



**Anderson da Silva Ramos**

**Aplicação de *business analytics* para  
seleção de indicadores e identificação de  
seus relacionamentos em um sistema de  
mensuração de desempenho**

**Dissertação de Mestrado**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-Graduação em Engenharia de Produção do Departamento de Engenharia Industrial da PUC-Rio.

Orientador: Prof. Fernando Luiz Cyrino Oliveira  
Coorientadora: Prof. Cristina Márcia Barros de Castro

Rio de Janeiro  
agosto de 2020



**Anderson da Silva Ramos**

**Aplicação de *business analytics* para  
seleção de indicadores e identificação de  
seus relacionamentos em um sistema de  
mensuração de desempenho**

Dissertação apresentada como requisito parcial para  
obtenção do grau de Mestre pelo Programa de Pós-  
Graduação em Engenharia de Produção do  
Departamento de Engenharia Industrial da PUC-Rio.  
Aprovada pela Comissão Examinadora abaixo:

**Prof. Fernando Luiz Cyrino Oliveira**

Orientador

Departamento de Engenharia Industrial - PUC-Rio

**Prof. Cristina Márcia Barros de Castro**

Coorientadora

Universidade Federal de Juiz de Fora - UFJF

**Prof. Fernanda Baião**

Departamento de Engenharia Industrial - PUC-Rio

**Prof. José Eugênio Leal**

Departamento de Engenharia Industrial - PUC-Rio

Todos os direitos reservados. A reprodução, total ou parcial, do trabalho é proibida sem a autorização da universidade, do autor e dos orientadores.

### Anderson da Silva Ramos

Graduou-se em Engenharia de Produção pela Universidade Federal de Juiz de Fora (UFJF) no ano de 2017. Possui formação técnica em Metalurgia pelo Instituto Federal de Educação, Ciência e Tecnologia do Sudeste de Minas Gerais - Campus Juiz de Fora. Trabalha na área de pesquisa sobre métodos quantitativos de apoio à decisão, *business analytics*, sistemas de mensuração de desempenho e gerenciamento de sistemas logísticos, sobretudo vinculados ao transporte ferroviário de cargas.

#### Ficha Catalográfica

Ramos, Anderson da Silva

Aplicação de *business analytics* para seleção de indicadores e identificação de seus relacionamentos em um sistema de mensuração de desempenho / Anderson da Silva Ramos ; orientador: Fernando Luiz Cyrino Oliveira ; coorientadora: Cristina Márcia Barros de Castro. – 2020.

53 f. : il. color. ; 30 cm

Dissertação (mestrado)–Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Industrial, 2020.

Inclui bibliografia

1. Engenharia Industrial – Teses. 2. Sistema de mensuração de desempenho. 3. *Business analytics*. 4. Seleção de variáveis. 5. Mapeamento de relações. I. Oliveira, Fernando Luiz Cyrino. II. Castro, Cristina Márcia Barros de. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Industrial. IV. Título.

CDD: 658.5

Aos meus pais, Adalton e Maria Heli, e ao meu irmão,  
Adalton Júnior.

## Agradecimentos

Agradeço, primeiramente, a Deus, autor e senhor de minha vida, por me conceder tudo aquilo que sou, tenho e vivo, permitindo-me alcançar junto com meus familiares e amigos essa grande e importante conquista.

Aos meus pais, Adalton e Maria Heli, pelo amor incondicional, traduzido desde o primeiro “sim” ao querer de Deus por minha existência até a doação diária de suas vidas como base para que eu possa caminhar. Ao meu amado irmão, Adalton Júnior, maravilhoso amigo e companheiro em todas as circunstâncias de minha história.

A todos os meus parentes e amigos, pelo carinho e pelas orações, essenciais, como sempre, nesta jornada.

Ao meu orientador, Fernando Cyrino, e à minha coorientadora, Cristina Castro, por toda a dedicação e atenção dispendidas para a construção deste trabalho.

A todos os docentes, a começar por aqueles de minha infância até aqueles que estiveram presentes diretamente neste curso de mestrado, pelos conhecimentos e experiências a mim transmitidos, que não só contribuíram para minha formação profissional, mas, com certeza, também para a pessoal.

Aos colaboradores da empresa que possibilitaram o desenvolvimento do estudo prático, por essa oportunidade.

Enfim, a todos que, de algum modo, fizeram e continuarão a fazer parte de minha história, pelo auxílio na caminhada, pois é certo de que sozinho ninguém é capaz de superar os desafios mais difíceis e alcançar grandes vitórias. Deus os abençoe!

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

## Resumo

Ramos, Anderson da Silva; Oliveira, Fernando Luiz Cyrino (Orientador). **Aplicação de *business analytics* para seleção de indicadores e identificação de seus relacionamentos em um sistema de mensuração de desempenho.** Rio de Janeiro, 2020. 53p. Dissertação de Mestrado – Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro.

Os sistemas de mensuração de desempenho buscam acompanhar o alcance dos objetivos estratégicos a partir de um conjunto de indicadores que suportem os processos de tomada de decisão. Várias iniciativas, entretanto, têm se mostrado ineficazes devido à subjetividade no desdobramento desses objetivos em indicadores. Métodos de *business analytics* vêm sendo utilizados para auxiliar esse desdobramento via análise de dados, com maior geração de valor para as organizações. O presente trabalho apresenta a aplicação das técnicas de *Random Forest* e *Bayesian Belief Network* para, respectivamente, selecionar indicadores e mapear suas relações em um estudo prático numa empresa do setor de transporte ferroviário de cargas, com foco no suporte ao indicador de disponibilidade de locomotivas. Para o processo de seleção de variáveis, observou-se que o algoritmo *Variable Selection Using Random Forest* obteve o melhor desempenho em acurácia e tempo de processamento. Na elaboração do mapa estratégico, a combinação do algoritmo *Tabu Search* com o critério estatístico *Bayesian Information Criteria* levou à escolha de um arranjo parcimonioso em suas relações, aderente à expectativa inicial associada ao critério estatístico utilizado. Foi observado um significativo vínculo entre a disponibilidade de locomotivas e indicadores operacionais da empresa em estudo, revelando o potencial de influência do modelo operacional nos resultados da disponibilidade. Verifica-se a oportunidade de emprego de técnicas de séries temporais para a previsão de desempenho a partir dos relacionamentos entre indicadores, bem como para aperfeiçoar a fase de seleção de variáveis, com a identificação de possíveis defasagens existentes nesses relacionamentos.

## Palavras-chave

Sistema de mensuração de desempenho; *business analytics*; seleção de variáveis; mapeamento de relações

## Abstract

Ramos, Anderson da Silva; Oliveira, Fernando Luiz Cyrino (Advisor). **Application of business analytics to select indicators and identify their relationships in a performance measurement system.** Rio de Janeiro, 2020. 53p. Dissertação de Mestrado – Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro.

Performance measurement systems seek to monitor the achievement of strategic objectives through a set of indicators that support decision-making processes. Several initiatives, however, have been shown to be ineffective due to the subjectivity in the unfolding of these objectives into indicators. Business analytics methods have been used to assist this deployment via data analysis, with greater value generation for organizations. The present work presents the application of Random Forest and Bayesian Belief Network techniques to, respectively, select indicators and map their relationships in a practical study in a company in the rail freight sector, with a focus on supporting the locomotive availability indicator. For the variable selection process, it was observed that the Variable Selection Using Random Forest algorithm obtained the best performance in accuracy and computation time. In the preparation of the strategic map, the combination of the Tabu Search algorithm with the Bayesian Information Criteria statistical criterion led to the choice of a parsimonious arrangement in its relations, adhering to the initial expectation associated with the statistical criterion used. A significant link was observed between the locomotive availability and operational indicators of the company under study, revealing the potential influence of the operational model on the availability results. There is an opportunity to use time series techniques to predict performance based on the relationships between indicators, as well as to improve the variable selection phase, with the identification of possible lags in these relationships.

## Keywords

Performance measurement system; business analytics; variable selection; relationship mapping

# Sumário

1. Introdução	14
2. Metodologia	17
3. Revisão da literatura	19
4. Fundamentos teóricos	25
4.1 <i>Random Forest</i>	25
4.2 <i>Bayesian Belief Network</i>	29
5. Estudo prático	32
5.1 Escolha de objetivos e indicadores e levantamento de dados históricos	32
5.2 Seleção de indicadores via <i>random forest</i>	34
5.3 Mapeamento de relacionamentos via <i>bayesian belief network</i>	42
5.4 Discussões	46
6. Conclusão	48
7. Referências bibliográficas	51

## Lista de figuras

Figura 1 - Erros médios do método RFCV (não recursivo e recursivo)	37
Figura 2 - Importância das variáveis e erros médios do método VSURF	39
Figura 3 - Erros de predição a partir dos preditores selecionados	41

## Lista de quadros

Quadro 1 - Referências de métodos de pesquisa operacional e gestão	20
Quadro 2 - Referências de métodos de <i>business analytics</i>	22
Quadro 3 - Principais passos das abordagens de seleção de variáveis	28
Quadro 4 - Lista de indicadores considerados no estudo prático	33
Quadro 5 - Parâmetros dos métodos de seleção de variáveis	35
Quadro 6 - Tempos de processamento na seleção de variáveis	36
Quadro 7 - Preditores selecionados nos métodos RFCV e VSURF	40
Quadro 8 - Parâmetros do algoritmo <i>tabu search</i>	42
Quadro 9 - Grafos com base nos parâmetros “tabu” e “max.tabu”	43

## Lista de tabelas

Tabela 1 - Medidas de processamento computacional no processo de *structure learning* 44

Tabela 2 - Redução do critério BIC pela remoção de arcos do grafo G3 45

## Lista de abreviaturas e siglas

AHP - *Analytic Hierarchy Process*

ANP - *Analytic Network Process*

ARIMA - *Autoregressive Integrated Moving Average*

BF - *Bayes Factor*

BIC - *Bayesian Information Criteria*

BPA - *Business Performance Analytics*

BSC - *Balanced Scorecard*

CART - *Classification and Regression Trees*

CLP - *Causal Loop Diagram*

DEA - *Data Envelopment Analysis*

DMU - *Decision Making Unit*

KPI - *Key Performance Indicator*

MACBETH - *Measuring Attractiveness by a Categorical-Based Evaluation Technique*

MLP - *Multilayer Perceptron*

MSE - *Mean Squared Error*

OOB - *Out-Of-Bag*

PCA - *Principal Component Analysis*

PLS - *Partial Least Squares*

PMA - *Performance Management Analytics*

PMS - *Performance Measurement System*

QMPMS - *Quantitative Model for Performance Measurement System*

QRPMS - *Quantitative Relationships at the Performance Measurement System*

RFCV - *Random Forest Cross-Validation*

SEM - *Structural Equation Model*

SVM - *Support Vector Machines*

VIF - *Variance Inflation Factor*

VSURF - *Variable Selection Using Random Forest*



# 1 Introdução

Os sistemas de mensuração de desempenho (PMSs, do inglês *performance measurement system*) buscam definir objetivos a partir da estratégia organizacional e desdobrá-los em um conjunto de indicadores (KPIs, do inglês *key performance indicator*) a serem acompanhados e cujos resultados devem apontar se os objetivos pretendidos estão sendo alcançados (Rodriguez-Rodriguez et al., 2009). Esses sistemas têm despertado grande interesse nos meios acadêmico e empresarial ao longo das últimas décadas, gerando diferentes estudos acerca de modelos para sua aplicação. Entretanto, várias iniciativas de implementação vêm apresentando-se limitadas quanto ao potencial de suporte aos processos de tomada de decisão (Santos et al., 2018). Muitas vezes, tem-se significativa dúvida se os indicadores escolhidos realmente são os mais adequados para fornecer as informações necessárias ao mapeamento das relações existentes entre os objetivos corporativos (Peral et al., 2017).

O modelo *Balanced Scorecard* (BSC), proposto por Kaplan & Norton (1992), é uma abordagem clássica para a confecção de um PMS, com atenção para o desdobramento dos objetivos corporativos em indicadores de desempenho abrangendo não só a perspectiva financeira, mas também outros horizontes de desempenho da organização (clientes, processos internos e aprendizado e crescimento). Embora seja largamente conhecido e utilizado, esse modelo é significativamente dependente da subjetividade advinda da opinião de especialistas e carente de métodos quantitativos para seleção de KPIs, identificação de relacionamentos entre eles e os objetivos estratégicos e projeção de cenários futuros (Wang et al., 2018). A subjetividade contribui para que indicadores e relações não efetivas à estratégia organizacional sejam acompanhados, prejudicando a efetividade esperada do PMS (Schláfke et al., 2013). Por outro lado, muitas relações de interdependência entre KPIs e objetivos ficam ocultas e poderiam fornecer informações relevantes para a tomada de decisão no contexto do gerenciamento de desempenho (Rodriguez-Rodriguez et al., 2009).

