

Ricardo César Gonçalves Sant'Ana

Moisés Lima Dutra

Guilherme Ataíde Dias

Organizadores

WIDaT 2018

II WORKSHOP DE INFORMAÇÃO,
DADOS E TECNOLOGIA

ANAIS
WIDaT 2018

Organização do WIDaT 2018

- **Organização Geral:**

Guilherme Ataíde Dias (PPGCI-UFPB) - Coordenador geral do evento
Moisés Lima Dutra (PPGCIN-UFSC) - Vice-coordenador

- **Coordenador da Comissão Científica:**

Ricardo César Gonçalves Sant'Ana (PPGCI-UNESP)

- **Comissão científica**

Adilson Luiz Pinto (PPGCIN-UFSC)
Ana Alice Baptista (Universidade do Minho, Portugal)
Ana Carolina Simionato (PPGCI-UFSCar)
Angela Maria Grossi de Carvalho (PPGCI-UNESP)
Bernardina Maria Juvenal Freire de Oliveira (PPGCI-UFPB)
Cristian Berrío-Zapata (PPGCI-UFPA)
Dalton Lopes Martins (FCI-UnB)
Denysson Axel Ribeiro Mota (PPGB-UFCA)
Douglas Dyllon Jeronimo de Macedo (PPGCIN-UFSC)
Ed Porto Bezerra (PPGI-UFPB)
Edgar Bisset Alvarez (PPGCIN-UFSC)
Edna Gusmão de Goés Brennand (MPGOA-UFPB)
Edna Gomes Pinheiro (DCI-UFPB)
Elaine Parra Affonso (FATEC-SP)
Elvis Fusco (UNIVEM-Marília)
Enrique Muriel Torrado (PPGCIN-UFSC)
Evandro de Barros Costa (IC-UFAL)
Fábio Paraguaçu (IC-UFAL)
Fernando de Assis Rodrigues (PPGCI-UNESP)
Gustavo Medeiros de Araújo (PPGCIN-UFSC)
Henry Pôncio Cruz de Oliveira (PPGCI-UFPB)
Joana Coeli Ribeiro Garcia (PPGCI-UFPB)
José Eduardo Santarém Segundo (USP-FFCLRP)
Leonardo Castro Botega (UNIVEM-Marília)
Luana Farias Sales Marques (PPGCI-IBICT-UFRJ)
Marckson Roberto Ferreira de Sousa (PPGCI-UFPB)
Luís Fernando Sayão (CNEN)
Marcelo Morandini (EACH-USP)
Márcio Matias (PPGCIN-UFSC)
Marcos Mucheroni (CBD-USP)
Marynice de Medeiros Matos Autran (PPGCI-UFPB)

Maurício Barcellos Almeida (PPGGOC-UFMG)
Moisés Lima Dutra (PPGCIN-UFSC)
Plácida Leopoldina V. da Costa Santos (PPGCI-UNESP)
Pedro Luiz Pizzigatti Corrêa (POLI-USP)
Renata Baracho (PPGGOC-UFMG)
Ricardo César Gonçalves Sant'Ana (PPGCI-UNESP)
Robson Rodrigues Lemos (UFSC-Araranguá)
Rogério Ramalho (PPGCI-UFSCar)
Ryan Ribeiro de Azevedo (UFRPE-UAG)
Sandra de Albuquerque Siebra (PPGCI-UFPE)
Sandro Rautenberg (DECOMP-UNICENTRO)
Silvana Aparecida Borsetti G. Vidotti (PPGCI-UNESP)
Virginia Bentes Pinto (PPGCI-UFC)
Wagner Junqueira de Araújo (PPGCI-UFPB)
Zaira Regina Zafalon (PPGCI-UFSCar)

- **Coordenador do Cerimonial:**

André Luiz Dias de França (PPGCI-UFPB)

- **Coordenador da Equipe Técnica Local:**

Laerte Pereira da Silva Júnior (CCHLA-UFPB)

- **Equipe Técnica Local:**

Adriana Alves Rodrigues (PPGCI-UFPB)
Antonio Felipe dos Santos (MPGOA-UFPB)
Débora Gomes de Araújo (PPGCI-UFPB)
Pedro Augusto de Lima Barroso (PPGCI-UFPB)
Pollianna Marys de Souza e Silva (PPGCI-UFPB)
Renata Lemos dos Anjos (PPGCI-UFPB)

