

# DASHBOARD PREDITIVO PARA APOIO À TOMADA DE DECISÃO: O CASO DA LOGÍSTICA DE UMA EMPRESA DO SETOR SIDERÚRGICO

**Luiza Santos Cunha**

[luiza.cunha@ufrgs.br](mailto:luiza.cunha@ufrgs.br)

**José Luis Duarte Ribeiro, PhD**

[ribeiro@producao.ufrgs.br](mailto:ribeiro@producao.ufrgs.br)

Universidade Federal do Rio Grande do Sul

Engenharia de Produção e Transportes

2017

---

Com o ambiente de negócios em constante mudanças, os recursos de *Business Intelligence* (BI) ganharam maior importância, uma vez que os desafios em gerenciar os dados gerados são cada vez maiores. As ferramentas de BI podem oferecer às empresas informações estruturadas acerca do negócio, baseadas em fontes de dados diversas, possibilitando uma tomada de decisão rápida e efetiva. O resultado desse estudo foi uma solução de *Business Intelligence* construída através do software *Power BI*, recurso de *BI Self-Service* fornecido pela Microsoft. A fim de introduzir funções preditivas e visuais, foram utilizados *scripts* da linguagem de programação R. O *dashboard* criado tem como objetivo apoiar a tomada de decisões dos executivos da área de logística da empresa estudada.

**Palavras chave:** *Business Intelligence*; linguagem R; Power BI.

---

## 1. Introdução

Na era da informação, uma expressiva quantidade de dados esta à disposição dos tomadores de decisão. O *big data* (BD) se refere a conjuntos de dados que não são apenas

grandes, mas também possuem alta variedade e velocidade de criação, o que os torna difíceis de manipular usando ferramentas e técnicas tradicionais. Devido ao rápido crescimento de tais dados, novas soluções precisam ser estudadas e fornecidas a fim de extrair valor e conhecimento a partir desses conjuntos de dados (Sivarajah et al., 2017). Esse conhecimento pode ser obtido usando *Big Data Analytics* (análise de grandes dados, em tradução livre), que é a aplicação de técnicas avançadas de análise em grandes dados. (McAfee & Brynjolfsson, 2012).

O *Data Science* (DS), também conhecido como Ciência de Dados no Brasil, foi possibilitado por novas linguagens de programação e processadores de alto desempenho. Este campo de estudo usualmente alia bases de dados históricos com a análise em tempo real de conjuntos de dados não estruturados. Assim, estas informações podem ter fontes distintas, que vão desde de sistemas integrados de gestão empresarial, como o SAP, aos dados gerados nas redes sociais. Dessa forma, os tomadores de decisão precisam ser capazes de obterem informações valiosas de dados variados e em rápida mudança (Schmidt et al., 2016; Sivarajah et al., 2017).

Segundo Makris & Chryssolouris (2013), apesar dos profissionais do *Supply Chain Management* (SCM), trabalharem com um grande número de dados para formar seus indicadores de desempenho, a partir da manipulação estatística das suas informações, a aplicação de DS é recente nessa área. Tendo em vista esse cenário, o *Journal of Business Logistics*, em 2013, convocou a comunidade científica para a realização de pesquisas acerca das possíveis aplicações do DS no SCM (Hazen et al., 2014). Assim, a partir deste ano, passos significativos foram dados em relação ao uso de DS na SCM e muitos dos conceitos estão sendo utilizados. A visão preditiva oriunda dessas aplicações se tornou um diferencial competitivo nas organizações, permitindo que o tomador de decisão tenha mais informações pertinentes ao negócio (Necula, 2015).

Segundo um estudo feito pela *BVL International* (instituto alemão especializado em promover a gestão da cadeia logística) em 2013, 60% das empresas mundiais que atuam na área de logística pretendem investir em ferramentas de análise de dados nos próximos 5 anos, o que representa um alerta aos gestores com relação à necessidade de conectar inteligência ao crescente número de dados gerados (Handfield et. al., 2013). Assim, como em quase todas as áreas de operações industriais, o DS está começando a fazer incursões na logística e no gerenciamento da cadeia de suprimentos.

Por outro lado, no mercado siderúrgico brasileiro, de acordo com o Instituto Aço Brasil (2016), as usinas brasileiras operaram em 2015, em média, com 68% da sua capacidade instalada de produção. É importante ressaltar que, a fim de alcançar a rentabilidade mínima de operação, esse valor deveria estar no patamar de 80%. Nesse interim, percebe-se a necessidade do setor de se reinventar, buscando soluções que visem redução de custos e respostas ágeis às variações do mercado.

Diante do exposto, este estudo visa propor a utilização de *dashboard* interativo com funções preditivas para orientar certas decisões de logística da empresa estudada, desenvolvidos a partir da linguagem estatística de programação R. Essa solução de *Business Intelligence* (BI) utilizou o software *Power BI* fornecido pela Microsoft, um dos principais players no campo do *BI Self-Service*. Assim, o resultado deste estudo foi uma solução de BI e o caso empírico apresentado pode ser usado por todos aqueles que estão interessados no relacionamento do *Power BI* com a linguagem R.

Este artigo está estruturado em cinco seções, a primeira referente à introdução e objetivos, a segunda com uma breve revisão de literatura sobre o planejamento estratégico logístico e os desafios enfrentados pelo setor e a utilização de recursos computacionais para o *business intelligence* e sua integração com a linguagem R de programação estatística. Na próxima seção, será feita uma descrição do método de trabalho com as características da empresa selecionada e dos dados trabalhados. Por fim, na seção quatro serão apresentados os resultados do estudo bem como as análises realizadas e, na última, as conclusões obtidas.

## **2. Referencial Teórico**

### **2.1 Contextualização: planejamento logístico**

A logística tem que fornecer as quantidades certas de bens, de forma eficiente, no lugar certo, na ordem correta e dentro do tempo estipulado (Tan, 2001). O atendimento a essas demandas requer um planejamento de logística antecipado, o qual cobre todos os processos logísticos e os recursos necessários. Este sistema pode ser separado em planejamento logístico estratégico (longo prazo), tático (médio prazo) e operacional (curto prazo) (Novaes, Takebayashi, & Briesemeister, 2015; Paiola et. al., 2013).

O primeiro, planejamento logístico estratégico, está focado em realizar as avaliações iniciais sobre a viabilidade das interações entre as partes interessadas de toda a *supply chain*. Enquanto que o planejamento tático da logística se concentra em alternativas de processos logísticos e na sua avaliação. Especialmente, na flexibilidade desses processos para adotar mudanças, como variações no volume, por exemplo. Por fim, no estágio operacional, esses processos e recursos de logística oriundos dos estágios anteriores serão continuamente detalhados e integrados (Paiola et al., 2013; Tan, 2001; Tseng, Yue, & Taylor, 2005).

