

**PROJETO DE PESQUISA: DESENVOLVIMENTO DE ABORDAGENS DE  
BUSINESS ANALYTICS NA PREVISÃO DE ATRASOS POR  
CONGESTIONAMENTO NO AEROPORTO DE CONGONHAS**

**PROPONENTE: Dr. RODRIGO ARNALDO SCARPEL**

**DEPARTAMENTO DE GESTÃO E SUPORTE À DECISÃO, DIVISÃO DE  
CIÊNCIAS FUNDAMENTAIS (IEF), INSTITUTO TECNOLÓGICO DE  
AERONÁUTICA (ITA)**

**RESUMO**

*Business Analytics* é definido como o estudo das competências, tecnologias e práticas usadas para avaliar continuamente as estratégias e operações das diferentes áreas das organizações, visando a obtenção de conhecimentos e orientações que auxiliem no seu planejamento. Tais estudos são executados utilizando diferentes métodos como análises operacionais e estatísticas, modelos preditivos e de classificação e técnicas de otimização. O presente projeto tem por objetivo desenvolver abordagens de *Business Analytics*, integrando métodos de previsão, de classificação e de formação de agrupamentos, por meio de técnicas de programação matemática, visando identificar padrões temporais na ocorrência de atrasos por congestionamento no aeroporto de Congonhas. Desde que o transporte aéreo no Brasil foi liberalizado, uma de suas consequências foi a concentração dos voos em alguns *hubs*. Embora a criação de *hubs* pareça benéfica às empresas de transporte aéreo e ofereça algumas vantagens aos viajantes, a concentração excessiva de voos em um *hub* pode resultar em impactos econômicos negativos denominados atrasos por congestionamento, os quais aumentam o tempo total de viagem dos passageiros e o custo operacional das empresas.

**Palavras-chave:** métodos de previsão, métodos de classificação, métodos de formação de agrupamentos, programação matemática, atrasos por congestionamento; análise de séries temporais.

**PROJECT TITLE: DEVELOPMENT OF BUSINESS ANALYTICS APPROACHES FOR PREDICTING CONGESTION DELAYS AT CONGONHAS AIRPORT**

**Dr. RODRIGO ARNALDO SCARPEL**

**DEPARTAMENTO DE GESTÃO E SUPORTE À DECISÃO, DIVISÃO DE CIÊNCIAS FUNDAMENTAIS (IEF), INSTITUTO TECNOLÓGICO DE AERONÁUTICA (ITA)**

**ABSTRACT**

Business Analytics is defined as the study of the skills, technologies and practices used to evaluate organization-wide strategies and operations continuously to obtain insights and guide the business planning of an organization. Such studies are performed using different approaches as statistical and operations analysis, predictive and classification modeling and optimization. The goal of this project is to develop Business Analytics approaches by integrating forecasting, classification and clustering methods using mathematical programming techniques in order to identify temporal patterns on the occurrence of congestion delays at Congonhas airport. Since the liberalization of air transport in Brazil, one of the consequences of this process was the concentration of flights in a few hubs. Although hubbing seems to benefit airlines and offers some advantages to travelers, the extent of excessive concentration at a hub can result in some negative economic impacts, namely, congestion delay which increases passenger's total travel time and airlines' operating costs.

**Keywords:** forecasting methods, classification methods, clustering methods, mathematical programming, congestion delays, time series analysis.

## 1. ENUNCIADO DO PROBLEMA

O ambiente competitivo atual, caracterizado pela contínua inovação dos produtos e dos processos e pelas constantes mudanças no ambiente interno e externo, demanda uma sistemática e consistente transformação de dados brutos, coletados através de sistemas de informação, pesquisas de mercado e, até, da internet, em conhecimento útil para a tomada de decisões. Foi nesse ambiente competitivo e com os propósitos elencados que surgiu a área de *Business Analytics* (BA). BA é definido como o estudo das competências, tecnologias e práticas usadas para avaliar continuamente as estratégias e operações das diferentes áreas das organizações visando a obtenção de conhecimentos e orientações que auxiliem no seu planejamento (Chen *et al.*, 2012). Tais estudos são executados utilizando diferentes métodos como análises operacionais e estatísticas, modelos preditivos e de classificação e técnicas de otimização. De acordo com Wang *et al.* (2016), os métodos de BA podem ser enquadrados em 3 categorias: descritivos, preditivos e prescritivos. O foco deste projeto está no desenvolvimento de abordagens que integrem métodos descritivos e preditivos.

