.]: Morgan Kaufmann, 2016. Citado na página 45.

BNVIEWER - AN R PACKAGE FOR INTERACTIVE VISUALIZATION OF BAYESIAN NETWORKS

O desenvolvimento deste pacote ocorreu devido as limitações encontradas em pacotes de visualização para Redes Bayesianas, por meio de algoritmos de aprendizado de estrutura automática em Redes Bayesianas na linguagem R. Os pacotes encontrados forneciam visualizações estáticas e de difícil interpretação das redes, principalmente em situações onde haviam muitos vértices e arestas nos modelos.

A.1 Introdução

O *bnviewer* <<http://robsonfernandes.net/bnviewer>>, é um pacote desenvolvido para linguagem R, para visualização interativa de Redes Bayesianas baseado nos pacotes *bnlearn* e *visNetwork*. Ele foi lançado em 31 de Julho de 2018 e ultrapassou mais dois mil downloads em seis meses de lançamento. O pacote *bnviewer* interpreta vários algoritmos de aprendizado de estrutura fornecidos pelo pacote *bnlearn*, tais como:

1. Algoritmos de aprendizado de estrutura baseada em restrições

- a) PC
- b) *Grow-Shrink* (GS)
- c) *Incremental Association Markov Blanket* (IAMB)
- d) *Fast Incremental Association* (Fast-IAMB)
- e) *Interleaved Incremental Association* (Inter-IAMB);
- f) *Max-Min Parents & Children* (MMPC)

g) *Semi-Interleaved Hiton-PC* (SI-HITON-PC)

2. Algoritmos de aprendizado de estrutura baseados em pontuação

a) *Hill Climbing* (HC)

b) *Tabu Search* (Tabu)

3. Algoritmos de aprendizado de estrutura híbrida

a) *Max-Min Hill Climbing* (MMHC)

b) *General 2-Phase Restricted Maximization* (RSMAX2)

4. Algoritmos de descoberta local

a) *Chow-Liu*

b) *ARACNE*

E, permite visualizá-los de forma interativa, por meio de layouts customizados, bem como realizar interações com operações de *drag and drop*, *zoom* e *click* nos vértices e arestas da rede. Além disso, o pacote *bnviewer* permite interpretar a força das relações de causalidade entre os vértices da Rede Bayesiana analisada.

A.2 Instalação

A versão estável do *bnviewer* encontra-se no repositório oficial do R, *Comprehensive R Archive Network* (CRAN). O Código 1, contém os procedimentos para baixar, instalar e carregar a versão atual.

Código-fonte 1 – Instalação via CRAN

```
1: install.packages("bnviewer")
```

O pacote *bnviewer* também está disponível para desenvolvedores, a instalação deve ser realizada por meio do GitHub.

Código-fonte 2 – Instalação via GitHub

```
1: install.packages("devtools")
2: devtools::install_github("robson-fernandes/bnviewer")
```

A.3 Como utilizar o bnviewer

Importe os pacotes *bnlearn* e *bnviewer*, conforme o Código 3.

Código-fonte 3 – Importação dos pacotes necessários

```
1: library(bnlearn)
2: library(bnviewer)
```

Importe o conjunto de dados desejado e aplique um algoritmo de aprendizado de estrutura. Exemplo (*Hill-Climbing* (HC)), conforme o Código 4.

Código-fonte 4 – Instalação via GitHub

```
1: data("alarm")
2: bn.learn.hc = hc(alarm)
```

Chame a função **viewer** com os parâmetros desejados, conforme o Código 5.

Código-fonte 5 – Visualização Interativa da Rede Bayesiana

```
1: viewer(bn.learn.hc,
2:   bayesianNetwork.width = "100%",
3:   bayesianNetwork.height = "800px",
4:   bayesianNetwork.layout = "layout_with_sugiyama",
5:   bayesianNetwork.title="Discrete Bayesian Network - Alarm",
```


Exemplo de visualização de Rede Bayesiana com vértices customizados e layout em grade.

Código-fonte 6 – Visualização Interativa da Rede Bayesiana - Layout em Grade

[illegible]

Na Figura 43, temos a visualização interativa da Rede Bayesiana Discreta - Alarme com layout em grade. Na Figura 44, temos o *zoom* na Rede Bayesiana para visualizar as relações causais entre vértices.

Figura 43 – Monitoramento de pacientes de atendimento de emergência. Rede Bayesiana Discreta - Alarme, layout em grade.

Figura 44 – Monitoramento de pacientes de atendimento de emergência. Rede Bayesiana Discreta - Alarme, layout em grade. Zoom na Rede Bayesiana para visualizar as relações causais entre os vértices.

