



Faculdade de Administração, Contabilidade e Economia  
Departamento de Economia  
Mestrado em Economia do Setor Público

**MÁQUINAS DE SUPORTE VETORIAL COMO  
INSTRUMENTO DE PRIORIZAÇÃO DE INVESTIMENTOS APLICADO AO  
PROGRAMA DE INVESTIMENTO E LOGÍSTICA – AEROPORTOS**

**RUBEM OLIVEIRA DE PAULA**

**ORIENTADOR: DR. DONALD MATTHEW PIANTO**

SETEMBRO, 2016.



**RUBEM OLIVEIRA DE PAULA**

**MÁQUINAS DE SUPORTE VETORIAL COMO  
INSTRUMENTO DE PRIORIZAÇÃO DE INVESTIMENTOS APLICADO AO  
PROGRAMA DE INVESTIMENTO E LOGÍSTICA – AEROPORTOS**

Dissertação apresentada como requisito para obtenção do grau de Mestre pelo Programa de Mestrado em Economia do Setor Público, do Departamento de Economia da Faculdade de Administração, Contabilidade e Economia, Universidade de Brasília e Escola de Administração Fazendária do Ministério da Fazenda.

Orientador: Dr. Donald Matthew Pianto

Brasília, setembro de 2016.



Dissertação de Mestrado, sob o título Máquinas de suporte vetorial como instrumento de priorização de investimentos aplicado ao programa de investimento e logística – aeroportos, defendida por Rubem Oliveira de Paula e aprovada em setembro de 2016, em Brasília, Distrito Federal, pela banca examinadora constituída pelos doutores:

Doutor, Donald Matthew Pianto  
Professor Orientador

Doutora, Maria Eduarda Tannuri-Pianto  
Professora-Examinadora

Doutor, Rodrigo Otávio Moreira da Cruz  
Examinador Externo



Para minha esposa Alice e  
todos os que atuam para  
um serviço público melhor.





## **Agradecimentos**

Agradeço a minha esposa, Alice, que nem mesmo durante a gravidez, deixou de mostrar capacidade irrestrita de incentivo e auxílio nos mais distintos momentos.

Agradeço à minha família, aos familiares originais, minha mãe, Marta, meu pai, Hermes, minha irmã, Dalva; e os familiares agregados, meus cunhados Raphael, Thaís e meu concunhado Alysson, meus sogros, Ana e Nelson. Todos sempre atentos às mais diversas mudanças e cenários, cada um em seu respectivo ramo, e que me estimulam aos mais diversos e profícuos debates.

Agradeço às equipes da Secretaria de Aviação Civil, Leandro, Marcelo e Victor, ao Secretário Leonardo Cruz, e ao então Diretor Rodrigo, que auxiliaram com invaloráveis debates e sugestões. Aos amigos da Secretaria Executiva, Guilherme Ramalho, Thiago Pedroso, Cláudia Kattar e Fabiana Todesco, que desde sempre apoiaram a iniciativa e incentivaram de diversas formas os estudos e a pesquisa.

Por fim agradeço ao corpo docente do MESP-UnB, pela atenção e desprendimento na aplicação desse programa de mestrado, especialmente ao Professor Donald Pianto, pela atenção e acolhimento, e que mesmo em momento de grande dificuldade pessoal atendeu às minhas solicitações com precisão e empenho em me mostrar novos caminhos e formas de se olhar para tudo o que eu colocava.



## Resumo

Este estudo aborda a utilização de modelos de máquinas de suporte vetorial (SVM) como instrumento de priorização de investimentos públicos. Parte-se da apresentação das principais políticas públicas de investimento em aeroportos do Governo Federal Brasileiro: Programa de Investimento em Logística – Aeroportos e Plano Aeroviário Nacional. Para formação das bases e aplicação da metodologia, foi realizada a agregação dos municípios e informações de nível municipal em Zonas Aeroviárias. Foram utilizados *kernels* do tipo radial, mínimos quadrados radial, polinomial, linear e processo gaussiano de base radial, com diferentes parâmetros. Formulou-se um indicador de classificação de aeroportos de acordo com a operação ou não de voos regulares no período de 2004 a 2015, para posterior classificação de aeródromos passíveis ou não de receberem voos regulares. A partir de uma análise comparativa, observou-se que, em geral, os modelos SVM apresentaram resultados superiores ao modelo de mínimos quadrados ordinários, com especial diferença em relação ao perfil das Zonas Aeroviárias classificadas como receptoras de voos regulares.

Palavras chave: máquinas de suporte vetorial, priorização de investimentos, aeroportos regionais.



## **Abstract**

This study discusses the use of support vector machines' (SVM) models as an instrument for prioritization of public investment. It starts presenting the main investment policies on airports carried by the Brazilian Federal Government: Programa de Investimento em Logística – Aeroportos e Plano Aeroviário Nacional. To constitute the data basis on which the application of the methodology was carried out, the municipalities and municipal level information are aggregated in Zonas Aeroviárias. Kernels of the radial type, least square radial, polynomial, linear and radial basis gaussian process, with different parameters were implemented. An index for airports classification according to the regularity of operation of flights has been formulated, taking the period from 2004 to 2015 as basis, to be applied to classify aerodromes by whether being subject or not to receive regular flights. After application on the databases and comparison of results from different models, it was observed that in general SVM models showed better results than an OLS model with particular difference to the profile of the Zonas Aeroviárias classified as receivers of scheduled flights.

Key words: support vector machines, investment prioritization, regional airports.



## Sumário

1	Introdução .....	17
2	O setor de aviação civil brasileiro, o Programa de Investimento em Logística: Aeroportos e o Plano Aeroviário Nacional .....	19
2.1	O Programa de Investimento em Logística: Aeroportos.....	22
2.2	O Plano Aeroviário Nacional .....	25
3	Aprendizado de máquina e máquinas de suporte vetorial .....	27
3.1	Aprendizado supervisionado e não supervisionado .....	28
3.2	Capacidade, overfitting e underfitting .....	29
3.3	Máquinas de suporte vetorial. ....	31
3.4	Classes misturadas e o artifício dos kernels.....	34
3.5	Métodos de reamostragem. ....	38
4	Aplicação da metodologia.....	40
4.1	Constituição da base de dados.....	40
4.2	O pacote <i>caret</i> desenvolvido para R. ....	46
4.3	Avaliação e escolha de modelos. ....	49
5	Resultados .....	53
5.1	Aprendizado de máquina como instrumento de tomada de decisão e priorização de investimentos.....	60
6	Conclusões.....	62
7	Bibliografia .....	64
8	Anexo I: programação utilizada em R para os modelos SVM. ....	66
9	Anexo II: Lista das ZAs, classificações e localização da sede.....	69
10	Anexo III: Modelo de Mínimos Quadrados Ordinários. ....	87





## 1 Introdução

Cabe aos servidores públicos a constante busca pela inovação na gestão pública, nas mais diversas áreas de atuação e com necessidade ainda maior quando se trata do aprimoramento de técnicas que auxiliam o planejamento e um melhor uso de recursos públicos. Este trabalho visa à exploração de potenciais de tecnologia de aprendizado de máquina como ferramenta de planejamento e priorização de investimentos, com enfoque no que se busca com a implementação do Programa de Investimento em Logística: Aeroportos (PIL) e na elaboração do Plano Aeroviário Nacional. O PIL tem como horizonte de investimentos de curto a médio prazo, um período de aproximadamente cinco anos, enquanto o Plano Aeroviário Nacional, ainda em construção, é um instrumento de gestão a ser utilizado no longo prazo, até 30 anos.

O aprendizado de máquina é um campo do conhecimento voltado ao desenvolvimento e utilização de técnicas, algoritmos que possibilitem que programas computacionais sejam capazes de aprimorar seus resultados com experiência (MITCHELL, 1997). O aprendizado de máquina vem ganhando atenção nas mais distintas áreas do conhecimento, em estudos de mapeamento genético, nas mais diferentes áreas que buscam interpretação de imagens, como na busca por reconhecimento facial em segurança, em finanças, na constituição de carteiras de investimento, por empresas de vendas pela internet, com reconhecimento de padrões de usuários e consumidores, e diversos outros ramos da academia, da indústria e dos negócios.

No setor público, como ferramental de auxílio a investimentos, encontram-se poucos trabalhos em literatura, como em Chou e Lin (2012), que aborda a questão de utilização de modelos de classificação para previsão de investimentos em projetos de parcerias público-privadas. Em investimentos privados, com ênfase em investimentos em ativos do mercado financeiro, em contraposição, há literatura abundante em técnicas de aprendizado de máquinas para previsão de valores de ativos, formulação de carteiras de investimento, entre outras aplicações (MARCELINO, 2016). Para o planejamento da aviação civil, ainda há pouca utilização (fora a já citada utilização em segurança aeroportuária, que é inclusive um dos grandes impulsionadores de métodos

de reconhecimento facial)<sup>1</sup>, como no caso de previsão de demanda em aeroportos com métodos de sistemas neuro-difusos (XIAO *et ali.*, 2014).

Este estudo apresenta brevemente o panorama da aviação civil e da evolução da movimentação de passageiros ao longo dos últimos 12 anos e aborda as principais políticas do governo federal para o setor, nomeadamente o PIL e o PAN de forma resumida no capítulo 2. O capítulo 3 trata dos principais conceitos de aprendizado de máquina, com foco em máquinas de suporte vetorial (SVM), utilização de *kernels*. É importante salientar desde já que o trabalho não tem enfoque na questão de ajustes finos ou proposição de novos métodos estatísticos ou de programação utilizados em aprendizado de máquina, tendo, portanto, enfoque na aplicação de modelos com ajustes básicos de parâmetros. Não há assim, necessidade de conhecimento aprofundado nas questões mais técnicas, matemáticas, estatísticas ou computacionais para entendimento do que é exposto.

No capítulo 4 é exibida a aplicação da metodologia, com descrição das bases de dados utilizadas, uma breve referência ao pacote estatístico *caret* desenvolvido para a plataforma R e são delineados alguns parâmetros que auxiliam na escolha de modelos de SVM. O capítulo 5 relata os resultados mais relevantes advindos da aplicação dos modelos, sugerindo formas de utilização de SVM como instrumento de seleção e priorização de investimentos em aeroportos. O capítulo 6 conclui expondo alguns comentários a respeito das possibilidades futuras de aplicação de modelos de aprendizado de máquina para o planejamento no setor público. A bibliografia compõe o capítulo 7.

No anexo I encontra-se a programação utilizada para rodar os modelos SVM em R, com utilização do pacote *caret*. O Anexo II traz a lista de aeroportos que compõem a rede atual de aeroportos que operaram voos regulares, conforme definido no capítulo 2.

---

<sup>1</sup> Considerando todos os grandes temas que influenciam a aviação civil, há ainda a utilização de aprendizado de máquina nas previsões climáticas pelo reconhecimento de mudanças atmosféricas, no reconhecimento por empresas aéreas e sites de turismo que direcionam determinadas formas de publicidade de acordo com o perfil de utilização da internet pelos usuários.

## 2 O setor de aviação civil brasileiro, o Programa de Investimento em Logística: Aeroportos e o Plano Aeroviário Nacional

A aviação civil brasileira constitui um setor que a despeito da história relativamente longa, que remonta ao início do século XX, só passou a ser regulada e tratada por uma entidade pública civil em meados da década de 2000, com a criação da ANAC pela Lei 11.182, de 27 de setembro de 2005, posteriormente vinculada à Secretaria de Aviação Civil da Presidência da República (SAC), esta criada pela Lei 12.462, de 4 de agosto de 2011. Em 2016 a Secretaria de Aviação Civil foi integrada ao Ministério dos Transportes, que passou a ser formalmente denominado Ministério dos Transportes, Portos e Aviação Civil, a partir da publicação da Medida Provisória nº 726 de 12 de maio de 2016, restando inalterada a vinculação da ANAC à SAC.

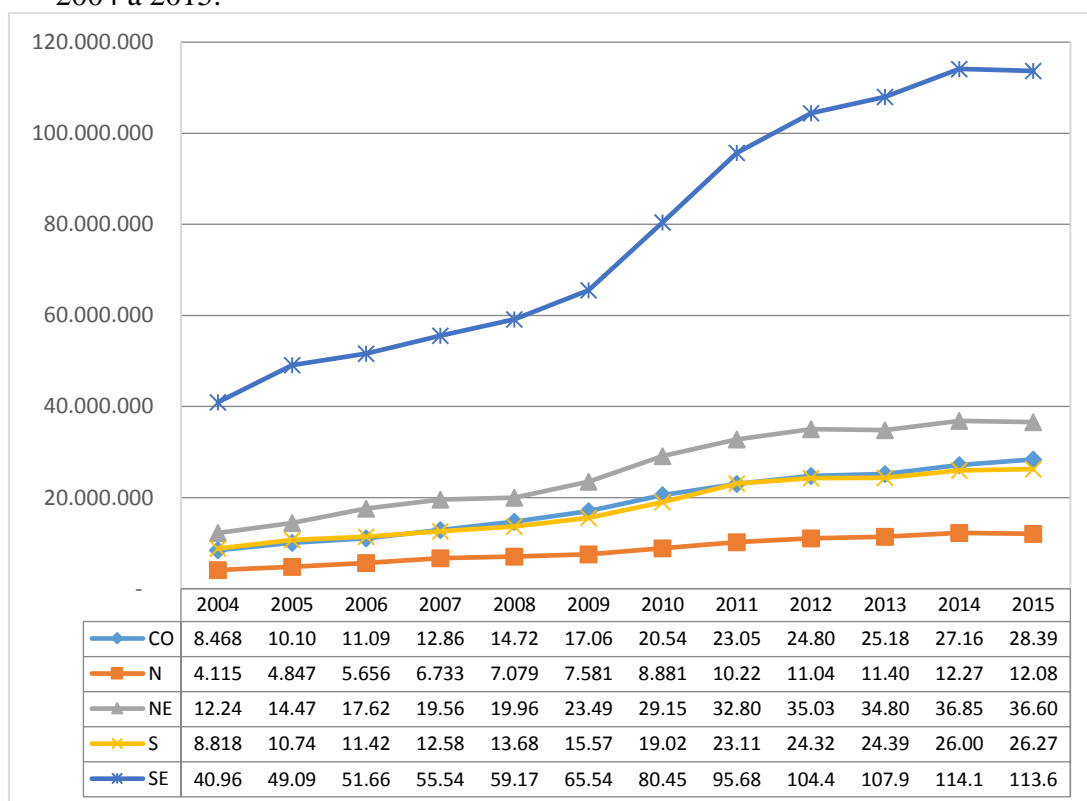
Com a criação e transferência das competências dos militares aos órgãos civis de Aviação Civil, o setor passou a desenvolver novas políticas e contar com maior autonomia para formulação e execução dessas políticas. O primeiro passo foi a implementação de regulação específica pela ANAC, procurando se adequar a padrões internacionais da Organização para a Aviação Civil Internacional (ICAO na sigla em inglês), braço vinculado à Organização das Nações Unidas (ONU) que trata do setor. O principal reflexo foi a instituição da liberdade tarifária e operação de rotas discricionariamente ofertadas pelas empresas.

Atribui-se a essa nova forma de atuação do estado no setor a redução do preço médio real das tarifas aéreas domésticas de 48,8% entre 2005 e 2014, o que, juntamente com o aumento real do salário mínimo e crescimento do PIB, refletiram em um aumento da movimentação de passageiros no mesmo período, de acordo com o Anuário do Transporte Aéreo de 2014 (ANAC, 2015). Em termos numéricos, a movimentação de passageiros saltou de menos de 75 milhões em 2004, para aproximadamente 217 milhões em 2015, ou seja, crescimento 191%, que equivale a um incremento médio anual de 10%. No mesmo período pode-se observar que o crescimento médio anual em países como China, Índia, Rússia, também com grandes áreas e populações, superou a casa dos dois dígitos, com respectivamente 12%, 14% e 10%. Países com maior nível de desenvolvimento e renda per capita, por sua vez,

tiveram crescimento menor. Tomando-se, como exemplo os países membros da Organização para Cooperação e Desenvolvimento Econômico tiveram um incremento anual médio de 3%, e os da União Europeia, 4%. (Banco Mundial, 2016)

A figura 1 abaixo ilustra o crescimento da movimentação de passageiros nos aeroportos brasileiros no período de 2004 a 2015, com desagregação por região. Por outro lado, a desregulamentação do setor e atuação livre por parte do setor privado deixou uma lacuna em relação à prestação de serviços aéreos regionais, antes subsidiados. Desenvolveu-se, então, o que é hoje o principal conjunto de políticas públicas voltados para o setor de aviação civil, o Programa de Investimento em Logística: Aeroportos (PIL).

Figura 1: Movimentação de passageiros nos aeroportos brasileiros por região, 2004 a 2015.



Fonte: elaboração própria com dados do sistema Hórus, da SAC/MT.

A Figura 1 apresenta a evolução dos valores de movimentação de passageiros nas diferentes regiões do país. Nota-se que o crescimento da movimentação se distribui de forma relativamente uniforme, com os valores para o ano de 2015 em torno de 3 vezes superiores aos valores para 2004, sendo o maior aumento relativo o da região Centro-Oeste (235%), e o menor o da região Sudeste (177%). Saliente-se que a movimentação

nesta última região é, durante todo o período, superior à soma da movimentação das demais regiões (em 2015, estes valores são de 113,6 milhões para o Sudeste e 103,3 milhões para as demais regiões somadas).

O PIL visa à ampliação da atuação em parceria com o setor privado na aviação civil, especialmente pela concessão de aeroportos, que se encontra ao tempo de realização deste trabalho, na quarta rodada, mudança da forma de atuação da Empresa Brasileira de Infraestrutura Aeroportuária, INFRAERO, e o investimento em aeroportos regionais. O PIL tem aplicação no médio prazo, ou seja, aproximadamente 5 anos. Para o longo prazo, está se desenvolvendo o Plano Aeroviário Nacional, que trará diretrizes para os próximos 20 a 30 anos para os investimentos em infraestrutura do setor. Ambas as políticas serão tratadas a seguir.

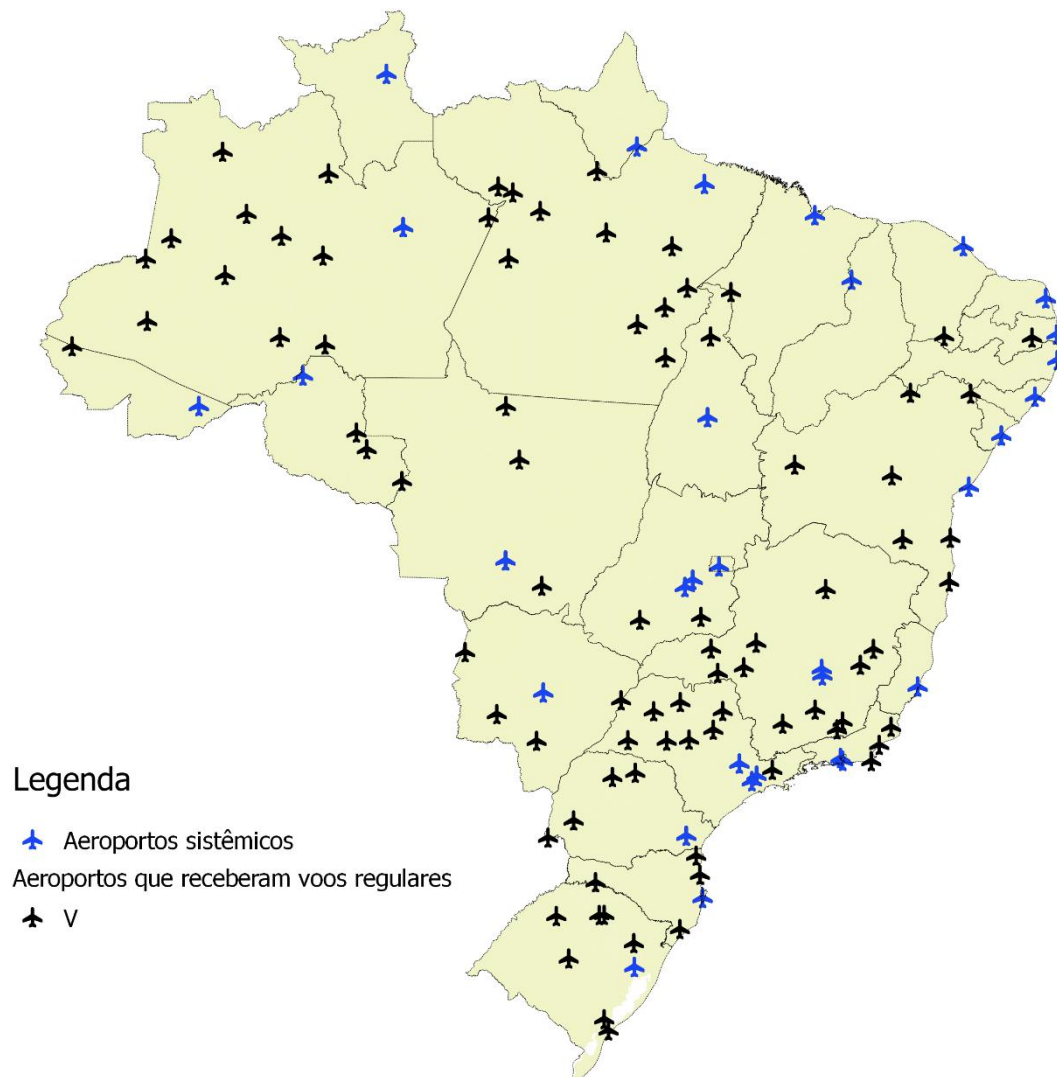
O programa procura ampliar a atual rede de aeroportos, constituída por 31 aeroportos de capitais e regiões metropolitanas, e aproximadamente 80 aeroportos regionais. Um número que varia diariamente à medida que se atualizam as informações do Hotran, sistema da ANAC utilizado para processamento de autorizações de voos.<sup>2</sup> Abaixo, o Mapa 1 apresenta a atual rede, considerando todos os aeroportos sistêmicos e os que receberam voos regulares em ao menos três anos ao longo do período de 2004 a 2015.

Aqui cabe esclarecer uma definição específica para este trabalho e que será abordada ao longo de todo o texto: entende-se por aeroporto, ou Zona Aeroviária, que recebe ou tem potencial para receber voos regulares, aquele que tem opera ao menos uma chegada e uma partida de uma aeronave modelo ATR-42, com taxa de ocupação de 80%. Essa escolha é decorrente da premissa inicial de que os aeroportos que constituem o PIL teriam infraestrutura que atendessem os requisitos mínimos à operação desse modelo de aeronave, e 80% é um valor aproximado da taxa de ocupação de empresas aéreas (ANAC, 2015).

---

<sup>2</sup> O Hotran é atualizado em tempo real, à medida que as empresas enviam solicitações ou cancelamentos de voos agendados para a ANAC, e está disponível para acesso público em: <http://www2.anac.gov.br/hotran/>.

Mapa 1 Aeroportos sistêmicos e que receberam voos regulares no período de 2004 a 2015.



Fonte: Fonte: elaboração própria com dados da ANAC e do sistema Hórus, SAC/MT.