Abordagens considerando o uso de técnicas de pesquisa operacional e gestão têm sido propostas como forma de aperfeiçoar, em termos de aplicação, o arcabouço conceitual apresentado no modelo BSC (Santos et al., 2018). No entanto, muitos modelos ainda requerem ponderações subjetivas em algumas etapas e não conseguem extrair informações diretamente dos dados históricos dos indicadores (Wang & Chien, 2016). Dessa forma, tem-se a necessidade de novos trabalhos relativos à seleção de KPIs e à identificação de suas relações de interdependência, sobretudo no contexto em que as organizações têm acesso a um volume crescente de dados, os quais possuem um enorme potencial para geração de valor aos negócios e devem ser trabalhados dentro dos sistemas de mensuração de desempenho (Peral et al., 2017; Mello & Martins, 2019).

Observa-se, neste sentido, uma oportunidade de emprego de métodos de *business analytics*, os quais contribuem para a ampliação da capacidade de análise de forma objetiva e, conseqüentemente, para o aprimoramento do uso dos sistemas de mensuração de desempenho (Mello & Martins, 2019). Em linhas gerais, esses métodos se baseiam na utilização de ferramentas estatísticas e de tecnologia da informação para não só descrever resultados históricos e diagnosticar as relações entre variáveis que os motivaram, mas também possibilitar análises de predição de desempenho, bem como a prescrição de ações a partir dos resultados e relações identificados (Wang et al., 2018).

As técnicas de *business analytics* tendem a propiciar melhor compreensão da dinâmica organizacional, suportando processos de tomada de decisão orientados pela análise de dados, que podem gerar ganhos competitivos não visualizados em outras abordagens de PMS (Schláfke et al., 2013). Apesar de seu potencial em preencher as lacunas existentes no processo de implementação dos sistemas de mensuração de desempenho, o emprego das técnicas de *business analytics* é ainda pouco explorado para esse fim (Schláfke et al., 2013; Mello & Martins, 2019).

Nesse contexto, o presente trabalho avalia a utilização das técnicas de *business analytics* em PMSs, verificando as abordagens já existentes na literatura e apresentando uma aplicação prática que abranja os processos de seleção de indicadores e identificação de relacionamentos e suporte para com os objetivos estratégicos no PMS de uma empresa do setor de transporte ferroviário de cargas. Tem-se por principal objetivo responder à pergunta: “Como empregar métodos de

*business analytics* na seleção de KPIs e na avaliação das relações existentes entre eles e os objetivos estratégicos dentro de sistemas de mensuração de desempenho?”. Espera-se, desta forma, contribuir para a literatura acerca de métodos quantitativos para a implementação efetiva dos sistemas de mensuração de desempenho, a partir da confecção de desdobramentos que melhor suportem os processos de tomada de decisão. Pretende-se colaborar também para a literatura no que tange à aplicação das técnicas de *business analytics* em processos estratégicos, no caso, via PMS, e ao seu consequente potencial de geração de valor para as organizações.

Quanto à estrutura, este trabalho é composto por seis capítulos, incluindo essa introdução. No Capítulo 2, é apresentada a metodologia aplicada para o desenvolvimento do estudo. No Capítulo 3, tem-se a descrição dos principais trabalhos identificados na fase de revisão da literatura acerca de métodos para a seleção de indicadores e o mapeamento de suas relações, com a escolha da abordagem direcionadora para o estudo prático. No Capítulo 4, discorre-se a respeito dos principais fundamentos teóricos relativos às técnicas de *business analytics* presentes na abordagem direcionadora. No Capítulo 5, é feito o detalhamento dos objetivos e indicadores avaliados na empresa em estudo, bem como dos procedimentos aplicados e resultados alcançados, discutindo os aspectos principais com base nos fundamentos teóricos anteriormente apresentados. Por fim, no Capítulo 6, é feita a consolidação das observações mais relevantes extraídas ao longo do estudo, com a proposição de trabalhos futuros.

## 2 Metodologia

Tendo em vista o objetivo de apresentar o emprego de métodos de *business analytics* para a seleção de indicadores e o mapeamento de suas relações no contexto dos sistemas de mensuração de desempenho, foram estabelecidas cinco principais fases para o desenvolvimento deste trabalho, as quais estão listadas a seguir:

- (i). Efetuar revisão da literatura sobre as abordagens de cunho quantitativo para a seleção de indicadores e a identificação de relações de interdependência e suporte para com os objetivos estratégicos em sistemas de mensuração de desempenho;
- (ii). Selecionar abordagens completas ou técnicas relativas a *business analytics* a serem utilizadas como direcionadoras para um estudo prático em uma empresa do setor de transporte ferroviário de cargas;
- (iii). Identificar os principais fundamentos teóricos associados às técnicas direcionadoras de *business analytics* com vista a embasar os procedimentos e análises do estudo prático;
- (iv). Aplicar as abordagens ou técnicas selecionadas para a escolha de KPIs e a identificação de relacionamentos no contexto do PMS da empresa em estudo;
- (v). Descrever e analisar os resultados alcançados a partir do emprego das abordagens ou técnicas de *business analytics* dentro do estudo prático.

Inicialmente, foi feita a avaliação de abordagens que tratam da seleção de indicadores e da identificação de relacionamentos no desdobramento dos objetivos corporativos em sistemas de mensuração de desempenho. A revisão da literatura foi conduzida através da análise de publicações disponíveis na base de dados *Scopus*, com a busca efetuada nos filtros relativos a título, resumo e palavras-chave pela combinação de dois grupos de termos, tendo-se a necessidade de que, ao menos, um termo de cada grupo estivesse presente nas publicações. O

primeiro grupo foi composto pelos termos "*performance measurement system*", "*performance management system*" e "*key performance indicator*", enquanto que o segundo por "*quantitative model*", "*operational research*", "*business analytics*" e "*big data*". Os materiais resultantes foram avaliados, inicialmente, pelo título e/ou resumo, sendo aqueles considerados mais representativos para o escopo deste trabalho analisados na íntegra pelo autor. Utilizou-se também a busca a partir das citações feitas entre as publicações para identificar outros trabalhos relevantes a este projeto.

Com base nos materiais avaliados, buscou-se escolher uma abordagem ampla nas vertentes de *business analytics* que discorresse sobre as fases de seleção de indicadores e mapeamento de suas relações, tendo em vista o direcionamento das atividades do estudo prático. Após essa escolha, trabalhou-se na identificação e análise dos principais fundamentos teóricos que subsidiassem a aplicação dos métodos com os dados da empresa em estudo, novamente através de publicações disponíveis na base de dados *Scopus*.

O estudo prático foi realizado em uma empresa do setor de transporte ferroviário de cargas, desde a fase de escolha dos objetivos e indicadores a serem analisados, considerando a estrutura do PMS existente na empresa. Em seguida, foram conduzidos a coleta e o tratamento das bases históricas dos indicadores para a posterior aplicação das técnicas de *business analytics*, tendo por base a abordagem direcionadora e os principais fundamentos associados. A partir dos resultados de aplicação, pôde-se identificar os indicadores mais relevantes aos objetivos e apresentar as relações existentes via mapa estratégico, bem como avaliar o desempenho dos métodos empregados.

### 3 Revisão da literatura

Considerando abordagens de cunho mais quantitativo para a estruturação de PMSs, pode-se efetuar uma divisão da literatura em duas grandes vertentes, uma abrangendo métodos de pesquisa operacional e gestão e outra de *business analytics*. A primeira contempla publicações que se estendem desde os primeiros trabalhos acerca da seleção de indicadores de desempenho e da identificação de relacionamentos de forma quantitativa até aplicações nos dias atuais. Já a vertente de *business analytics* é mais nova, com trabalhos sendo publicados especialmente na última década.

Para os métodos de pesquisa operacional e gestão, Santos et al. (2018) apresentaram uma breve revisão acerca dos principais modelos já publicados, segregando-os em dois grandes grupos, um que utiliza métodos originalmente não concebidos para o contexto de PMSs, mas que combinados podem ser aplicados para esse fim, e outro que contempla o uso de técnicas voltadas para a análise de desempenho, como o DEA (do inglês, *Data Envelopment Analysis*), em conjunto com outros métodos de pesquisa operacional. Essa revisão serve como base para a exposição de trabalhos dentro da vertente de pesquisa operacional e gestão efetuada nos dois parágrafos seguintes, com a restrição em discorrer apenas sobre referências do primeiro grupo, tendo em vista que as abordagens vinculadas à técnica DEA se direcionam mais para a análise comparativa entre unidades de desempenho (DMUs, do inglês *Decision Making Unit*), o que excede ao escopo desta dissertação. No Quadro 1, são apresentadas as principais referências na vertente de métodos de pesquisa operacional e gestão.

Uma das primeiras abordagens de pesquisa operacional e gestão para PMSs foi apresentada por Suwignjo et al. (2000), baseada na combinação de mapas cognitivos, diagramas de causa e efeito e diagramas de árvore com o método AHP (do inglês, *Analytic Hierarchy Process*). Essa abordagem foi intitulada QMPMS (do inglês, *Quantitative Model for Performance Measurement System*). Os mapas e diagramas se direcionam para a identificação de fatores de desempenho e a organização de suas relações de forma hierárquica, enquanto o método AHP

possibilita a quantificação dos efeitos de cada fator na performance final. Bititci et al. (2001) apresentaram a aplicação dessa abordagem considerando análises de cenários para o posicionamento estratégico. Sarkis (2003) propôs um modelo ajustado com a substituição do método AHP pelo ANP (do inglês, *Analytic Network Process*), dado que este é menos susceptível a reversões de classificação e permite o cálculo combinado do efeito dos fatores na performance final.

Quadro 1 - Referências de métodos de pesquisa operacional e gestão

Referência	Título	Periódico	Abordagem / Técnicas
Suwignjo et al. (2000)	<i>Quantitative models for performance measurement system</i>	<i>International Journal of Production Economics</i>	QMPMS (AHP, Mapa Cognitivo, Diagrama de Causa e Efeito e Diagrama de Árvore)
Bititci et al. (2001)	<i>Strategy management through quantitative modelling of performance measurement systems</i>	<i>International Journal of Production Economics</i>	QMPMS e Análise de Cenários
Sarkis (2003)	<i>Quantitative models for performance measurement systems - alternate considerations</i>	<i>International Journal of Production Economics</i>	QMPMS (ANP, Mapa Cognitivo, Diagrama de Causa e Efeito e Diagrama de Árvore)
Santos et al. (2002)	<i>Adding value to performance measurement by using system dynamics and multicriteria analysis</i>	<i>International Journal of Operations &amp; Production Management</i>	CLP, Mapeamento Oval e Análise de Decisão Multicritério
Santos et al. (2008)	<i>Enhanced performance measuring using OR: a case study</i>	<i>Journal of the Operational Research Society</i>	
Santos et al. (2018)	<i>Measuring organisational performance using a mix of OR methods</i>	<i>Technological Forecasting &amp; Social Change</i>	
Ferreira et al. (2011)	<i>Adding value to bank branch performance evaluation using cognitive maps and MCDA: a case study</i>	<i>Journal of the Operational Research Society</i>	Mapa Cognitivo e MACBETH
Joshi et al. (2011)	<i>A Delphi-AHP-TOPSIS based benchmarking framework for performance improvement of a cold chain</i>	<i>Expert Systems with Applications</i>	Delphi

Nota: QMPMS - *Quantitative Model for Performance Measurement System*; AHP - *Analytic Hierarchy Process*; ANP - *Analytic Network Process*; CLP - *Causal Loop Diagram*; MACBETH - *Measuring Attractiveness by a Categorical-Based Evaluation Technique*.

Santos et al. (2002, 2008, 2018), por sua vez, trataram da aplicação de técnicas como CLP (do inglês, *Causal Loop Diagram*), mapeamento oval e análise de decisão multicritério na seleção de indicadores relevantes e na

mensuração de relações de dependência. Ferreira et al. (2011) empregaram o mapeamento cognitivo junto ao método MACBETH (do inglês, *Measuring Attractiveness by a Categorical-Based Evaluation Technique*), sendo este usado para a quantificação dos relacionamentos entre fatores. Joshi et al. (2011), dentre outros aspectos, apresentaram a aplicação do método Delphi para o processo de definição de indicadores, a partir da convergência das opiniões de especialistas.