PREVISÕES DE AUSÊNCIA EM ATENDIMENTO ESPECIALIZADO EM UNIDADES DE SAÚDE PÚBLICA DE FLORIANÓPOLIS

ABSENCE FORECAST IN SPECIALIZED CARE IN PUBLIC HEALTH UNITS OF FLORIANÓPOLIS

Milena Teixeira¹, Leandro Pereira Garcia^(2,3), Ingo Ramos⁽¹⁾, Lucas Alexandre Pedebôs⁽³⁾, Douglas Macedo⁽¹⁾, Gustavo Medeiros de Araujo⁽¹⁾

(1) Universidade Federal de Santa Catarina, R. Eng. Agrônomo Andrei Cristian Ferreira, s/n - Trindade, Florianópolis - SC, 88040-900, milena.correia@grad.ufsc.br, ingoramos12@gmail.com
douglas.macedo@ufsc.br, gustavo.araujo@ufsc.br.

(2) Universidade do Sul de Santa Catarina, Avenida Pedra Branca, 25, Cidade Universitária Pedra Branca, CEP 88137-272, Palhoça, SC lpgarcia18@gmail.com

(3) Secretaria Municipal de Saúde de Florianópolis, Avenida Professor Henrique da Silva Fontes, 6100, Florianópolis, CEP 88036-700, Florianópolis, SC geinfo.sms@gmail.com

Resumo:

O absenteísmo em serviços de atendimento de saúde de média e alta complexidade do Sistema Único de Saúde está associado ao desperdício e ineficiência, uma vez que estes procedimentos envolvem grande densidade tecnológica. Nesse sentido, esse estudo busca desenvolver um modelo que possa identificar pacientes com grande probabilidade de falta em consultas especializadas, como base em aprendizagem de máquina. Para realização dessa pesquisa foram usados, como estudo-piloto, dados de atendimentos nos Centros de Especialidades Odontológicas da Secretaria Municipal de Saúde de Florianópolis entre os anos de 2016 a junho de 2018.

Palavras-chave: Absenteísmo. Consultas médicas especializadas. Faltas preditivas. Unidades de Saúde Pública de Florianópolis.

Abstract:

Absenteeism in health care services of medium and high complexity of the Unified Health System is associated with waste and inefficiency, since this procedure involves a high technological density. In this sense, this study seeks to develop a model that can identify patients with a high probability of lack in specialized consultations, as a basis in machine learning. In order to carry out this research, we used as a pilot study data, between 2016 and June 2018, from the Dental Specialties Centers of the Municipal Health Department of Florianópolis.

Keywords: Absenteeism. Specialized medical consultations. Predictive absence. Public Health Units of Florianópolis.

I INTRODUÇÃO

O sistema único de saúde (SUS) tem convivido com problemas de financiamento desde que foi criado. Iniciativas foram tomadas para minorar, pelo menos em parte, o problema. A aprovação da CPMF¹, em 1996, e da Emenda Constitucional 29², em 2000, foram as mais importantes” (Piola et al., 2013) regulamentando e potencializando a atenção ao sistema único de saúde no País.

Contudo, é preciso ressaltar que os recursos destinados à saúde no Brasil, embora não ideais, não diferem de alguns países em desenvolvimento que lograram obter melhor assistência à população do que aquela que hoje, em média, é fornecida aos brasileiros (Piola et al., 2013).

Em épocas de escassez de recursos, duas ações são fundamentais para que o SUS consiga ofertar melhores serviços: 1) forçar um maior repasse para a saúde pública; e 2) melhorar os gastos em saúde, aumentando a eficiência do sistema como um todo.

Ações de advocacy para o aumento do repasse têm surgido de entidades como a ABRASCO. O aumento da eficiência, por sua vez, pode ser alcançado com o fortalecimento da atenção primária e com a redução de desperdícios na média e na alta complexidade. A atenção primária à saúde é o atendimento inicial, cujo principal objetivo é a prevenção de doenças, tratamento de agravos mais prevalentes e o direcionamento de casos menos prevalentes e mais graves para outros níveis de complexidade (SALDIVA; VERAS, 2018). Dessa forma, consegue racionalizar a utilização dos recursos em saúde, fazendo com que a entrada dos indivíduos no sistema de saúde ocorra em seu ponto mais seguro e menos dispendioso, as Unidades Básicas de Saúde (UBS). Uma boa atenção primária, deve oferecer acesso amplo e oportuno a população, fazendo com que as pessoas consigam rapidamente sanar suas principais necessidades em saúde.