## 2.1 O Programa de Investimento em Logística: Aeroportos.

Em dezembro de 2012 o Governo Federal Brasileiro lançou o Programa de Investimento em Logística: Aeroportos. O PIL contém, entre as áreas de atuação, incentivos à aviação regional, divididos em três diferentes frentes de atuação: i) investimento em infraestrutura aeroportuária; ii) subsídios a rotas regionais; e iii) aperfeiçoamento da gestão com atuação em parceria com Estados e Municípios e investimentos em formação de profissionais do setor.

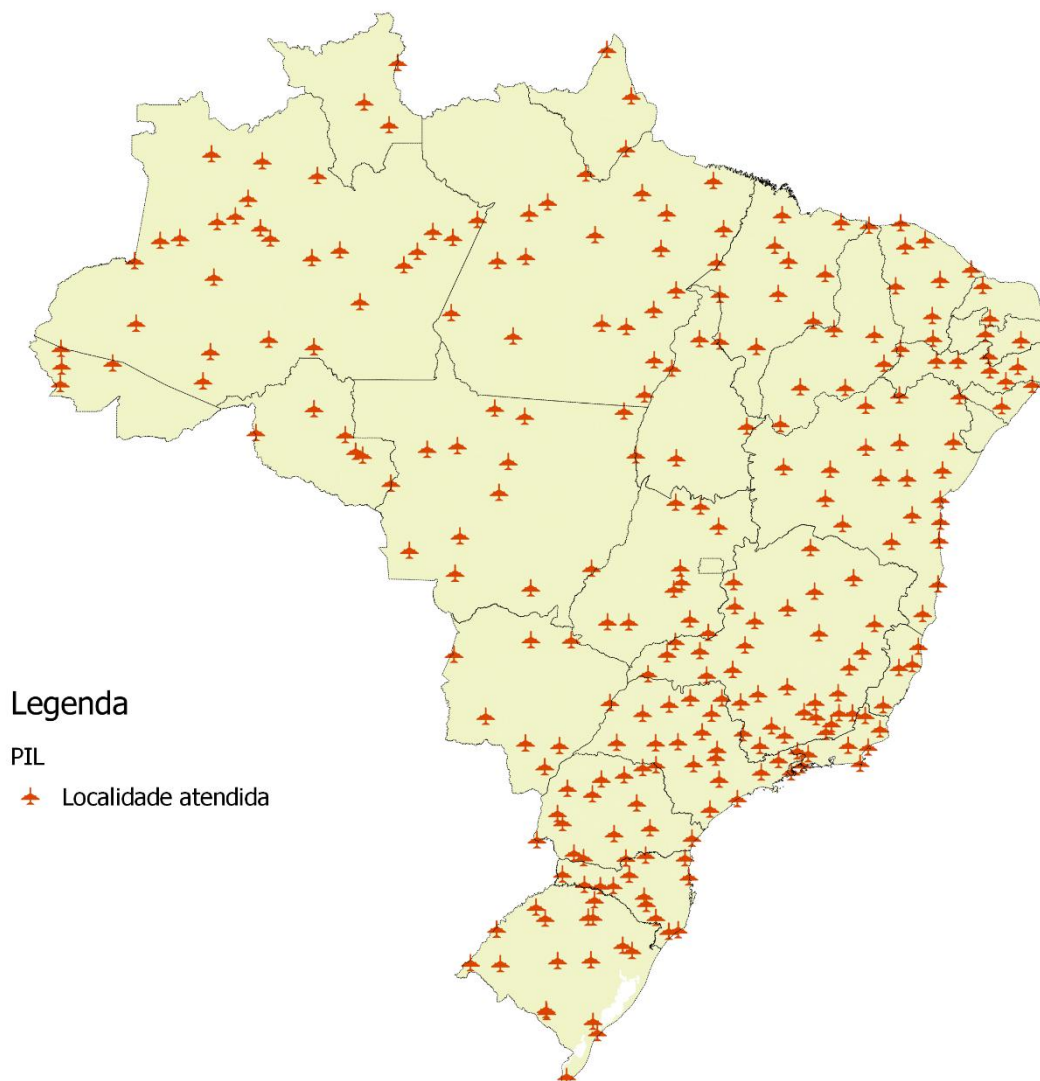
Os investimentos previstos serão, inicialmente, voltados a 270 localidades que deveriam ser atendidas diretamente por aeroportos selecionadas a partir de quatro critérios pertinentes aos municípios em que se localizam, quais sejam: a) municípios indutores do turismo; b) municípios polo de desenvolvimento de acordo com a definição dada pelo estudo Regiões de Influência das Cidades, REGIC-IBGE; c) municípios isolados; e d) municípios que já possuem voos regulares ou nos quais empresas aéreas apresentaram interesse em atuar.

O programa, na vertente à infraestrutura, objetiva adequar os aeródromos dessas 270 localidades, com reforma de aeroportos existentes ou construção de novos aeroportos de acordo com os usos potenciais determinados pela dinâmica que se espera nas respectivas regiões. Para tanto, são previstos investimentos no montante de R\$ 7,2 bilhões ao longo de aproximadamente 5 anos, portanto o PIL é voltado ao médio prazo. Os recursos serão provenientes do Fundo Nacional de Aviação Civil (FNAC). O Mapa 2 traz as localidades inicialmente incluídas no PIL.

Interessa apontar que antes do modelo de investimento proposto no âmbito do PIL, os aeroportos regionais recebiam recursos financeiros da esfera federal apenas por meio do Programa Federal de Auxílio a Aeroportos – PROFAA, instituído pela Lei 8.399, de 7 de janeiro de 1992, para o atendimento às necessidades de implantação, melhoramento, reaparelhamento, reforma ou ampliação dos aeroportos de interesse estadual ou regional. No PROFAA, cuja atuação se dá com a celebração de convênios, os estados e municípios acionam o Governo Federal por meio de propostas, que são avaliadas e priorizadas pela SAC. (BRASIL, 2016)

Em relação à frente de subsídios a rotas regionais, foi instituído, pela lei 13.097, de 19 de janeiro de 2015, o Programa de Desenvolvimento da Aviação Regional (PDAR), que tem como principal objetivo, aumentar o acesso da população brasileira ao sistema de transporte aéreo, com prioridade aos residentes nas regiões menos desenvolvidas do País, considerando tanto o aumento do número de Municípios e rotas atendidos por transporte aéreo regular, como o número de frequências das rotas regionais operadas regularmente. O programa está em fase de detalhamento, e para o ano de 2016 conta com um orçamento de cinquenta milhões de reais.

Mapa 2 - 270 localidades a serem atendidas que compõem a seleção do Programa de investimento em Logística: Aeroportos.



Fonte: Fonte: elaboração própria com dados da ANAC e do sistema Hórus, SAC/MT.

Quanto à gestão, a SAC desenvolve diversos cursos voltados à ampliação da oferta de mão de obra especializada para atuação nos aeródromos, abrangendo áreas como gestão e administração aeroportuária, segurança, bombeiros de aeródromos, entre outras. Foi instituído também o Plano Geral de Outorgas, pela portaria SAC, de 14 de agosto de 2014, que facilita o processo de outorga junto à União de aeródromos públicos por Estados e Municípios. Ainda em fase de estudos diversos, o PIL deve, em um futuro próximo necessitar de documentos que subsidiem tomadas de decisão,



especialmente as relacionadas a priorização de investimentos, para que se otimizem os gastos públicos em infraestrutura no setor.

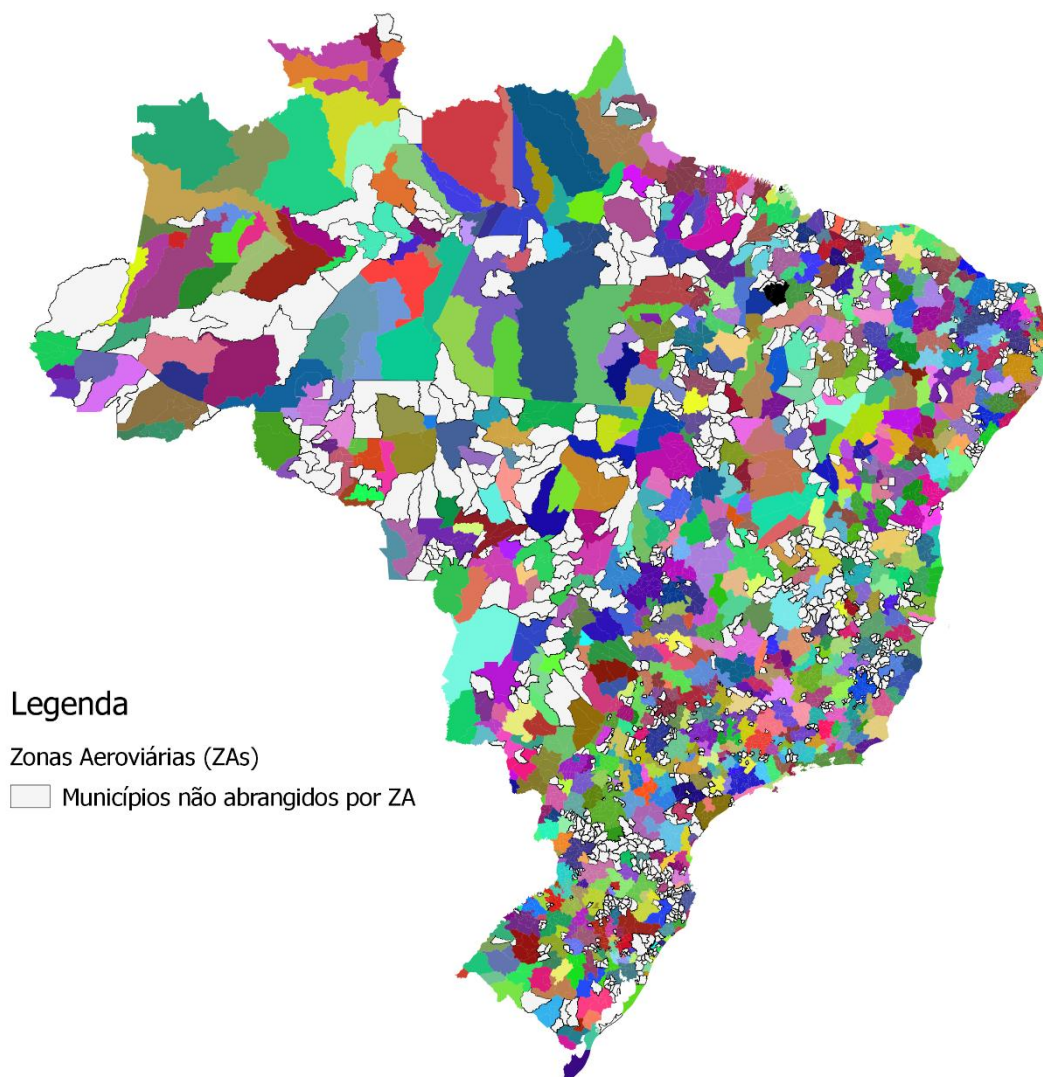
## 2.2 O Plano Aeroviário Nacional

A SAC está elaborando o novo Plano Aeroviário Nacional, que englobará uma série de parâmetros, diretrizes e objetivos para o planejamento do setor mais especificamente voltado ao planejamento de longo prazo voltado às infraestruturas, como as edificações, pistas de pouso, instalações necessárias à operação aérea, serviços dentro e fora da área dos aeroportos e aeródromos, destinados a facilitar e tornar seguro a navegação aérea e o tráfego aéreo. Além disso, o PAN tratará da integração aos demais modais de transporte do país, notadamente o transporte fluvial, marítimo, ferroviário e rodoviário.

Para fins de elaboração do PAN, o Departamento de Planejamento e Estudos da Secretaria de Aeroportos da SAC/MT (DPE/SEAP/SAC/MP) desenvolveu uma classificação dos municípios brasileiros em Zonas Aeroviárias (ZAs). As ZAs foram constituídas a partir da agregação de municípios inicialmente tomando como ponto de partida os Arranjos Populacionais, conforme classificação do Instituto Brasileiro e Geografia e Estatística (IBGE), a partir dos quais foram estipuladas áreas de influência de aeródromos ou em que de vislumbrou a possibilidade de que determinado conjunto de municípios pudesse ser elegível ao recebimento de um aeródromo.

As ZAs agregam 4.516 municípios em 775 aglomerações, constituindo hodiernamente o maior nível de desagregação para efeitos de elaboração do Plano Aeroviário Nacional. Ao tempo da elaboração deste estudo, a composição das zonas não era definitiva, mas permitia muito boa ferramenta de análise e aplicação das bases de dados para efeitos de averiguação dos potenciais das ferramentas estatísticas e computacionais que serão apresentadas, para fins de auxílio à priorização de investimentos. O Mapa 3 apresenta de forma visual a disposição das ZAs e municípios que não são abrangidos por qualquer das zonas ao longo do território nacional em branco.

Mapa 3: Distribuição das Zonas Aeroviárias brasileiras ao longo do território brasileiros



Fonte: Fonte: elaboração própria com dados do DPE/SEAP/SAC/MT e do sistema Hórus, SAC/MT.

Como será apresentado posteriormente, as Zonas Aeroviárias forma utilizadas neste trabalho como elementos de agregação de características municipais, tais como população, produto interno bruto (PIB), gastos municipais, entre outros. Tais informações constituem o cerne da avaliação por parte das máquinas de suporte vetorial e aprendizado de máquina, conceitos que serão abordados no próximo capítulo.

### 3 Aprendizado de máquina e máquinas de suporte vetorial

Segundo Goodfellow et al (2016), aprendizado de máquina é essencialmente uma forma de estatística aplicada com maior ênfase no uso de computação para estimar estatisticamente funções complexas e menor ênfase em determinar os intervalos de confiança em torno dessas funções. Este aprendizado se dá utilizando-se algoritmos, também conhecidos como algoritmos de aprendizado de máquina. Pode-se dividir os algoritmos de aprendizado de máquina em dois grupos: aprendizado supervisionado e aprendizado não supervisionado.

Mitchell (1997) define o aprendizado de máquina como:

“diz-se que um programa computacional aprende com uma experiência  $E$  em relação a determinada classe de tarefas  $T$  e à medida de performance  $P$ , se a performance nas tarefas  $T$ , mensuradas por  $P$ , melhora com a experiência  $E$ ”. (MITCHELL, 1997)[Tradução Livre]

Uma tarefa,  $T$ , é definida como a maneira que um sistema processa um exemplo, constituído por um conjunto de características, normalmente quantitativamente mensuradas, de algum objeto ou evento que se deseja que o sistema de aprendizado de máquina processe. Um exemplo é tipicamente representado como um vetor  $\mathbf{x} \in \mathbb{R}^n$  em que cada elemento  $x_i$  do vetor é uma característica. Neste trabalho os tipos de tarefa abordados são basicamente o de i) classificação, ii) classificação com dados de entrada inexistentes; e iii) regressão.<sup>3</sup> Esses grupos de tarefa são assim definidos segundo Goodfellow et al (2016):

- i) Classificação: é o tipo de tarefa em que um programa é requerido a especificar qual dentre  $k$  categorias um grupo de entradas (*inputs*) pertence. Para resolver essa tarefa, o algoritmo de aprendizado é exigido de maneira a produzir um função  $f : \mathbb{R}^n \rightarrow \{1, \dots, k\}$ . Quando  $y = f(\mathbf{x})$  o

---

<sup>3</sup> Goodfellow et al (2016) lista algumas das tarefas mais comuns: i) classificação; ii) classificação com dados de entrada inexistentes; iii) regressão, iv) transcrição; v) tradução automatizada; vi) saídas estruturadas; vii) detecção de anomalias; viii) síntese e amostragem; ix) preenchimento de valores inexistentes; x) eliminação de ruídos; e xi) estimação de função de densidade ou de massa de probabilidade.

modelo atribui *inputs* descritos por um vetor  $\mathbf{x}$  à categoria identificada pelo código numérico  $y$ .

- ii) Classificação com dados de entrada inexistentes: quando os *inputs* não são completos, ou seja, há dados faltantes, o algoritmo deve aprender um conjunto de funções, ao invés de uma só, em que cada função corresponde à classificação de  $\mathbf{x}$  baseada em diferentes subconjuntos de *inputs* faltantes.
- iii) Regressão: neste tipo de tarefa o programa computacional é levado a prever um valor numérico baseado nos *inputs*. Para tanto o algoritmo de aprendizado deve gerar um resultado (*output*) normalmente do tipo  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ .

A medida de performance,  $P$ , é utilizada para medir o nível de acurácia do modelo. A acurácia é em geral dada pela proporção do número de *outputs* gerados corretamente pelo modelo.

A experiência,  $E$ , define-se como a maneira que um algoritmo utiliza para analisar um conjunto de dados, podendo ser enquadrada em duas categorias, supervisionada e não supervisionada. Um conjunto de dados é uma coleção de diversos exemplos, por sua vez formados por diferentes características ou atributos.

### 3.1 Aprendizado supervisionado e não supervisionado

Aprendizado supervisionado e aprendizado não supervisionado não são termos formalmente definidos. O termo supervisionado é derivado da ideia de que há uma variável, ou vetor de variáveis  $\mathbf{y}$ , utilizada como um instrutor, ou professor, que mostra à máquina o que deve ser feito, qual o resultado que se deseja encontrar, ou ao qual se deseja relacionar as características observadas quando da análise dos exemplos. No aprendizado não supervisionado não há esse professor, e o algoritmo utilizado pela máquina deve tirar conclusões ou sentido dos exemplos sem uma variável guia pré especificada (Goodfellow et al 2016).

Segundo Hastie et al (2009) o aprendizado é tido como supervisionado quando, para o conjunto de exemplos apresentado ao algoritmo de aprendizado, há variáveis previamente denotadas como *inputs* e outras denotadas como *outputs*. Para cada exemplo o objetivo é usar os *inputs* para prever os *outputs*. Os *inputs* são também chamados de variáveis independentes, variáveis previsoras ou regressoras. *Outputs*

são designados variáveis dependentes ou variáveis resposta. (HASTIE e colab., 2009)

Diz-se que o aprendizado é não supervisionado quando o algoritmo experimenta os dados sem que se atribua variável de referência como resultado. Ou seja, o objetivo é inferir as propriedades da distribuição da densidade de probabilidade sem ajuda de uma variável “professora” para cada exemplo. (HASTIE e colab., 2009)

O aprendizado não supervisionado diz respeito à observação de vários exemplos de um vetor aleatório  $\mathbf{x}$ , com a tentativa implícita ou explícita aprender a probabilidade de distribuição  $p(\mathbf{x})$ , ao passo que o aprendizado supervisionado diz respeito à observação de vários exemplos de um vetor aleatório  $\mathbf{x}$  e um valor associado ou vetor  $\mathbf{y}$ , com o objetivo de prever o valor de  $\mathbf{y}$  dado o valor de  $\mathbf{x}$ , normalmente por meio da estimação  $p(\mathbf{y} | \mathbf{x})$ .

Para efeitos do que se vai apresentar posteriormente, tratar-se-á aqui, basicamente, do aprendizado de máquina supervisionado. Há hoje na literatura grande diversidade de modelos e algoritmos de aprendizado supervisionado, Kuhn e Johnson (2013), Bishop (2013), Hastie et ali (2009), Barber (2011), Rasmussen e Williams (2006) e Mitchell (1997) reúnem os principais tipos de aprendizado supervisionado sob perspectivas diferentes. A discussão sobre aprendizado não supervisionado não faz parte do escopo deste trabalho e portanto não será abordada em profundidade (BARBER, 2011; BISHOP, 2013; HASTIE e colab., 2009; MITCHELL, 1997; RASMUSSEN, C.E. E WILLIAMS, 2006).

### 3.2 Capacidade, overfitting e underfitting.

O principal objetivo do aprendizado de máquina é a realização de previsões que são realizadas, por definição, sobre coisas ainda não visualizadas, assumindo-se, eventualmente, características que se pode observar. Destarte, o desafio central do aprendizado de máquina é a capacidade de efetuar previsões precisas em relação a *inputs* não observados previamente. Essa capacidade de efetuar boas previsões sobre *inputs* não observados se denomina generalização (Goodfellow et ali, 2016).

Para realizar o treinamento de um modelo de aprendizado de máquina utiliza-se um conjunto de treinamento. A partir do conjunto de treinamento e do modelo aprendido após aplicado determinado algoritmo, pode-se aferir o erro de

treinamento. Por exemplo, supondo que se queira descobrir qual o próximo valor de uma variável  $v \in \mathbb{I}$  coletada ao longo de  $t$  períodos e assumindo que o algoritmo utilizado é muito simples: a média  $\frac{1}{t} \sum v$ , temos, portanto,  $f(v) : \mathbb{I} \rightarrow \mathbb{R}$ . Neste caso o erro de treinamento será simplesmente o desvio padrão,  $\sigma$ , e é simples observar que, sabendo-se que o resultado pode ser um número não inteiro, e supondo que o seja, o valor previsto será necessariamente errôneo para o período  $t + 1$ . A diferença entre o erro observado *a posteriori* em relação a  $v$  em  $t + 1$ , ou entre  $\frac{1}{t} \sum_{t=t+1}^n v$ , e  $\sigma$  é chamado de erro de teste, ou erro de generalização.

No exemplo ora apresentado, o erro de generalização é decorrente da excessiva simplificação do modelo. Nesses casos em que há erro dessa natureza, ocorre o que se denomina *underfitting*, ou um subajuste. Entretanto, dados os níveis alcançados por algoritmos, somado à alta capacidade computacional, é muito recorrente o problema inverso, denominado *overfitting*, ou sobreajuste, e ocorre quando o modelo aprendido no processo de treinamento é capaz de replicar cada caso do grupo treino com grande precisão, porém quando aplicado em um grupo não observado, apresenta erros de teste elevados.

Como ensina Goodfellow et ali (2016), o fator que determina quão boa é a atuação de um algoritmo de aprendizado é a habilidade de ele:

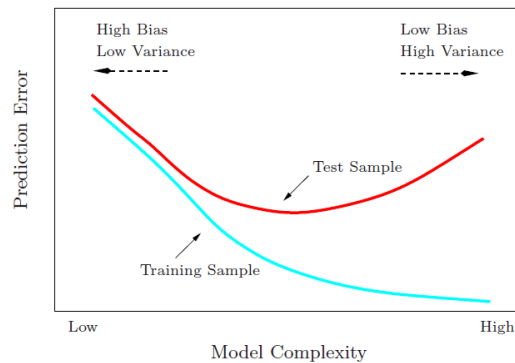
1. Fazer com que o erro de treinamento seja pequeno.
2. fazer com que a diferença entre o erro de treinamento e o de teste seja pequena.

Esses dois fatores correspondem aos dois desafios centrais no aprendizado de máquina: *underfitting* e *overfitting*. *Underfitting* ocorre quando o modelo não consegue obter erro suficientemente pequeno para o grupo de treino. *Overfitting* acontece quando, sendo o erro de treinamento pequeno, a diferença entre este e o erro de teste é grande. [...]Deriva daí o conceito de capacidade do modelo, que é a habilidade dele de se adequar a uma grande variedade de funções. ” Goodfellow et ali (2016, p. 112)[tradução livre]

Hastie et ali (2009, p. 38) abordam a questão do *tradeoff* entre viés e variância, ou *tradeoff* viés-variância, quando da escolha entre modelos, que decorre basicamente do empasse entre maior complexidade e menor capacidade de explicação do modelo, o que está intrinsicamente ligado aos problemas de *overfitting* e *underfitting*, como ilustra a figura abaixo. Quando se aumenta a

complexidade do modelo, ou seja, à medida que o adequamos ao conjunto de treinamento, o erro de treinamento (refletido pela variância) diminui. Entretanto, quando há adequação excessiva ao conjunto de treinamento é possível que não haja boa generalização, havendo assim maior erro de teste (refletido pelo viés). (HASTIE e colab., 2009)

Figura 2: Tradeoff viés-variância.