Embora os métodos de pesquisa operacional e gestão já contribuam para a implementação de PMSs, por melhor direcionarem a escolha de indicadores e a visualização de relacionamentos, boa parte dos modelos é dependente de considerações subjetivas em alguma etapa, a citar os pesos atribuídos nas análises de relevância entre fatores dentro do método AHP. O potencial fornecido pelas técnicas de *business analytics* para extração de informações diretamente das bases históricas dos indicadores tem sido um importante fator para a realização de trabalhos nessa área. A seguir, são apresentadas as principais referências dentro dessa vertente de métodos quantitativos para PMSs, sumarizadas no Quadro 2.

Schläfke et al. (2013) apresentaram uma abordagem intitulada PMA (do inglês, *Performance Management Analytics*), caracterizada pelo uso intensivo de dados e ferramentas analíticas para compreender a dinâmica organizacional, controlar indicadores chave e aumentar o desempenho operacional. Apesar de não se configurar como um modelo com a especificação de técnicas para a implementação de PMSs, essa abordagem ressalta a relevância dos métodos de *business analytics* como potencial fator de sucesso para os sistemas de mensuração de desempenho, capaz de preencher a lacuna existente entre a estruturação desses sistemas e sua efetiva implementação.

Também de forma mais conceitual, Raffoni et al. (2018) destacaram uma estrutura chamada BPA (do inglês, *Business Performance Analytics*) que direciona o emprego sistemático de dados e técnicas de análise no gerenciamento de desempenho. Essa estrutura é composta de cinco passos que tratam da avaliação da estratégia organizacional e de seus objetivos; da definição de questões chave a serem acompanhadas frente à estratégia; da identificação das necessidades de dados e formas de coleta; da escolha e uso das ferramentas mais adequadas de análise, considerando as vertentes características das abordagens de *business analytics* (descrição, diagnóstico, predição e prescrição), e da promoção de ações com base nos resultados e relações mapeados para os indicadores. Os

autores destacaram o potencial da estrutura para auxiliar na identificação das variáveis mais relevantes a serem controladas, bem como dos vínculos existentes entre elas.

Quadro 2 - Referências de métodos de *business analytics*

Referência	Título	Periódico	Abordagem / Técnicas
Schläfke et al. (2013)	<i>A framework for business analytics in performance management</i>	<i>International Journal of Productivity and Performance Management</i>	<i>Performance Management Analytics</i>
Raffoni et al. (2018)	<i>Business Performance Analytics: exploring the potential for Performance Management Systems</i>	<i>Production Planning &amp; Control</i>	<i>Business Performance Analytics</i>
Rodriguez-Rodriguez et al. (2009)	<i>Quantitative relationships between key performance indicators for supporting decision-making processes</i>	<i>Computers in Industry</i>	QRPMS (PCA / SEM e PLS)
Rodriguez-Rodriguez et al. (2010)	<i>Building internal business scenarios based on real data from a performance measurement system</i>	<i>Technological Forecasting &amp; Social Change</i>	PCA e Análise de Cenários
Wang & Chien (2016)	<i>Combining balanced scorecard with data envelopment analysis to conduct performance diagnosis for Taiwanese LED manufacturers</i>	<i>International Journal of Production Research</i>	Teste de Correlação, VIF, Regras de Árvore de Decisão e Regressão Logística
Peral et al. (2017)	<i>Application of Data Mining techniques to identify relevant Key Performance Indicators</i>	<i>Computer Standards &amp; Interfaces</i>	ARIMA, SVM, <i>Random Forest</i> e MLP
Wang et al. (2018)	<i>Using Bayesian belief network and time-series model to conduct prescriptive and predictive analytics for computer industries</i>	<i>Computers &amp; Industrial Engineering</i>	<i>Random Forest</i> , <i>Bayesian Belief Network</i> e ARIMA

Nota: QRPMS - *Quantitative Relationships at the Performance Measurement System*; PCA - *Principal Component Analysis*; SEM - *Structural Equation Model*; PLS - *Partial Least Squares*; VIF - *Variance Inflation Factor*; ARIMA - *Autoregressive Integrated Moving Average*; SVM - *Support Vector Machines*; MLP - *Multilayer Perceptron*.

Rodriguez-Rodriguez et al. (2009) descreveram a proposta de um modelo chamado QRPMS (do inglês, *Quantitative Relationships at the Performance Measurement System*), direcionado para selecionar KPIs, elicitare relacionamentos entre eles e projetar os impactos nos objetivos estratégicos. O modelo se constitui em quatro etapas, a começar por uma estruturação/análise do PMS voltada para o vínculo conceitual entre os indicadores e objetivos pré-estabelecidos. Em seguida,

é feito o pré-processamento dos dados históricos, considerando, por exemplo, a filtragem de valores anormais, homogeneização da frequência temporal, normalização e centralização dos dados. Após esses tratamentos, são identificados e quantificados os relacionamentos entre KPIs, sendo aplicadas as técnicas PCA (do inglês, *Principal Component Analysis*) ou SEM (do inglês, *Structural Equation Model*) para a identificação de KPIs, dependendo da razão entre número de indicadores e tamanho da amostra, e a técnica PLS (do inglês, *Partial Least Squares*) para a mensuração das relações. Por fim, estas são projetadas ao nível dos objetivos, considerando os vínculos conceituais avaliados na primeira etapa. Rodriguez-Rodriguez et al. (2010) apresentaram um modelo para análise de cenários empregando a evolução histórica dos aspectos de negócio (componentes) extraídos do PCA.

Wang & Chien (2016) descreveram outra abordagem considerando *business analytics* ao empregar técnicas para seleção inicial de indicadores, seguida pela aplicação de classificadores que indicam os relacionamentos significativos para a predição de desempenho. Os autores apresentaram o uso de teste de correlação supervisionado e do VIF (do inglês, *Variance Inflation Factor*) não supervisionado para remover indicadores redundantes dentro de uma listagem pré-existente. Em seguida, trataram da calibração de dois classificadores oriundos de regras de árvore de decisão e de regressão logística, efetuando comparativos em termos de acurácia, sensibilidade e especificidade.

Peral et al. (2017) destacaram como a estrutura orientada em *data mining* pode contribuir para que os objetivos estratégicos sejam devidamente acompanhados empregando as ferramentas de tecnologia da informação. Nesse contexto, faz-se necessária a seleção adequada de indicadores com base em suas relações internas e para com os objetivos. Por isso, os autores apresentaram uma abordagem na qual a definição de KPIs é feita em cinco etapas: pré-processamento, detecção de potenciais anomalias, cálculo de séries de diferenças, análise de relações entre pares e análise de relações compostas (entre vários indicadores). Em termos de técnicas, pode-se salientar a exposição dos autores relativa a modelos ARIMA (do inglês, *Autoregressive Integrated Moving Average*) para verificar o relacionamento entre as variáveis a partir das séries de diferenças calculadas e a avaliação dos modelos SVM (do inglês, *Support Vector*

*Machines*), *Random Forest* e MLP (do inglês, *Multilayer Perceptron*) dentro de um estudo de caso na etapa de análise de relações compostas.

Por fim, Wang et al. (2018) descreveram a combinação dos métodos *Random Forest*, *Bayesian Belief Network* e ARIMA para, respectivamente, selecionar KPIs, identificar relacionamentos com vista à construção de mapas estratégicos e simulações e efetuar previsão de desempenho considerando o efeito temporal existente entre os resultados dos indicadores. Os autores salientaram que esses procedimentos possibilitam o atendimento às quatro vertentes de análise características das abordagens de *business analytics*: descrição, diagnóstico, previsão e prescrição. A seleção de indicadores compõe as vertentes de descrição e diagnóstico, a previsão de desempenho associa-se à natureza de previsão e as simulações baseadas em mapas estratégicos contribuem dentro da vertente de prescrição.

Uma vez que a abordagem de Wang et al. (2018) se configura como a mais abrangente em termos das quatro vertentes de *business analytics*, atendendo ao escopo inicial deste trabalho associado à seleção de KPIs e à análise de suas relações, optou-se por empregá-la como a abordagem direcionadora para o estudo prático desenvolvido dentro de uma empresa de transporte ferroviário. No capítulo seguinte, é feita a apresentação dos principais aspectos vinculados aos métodos *random forest* e *bayesian belief network* que nortearam a fase do estudo prático.

## 4 Fundamentos teóricos

Como mencionado na seção anterior, dada a abrangência da proposta apresentada por Wang et al. (2018) relativa ao emprego de técnicas de *business analytics* para o estudo de seleção de indicadores, mapeamento de suas relações e projeção de resultados, essa abordagem foi considerada como a direcionadora para o estudo prático na empresa de transporte ferroviário. Tendo em vista ainda que a análise de previsão de desempenho dos indicadores excede ao escopo definido para este trabalho, foca-se neste capítulo em discorrer a respeito dos métodos *Random Forest* e *Bayesian Belief Network*, tratando dos principais aspectos associados a esses métodos que direcionaram o estudo prático e a análise de resultados.

### 4.1 **Random Forest**

O método *Random Forest*, apresentado por Breiman (2001), é um modelo de *ensemble* para predição caracterizado por uma estrutura composta por árvores de classificação ou regressão, na qual cada árvore é gerada com o emprego de uma amostra de *bootstrap*, processo de reamostragem com repetição, a partir dos dados originais utilizados para a calibração do modelo. As árvores são formadas por uma sequência de partições binárias de variáveis direcionadas à predição de uma determinada variável de interesse, sendo que o modelo considera apenas certa quantidade de variáveis preditoras, selecionadas de forma aleatória, como candidatas em cada partição. Dois parâmetros relevantes do *random forest* são a quantidade de preditores sorteados em cada partição e a quantidade de árvores que compõem a estrutura de predição, comumente referenciados como *mtry* e *nree*, respectivamente. Em geral, as predições de um modelo *random forest* consideram a categoria mais apresentada entre as árvores para problemas de classificação ou a média dos valores preditos em casos de regressão (Svetnik et al., 2004; Speiser et al., 2019).

Devido à construção com base em amostras de *bootstrap*, cada árvore de um modelo *random forest* possui a si associada dados chamados de *out-of-bag* (OOB), pertencentes ao conjunto inicial de dados de treino do modelo, mas não selecionados no processo de reamostragem para a calibração de determinada árvore. Os dados *out-of-bag* são comumente utilizados para fornecer dois tipos de informações relevantes: estimativa do erro de predição do modelo e referências de importância das variáveis preditoras. A estimativa do erro de predição é calculada com base nos resíduos das predições geradas a partir das árvores nas quais cada amostra se configura como *out-of-bag*. Já as referências de importância estão geralmente relacionadas à diferença no erro médio das árvores para a predição das amostras *out-of-bag* quando se efetua uma permutação dos valores de determinada variável nessas amostras. As variáveis mais importantes são aquelas para as quais a permutação gera maiores alterações no erro (Breiman, 2001; Svetnik et al., 2004; Genuer et al., 2010).

As referências de importância conferem ao método *random forest* sua aplicabilidade no processo de *feature selection* (também conhecido como *variable selection* ou *variable reduction*), o qual se direciona para a identificação dos preditores mais relevantes a serem considerados em um modelo de predição. O trabalho com um número reduzido e adequado de variáveis tende a propiciar maior eficiência ao modelo, por meio do aumento da acurácia, simplificação na coleta de dados e análise de resultados e maior agilidade no processamento computacional. Por isso, o método *random forest* possui grande relevância em muitas aplicações e sistemas reais para apoio à tomada de decisão, seja por sua capacidade de predição ou sua atuação no processo de seleção de variáveis (Genuer et al., 2015; Wang et al., 2018; Speiser et al., 2019).

Speiser et al. (2019) apresentaram um comparativo entre diferentes abordagens de emprego do *random forest* para a escolha de preditores, com foco na análise de desempenho para problemas de classificação. Embora no estudo prático da presente dissertação, descrito no Capítulo 5, tenha-se trabalhado com modelos de regressão, a ausência na literatura de um comparativo similar de desempenho para esse tipo de modelo, conforme mencionado por Speiser et al. (2019), e a aplicabilidade de boa parte das abordagens para o contexto de regressão motivaram a consideração das informações apresentadas naquele trabalho para a escolha das propostas aqui utilizadas.