Porém, a atenção primária, isoladamente, não é capaz de oferecer um cuidado integral ao cidadão, sendo, muitas vezes necessária sua articulação com serviços de saúde de média e alta complexidade. Esta articulação deve ser feita por meio de regulação, fazendo com que casos mais graves sejam atendidos de forma prioritária.

Nesse sentido, um grande problema em termos de eficiência são os desperdícios ocasionados por faltas em consultas médicas e odontológicas nos níveis mais altos de complexidade. Diferente do que ocorre na atenção primária, onde o paciente deve, preferencialmente, ser escutado no dia em que procura a UBS, reduzindo muito o número de faltas, o acesso à média e alta complexidade dá-se por encaminhamento, com marcação antecipada, fazendo com que os pacientes precisem aguardar em fila para a consulta. Desta forma, o número de faltas nestes níveis de complexidade tendem a ser maior que na atenção primária.

Em Florianópolis, por exemplo, o percentual de faltas é de 3% para consultas médicas e 12% para consultas odontológicas na atenção primária (FLORIANÓPOLIS, 2018a). Na média complexidade municipal, a mediana de faltas é de 26% (FLORIANÓPLIS, 2018b).

O absenteísmo em serviços de atendimento de média e alta complexidade vinculam-se a um alto custo, uma vez que estes são procedimento geralmente demorados e com o envolvimento de uma maior densidade tecnológica. (CONASS, 2007)

1 CPMF Disponível em: <<https://www12.senado.leg.br/noticias/entenda-o-assunto/cpmf>>. Acesso em: 19 set. 2018

2 Regulamentação da Emenda Constitucional Nº 29, De 2000. Disponível em: <<http://www2.camara.leg.br/orcamento-da-uniao/estudos/2012/nt14.pdf>>. Acesso em: 19 set. 2018

Neste sentido, o presente trabalho visa a contribuir com a redução do desperdício no SUS, buscando produzir um modelo que possa identificar pacientes com grande probabilidade de falta às consultas especializadas, baseado em algoritmos de aprendizagem de máquina.

2 OBJETIVO

Construir um modelo de predição de possíveis faltantes à consultas especializadas no Sistema Único de Saúde.

3 PROCEDIMENTOS METODOLÓGICOS

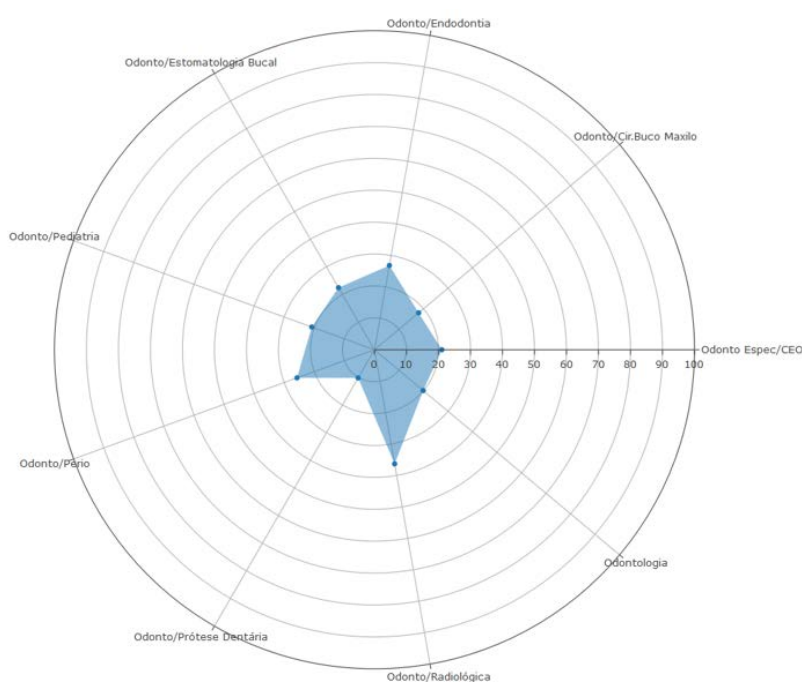
O desenvolvimento do modelo, partiu de estudo-piloto com dados secundários fornecidos pela Gerência de Inteligência e Informação da Secretaria Municipal de Saúde de Florianópolis.