Fonte: Hastie *et ali*(2009)

Portanto, os maiores desafios quando da aplicação do aprendizado de máquina se encontram normalmente na otimização dos modelos e na ampliação da capacidade de generalização deles, buscando se evitar o *overfitting* e o *underfitting*.

### 3.3 Máquinas de suporte vetorial.

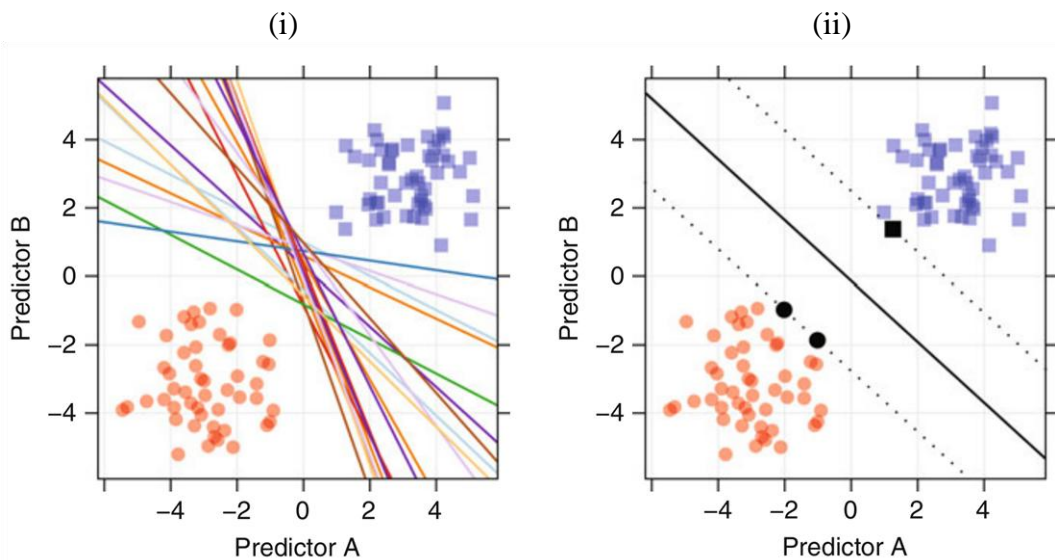
Kuhn (2013) ensina que máquina de suporte vetorial é uma classe de modelos estatísticos desenvolvidas inicialmente por Vladimir Vapnik na década de 1960 e que vem ganhando atenção gradativa nos últimos anos, e tem se mostrado um dos instrumentos mais versáteis e efetivos de aprendizado. Apreende-se de Hastie (2009) que as máquinas de suporte vetorial constituem técnicas de criação de hiperplanos que separam duas classes, ou mais, construindo uma linha de fronteira em uma versão transformada do espaço de características.

Considerando o exemplo de um problema em que são utilizadas duas variáveis para prever duas classes de pontos completamente separáveis em um espaço bidimensional, como ilustrado na Figura 3 abaixo, em que há duas variáveis predictoras, *Predictor A* e *Predictor B*, e duas classes, azul (Z) e vermelha (V). É simples perceber que há infinitas possibilidades de fronteiras capazes de separar perfeitamente as duas classes, como se ilustra em 3(i). Vapnik definiu uma medida

denominada margem, que auxilia na escolha das fronteiras elegíveis para separação das classes. A margem é a distância entre a fronteira de classificação e o ponto mais próximo de um conjunto de treinamento (VAPNIK *apud* KUHN e JOHNSON, 2013, p. 343).

Como se observa na figura 3(ii) as linhas pontilhadas, de ambos os lados da fronteira sólida que separa as duas classes, estão à distância máxima entre a fronteira e o ponto mais próximo que representa cada classe. Os três pontos destacados na cor preta, que se encontram sobre as margens, estão equidistantes da fronteira. A margem definida por estes pontos pode ser utilizada para elaborar um modelo de classificação (HASTIE, 2009, p. 418; e KUHN e JOHNSON, 2013, p. 344).

Figura 3: representação de dados perfeitamente separáveis em duas classes.



Fonte: Kuhn e Johnson (2013, p. 344)

Em termos mais específicos, a inclinação e intercepto da fronteira que maximizam o “amortecimento” entre essa fronteira e os dados é conhecido como margem máxima de classificação. Formalmente pode-se atribuir valores às classes, suponha-se  $Z = -1$  para uma classe azul e  $V = 1$  para a classe vermelha. Assumindo, ainda, que  $\mathbf{x}_i$  é a matriz formada pelos vetores com os valores das variáveis de previsão. A margem máxima de classificação determina uma função de decisão  $D(\mathbf{x})$  para classificação de amostras de forma que quando  $D(\mathbf{x}) < 0$  as amostras são classificadas como -1, azul, e quando  $D(\mathbf{x}) > 0$  as amostras são classificadas como 1, vermelho. Supondo agora uma amostra desconhecida  $\mathbf{u}$ , a



função de decisão pode ser escrita em termos de um intercepto e das inclinações referentes a cada variável de previsão como:

$$D(\mathbf{u}) = \beta_0 + \boldsymbol{\beta}^T \mathbf{u} \quad (4.1)$$

Uma vez que a função acima é escrita com foco nas variáveis de previsão, podemos transformá-la de maneira que seja escrita em termo de cada ponto da amostra, na seguinte forma:

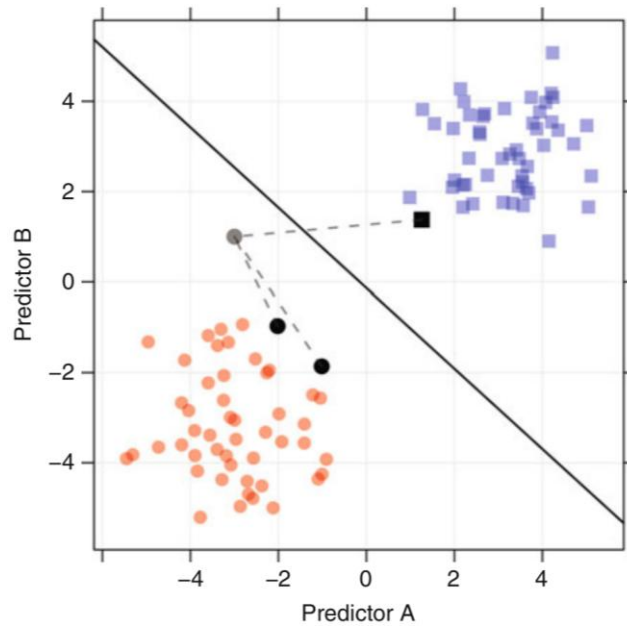
$$D(\mathbf{u}) = \beta_0 + \sum_{j=1}^P \beta_j u_j$$

$$D(\mathbf{u}) = \beta_0 + \sum_{i=1}^n y_i \alpha_i \mathbf{x}_i^T \mathbf{u} \quad (4.2)$$

Sendo  $y_i$  a classe do ponto  $i$ , e  $\alpha \geq 0$  um parâmetro calculado a partir do grupo de treinamento para cada ponto que compõe a margem;  $\alpha = 0$  para os pontos fora da margem, e  $\alpha > 0$  para os pontos sobre os limites da margem. Portanto, a equação de previsão não é, ao final, formada por todos os valores observados no conjunto treino, mas apenas os valores que estão sobre os limites das margens, ou seja, os pontos destacados em preto na figura X(ii). Como a equação é “suportada” apenas por estes pontos, ou vetores, a margem máxima de classificação é também conhecida como máquina de suporte vetorial (KUHNS e JOHNSON, 2013, p. 345).

Para exemplificar como é feita a classificação de novas amostras, suponhamos o acréscimo de um novo ponto cuja classe é desconhecida, representado na cor cinza na Figura 4 abaixo. A distância entre cada vetor de suporte e os novos pontos é representada pelas linhas intermitentes em cinza.

Figura 4: previsão de uma nova amostra utilizando uma máquina de suporte vetorial.



Fonte: Kuhn e Johnson (2013, p. 346)

Temos aqui, para cada um dos três vetores, portanto as seguintes informações

Vetor de Suporte (VS)	Classe	$y_i$	$\alpha_i$	$x_i^T u$	$y_i \alpha_i x_i^T u$
VS 1	Azul	-1	1	-2,4	2,4
VS 2	Vermelho	1	0,34	5,1	1,734
VS 3	Vermelho	1	0,66	1,2	0,792

tabuladas:

Fonte: adaptado de Kuhn e Johnson (2013, p. 345)

Note-se que  $x_i^T u$  é determinado pela distância entre  $x_i$  e a origem, a distância entre  $u$  e a origem, e o cosseno do ângulo entre  $x_i$  e  $u$ . (DILLON e GOLDSTEIN *apud* KUHN e JOHNSON, 2013). O intercepto,  $\beta_0$ , é -4,372, assim, o valor final da equação de decisão, conforme dado por (4.2) é  $-4,372 + 4,926 = 0,553$ . Sendo assim,  $D(u) > 0$ , portanto, classifica-se o novo ponto com 1, vermelho.

### 3.4 Classes misturadas e o artifício dos kernels.

O modelo apresentado na seção anterior é bastante simples, uma vez as classes são claramente separadas e de fácil delimitação. Entretanto, é muito comum que não seja possível separar as classes com tanta precisão, com utilização de um plano tão simples como uma reta. Nesses casos em que as classes se misturam, gradativamente, ou a partir de padrões mais complexos, são realizadas alterações no

modelo. Cortes e Vapnik (1995, p. 290) propõem um modelo em que se utilizam redes de vetores de suporte a partir de transformações polinomiais dos *inputs* que ajudam a enfrentar esse problema. Segundo Kuhn e Johnson (2009, p. 346) formulação deles traz o conceito de custo (C) para a soma das características dos pontos que constituem os limites das margem ou do lado oposto em relação ao que é esperado para a respectiva classe. Quando da estimação dos parâmetros  $\alpha$ , a margem é penalizada se houverem pontos dentro da margem ou pontos do lado não esperado.

Uma das formas primeiramente visualizadas para ampliar a capacidade de predição dos modelos, especialmente pela extensão dos modelos lineares para modelos não lineares foi proposta por Boser *et ali* (1992, p. 144). Estes autores propuseram um algoritmo que ajusta automaticamente a função de classificação e maximiza as margens entre diferentes classes em pontos distintos, portanto, com ampliação da complexidade dos modelos, e apresentando uma solução expressada como uma combinação linear (BOSER e colab., 1992).

A principal alteração em relação ao que foi abordado no modelo anterior é a utilização de uma normalização, ou alteração do núcleo do modelo, denominada *kernel* (no termo em inglês que será utilizado doravante), da equação apresentada 4.2, da seguinte forma.

$$D(\mathbf{u}) = \beta_0 + \sum_{i=1}^n y_i \alpha_i \mathbf{x}_i^T \mathbf{u}$$

$$D(\mathbf{u}) = \beta_0 + \sum_{i=1}^n y_i \alpha_i K(\mathbf{x}_i, \mathbf{u}) \quad (4.3)$$

Em que  $K(\mathbf{x}_i, \mathbf{u})$  deixa de ser limitado ao simples produto vetorial  $\mathbf{x}_i^T \mathbf{u}$ , que representa a função linear dos vetores, e passa a ser uma função com diferentes possibilidades: formas polinomiais, funções de base radial, tangentes hiperbólicas entre outras. Algumas formas de *kernel* são as seguintes (Kuhn, 2013, p. 347):

$$\text{polinomial} = (\text{escalar}(\mathbf{x}^T \mathbf{u}) + 1)^{\text{grau}}$$

$$\text{função de base radial} = \exp(-\sigma \|\mathbf{x} - \mathbf{u}\|^2)$$

$$\text{tangente hiperbólica} = \tanh(\text{escalar}(\mathbf{x}^T \mathbf{u}) + 1)$$

Outro *kernel* que se mostrou relevante, especialmente após o início da aplicação de diferentes formas de classificação por máquina de suporte vetorial, foi o classificador por mínimos quadrados (ou lssvm, sigla do inglês *least square support vector machine classifier*), apresentado primeiramente por Suykens e Vandewalle (1999). Como resultado da aplicação deste *kernel* a solução é derivada diretamente da resolução de um conjunto de equações lineares. Enquanto em outros modelos clássicos de SVM muitos dos valores de vetores de suporte são iguais a zero, na abordagem lssvm os valores dos vetores de suporte são proporcionais aos erros (SUYKENS e VANDEWALLE, 1999, p. 294).

Os *kernels* passaram a exercer, portanto, função importante em diversas áreas de aprendizado de máquina, tanto em regressões como em classificações, sendo utilizado o termo artifício de *kernel* (*Kernel Trick*) para se referir à sua utilização. Em resumo, a utilização de *kernels* permitem aumentar a flexibilidade de procedimentos, ampliando-se o espaço de características ou atributos com utilização de polinômios ou splines.<sup>4</sup> Geralmente fronteiras lineares em um espaço aumentado permitem melhor separação de classes em treinos e se traduzem em fronteiras não lineares no espaço original (HASTIE e colab., 2009, p. 423).

Barber (2011, p.324) exemplifica de forma bastante ilustrativa: a utilização de *kernels* é “uma forma de expandir o método para uma fronteira de classificação não linear e considerar os inputs  $\mathbf{x}$  de uma maneira não linear  $\phi(\mathbf{x})$ . [...] Por exemplo, um *input* unidimensional  $x$  poderia ser mapeado como um vetor bidimensional  $(x^2, \text{sen}(x))$ ”. Mapear pontos em dimensões mais elevadas pode tornar mais fácil a definição de um hiperplano que separe as classes de forma mais precisa.

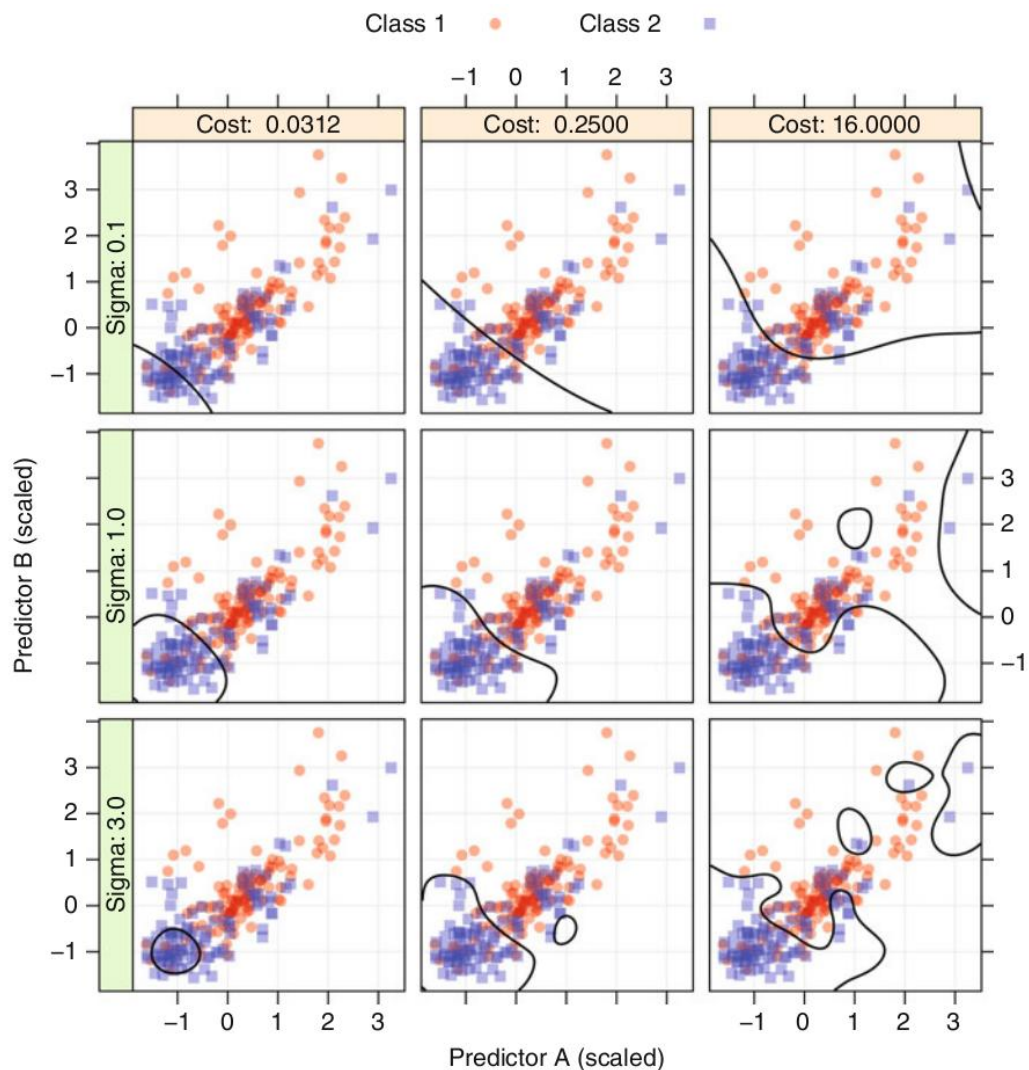
Kuhn *et ali* (2013, p. 347) salienta que a escolha dos parâmetros da função *kernel* e o valor de custo (C) atribuído controlam a complexidade e são os valores que devem ser ajustados de maneira a evitar o *overfitting* do grupo de treinamento. Normalmente quando C é baixo o modelo é *underfited*, subajustado, ao passo que quando o valor de C é elevado pode ocorrer *overfitting*, sobreajuste, a depender do

---

<sup>4</sup> Spline é uma curva definida por dois ou mais pontos de controle. Pontos de controle que pertencem à curva são denominados nós. E os que não pertencem à curva mas auxiliam o delineamento dela definem pontos de tangência.

valor dos parâmetros do *kernel*. Para suplantar esse problema costuma-se utilizar técnicas de reamostragem de maneira a se encontrar valores equilibrados de C e dos parâmetros da função *kernel* K. A figura 4 abaixo apresenta diferentes resultados de fronteiras de classificação baseadas em um *kernel* de função de base radial com diferentes ajustes dos parâmetros C e  $\sigma$ .

Figura 4: Impacto de diferentes ajustes dos parâmetros C e  $\sigma$  em um modelo que utiliza *kernel* de função de base radial, nas respectivas fronteiras de classificação.



Fonte: Kuhn e Johnson (2013, p. 348)

Nota-se claramente que à medida que o valor de  $\sigma$  (Sigma) aumenta, mais específico é o grupo abrangido pela fronteira, e à medida que C (Cost) aumenta, mais sinuosa se torna a fronteira. Nos dois primeiros gráficos da linha superior, observam-se basicamente linhas com curvatura leve (o que pode indicar

*underfitting*), enquanto nos últimos gráficos da linha inferior e intermediária, observam-se curvas bastante sinuosas e que chegam a classificar pontos isolados (o que pode indicar *overfitting*). É de se frisar, ainda, que, em decorrência da ampliação das dimensões derivadas da expansão levada a cabo pelo *kernel*, são apresentadas inclusive fronteiras fechadas no espaço bidimensional de alguns dos gráficos apresentados.

### 3.5 Métodos de reamostragem.

Há maneiras distintas de fazer reamostragem, Kuhn e Johnson (2013) resumem as principais técnicas em quatro grupos, e explicam que elas operam de forma similar. As técnicas abordadas por estes autores são:

1) Validação Cruzada com  $k$ -Dobras (*k-Fold Cross Validation*): a amostra é dividida em  $k$  conjuntos fixos de mesmo tamanho, sendo um primeiro conjunto deixado de fora, e o modelo é rodado a partir dos conjuntos remanescentes e utilizado para estimar o conjunto externo para estimar a performance do modelo. Posteriormente utiliza-se o primeiro conjunto retirado é incluído no grupo utilizado para estimação de um modelo, sendo outro deixado de fora da amostra e utilizado para verificação da performance, e assim sucessivamente. As  $k$  estimativas com as reamostragens são sumarizadas de forma a gerar o modelo final. No extremo, pode-se fazer  $k$  igual ao número de exemplos da amostra, neste caso, o método recebe o nome de validação cruzada com um de fora (em inglês, *leave-one-out cross-validation* LOOCV).

2) Validação Cruzada Generalizada: este método utiliza uma estatística de validação cruzada a partir de uma fórmula que retira a necessidade de reajustes com diferentes iterações de subconjuntos, baseando-se apenas nos erros e nos graus de liberdade dos modelos estimados para cada subconjunto.

3) Repetidas Divisões Treino/Teste: este método é muito semelhante ao *k-Fold*, entretanto, os subconjuntos têm os elementos que ficam de fora do subgrupo treino sorteados aleatoriamente e recolocados no conjunto inicial a cada rodada.

4) *Bootstrap*: nesta metodologia os elementos do subconjunto de referência, a ser previsto, são retirados aleatoriamente a cada rodada do subconjunto de treino e são substituídos no subconjunto de treino por outros elementos deste conjunto que

têm seus valores repetidos, de maneira que o tamanho do subconjunto de treino permanece com o mesmo tamanho do conjunto original. Posteriormente os elementos do subconjunto a ser previsto são repostos no conjunto inicial e uma nova subamostra é retirada para compor um subconjunto de referência e o processamento é repetido, e assim sucessivamente pelo número de rodadas estipulado.

Cada um dos métodos tem diferentes variações e por vezes são utilizados concomitantemente, podendo variar diversos aspectos, como o tamanho de cada conjunto a cada rodada, utilização de diferentes modelos em rodadas distintas, entre outros.