Os métodos descritivos têm por objetivo identificar problemas e oportunidades nos processos e funções existentes na organização. Para esse fim, destacam-se os métodos de redução de dimensão, de visualização de dados e de formação de agrupamentos. Os métodos de redução de dimensão auxiliam na eliminação da redundância dos dados e na geração de bases de dados com menos variáveis. Métodos estatísticos multivariados como análise de componentes principais e análise fatorial são muito utilizados para esse fim (Mingoti, 2005). Em relação aos métodos de visualização de dados, estes também geram redução de dimensão permitindo a representação de observações em gráficos com poucas dimensões (Cardoso Jr. e Scarpel, 2012; Cardoso Jr. e Scarpel, 2014). No que diz respeito aos métodos de formação de agrupamentos, estes lidam com a identificação de grupos homogêneos (Scarpel *et al.*, 2015). Por esses métodos agrupam-se observações de um conjunto de dados formando subgrupos, nos quais as observações são mais similares entre si do que as observações contidas em outros subgrupos. Segundo Duda *et al.* (2001), o problema de geração de agrupamentos é o de alocar  $n$  observações ou pontos  $\{x^1, x^2, \dots, x^n\}$  no espaço  $p$ -dimensional  $R^p$  em  $k$  agrupamentos. São vários os métodos de geração de agrupamentos, como: os métodos aglomerativos ou hierárquicos, nos quais os agrupamentos são criados baseando-se em uma matriz de dissimilaridade; os métodos partitivos (Resende e Scarpel, 2008), nos

quais dividem-se os pontos objetivando minimizar algum critério de erro, os métodos de mistura e os métodos fuzzy, em que as observações pertencem a todos os agrupamentos com diferentes graus de pertinência.

Nos métodos preditivos, o objetivo é identificar padrões nos dados disponíveis (gerados pelos processos e funções existentes nas organizações) visando a explicação e/ou a previsão dos fenômenos identificados. Para esse fim, destacam-se os métodos de previsão e de classificação. Os métodos de previsão tratam da geração de funções ou modelos numéricos que representem os fenômenos de interesse. Segundo Duda *et al.*(2001), em métodos de previsão o objetivo é estimar uma função  $g$  que mapeie os pontos ou vetores de atributos do espaço de entrada  $X$  para o espaço de saída  $Y$ , a partir de um conjunto finito de amostras,  $\{x_i, y(x_i)\}_{i=1}^N$ . Portanto, pretende-se prever o valor da variável independente ( $Y$ ) utilizando o valor dos atributos, ou variáveis dependentes ( $X$ ). O conjunto finito de amostras também é conhecido como conjunto de treinamento. Um grande número de abordagens para tratar esse problema foi proposto como os métodos regressão (linear simples, linear múltipla, não-linear, entre outros) e, mais recentemente, o uso de árvores de regressão (Scarpel, 2014), de composição de especialistas locais (Scarpel, 2013) e de composição de especialistas globais integrados (Scarpel e Milioni, 2007; Scarpel e Milioni, 2012; Scarpel, 2015).

Já os métodos de classificação buscam a identificação de modelos que atribuem observações às classes pré-especificadas. Segundo Duda *et al.* (2001), os problemas em modelagem preditiva podem ser diferenciados pela forma do espaço de saída  $Y$ . Se a variável a ser prevista for contínua, i.e.,  $Y \in \mathbb{R}$ , então, o problema de previsão é um problema de regressão. Já se a variável a ser prevista for discreta, i.e.,  $Y = \{0, 1, \dots, v-1\}$  em que  $v$  é o número de classes, então, o problema de previsão é um problema de classificação. São vários os métodos de classificação disponíveis na literatura como a análise discriminante paramétrica e não paramétrica, a regressão logística (Scarpel e Milioni, 2001), programação inteira (Loucopoulos, 2001) e o de support vector machine (Bradley e Mangasarian, 2000; Scarpel, 2008; Silva e Scarpel, 2007).