Segundo os testes feitos por Speiser et al. (2019), duas eficientes abordagens de seleção foram descritas por Svetnik et al. (2004) e Genuer et al. (2010, 2015). A primeira se baseia na repetição de um processo de *k-fold cross-validation*, sendo feita a avaliação, em termos do erro de predição desse processo, de modelos com diferentes quantidades de variáveis preditoras, reduzidas progressivamente considerando a ordem de importância de cada uma. Já a proposta de Genuer et al. (2010, 2015), chamada VSURF (do inglês, *Variable Selection Using Random Forest*), é um algoritmo com uma eliminação inicial das variáveis menos importantes, seguida por avaliações, em termos da estimativa do erro de predição pelos dados *out-of-bag*, de alternativas para o modelo contemplando apenas a variável mais importante até o total de variáveis pré-selecionadas na primeira etapa. São gerados dois conjuntos de preditores: um direcionado para interpretação, mostrando os preditores mais relacionados à variável de interesse, e outro voltado para predição, com o tratamento de casos de redundância no conjunto anterior. No Quadro 3, é apresentado um detalhamento dos principais passos efetuados em cada uma das abordagens, as quais serviram como base para a fase de seleção de indicadores dentro do estudo prático.

Algumas características dessas abordagens são relevantes para serem mencionadas. A primeira se refere à classificação como métodos de *wrapper selection*, tendo em vista que a escolha das variáveis é feita através da avaliação de medidas de acurácia do modelo de predição. A segunda está relacionada ao fato de que a abordagem de Svetnik et al. (2004) se direciona, inicialmente, para a identificação da quantidade mais adequada de preditores, sem a explicitação dos mesmos, ao passo que o algoritmo VSURF já seleciona um conjunto específico de variáveis. Para a associação dos preditores na primeira abordagem, pode-se ordená-los com base na importância média de diferentes modelos de *random forest* (similar ao passo 1 do método VSURF) e filtrar os mais importantes até a quantidade associada ao menor erro médio, obtida no passo 7 da respectiva abordagem.

Outro ponto de destaque é que ambos os procedimentos são não recursivos, dado que o cálculo da importância das variáveis considera o conjunto completo das mesmas, não sendo refeito para cada subconjunto de preditores avaliado. Svetnik et al. (2004) destacaram que a realização de um processo recursivo, conhecido como *recursive feature elimination*, tende a reduzir o desempenho do

trabalho de seleção devido à maior susceptibilidade a sobreajuste. Essa questão foi analisada com os dados do estudo prático e será discutida em maiores detalhes adiante.

Quadro 3 - Principais passos das abordagens de seleção de variáveis

Svetnik et al. (2004)	Genuer et al. (2010, 2015)
<ol style="list-style-type: none"> <li>1) Particionar os dados em <i>k folds</i> dentro do processo de <i>cross-validation</i>;</li> <li>2) Para cada conjunto de treino, construir um modelo <i>random forest</i> com todas as variáveis, armazenando a importância de cada uma e a medida de erro para o <i>fold</i> de teste;</li> <li>3) Reconstruir o modelo de <i>random forest</i> para subconjuntos das variáveis mais importantes mapeadas no passo 2, reduzindo progressivamente o tamanho do subconjunto por um determinado fator e armazenando as medidas de erro para o <i>fold</i> de teste;</li> <li>4) Repetir os passos 2 e 3 até que todos os <i>k folds</i> tenham sido empregados como teste;</li> <li>5) Agregar as medidas de erro dos <i>k folds</i> de <i>cross-validation</i> para cada tamanho de subconjunto de variáveis preditoras;</li> <li>6) Repetir os passos 1 a 5 várias vezes com a vista à suavização da variabilidade;</li> <li>7) Agregar novamente as medidas de erro para cada tamanho de subconjunto de preditores, selecionando o tamanho com menor erro associado.</li> </ol>	<p><u>Ordenação e exclusão preliminar</u></p> <ol style="list-style-type: none"> <li>1) Calcular a importância média de cada variável com base em várias repetições do modelo <i>random forest</i>, ordenando as variáveis de forma decrescente por essa medida;</li> <li>2) Excluir as variáveis menos importantes (aquelas com a medida de importância média inferior a um patamar limite, relacionado ao seu desvio-padrão);</li> </ol> <p><u>Seleção de variáveis (interpretação)</u></p> <ol style="list-style-type: none"> <li>3) Construir modelos repetidos de <i>random forest</i> para diferentes quantidades de variáveis preditoras, iniciando com a mais importante (mapeada no passo 1) até contemplar todas as variáveis que permaneceram após o passo 2;</li> <li>4) Calcular o erro médio das amostras <i>out-of-bag</i> para os modelos associados a cada conjunto de variáveis, selecionando o menor conjunto com erro médio inferior ao menor erro acrescido do seu desvio padrão;</li> </ol> <p><u>Seleção de variáveis (predição)</u></p> <ol style="list-style-type: none"> <li>5) Construir uma sequência de modelos <i>random forest</i>, acrescentando, em ordem de importância, cada variável, desde que ela contribua significativamente para a redução do erro médio <i>out-of-bag</i>, considerando os preditores selecionados no passo 4;</li> <li>6) Selecionar o conjunto de variáveis do último modelo do passo 5 (com menor erro).</li> </ol>

O quarto aspecto relevante está associado à medida de erro considerada em cada abordagem (advinda do processo de *cross-validation* ou das amostras *out-of-bag*). Svetnik et al. (2004) apresentaram que a utilização do erro *out-of-bag* contribuiria fortemente para o sobreajuste, gerando resultados inconsistentes, especialmente, em situações nas quais os preditores não fossem correlacionados à variável de interesse. Genuer et al. (2015) salientaram que a medida *out-of-bag* realmente tende a selecionar algum grupo de preditores mesmo que não exista

relação significativa com a variável de interesse. Assim como o aspecto anterior, este também foi avaliado com os dados do estudo prático e será tratado mais a frente.

Por fim, vale ressaltar que, conforme apresentado por Ambroise & McLachlan (2002) e Reunanen (2003), o comparativo entre o desempenho de distintos processos de *variable selection* requer a avaliação de acurácia sobre dados específicos de teste ou gerados por reamostragem (*cross-validation* ou *bootstrap*, por exemplo) externa ao processo de seleção. Esse procedimento visa a evitar a análise por estimativas de acurácia supervalorizadas, que podem levar a conclusões inadequadas de melhor desempenho para modelos mais complexos, resultantes de um problema de sobreajuste conhecido como *selection bias*.

## 4.2 **Bayesian Belief Network**

O modelo *Bayesian Belief Network*, apresentado por Pearl (1986), se baseia em uma estrutura de grafo acíclico orientado, no qual os nós representam as variáveis avaliadas e os arcos mostram as relações de dependência existentes entre elas, sendo a ausência de conexão entre dois nós indicativo de independência em termos de probabilidade condicional. As variáveis de uma rede bayesiana podem, em princípio, apresentar qualquer distribuição de probabilidade. No entanto, os estudos presentes na literatura se direcionam, em especial, para redes com variáveis discretas ou normalmente distribuídas e conectadas entre si via relações lineares, as quais são conhecidas, respectivamente, como redes discretas e redes gaussianas (Scutari et al., 2019).

A definição completa de uma rede bayesiana a partir de um conjunto de dados passa, necessariamente, pela identificação da estrutura mais provável de arranjo dos nós e arcos frente aos dados e, em seguida, pela estimativa dos parâmetros de distribuição associados a cada variável, considerando os dados e a estrutura mapeada. Essas duas fases de aprendizado são conhecidas como *structure learning* e *parameter learning* (Scutari et al., 2019). Como o foco principal da aplicação da rede bayesiana no presente trabalho é a identificação das relações entre indicadores de forma objetiva com vista à construção de um mapa

estratégico, sem a simulação de impacto entre variáveis feita por Wang et al. (2018), optou-se por restringir ao detalhamento apenas da primeira fase.

O processo de *structure learning* é realizado pela aplicação de um determinado critério estatístico por meio de um algoritmo que direciona a avaliação de alternativas de estrutura dentro do espaço de soluções. O desempenho de cada abordagem de aprendizado, em termos de acurácia ou processamento computacional, é influenciado diretamente tanto pelo critério estatístico quanto pelo algoritmo utilizado (Scutari et al., 2019).

Três são os principais grupos de abordagens de *structure learning* presentes na literatura: *constraint-based*, *score-based* e *hybrid*. O primeiro se baseia no emprego de testes de independência condicional entre os nós para a composição dos arcos, ao passo que o segundo associa uma pontuação representativa da qualidade de ajuste de cada estrutura, buscando encontrar aquela com o melhor valor associado. Já o último grupo trabalha com uma fase preliminar de identificação de possíveis relações entre as variáveis via testes de independência condicional, seguida pela busca da melhor estrutura, em termos de qualidade de ajuste, que contenha um subconjunto dos arcos mapeados na fase anterior (Scutari et al., 2019).

Scutari et al. (2019) realizaram um comparativo entre algoritmos de *structure learning* dos três grupos acima mencionados, considerando os mesmos critérios estatísticos e amostras com diferentes tamanhos e níveis de complexidade. A complexidade está relacionada tanto à quantidade de dados quanto à de parâmetros a serem estimados para a construção da rede bayesiana. Os autores apresentaram que o algoritmo *tabu search*, do grupo *score-based*, se destacou em termos de acurácia e processamento computacional. De modo geral, esse algoritmo se baseia na montagem de uma lista das últimas estruturas avaliadas, que direciona o processo de busca pelo melhor arranjo de forma a não se limitar a um determinado ótimo local, explorando a vizinhança de uma estrutura fora da listagem das últimas avaliadas. O processo de busca base, conhecido como *hill climbing*, é feito através da adição, remoção ou reversão de um arco com vista a selecionar apenas uma alternativa que melhore a solução já encontrada.

No que tange ao critério estatístico, Chua & Ong (2017) avaliaram cinco possibilidades em algoritmos *score-based* aplicados sobre amostras com distintos

tamanhos e quantidades de variáveis, observando o melhor desempenho, em acurácia, associado ao *Bayesian Information Criteria* (BIC), proposto inicialmente por Schwarz (1978). Esse critério, que busca identificar a estrutura mais adequada de relação entre as variáveis priorizando arranjos com menores quantidades de arcos e, conseqüentemente, de parâmetros a serem estimados, é comumente empregado nos métodos de *structure learning* (Scutari et al., 2019).

Para a comparação entre duas estruturas dado um critério estatístico, pode-se empregar o fator BF (do inglês, *Bayes Factor*), que indica a razão entre as probabilidades condicionais do conjunto de dados em relação às estruturas. Num cenário em que se possa considerar que os arranjos, *a priori*, tenham iguais probabilidades de ocorrência, o fator BF representa a razão entre as probabilidades condicionais de cada estrutura em relação aos dados, indicando a preferência pela estrutura apresentada no numerador caso seu valor seja maior do que 1 (Kass & Raftery, 1995).

No próximo capítulo, são apresentados os resultados de implementação do algoritmo *tabu search* junto ao critério BIC para a elaboração de um mapa estratégico a partir das relações entre variáveis do estudo prático, bem como das abordagens de *random forest* para a seleção de variáveis.

## 5 Estudo prático

Neste capítulo, são apresentados os procedimentos de implementação e os resultados alcançados com as técnicas de *random forest* e *bayesian belief network* para a avaliação de indicadores e seus relacionamentos no contexto do PMS de uma empresa de transporte ferroviário de cargas, tendo por base a abordagem proposta por Wang et al. (2018). É importante salientar que as implementações foram efetuadas via linguagem R, versão 3.5.3 (R Core Team, 2019), dentro do ambiente RStudio, versão 1.2.5033, em um computador com a configuração Intel(R) Core\_(TM) i5-8250U CPU @ 1.60GHz, 4 núcleos, 8 processadores lógicos e 8.0GB de RAM.

### 5.1 Escolha de objetivos e indicadores e levantamento de dados históricos

Em linhas gerais, a empresa avaliada trabalha atualmente com seis principais objetivos estratégicos desdobrados em indicadores nas diferentes áreas funcionais, abrangendo desde o nível estratégico até o operacional. No escopo deste estudo, limitou-se à análise de três objetivos: (i) uso eficiente dos recursos, (ii) ampliação do volume transportado e (iii) desenvolvimento de novos negócios. Numa perspectiva da gestão de ativos ferroviários, um dos pilares comumente considerado para o suporte a esses objetivos é o indicador de disponibilidade de locomotivas, o qual corresponde ao percentual do tempo em que esses ativos, responsáveis por fornecer o esforço de tração para a movimentação dos trens, não estão direcionados à atuação da área de manutenção e, conseqüentemente, podem ser utilizados pela operação em atividades de transporte de carga para os clientes ou em serviços internos. Dada a relevância desse indicador, optou-se por tratá-lo como o indicador predito, para o qual foram conduzidas as análises de seleção de indicadores preditores e de relacionamentos existentes.