## 4 Aplicação da metodologia

### 4.1 Constituição da base de dados

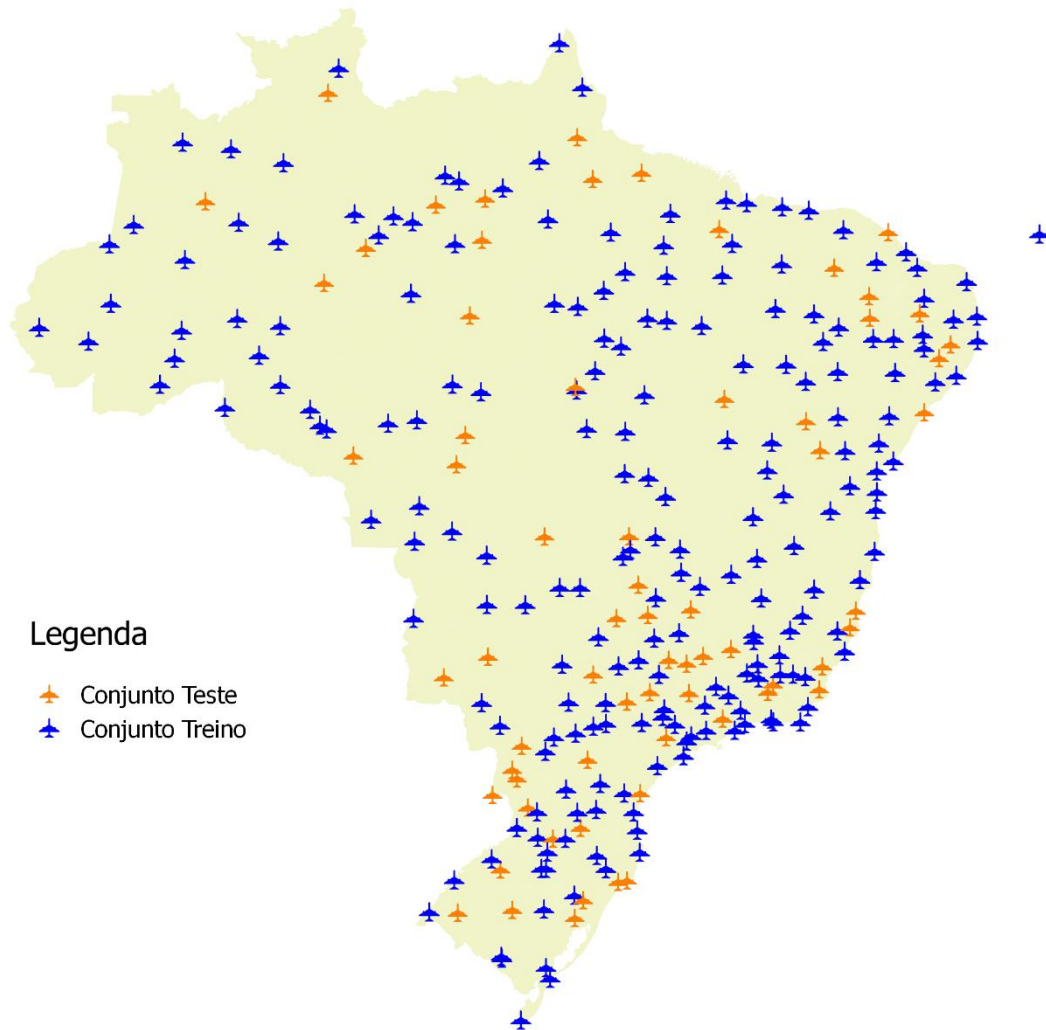
A base de dados utilizada contém dados oriundos de diferentes fontes e compreende características distintas. As variáveis são obtidas em sua maioria para o nível municipal e agrupadas para o nível de agregação de Zonas Aeroviárias. Nesta seção serão detalhadas as variáveis, sua agregação, as fontes e a forma de construção de algumas delas.

As zonas aeroviárias (ZAs), como descritas no capítulo 2, foram formuladas a partir de agregação feita no âmbito do Departamento de Planejamento e Estudos da Secretaria de Aviação Civil do Ministério dos Transportes, Portos e Aviação Civil, e que tem como ponto de partida os arranjos populacionais do IBGE (2015). Para efeitos de estimação e teste dos modelos, foram utilizadas na base inicial 266 ZAs, sendo que 200 delas compuseram o conjunto de treinamento e 66 o conjunto teste. Este primeiro grupo (de 266) é formado pelas ZAs que agregavam ao menos uma das três características a seguir: (a) contém um aeroporto sistêmico ou (b) contém ao menos um aeroporto que apresentou movimentação de passageiros no período de 2004 a 2015, (c) ZAs que integram o grupo das 270 localidades abrangidas pelo PIL Aeroportos e contém ao menos um aeródromo público homologado. Abaixo o Mapa 4 traz a divisão dos conjuntos de treinamento e de teste com visualização facilitada.

No anexo II estão listadas as ZAs constantes da base de dados com a classificação nos conjuntos treino, teste e novo (este referente à segunda base de aplicação dos dados com 509 ZAs), indicação de inserção entre as 270 localidades selecionadas no âmbito do PIL, indicação da condição de aeroporto sistêmico, e o município e unidade da federação (UF) em que se encontra a sede da ZA.



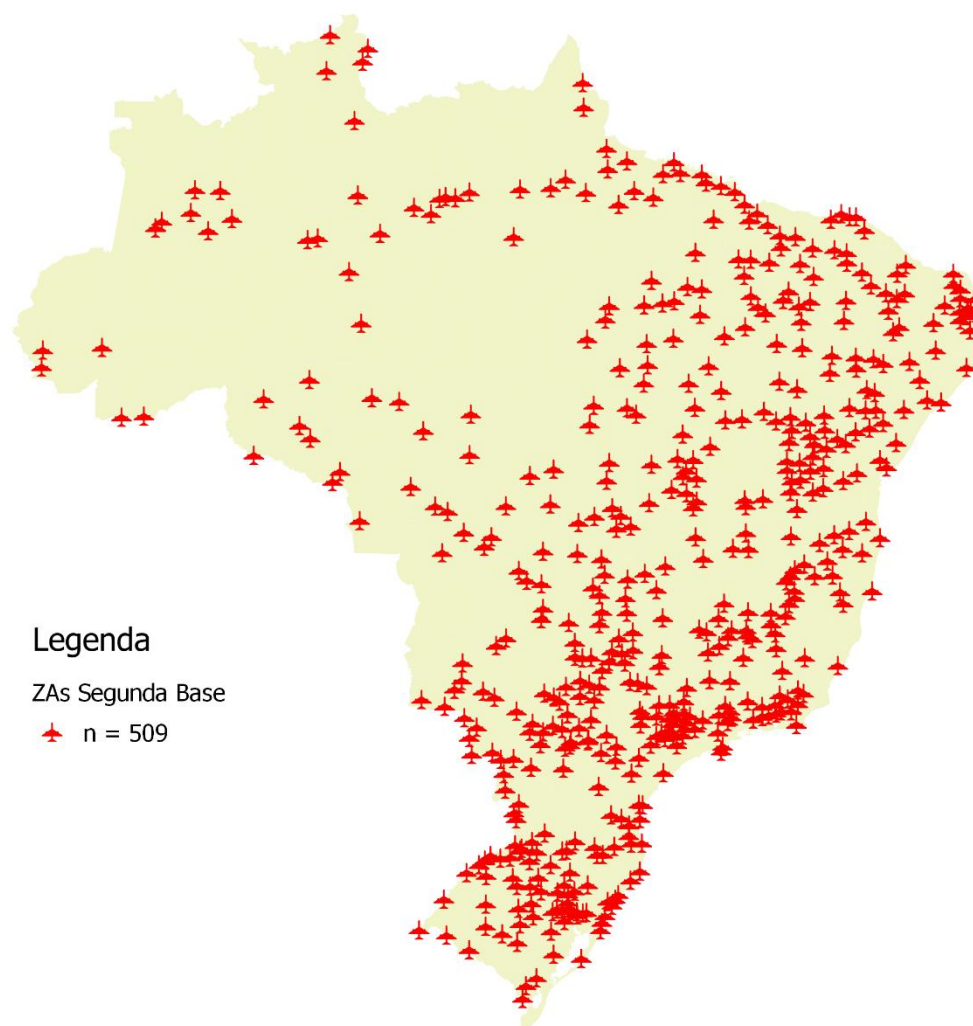
Mapa 4: Distribuição dos aeroportos em conjunto de treinamento e conjunto de teste, submetidos à classificação por SVM.



Fonte: elaboração própria.

Posteriormente foi utilizada uma segunda base, composta por 509 ZAs, para efeitos de aplicação dos modelos e visualização dos resultados em um conjunto totalmente distinto do inicial. Esta segunda base é constituída por 22 localidades do PIL que não possuem aeródromo público homologado e outras 487 localidades para as quais foi possível agregar os dados que constituem a base e serão descritos doravante.

Mapa 5: ZAs que compõem a segunda base de dados, submetida a posteriori à classificação por SVM para fins de verificação da aplicabilidade dos modelos.



Fonte: elaboração própria.

Com intuito de facilitar a apresentação das variáveis, elas foram segregadas em quatro grupos: a) variáveis de características de aeroportos, b) variáveis socioeconômicas e regionais, c) indicadores de gastos municipais e d) indicadores de turismo. As fontes de informações utilizadas são todas de órgãos públicos do Governo Federal, abrangendo por vezes períodos distintos.

a) Variáveis relativas aos aeroportos. Este grupo abarca três variáveis, relacionadas à localização dos aeroportos e sua localização geográfica.

- 1) Variável dependente (tresanosatr) que indica se uma ZA recebeu ou não voo regular. É a variável dependente utilizada na estimação dos modelos, ou seja, é a variável utilizada como referência para aplicação do aprendizado supervisionado. Ela é construída a partir da média semanal de movimentação de passageiros nos aeroportos das zonas aeroviárias. Assume valor V para os aeroportos cuja movimentação média anual foi superior a 4056, por no mínimo 3 anos não necessariamente consecutivos, no período de 2004 a 2015. Esse valor equivale a duas movimentações semanais, uma chegada e uma partida, de uma aeronave ATR-42 com aproximadamente 80% de taxa de ocupação. Essa é a menor aeronave de referência para investimentos anunciada no âmbito do PIL. A variável recebe valor F para qualquer valor inferior a esse patamar. Os dados são oriundos da ANAC, da Lista de Aeródromos Públicos, disponível em <http://www.anac.gov.br/Anac/assuntos/setor-regulado/aerodromos> e foram obtidos em 20 de julho de 2016 para composição da base final utilizada neste trabalho. Os dados referentes à ZA de Confresa foram obtidos da Lista de Aeródromos Privados disponível no mesmo sítio da internet, e representa o único aeródromo privado que entrou na base como tendo recebido voos regulares.
  - 2) Uma variável que representa a distância linear entre o ponto que determina o posicionamento geográfico do principal aeródromo da ZA e o aeródromo sistêmico mais próximo (distsistêmico). Para a segunda base (509 ZAs), o ponto de referência é o ponto médio entre aqueles que indicam as sedes municipais.
  - 3) A variável regional (regiao) indica a grande região brasileira na qual se encontra a ZA, Norte, Nordeste, Sudeste, Sul ou Centro-Oeste.
- b) Variáveis socioeconômicas, o grupo é composto por dois conjuntos de
- 1) População total da ZA para os anos de 2004 a 2015 (popza2004-2015). A população é dada pela agregação das populações dos municípios que compõem a respectiva ZA. Para o ano de 2007 os dados são da Contagem Populacional 2007, disponível em:

<http://ibge.gov.br/home/estatistica/populacao/contagem2007/default.shtm> ; para o ano de 2010 os dados são do Censo Demográfico do mesmo ano, disponível em:

<http://ibge.gov.br/home/estatistica/populacao/censo2010/default.shtm> ; para os demais anos o valor é da estimativa populacional, publicada anualmente pelo IBGE e disponível em:

<http://ibge.gov.br/home/estatistica/populacao/estimativa2015/default.shtm> .

- 2) Produto interno bruto (PIB) real per capita da ZA (pibpcza2004-2015). Agregação dos PIB municipais da ZA dividida pela população da ZA, para os anos de 2004 a 2013. Ressalte-se que para os anos de 2010 a 2013, foram utilizados os dados de PIB de acordo com nova metodologia. Os valores foram deflacionados utilizando-se o IPCA calculado pelo IBGE para o período. Os PIB municipais são disponibilizados pelo IBGE para livre acesso em:

[http://ibge.gov.br/home/estatistica/economia/pibmunicipios/2010\\_2013/default.shtm](http://ibge.gov.br/home/estatistica/economia/pibmunicipios/2010_2013/default.shtm) .

- c) Indicadores de gastos municipais: são três os indicadores de gastos dos governos municipais, incluídos para trazer uma indicação da disposição do governo municipal investir, especialmente no setor de transporte e de transporte aéreo. Não entram nessa conta gastos diretos dos Estados e da União nos municípios, mas apenas as rubricas municipais em sentido estrito. Os dados são anuais de 2004 a 2015 e estão disponíveis nas base do Finanças do Brasil – Dados Contábeis dos Municípios, em:

[http://www.tesouro.fazenda.gov.br/pt\\_PT/contas-aneais](http://www.tesouro.fazenda.gov.br/pt_PT/contas-aneais). Esses dados não constituem bases completas para todos os municípios, porém avaliou-se que sua utilização contribuiu para os resultados.

- 1) Despesas orçamentárias total dos municípios da ZA (desptot2004-2015pc).
- 2) Despesas orçamentárias dos municípios da ZA com investimentos em transporte (desptransp2004-2015pc). Para se ter uma ideia de ordem de

grandeza, o valor agregado desta rubrica no ano de 2006 para todos os municípios é de cerca de 3% do valor agregado do indicador anterior.

- 3) Despesas orçamentárias dos municípios da ZA com investimento em transporte aéreo (despaero2004-2015pc). Aqui também cabe ressaltar que essa rubrica agregada para todos os municípios equivale a 0,01% de todas as despesas orçamentárias dos municípios.
- d) Indicadores de turismo: os indicadores de turismo são produzidos pelo ministério do turismo e entram no estudo para auxiliar na captação da importância do turismo como elemento indutor de viagens e utilização do modal aéreo. Os valores são relativos ao ano de 2015. As variáveis são:
  - 1) Quantidade de empregos gerados pelo setor de turismo nos municípios que compõem a ZA dividida pela população da ZA (qt\_emprego\_pc).
  - 2) Quantidade de estabelecimentos comerciais nos municípios da ZA ligados ao setor de turismo dividida pela população dos municípios da ZA (qt\_estabelecimentopc).
  - 3) Quantidade estimada de visitantes de origem brasileira estimados que viajaram para municípios da ZA dividida pela população dos municípios da ZA (qt\_visita\_nacional\_estimadapc)
  - 4) Quantidade estimada de visitantes de origem internacional que viajaram para municípios da ZA dividida pela população dos municípios da ZA (qt\_visita\_internacional\_estimadapc).
  - 5) Cinco indicadores que apresentam número de municípios da ZA que receberam a respectiva classificação do Ministério do Turismo de A (município de grande valor turístico) a E (município de menor valor turístico), havendo as classificações intermediárias B, C e D (mturA a mturE).

Como será apresentado na seção seguinte, a base de dados cujas variáveis foram descritas acima foi utilizada na formulação e avaliação dos modelos de Máquina de suporte vetorial, com aplicação de um pacote específico disponível para a plataforma estatística R.

## 4.2 O pacote *caret* desenvolvido para R.

R, ou projeto R, é um software livre com grande capacidade de processamento e tem como principal vantagem o grande número de pacotes programados por usuários que compõem uma comunidade formada por alunos, professores e uma grande diversidade de profissionais tanto da iniciativa privado como de órgãos públicos em todo o mundo. Para o desenvolvimento deste trabalho foi utilizado especialmente o pacote *caret* (nome derivado do termo em inglês *classification and regression training*). O pacote é de autoria de Max Kuhn, com contribuições de outros autores, e reúne grande quantidade de funções, além de se utilizar de outros pacotes estatísticos disponíveis para R.

Kuhn (2008) afirma que modelos de regressão e classificação são cada vez mais utilizados em diversas áreas da ciência e que há uma miríade de modelos e funções disponíveis na linguagem R, é cada vez mais difícil acompanhar e entender as nuances sintáticas de cada função. O pacote *caret* foi criado com os seguintes objetivos: “i) eliminar as diferenças sintáticas entre muitas funções para construção previsão de modelos; ii) desenvolver um conjunto semiautomatizado e de abordagem compreensível para otimizar os valores de parâmetros ajustáveis para esses modelos; e iii) criar um pacote que seja facilmente extensível para sistemas de processamento paralelo”<sup>5</sup>(KUHN, 2008).

O *caret* apresenta, portanto, uma série de funcionalidades que permitem aplicar e controlar os principais parâmetros abordados no capítulo anterior, como a divisão de grupos de treinamento e de teste, seleção entre diferentes modelos de regressão, métodos de amostragem e reamostragem, auxílio a diagnóstico de *overfitting*, entre outros. Destaque-se que as funcionalidades do *caret* vão além do SVM, há grande diversidade de modelos consolidados e cuja forma de utilização foi unificada para trabalho por dentro do pacote. Ao tempo da execução deste trabalho, contabilizavam-se 217 modelos distintos, de mais de cerca de cem pacotes acionados pelo *caret*, utilizáveis para modelos de aprendizado em regressão ou classificação. Além de diversas formas de SVM, encontram-se modelos de

---

<sup>5</sup> O autor do pacote *caret* mantém uma página na internet voltada à divulgação da ferramenta, com diversos tutoriais e exemplos sobre as funcionalidades. O endereço para acesso é <http://topepo.github.io/caret/index.html>.

classificação bayesiana ingênua, florestas randômicas, identificação neuro-difusa, entre outros grandes grupos.

A função `createDataPartition` é utilizada para criar subamostras aleatórias de uma base de dados. A função cria estratificações aleatórias dentro de cada classe, de forma que a distribuição das classes seja preservada o quanto possível (KUHN, 2008). A separação da base inicial, descrita em seção anterior, com 266 observações, foi realizada utilizando-se essa função, em uma proporção de  $\frac{3}{4}$  para o conjunto de treinamento e  $\frac{1}{4}$  para o conjunto teste. O conjunto treinamento ficou, assim, com 200 observações e o conjunto teste com 66 observações.

Como resultado da aplicação do programa à base, foram obtidos diversos resultados, dentre os quais se selecionaram os apresentados neste capítulo. Foram escolhidos cinco modelos, todos com níveis semelhantes de performance após a aplicação sobre o grupo teste. Todos os modelos foram aplicados, em versão final, utilizando-se 200 amostras do conjunto de treinamento, com uma variável dependente, no caso a variável que representa os voos regulares, e 69 variáveis explicativas.

Para aplicação às bases, optou-se especificamente pelos modelos SVM contidos no `kernlab`, principal pacote, dentre os consolidados no âmbito do `caret`, acionado para cálculos de modelos SVM. O `kernlab` objetiva prover aos usuários do R algumas das principais funcionalidades de *kernels*, com possibilidade de escolha entre diferentes tipos de *kernel* e flexibilidade para alteração de parâmetros (KARATZOGLOU e colab., 2004).

Foram utilizados modelos SVM de classificação com distintos *kernels*, sendo que a variável dependente era originalmente classificada em duas classes, conforme definição já apresentada: F, atribuída às ZAs que não receberam voos regulares em ao menos três anos distintos; e V, atribuída às ZAs que receberam voos regulares em ao menos três anos distintos. Todos os modelos foram calculados utilizando-se o método de reamostragem de *bootstraps*, com 25 rodadas, e demais parâmetros em programação predefinidos pelo pacote `caret`. O Anexo I contém a programação utilizada para rodar os modelos.

Os modelos resultantes são:

1 – Modelo de SVM com *kernel* de função de base radial:

Resultados após reamostragem e cálculos com diferentes valores para os parâmetros:

sigma	C	Acurácia	Kappa
0.002866012	0.25	0.6731143	0.02612219
0.002866012	0.50	0.6830380	0.09892993
0.002866012	1.00	0.6860792	0.13916315
0.021429754	0.25	0.6944682	0.11363561
0.021429754	0.50	0.7123593	0.21831568
0.021429754	1.00	0.7085999	0.24309420
0.039993495	0.25	0.6874445	0.07849840
0.039993495	0.50	0.7086425	0.20198486
0.039993495	1.00	0.7209271	0.27680859

A acurácia foi utilizada para seleção do modelo final, com  $\sigma = 0.03999349$  e  $C = 1$ .

2 – Modelo SVM com *kernel* de processo gaussiano com função de base radial:

Resultados após reamostragem e cálculos:

Acurácia	Kappa
0.6990432	0.2321474

Tuning parameter 'sigma' was held constant at a value of 0.02142975

3 – Modelo SVM com *kernel* de mínimos quadrados de função de base radial:

Resultados após reamostragem e cálculos com diferentes valores para os parâmetros:

sigma	tau	Acurácia	Kappa
0.002866012	0.0625	0.7166780	0.3026656
0.002866012	0.1250	0.7102418	0.2736274
0.002866012	0.2500	0.7119978	0.2562608
0.021429754	0.0625	0.7070532	0.2816812
0.021429754	0.1250	0.7080745	0.2809005
0.021429754	0.2500	0.7108276	0.2753349
0.039993495	0.0625	0.7258829	0.2721930
0.039993495	0.1250	0.7222001	0.2592655
0.039993495	0.2500	0.7264999	0.2669105

A acurácia foi utilizada para seleção do modelo final, com  $\sigma = 0,03999349$  e  $\tau = 0,25$ .

4 – Modelo SVM de *kernel* polinomial

Support Vector Machines with Polynomial Kernel

Resultados após reamostragem e cálculos com diferentes valores para os parâmetros:

degree	scale	C	Accuracy	Kappa
1	0.001	0.25	0.6805294	0.02125845
1	0.001	0.50	0.6877361	0.04797984
1	0.001	1.00	0.6833979	0.05311271
1	0.010	0.25	0.6855667	0.10535444
1	0.010	0.50	0.6853964	0.13669044
1	0.010	1.00	0.6897483	0.18782515



1	0.100	0.25	0.6838059	0.21362111
1	0.100	0.50	0.6860384	0.25777616
1	0.100	1.00	0.6931292	0.30061894
2	0.001	0.25	0.6877361	0.04797984
2	0.001	0.50	0.6845780	0.05402857
2	0.001	1.00	0.6820440	0.08434402
2	0.010	0.25	0.6862284	0.14370704
2	0.010	0.50	0.6868388	0.18037877
2	0.010	1.00	0.6830294	0.20851369
2	0.100	0.25	0.7317392	0.37747719
2	0.100	0.50	0.7402480	0.40450909
2	0.100	1.00	0.7393654	0.40972462
3	0.001	0.25	0.6827833	0.04282273
3	0.001	0.50	0.6773253	0.05493460
3	0.001	1.00	0.6842107	0.11015534
3	0.010	0.25	0.6901572	0.17897885
3	0.010	0.50	0.6951023	0.23220516
3	0.010	1.00	0.7091914	0.29291842
3	0.100	0.25	0.7389544	0.40534237
3	0.100	0.50	0.7410489	0.41005695
3	0.100	1.00	0.7410489	0.41005695

A acurácia foi utilizada para seleção do modelo final. Os valores do modelo final são grau = 3, scale = 0,1 e C = 0,5.

#### 5 – Modelo SVM de *kernel* linear

Resultados após reamostragem e cálculos:

Acurácia	Kappa
0.678011	0.295481

C foi mantido constante em 1.

### 4.3 Avaliação e escolha de modelos.

Um dos principais elementos citados na literatura quando se disserta sobre a necessidade de escolha entre modelos é o da Navalha de Occam, ou princípio da parcimônia, atribuído a William of Ockham, frei franciscano britânico nascido no século XIII que dá nome ao princípio. Ele estabelece que quando há concorrência entre hipóteses equivalentes, deve-se optar pela mais simples, com menor número de hipóteses. Para fins de utilização deste princípio no âmbito do aprendizado de máquina, consideram-se questões como aplicabilidade do modelo a diferentes bases de dados, capacidade computacional e tempo de processamento de bases de dados por diferentes modelos que produzem resultados semelhantes. Portanto, a Navalha de Occam deve ser sempre um instrumento da razoabilidade quando há necessidade de se optar pela utilização de um ou outro modelo (BARBER, 2011; CORTES e

VAPNIK, 1995; GOODFELLOW, 2016; HASTIE, 2009; KHUN E JOHNSON, 2013; MITCHELL, 1997; RASMUSSEN E WILLIAMS, 2006).