Existem grandes desafios de pesquisa no desenvolvimento de abordagens de *Business Analytics* no auxílio a tomada de decisão em sistemas complexos. No que diz respeito aos atrasos por congestionamento, em aeroportos brasileiros, uma das consequências da liberalização do transporte aéreo no Brasil foi a concentração dos voos em alguns hubs. De acordo com Wensveen (2011), embora a criação de hubs

pareça benéfica às empresas de transporte aéreo e ofereça algumas vantagens aos viajantes, a concentração excessiva de voos em um hub pode resultar em alguns impactos econômicos negativos denominados atrasos por congestionamento, os quais aumentam o tempo total de viagem dos passageiros e o custo operacional das empresas. Picos de congestionamento também causam grandes problemas aos aeroportos e ao pessoal das empresas aéreas e demandam trabalho adicional aos controladores de tráfego aéreo (Wensveen, 2011). Diferentes alternativas podem ser empregadas para prevenir atrasos por congestionamento como a redução de passageiros em conexão, a imposição de restrições à utilização dos *Slots* e a limitação de voos, de acordo com a capacidade máxima do aeroporto. Porém, para que essas medidas sejam aplicadas faz-se necessária a previsão na ocorrência de atraso por congestionamento.

## **2. RESULTADOS ESPERADOS**

O presente projeto tem por objetivo desenvolver abordagens de *Business Analytics*, integrando métodos de previsão, de classificação e de formação de agrupamentos, por meio de técnicas de programação matemática, visando identificar padrões temporais na ocorrência de atrasos no Aeroporto de Congonhas.

Esse tipo de integração já foi empregado no meu projeto de pesquisa do doutorado e na ocasião foi proposta uma forma de otimizar a formação de agrupamentos em composição de especialistas locais. Tal abordagem se mostrou bastante útil na criação de modelos de previsão e foi empregada em previsão de passageiros aéreos (Scarpel, 2013), no desenvolvimento de um modelo de classificação para detecção precoce da ocorrência de atrasos por congestionamento no Aeroporto Internacional de São Paulo (Scarpel e Pelicioni, 2018) e em planejamento de demanda (Scarpel, 2015).

No que diz respeito ao Aeroporto de Congonhas, tradicionalmente seus passageiros são viajantes a negócios e altamente sensíveis ao fator tempo (Miranda e Oliveira, 2018) e este foi o primeiro aeroporto a ter coordenação de *Slots*. Segundo a Agência Nacional de Aviação Civil (ANAC, 2019), a coordenação de *Slots* tem como objetivo reger o acesso à infraestrutura em aeroportos saturados, buscando harmonizar a demanda por operações aéreas com a capacidade aeroportuária disponível. No Aeroporto de Congonhas, após o acidente ocorrido com o voo 3054 da TAM, no dia 17 de julho de 2007, as autoridades da aviação introduziram um conjunto de restrições na sua operação, dentre as quais a redução do número de *Slots*, permitindo a realização de

30 movimentações (pousos e decolagens) por hora, visando a redução do seu congestionamento. A Figura 1 mostra a evolução temporal do número de movimentações diárias ocorridas no Aeroporto de Congonhas entre 2.000 e 2.018.