No Quadro 4, estão listados todos os indicadores considerados no estudo, sendo válido destacar a busca por não incluir as parcelas diretas de indisponibilidade como possíveis preditores, tendo em vista a óbvia relação com o indicador de interesse e, por isso, o provável impacto substancial na fase de seleção de variáveis via acurácia de predição do modelo *random forest*. Dessa forma, oportuniza-se que outras relações possam ser identificadas e assim o processo de escolha de indicadores seja mais efetivo, principalmente no âmbito estratégico, gerando novos *insights* para a tomada de decisão. Importante salientar que isso não elimina a necessidade de acompanhamento das parcelas de indisponibilidade dentro do PMS da empresa em estudo.

Quadro 4 - Lista de indicadores considerados no estudo prático

Identificador	Indicador
Y, $X_Y^S$	Percentual de disponibilidade de locomotivas
$X_1, X_1^S$	Quantidade de locomotivas ativas
$X_2, X_2^S$	Idade média das locomotivas ativas
$X_3, X_3^S$	Capacidade média de tração das locomotivas ativas
$X_4, X_4^S$	Quantidade de manutenções finalizadas por locomotiva ativa
$X_5, X_5^S$	Quantidade de ordens de manutenção finalizadas por locomotiva ativa
$X_6, X_6^S$	Homem x hora aplicado em manutenções finalizadas por locomotiva ativa
$X_7, X_7^S$	Tempo decorrido em manutenções finalizadas por locomotiva ativa
$X_8$	Quantidade de solicitações de manutenção abertas por locomotiva ativa
$X_9$	Quilometragem percorrida por solicitação de manutenção aberta
$X_{10}$	Locomotiva x hora em restrição por locomotiva ativa
$X_{11}$	Capacidade em Homem x hora das oficinas de manutenção
$X_{12}$	Percentual de Homem x hora das oficinas aplicado em manutenção
$X_{13}$	Homem x hora planejado das manutenções finalizadas
$X_{14}$	Quantidade normalizada de manutenções finalizadas por <i>head count</i> em oficina
$X_{15}, X_{15}^S$	Volume transportado
$X_{16}$	Tonelada quilômetro bruta transportada
$X_{17}, X_{17}^S$	Quilometragem percorrida por locomotiva ativa
$X_{18}$	Consumo de diesel por locomotiva ativa
$X_{19}$	Quilometragem percorrida por consumo diesel
$X_{20}$	Consumo de diesel por tonelada quilômetro bruta transportada
$X_{21}, X_{21}^S$	Custo de manutenção de locomotivas
$X_{22}$	Custo de manut. de locomotivas por tonelada quilômetro bruta transportada
$X_{23}$	Custo de manutenção por locomotiva disponível

Como pode ser observado no Quadro 4, para facilitar a menção e acompanhamento dos resultados do estudo prático, associou-se um identificador a

cada uma das variáveis avaliadas, empregando a letra Y para o indicador a ser predito (disponibilidade de locomotivas) e a letra X, com índices específicos, para referenciar os possíveis preditores.

Alguns dos indicadores possuem mais de um identificador por estarem vinculados a valores considerados não apenas na visão consolidada, com abrangência corporativa, mas também com um nível de segregação pertinente aos processos da empresa, na forma de um resultado percentual em relação ao consolidado (referenciado pelo sobrescrito <sup>S</sup>). O indicador de disponibilidade, por exemplo, é avaliado em sua abrangência corporativa como a variável a ser predita, tendo o resultado segregado de um determinado grupo de locomotivas, direcionado ao transporte em certa região, como um dos possíveis preditores.

Esse procedimento visa a atender às especificidades dos processos em estudo, incluindo os níveis de detalhamento aderentes à realidade organizacional, sem prejudicar a configuração dos indicadores para as análises de importância, ao evitar a presença de preditores muito correlacionados entre si. Ressalta-se aqui a necessidade de atenção aos procedimentos de *feature engineering*, os quais se relacionam à forma como os dados são organizados para o uso nos modelos, sendo dependentes das características do caso avaliado (Kuhn & Johnson, 2013).

Quanto ao levantamento de dados históricos, foram coletados os resultados de cada um dos 35 indicadores (considerando os valores segregados) no período de Jan/2016 a Out/2019 (46 meses).

## 5.2 Seleção de indicadores via *random forest*

Conforme mencionado no início deste capítulo, as implementações dos métodos de *random forest* e *bayesian belief network* foram realizadas com o emprego da linguagem R. A abordagem de seleção de indicadores via *random forest* apresentada por Svetnik et al. (2004) foi conduzida através do pacote “randomForest” (Breiman et al., 2018), no qual uma função chamada “rfcv” (do inglês, *random forest cross-validation*) possibilita a execução do procedimento de *cross-validation* para diferentes quantitativos de possíveis preditores, com o cálculo dos erros associados. Esse processo, que compreende os passos 1 a 5 elucidados no Quadro 3, deve ser executado de forma repetida por uma função

externa ao pacote “randomForest” para a suavização da variabilidade, seguida pela agregação da medida de erro e escolha do tamanho de subconjunto com a menor medida associada (passos 6 e 7). Com vista a facilitar as menções ao longo do texto, é empregada a referência RFCV para representar essa abordagem.

Já a proposta desenvolvida por Genuer et al. (2010, 2015), encontra-se estruturada no pacote “VSURF” (Genuer et al., 2019), conforme o próprio nome do método, sendo possível avaliar os resultados de cada uma das três fases do processo de seleção, com a importância das variáveis, a evolução na listagem de preditores considerados, as medidas de erro *out-of-bag*, os tempos de processamento, entre outras informações. No Quadro 5, são apresentados os valores base empregados para os principais parâmetros das duas abordagens de seleção de indicadores.

Quadro 5 - Parâmetros dos métodos de seleção de variáveis

Abordagem	Parâmetro	Valor
RFCV	<b>ntree:</b> nº de árvores em cada modelo <i>random forest</i>	2000
	<b>mtry:</b> nº de variáveis sorteadas em cada partição de uma árvore	p/3*
	<b>k:</b> nº de <i>folds</i> utilizados no processo de <i>cross-validation</i>	5
	<b>r:</b> nº de repetições do processo de <i>cross-validation</i>	20
	<b>step:</b> nº de variáveis do próximo tamanho de subconjunto avaliado	0,5p*
VSURF	<b>ntree:</b> nº de árvores em cada modelo <i>random forest</i>	2000
	<b>nfor.thres:</b> nº de modelos <i>random forest</i> gerados na fase de ordenação e exclusão preliminar ( <i>thresholding</i> )	50
	<b>nfor.interp:</b> nº de modelos <i>random forest</i> gerados na fase de seleção de variáveis para interpretação	25
	<b>nfor.pred:</b> nº de modelos <i>random forest</i> gerados na fase de seleção de variáveis para predição	25
	<b>mtry.thres:</b> nº de variáveis sorteadas em cada partição de uma árvore na fase de ordenação e exclusão preliminar ( <i>thresholding</i> )	p/3*
	<b>mtry.interp:</b> nº de variáveis sorteadas em cada partição de uma árvore na fase de seleção de variáveis para interpretação	p/3*
	<b>mtry.pred:</b> nº de variáveis sorteadas em cada partição de uma árvore na fase de seleção de variáveis para predição	p/3*

\* p: nº de variáveis consideradas para a construção do modelo *random forest*. Na prática, os valores “p/3” e “0,5p” são utilizados arredondados para baixo, sem casas decimais e limitados ao mínimo de 1.

Embora os valores dos parâmetros possam ser um fator de análise em cada caso de aplicação, optou-se, no presente trabalho, em considerar as configurações propostas por Svetnik et al. (2004) e Genuer et al. (2015), sendo o valor de *ntree* da abordagem RFCV o mesmo empregado em VSURF. Svetnik et al. (2004) não

especificaram um valor para o parâmetro *n<sub>tree</sub>*. O valor de *mtry* como um terço da quantidade de variáveis, independente do método, é a referência geralmente utilizada para problemas de regressão (Kuhn & Johnson, 2013).

Além das implementações dos dois métodos em suas estruturas básicas, avaliou-se também a alternativa com processo recursivo discutida por Svetnik et al. (2004), na qual o cálculo da importância das variáveis é feito a cada novo tamanho de subconjunto analisado. É importante destacar que foram utilizados todos os dados históricos nas implementações de seleção de variáveis, considerando os 34 possíveis preditores para o indicador de disponibilidade de locomotivas em seus 46 resultados mensais.

Conforme salientado por Speiser et al. (2019), o estudo de tempos de processamento é relevante no contexto da análise de desempenho de métodos de *feature selection*. De forma a orientar uma comparação ilustrativa entre os métodos utilizados, são apresentados, no Quadro 6, os tempos associados a uma execução de cada um deles.

Quadro 6 - Tempos de processamento na seleção de variáveis

	<b>Tempo (s)</b>
<b>RFCV</b>	39,4
<b>RFCV (recursivo)</b>	44,9
<b>VSURF</b>	38,8
Ordenação e exclusão preliminar	17,7
Seleção de variáveis (interpretação)	18,6
Seleção de variáveis (predição)	2,4

Como mencionado no Capítulo 4, a abordagem RFCV aponta o quantitativo mais adequado de preditores para uma determinada variável de interesse, sem nomeá-los. Já a VSURF explicita os preditores em cada uma de suas fases de seleção. Esse aspecto é importante ao se avaliar os tempos de processamento do Quadro 6, visto que, aos valores das implementações de RFCV (não recursivo ou recursivo), deve-se adicionar um tempo para um procedimento de identificação das variáveis, caso o objetivo seja obter uma listagem das mesmas. Esse procedimento pode ser similar à primeira fase do processo VSURF, escolhendo-se o quantitativo mapeado dos preditores mais importantes. Independente dessa necessidade de acréscimo de tempo às implementações da abordagem RFCV, observa-se que a VSURF apresentou menor tempo de execução, detalhado em

suas três fases diretamente pelo pacote “VSURF”, na linguagem R. Dentro da abordagem RFCV, a alternativa com processo recursivo demandou maior tempo, como já esperado em razão dos recálculos de importância das variáveis.

Antes de adentrar nos resultados da seleção de preditores, é válido salientar que, devido à aleatoriedade presente nas abordagens utilizadas, os testes foram inicializados com uma semente para a geração de números aleatórios, tendo em vista a reprodução dos resultados. Outro ponto a se destacar é que as medidas de erro utilizadas se baseiam no erro quadrático médio (MSE, do inglês *mean squared error*).

Na Figura 1, é apresentado um comparativo entre os erros médios obtidos em 20 repetições do processo de *5-fold cross-validation* dentro do método RFCV, considerando as alternativas com e sem procedimento recursivo. Com base no parâmetro *step* mostrado no Quadro 5, que indica o tamanho de cada subconjunto no processo de redução de variáveis e cujo valor de “0,5p” representa a manutenção de metade dos preditores contabilizados no subconjunto anterior, foram avaliados os seguintes quantitativos de variáveis: 1, 2, 4, 8, 17 e 34, partindo, inicialmente, do conjunto com todos os possíveis preditores até se chegar a apenas um.

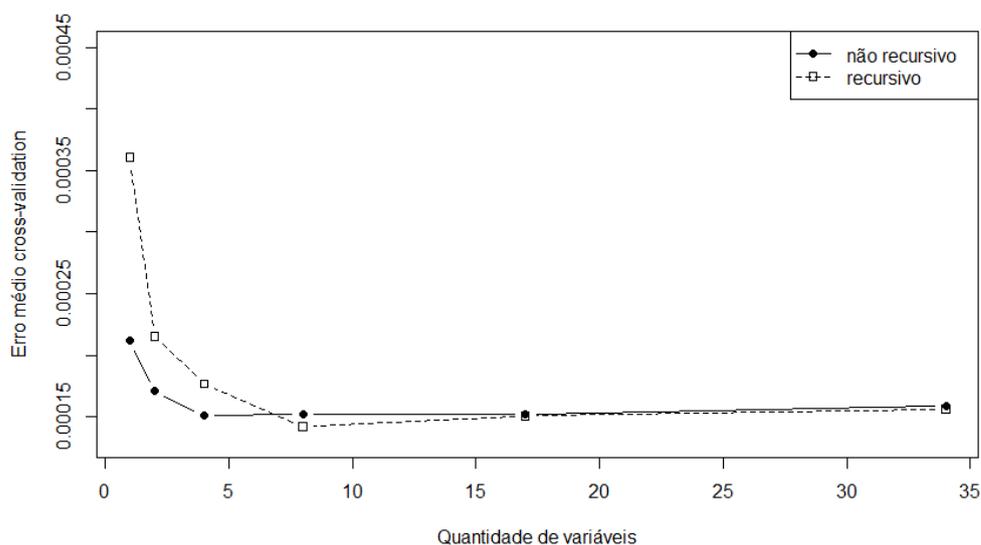


Figura 1 - Erros médios do método RFCV (não recursivo e recursivo)

Como pode ser observado na Figura 1, para a alternativa não recursiva, o menor erro médio foi obtido com o quantitativo de 4 indicadores, ao passo que, com o processo recursivo, o menor erro ficou associado ao subconjunto com 8

preditores. Verifica-se que em ambas as estruturas o aumento de variáveis teve baixa interferência na acurácia do modelo, fato relevante para o objetivo de se selecionar um conjunto mais parcimonioso de indicadores que consiga retratar uma variável de interesse e simplifique o acompanhamento de informações dentro de um sistema de mensuração de desempenho. Em linha com os resultados avaliados por Svetnik et al. (2004), para os menores subconjuntos de preditores, a alternativa recursiva apresentou maiores erros, o que, segundo os autores, seria motivado por uma tendência mais significativa dessa estrutura ao sobreajuste.