Há determinados estatísticas que auxiliam na escolha do melhor dentre um grupo modelos estimados quando da aplicação desses em um conjunto teste. A tabela 1 abaixo resume esses valores calculados para cada um dos modelos apresentado. São apresentados, ainda, alguns valores de referência; i) intervalo de confiança de 95% para o nível de acurácia; ii) p-valor para o nível de acurácia; e iii) estatística Kappa, uma relação entre a chance de a classe ocorrer na amostra e a acurácia esperada, valores entre 0,3 e 0,5 indicam concordância razoável.

Para auxiliar na explanação das variáveis, suponhamos um quadro que resume os valores de previsões certas ou erradas do modelo, conhecido como matriz de confusão ou tabela de contingência. Abaixo uma tabela de contingência que apresenta os valores *A, B, C* e *D* em relação à quantidade de valores previstos (Pre) e valores de referência (Ref) reais, para duas classes, F e V, ou seja, para o quadrante em que as previsões F foram coincidentes com a realidade, o valor é *A* para o quadrante em se previu que uma observação era da classe V, mas que na verdade ela era F, têm-se a quantidade *C*, e assim por diante.

#### Matriz de Confusão

		Ref	
		F	V
P r e	F	<i>A</i>	<i>B</i>
	V	<i>C</i>	<i>D</i>

A partir da tabela de contingência calcula-se entre outros parâmetros, a acurácia:

$$Acurácia = \frac{A + D}{A + B + C + D}$$

Tabela 1: Comparativo dos parâmetros dos modelos para o conjunto teste.

Parâmetros	svmRadial	lssvmRadial	svmPoly	svmLinear
Prev F Real F (A)	27	29	30	29
Prev F Real V (B)	8	9	9	8
Prev V Real F (C)	12	10	9	10
Prev V Real V (D)	19	18	18	19
Acurácia	0,6970	0,7121	0,7273	0,7273
Intervalo de Confiança de 95%	(0,5715 ; 0,8041)	(0,5875 ; 0,817)	(0,6036 ; 0,8297)	(0,6036 ; 0,8297)
Kappa	0,3872	0,4079	0,4359	0,4423

É interessante, ainda, utilizar-se de um parâmetro mais simples para avaliação dos modelos. No caso, após uma série de testes, optou-se por um modelo de regressão linear simples, por mínimos quadrados ordinários (MQO), tomando como variável dependente e independentes, as mesmas variáveis dos demais modelos, tomados as 266 observações da primeira base já comentada, utilizando-se o software estatístico Gretl<sup>6</sup>. O modelo MQO estimado está disponível para consulta no Anexo III.

Para fins de aplicação e comparação dos resultados dos modelos com os resultados do modelo MQO, atribuiu-se valor de 1 para os aeroportos que não realizaram voos regulares e 2 para os que realizaram voos regulares. Para fins de regularização dos resultados, consideraram-se os valores resultantes abaixo de 1,5 como 1 e os acima de 1,5 como 2. Como parâmetro para comparação, temos que, após aplicação do modelo na primeira base, o nível de acerto, ou seja, a acurácia, do modelo MQO foi de 85,7%, enquanto para a segunda base, de 509 ZAs, o nível de acurácia, foi de 85,0%.

A comparação com o modelo MQO é interessante por este ser o modelo econométrico mais simples, com possibilidade de aplicação em diversos softwares de uso mais amplo e popular, e cujos resultados são bastante replicáveis. O MQO, portanto, é, seguindo o que foi explanado em relação ao princípio da parcimônia, um modelo simples e que serve como ponto de partida à comparação dos resultados entre os diferentes modelos ora elencados.

<sup>6</sup> Gretl é um software livre, com licença do tipo Gnu de documentação livre, voltado para econometria, séries temporais e regressões de diversos tipos, em plataforma amigável, com aplicação simples e intuitiva, disponível em <http://gretl.sourceforge.net/>.

No próximo capítulo serão tratados os resultados dos modelos SVM e a comparação deles entre si e com o resultado apresentado pelo modelo MQO, com especial atenção à acurácia em cada um dos modelos e nos diferentes conjuntos tratados dentro de cada base.

## 5 Resultados

A principal forma de utilização dos resultados dos modelos no caso ora apresentado se dá especialmente pela avaliação da capacidade de eles preverem eventuais aeroportos, ou zonas aeroviárias que têm potencial de receber voos regulares, conforme definido anteriormente. Primeiramente, cabe destacar que o programa computacional acusou incapacidade de rodar o modelo SVM de *Kernel* de Processo Gaussiano de Base Radial (Gauss Radial) por produzir iterações infinitas. O que, a partir do princípio da parcimônia, permite que se elimine este modelo do conjunto de elegíveis. Partindo-se à análise da aplicação dos modelos às diferentes bases observam-se os seguintes resultados

Há 110 zonas aeroviárias que receberam voos regulares em pelo menos três anos durante o período de 2004 a 2015, e sabendo que consideradas todas as zonas aeroviárias cujos aeroportos movimentaram essa quantidade de passageiros em ao menos um ano durante este período, tem-se um total de 135 zonas aeroviárias. Para efeitos de utilização para formulação de políticas públicas o que se espera da aplicação dos modelos é justamente que eles indiquem zonas aeroviárias que tenham as características necessárias que permitam a movimentação de ao menos duas movimentações semanais (uma chegada e uma partida) de uma aeronave ATR-42, com taxa de ocupação de aproximadamente 80%.<sup>7</sup> Esse indicador é interessante por sinalizar aonde se poderia investir em aeroportos em localidades em que há um nível de demanda mínimo para voos regulares.

Há, portanto, especial interesse nos erros de previsão em relação às observações em que o modelo indica que em dada ZA haveria movimentação de passageiros no patamar de corte estipulado, mas, em verdade, essa movimentação não ocorreu por motivos diversos que não estão abordados neste trabalho (e.g. restrições de infraestrutura aeroportuária).

A tabela 2 abaixo traz a proporção de previsões corretas para cada um dos grupos de zonas aeroviárias abordados, quais sejam conjunto treino, conjunto teste e a base com ZAs agregadas a posteriori. Apresentam-se, ainda, a quantidade de zonas cujos modelos preveem que teriam recebido uma movimentação semanal por pelo

---

<sup>7</sup> Os aviões ATR 42 têm configurações que de assentos que comportam até cinquenta passageiros, considerou-se, um valor de 39 assentos como de referência para cálculo dos 80% de taxa de ocupação. [http://www.atr-aircraft.com/datas/download\\_center/81/15c2188\\_fiche\\_42\\_web\\_81.pdf](http://www.atr-aircraft.com/datas/download_center/81/15c2188_fiche_42_web_81.pdf).

menos três anos, mas que não tiveram esse desempenho durante o período analisado, ou seja, previu-se V, mas o realizado foi F. Depreende-se da tabela, que os modelos sinalizaram acréscimo de diferentes quantidades de aeroportos com razoável discrepância. Por exemplo, o modelo SVM Linear indicou 8 zonas a mais no grupo treino, 10 no grupo de teste e 42 na segunda base, em relação aos 110 que já apresentaram a característica avaliada.

Tabela 2: Acurácia e número de erros do tipo Previsto V e Realizado F, por modelo e conjunto de aplicação de previsões.

Modelos SVM	Treino (n=200)		Teste (n=66)		Segunda Base (n=509)	
	Acurácia	Previsto V Real F	Acurácia	Previsto V Real F	Acurácia	Previsto V Real F
Radial	0,89	5	0,70	12	0,98	11
LS Radial	0,90	5	0,71	10	0,96	20
Polinomial	0,89	7	0,73	9	0,87	67
Linear	0,92	8	0,73	10	0,92	42

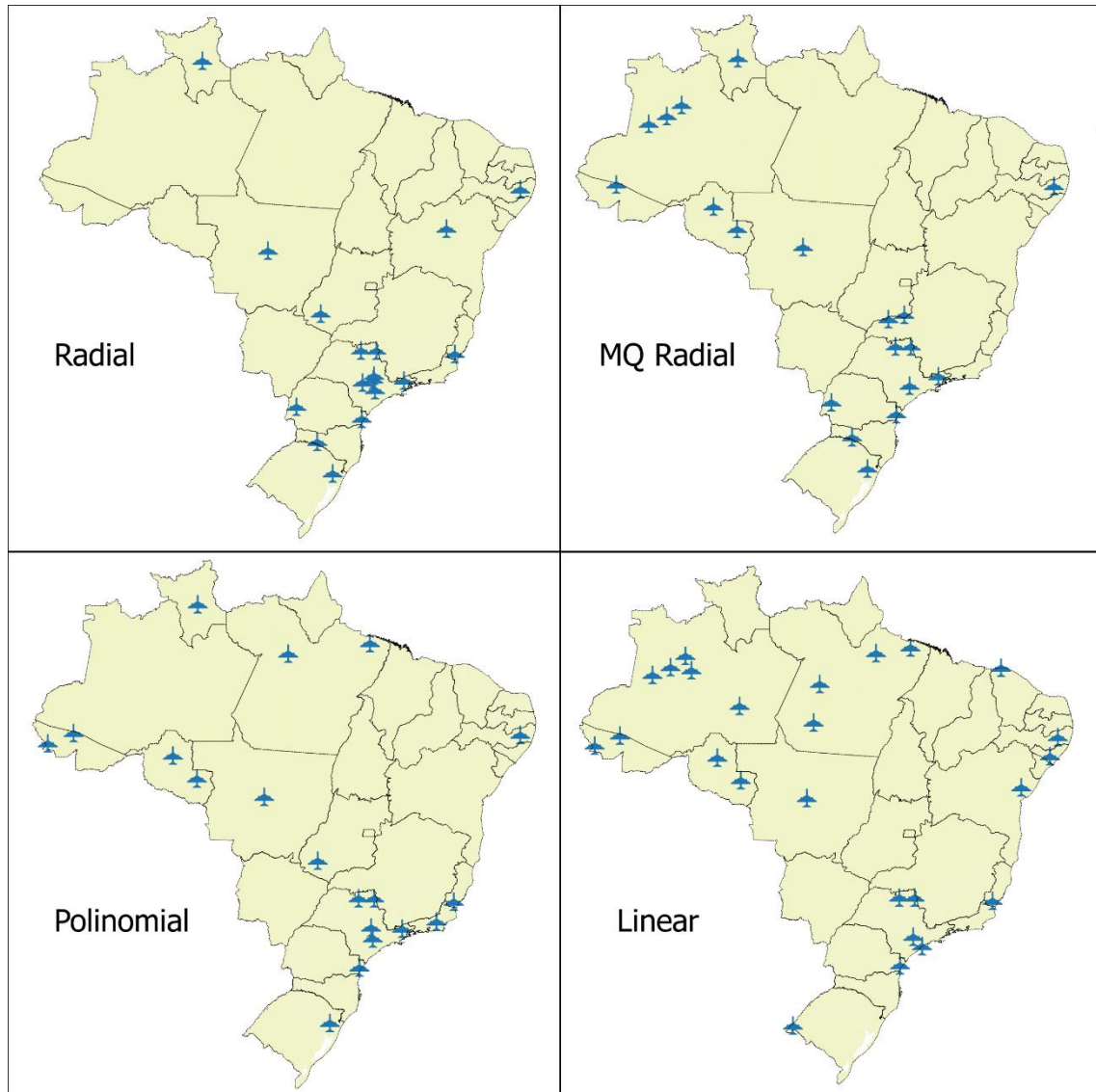
Observa-se que os resultados de acertos para os modelos SVM na base de 266 ZAs foram, quando feita a ponderação dos grupos treino e teste, de 84,3%, 85,3%, 85,0% e 87,3%, respectivamente. Lembre-se que este valor para o modelo MQO foi de 85,7%, portanto, para este grupo, o modelo MQO tem um resultado semelhante. Entretanto, quando aplicado à base de 509 ZAs, o modelo tem um nível de acerto de 85,0%, valor inferior aos apresentados pelos modelos SVM. O modelo MQO identifica 76 ZAs como sendo receptoras de voos regulares, dentre as 509 ZAs da segunda base.

Há de se avaliar, entretanto quais as ZAs que são indicadas pelos modelos. Como se observou anteriormente, é de 110 o número de ZAs que receberam em média duas movimentações de passageiros semanais em três anos distintos, e 135 as ZAs que receberam essa movimentação em ao menos um ano. Após análise mais criteriosa da aplicação dos modelos nas 775 ZAs vê-se que os resultados dos modelos não coincidem completamente sobre as ZAs que teriam movimentação, mas não têm.

O Mapa 6 traz um conjunto de quatro mapas que mostram as diferenças entre as previsões realizadas pelos diferentes modelos de SVM. Nele são destacadas as ZAs indicadas por cada um dos modelos, considerando apenas ZAs que fazem parte do PIL e que não receberam voos regulares, ou seja, são apresentados os aeroportos

classificados como de alta probabilidade de receber voo regular por cada um dos modelos.

Mapa 6: ZAs classificadas como receptoras de voo regular pelos diferentes métodos de SVM.



Fonte: Fonte: elaboração própria.

É de se destacar que apenas oito ZAs foram acrescentadas concomitantemente pelos quatro modelos: Barretos – SP, Sorocaba – SP, Taubaté – SP, Franca – SP, Caruaru – PE, Lucas do Rio Verde – MT, Paranaguá - PR e Óbidos – PA, observando-se que quatro delas já haviam atingido o patamar mínimo de movimentação em pelo menos um ano: Sorocaba, Franca, Lucas do Rio Verde e Caruaru. Um segundo grupo que chama atenção quando da análise é o das ZAs indicadas por três modelos distintos, são nove que se enquadram nesse grupo: Tarauacá – AC, Ariquemes – RO, Pimenta Bueno – RO, Guaratinguetá – SP,

Cachoeiro do Itapemirim – ES, Canela – RS, Caracaraí – RR, Novo Hamburgo – RS e Guararapes – SP, dessas, apenas a primeira teve a movimentação mínima de corte por pelo menos um ano. O Mapa 5 sinaliza as ZAs classificadas como receptoras de voo regular indicando quantas vezes elas foram selecionadas por diferentes modelos, 1, em azul claro esverdeado, indica que elas receberam classificação V por apenas um modelo e 4, em azul escuro, indica que receberam tal classificação dos quatro modelos distintos.

Mapa 7: ZAs classificadas como receptoras de voo regular pelos diferentes métodos de SVM.



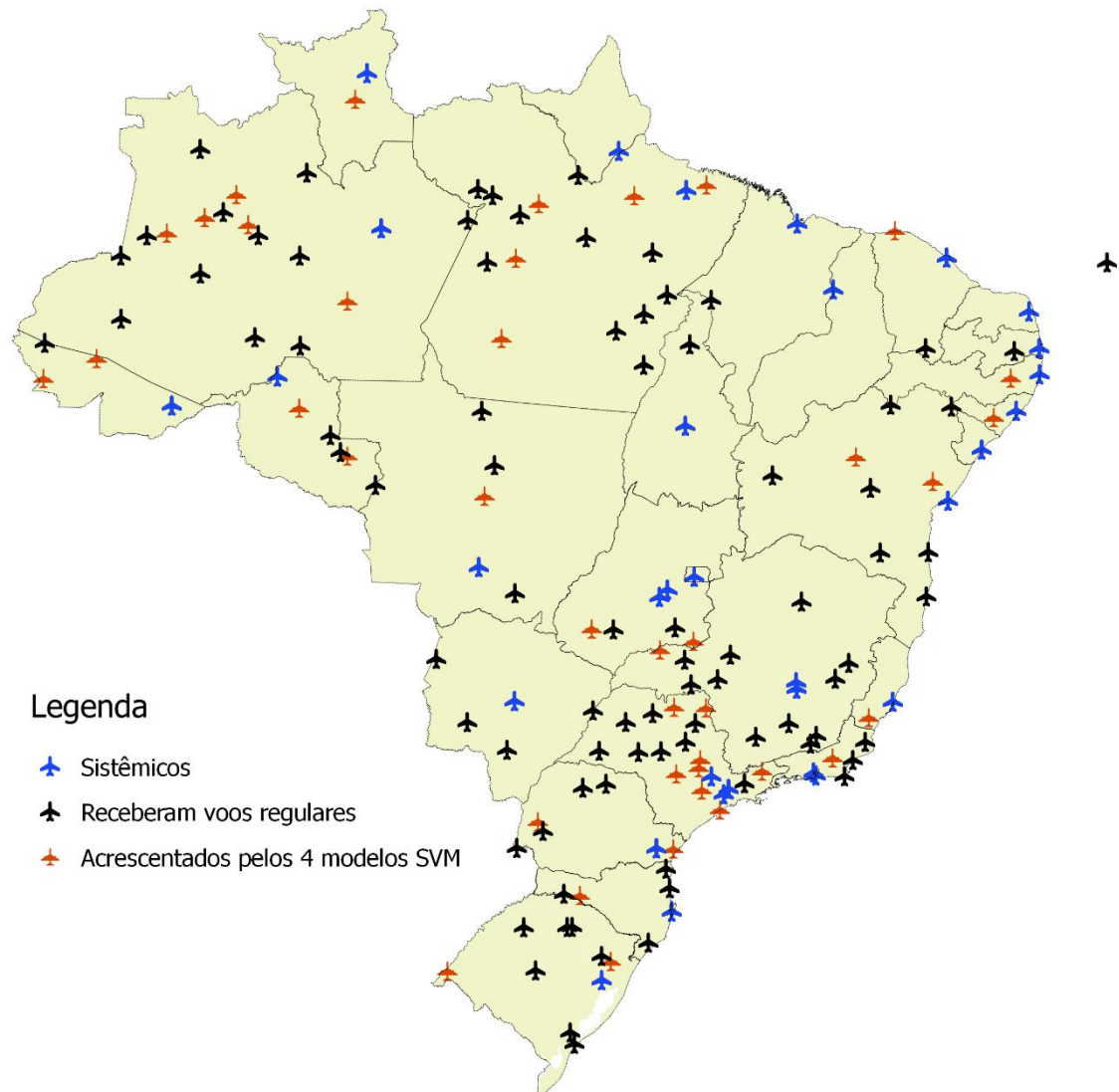
Fonte: Fonte: elaboração própria.

Essa interseção dos conjuntos formados pelos resultados dos diferentes modelos indica que há a opção de utilização conjunta dos modelos para efeitos de seleção ou



priorização de investimentos, uma vez que isso pode sinalizar que há maior probabilidade de os resultados refletirem a realidade daquelas ZAs que tem potencial mais elevado à movimentação de passageiros. Isso nos permite criar um cenário hipotético em que se aceitariam os quatro modelos, estipulando, por exemplo pesos distintos para cada uma das ZAs acrescentadas ao cenário atual. Partindo-se desta premissa, teríamos um cenário de investimento que vislumbrasse aeroportos com capacidade de receber voos regulares como apresentado no Mapa 6, em que se somam aos aeroportos sistêmicos e que já operaram voos regulares todos aqueles indicados pelos modelos, considerados apenas aqueles que fazem parte do PIL.

Mapa 8: Rede hipotética de aeroportos, constituída pela junção de resultados dos quatro modelos SVM abordados.



Fonte: Fonte: elaboração própria.

Em contraposição, após a análise mais criteriosa dos resultados, podem-se estabelecer critérios para escolha de um modelo singular que sirva de fato aos fins de seleção ou priorização de investimentos. Uma dessas variáveis que auxilia na escolha de um único modelo, no presente caso, é a distância em relação a um aeroporto sistêmico (distsistêmico), uma vez que ela sinaliza a possibilidade de competição entre aeroportos que se encontram muito próximos um do outro, o que pode exacerbar aspectos de inviabilidade de um aeroporto derivados de um excesso de oferta de infraestrutura aeroportuária.

Após a visualização dos resultados nota-se que o modelo com menor acréscimo de ZAs com potencial de recebimento de voos regulares a uma distância de até 100 km de um aeroporto sistêmico é o SVM Linear, com 7 aeroportos nesta faixa. A tabela 3, mostra a quantidade de ZAs classificadas como potenciais recebedores de voos regulares por modelo, e o total de novas zonas que cada modelo acrescenta.

Tabela 3: Total de ZAs acrescentadas, ZAs acrescentadas a menos de 100 km de um aeroporto sistêmico, e razão entre esses dois grupos.

Modelo	Radial	LsRadial	Polinomial	Linear
Total (A)	28	35	83	60
Menos de 100km de sistêmico (B)	10	9	11	7
Proporção (B/A)	0,36	0,26	0,13	0,12

Outro fator importante é a visualização de quais são essas ZAs acrescentadas a uma distância de até 100 km, aqui se destaca novamente a capacidade do modelo Linear de fazer previsões aparentemente mais promissoras, uma vez que as sete ZAs acrescentadas contém grande população e PIB per capita elevado, o que contribui para uma maior movimentação. Ou seja, é de se acreditar que as zonas aeroviárias acrescentadas pelo modelo linear têm maior potencial para de desenvolvimento de uma demanda consistente para aviação regular. A tabela 4 traz as zonas acrescentadas após os resultados de cada modelo e indicadores selecionados para as ZAs.