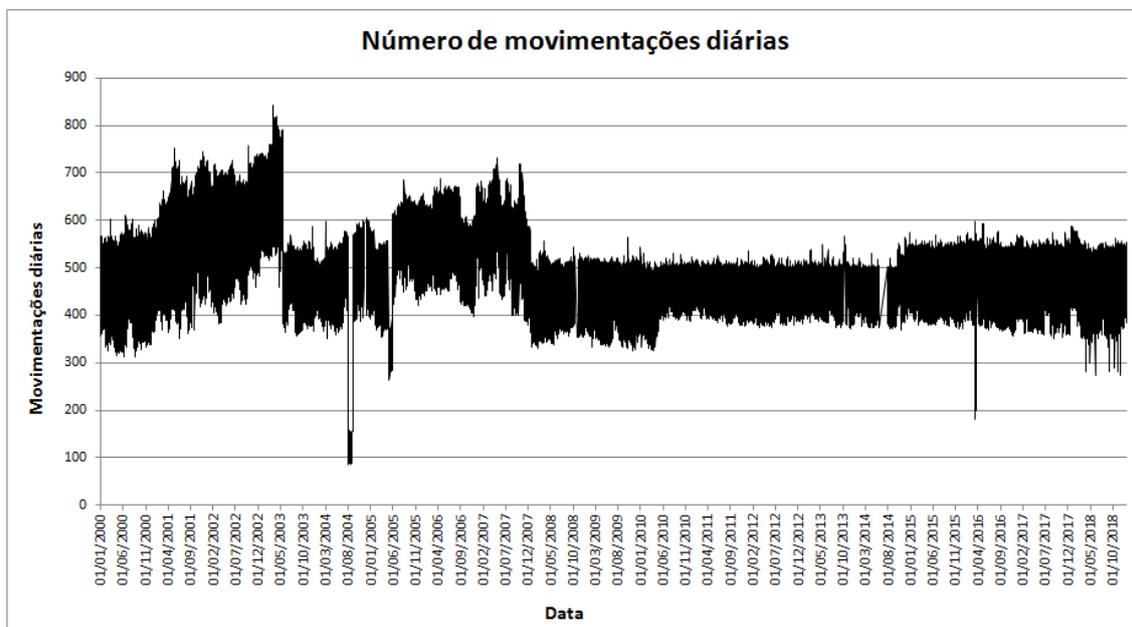


Figura 1 – Evolução temporal do número de movimentações diárias no Aeroporto de Congonhas entre 01/2000 e 12/2018.

Percebe-se, a partir da Figura 1, a formação de 3 períodos distintos (padrões): do início do ano 2.000 até agosto de 2.007, de setembro de 2.007 a novembro de 2.014 e de dezembro de 2.014 até o final de 2.018. No primeiro período indicado não havia coordenação dos *Slots* e verifica-se um crescente número de movimentações diárias. No segundo período, com o início da coordenação dos *Slots*, verifica-se uma redução do número diário de movimentações, que se torna um processo estacionário com uma média de 478,83 movimentações por dia (29,92 movimentações por hora, abaixo do limite estabelecido de 30 movimentações por hora) e desvio-padrão de 61,4 movimentações por dia. No terceiro período a ANAC aumentou o número de movimentações permitidas nos horários de pico de 30 para 33 movimentações por hora e, desta forma, verifica-se um aumento do número diário de movimentações, mas ainda caracterizado como um processo estacionário, com média de 505,97 movimentações por dia (31,62 movimentações por hora) e desvio-padrão de 69,0 movimentações por dia.

A Figura 2 mostra o *boxplot* da fatia de movimentações atrasadas (com atraso maior que 15 minutos) do Aeroporto de Congonhas, nos 3 períodos indicados e o

histograma da fatia de movimentações atrasadas no terceiro período (entre dezembro de 2.014 e dezembro de 2.018).