Quanto ao método VSURF, a Figura 2 contém os principais gráficos relativos às suas três fases de seleção, iniciando, nos gráficos (a) e (b), com as informações de importância média das variáveis e a definição, via valores de desvio padrão, do patamar de exclusão (destacado em vermelho) daqueles preditores menos relevantes. Nessa primeira fase, foram calculados os valores de importância a partir de 50 modelos de *random forest* (parâmetro *nfor.thres* do Quadro 5), que possibilitaram a ordenação dos preditores seguida pela exclusão de 11 variáveis com importância média inferior ao patamar de exclusão obtido como o menor valor de predição dado por um modelo CART (do inglês, *Classification and Regression Trees*), em verde no gráfico (b), ajustado sobre a curva de desvio padrão dos valores de importância (Genuer et al., 2015).

No gráfico (c), tem-se a curva com os erros médios das amostras *out-of-bag* resultantes da construção de 25 modelos *random forest* (parâmetro *nfor.interp* do Quadro 5) para cada subconjunto das variáveis mais importantes oriundas da primeira fase, acrescentando-se uma de cada vez segundo a ordem de importância já estabelecida. A linha vermelha sinaliza a quantidade de preditores selecionados na segunda fase (5), correspondendo ao menor modelo com erro médio associado inferior ao menor erro acrescido do desvio padrão, critério base empregado nessa etapa de seleção.

No gráfico (d), é apresentada a curva com os erros médios *out-of-bag* obtidos a partir de outros 25 modelos *random forest* (parâmetro *nfor.pred* do Quadro 5) para cada subconjunto dos preditores escolhidos na fase anterior em que o acréscimo de uma nova variável, de acordo com a mesma ordem de importância das etapas anteriores, contribua efetivamente para a redução do erro médio *out-of-bag*. Verifica-se que esse processo resultou na seleção de um modelo com 4 variáveis, que não correspondem necessariamente às 4 variáveis

mais importantes da primeira fase, dada a premissa de adição progressiva somente daqueles preditores que exerçam significativa influência na redução do erro médio.

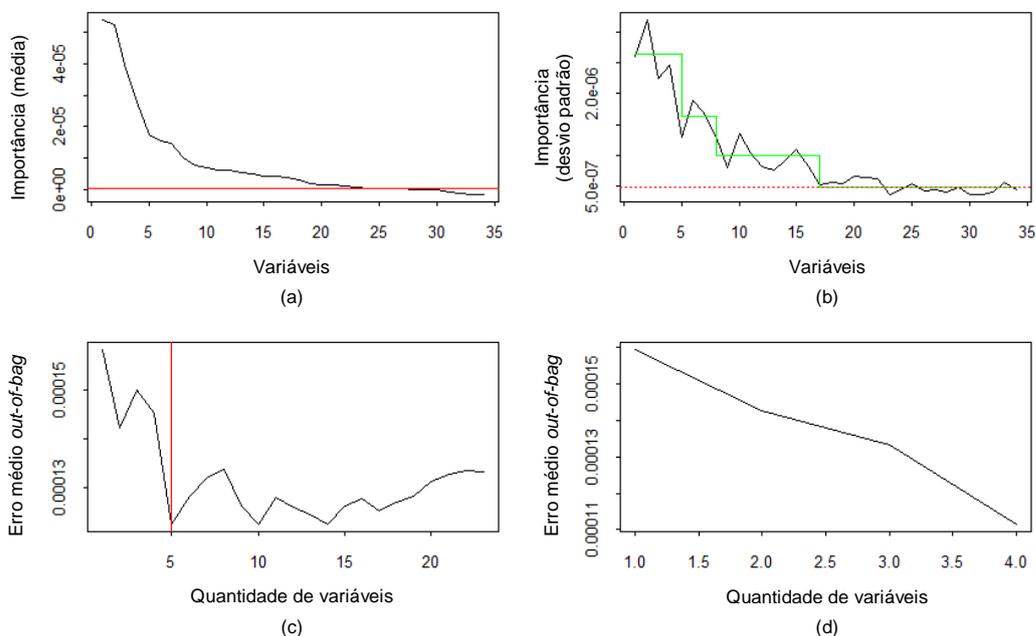


Figura 2 - Importância das variáveis e erros médios do método VSURF

Tendo em vista estabelecer um comparativo de desempenho entre os métodos RFCV e VSURF em termos de acurácia de predição, buscou-se associar indicadores específicos aos quantitativos apresentados como mais adequados no primeiro método, empregando a possibilidade já mencionada relativa à ordem de importância da primeira fase do algoritmo VSURF. No Quadro 7, encontram-se as listagens de indicadores selecionados via RFCV e VSURF, sendo feita a segregação dos resultados das fases de interpretação e predição da abordagem VSURF para posterior análise do ganho associado a um modelo mais parcimonioso. Observa-se que a distinção entre as listagens dessas duas fases é dada pela retirada do indicador  $X_{17}^s$  na fase de predição, o qual, conceitualmente, está muito relacionado ao indicador  $X_{15}^s$ , evidenciando a perspectiva de eliminação de redundância almejada com essa última etapa do algoritmo VSURF.

Outro ponto a se destacar no Quadro 7 é que, embora a alternativa não recursiva do método RFCV e a etapa de predição do algoritmo VSURF resultem no mesmo quantitativo de indicadores, as listagens associadas não são iguais, fato

também motivado pela premissa de adição mais restritiva durante a fase final do método VSURF.

Quadro 7 - Preditores selecionados nos métodos RFCV e VSURF

	n° de preditores	Preditores
<b>RFCV</b>	4	$X_{15}^S, X_{22}, X_{17}^S, X_2$
<b>RFCV (recursivo)</b>	8	$X_{15}^S, X_{22}, X_{17}^S, X_2, X_Y^S, X_{19}, X_{16}, X_{23}$
<b>VSURF - interpretação</b>	5	$X_{15}^S, X_{22}, X_{17}^S, X_2, X_Y^S$
<b>VSURF - predição</b>	4	$X_{15}^S, X_{22}, X_2, X_Y^S$

Para avaliar o desempenho de cada listagem de preditores, com atenção voltada a evitar o efeito de *selection bias*, mencionado no Capítulo 4, foram calculados os erros a partir de 50 repetições do processo de *10-fold cross-validation*, empregando todos os 46 resultados históricos mensais dos indicadores. Cada modelo de *random forest* foi construído com os parâmetros  $n_{tree} = 2000$  e  $m_{try} = p/3$  (sendo “p” a quantidade de preditores). Tendo em vista à reprodução dos resultados dos métodos, todos os testes foram inicializados com uma semente para a geração de números aleatórios. A Figura 3 mostra a distribuição dos erros associados a cada uma das listagens de preditores.

Pode-se observar que, em termos de acurácia de predição, o desempenho da abordagem VSURF foi superior ao da RFCV, dados os menores erros no processo de *cross-validation*. A listagem obtida na fase de interpretação já se destaca em relação às alternativas RFCV, sendo ainda melhorada com a tratativa presente na etapa de seleção para predição e a consequente geração de um modelo mais parcimonioso. Vinculando ao aspecto de tempo de processamento, anteriormente apresentado, tem-se que a abordagem VSURF sobressai nos dois aspectos, configurando-se, dessa forma, na mais adequada para o contexto estudado. É válido enfatizar o trabalho com um problema de regressão e poucos preditores (34), sendo essa última classificação baseada na divisão feita por Speiser et al. (2019), que usa o critério da presença mais de 50 variáveis para distinguir as bases de dados com muitos preditores.

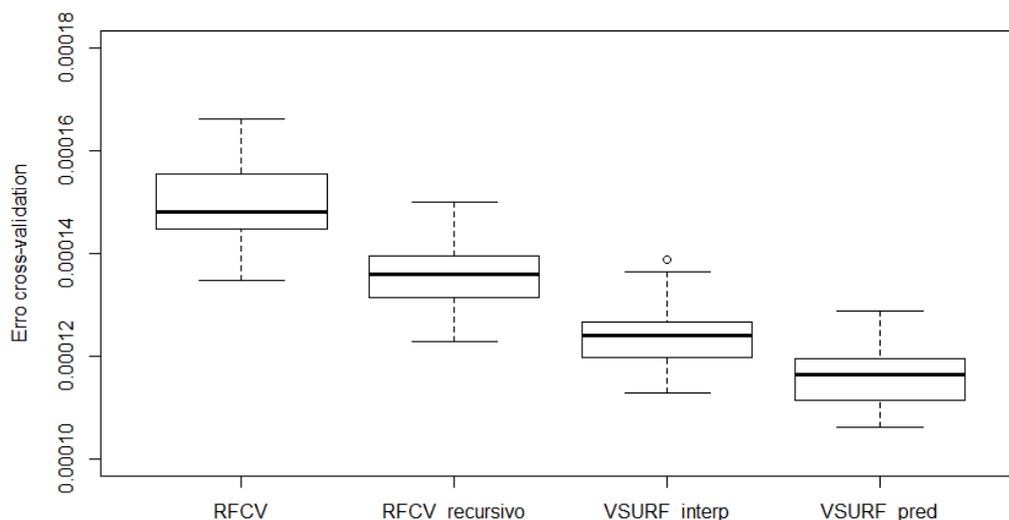


Figura 3 - Erros de predição a partir dos preditores seleccionados

Há de se salientar ainda que a listagem oriunda da alternativa com processo recursivo do método RFCV apresentou melhor desempenho, em termos do erro de predição, quando comparado ao método tradicional, cenário inverso ao observado no tempo de processamento. Esse desempenho superior em acurácia é interessante, especialmente, devido à exposição feita por Svetnik et al. (2004) relativa à maior susceptibilidade da alternativa recursiva a sobreajuste, com base nos erros calculados para baixos quantitativos de preditores durante o processo de *feature selection*.

Por fim, é importante discorrer sobre o emprego dos erros associados às amostras *out-of-bag* no método VSURF. Conforme mencionado no Capítulo 4, Svetnik et al. (2004) abordaram uma fragilidade dessas medidas de erro, corroborada por Genuer et al. (2015), quando utilizadas em um cenário no qual os possíveis preditores não sejam correlacionados à variável de interesse. Mesmo nessas condições, o método tende a selecionar algum grupo de preditores, gerando uma expectativa errônea de identificação de relações existentes. No caso estudado, a própria acurácia do método VSURF evidencia a aplicabilidade das medidas de erro *out-of-bag*, sendo válido destacar que, no contexto amplo dos sistemas de mensuração de desempenho, é esperado trabalhar-se com indicadores vinculados a processos de uma organização, que naturalmente devem possuir pontos de convergência. Caso seja necessário, pode-se recorrer a uma análise preliminar de correlação entre os indicadores.

### 5.3

#### Mapeamento de relacionamentos via *bayesian belief network*

Conforme mencionado anteriormente, assim como para os métodos de *random forest*, as implementações relativas ao modelo *bayesian belief network* foram efetuadas com a linguagem R, neste caso via o pacote “bnlearn” (Scutari & Ness, 2019), no qual estão disponíveis funções para os processos de *structure learning*. É importante salientar que, com o objetivo de empregar a seleção feita via *random forest* e, ao mesmo tempo, trabalhar com um quantitativo de variáveis que possibilitasse a visualização de mais relações dentro do mapa estratégico, optou-se por considerar, nesta seção, os 8 preditores identificados através da abordagem recursiva do método RFCV, juntamente à variável de interesse. Como destacado no Capítulo 4, o processo de *structure learning* foi conduzido através da combinação entre o algoritmo *tabu search* e o critério estatístico BIC.