Os dados correspondem a consultas agendadas no período de 1 de janeiro de 2016 a 30 de junho de 2018, realizadas nos dois Centros de Especialidades Odontológicas - CEOs do município: CEO Centro e CEO Continente. Estes serviços atendem nove especialidades odontológicas, com agendamento a partir de encaminhamentos de dentistas da atenção primária do município.

Para cada uma das especialidades, foram extraídos dos bancos: a data e o período do dia do atendimento, a Unidade de Referência (Centro de Saúde que realizou o encaminhamento), escolaridade, idade e sexo do paciente, além da indicação de primeira consulta ou retorno e de falta ou comparecimento.

Baseando-se nas especialidades com maior porcentagem de faltas neste período, foram selecionadas apenas três delas para o estudo: endodontia, periodontia e radiológica.

Gráfico I: Percentual de faltas por especialidade nos anos 2016 /2017.



Para a etapa de predição de faltas nos atendimentos, foram aplicados os modelos de aprendizagem de máquina nos bancos de dados das especialidades. Por se tratar de um problema de classificação binária, que neste caso, será classificar se o usuário irá faltar ou não, utilizamos a matriz de confusão para aferir o preditor. Para o problema em questão, utilizamos como métrica de avaliação do preditor, a divisão do número de verdadeiros positivos pela soma dos falsos positivos e falsos negativos ($VP / FP + FN$). O propósito em usar essa métrica foi selecionar o algoritmo que maximiza a relação entre acertos e erros nas classificações dos faltantes.

É importante salientar, que segundo a Resolução N° 510, de 7 de abril de 2016 (CNS, 2016), do Conselho Nacional de Saúde do Brasil, pesquisas com bancos de dados, cujas informações são agregadas, sem possibilidade de identificação individual não precisam ser submetidas a Comitês de Ética. Desta forma, não houve submissão deste projeto a um Comitê de Ética em pesquisa.

3.1 Tratamento de Dados

Os dados entre os anos de 2016 e 2017 foram usados para o treinamento dos modelos. Os modelos treinados e validados utilizando o método K-fold, com número de *folds* igual a 10. Além disso, os conjuntos de dados foram balanceados para aumentar a classe minoritária. Para os testes dos algoritmos, utilizou-se dados do ano de 2018 para verificar a capacidade de predição do melhor modelo escolhido.

3.2 Algoritmos de Aprendizado de Máquina

Utilizamos os seguintes algoritmos de aprendizagem de máquina para realização da classificação dos pacientes como faltantes ou não faltantes: Generalized Linear Model, Random Forest, k-Nearest Neighbors, Neural Network, Stacked AutoEncoder Deep Neural Network, eXtreme Gradient Boosting, Stochastic Gradient Boosting, AdaBoost Classification Trees. A implementação destes algoritmos deu-se por meio do pacote Caret do software R (R, 2018).

Algoritmos foram parametrizados da seguinte forma:

Figura 1: Parâmetros para a função de controle dos modelos.

Função "trainControl": controle que deve ser utilizado em todos os modelos para que sejam comparáveis. No código é atribuída a uma variável de nome "myControl", para poder ser utilizada da função de treino.		
Parâmetro	Valor	Descrição
method	"repeatedcv"	Dividir o banco em treinamento e teste diversas vezes. Reamostragem.
number	10	Define o número de folds (dobras) para o método de reamostragem.
summaryFunction	twoClassSummary	Calcular métricas de desempenho.
classProbs	TRUE	Computa as probabilidades dos modelos.
verboseIter	FALSE	Não imprime o log de treinamento.
savePredictions	TRUE	Quanto das predições deve ser salvo. (Tudo)
returnResamp	"all"	Quantidade das métricas que devem ser salvas. (Tudo)
sampling	"up"	Aumenta aleatoriamente a classe minoritária. Balanceamento dos dados.
allowParallel	TRUE	Permite a computação paralela

Figura 2: Parâmetros para a função de treino dos modelos.