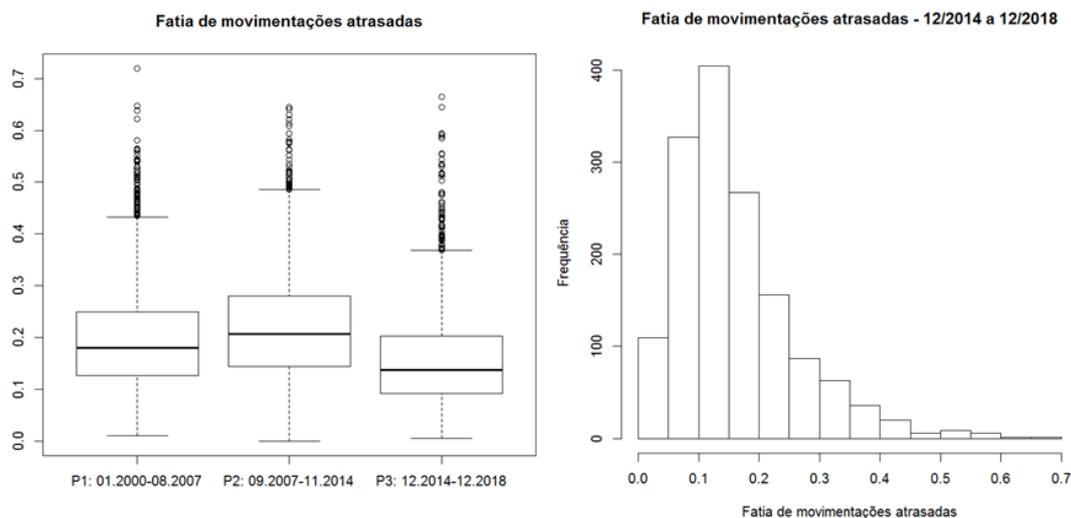


Figura 2 – *Boxplot* da fatia de movimentações atrasadas nos períodos 01/2000 a 08/2007, 09/2017 a 11/2014 e 12/2014 a 12/2018 e histograma da fatia de movimentações atrasadas entre 12/2014 e 12/2018.

Embora, de acordo com Miranda e Oliveira (2018), o sistema de coordenação de *Slots* implantado pela ANAC no Aeroporto de Congonhas tenha reduzido a ocorrência de atrasos de voos, se considerarmos o critério internacional de classificar como atrasada uma movimentação que ocorreu mais de 15 minutos depois do previsto, verifica-se a partir da Figura 2 que mesmo depois da coordenação dos *Slots* há uma grande parcela dos dias com mais de 25% do voos atrasados. Portanto, é de grande relevância a identificação da ocorrência de padrões temporais de atraso e a identificação dos fatores que geram esses atrasos e a antecipação, por meio de previsão, da ocorrência de dias com alta concentração de voos atrasados. Quanto aos fatores geradores de atrasos, nesse projeto serão considerados os efeitos que práticas realizadas pelas empresas de transporte aéreo, como a concentração de *Slots*, o entesouramento de *Slots*, do inglês *Slots hoarding* (Sheng *et al.*, 2019) e a crescente prática de aumentar artificialmente a duração dos voos (do inglês *schedule padding*), resultam na ocorrência de dias com alta concentração de voos atrasados.

### 3. DESAFIOS CIENTÍFICOS E TECNOLÓGICOS E OS MEIOS E MÉTODOS PARA SUPERÁ-LOS

Para alcançar os objetivos delineados, pretende-se fazer uso de uma estrutura metodológica devidamente alinhada a estes (desenvolver abordagens de *Business Analytics*, integrando métodos de previsão, de classificação e de formação de agrupamentos, por meio de técnicas de programação matemática, para identificar padrões temporais na ocorrência de atrasos no Aeroporto de Congonhas). No que diz respeito aos desafios científicos e tecnológicos, o projeto pretende gerar contribuições tanto em termos de métodos de BA, visto que são raros, na literatura, trabalhos que lidam com métodos de classificação e de formação de agrupamentos utilizando dados de séries temporais, quanto na área de apoio a decisão em transporte aéreo, pois a literatura é bastante escassa em trabalhos que tratam do impacto gerado pelas práticas realizadas pelas empresas de transporte aéreo, como o entesouramento de *Slots*, na ocorrência de atrasos. A orientação da investigação segue uma linha fundamentalmente quantitativa e serão seguidos quatro estágios centrais:

i. *Pesquisa bibliográfica*: a partir do tema de pesquisa indicado, pretende-se realizar uma pesquisa bibliográfica completa dos métodos de BA existentes na literatura e das práticas realizadas pelas empresas de transporte aéreo (como a concentração dos *Slots*, o entesouramento dos *Slots* e o *Schedule padding*).

ii. *Análise crítica dos métodos existentes*: a partir da pesquisa bibliográfica realizada, pretende-se identificar lacunas na literatura e oportunidades para integração de métodos de previsão, classificação e de formação de agrupamentos.

iii. *Proposição de novas abordagens*: visando suprir as lacunas identificadas, pretende-se propor abordagens de BA integrando, por meio de programação matemática, métodos de previsão, de classificação e de formação de agrupamentos. Essa forma de integração é possível, pois os métodos de BA se baseiam fortemente em métodos estatísticos, em inteligência artificial e em aprendizagem de máquina, e podem ser formulados como problemas de otimização (Padmanbhan e Tuzhilin, 2003).