Segundo Paiola et al. (2013), é no estágio operacional que o maior número de dados é gerado e, ainda, representa os maiores desafios de integração e extração de conhecimento. Os desenvolvimentos recentes na área da tecnologia da informação oferecem a possibilidade de integrar os dados históricos existentes para impulsionar a qualidade do planejamento logístico.

## **2.2 Desafios de Ciência de Dados no setor logístico**

Hazen et al., (2014) propuseram uma definição de *Data Science* (DS) no âmbito da *Supply Chain Management*: o DS em SCM é a aplicação de métodos quantitativos e qualitativos de uma variedade de disciplinas em combinação com a teoria SCM para resolver problemas relevantes de SCM e prever resultados, levando em conta a qualidade dos dados e os problemas de disponibilidade.

A complexidade para planejar uma quantidade crescente de processos com base em informações que mudam frequentemente é um desafio considerável. (Matende & Ogao, 2013). De acordo com Schallow et al. (2014), 50% do tempo de um planejador na indústria é usado para coletar e preparar informações. Enquanto apenas 20% do tempo é usado para tarefas de planejamento. As causas identificadas nesse contexto são (1) falta de suporte de software de planejamento, (2) falta de conexão e consistência de dados e informações e (3) não utilização do conhecimento gerado anteriormente.

Na indústria 4.0, há um aumento de dados gerados em toda a cadeia de suprimentos, como, por exemplo, os códigos de barras e a tecnologia RFID (Attaran, 2007; Prater, Frazier, & Reyes, 2005). Atualmente, essa quantidade crescente de dados dificilmente é utilizada para tarefas de planejamento, embora o *Machine Learning* (ML),

conteúdo do DS, possa ser usado como parte integrante do planejamento da cadeia de suprimentos (Bertoni, Colombo, & Grilli, 2011; Schallow et al., 2014). Além de automatizar as tarefas de planejamento, o ML oferece potencial para uma maior utilização do conhecimento gerado anteriormente nos processos logísticos implementados com sucesso.

### 2.3 Business Intelligence e o Power BI

A fim de garantir a sua permanência em mercados competitivos, as organizações precisam de um processo decisório rápido e eficaz. Para responder a essa demanda, os fornecedores de *softwares* começaram a criar sistemas que englobam informações de toda a empresa. Esses tipos de produtos ou serviços estão sob o termo de *Business Intelligence* (BI) e são soluções que contam com visualizações dos dados, geração de alertas e cálculos de indicadores de desempenho. (Lu, 2014; Sivarajah et al., 2017)

Segundo Lu (2014), o termo *Business Intelligence* foi proposto por Howard Dresner, do Grupo Gartner, no ano de 1996, quando foi definido como o uso de tecnologias e aplicações que consistem em *data warehouse* ou *data mart*, relatórios de consulta, análise de dados, mineração de dados e *backup* de dados.

A fim de ilustrar a importância do BI, Lu (2014) propõe dois cenários: o primeiro com fósforos espalhados e o segundo enfileirados em grupos de cinco unidades. O conjunto de fósforos é o mesmo, porém, é possível obter informações de forma mais rápida e eficaz com o último grupo. Com essa analogia, pode-se dizer que o primeiro grupo representa a vasta quantidade de dados disponíveis às empresas, com informações de produção, vendas, logística entre outros. Dessa forma, há a necessidade de organizar os dados não estruturados de tal maneira a gerar valor ao negócio. Assim, as soluções de BI são ferramentas que auxiliam na análise dos dados fazendo com que eles possuam significado ao negócio e melhorem o processo decisório.

Com o ambiente de negócios em constantes mudanças, o uso do BI ganhou notoriedade por fornecer às organizações a possibilidade de uma análise rápida aos seus dados. Com o objetivo de aumentar a autonomia dos analistas e permitir a customização dos relatórios dentro das empresas, foi proposta uma nova abordagem de BI, o *BI Self-*

*Service* e, durante os últimos anos, o número de *players* de mercado que oferecem essa abordagem aumentou rapidamente (Clark, 2017; Lu, 2014).

O *software Power BI* (PBI) fornece recursos de análise e visualização de dados orientados a usuários empresariais para criar relatórios interativos e apoiar o processo de tomada de decisão. O PBI é uma solução de BI *self-service* baseada em nuvem, o que possibilita a atualização dos dados em tempo real. Com o seu uso, é possível a criação de relatórios por pessoas que não possuem conhecimentos técnicos específicos de programação. (Clark, 2017)

Segundo Imhoff & White (2011), o BI *self-service* pode ser definido como as aplicações de BI que permitem aos usuários maior autonomia e menor dependência de profissionais da tecnologia da informação (TI). Essas soluções se concentram em quatro principais objetivos (Imhoff & White, 2011; Lu, 2014):

- O primeiro objetivo centra na implantação rápida e no gerenciamento do *data warehouse*. A solução é usada para garantir o desempenho e a escalabilidade de dados. Também permite que os usuários modifiquem suas próprias aplicações de acordo com suas necessidades específicas, o que ajuda no aumento da satisfação do consumidor;

- O segundo objetivo contempla a facilidade de uso. Como os usuários finais não são profissionais de TI e precisam de uma ferramenta que os ajude na criação de relatórios e de análises, a usabilidade é um dos fatores mais significativos do BI *Self-Service*. Com a interface visual fácil e intuitiva, os usuários podem se tornar mais auto-suficientes;

- O terceiro objetivo compreende a visão que os relatórios e os resultados devem ser entendidos facilmente, e serem exibidos em diferentes dispositivos. Do ponto de vista técnico, é necessário que as informações estejam claras e que sejam facilmente acessíveis, bem como a fonte de dados deve ser rastreável e estar documentada. A organização melhora sua tomada de decisão rastreando as interações e as decisões, capturando e disseminando as melhores práticas;

- Por fim, o quarto objetivo é facilitar a acessibilidade à fonte de dados. Os usuários finais podem acessar as fontes de dados de uma maneira mais fácil do que com o BI tradicional, o que pode acelerar todo o processo decisório. No entanto, isso não significa que o departamento de TI não tenha controle sobre a fonte de dados no BI *Self-*

*Service*, usualmente, essa área será responsável por permitir o acesso aos dados de acordo com o perfil do usuário.

## **2.3 Power BI e sua integração com a linguagem R**

Segundo Schmidt et al. (2016), R foi criado na Universidade de Auckland por George Ross Ihaka e Robert Gentleman, em 1994, e foi projetado para pequenas estações de trabalho, como desktops e laptops. Desde a criação da linguagem R, muitos cientistas de dados e estatísticos alimentaram o sucesso da R, ao contribuir com pacotes que permitem que os usuários possam utilizar as técnicas mais recentes de análise estatística, visualização e mineração de dados. Isso levou a R a tornar-se uma das línguas mais utilizadas pelos cientistas de dados. Muitos dos algoritmos necessários para o trabalho de DS não estão incluídos como parte da instalação básica, e devem ser adicionados à biblioteca do usuário (R. J. Hyndman & Athanasopoulos, 2013; Lantz, 2015).