Função "train": configuração de parâmetros para os modelos de classificação e regressão e calculo de desempenho.		
Parâmetro	Valor	Descrição
y	FALTA	Variável preditora
data	train_data	Banco no qual serão aplicados os modelos.
preProcess	c("center", "scale")	Centraliza e escala os dados de predição.
metric	ROC	Métrica para comparação dos modelos.
method	modelos[index]	Lista com índice, retorna um string com o nome do modelo.
trControl	myControl	Aplica o controle (trainControl).

Todas as análises foram executadas com o uso dos softwares estatísticos R (R, 2018) com a IDE R-Studio (RSTUDIO, 2018).

3.3 Aplicação de Ciência de Dados à previsão de Faltas

Na tentativa de encontrar padrões nos dados referentes aos anos de 2016 a junho de 2018 de atendimentos nos Centros de Especialidade Odontológicas para prever usuários que faltam em consultas, foram feitas análises e descrições de algumas das variáveis disponíveis, sendo elas, dia da semana e período dos atendimentos, e também escolaridade e sexo dos pacientes.

4 RESULTADOS

Primeiramente, realizamos um trabalho de análise descritiva para verificar a relevância das características em cada especialidade. Posteriormente, realizamos a avaliação dos diversos modelos de aprendizado automático para verificar a eficiência na predição das faltas.

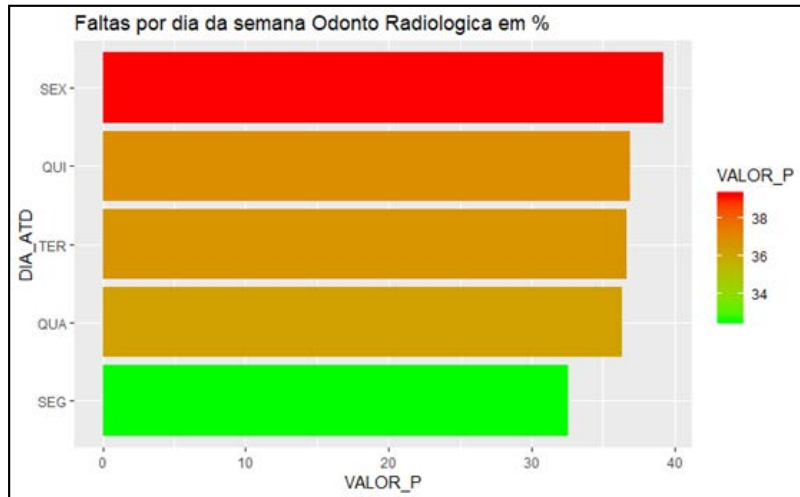
4.1 Análise Estatística Descritiva

Com as especialidades já escolhidas, obtiveram-se dados de 18.503 consultas, tendo registradas 5.944 faltas para o conjunto.

Sendo que ao separar os dados por especialidade, periodontia 2.531 atendimentos e 652 faltas, endodontia com 5.360 registros e 1.438 faltantes e por fim, radiológica com a maior quantidade de atendimentos, 10.612, assim como a maior quantidade de faltas, com 3.854.