iv. *Aplicação e teste das abordagens propostas* (visando avaliar tanto a consistência, quanto a aplicabilidade das abordagens propostas no auxílio a tomada de decisão para o problema de pesquisa identificado). Para a execução desse estágio se faz necessária a formação das bases de dados e analisar sua consistência, uma vez que, os

dados que serão empregados estão divididos em bases de dados anuais com todos os pousos e decolagens programados e ocorridos ou não (cancelados) para todos os aeroportos brasileiros. Assim, esta primeira etapa demandará um grande trabalho de ETL (extração, transformação e carga dos dados) para selecionar apenas os dados do Aeroporto de Congonhas, para unir todos os anos que serão empregados, eliminar os voos cancelados, estimar e identificar a ocorrência de atrasos por movimentação, estimar e identificar as práticas realizadas pelas empresas de transporte aéreo e fazer análises para avaliar a consistência desses dados. Com a base de dados pronta, serão feitas transformações dos dados e análises para caracterizar as séries temporais e posteriormente fazer uso das abordagens propostas.

#### **4. CRONOGRAMA**

1. PESQUISA BIBLIOGRÁFICA
2. ANÁLISE CRÍTICA DOS MÉTODOS EXISTENTES
3. PROPOSIÇÃO DE NOVAS ABORDAGENS
4. APLICAÇÃO E TESTE DAS ABORDAGENS PROPOSTAS:
  - 4.1 Formação e consistência da base de dados
  - 4.2 Transformação dos dados
  - 4.3 Análises preliminares e discussão dos resultados obtidos
  - 4.4 Utilização dos métodos de BA propostos
5. TRABALHOS PARA PUBLICAÇÃO: deseja-se confeccionar trabalhos para publicação ao longo do segundo ano do projeto.
6. RELATÓRIOS:
  - 6.1 Relatório parcial
  - 6.2 Relatório final: será confeccionado nos últimos dois meses do projeto.

ATIVIDADE	BIMESTRES											
	1	2	3	4	5	6	7	8	9	10	11	12
PESQUISA BIBLIOGRÁFICA	*	*	*					*	*			
ANÁLISE CRÍTICA DOS MÉTODOS EXISTENTES		*	*	*				*	*			
PROPOSIÇÃO DE NOVAS ABORDAGENS				*	*	*			*	*		
<b>APLICAÇÃO E TESTE DAS ABORDAGENS PROPOSTAS:</b>						*	*	*	*	*	*	*
Formação e consistência da base de dados						*	*					
Transformação dos dados							*					
Análises preliminares e discussão dos resultados obtidos								*	*			
Utilização dos métodos propostos									*	*	*	
RELATÓRIO PARCIAL						*						
TRABALHOS PARA PUBLICAÇÃO							*	*	*	*	*	*
RELATÓRIO FINAL												*

## 5. DISSEMINAÇÃO E AVALIAÇÃO

O desenvolvimento do projeto de pesquisa prevê orientações, no horizonte de 2 anos, de:

- 2 ou mais alunos fazendo trabalho de conclusão do curso de graduação
- 1 ou mais alunos de iniciação científica
- 1 ou mais alunos de mestrado

Em termos de publicações estimadas, cada trabalho de graduação e cada iniciação científica deverá gerar pelo menos uma publicação em um congresso nacional, com publicação do artigo completo nos anais dos referidos congressos (dependendo de aceite e/ou da política da entidade organizadora do evento).

Das dissertações de mestrado espera-se uma produção mais ambiciosa, na forma de um texto que deverá ser submetido para apresentação em um congresso internacional, com publicação integral nos anais e se a qualidade do texto assim justificar, ele poderá ser submetido a uma revista técnica especializada.

## 6. OUTROS APOIOS

Não há outros apoios a esse projeto.

## 7. BIBLIOGRAFIA

AGÊNCIA NACIONAL DE AVIAÇÃO CIVIL. **Coordenação de Slots ANAC.**

Disponível em: < <https://www.anac.gov.br/assuntos/setor-regulado/empresas/slot/coordenacao-de-slot>>. Acesso em: 21 fev. 2019.

BRADLEY, P. S. and MANGASARIAN, O. L. Massive data discrimination via linear support vector machines. **Optimization Methods and Software**, 13, p.1–10, 2000.