R é uma linguagem de programação estatística de código aberto, nos termos da GNU GPL (*General Public License*), e pode ser acessada através da IDE (*integrated development environment*) RStudio. Comumente usada pela comunidade DS para resolver uma extensa variedade de problemas de negócios, os quais abrangem áreas como: Ciências atuariais (para descobrir riscos expostos a seguros e finança); Telecomunicações (uso fraudulento do cartão SIM ou padrões de uso móvel); Finanças e serviços bancários (para identificar fraude em transações financeiras) (Fontama, Barga, & Tok, 2015).

Com a funcionalidade dos *scripts* R serem executados no *Power BI*, é possível conciliar a capacidade do primeiro de gerar análises estatísticas avançadas com o desempenho do segundo em acessar fontes de dados diversas e compartilhar os relatórios *online*.

## **3. Método**

### **3.1 Cenário e características da pesquisa**

A empresa siderúrgica em estudo é uma das cinco maiores no panorama mundial e possui diversos segmentos de clientes, como o da construção civil, do setor agropecuário, das indústrias de bens duráveis e da automobilística. Ela está dividida em

cinco grandes operações de negócios e possui unidades voltadas para a produção de aços especiais, de aços longos e de aços planos, além das atividades de mineração.

A operação mais significativa encontra-se no Brasil e possui, atualmente, 8 usinas, 3 fábricas, 76 unidades de distribuição, 12 unidades de corte e dobra voltadas para a construção civil, e, por fim, duas operações portuárias. Tendo em vista a representatividade e percepção de potenciais melhorias por parte dos gestores de logística da empresa, as unidades estudadas serão as usinas e fábricas.

### **3.2 Tipo de pesquisa**

Com o objetivo de identificar e propor o *dashboard* com funções preditivas para orientar as decisões de logística da empresa estudada, o método empregado neste trabalho é uma pesquisa quantitativa, análise de dados numéricos através de procedimentos estatísticos, descritiva, identificação de oportunidades de melhoria e a análise dos dados, e aplicada, aplicação prática e solução de problemas reais. (Gerhardt & Silveira, 2009)

O estudo é considerado uma pesquisa-ação, uma vez que o *dashboard* será desenvolvido e proposto através de trabalho colaborativo, envolvendo profissionais da empresa e a própria autora do trabalho. De acordo com Tripp (2005), a pesquisa-ação consiste essencialmente em unir pesquisa e ação em um único processo, no qual os atores implicados participam junto com os pesquisadores na identificação do problema, e buscam soluções para situações reais. Simultaneamente, há produção e uso de conhecimento.

A pesquisa-ação deve ser interativa e requer a cooperação entre o pesquisador e os profissionais da empresa em estudo. Estes estudos são baseados em constantes ajustes oriundos de novas informações e de novos eventos, objetivando um entendimento holístico de um projeto de reconhecida complexidade (Costa, Politano, & Pereira, 2013). Neste caso, o método serve para ilustrar o contexto e compreender o comportamento dos dados gerados pela logística da empresa, e, assim, orientar os gestores na tomada de decisão.

### **3.3 Etapas da pesquisa**



O método de pesquisa utilizado é uma adaptação da abordagem em três grandes fases propostas por Sivarajah et al. (2017) para problemas que envolvem grandes volumes de dados. A primeira fase contemplou: a definição dos objetivos da pesquisa, os quais estão salientados na introdução deste artigo; as limitações da revisão de literatura, que foi baseada em artigos e livros publicados entre 2005 e 2017; método de definição dos objetivos do *dashboard* em conjunto com a liderança da logística da empresa, apresentados na subseção 3.3.1 deste artigo; e a forma de definição das variáveis envolvidas, subseção 3.3.2.

A segunda fase focou em: entrevistas para definição dos objetivos, apresentadas na subseção 4.1; importação da base de dados, subseção 4.2; e revisão dos dados disponíveis pela empresa, identificando, selecionando, avaliando e sintetizando as bases de dados pertinentes. A última etapa da segunda fase foi dividida em duas partes, cada uma contemplando um dos objetivos selecionados na etapa anterior, e são apresentadas nas subseções 4.3 e 4.4.

Por fim, a terceira fase envolveu a divulgação dos resultados gerais da pesquisa e visualização do *dashboard* criado, que estão apresentadas na seção 4.5 do artigo.

### **3.3.1 Definição dos objetivos do dashboard**

A fim de delimitar qual seria o escopo da pesquisa-ação e identificar as variáveis da análise, questionou-se a liderança da logística da empresa sobre os possíveis aperfeiçoamentos nos *dashboards* já existentes. Foram feitas entrevistas presenciais e por Skype com profissionais de diferentes expertises: chefes de transportes rodoviários; assessor de cabotagem; assessor de limítrofes; assessor de ferrovia; gerente de porto; chefes de logística e assessores de planejamento logístico. A pluralidade de funções é necessária para diminuir possíveis distorções que a visão de uma área específica poderia incutir.

### **3.3.2 Definição das variáveis envolvidas**

A partir dos objetivos definidos, as variáveis que interferem nos problemas expostos pela liderança foram selecionadas. O banco de dados que embasa o estudo foi constituído de informações internas referentes aos processos logísticos das usinas

estudadas e são oriundas do SAP da empresa. De acordo com as variáveis selecionadas e de forma alinhada com os objetivos, foi definido o tipo de importação que precisa ser realizado no PBI.

## **4. Resultados**

### **4.1 Objetivos específicos do dashboard**

Com o propósito de definir os objetivos do *dashboard*, foram realizadas entrevistas com a liderança da logística, durante os meses de abril e maio de 2017. Durante essas reuniões, foram listados os *dashboards* já existentes, que incluem relatórios como: volumes faturados para mercado interno e para mercado externo; tempo de permanência dos veículos nas usinas; participação nas entregas a partir dos modais ferroviários e de cabotagem; custo com frete por grupo de produto; custo com frete para mercado interno, para mercado externo e para transferências entre unidades; e preço médio de combustíveis.

A partir desses encontros, relacionando o material então disponível com os objetivos futuros da área, foram salientadas duas necessidades de melhoria: orientação para a priorização de entregas e previsão da necessidade de entregas durante o mês. A caracterização dos dados disponíveis e a análise utilizando a linguagem de programação R acerca do primeiro objetivo, são apresentadas na subseção 4.3 desse artigo, enquanto que as mesmas informações referentes ao segundo objetivo são apresentadas na subseção 4.4.