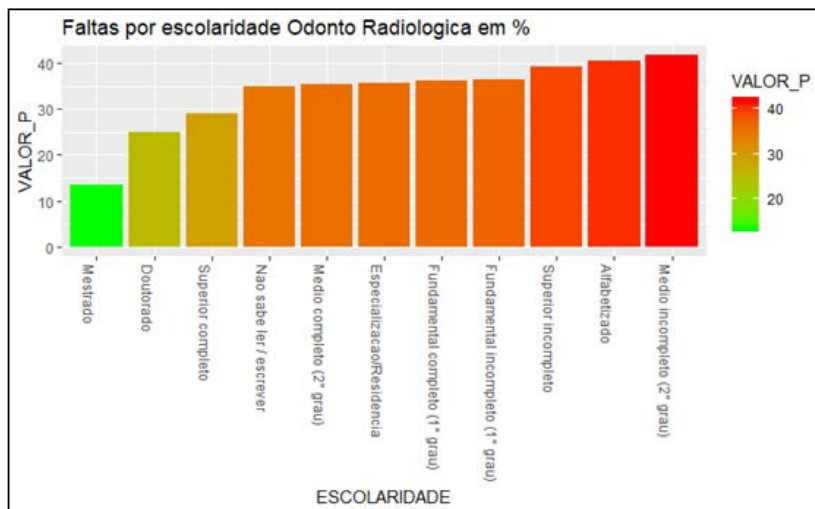
Quanto às variáveis usadas, odontologia radiológica e odontologia periodontia apresentaram sexta e sábado respectivamente como dias da semana com maior percentual de faltas, enquanto segunda-feira foi o dia com menos faltantes. Na odontologia radiológica, 40% de todos os pacientes que marcaram na sexta-feira, faltaram. A endodontia foi a única das três que não apresentou sexta ou sábado como dia da semana com mais faltas, mas sim terça-feira.

Gráfico 1: Maior percentual de faltas em um dia da semana. (Radiológica)



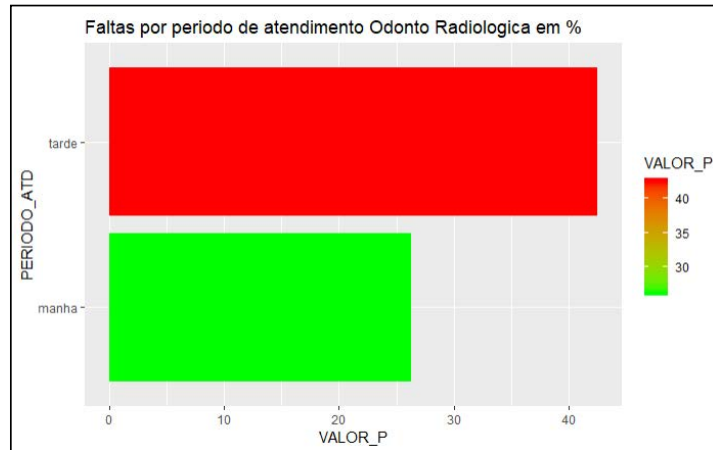
A escolaridade segue um padrão, para as três especialidades. Foi possível perceber que o de faltantes diminui com relação a pacientes que tenham ensino superior, mestrado ou doutorado. Da mesma forma, que ensino médio incompleto aparece como grau de escolaridade com o maior percentual, assim como, não alfabetizados, fundamental (completo e incompleto).

Gráfico 2: Percentual de faltas por escolaridade odontologia radiológica.



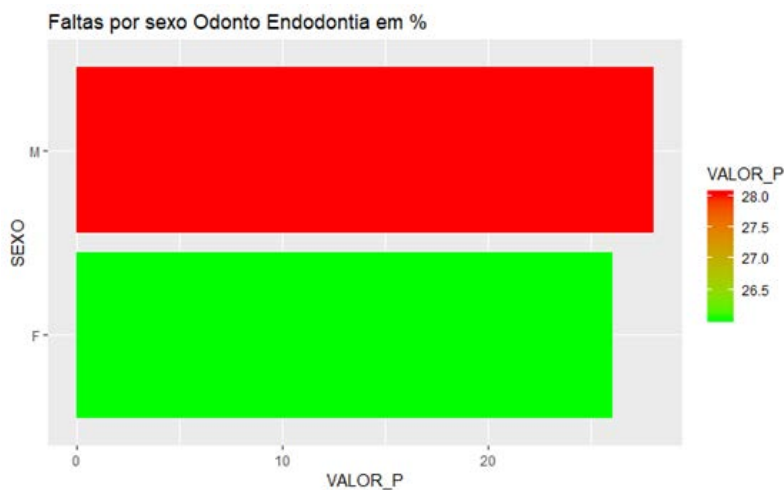
O percentual de faltas não teve muita variação na especialidade periodontia de acordo com o período de atendimento, sendo 25,8% dos pacientes que tinha consulta agendada para tarde faltará e 25,5% daqueles com consultas agendadas pela manhã. Padrão semelhante ao encontrado na Endodontia, com 28% de faltas a tarde e 25% no período matutino. A especialidade com maior discrepância de faltas por período de atendimento foi radiológica, que variou mais de 10% entre os períodos, sendo que de tarde aconteceram mais faltas.