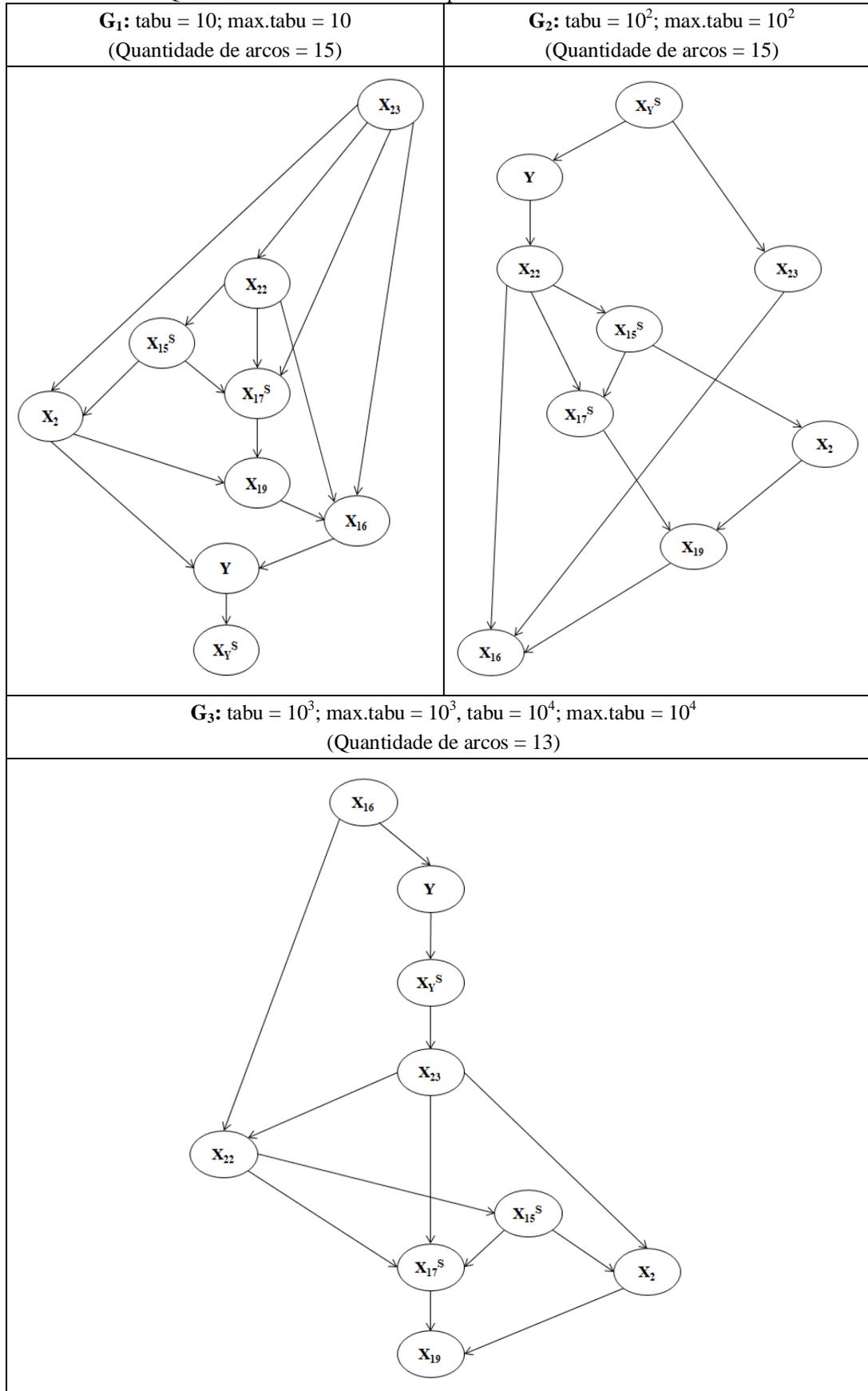
O algoritmo *tabu search* possui quatro principais parâmetros, cujos valores utilizados estão apresentados no Quadro 8. O parâmetro “tabu” indica o tamanho da listagem que armazena as últimas estruturas avaliadas para o direcionamento do processo de busca em diferentes vizinhanças, a fim de se evitar a limitação a um determinado ótimo local. Já o parâmetro “max.tabu” mostra a quantidade máxima de iterações que podem ser executadas pelo algoritmo sem que haja melhoria na solução. Embora esses dois parâmetros possam receber valores distintos, a configuração padrão presente no pacote “bnlearn” utiliza o mesmo valor, sendo essa configuração mantida nos testes realizados. Quanto aos termos “max.iter” e “maxp”, correspondentes às quantidades máximas de iterações e de pais para cada nó, foram empregados com o valor infinito (configuração padrão).

Quadro 8 - Parâmetros do algoritmo *tabu search*

Parâmetro	Valor
<b>tabu:</b> n° de estruturas na lista tabu (últimas estruturas avaliadas)	10/10 <sup>2</sup> /10 <sup>3</sup> /10 <sup>4</sup>
<b>max.tabu:</b> n° máximo de iterações sem melhoria na solução	10/10 <sup>2</sup> /10 <sup>3</sup> /10 <sup>4</sup>
<b>max.iter:</b> n° máximo de iterações	Infinito
<b>maxp:</b> n° máximo de pais para cada nó	Infinito

Tendo em vista avaliar o impacto na estrutura da rede bayesiana, foram empregados quatro valores para os parâmetros “tabu” e “max.tabu” (10, 10<sup>2</sup>, 10<sup>3</sup> e 10<sup>4</sup>). Os grafos resultantes estão apresentados no Quadro 9.

Quadro 9 - Grafos com base nos parâmetros “tabu” e “max.tabu”



Conforme pode ser verificado no Quadro 9, a alteração nos parâmetros gerou diferentes arranjos entre as variáveis, sendo interessante a identificação da mesma estrutura para os maiores valores utilizados ( $10^3$  e  $10^4$ ). Esse fato aponta para a relevância do emprego de diferentes alternativas para os parâmetros com vista ao mapeamento do melhor arranjo, considerando a tendência de estabilização à medida que os valores propiciam a avaliação de um maior número de vizinhanças no espaço de soluções.

Para uma breve comparação entre o processamento computacional requerido por cada valor dos parâmetros avaliados, são apresentados, na Tabela 1, os dados do tempo de execução e do número de testes realizados com o critério estatístico (BIC) dentro do processo de *structure learning*. Percebe-se, como esperado, um aumento na complexidade computacional para maiores valores dos parâmetros, embora o processamento ainda tenha sido efetuado em tempos muito baixos. A variação deve ser mais significativa em condições com maior número de variáveis, tendo em vista o consequente aumento de possibilidades de estruturas analisadas em cada iteração.

Tabela 1 - Medidas de processamento computacional no processo de *structure learning*

“tabu” e “max.tabu”	Tempo de processamento (s)	Testes com o critério estatístico (qtd)
10	0,06	412
$10^2$	0,14	2700
$10^3$	0,62	18596
$10^4$	16,80	165724

Dois pontos valem ser mencionados antes de adentrar na análise das estruturas obtidas, a começar pelo fato de que, devido à natureza contínua dos indicadores, as estruturas foram construídas com o arcabouço relativo às redes gaussianas. Além disso, para garantir a reprodução dos resultados, as implementações foram inicializadas com uma mesma semente para a geração de números aleatórios, com foco na fase de adição, remoção ou reversão de um arco dentro do processo *hill climbing*, considerando a possibilidade de que a escolha do arco avaliado seja feita de forma aleatória.

Para comparar os três grafos resultantes das implementações (Quadro 9), empregou-se o fator BF, descrito no Capítulo 4. Os valores desse fator foram calculados considerando o critério estatístico BIC e a premissa de que todos os

grafos possuem, *a priori*, a mesma probabilidade de ocorrência. Dessa forma, um valor superior a 1 indica que, a partir do conjunto de dados associados às variáveis, a estrutura utilizada no numerador (1º índice) é mais provável do que aquela referenciada no denominador (2º índice). Os resultados obtidos sinalizaram que, no caso do estudo prático, a estrutura mais provável é aquela representada pelo grafo  $G_3$ . Detalhando os valores, o fator  $BF_{21} = 1,24$  indica a preferência, em termos de probabilidade, por  $G_2$  em relação a  $G_1$ , ao passo que o fator  $BF_{32} = 1,08$  mostra a preferência de  $G_3$  em relação a  $G_2$ . O fator  $BF_{31} = 1,34$ , dado como o produto entre  $BF_{21}$  e  $BF_{32}$ , reforça  $G_3$  como a estrutura mais provável.

Avaliando o Quadro 9, observa-se que o quantitativo de arcos do arranjo  $G_3$  (13) é menor que o associado às estruturas  $G_1$  e  $G_2$  (15), correspondendo a uma menor quantidade de parâmetros a serem estimados. Esse fato, somado à indicação do grafo  $G_3$  como o mais provável via o fator BF com o critério estatístico BIC, aponta para a busca desse critério em identificar uma estrutura parcimoniosa em seus arcos, privilegiando as relações mais significativas.

A fim de mensurar a força dos relacionamentos probabilísticos entre as variáveis do grafo  $G_3$ , foram calculados os valores de redução do critério BIC associados à remoção de cada um dos arcos dessa estrutura, sendo os resultados apresentados na Tabela 2. As variáveis de origem e destino de cada arco são referenciadas pelos termos “De” e “Para”, respectivamente. Quanto maior a redução do critério BIC, mais forte é a relação entre as variáveis. Pode-se observar que os quatro primeiros arcos elencados se destacam frente aos demais, dois deles representando vínculos diretos por termos na fórmula de cálculo dos indicadores e os outros dois evidenciando aspectos de relacionamento característicos dos processos da empresa em estudo.

Tabela 2 - Redução do critério BIC pela remoção de arcos do grafo  $G_3$

De	Para	$\Delta$ BIC	De	Para	$\Delta$ BIC
$X_{23}$	$X_{22}$	-68,9	$X_{15}^S$	$X_2$	-7,6
$X_{16}$	$X_{22}$	-50,5	$X_{22}$	$X_{17}^S$	-7,5
$X_{15}^S$	$X_{17}^S$	-39,1	$X_{17}^S$	$X_{19}$	-7,1
$X_{22}$	$X_{15}^S$	-29,2	$X_Y^S$	$X_{23}$	-3,2
$X_{23}$	$X_{17}^S$	-10,5	$X_2$	$X_{19}$	-2,1
$X_{16}$	Y	-8,9	$X_{23}$	$X_2$	-1,1
Y	$X_Y^S$	-8,4			

O grafo  $G_3$ , auxiliado pelas medidas de força dos arcos, compõe um mapa estratégico com as principais relações existentes entre os indicadores. Conforme salientado por Wang et al. (2018), essa forma de construção via estrutura de rede bayesiana é orientada pelos dados históricos dos indicadores, conferindo mais objetividade ao processo de elaboração dos mapas. A atuação dos especialistas fica direcionada à análise das relações identificadas, com o vínculo prático entre elas e as realidades do contexto em estudo.

## 5.4 Discussões

Ao avaliar o processo de *feature selection* via *random forest*, a redução no erro de predição da variável de interesse  $Y$  (disponibilidade de locomotivas) aponta para o potencial de um conjunto mais parcimonioso de indicadores em simplificar e melhorar o acompanhamento dos sistemas de mensuração de desempenho. Isso não exclui a possibilidade de se calcular e avaliar um maior número de variáveis, mas direciona o foco para itens específicos mais relevantes a determinados objetivos.

Como mencionando anteriormente, optou-se por não considerar as parcelas de indisponibilidade nas análises do indicador de disponibilidade de locomotivas, tendo em vista que a relação direta de desdobramento tenderia a concentrar a fase de seleção somente nessas parcelas. Essa medida, com foco de avaliação a um nível mais estratégico, possibilitou, por exemplo, a visualização do significativo relacionamento de indicadores operacionais com a variável de interesse. Na fase de *feature selection*, o indicador  $X_{15}^S$ , relacionado a um percentual do volume transportado, foi escolhido dentro do conjunto de preditores no método VSURF (Quadro 7), apresentando o maior valor de importância média dentre todas as variáveis consideradas.

No mapa estratégico do grafo  $G_3$  gerado a partir do método *bayesian belief network* e apresentado no Quadro 9, pode-se observar novamente o vínculo de indicadores operacionais com a disponibilidade de locomotivas, através, por exemplo, do arco com o nó  $X_{16}$  (tonelada quilômetro bruta transportada). Este representa uma importante medida de desempenho operacional no contexto do transporte ferroviário de cargas, por efetuar o cruzamento entre a distância

percorrida por cada trem e o volume transportado (considerando também o peso dos ativos - locomotivas e vagões - que compõem os trens).