Tabela 4: Zonas Aeroviárias que seriam acrescentadas por critério de priorização utilizando diferentes modelos de SVM, que se encontram a uma distância inferior a 100 km de um aeroporto sistêmico

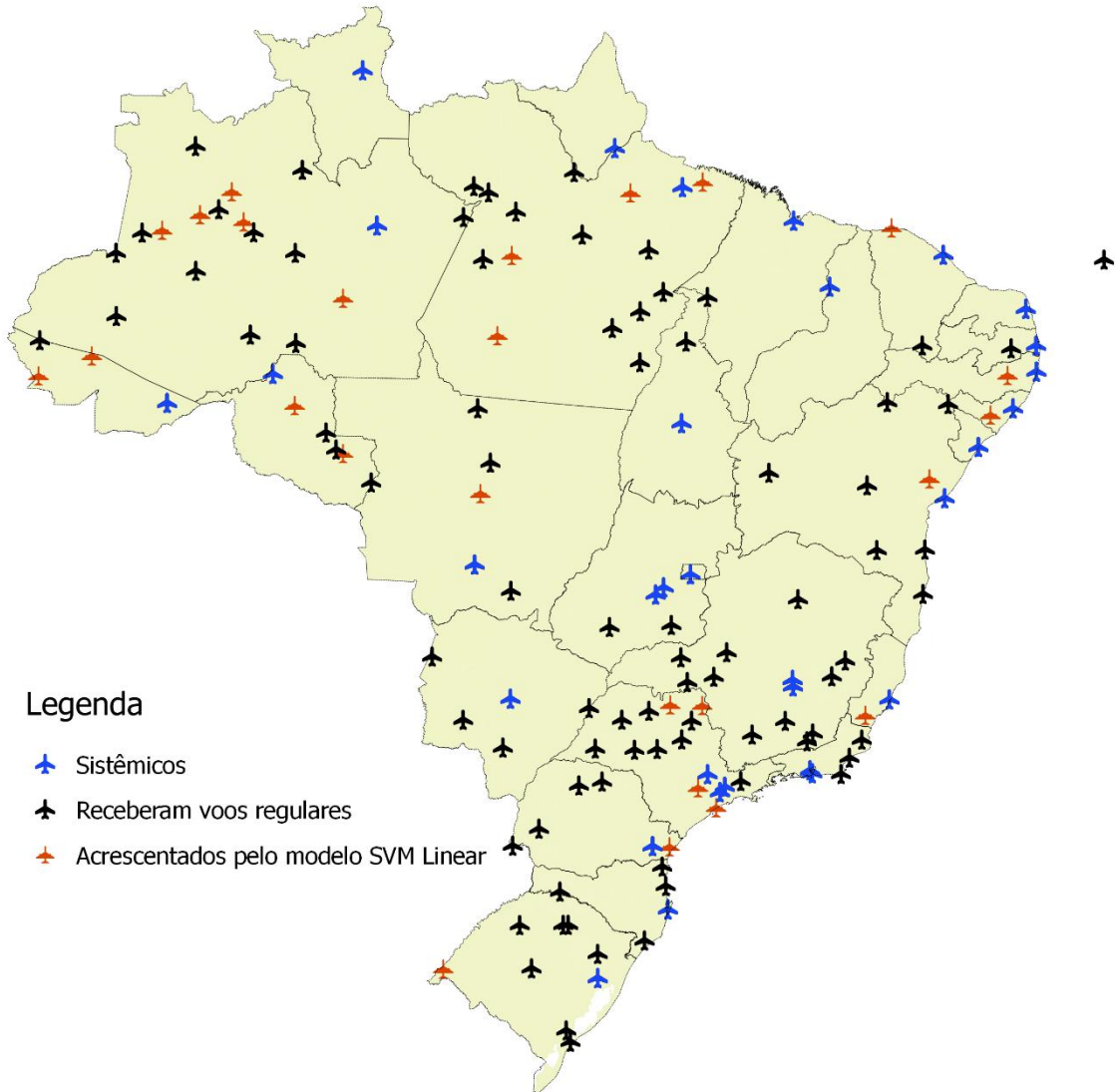
Radial	LsRadial	Poly	Linear	Grupo	SEDEZA	regiao	distsitemico	pibpc2013	popza2015
V	V	F	F	NOVA	Itu - SP	SE	30,7	49,7	281.266
F	F	F	V	NOVA	Jundiaí - SP	SE	31,5	94,0	771.335
V	V	V	F	NOVA	Novo Hamburgo - RS	S	35,1	33,4	663.137
V	V	F	F	NOVA	Itatiba - SP	SE	36,2	47,0	126.218
V	V	F	F	NOVA	Americana - SP	SE	36,5	43,8	476.225
F	F	F	V	NOVA	Limeira - SP	SE	60,4	41,1	318.997
F	F	V	F	NOVA	Taquara - RS	S	61,1	28,0	173.157
F	F	F	V	TREINO	Santos - SP	SE	61,6	32,2	1.732.274
V	F	V	F	TREINO	Piracicaba - SP	SE	63,2	56,0	475.843
V	V	V	V	TESTE	Sorocaba - SP	SE	65,8	43,1	958.177
F	V	F	F	NOVA	Cerquillo - SP	SE	71,4	39,3	115.862
V	V	V	V	TESTE	Paranaguá - PR	S	71,7	48,9	150.660
V	V	F	F	NOVA	Mogi Guaçu - SP	SE	78,4	39,0	347.486
V	V	V	F	TESTE	Canela - RS	S	78,9	32,0	77.016
V	F	F	F	TREINO	Rio Claro - SP	SE	79,9	55,5	258.882
F	F	V	F	NOVA	São Bento do Sul - SC	S	81,4	34,3	134.530
F	F	V	V	NOVA	Castanhal - PA	N	88,9	12,1	535.313
F	F	V	F	NOVA	Mafra - SC	S	94,6	28,2	88.708
F	F	V	F	NOVA	Bento Gonçalves - RS	S	94,9	51,8	184.529
F	F	F	V	TREINO	Arapiraca - AL	NE	98,5	10,5	984.162
F	F	V	F	NOVA	Nova Friburgo - RJ	SE	99,0	23,8	184.786

Há, após o que foi até aqui apresentado, indicações de que o melhor modelo para utilização individual é o modelo Linear. Primeiramente, recorrendo-se ao princípio da parcimônia, esse é o modelo mais simples. Em segundo lugar, é o modelo que trouxe menor proporção de resultados acrescentados a uma distância inferior a 100 km de um aeroporto sistêmico. Em terceiro, e por fim, estas zonas apresentam indicadores de população e PIB per capita em geral elevados o que permite inferir que tenham de fato maior potencial para receber voos regulares.

Da mesma forma que se apresentou a possibilidade de uma rede de aeroportos que abrange os aeroportos sistêmicos, os aeroportos que já operaram aviação regular

de passageiros e os que são incluídos pela classificação do modelo SVM linear, considerando-se novamente os que fazem parte do PIL, o Mapa 7 traz esse cenário.

Mapa 7: Rede hipotética de aeroportos, constituída pela junção de resultados do modelo SVM linear.



Fonte: Fonte: elaboração própria.

## 5.1 Aprendizado de máquina como instrumento de tomada de decisão e priorização de investimentos.

O aprendizado de máquina é utilizado atualmente como ferramenta de tomada de decisão em diversos setores, em especial no setor financeiro, por exemplo máquinas de suporte vetorial que se utilizam de avaliação de séries temporais de valores de ativos e outros elementos para classificação de ativos na formação de carteiras de

investimento. No setor público federal brasileiro a utilização de aprendizado de máquina ainda é incipiente no aperfeiçoamento de políticas públicas ou tomada de decisão. Mostra-se aqui uma janela interessante para utilização de aprendizado de máquina para a tomada de decisão tanto no longo prazo, na seleção de municípios, no caso específico ZAs, com maior potencial de utilização efetiva da infraestrutura pública.

Um exemplo que se vislumbra para fins de priorização de investimentos: é possível utilizar um dos dois cenários hipotéticos aqui apresentados. No primeiro, seriam selecionados todos os aeroportos classificados como recebedores de voo regular, podendo-se dar prioridade aos que receberam tal classificação por uma maior quantidade de modelos, ao exemplo daqueles que foram classificados com V pelos quatro modelos: Barretos – SP, Sorocaba – SP, Taubaté – SP, Franca – SP, Caruaru – PE, Lucas do Rio Verde – MT, Paranaguá - PR e Óbidos – PA. Outra forma de utilização de SVM seria com a utilização de um único modelo, com resultados aparentemente mais consistentes, como foi o caso do modelo linear, posteriormente seria necessário estabelecer novo parâmetro, ou conjunto de parâmetros, para priorização, exemplificando de forma simplificada, é possível considerar, prioritário o investimento em um aeroporto na Zona Aeroviária de Santos, uma vez que esta é a ZA de maior população sem acesso direto a um aeroporto, possuindo ainda renda *per capita* relativamente elevada.

Assim, a utilização de modelos SVM, pode se apresentar como uma alternativa a modelos tradicionalmente utilizados para priorização de investimentos em infraestrutura, ou mesmo em setores não ligados às áreas de infraestrutura, como a AHP, ou processo de análise hierárquica, desenvolvido por Thomas Saaty na década de 1970 consolidado em Saaty (1980), que é utilizado em diversos setores.

## 6 Conclusões

Foi abordado ao longo desta dissertação o atual estágio da aviação civil brasileira em relação ao atendimento da população em termos agregados e formas que o Estado vem buscando atuar para alterar essa realidade, especialmente o Programa de Investimento em Logística: Aeroportos e o Plano Aeroviário Nacional. O primeiro uma política de investimento de curto e médio prazo e o segundo uma política que visa à auxiliar a tomada de decisão no longo prazo. Ambos necessitam de ferramentas que permitam avaliação objetiva em relação às necessidades de investimento e potenciais regionais para desenvolvimento da aviação.

O aprendizado de máquina e as máquinas de suporte vetorial são pouco utilizados no âmbito do Governo Federal Brasileiro a despeito de já serem utilizados em diversos setores, inclusive para tomada de decisão e direcionamento de investimentos, ressaltando-se sua utilização no setor financeiro. Essas ferramentas, em diferentes formas de aplicação e com metodologias distintas, mostram-se promissoras para o desenvolvimento de políticas públicas e aperfeiçoamento do gasto público. Em termos práticos, este trabalho apresentou como se pode aplicar máquinas de suporte vetorial como auxiliar na tomada de decisão para fins de investimentos públicos em aeroportos.

Partiu-se do pressuposto de que o PIL: Aeroportos visa ao investimento em aeroportos regionais com intuito de que eles possuam infraestrutura condizente com a operação de no mínimo aeronaves modelo ATR-42. Definindo um voo regular como um voo semanal feito por um ATR-42, têm-se os aeroportos que fizeram, em média, voos regulares ao longo de 3 anos distintos, não necessariamente consecutivos. A partir daí, e de uma cesta de indicadores socioeconômicos e regionais, utilizou-se de aprendizado de máquina para classificar Zonas Aeroviárias em receptoras ou não receptoras de voos regulares com distintos modelos SVM. Aplicando-se os modelos a conjuntos distintos de aeroportos tem-se uma rede hipotética de aeródromos que receberiam voos regulares, se tivessem infraestrutura para tal.

Verificando quais as zonas que de fato são indicadas como receptoras de voos regulares é possível montar uma carteira de investimentos, com prioridades de investimentos, ou que aponte os locais em que de fato faz sentido se investir em uma infraestrutura aeroportuária capaz de receber aeronaves ATR-42, ou superiores. Por outro lado, é de se utilizar a ferramenta para que se evitem gastos desnecessários com infraestrutura ociosa, em localidades, que no exemplo em tela, não tem potencial de receber voos regulares, dada uma pletera de parâmetros.

Não obstante, é de se salientar, os programas computacionais estatísticos e econométricos utilizados para este trabalho, apesar de isto não ser algo mencionado até aqui ao logo do trabalho, a redução de custos permitida por meio da utilização de softwares livres, pode estar na casa dos milhares de reais anuais para cada licença utilizada individualmente. Uma conclusão secundária deste trabalho é de que, hodiernamente, os aplicativos e softwares de custo zero podem produzir resultados excelentes.

Conclui-se que o Aprendizado de máquina se mostra como uma ferramenta de trabalho poderosa, com muitas possibilidades a serem empreendidas, à boa execução dos gastos públicos.

## 7 Bibliografia

ANAC (2016), **Anuário do Transporte Aéreo 2014**, volume único, 1ª edição. Agência Nacional de Aviação Civil – ANAC. Brasília – Brasil, 2015.

BANCO MUNDIAL, **World DataBank: World Development Indicators**.

<http://databank.worldbank.org/data/reports.aspx?source=2&series=IS.AIR.PSGR&country=#> , Acessado em 04 de outubro de 2016.

BARBER, David. **Bayesian Reasoning and Machine Learning**. Cambridge University Press, 2011. Disponível em: <<http://eprints.pascal-network.org/archive/00007920/>\n<http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Bayesian+Reasoning+and+Machine+Learning#0>>.

BISHOP, Christopher M. **Pattern Recognition and Machine Learning**. Springer Science+Business Media, LLC. Printed in Singapore. 2006.

BOSER, Bernhard E. e GUYON, Isabelle M. e VAPNIK, Vladimir N. **A Training Algorithm for Optimal Margin Classifiers**. Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory, p. 144–152, 1992. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.21.3818>>.

BRASIL, Ministério dos Transportes, Portos e Aviação Civil. **Sítio oficial do Programa federal de auxílio a aeroportos**, <http://www.aviacao.gov.br/acesso-a-informacao/investimentos/programas/programa-federal-de-auxilio-a-aeroportos-profaa> . Acessado em 04 de outubro de 2016.

CHOU, Jui-Sheng (Rayson) e LIN, Chieh. **Predicting Disputes in Public-Private Partnership Projects: Classification and Ensemble Models**. Journal of Computing in Civil Engineering, v. 27, n. February, p. 51–60, 2012.

CORTES, Corinna e VAPNIK, Vladimir. **Support-Vector Networks**. Machine Learning, v. 20, n. 3, p. 273–297, 1995.

COTTRELL, Allin e LUCCHETTI, Riccardo. **Gretl: Gnu regression, econometrics and time-series library**. 2016. Disponível em <http://gretl.sourceforge.net/> .

HASTIE, Trevor e TIBSHIRANI, Robert e FRIEDMAN, Jerome. **The Elements of Statistical Learning**. [S.l: s.n.], 2009. v. 1. Disponível em: <<http://www.springerlink.com/index/10.1007/b94608>>.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Arranjos Populacionais e Concentrações Urbanas do Brasil**. 2015. Disponível em: <[http://www.ibge.gov.br/apps/arranjos\\_populacionais/2015/pdf/publicacao.pdf](http://www.ibge.gov.br/apps/arranjos_populacionais/2015/pdf/publicacao.pdf)>.

KARATZOGLU, Alexandros *et al.* **kernlab – An S4 Package for Kernel Methods in R**. Journal of Statistical Software, v. 11, n. 9, p. 1–20, 2004. Disponível em: <<http://www.jstatsoft.org/v11/i09/paper>>.

KUHN, Max. **Building Predictive Models in R Using the caret Package**. Journal Of



Statistical Software, v. 28, n. 5, p. 1–26, 2008. Disponível em:  
<<http://www.jstatsoft.org/v28/i05/>>.

KUHN, Max e Colab. **caret: Classification and Regression Training. R package version 6.0-71**. 2016 <https://CRAN.R-project.org/package=caret>

MARCELINO, Sarah S.F. **Formação De Portfólio Por Meio De Máquinas De Suporte Vetorial e Deep Learning**. 2016. 2016.

MITCHELL, Tom M. **Machine Learning**. New York, USA. McGraw-Hill Science/Engineering/Math. 1997.

R Core Team 2015. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria., 2015. Disponível em <https://www.R-project.org/> .

RASMUSSEN, C.E. E WILLIAMS, C.K.I. **Gaussian processes for machine learning**. the MIT press, Massachusetts Institute of Technology, 2006. Disponível em:  
<<http://www.gaussianprocess.org/gpml/chapters/RW.pdf>>.

SAATY, T.L. **Decision making: the analytic hierarchy process: planning, priority setting, resource allocation**. New York, USA. McGraw Hill, 1980.

SUYKENS, J e .A.K.; VANDEWALLE, J. **Least Squares Support Vector Machine Classifiers**. Neural Processing Letters, v. 9, n. 3, p. 293–300, 1999.

XIAO, Yi e colab. **A neuro-fuzzy combination model based on singular spectrum analysis for air transport demand forecasting**. Journal of Air Transport Management, v. 39, p. 1–11, 2014.

## 8 Anexo I: programação utilizada em R para os modelos SVM.

```
# Abertura da base de dados.
base = read.csv2("C:\\Users\\SVMCOMP1.csv", head = TRUE)
attach(base)

# Ativação do pacote caret.
library(caret)

# Definição de semente que permita a replicação do modelo.
set.seed(1)

# Separação da base de dados em conjuntos de treinamento
(treino), 3/4, e de teste (teste), 1/4 e designação da variável
classe.

treino = createDataPartition(tresanosatr, p = .75, list = FALSE)

basetreino = base[treino,]
baseteste = base[-treino,]
classetreino = tresanosatr[treino]
classeteste = tresanosatr[-treino]

# Verificação das probabilidades das classes da variável
tresanosreg, que distingue os aeroportos que tiveram voo regular
segundo definição apresentada de ao menos três anos com mais de
um voo semanal de média, após a partição dos dados nos
subconjuntos treino e teste.

prop.table(table(tresanosatr))
prop.table(table(classetreino))
prop.table(table(classeteste))

# Rodando os diferentes modelos após adequação da base,
escolhendo utilizar a base total e não a reduzida.

set.seed(1)

modelosvm1 = train(tresanosatr ~., data = basetreino, method =
"svmRadial", preProc = c("center", "scale"))

set.seed(1)

modelosvm2 = train(tresanosatr ~., data = basetreino, method =
"gaussprRadial", preProc = c("center", "scale"))

set.seed(1)

modelosvm3 = train(tresanosatr ~., data = basetreino, method =
"lssvmRadial", preProc = c("center", "scale"))

set.seed(1)

modelosvm4 = train(tresanosatr ~., data = basetreino, method =
"svmPoly", preProc = c("center", "scale"))

set.seed(1)

modelosvm5 = train(tresanosatr ~., data = basetreino, method =
"svmRadialCost", preProc = c("center", "scale"))
```

```

set.seed(1)

modelosvm6 = train(tresanosatr ~., data = basetreino, method =
"svmRadialSigma", preProc = c("center", "scale"))

set.seed(1)

modelosvm7 = train(tresanosatr ~., data = basetreino, method =
"svmLinear", preProc = c("center", "scale"))

set.seed(1)

modelosvm8 = train(tresanosatr ~., data = basetreino, method =
"svmRadialWeights", preProc = c("center", "scale"))

#Previsão svmRadialSigma customizado.
Prev1 = predict(modelosvm1, baseteste[,-1])

#Previsão gaussprRadial.
prev2 = predict(modelosvm2, baseteste[,-1])

#Previsão lssvmRadial.
prev3 = predict(modelosvm3, baseteste[,-1])

#Previsão svmPoly.
prev4 = predict(modelosvm4, baseteste[,-1])

#Previsão svmLinear customizado.
Prev5 = predict(modelosvm5, baseteste[,-1])

#Matrizes e estatísticas de resultado de cada modelo:
matriz1 = confusionMatrix(prev1, classeteste)
matriz2 = confusionMatrix(prev2, classeteste)
matriz3 = confusionMatrix(prev3, classeteste)
matriz4 = confusionMatrix(prev4, classeteste)
matriz5 = confusionMatrix(prev5, classeteste)

matriz1
matriz2
matriz3
matriz4
matriz5

modelosvm1
modelosvm2
modelosvm3
modelosvm4
modelosvm5

#Previsões dos TREINOS.
tprev1 = predict(modelosvm1, basetreino[,-1])

tprev2 = predict(modelosvm2, basetreino[,-1])

tprev3 = predict(modelosvm3, basetreino[,-1])

tprev4 = predict(modelosvm4, basetreino[,-1])

tprev5 = predict(modelosvm5, basetreino[,-1])

```

```
cbind(basetreino[,1:3], tprev1, tprev2, tprev3, tprev4, tprev5)
cbind(baseteste[,1:3], prev1, prev2, prev3, prev4, prev5)
base2 = read.csv2("C:\\Users\\SVMCOMP2.csv", head = TRUE)
attach(base2)

# Previsões usando nova base.
nprev1 = predict(modelosvm1, base2[,-1])
nprev2 = predict(modelosvm2, base2[,-1])
nprev3 = predict(modelosvm3, base2[,-1])
nprev4 = predict(modelosvm4, base2[,-1])
nprev5 = predict(modelosvm5, base2[,-1])
cbind(base2[,1:3], nprev1, nprev3, nprev4, nprev5)
```

## 9 Anexo II: Lista das ZAs, classificações e localização da sede.

<b>CLASSE</b>	<b>ZA</b>	<b>PIL-270</b>	<b>SISTEMICO</b>	<b>Município Sede - UF</b>
TREINO	2	Sim	Não	Açailândia - MA
TREINO	20	Sim	Não	Araguaína - TO
TREINO	22	Sim	Não	Arapiraca - AL
TREINO	27	Sim	Não	Araripina - PE
TREINO	29	Sim	Não	Araxá - MG
TREINO	32	Sim	Não	Bacabal - MA
TREINO	33	Sim	Não	Bagé - RS
TREINO	34	Sim	Não	Santos - SP
TREINO	35	Sim	Não	Barbacena - MG
TREINO	38	Sim	Não	Barreiras - BA
TREINO	39	Sim	Não	Barretos - SP
TREINO	42	Não	Sim	Belo Horizonte - MG
TREINO	47	Não	Sim	Boa Vista - RR
TREINO	49	Sim	Não	Bom Jesus do Itabapoana - RJ
TREINO	50	Sim	Não	Botucatu - SP
TREINO	53	Não	Sim	Brasília - DF
TREINO	56	Sim	Não	Cabo Frio - RJ
TREINO	59	Sim	Não	Caetés - PE
TREINO	65	Sim	Não	Campina Grande - PB
TREINO	66	Não	Sim	Campinas - SP
TREINO	70	Sim	Não	Canoinhas - SC
TREINO	77	Sim	Não	Carolina - MA
TREINO	81	Sim	Não	Catalão - GO
TREINO	83	Sim	Não	Caxambu - MG
TREINO	84	Sim	Não	Caxias do Sul - RS
TREINO	88	Sim	Não	Chapecó - SC
TREINO	92	Sim	Não	Colatina - ES
TREINO	93	Sim	Não	Conceição do Almeida - BA
TREINO	94	Sim	Não	Conselheiro Lafaiete - MG
TREINO	98	Não	Sim	Cuiabá - MT
TREINO	100	Não	Sim	Curitiba - PR
TREINO	103	Sim	Não	Dourados - MS
TREINO	107	Sim	Não	Feira de Santana - BA
TREINO	108	Sim	Não	Fernandópolis - SP
TREINO	109	Sim	Não	Floriano - PI
TREINO	110	Não	Sim	Florianópolis - SC
TREINO	117	Sim	Sim	Goiânia - GO
TREINO	118	Sim	Não	Governador Valadares - MG
TREINO	121	Sim	Não	Guarapuava - PR
TREINO	126	Sim	Não	Ilhéus - BA
TREINO	128	Sim	Não	Imperatriz - MA
TREINO	130	Sim	Não	Santa Vitória do Palmar - RS
TREINO	138	Sim	Não	Corumbá - MS

TREINO	143	Sim	Não	Guajará-Mirim - RO
TREINO	146	Sim	Não	Oiapoque - AP
TREINO	154	Sim	Não	São Borja - RS
TREINO	155	Sim	Não	Tabatinga - AM
TREINO	156	Sim	Não	Uruguaiana - RS
TREINO	157	Sim	Não	Ipatinga - MG
TREINO	160	Sim	Não	Itajaí - SC
TREINO	176	Sim	Não	Janaúba - MG
TREINO	181	Sim	Não	Jequié - BA
TREINO	182	Sim	Não	Ji-Paraná - RO
TREINO	183	Sim	Não	Joaçaba - SC
TREINO	185	Não	Sim	João Pessoa - PB
TREINO	186	Sim	Não	Joinville - SC
TREINO	191	Sim	Não	Lages - SC
TREINO	194	Sim	Não	Laranjal do Jari - AP
TREINO	199	Sim	Não	Londrina - PR
TREINO	201	Sim	Não	Macaé - RJ
TREINO	203	Não	Sim	Maceió - AL
TREINO	206	Não	Sim	Manaus - AM
TREINO	208	Sim	Não	Marabá - PA
TREINO	210	Sim	Não	Marília - SP
TREINO	211	Sim	Não	Maringá - PR
TREINO	218	Sim	Não	Montes Claros - MG
TREINO	220	Sim	Não	Mossoró - RN
TREINO	221	Sim	Não	Muriaé - MG
TREINO	222	Não	Sim	Natal - RN
TREINO	230	Sim	Não	Ourinhos - SP
TREINO	231	Não	Sim	Palmas - TO
TREINO	237	Sim	Não	Parauapebas - PA
TREINO	239	Sim	Não	Parnaíba - PI
TREINO	240	Sim	Não	Passo Fundo - RS
TREINO	242	Sim	Não	Pato Branco - PR
TREINO	246	Sim	Não	Paulo Afonso - BA
TREINO	251	Sim	Não	Pelotas - RS
TREINO	253	Sim	Não	Petrolina - PE
TREINO	255	Sim	Não	Picos - PI
TREINO	256	Sim	Não	Piracicaba - SP
TREINO	257	Sim	Não	Pirapora - MG
TREINO	263	Sim	Não	Pombal - PB
TREINO	264	Sim	Não	Ponta Grossa - PR
TREINO	265	Sim	Não	Ponte Nova - MG
TREINO	268	Sim	Não	Porto Seguro - BA
TREINO	269	Não	Sim	Porto Velho - RO
TREINO	270	Sim	Não	Pouso Alegre - MG
TREINO	274	Sim	Não	Presidente Prudente - SP
TREINO	278	Não	Sim	Recife - PE