A justificação das carências levantadas se dá pela falta de padronização e pela rotatividade que esse setor da empresa apresenta. Assim, as decisões acerca da prioridade de entregas e de alocação de recursos para suprir a demanda de volumes a serem expedidos se dá pelo conhecimento tácito dos colaboradores. Dessa forma, o *dashboard* desenvolvido nesta pesquisa é destinado aos chefes de logística, contemplando o planejamento tático e de médio prazo no primeiro objetivo, e aos analistas de transportes, nível operacional, para o segundo objetivo.

### **4.2 Importação da base de dados**

A partir da definição dos objetivos do estudo, foram selecionadas, em conjunto com os assessores de planejamento logístico, quais seriam as variáveis a serem estudadas. Utilizou-se o banco de dados do SAP HANA da empresa e os itens considerados foram: ano da fatura; mês; centro emissor; data de faturamento; data desejada de recebimento da mercadoria; grupo de produtos; modal de envio; código do transporte; tipo de venda; volume e valor líquido da fatura. O período de análise contemplou 17 meses, entre janeiro de 2016 e maio de 2017.

Foram utilizadas 3 *views*, a primeira referente aos transportes para entregas dentro do Brasil e para o exterior, a segunda com os dados das transferências entre as plantas da empresa e, por fim, a *view* com as informações de custos de frete que foram provisionados. Como o período de análise contempla 17 meses, foi necessário realizar a importação de dados para cada ano separadamente e combinar as consultas, a fim de abranger os dois anos de análise. O próximo passo foi a criação de relacionamento entre as tabelas criadas, a partir do uso de chaves comuns.

O PBI possui duas formas de obtenção dos dados a partir do SAP HANA, importação total dos dados ou utilização do *DirectQuery*. Na primeira todas as variáveis selecionadas são copiadas para o PBI *Desktop*, e é possível refinar a busca utilizando o editor de consultas, porém, isso implica em uma nova importação da base. A segunda opção não copia nem importa os dados, o que pode resultar em consultas mais rápidas e garante que as informações apresentadas sejam atuais. Entretanto, as transformações de dados e modelagens são limitadas. Dessa forma, utilizou-se a primeira opção, por possibilitar a criação de novas colunas calculadas e de medidas.

Antes da etapa de análise dos dados, foi necessário retirar todos os dados que não seriam utilizados, como entregas realizadas por centros que não estão no escopo do estudo e erros de digitação (datas 00/00/0000). Posteriormente, deu-se a criação de medidas, como:  $\text{frete} = (\text{custo do frete}) / (\text{volume})$ , em R\$/tonelada; e  $\text{frete/RLV} = (\text{custo do frete}) / (\text{receita líquida de vendas})$ , em %. Paralelamente, foram geradas colunas calculadas, como: tipo de mercado atendido, a partir da organização de vendas; semana de saída da mercadoria, com o uso da função “*weeknum*” do PBI; e nível de serviço, obtido pela subtração da data de fatura na usina pela data solicitada de saída da mercadoria.

### **4.3 Análise do nível de serviço oferecido**

A fim de orientar a priorização das entregas, foi realizada uma análise do nível de serviço oferecido. Primeiramente investigou-se o comportamento das variáveis, explorando o impacto do tempo de entrega nos custos e, posteriormente, foi empregado um algoritmo baseado em *machine learning* para determinar as variáveis mais significativas.

De acordo com a figura 1, o custo com o frete para clientes finais é maior quando há adiantamento ou postergação das entregas, enquanto que o ponto ótimo se encontra próximo à data solicitada de saída da mercadoria.

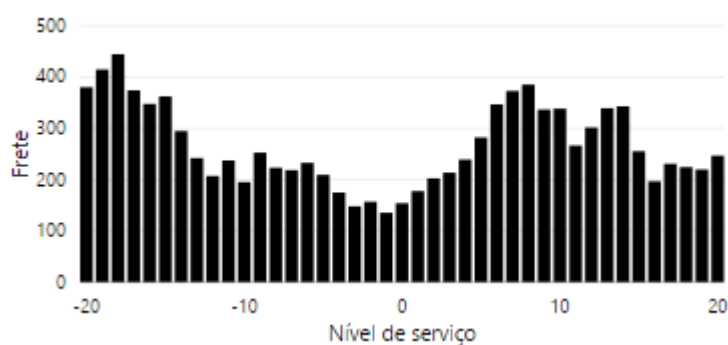


Figura 1 - Custo com frete em R\$/t por nível de serviço

Assim sendo, a representatividade do custo de frete sobre a receita líquida de vendas é menor quando a entrega é realizada na data solicitada, como pode ser observado na figura 2. Quanto mais próxima a saída da usina é da data planejada, melhor é a razão do preço do frete pela receita líquida de vendas (RLV) das entregas para a empresa. Assim, percebe-se que o nível de serviço oferecido pelas usinas influencia tanto na qualidade percebida pelo cliente, sendo atendido no prazo solicitado, quanto na representatividade do custo com frete na RLV.

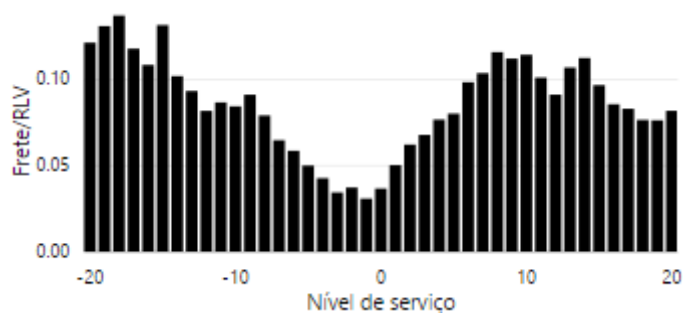


Figura 2 - Frete/RLV por nível de serviço

A fim de auxiliar na priorização das entregas, utilizou-se a biblioteca *random forest* do R para testar quais variáveis têm maiores impactos no nível de serviço (comparação entre a data solicitada pelo cliente e a data de fatura na usina). De acordo com Lantz (2015), este é um dos métodos de *machine learning* mais utilizados atualmente, devido a sua versatilidade e facilidade de uso, quando comparados a outros métodos.

O *random forest* foi defendido por Leo Breiman e Adele Cutler e é um algoritmo classificador baseado no conceito de árvores de decisão. Ele cria várias árvores usando um subconjunto de atributos selecionados aleatoriamente a partir da base de dados original. Este processo gera várias *decision trees* de forma aleatória e sem intervenção humana, cada uma com suas particularidades para realizar a análise (Fontama et al., 2015; Lantz, 2015).

Neste estudo, foram considerados os quesitos: tipo de mercado atendido; modal utilizado para a entrega; grupo de produto e o volume que foi entregue. Os possíveis valores para estes atributos podem ser observados no quadro 1 a seguir.