Nos processos da empresa em estudo, o relacionamento entre os indicadores  $X_{16}$  e  $Y$  se baseia nas características dos modelos operacionais presentes nas diferentes regiões de atuação, sendo observados melhores resultados de disponibilidade em períodos de concentração de transporte na região onde circulam os trens com maiores valores de tonelada quilômetro bruta transportada. Os indicadores  $X_{15}^S$  e  $X_{17}^S$ , que podem ser associados, respectivamente, aos percentuais de volume e quilometragem percorrida nessa região também aparecem no mapa estratégico apresentado no grafo  $G_3$ .

Considerando a relação do indicador de disponibilidade de locomotivas com os três objetivos estratégicos da empresa destacados na Seção 5.1, vale salientar que o patamar superior em disponibilidade na região acima mencionada é fruto das melhorias promovidas pela empresa nos processos de operação e manutenção ao longo dos anos de atuação na região, refletindo a busca pelo alcance do objetivo de uso eficiente dos recursos. No grafo  $G_3$ , o arco entre os indicadores  $X_{15}^S$  e  $X_2$  (idade média da frota de locomotivas) sinaliza para um relacionamento a princípio inexistente, em termos conceituais. No entanto, o indicador  $X_2$  carrega consigo um efeito de evolução temporal aderente ao processo de diversificação de portfólio da empresa em estudo, diretamente relacionado aos objetivos de ampliação do volume transportado e de desenvolvimento de novos negócios.

## 6 Conclusão

Como observado na revisão da literatura, embora os estudos sobre a aplicação de métodos de pesquisa operacional e gestão já tenham contribuído para melhorar o processo de implementação dos sistemas de mensuração de desempenho, a subjetividade ainda presente em algumas etapas dos métodos é um desafio a ser enfrentado. Técnicas de *business analytics* são uma importante alternativa para a condução, de forma objetiva, dos processos de seleção de KPIs e de mapeamento de suas relações, além da projeção de desempenho futuro com base nos resultados históricos já observados dos indicadores e suas relações.

No presente trabalho, foi utilizada uma abordagem de *business analytics* em um estudo prático considerando o PMS de uma empresa do setor de transporte ferroviário de cargas. Buscou-se discorrer sobre os principais aspectos teóricos e de implementação das técnicas de *random forest* e *bayesian belief network*, de forma a avaliar o emprego, respectivamente, nas fases de seleção de indicadores e de construção de mapas estratégicos com base em suas relações.

Avaliando, com o foco na gestão de ativos ferroviários, o suporte para o alcance dos objetivos estratégicos de uso eficiente dos recursos, ampliação do volume transportado e desenvolvimento de novos negócios, pôde-se trabalhar na seleção dos principais indicadores vinculados à disponibilidade de locomotivas, bem como na elaboração de um mapa estratégico com a representação dos relacionamentos entre as variáveis. Foi observado um significativo vínculo entre indicadores operacionais e a disponibilidade, não motivado pelo fato de que a oferta de mais ativos tende a subsidiar um maior volume de produção, mas revelando que o modelo operacional pode também contribuir para os resultados de disponibilidade de locomotivas.

Quanto aos métodos de *random forest* utilizados, verificou-se um melhor desempenho, em termos de acurácia e processamento computacional, da abordagem VSURF, que gerou também o menor conjunto de preditores com tratamento de redundância entre eles. A fragilidade dessa abordagem em selecionar variáveis mesmo em situações nas quais os possíveis preditores não

sejam correlacionados à variável de interesse tende a não ser representativa no contexto dos sistemas de mensuração de desempenho, visto que os indicadores retratam os processos de uma mesma organização e devem, naturalmente, possuir pontos de convergência.

Ao avaliar a abordagem de seleção proposta por Svetnik et al. (2004), aqui referenciada por RFCV, o desempenho superior, em acurácia, da alternativa recursiva, caracterizada inicialmente como mais susceptível a sobreajuste, em relação a não recursiva aponta para uma possibilidade de ampliação desse comparativo. Sugere-se a consideração de distintos tamanhos para o conjunto de variáveis avaliadas, tendo, como medida de acurácia, o erro de predição sobre dados de treino ou reamostragem posterior ao processo de seleção.

Na linha do comparativo entre propostas de *feature selection* via *random forest*, observa-se a oportunidade de desenvolvimento de um trabalho com uma análise ampla para problemas de regressão, considerando um maior número de propostas em diferentes conjuntos de dados. Na presente dissertação, foram utilizadas algumas abordagens sobre uma instância específica oriunda da empresa em estudo. Vale salientar a importância da avaliação do tempo de processamento junto à acurácia de predição para a análise comparativa de desempenho entre métodos de *feature selection*.

Em termos da aplicação do método *bayesian belief network* para a identificação de relacionamentos e a construção de mapa estratégico entre as variáveis, empregou-se a combinação do algoritmo *tabu search* com o critério BIC no processo de *structure learning*, com a escolha feita, via *Bayes factor*, entre as estruturas geradas para diferentes valores de parâmetros do algoritmo. Foi verificada a tendência de estabilização da estrutura à medida que os parâmetros possibilitaram explorar um maior conjunto de vizinhanças dentro do espaço de soluções, sendo priorizado o arranjo com a menor quantidade de arcos, o que se mostrou em linha com a expectativa inicial de um modelo mais parcimonioso em suas relações dado o critério estatístico utilizado.

Embora não tenha sido considerada no escopo do presente trabalho, a análise de previsão de desempenho futuro a partir dos relacionamentos entre indicadores é uma importante fase para os sistemas de mensuração de desempenho. Nesse sentido, há a oportunidade de desenvolvimento de estudos

que avaliem os aspectos teóricos e de implementação também dentro desse processo.

Além disso, é observada a possibilidade de se empregar o arcabouço de técnicas de séries temporais para aperfeiçoar a fase de *feature selection*, tendo em vista que os indicadores podem apresentar defasagens em seus relacionamentos. Estas precisam ser identificadas de forma adequada e não devem ser perdidas dentro do processo de seleção de variáveis, normalmente focado em vínculos em uma mesma amostra ou período de tempo. Acredita-se que este possa ser um relevante tema de pesquisa no contexto dos sistemas de mensuração de desempenho, uma vez que os resultados dos indicadores se apresentam, geralmente, na forma de séries históricas.

## 7

### Referências bibliográficas

AMBROISE, C.; MCLACHLAN, J. Selection Bias in Gene Extraction on the Basis of Microarray Gene-Expression Data. **Proceedings of the National Academy of Sciences of the United States of America**, v. 99, n. 10, p. 6562-6566. 2002.

BITITCI, U. S.; SUWIGNJO, P.; CARRIE, A. S. Strategy management through quantitative model-ing of performance measurement systems. **International Journal of Production Economics**, v. 69, n. 1, p. 15-22. 2001.

BREIMAN, L. Random Forests. **Machine Learning**, v. 45, p. 5-32. 2001.

BREIMAN, L. et al. Package ‘randomForest’- Breiman and Cutler's Random Forests for Classification and Regression. 2018. Disponível em: <<https://cran.r-project.org/web/packages/randomForest/randomForest.pdf>>. Acesso em: 07 jan. 2020.

CHUA, C.; ONG, H. Comparison of Scoring Functions on Greedy Search Bayesian Network Learning Algorithms. **Science & Technology**, v. 25, n. 3, p. 719-734. 2017.

FERREIRA, F. A. F.; SANTOS, S. P.; RODRIGUES, P. M. M. Adding value to bank branch performance evaluation using cognitive maps and MCDA: a case study. **Journal of the Operational Research Society**, v. 62, n. 7, p. 1320-1333. 2011.

GENUER, R.; POGGI, J. M.; TULEAU-MALOT, C. Variable selection using random forests. **Pattern Recognition Letters**, v. 31, p. 2225-2236. 2010.

\_\_\_\_\_. VSURF: An R Package for Variable Selection Using Random Forests. **The R Journal**, v. 7, n. 2, p. 19-33. 2015.

\_\_\_\_\_. Package ‘VSURF’- Variable Selection Using Random Forests. 2019. Disponível em: <<https://cran.r-project.org/web/packages/randomForest/randomForest.pdf>>. Acesso em: 07 jan. 2020.

JOSHI, R.; BANWET, D. K.; SHANKAR, R. A. Delphi-AHP-TOPSIS based benchmarking framework for performance improvement of a cold chain. **Expert Systems with Applications**, v. 38, n. 8, p. 10170-10182. 2011.

KUHN, M.; JOHNSON, K. **Applied Predictive Modeling**. New York: Springer, 2013. p. 27-28, p. 199.

KAPLAN, R. S.; NORTON, D. P. The balanced scorecard – measures that drive performance. **Harvard Business Review**, p. 71-79, jan./fev. 1992.

KASS, R. E.; RAFTERY, A. E. Bayes Factors. **Journal of the American Statistical Association**, v. 90, n. 430, p. 773-775. 1995.

MELLO, R.; MARTINS, R. A. Can Big Data Analytics Enhance Performance Measurement Systems?. **IEEE Engineering Management Review**, v. 47, n. 1, p. 52-57. 2019.

PEARL, J. Fusion, Propagation, and Structuring in Belief Networks\*. **Artificial Intelligence**, v. 29, 241-288. 1986.

PERAL, J.; MATÉ, A.; MARCO, M. Application of Data Mining techniques to identify relevant Key Performance Indicators. **Computer Standards & Interfaces**, v. 50, p. 55-64. 2017.

R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>. 2019.

RAFFONI, A. et al. Business Performance Analytics: exploring the potential for Performance Management Systems. **Production Planning & Control**, v. 29, n. 1, p. 51-67. 2018.

REUNANEN, J. Overfitting in Making Comparisons Between Variable Selection Methods. **Journal of Machine Learning Research**, v. 3, p. 1371-1382. 2003.

RODRIGUEZ-RODRIGUEZ, R.; SAIZ, J. J. A.; BAS, A. O. Quantitative relationships between key performance indicators for supporting decision-making processes. **Computers in Industry**, v. 60, n. 2, p. 104-113. 2009.

RODRIGUEZ-RODRIGUEZ, R. et al. Building internal business scenarios based on real data from a performance measurement system. **Technological Forecasting & Social Change**, v. 77, n. 1, p. 50-62. 2010.

SANTOS, S. P. et al. Adding value to performance measurement by using system dynamics and multicriteria analysis. **International Journal of Operations & Production Management**, v. 22, n. 11, p. 1246-1272. 2002.

\_\_\_\_\_. Measuring organizational performance using a mix of OR methods. **Technological Forecasting & Social Change**, v. 131, p. 18-30. 2018.

SANTOS, S. P.; BELTON, V.; HOWICK, S. Enhanced performance measuring using OR: a case study. **Journal of the Operational Research Society**, v. 59, n. 6, p. 762-775. 2008.

SARKIS, J. Quantitative models for performance measurement systems - alternate considerations. **International Journal of Production Economics**, v. 86, n. 1, p. 81-90. 2003.

SCHLÄFKE, M.; SILVI, R.; MÖLLER, K. A framework for business analytics in performance management. **International Journal of Productivity and Performance Management**, v. 62, n. 1, p. 110-122. 2013.

SCHWARZ, G. Estimating the dimension of a model. **The Annals of Statistics**, v. 6, n. 2, p. 461-464. 1978.

SCUTARI, M.; GRAAFLAND, C. E.; GUTIÉRREZ, J. M. Who learns better Bayesian network structures: Accuracy and speed of structure learning algorithms. **International Journal of Approximate Reasoning**, v. 115, 235-253. 2019.

SCUTARI, M.; NESS, R. Package ‘bnlearn’ - Bayesian Network Structure Learning, Parameter Learning and Inference. 2019. Disponível em: <<https://cran.r-project.org/web/packages/bnlearn/bnlearn.pdf>>. Acesso em: 07 jan. 2020.

SPEISER, J. L. et al. A comparison of random forest variable selection methods for classification prediction modeling. **Expert Systems With Applications**, v. 134, p. 93-101. 2019.

SUWIGNJO, P.; BITITCI, U. S.; CARRIE, A. S. Quantitative models for performance measurement system. **International Journal of Production Economics**, v. 64, n. 1, p. 231-241. 2000.

SVETNIK, V. et al. Application of Breiman’s random forest to modeling structure-activity relationships of pharmaceutical molecules. **International Workshop on Multiple Classifier Systems**, p. 334-343. 2004.

WANG, C.-H.; CHENG, H.-Y.; DENG, Y.-T. Using Bayesian belief network and time-series model to conduct prescriptive and predictive analytics for computer industries. **Computers & Industrial Engineering**, v. 115, p. 486-494. 2018.

WANG, C.-H.; CHIEN, Y.-W. Combining balanced scorecard with data envelopment analysis to conduct performance diagnosis for Taiwanese LED manufacturers. **International Journal of Production Research**, v. 54, n. 17, p. 5169-5181. 2016.