CARDOSO JR., M. M. and SCARPEL, R. A. Cognitive structure of occupational risks represented by a perceptual map. **WORK-A Journal of Prevention Assessment & Rehabilitation**, 41 (1 - Supplement IEA 2012), p. 3196–3201, 2012.

CARDOSO JR., M. M. and SCARPEL, R. A. Multiple Perceptual Map Generation using MDSvarext. **Global Journal of Researches in Engineering**, 14, 1–8, 2014.

CHEN, H., CHIANG, R.H. and STOREY, V.C. Business intelligence and analytics: from big data to big impact. **MIS Quarterly**, 4 (36), 1165–1188, 2012.

DUDA R.O., HART, P.E. and STORK, D.G. **Pattern Classification**, 2<sup>nd</sup> edition, John Wiley & Sons, Inc. New York, 2001.

LOUCOPOULOS, C. Three-group classification with unequal misclassification costs: a mathematical programming approach. **Omega**, 29, 291–297, 2001.

MINGOTI, S. **Análise de Dados Através de Métodos de Estatística Multivariada: Uma Abordagem Aplicada**. Editora UFMG, 2005.

MIRANDA, V. A. P. OLIVEIRA, A.V. M. Airport slots and the internalization of congestion by airlines: An empirical model of integrated flight disruption management in Brazil. **Transportation Research. Part A**, 116, 201–219, 2018.

PADMANABHAN, B. e TUZHILIN, A., On the use of optimization for data mining: Theoretical interactions and eCRM opportunities. **Management Science**, vol. 49(10), p. 1327-1343, 2003.

RESENDE, C. B. and SCARPEL, R. A. A New Mathematical Programming Formulation to Deal with the Clustering Problem. **In: 4<sup>th</sup> International Conference on Production Research – ICPR Americas'**, 2008.

SCARPEL, R. A. Previsão de insolvência de empresas utilizando support vector machine. **Revista de Economia e Administração**, 7, 281–295, 2008.

- SCARPEL, R.A. Forecasting Air Passengers at São Paulo International Airport using a Mixture of Local Experts Model. **Journal of Air Transport Management**, 26, 35–39, 2013.
- SCARPEL, R. A. A demand trend change early warning forecast model for the city of São Paulo multi-airport system. **Transportation Research. Part A**, 65, 23–32, 2014.
- SCARPEL, R. A. An integrated mixture of local experts model for demand forecasting. **International Journal of Production Economics**, 164, 35–42, 2015.
- SCARPEL, R. A., LADAS, A. and AICKELIN, U. Indebted Households Profiling: A Knowledge Discovery from Database Approach. **Annals of Data Science**, 2, 43–59, 2015.
- SCARPEL, R. A. e MILIONI, A. Z. Aplicação de modelagem econométrica à análise financeira de empresas. **Revista de Administração (USP)**, 36 (2), 80–88, 2001.
- SCARPEL, R. A. e MILIONI, A. Z. Otimização na formação de agrupamentos em problemas de composição de especialistas. **Pesquisa Operacional**, 27, 85–104, 2007.
- SCARPEL, R. A. and MILIONI, A. Z. Integrated mixture of local experts model for forecasting. **International Journal of Business, Humanities and Technology**, 2, 2012.
- SCARPEL, R. A. and PELICIONI, L. C. A data analytics approach for anticipating congested days at the São Paulo International Airport. **Journal of Air Transport Management**, 72, 1–10, 2018.
- SILVA, V. D. e SCARPEL, R. A. Detecção de fraudes na distribuição de energia elétrica utilizando support vector machine. **Investigação Operacional**, 27, 139–150, 2007.
- SHENG, D., LI, Z. C. and FU, X. Modeling the effects of airline slot hoarding behavior under the grandfather rights with use-it-or-lose-it rule. **Transportation Research Part E**, 122, 48–61, 2019.
- WANG, G., GUNASEKARAN, A. NGAI, E. W.T. and PAPADOPOULOS, T. Big data analytics in logistics and supply chain management: Certain investigations for research and applications. **International Journal of Production Economics**, 176, 98–110, 2016.
- WENSVEEN, J. G., **Air transportation: A management perspective**, 7<sup>th</sup> edition. Aldershot, UK: Ashgate, 2011.