<b>Fator</b>	<b>Valores</b>
Mercado	Mercado interno, mercado externo, transferência
Modal	Rodoviário, ferrovia, cabotagem, limítrofes
Grupo de produto (GPM)	Ampliados, arames agropecuários, arames industriais, barras & perfis, fio máquina, perfis estruturais, placas, chapas grossas, planos laminados a quente, pregos, tarugo, vergalhão.
Volume	Numérico

Quadro 1 - Valores possíveis para cada fator analisado

A fim de que cada fator seja considerado em diferentes modelos, é necessário que se treine diversas *decision trees*. Assim, foram criadas 500 árvores de decisão utilizando o algoritmo *random forests* para criar o algoritmo. A análise foi feita individualmente para cada usina em estudo.

O gráfico de importância variável é uma ferramenta útil para identificar qual variável, se modificada, tem maior impacto no modelo e pode ser plotado usando a função *varImpPlot*. As principais variáveis são selecionadas e plotadas com base na influência

que apresentam sobre a precisão do modelo, e os resultados para a U8 são apresentados na figura 3 a seguir.

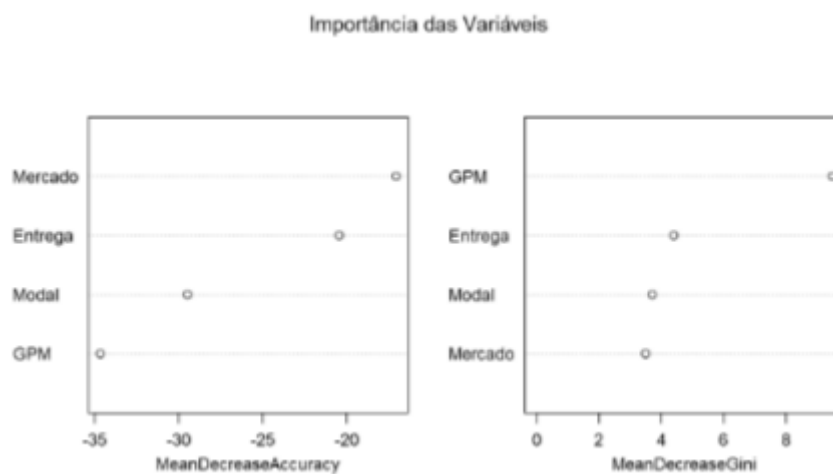


Figura 3 - Importância das variáveis estudadas na U8.

Na figura 3, “*mean decrease accuracy*” refere-se a diminuição média na precisão que uma variável causa quando excluída ou permutada no modelo. Quanto mais a precisão da floresta aleatória diminui devido à remoção de uma única variável, mais importante é a variável e, portanto, as variáveis com grande diminuição média de precisão são mais importantes para a classificação dos dados. Enquanto que “*mean decrease gini*” é uma medida de como cada variável contribui para a homogeneidade dos nós na *random forest* resultante (Fontama et al., 2015; Lantz, 2015).

Como neste estudo o objetivo foi decidir qual variável é mais significativa na priorização de entregas, utilizou-se o “*mean decrease accuracy*” a fim de definir qual variável é a mais importante no modelo. Assim, é possível observar que o fator mais influente, para a oitava usina estudada (U8), no desvio em relação à data estabelecida é o tipo de mercado que será atendido, clientes do mercado interno, mercado externo ou transferências.

As análises foram repetidas para todas as usinas contempladas no estudo, e o resumo pode ser observado na tabela a seguir:

Usina	Variável mais significativa no modelo
U1	Mercado
U2	Mercado
U3	Modal

U4	Modal
U5	Mercado
U6	Modal
U7	Mercado
U8	Mercado
U9	Mercado
U10	Modal

*Tabela 2 - Random Forests - importância das variáveis*

Considerando que nem todas as usinas estudadas utilizam o modal ferroviário ou a cabotagem, é possível observar que as características inerentes à cada usina influenciam na qualidade do nível de serviço oferecido e impacta nos custos de entrega. Com o resultado do *random forest*, cada analista de transporte nas usinas é capaz de priorizar as entregas de modo a progredir com o nível de serviço oferecido.

#### **4.4 Análise da capacidade de expedição diária**

A fim de auxiliar no planejamento dos volumes a serem atendidos, de forma que a usina não ultrapasse o seu limite de expedição, foi realizada uma análise das capacidades. Primeiramente investigou-se o comportamento dos volumes de expedição no período, observando o impacto das entregas no fim de cada mês, e posteriormente foi empregado *HoltWinters* para determinar a previsão de volumes de retirada nos próximos dias.

É possível observar, na figura 4, as características dos envios diários, com vales durante os finais de semana e picos no fim de cada mês, com exceção para dezembro, quando essa característica não é observada.

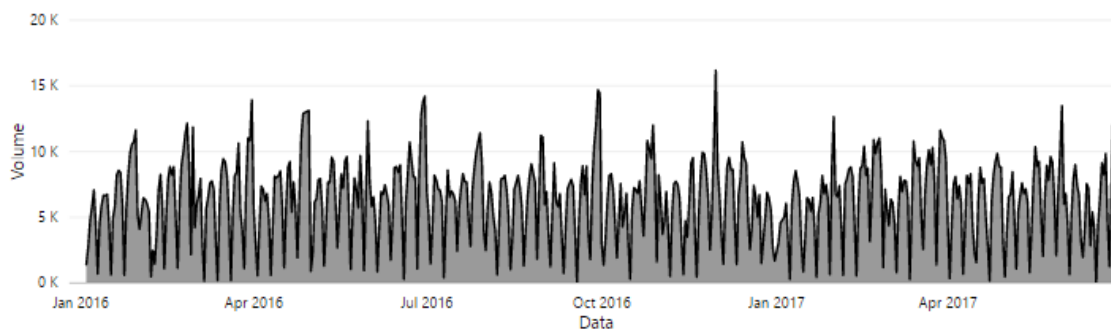


Figura 4 - Volumes de entregas diárias no período

A fim de entender melhor o fenômeno, foi gerado um gráfico que mostra o volume de entregas por dia do mês, somando todas as entregas realizadas no período estudado pelas usinas selecionadas. É válido mencionar que os meses podem acabar nos dias 28, 30 ou 31, assim, as entregas no fim do mês devem abranger esse intervalo. Este gráfico está representado na figura 5. Observa-se que há picos com concentração do volume de entregas realizadas nos últimos dias do mês, representando uma característica comum no setor de postergar as compras e o anseio por aumento de preços no mês seguinte, entre outras razões.