TREINO	279	Sim	Não	Redenção - CE
TREINO	280	Sim	Não	Resende - RJ
TREINO	281	Sim	Não	Ribeirão Preto - SP
TREINO	282	Não	Sim	Rio Branco - AC
TREINO	283	Sim	Não	Rio Claro - SP
TREINO	284	Não	Sim	Rio de Janeiro - RJ
TREINO	286	Sim	Não	Rio Grande - RS
TREINO	287	Sim	Não	Rio Verde - GO
TREINO	289	Sim	Não	Rondonópolis - MT
TREINO	291	Não	Sim	Salvador - BA
TREINO	293	Sim	Não	Santa Cruz do Sul - RS
TREINO	295	Sim	Não	Santa Helena - MA
TREINO	296	Sim	Não	Santa Maria da Vitória - BA
TREINO	309	Sim	Não	São João del Rei - MG
TREINO	311	Sim	Não	São José do Rio Preto - SP
TREINO	312	Sim	Não	São José dos Campos - SP
TREINO	314	Não	Sim	São Luís - MA
TREINO	316	Não	Sim	São Paulo - SP
TREINO	321	Sim	Não	Serra Talhada - PE
TREINO	325	Sim	Não	Sobral - CE
TREINO	334	Sim	Não	Teixeira de Freitas - BA
TREINO	335	Sim	Não	Teófilo Otoni - MG
TREINO	336	Não	Sim	Teresina - PI
TREINO	344	Sim	Não	Três Lagoas - MS
TREINO	349	Sim	Não	Ubá - MG
TREINO	350	Sim	Não	Ubaitaba - BA
TREINO	351	Sim	Não	Uberaba - MG
TREINO	354	Sim	Não	União da Vitória - PR
TREINO	358	Sim	Não	Varginha - MG
TREINO	362	Sim	Não	Vitória da Conquista - BA
TREINO	363	Não	Sim	Serra - ES
TREINO	374	Sim	Não	Alta Floresta - MT
TREINO	375	Sim	Não	Altamira - PA
TREINO	378	Sim	Não	Alto Paraíso de Goiás - GO
TREINO	381	Sim	Não	Amapá - AP
TREINO	388	Sim	Não	Aracati - CE
TREINO	395	Sim	Não	Ariquemes - RO
TREINO	400	Sim	Não	Balsas - MA
TREINO	402	Sim	Não	Bandeirantes - PR
TREINO	403	Sim	Não	Barcelos - AM
TREINO	405	Sim	Não	Barra do Corda - MA
TREINO	407	Sim	Não	Barreirinhas - MA
TREINO	411	Sim	Não	Boca do Acre - AM
TREINO	412	Sim	Não	Bom Jesus - PI
TREINO	428	Sim	Não	Cáceres - MT
TREINO	430	Sim	Não	Cacoal - RO

TREINO	440	Sim	Não	Campo Mourão - PR
TREINO	452	Sim	Não	Carauari - AM
TREINO	469	Sim	Não	Cipó - BA
TREINO	470	Sim	Não	Coari - AM
TREINO	478	Sim	Não	Conceição do Araguaia - PA
TREINO	485	Sim	Não	Costa Rica - MS
TREINO	486	Sim	Não	Coxim - MS
TREINO	492	Sim	Não	Cruzeiro do Sul - AC
TREINO	497	Sim	Não	Diamantina - MG
TREINO	500	Sim	Não	Eirunepé - AM
TREINO	503	Sim	Não	Erechim - RS
TREINO	508	Sim	Não	Fernando de Noronha - PE
TREINO	520	Sim	Não	Guanambi - BA
TREINO	526	Sim	Não	Gurupi - TO
TREINO	528	Sim	Não	Humaitá - AM
TREINO	533	Sim	Não	Ibotirama - BA
TREINO	539	Sim	Não	Itaberaba - BA
TREINO	540	Sim	Não	Itacoatiara - AM
TREINO	541	Sim	Não	Itaituba - PA
TREINO	547	Sim	Não	Jacareacanga - PA
TREINO	548	Sim	Não	Jacobina - BA
TREINO	555	Sim	Não	Jataí - GO
TREINO	558	Sim	Não	João Pinheiro - MG
TREINO	559	Sim	Não	Juara - MT
TREINO	561	Sim	Não	Juína - MT
TREINO	567	Sim	Não	Lábrea - AM
TREINO	585	Sim	Não	Matupá - MT
TREINO	586	Sim	Não	Maués - AM
TREINO	588	Sim	Não	Minaçu - GO
TREINO	592	Sim	Não	Monte Alegre - PA
TREINO	595	Sim	Não	Monteiro - PB
TREINO	604	Sim	Não	Naviraí - MS
TREINO	610	Sim	Não	Nova Olinda do Norte - AM
TREINO	618	Não	Não	Oriximiná - PA
TREINO	619	Sim	Não	Ourilândia do Norte - PA
TREINO	625	Sim	Não	Paracatu - MG
TREINO	626	Sim	Não	Paragominas - PA
TREINO	632	Sim	Não	Pauini - AM
TREINO	633	Sim	Não	Paulistana - PI
TREINO	638	Sim	Não	Pimenta Bueno - RO
TREINO	647	Sim	Não	Pontes e Lacerda - MT
TREINO	648	Sim	Não	Porangatu - GO
TREINO	662	Sim	Não	Redenção - PA
TREINO	663	Sim	Não	Registro - SP
TREINO	675	Sim	Não	Salgueiro - PE
TREINO	676	Sim	Não	Salinas - MG



TREINO	682	Sim	Não	Santa Isabel do Rio Negro - AM
TREINO	687	Sim	Não	Santa Rosa - RS
TREINO	689	Sim	Não	Santana do Araguaia - PA
TREINO	695	Sim	Não	São Félix do Araguaia - MT
TREINO	696	Sim	Não	São Félix do Xingu - PA
TREINO	700	Sim	Não	São Gabriel da Cachoeira - AM
TREINO	706	Sim	Não	São Joaquim - SC
TREINO	714	Sim	Não	São Miguel do Oeste - SC
TREINO	715	Sim	Não	São Paulo de Olivença - AM
TREINO	716	Sim	Não	São Raimundo Nonato - PI
TREINO	720	Sim	Não	Sento Sé - BA
TREINO	729	Sim	Não	Tangará da Serra - MT
TREINO	731	Sim	Não	Tarauacá - AC
TREINO	734	Sim	Não	Tefé - AM
TREINO	744	Sim	Não	Unaí - MG
TREINO	767	Sim	Não	Angra dos Reis - RJ
TREINO	770	Sim	Não	Paraty - RJ
TREINO	781	Sim	Não	Tucuruí - PA
TREINO	782	Não	Não	Confresa - MT
TESTE	19	Não	Sim	Aracaju - SE
TESTE	25	Sim	Não	Araraquara - SP
TESTE	37	Sim	Não	Barra do Garças - MT
TESTE	40	Sim	Não	Bauru - SP
TESTE	41	Não	Sim	Belém - PA
TESTE	58	Sim	Não	Cachoeiro de Itapemirim - ES
TESTE	62	Sim	Não	Caldas Novas - GO
TESTE	67	Não	Sim	Campo Grande - MS
TESTE	68	Sim	Não	Campos dos Goytacazes - RJ
TESTE	69	Sim	Não	Canela - RS
TESTE	78	Sim	Não	Caruaru - PE
TESTE	79	Sim	Não	Cascavel - PR
TESTE	96	Sim	Não	Criciúma - SC
TESTE	102	Sim	Não	Divinópolis - MG
TESTE	111	Não	Sim	Fortaleza - CE
TESTE	112	Sim	Não	Franca - SP
TESTE	113	Sim	Não	Garanhuns - PE
TESTE	122	Sim	Não	Guaratinguetá - SP
TESTE	140	Sim	Não	Foz do Iguaçu - PR
TESTE	187	Sim	Não	Juazeiro do Norte - CE
TESTE	188	Sim	Não	Jucás - CE
TESTE	189	Sim	Não	Juiz de Fora - MG
TESTE	198	Sim	Não	Linhares - ES
TESTE	202	Sim	Sim	Macapá - AP
TESTE	234	Sim	Não	Paranaguá - PR
TESTE	238	Sim	Não	Parintins - AM
TESTE	241	Sim	Não	Passos - MG

TESTE	243	Sim	Não	Patos de Minas - MG
TESTE	244	Sim	Não	Patos - PB
TESTE	262	Sim	Não	Poços de Caldas - MG
TESTE	267	Não	Sim	Porto Alegre - RS
TESTE	297	Sim	Não	Santa Maria - RS
TESTE	298	Sim	Não	Santarém - PA
TESTE	299	Sim	Não	Santo Ângelo - RS
TESTE	301	Sim	Não	Araçatuba - SP
TESTE	315	Sim	Não	São Mateus - ES
TESTE	324	Sim	Não	Sinop - MT
TESTE	328	Sim	Não	Sorocaba - SP
TESTE	340	Sim	Não	Toledo - PR
TESTE	352	Sim	Não	Uberlândia - MG
TESTE	353	Sim	Não	Umuarama - PR
TESTE	372	Sim	Não	Alegrete - RS
TESTE	414	Sim	Não	Bonito - MS
TESTE	416	Sim	Não	Borba - AM
TESTE	420	Sim	Não	Breves - PA
TESTE	426	Sim	Não	Caçador - SC
TESTE	451	Sim	Não	Caracaraí - RR
TESTE	480	Sim	Não	Concórdia - SC
TESTE	483	Sim	Não	Corrente - PI
TESTE	487	Sim	Não	Crateús - CE
TESTE	509	Sim	Não	Fonte Boa - AM
TESTE	512	Sim	Não	Francisco Beltrão - PR
TESTE	538	Sim	Não	Irecê - BA
TESTE	546	Sim	Não	Ituiutaba - MG
TESTE	551	Sim	Não	Jaguaruna - SC
TESTE	569	Sim	Não	Lençóis - BA
TESTE	572	Sim	Não	Lucas do Rio Verde - MT
TESTE	577	Sim	Não	Manicoré - AM
TESTE	615	Sim	Não	Novo Progresso - PA
TESTE	641	Sim	Não	Pirenópolis - GO
TESTE	645	Sim	Não	Piumhi - MG
TESTE	673	Sim	Não	Rurópolis - PA
TESTE	681	Sim	Não	Santa Inês - MA
TESTE	735	Sim	Não	Telêmaco Borba - PR
TESTE	758	Sim	Não	Vila Rica - MT
TESTE	759	Sim	Não	Vilhena - RO
NOVO	1	Não	Não	Abaetetuba - PA
NOVO	3	Não	Não	Adamantina - SP
NOVO	4	Não	Não	Adrianópolis - PR
NOVO	5	Não	Não	Águas de Lindóia - SP
NOVO	6	Não	Não	Alagoa Nova - PB
NOVO	7	Não	Não	Alagoinhas - BA
NOVO	8	Não	Não	Além Paraíba - MG

NOVO	9	Não	Não	Alto Araguaia - MT
NOVO	10	Não	Não	Alto Parnaíba - MA
NOVO	11	Não	Não	Alvorada do Norte - GO
NOVO	12	Não	Não	Amarante - PI
NOVO	13	Não	Não	Americana - SP
NOVO	14	Não	Não	Amparo - SP
NOVO	15	Não	Não	Angical do Piauí - PI
NOVO	16	Não	Não	Antônio Prado - RS
NOVO	17	Não	Não	Apucarana - PR
NOVO	18	Não	Não	Aquidauana - MS
NOVO	21	Não	Não	Araguari - MG
NOVO	23	Não	Não	Arapongas - PR
NOVO	24	Não	Não	Araranguá - SC
NOVO	26	Não	Não	Araras - SP
NOVO	28	Não	Não	Araruama - RJ
NOVO	30	Não	Não	Arenópolis - MT
NOVO	31	Não	Não	Artur Nogueira - SP
NOVO	36	Não	Não	Barra Bonita - SP
NOVO	43	Não	Não	Bento Gonçalves - RS
NOVO	44	Não	Não	Bicas - MG
NOVO	45	Não	Não	Birigui - SP
NOVO	46	Não	Não	Blumenau - SC
NOVO	48	Não	Não	Boituva - SP
NOVO	51	Não	Não	Bragança Paulista - SP
NOVO	52	Não	Não	Bragança - PA
NOVO	54	Não	Não	Brejões - BA
NOVO	55	Não	Não	Brusque - SC
NOVO	57	Não	Não	Cachoeira - BA
NOVO	60	Não	Não	Caiçara - PB
NOVO	61	Não	Não	Caiçara do Norte - RN
NOVO	63	Não	Não	Cambuí - MG
NOVO	64	Sim	Não	Cametá - PA
NOVO	71	Não	Não	Capão Bonito - SP
NOVO	72	Não	Não	Capão da Canoa - RS
NOVO	73	Não	Não	Capinzal - SC
NOVO	74	Não	Não	Capivari - SP
NOVO	75	Não	Não	Caraguatatuba - SP
NOVO	76	Não	Não	Carai - MG
NOVO	80	Sim	Não	Castanhal - PA
NOVO	82	Não	Não	Catanduva - SP
NOVO	85	Sim	Não	Caxias - MA
NOVO	86	Não	Não	Ceres - GO
NOVO	87	Não	Não	Cerquillo - SP
NOVO	89	Não	Não	Charqueadas - RS
NOVO	90	Não	Não	Cidreira - RS
NOVO	91	Não	Não	Codó - MA

NOVO	95	Não	Não	Cordeiro - RJ
NOVO	97	Não	Não	Cruzeiro - SP
NOVO	99	Não	Não	Cuité - PB
NOVO	101	Não	Não	Curvelo - MG
NOVO	104	Não	Não	Duas Estradas - PB
NOVO	105	Não	Não	Estreito - MA
NOVO	106	Não	Não	Eunápolis - BA
NOVO	114	Não	Não	Gastão Vidigal - SP
NOVO	115	Não	Não	Getúlio Vargas - RS
NOVO	116	Não	Não	Goianésia - GO
NOVO	119	Não	Não	Guarabira - PB
NOVO	120	Não	Não	Guarapari - ES
NOVO	123	Não	Não	Guariba - SP
NOVO	124	Não	Não	Igarapava - SP
NOVO	125	Não	Não	Iguape - SP
NOVO	127	Não	Não	Imbituba - SC
NOVO	129	Não	Não	Indiara - GO
NOVO	131	Não	Não	Quaraí - RS
NOVO	132	Não	Não	Assis Brasil - AC
NOVO	133	Não	Não	Bela Vista - MS
NOVO	134	Não	Não	Barra do Quaraí - RS
NOVO	135	Sim	Não	Bonfim - RR
NOVO	136	Não	Não	Coronel Sapucaia - MS
NOVO	137	Não	Não	Brasiléia - AC
NOVO	139	Não	Não	Dionísio Cerqueira - SC
NOVO	141	Não	Não	Garruchos - RS
NOVO	142	Não	Não	Guaíra - PR
NOVO	144	Não	Não	Itaqui - RS
NOVO	145	Não	Não	Jaguarão - RS
NOVO	147	Não	Não	Paranhos - MS
NOVO	148	Não	Não	Ponta Porã - MS
NOVO	149	Não	Não	Porto Mauá - RS
NOVO	150	Não	Não	Porto Vera Cruz - RS
NOVO	151	Não	Não	Porto Xavier - RS
NOVO	152	Não	Não	Sant'Ana do Livramento - RS
NOVO	153	Não	Não	Santo Antônio do Sudoeste - PR
NOVO	158	Não	Não	Itabira - MG
NOVO	159	Não	Não	Itabuna - BA
NOVO	161	Não	Não	Itajubá - MG
NOVO	162	Não	Não	Itambé - PE
NOVO	163	Não	Não	Itaocara - RJ
NOVO	164	Não	Não	Itapema - SC
NOVO	165	Não	Não	Itapetininga - SP
NOVO	166	Sim	Não	Itapipoca - CE
NOVO	167	Não	Não	Itapiranga - SC
NOVO	168	Não	Não	Itatiba - SP

NOVO	169	Não	Não	Itaúna - MG
NOVO	170	Não	Não	Itu - SP
NOVO	171	Sim	Não	Itumbiara - GO
NOVO	172	Não	Não	Ivoti - RS
NOVO	173	Não	Não	Jacaraú - PB
NOVO	174	Não	Não	Jaciara - MT
NOVO	175	Não	Não	Jales - SP
NOVO	177	Não	Não	Jandaia do Sul - PR
NOVO	178	Não	Não	Jaraguá do Sul - SC
NOVO	179	Não	Não	Jardim - MS
NOVO	180	Não	Não	Jaú - SP
NOVO	184	Não	Não	João Monlevade - MG
NOVO	190	Não	Não	Jundiaí - SP
NOVO	192	Não	Não	Lagoa Vermelha - RS
NOVO	193	Não	Não	Lajeado - RS
NOVO	195	Não	Não	Lavras - MG
NOVO	196	Não	Não	Leme - SP
NOVO	197	Não	Não	Limeira - SP
NOVO	200	Não	Não	Lorena - SP
NOVO	204	Não	Não	Mafra - SC
NOVO	205	Não	Não	Mamanguape - PB
NOVO	207	Não	Não	Manhuaçu - MG
NOVO	209	Sim	Não	Maragogi - AL
NOVO	212	Não	Não	Matinhos - PR
NOVO	213	Não	Não	Matozinhos - MG
NOVO	214	Não	Não	Maurilândia - GO
NOVO	215	Não	Não	Mendes - RJ
NOVO	216	Não	Não	Miracema do Tocantins - TO
NOVO	217	Não	Não	Mogi Guaçu - SP
NOVO	219	Não	Não	Morros - MA
NOVO	223	Não	Não	Nhamundá - AM
NOVO	224	Sim	Não	Nova Friburgo - RJ
NOVO	225	Não	Não	Nova Granada - SP
NOVO	226	Não	Não	Nova Londrina - PR
NOVO	227	Não	Não	Nova Serrana - MG
NOVO	228	Não	Não	Novo Hamburgo - RS
NOVO	229	Não	Não	Otacílio Costa - SC
NOVO	232	Não	Não	Paraguaçu Paulista - SP
NOVO	233	Não	Não	Paranacity - PR
NOVO	235	Não	Não	Paranavaí - PR
NOVO	236	Não	Não	Paraopeba - MG
NOVO	245	Não	Não	Paty do Alferes - RJ
NOVO	247	Não	Não	Paulo de Faria - SP
NOVO	248	Não	Não	Pedreiras - MA
NOVO	249	Não	Não	Pedro Afonso - TO
NOVO	250	Não	Não	Pedro Osório - RS

NOVO	252	Não	Não	Penedo - AL
NOVO	254	Não	Não	Petrópolis - RJ
NOVO	258	Não	Não	Pirapozinho - SP
NOVO	259	Não	Não	Ipira - SC
NOVO	260	Não	Não	Pitangui - MG
NOVO	261	Não	Não	Planura - MG
NOVO	266	Não	Não	Ponto Belo - ES
NOVO	271	Não	Não	Pouso Alto - MG
NOVO	272	Não	Não	Praia Grande - SC
NOVO	273	Não	Não	Presidente Juscelino - MA
NOVO	275	Não	Não	Presidente Venceslau - SP
NOVO	276	Não	Não	Propriá - SE
NOVO	277	Não	Não	Quirinópolis - GO
NOVO	285	Não	Não	Rio do Sul - SC
NOVO	288	Não	Não	Riolândia - SP
NOVO	290	Não	Não	Salvador do Sul - RS
NOVO	292	Não	Não	Santa Adélia - SP
NOVO	294	Não	Não	Santa Fé do Sul - SP
NOVO	300	Não	Não	Santo Antônio de Pádua - RJ
NOVO	302	Não	Não	Santo Antônio - RN
NOVO	303	Não	Não	São Bento do Sul - SC
NOVO	304	Não	Não	São Carlos - SC
NOVO	305	Não	Não	São Carlos - SP
NOVO	306	Não	Não	São Gabriel - RS
NOVO	307	Não	Não	São Geraldo do Araguaia - PA
NOVO	308	Não	Não	São João da Boa Vista - SP
NOVO	310	Não	Não	São José de Mipibu - RN
NOVO	313	Não	Não	São Lourenço - MG
NOVO	317	Não	Não	São Pedro - SP
NOVO	318	Não	Não	São Roque - SP
NOVO	319	Não	Não	São Sebastião da Amoreira - PR
NOVO	320	Não	Não	Sapiranga - RS
NOVO	322	Não	Não	Serranos - MG
NOVO	323	Não	Não	Sete Lagoas - MG
NOVO	326	Não	Não	Solânea - PB
NOVO	327	Não	Não	Sombrio - SC
NOVO	330	Não	Não	Taquara - RS
NOVO	331	Não	Não	Taquaruçu do Sul - RS
NOVO	332	Não	Não	Tatuí - SP
NOVO	333	Não	Não	Taubaté - SP
NOVO	337	Não	Não	Teresópolis - RJ
NOVO	338	Não	Não	Teutônia - RS
NOVO	339	Não	Não	Porto Franco - MA
NOVO	341	Não	Não	Baliza - GO
NOVO	342	Não	Não	Torres - RS
NOVO	343	Não	Não	Tramandaí - RS