Para visualizar a força das arestas entre vértices da Rede Bayesiana, é necessário utilizar a função *boot.strength* do pacote *bnlearn*. Esta função utiliza *bootstrap* não paramétrico para avaliar a força e a direções das arestas. O método *bootstrap*, definido por [Efron e Tibshirani \(1994\)](#), é um método de reamostragem baseado na construção de sub-amostras a partir de uma amostra inicial. O método *bootstrap* pode ser construído com abordagem paramétrica ou não-paramétrica. No *bootstrap* paramétrico, a distribuição é realizada por meio de suposições, e assim, posteriormente reamostramos observações da distribuição postulada, mas usando os valores das estimativas dos parâmetros no processo de geração de pseudo-amostras. No *bootstrap* não-paramétrico, o processo de reamostragem se dá a partir da função de distribuição empírica dos dados (ou dos resíduos quando há uma estrutura de regressão).

Posteriormente, é possível construir uma Rede Bayesiana com base na avaliação de forças e direção das arestas fornecidas pela função *boot.strength*. Uma vez, tendo a estrutura da Rede Bayesiana, é possível visualizá-la por meio da função *strength.viewer* do pacote *bnviewer*, conforme o Código 7.

Código-fonte 7 – Visualização Interativa da Rede Bayesiana - Avaliação da força das relações de causalidade entre os vértices

```
1: bayesianNetwork.boot.strength = boot.strength(alarm, R = 20,
    algorithm = "hc")
2:
3: avg.bayesianNetwork = averaged.network(bayesianNetwork.boot.
    strength, threshold = 0.2)
4:
5: strength.viewer(
6:   avg.bayesianNetwork,
7:   bayesianNetwork.boot.strength,
8:   bayesianNetwork.background = "transparent",
9:   bayesianNetwork.arc.strength.threshold.expression = c("
    @threshold > 0 & @threshold < 0.5",
10:                                                         "
    @threshold >= 0.5 & @threshold <= 0.8",
11:                                                         "
    @threshold > 0.8 & @threshold <= 1"),
12:
13:   bayesianNetwork.arc.strength.threshold.expression.color = c(
    "red", "yellow", "blue"),
14:   bayesianNetwork.arc.strength.threshold.alternative.color = "
    white",
15:
16:   bayesianNetwork.arc.strength.label = TRUE,
17:   bayesianNetwork.arc.strength.label.prefix = "",
18:   bayesianNetwork.arc.strength.label.color = "black",
19:
20:   bayesianNetwork.arc.strength.tooltip = TRUE,
21:
22:   bayesianNetwork.edge.scale.min = 1,
23:   bayesianNetwork.edge.scale.max = 3,
24:
25:   bayesianNetwork.edge.scale.label.min = 14,
26:   bayesianNetwork.edge.scale.label.max = 14,
27:
28:   bayesianNetwork.width = "100%",
29:   bayesianNetwork.height = "800px",
30:   bayesianNetwork.layout = "layout_with_sugiyama",
31:   node.colors = list(background = "black",
32:                       border = "#2b7ce9",
```

```

33:             highlight = list(background = "#e91eba",
34:                               border = "#2b7ce9")),
35:
36:   node.font = list(color = "black", face="Arial"),
37:   edges.dashes = FALSE
38: )

```

Na Figura 45, temos a visualização interativa da Rede Bayesiana discreta baseada no conjunto de dados do sistema de monitoramento de alarme (sintético) definido por [Beinlich et al. \(1989\)](#). A Rede Bayesiana projetada para fornecer um sistema de mensagens de alarme para monitoramento do paciente. Nesta visualização é possível avaliar a força das relações de causalidade entre os vértices. Na Figura 46, temos o *zoom* na Rede Bayesiana para visualizar as relações causais entre vértices.

Na Rede Bayesiana representada pelas Figuras, 45 e 46, os vértices são apresentados na cor preta, e ao realizar o evento *click* sobre um vértice, o mesmo se apresenta na cor magenta, dando ênfase ao mesmo, e os demais vértices são inativados visualmente indicados na cor cinza. As arestas com força no intervalo, $threshold > 0$ and $threshold < 0.5$, estão indicadas na cor vermelha, arestas com força no intervalo, $threshold \geq 0.5$ and $threshold < 0.8$, estão indicadas na cor amarela e arestas com força no intervalo, $threshold \geq 0.8$ and $threshold < 1$, estão indicadas na cor azul.

Figura 45 – Monitoramento de pacientes de atendimento de emergência. Rede Bayesiana Discreta - Alarme. Avaliação da força das relações de causalidade entre os vértices.

Figura 46 – Monitoramento de pacientes de atendimento de emergência. Rede Bayesiana Discreta - Alarme. Avaliação da força das relações de causalidade entre os vértices. *Zoom* na Rede Bayesiana para visualizar as relações causais entre os vértices.