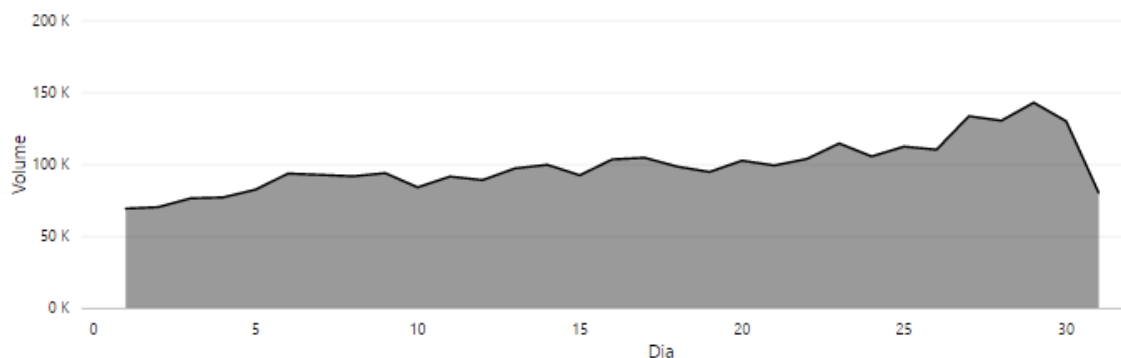


Figura 5 - Volumes de expedição por dia do mês

A fim de auxiliar no planejamento do atendimento às remessas, que são os pedidos de clientes que já estão disponíveis para entrega, de forma que possibilite atender à demanda sem exceder a capacidade de expedição diária das usinas, foi analisada a previsão de volumes para os próximos dias.

Foi utilizada a biblioteca *stats* do R para analisar a previsão de retirada das usinas, e a função *HoltWinters* foi aplicada à serie temporal dos volumes. O método de Holt-



Winters foi sugerido por Holt e Winters, e usa médias móveis ponderadas exponencialmente para atualizar as estimativas da média ajustada sazonalmente (nível), inclinação e sazonalidade. Esses parâmetros podem ser fixados no algoritmo disponível em R ou estimados de modo a minimizar o erro de previsão. Assim, quando os parâmetros não são informados, o algoritmo procura pelos valores que minimizam o erro de previsão (Hyndman et al., 2017). A figura 6 representa em preto os volumes efetivamente realizados na U6, enquanto os valores apresentados na linha vermelha foram previstos pelo modelo.

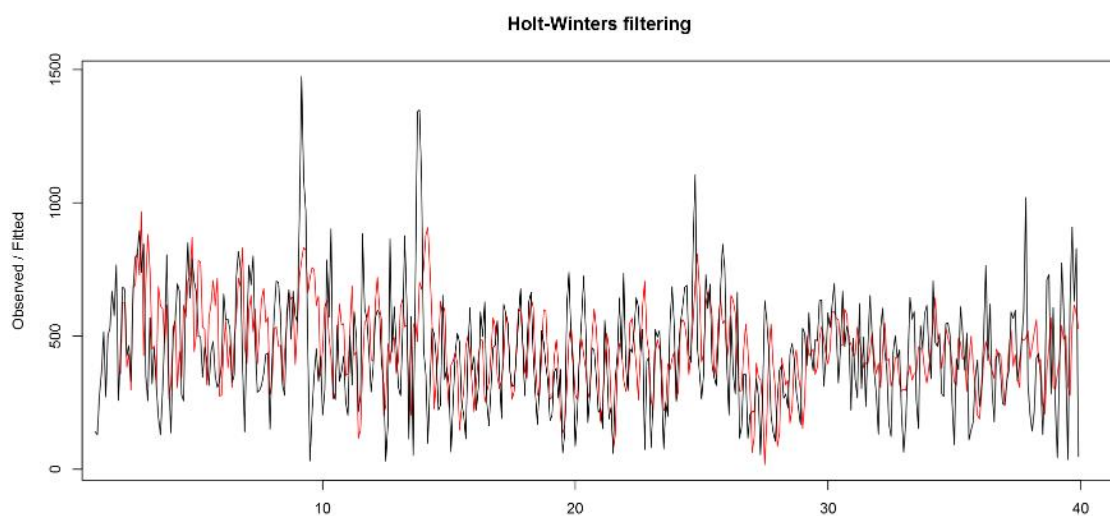


Figura 6 - Volumes realizados X volumes previstos no modelo (U6), em semanas.

A decomposição da série temporal é realizada pela função *fitted* e está apresentada na figura 7, representando a U6. É possível observar em *trend* a tendência de diminuição dos volumes, reflexo da baixa produção do setor siderúrgico brasileiro. Ainda, o gráfico *season* mostra os picos de concentração dos volumes no fim de cada mês.

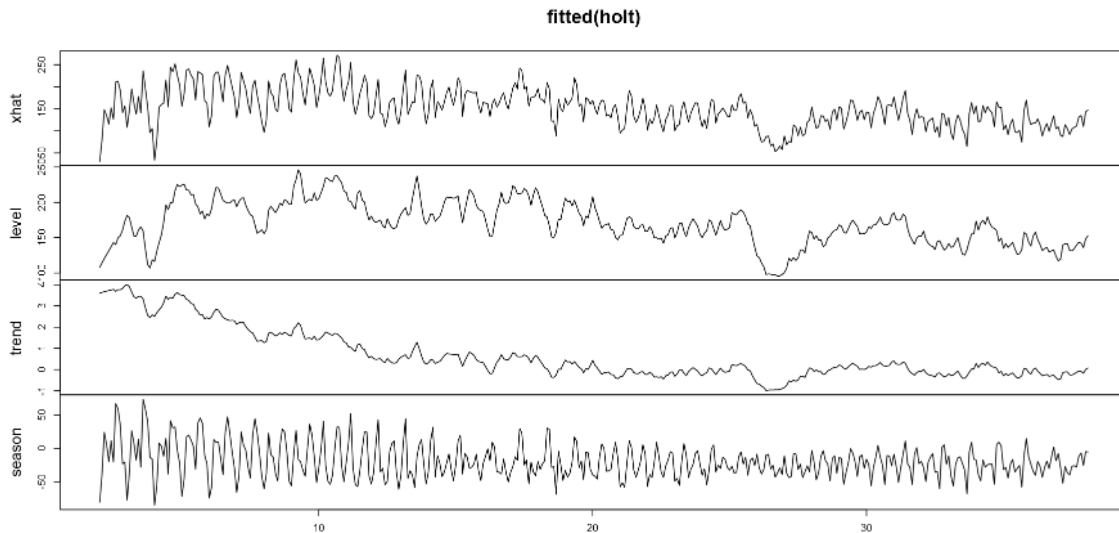


Figura 7 - Decomposição da série temporal em semanas para a U6

Por fim, foi gerado um gráfico para representar a previsão de volumes de retirada da U6 para os próximos 10 dias, utilizando intervalo de confiança de 90%. Representado na figura 8 deste artigo.