NOVO	345	Não	Não	Três Rios - RJ
NOVO	346	Não	Não	Tubarão - SC
NOVO	347	Não	Não	Tupã - SP
NOVO	348	Não	Não	Tupi Paulista - SP
NOVO	355	Não	Não	Uruçuí - PI
NOVO	356	Não	Não	Valença - RJ
NOVO	357	Não	Não	Valparaíso - SP
NOVO	359	Não	Não	Venâncio Aires - RS
NOVO	360	Não	Não	Vera Cruz - BA
NOVO	361	Não	Não	Viçosa - MG
NOVO	364	Sim	Não	Volta Redonda - RJ
NOVO	365	Não	Não	Votuporanga - SP
NOVO	366	Não	Não	Abaré - BA
NOVO	367	Não	Não	Adustina - BA
NOVO	368	Não	Não	Afrânio - PE
NOVO	369	Não	Não	Água Boa - MG
NOVO	370	Não	Não	Águas Formosas - MG
NOVO	371	Não	Não	Alcântara - MA
NOVO	373	Não	Não	Almenara - MG
NOVO	376	Não	Não	Alto Alegre - RR
NOVO	377	Não	Não	Alto Garças - MT
NOVO	379	Não	Não	Alvorada D'Oeste - RO
NOVO	380	Não	Não	Amambai - MS
NOVO	382	Sim	Não	Amaturá - AM
NOVO	383	Não	Não	Anajás - PA
NOVO	384	Não	Não	Andirá - PR
NOVO	385	Não	Não	Anori - AM
NOVO	386	Não	Não	Aparecida do Taboado - MS
NOVO	387	Não	Não	Apuí - AM
NOVO	389	Não	Não	Araçuaí - MG
NOVO	390	Não	Não	Araguacema - TO
NOVO	391	Não	Não	Araguaçu - TO
NOVO	392	Não	Não	Araguatins - TO
NOVO	393	Não	Não	Arapoti - PR
NOVO	394	Não	Não	Aripuanã - MT
NOVO	396	Não	Não	Arraias - TO
NOVO	397	Não	Não	Arroio Grande - RS
NOVO	398	Não	Não	Aruanã - GO
NOVO	399	Não	Não	Autazes - AM
NOVO	401	Não	Não	Bambuí - MG
NOVO	404	Não	Não	Barra - BA
NOVO	406	Não	Não	Barreirinha - AM
NOVO	408	Não	Não	Bataguassu - MS
NOVO	409	Não	Não	Belém do São Francisco - PE
NOVO	410	Não	Não	Boa Viagem - CE
NOVO	413	Não	Não	Bonfinópolis de Minas - MG

NOVO	415	Não	Não	Boquira - BA
NOVO	417	Não	Não	Brasília de Minas - MG
NOVO	418	Não	Não	Brejinho de Nazaré - TO
NOVO	419	Não	Não	Brejo - MA
NOVO	421	Não	Não	Brotas de Macaúbas - BA
NOVO	422	Não	Não	Buriti Alegre - GO
NOVO	423	Não	Não	Buriticupu - MA
NOVO	424	Não	Não	Buritirama - BA
NOVO	425	Não	Não	Buritis - RO
NOVO	427	Não	Não	Caçapava do Sul - RS
NOVO	429	Não	Não	Cachoeira do Sul - RS
NOVO	431	Não	Não	Caculé - BA
NOVO	432	Não	Não	Caetité - BA
NOVO	433	Não	Não	Calçoene - AP
NOVO	434	Não	Não	Camapuã - MS
NOVO	435	Não	Não	Camocim - CE
NOVO	436	Não	Não	Campina Verde - MG
NOVO	437	Não	Não	Campo Alegre de Lourdes - BA
NOVO	438	Não	Não	Campo do Meio - MG
NOVO	439	Não	Não	Campo Maior - PI
NOVO	441	Não	Não	Campo Novo - RS
NOVO	442	Não	Não	Campo Novo do Parecis - MT
NOVO	443	Não	Não	Campos Belos - GO
NOVO	444	Não	Não	Campos Sales - CE
NOVO	445	Não	Não	Canaã dos Carajás - PA
NOVO	446	Não	Não	Canarana - BA
NOVO	447	Não	Não	Canavieiras - BA
NOVO	448	Não	Não	Canto do Buriti - PI
NOVO	449	Não	Não	Canudos - BA
NOVO	450	Não	Não	Capelinha - MG
NOVO	453	Não	Não	Caravelas - BA
NOVO	454	Não	Não	Carazinho - RS
NOVO	455	Não	Não	Carinhanha - BA
NOVO	456	Não	Não	Carutapera - MA
NOVO	457	Não	Não	Casa Nova - BA
NOVO	458	Não	Não	Cassilândia - MS
NOVO	459	Não	Não	Castelo do Piauí - PI
NOVO	460	Não	Não	Castro - PR
NOVO	461	Não	Não	Cavalcante - GO
NOVO	462	Não	Não	Centenário do Sul - PR
NOVO	463	Não	Não	Cerejeiras - RO
NOVO	464	Não	Não	Chapada dos Guimarães - MT
NOVO	465	Não	Não	Chapadão do Céu - GO
NOVO	466	Não	Não	Chapadão do Sul - MS
NOVO	467	Não	Não	Chaves - PA
NOVO	468	Não	Não	Cianorte - PR



NOVO	471	Não	Não	Cocos - BA
NOVO	472	Sim	Não	Codajás - AM
NOVO	473	Não	Não	Coelho Neto - MA
NOVO	474	Não	Não	Colíder - MT
NOVO	475	Não	Não	Colinas - MA
NOVO	476	Não	Não	Colinas do Tocantins - TO
NOVO	477	Não	Não	Conceição - PB
NOVO	479	Não	Não	Conceição do Mato Dentro - MG
NOVO	481	Não	Não	Cornélio Procópio - PR
NOVO	482	Sim	Não	Correia Pinto - SC
NOVO	484	Não	Não	Costa Marques - RO
NOVO	488	Não	Não	Cristalina - GO
NOVO	489	Não	Não	Crixás - GO
NOVO	490	Não	Não	Cruz - CE
NOVO	491	Não	Não	Cruz Alta - RS
NOVO	493	Não	Não	Cumarú do Norte - PA
NOVO	494	Não	Não	Curaçá - BA
NOVO	495	Não	Não	Curitiba - SC
NOVO	496	Não	Não	Cururupu - MA
NOVO	498	Não	Não	Diamantino - MT
NOVO	499	Não	Não	Dianópolis - TO
NOVO	501	Não	Não	Encruzilhada - BA
NOVO	502	Não	Não	Encruzilhada do Sul - RS
NOVO	504	Não	Não	Esperantina - PI
NOVO	505	Não	Não	Espumoso - RS
NOVO	506	Não	Não	Euclides da Cunha - BA
NOVO	507	Não	Não	Feijó - AC
NOVO	510	Não	Não	Formosa do Rio Preto - BA
NOVO	511	Não	Não	Fortaleza dos Nogueiras - MA
NOVO	513	Não	Não	Frederico Westphalen - RS
NOVO	514	Não	Não	Gentio do Ouro - BA
NOVO	515	Não	Não	Gilbués - PI
NOVO	516	Não	Não	Goiatins - TO
NOVO	517	Não	Não	Goioerê - PR
NOVO	518	Não	Não	Grajaú - MA
NOVO	519	Não	Não	Guadalupe - PI
NOVO	521	Não	Não	Guanhães - MG
NOVO	522	Não	Não	Guaporé - RS
NOVO	523	Não	Não	Guaratuba - PR
NOVO	524	Não	Não	Guaxupé - MG
NOVO	525	Não	Não	Gurupá - PA
NOVO	527	Não	Não	Horizontina - RS
NOVO	529	Não	Não	Iaciara - GO
NOVO	530	Não	Não	Ibaiti - PR
NOVO	531	Não	Não	Ibimirim - PE
NOVO	532	Não	Não	Ibirubá - RS

NOVO	534	Não	Não	Iguatemi - MS
NOVO	535	Não	Não	Ijuí - RS
NOVO	536	Não	Não	Ipameri - GO
NOVO	537	Não	Não	Iporá - GO
NOVO	542	Não	Não	Itapetinga - BA
NOVO	543	Não	Não	Itaporanga - PB
NOVO	544	Não	Não	Itarema - CE
NOVO	545	Não	Não	Ituaçu - BA
NOVO	549	Não	Não	Jaguarari - BA
NOVO	550	Não	Não	Jaguaribe - CE
NOVO	552	Não	Não	Januária - MG
NOVO	553	Não	Não	Japurá - AM
NOVO	554	Não	Não	Jardim de Angicos - RN
NOVO	556	Não	Não	Jequitinhonha - MG
NOVO	557	Sim	Não	Jijoca de Jericoacoara - CE
NOVO	560	Não	Não	Juazeirinho - PB
NOVO	562	Não	Não	Júlio de Castilhos - RS
NOVO	563	Não	Não	Juruá - AM
NOVO	564	Não	Não	Juruena - MT
NOVO	565	Não	Não	Juruti - PA
NOVO	566	Sim	Não	Jutaí - AM
NOVO	568	Não	Não	Lagoa da Confusão - TO
NOVO	570	Não	Não	Lizarda - TO
NOVO	571	Não	Não	Loanda - PR
NOVO	573	Não	Não	Luciara - MT
NOVO	574	Não	Não	Luís Eduardo Magalhães - BA
NOVO	575	Não	Não	Macaúbas - BA
NOVO	576	Não	Não	Machadinho D'Oeste - RO
NOVO	578	Não	Não	Manoel Ribas - PR
NOVO	579	Sim	Não	Maraã - AM
NOVO	580	Não	Não	Maracaju - MS
NOVO	581	Não	Não	Maracás - BA
NOVO	582	Não	Não	Marechal Cândido Rondon - PR
NOVO	583	Sim	Não	Marechal Thaumaturgo - AC
NOVO	584	Sim	Não	Mateiros - TO
NOVO	587	Não	Não	Medianeira - PR
NOVO	589	Não	Não	Minas Novas - MG
NOVO	590	Não	Não	Mineiros - GO
NOVO	591	Não	Não	Montalvânia - MG
NOVO	593	Não	Não	Monte Alegre de Goiás - GO
NOVO	594	Não	Não	Monte Santo - BA
NOVO	596	Não	Não	Montenegro - RS
NOVO	597	Não	Não	Morada Nova - CE
NOVO	598	Não	Não	Morada Nova de Minas - MG
NOVO	599	Não	Não	Morro do Chapéu - BA
NOVO	600	Não	Não	Mostardas - RS

NOVO	601	Não	Não	Mozarlândia - GO
NOVO	602	Não	Não	Mucugê - BA
NOVO	603	Não	Não	Mundo Novo - BA
NOVO	605	Não	Não	Nioaque - MS
NOVO	606	Não	Não	Niquelândia - GO
NOVO	607	Não	Não	Nonoai - RS
NOVO	608	Não	Não	Normandia - RR
NOVO	609	Sim	Não	Nova Andradina - MS
NOVO	611	Não	Não	Nova Prata - RS
NOVO	612	Não	Não	Nova Xavantina - MT
NOVO	613	Não	Não	Novo Aripuanã - AM
NOVO	614	Não	Não	Novo Horizonte - BA
NOVO	616	Não	Não	Óbidos - PA
NOVO	617	Não	Não	Oeiras - PI
NOVO	620	Não	Não	Ourolândia - BA
NOVO	621	Não	Não	Pacaraima - RR
NOVO	622	Não	Não	Palmeira das Missões - RS
NOVO	623	Não	Não	Palmeiras - BA
NOVO	624	Não	Não	Palotina - PR
NOVO	627	Não	Não	Paramirim - BA
NOVO	628	Não	Não	Paraná - TO
NOVO	629	Não	Não	Paranaíba - MS
NOVO	630	Não	Não	Paranatinga - MT
NOVO	631	Não	Não	Pau dos Ferros - RN
NOVO	634	Não	Não	Pedra Azul - MG
NOVO	635	Não	Não	Pesqueira - PE
NOVO	636	Não	Não	Piatã - BA
NOVO	639	Não	Não	Pimenteiras do Oeste - RO
NOVO	640	Não	Não	Piracanjuba - GO
NOVO	642	Não	Não	Pires do Rio - GO
NOVO	643	Não	Não	Piripiri - PI
NOVO	644	Não	Não	Piritiba - BA
NOVO	646	Não	Não	Poconé - MT
NOVO	649	Não	Não	Portel - PA
NOVO	650	Não	Não	Porto de Moz - PA
NOVO	651	Não	Não	Porto dos Gaúchos - MT
NOVO	652	Não	Não	Porto Murtinho - MS
NOVO	653	Sim	Não	Porto Walter - AC
NOVO	654	Não	Não	Posse - GO
NOVO	655	Não	Não	Prainha - PA
NOVO	656	Não	Não	Prata - MG
NOVO	657	Não	Não	Presidente Dutra - MA
NOVO	658	Não	Não	Presidente Epitácio - SP
NOVO	659	Não	Não	Presidente Figueiredo - AM
NOVO	660	Não	Não	Primavera do Leste - MT
NOVO	661	Não	Não	Realeza - PR

NOVO	664	Não	Não	Remanso - BA
NOVO	665	Não	Não	Ribeiro Gonçalves - PI
NOVO	666	Não	Não	Rio Brilhante - MS
NOVO	667	Não	Não	Rio de Contas - BA
NOVO	668	Não	Não	Rio Paranaíba - MG
NOVO	669	Não	Não	Rio Pardo de Minas - MG
NOVO	670	Não	Não	Rolim de Moura - RO
NOVO	671	Sim	Não	Rorainópolis - RR
NOVO	672	Não	Não	Rosário do Sul - RS
NOVO	674	Não	Não	Russas - CE
NOVO	677	Não	Não	Salinópolis - PA
NOVO	678	Não	Não	Salto do Jacuí - RS
NOVO	679	Não	Não	Santa Cruz do Arari - PA
NOVO	680	Não	Não	Santa Helena de Goiás - GO
NOVO	683	Não	Não	Santa Maria da Boa Vista - PE
NOVO	684	Não	Não	Santa Maria do Suaçuí - MG
NOVO	685	Não	Não	Santa Quitéria - CE
NOVO	686	Não	Não	Santa Rita de Cássia - BA
NOVO	688	Não	Não	Santa Vitória - MG
NOVO	690	Não	Não	Santana do Ipanema - AL
NOVO	691	Não	Não	Santiago - RS
NOVO	692	Não	Não	Santo Antônio do Içá - AM
NOVO	693	Não	Não	São Benedito - CE
NOVO	694	Não	Não	São Domingos - GO
NOVO	697	Não	Não	São Francisco - MG
NOVO	698	Não	Não	São Francisco do Sul - SC
NOVO	699	Não	Não	São Gabriel - BA
NOVO	701	Não	Não	São Gabriel do Oeste - MS
NOVO	702	Não	Não	São Gotardo - MG
NOVO	703	Não	Não	São João do Piauí - PI
NOVO	704	Não	Não	São João do Rio do Peixe - PB
NOVO	705	Sim	Não	São João dos Patos - MA
NOVO	707	Não	Não	São Joaquim da Barra - SP
NOVO	708	Não	Não	São José do Jacuípe - BA
NOVO	709	Não	Não	São Lourenço do Sul - RS
NOVO	710	Não	Não	São Luís de Montes Belos - GO
NOVO	711	Não	Não	São Luiz Gonzaga - RS
NOVO	712	Não	Não	São Manuel - SP
NOVO	713	Não	Não	São Miguel do Araguaia - GO
NOVO	717	Não	Não	São Sepé - RS
NOVO	719	Não	Não	Senador Pompeu - CE
NOVO	721	Não	Não	Serro - MG
NOVO	722	Não	Não	Sertanópolis - PR
NOVO	723	Não	Não	Siqueira Campos - PR
NOVO	724	Não	Não	Sítio Novo - MA
NOVO	725	Não	Não	Sorriso - MT

NOVO	726	Não	Não	Souto Soares - BA
NOVO	727	Não	Não	Taguatinga - TO
NOVO	728	Não	Não	Tamboril - CE
NOVO	730	Não	Não	Tanque Novo - BA
NOVO	732	Não	Não	Tartarugalzinho - AP
NOVO	733	Não	Não	Tauá - CE
NOVO	736	Não	Não	Terra Santa - PA
NOVO	737	Não	Não	Tomé-Açu - PA
NOVO	738	Não	Não	Três Marias - MG
NOVO	739	Não	Não	Tucano - BA
NOVO	740	Não	Não	Turiação - MA
NOVO	741	Não	Não	Turmalina - MG
NOVO	742	Sim	Não	Uarini - AM
NOVO	743	Não	Não	Uauá - BA
NOVO	745	Não	Não	Urbano Santos - MA
NOVO	746	Não	Não	Uruaçu - GO
NOVO	747	Não	Não	Uruará - PA
NOVO	748	Não	Não	Urucará - AM
NOVO	749	Não	Não	Utinga - BA
NOVO	750	Não	Não	Vacaria - RS
NOVO	751	Não	Não	Valença do Piauí - PI
NOVO	752	Não	Não	Valente - BA
NOVO	753	Não	Não	Várzea Grande - PI
NOVO	754	Não	Não	Veranópolis - RS
NOVO	755	Não	Não	Viçosa - RN
NOVO	756	Não	Não	Videira - SC
NOVO	757	Não	Não	Vila Bela da Santíssima Trindade - MT
NOVO	760	Não	Não	Virgem da Lapa - MG
NOVO	761	Não	Não	Viseu - PA
NOVO	762	Não	Não	Xanxerê - SC
NOVO	763	Não	Não	Xinguara - PA
NOVO	764	Não	Não	Xique-Xique - BA
NOVO	765	Não	Não	Zé Doca - MA
NOVO	766	Não	Não	Ilhabela - SP
NOVO	768	Não	Não	Guararapes - SP
NOVO	769	Não	Não	Nanuque - MG
NOVO	771	Não	Não	Pompéu - MG
NOVO	772	Não	Não	Santa Maria - RN
NOVO	783	Não	Não	Buritis - MG
NOVO	784	Não	Não	Canarana - MT
NOVO	785	Não	Não	Gaúcha do Norte - MT
NOVO	786	Não	Não	Santa Terezinha - MT

*Página intencionalmente deixada em branco.*

## 10 Anexo III: Modelo de Mínimos Quadrados Ordinários.

Modelo MQO, usando as observações 1-266  
 Variável dependente: tresanosatr  
 Erros padrão robustos à heteroscedasticidade, variante HC1

	<i>Coefficiente</i>	<i>Erro Padrão</i>	<i>razão-t</i>	<i>p-valor</i>	
const	1,40777	0,163414	8,6147	<0,0001	***
regiao	-0,0628318	0,027409	-2,2924	0,0229	**
distsitemico	0,000723548	0,000198062	3,6531	0,0003	***
popza2004	-2,83812e-06	8,60677e-06	-0,3298	0,7419	
popza2005	0,000164373	6,37887e-05	2,5768	0,0107	**
popza2006	-0,000135622	5,36617e-05	-2,5273	0,0123	**
popza2007	1,95863e-05	7,19518e-06	2,7221	0,0071	***
popza2008	-0,00023602	0,000103543	-2,2794	0,0237	**
popza2009	0,000195525	9,50291e-05	2,0575	0,0410	**
popza2010	-0,000140015	6,77867e-05	-2,0655	0,0402	**
popza2011	0,000136155	7,01902e-05	1,9398	0,0538	*
popza2012	-4,22642e-07	1,41219e-05	-0,0299	0,9762	
popza2013	8,9976e-06	1,50799e-05	0,5967	0,5514	
popza2014	3,34901e-07	2,54115e-07	1,3179	0,1891	
popza2015	-9,84965e-06	1,32269e-05	-0,7447	0,4574	
pibpcza2004	0,00411244	0,0210922	0,1950	0,8456	
pibpcza2005	-0,0586529	0,0348795	-1,6816	0,0942	*
pibpcza2006	0,015863	0,0249502	0,6358	0,5257	
pibpcza2007	0,00502012	0,0175682	0,2858	0,7754	
pibpcza2008	0,00427583	0,0137221	0,3116	0,7557	
pibpcza2009	0,0183484	0,0112141	1,6362	0,1034	
pibpcza2010	0,0265139	0,0170098	1,5587	0,1207	
pibpcza2011	-0,010937	0,0190091	-0,5754	0,5657	
pibpcza2012	0,00684493	0,0219019	0,3125	0,7550	
pibpcza2013	-0,00818664	0,0167677	-0,4882	0,6259	
desptot2004pc	0,000715857	0,000282372	2,5352	0,0120	**
desptot2005pc	-7,17181e-05	0,000229563	-0,3124	0,7551	
desptot2006pc	-0,00040433	0,000279051	-1,4489	0,1490	
desptot2007pc	0,000235974	0,0003317	0,7114	0,4777	
desptot2008pc	0,000207973	0,000293127	0,7095	0,4789	
desptot2009pc	-0,000237195	0,000250291	-0,9477	0,3445	
desptot2010pc	-0,00040526	0,000222439	-1,8219	0,0700	*
desptot2011pc	-5,7448e-05	0,000190407	-0,3017	0,7632	
desptot2012pc	0,000106476	0,000117112	0,9092	0,3644	
desptot2013pc	-9,44821e-06	9,60676e-05	-0,0983	0,9218	
desptot2014pc	1,40821e-05	1,87302e-05	0,7518	0,4530	
desptot2015pc	-2,48618e-05	6,31203e-05	-0,3939	0,6941	
desptransp2004pc	-0,00383365	0,00220452	-1,7390	0,0836	*
desptransp2005pc	0,000764277	0,00211826	0,3608	0,7186	
desptransp2006pc	-0,00215722	0,00120532	-1,7897	0,0750	*
desptransp2007pc	0,000837355	0,00154698	0,5413	0,5889	
desptransp2008pc	-0,000312605	0,00112132	-0,2788	0,7807	
desptransp2009pc	-0,000222426	0,000546493	-0,4070	0,6844	

desptransp2010pc	0,00341132	0,000980264	3,4800	0,0006	***
desptransp2011pc	-0,00223471	0,00158774	-1,4075	0,1609	
desptransp2012pc	-0,000958923	0,00125961	-0,7613	0,4474	
desptransp2013pc	0,00186585	0,00142313	1,3111	0,1914	
desptransp2014pc	0,00115204	0,00105148	1,0956	0,2746	
desptransp2015pc	-0,0016155	0,00114597	-1,4097	0,1602	
despaero2004pc	-0,000715672	0,023639	-0,0303	0,9759	
despaero2005pc	0,00387706	0,0165592	0,2341	0,8151	
despaero2006pc	-0,118243	0,0437306	-2,7039	0,0075	***
despaero2007pc	0,0557877	0,0439183	1,2703	0,2055	
despaero2008pc	-0,0328806	0,0201767	-1,6296	0,1048	
despaero2009pc	0,0105653	0,0666685	0,1585	0,8742	
despaero2010pc	0,000948831	0,0110349	0,0860	0,9316	
despaero2011pc	0,0808022	0,0302869	2,6679	0,0083	***
despaero2012pc	0,00755698	0,00819515	0,9221	0,3576	
despaero2013pc	0,00431324	0,00519881	0,8297	0,4077	
despaero2014pc	-0,0306871	0,0140709	-2,1809	0,0304	**
despaero2015pc	-0,00639487	0,0375074	-0,1705	0,8648	
qt_empregopc	15,0147	10,4433	1,4377	0,1521	
qt_estabelecimentopc	-1392,92	871,603	-1,5981	0,1116	
qt_visita_nacional_est	-0,0473003	0,192127	-0,2462	0,8058	
imadapc					
qt_visita_internacional	4,54335	1,7922	2,5351	0,0120	**
_estimadpc					
mturA	0,189966	0,16337	1,1628	0,2463	
mturB	0,0869961	0,0669632	1,2992	0,1954	
mturC	-0,0528713	0,0368402	-1,4352	0,1528	
mturD	0,036809	0,0176277	2,0881	0,0381	**
mturE	0,0205591	0,0366386	0,5611	0,5753	
Média var. dependente	1,413534	D.P. var. dependente		0,493395	
Soma resíd. quadrados	30,33868	E.P. da regressão		0,393433	
R-quadrado	0,529715	R-quadrado ajustado		0,364156	
F(69, 196)	14,35914	P-valor(F)		1,17e-48	