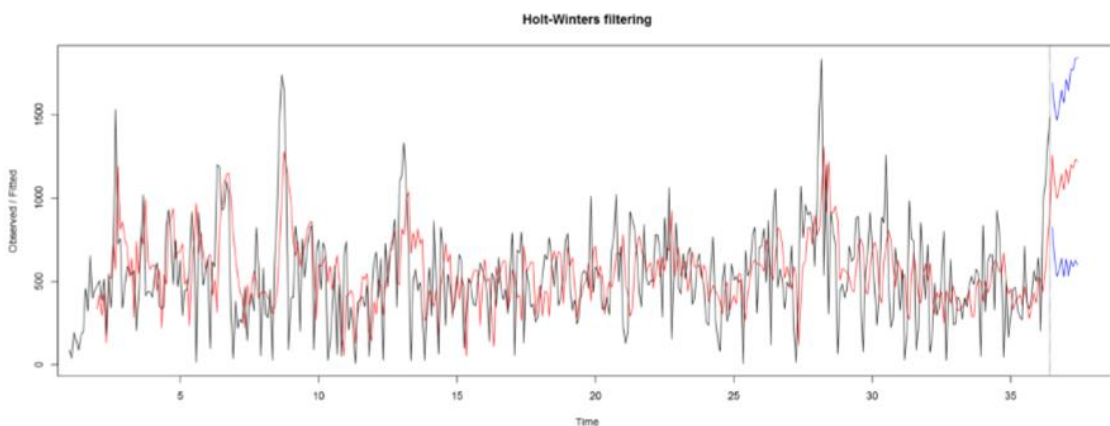


Figura 8 - Previsão de volumes para a U6 nos próximos 10 dias

#### 4.5 Criação final do dashboard

A partir das análises apresentadas nas subseções 4.3 e 4.4, foi elaborado um *dashboard* com previsão para os próximos dias, comparando com a capacidade de expedição diária por usina. Na figura 9 está representada a primeira página do relatório criado.

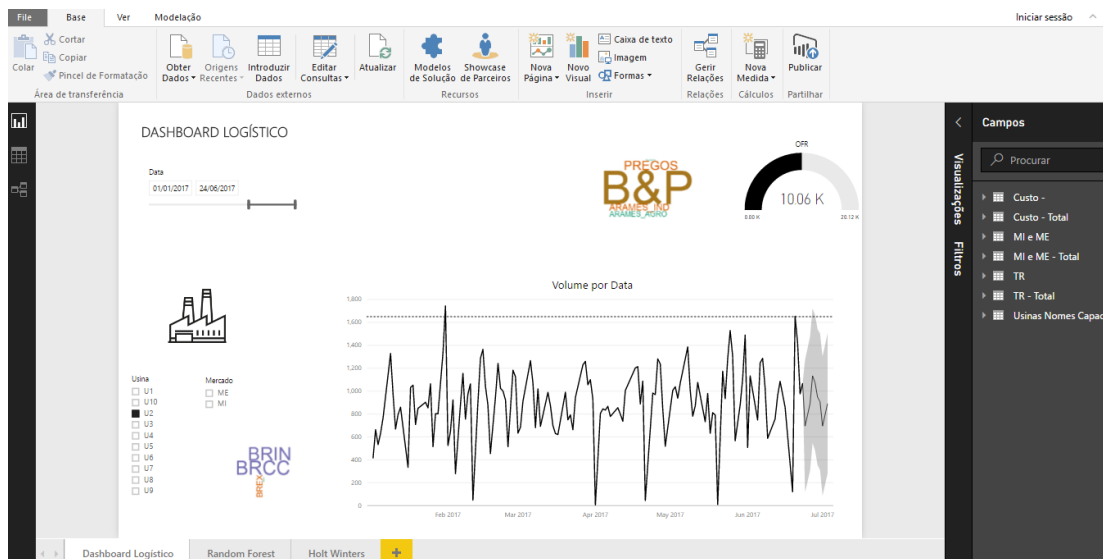


Figure 9 - Visualização do dashboard criado

É possível perceber que, de acordo com a análise temporal dos dados da U2, há a possibilidade de atingir a capacidade máxima de expedição diária da usina. Na página *Random Forest*, é possível filtrar a unidade para obter a variável que mais está impactando no nível de serviço, analisando o gráfico gerado, e assim orientar a prioridade de entregas de forma única e atualizada. Na terceira página, as análises para previsão de demanda a partir da função *HoltWinters* estão disponíveis, sendo necessário apenas selecionar a usina desejada e o período de avaliação.

## 5. Conclusão

No ambiente de negócios dinâmico de hoje, os dados, como um recurso, são uma das principais chaves para sobrevivência no mercado. Os estudos de como utilizá-los para gerar valor tornam-se vitais nas organizações, e os recursos de BI são uma forma de estabelecer um processo de criação de valor mais efetivo e eficiente usando uma combinação de tecnologias modernas.

De acordo com McAfee & Brynjolfsson (2012), uma série de executivos seniores são genuinamente *data driven* (orientados por dados) e dispostos a anular sua própria intuição quando os dados não concordam com eles. Entretanto, em 32% das empresas estudadas, os tomadores de decisão são guiados pela sua experiência e intuição e não nos dados. Com o *dashboard* criado neste estudo, espera-se que os profissionais de logística

da empresa tenham informações pertinentes ao negócio e que auxiliem na tomada de decisão. Visto que este foi um trabalho conjunto com os interessados, uma pesquisa-ação, onde objetivou-se a previsão da demanda futura de volumes para os chefes de logística e a orientação em relação à priorização das remessas para os analistas de transportes, é esperado que as funcionalidades atendam às necessidades e que o engajamento com o relatório seja gradual e sólido.

Os primeiros usuários do *dashboard* criado relataram que ele será útil na priorização de entregas, ao considerar qual sequência trará mais benefício em relação ao frete pago, e na adequação dos recursos necessários para atender a demanda, alinhando os anseios da área de vendas e de transportes. Também, eles mencionaram que desconheciam a possibilidade da integração da linguagem estatística R com o PBI e, ainda, salientaram outras funcionalidades que poderiam ser acrescentadas em relatórios futuros, como: comparativo entre o frete pago pela empresa com o praticado no mercado; clusterização dos clientes para a criação de regiões de entregas mais homogêneas e projeção do *rolling forecast* para pagamento de fretes.

Antes da criação desse *dashboard*, cada gestor de logística alocava os recursos que acreditava serem necessários para atender a demanda, baseados no seu conhecimento tácito. Ainda, seguiam uma ordem de priorização de demanda que é igual para todas as usinas, a qual depende apenas do tipo de mercado que deve ser atendido. Com o relatório, espera-se que os gestores da logística da empresa, ao longo do tempo, percebam valor nas informações e possam tomar decisões baseadas em dados reais e atualizados.

Como solução de *BI Self-Service*, o *Power BI* oferece grande capacidade de análise de dados e de visualização, e atende aos quatro objetivos propostos por Imhoff & White (2011) para essa categoria de software. Entretanto, o PBI apresentou diversas instabilidades ao processar os dados, como o equipamento utilizado possui capacidade de 32 bits, ele não forneceu memória suficiente para algumas análises com o aplicativo. Sugere-se o uso da versão mais recente do programa, publicada em 21/05/2017, em recursos que comporte a versão 64 bits.

A linguagem R, com seu rico conjunto de *libraries* e natureza altamente dinâmica, apresenta desafios únicos para os desenvolvedores de *Self-Service BI*. Um deles é a integração da IDE *RStudio* com o *Power BI*, na qual apenas é possível gerar o código na IDE sendo necessário executá-lo e testá-lo no próprio PBI. Além disso, o tempo máximo

para a execução dos *scripts* não pode exceder cinco minutos, assim, nem todas as técnicas tradicionais de compilação podem ser aplicadas.

Comparando o BI tradicional com o PBI da empresa, pode-se afirmar que o primeiro possui alta escalabilidade e menor tempo de resposta enquanto que o segundo apresentou limitações de processamento já mencionadas. Além disso, nem todos os dados da empresa estão disponíveis para consulta pelo PBI. Assim, embora o *BI Self-Service* possa ajudar a reduzir a dependência do departamento de TI e tornar os usuários mais autossuficientes, ele não é um produto substituto ao BI tradicional.

### Referências Bibliográficas:

- Attaran, M. (2007). RFID: an enabler of supply chain operations. *Supply Chain Management: An International Journal*, 12(4), 249–257. <https://doi.org/10.1108/13598540710759763>
- Bertoni, F., Colombo, M. G., & Grilli, L. (2011). Venture capital investor type and the growth mode of new technology-based firms. *Small Business Economics*, 40(3), 527–552. <https://doi.org/10.1007/s11187-011-9385-9>
- Clark, D. (2017). Introducing Power BI Desktop. In *Beginning Power BI: A Practical Guide to Self-Service Data Analytics with Excel 2016 and Power BI Desktop* (pp. 193–216). Berkeley, CA: Apress. [https://doi.org/10.1007/978-1-4842-2577-6\\_9](https://doi.org/10.1007/978-1-4842-2577-6_9)
- Costa, E. P., Politano, P. R., & Pereira, N. A. (2013). *Exemplo de aplicação do método de Pesquisa - ação para a solução de um problema de sistema de informação em uma empresa produtora. Gestão e Produção*.
- Fontama, V., Barga, R., & Tok, W. H. (2015). *Predictive Analytics with Microsoft Azure Machine Learning 2nd Edition*. <https://doi.org/10.1017/CBO9781107415324.004>
- Gerhardt, T. E., & Silveira, D. T. (2009). *Métodos de Pesquisa* (1ª). Porto Alegre: Editora da UFRGS.
- Handfield, R., Straube, F., Pfohl, H.-C., & Wieland, A. (2013). *Trends and strategies in logistics and supply chain management. DVV Media Group GmbH*. Hamburgo, Alemanha: DVV Media Group GmbH. <https://doi.org/10.1086/596757>
- Hazen, B. T., Boone, C. A., Ezell, J. D., & Jones-Farmer, L. A. (2014). Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *International Journal of Production Economics*, 154, 72–80. <https://doi.org/10.1016/j.ijpe.2014.04.018>
- Hyndman, R. J., & Athanasopoulos, G. (2013). *Forecasting: principles and practice*. Melbourne, Australia: OTexts: <http://otexts.org/fpp/> (acesso em 26/05/2017).
- Hyndman, R., O'Hara-Wild, M., Bergmeir, C., Razbash, S., & Wang, E. (2017).

- Forecasting Functions for Time Series and Linear Models Description*. Retrieved from <http://github.com/robjhyndman/forecast> VignetteBuilder
- Imhoff, B. C., & White, C. (2011). *Self-Service Business Intelligence: Empowering Users to Generate Insights*.
- Instituto Aço Brasil. (2016). *Relatório de Sustentabilidade*. Rio de Janeiro.
- Lantz, B. (2015). *Machine Learning with R. Packt* (2nd ed., Vol. 1). Birmingham, Reino Unido. <https://doi.org/10.1017/CBO9781107415324.004>
- Lu, M. (2014). *Discovering Microsoft Self-service BI solution : Power BI*. Haaga-Helia - University of Applied Sciences.
- Makris, S., & Chryssolouris, G. (2013). Web-services-based supply-chain-control logic: an automotive case study. *International Journal of Production Research*, *51*(7), 2077–2091. <https://doi.org/10.1080/00207543.2012.701781>
- Matende, S., & Ogao, P. (2013). Enterprise Resource Planning (ERP) System Implementation: A Case for User Participation. *Procedia Technology*, *9*, 518–526. <https://doi.org/10.1016/j.protec.2013.12.058>
- McAfee, A., & Brynjolfsson, E. (2012). Big Data: The Management Revolution. Retrieved April 15, 2017, from <https://hbr.org/2012/10/big-data-the-management-revolution>
- Necula, E. (2015). Analyzing traffic patterns on street segments based on GPS data using R. *Transportation Research Procedia*, *10*(July), 276–285. <https://doi.org/10.1016/j.trpro.2015.09.077>
- Novaes, A. N., Takebayashi, F., & Briesemeister, R. (2015). Cross-Docking em centros logísticos de distribuição urbana: considerações sobre operação e modelagem. *Transportes*, *23*(1), 47. <https://doi.org/10.14295/transportes.v23i1.795>
- Paiola, M., Sacconi, N., Perona, M., & Gebauer, H. (2013). Moving from products to solutions: Strategic approaches for developing capabilities. *European Management Journal*, *31*(4), 390–409. <https://doi.org/10.1016/j.emj.2012.10.002>
- Prater, E., Frazier, G. V., & Reyes, P. M. (2005). Future impacts of RFID on e-supply chains in grocery retailing. *Supply Chain Management: An International Journal*, *10*(2), 134–142. <https://doi.org/10.1108/13598540510589205>
- Schallow, J., Ludevig, J., Schmidt, M., & Deuse, J. (2014). Zukunftsperspektiven der Digitalen Fabrik. *Digitale Fabrik, Planung, Produktionsmanagement*, *104*(H.3), 139–145.
- Schmidt, D., Chen, W. C., Matheson, M. A., & Ostrouchov, G. (2016). Programming with BIG Data in R: Scaling Analytics from One to Thousands of Nodes. *Big Data Research*, *8*, 1–11. <https://doi.org/10.1016/j.bdr.2016.10.002>
- Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, *70*, 263–286. <https://doi.org/10.1016/j.jbusres.2016.08.001>
- Tan, K. C. (2001). A framework of supply chain management literature. *European Journal of Purchasing & Supply Management* *7* (2001) 39-48, *7*(February 1999).
- Tripp, D. (2005). Action research: a methodological introduction. *Educação E Pesquisa*,

*São Paulo*, 31(3), 443–466.

Tseng, Y., Yue, W. L., & Taylor, M. A. P. (2005). The Role of Transportation in Logistic Chain. *Proceedings of the Eastern Asia Society for Transportation Studies*, 5, 1657–1672.