Gráfico 3: Percentual de faltas por período em radiologia.



Não se observaram grandes variações entre o percentual de faltas entre homens e mulheres. Na Periodontia o percentual de faltas entre as mulheres tende a ser ligeiramente maior. Na Radiologia e na Endodontia homens tendem a faltar mais. Em todas as especialidades, porém, a diferença do percentual de faltas entre os sexos não passa de dois por cento.

Gráfico 4: Percentual de faltas por sexo em Endodontia



4.2 Aprendizado Automático

Em relação à seleção do modelo para predição dos faltantes, para cada uma das especialidades, usamos os dados entre 2016 a 2017 para treino e os dados de 2018 para teste. Quanto às matrizes de confusão do processo de treinamento, nenhum dos resultados apresentou especificidade boa, não passando de 0,4, o que também aconteceu para a sensibilidade, chegando à 0,739. A média entre os resultados dos três modelos com melhor desempenho de acurácia foi de aproximadamente 0,588. Entretanto, a acurácia não foi usada como método de avaliação dos algoritmos, uma vez que por se tratarem de bancos com dados não balanceados. Os modelos aplicados ao banco de dados

não balanceados predizem de forma discrepante apenas um valor da variável preditora - neste caso, que o usuário não faltará, enquanto prediz pouco que irá faltar, tornando essa métrica inválida para a avaliação. Mesmo com o balanceamento dos dados no treinamento, os resultados não melhoram na predição, a especificidade se mantém baixa e a taxa de não informação foi maior do que a acurácia para todos os resultados. Para todos os bancos, o modelo de aprendizado automático que teve o melhor desempenho foi o adaboost, alcançando sensibilidade de 0,761 e especificidade de 0,273 no banco Odonto_Endodontia (Apêndice A). Para o banco Odonto_Radiologica (Apêndice B) a sensibilidade foi a menor, chegando apenas em 0,636, mas a especificidade foi a melhor com 0,423. Já para o banco da especialidade Odonto_Perio (Apêndice C), a sensibilidade foi a maior, 0,802, mas a especificidade foi também baixa, 0,244.

5. CONCLUSÃO

Em cada cinco consultas médicas agendadas, um paciente falta e gera prejuízo de 13,4 milhões em SC em 2016 (WEISS, 2017). Apesar de implantadas medidas mais rigorosas, e exigências nas marcações de consultas ainda não impedem o absenteísmo, trazendo prejuízo à população. Nesse sentido, um modelo de previsão de faltas é imprescindível devendo diminuir os custos, pois o algoritmo pode prever se um paciente irá faltar ou não.

Este estudo-piloto indica um possível caminho para a utilização de técnicas de aprendizagem de máquina no aumento da eficiência do Sistema Único de Saúde. Os resultados preliminares dos algoritmos, podem estar relacionados à baixa capacidade preditora das *features* selecionadas. Desta forma, se faz necessário realizar mais estudos na seleção de características com maior significância no modelo de dados. Com um subconjunto de características mais significantes, é possível melhorar os preditores e por consequência, auxiliar as Secretarias de Saúde do Brasil a reduzirem o absenteísmo em consultas especializadas.

REFERÊNCIAS

ARAÚJO, Edson C.; LOBO, Maria Stella. Desafios para Sustentabilidade do Sistema Único de Saúde: **Draft for Review**. p.13, 2018

BRASIL. **Florianópolis atinge 100% de cobertura da Atenção Básica com a Estratégia de Saúde da Família**. Departamento de Atenção Primária. Ministério da Saúde. Disponível em: <http://dab.saude.gov.br/portaldab/noticias.php?conteudo=_&cod=1998>. Acessado em: 01/10/2018.

CNS. Resolução no 510, de 07 de Abril de 2016. 2016. Disponível em: <<http://conselho.saude.gov.br/resolucoes/2016/Reso510.pdf>>. Acessado em: 01/10/2018.

BRASIL. (CONASS)CONSELHO NACIONAL DE SECRETÁRIOS DE SAÚDE. **Assistência de média e alta complexidade no SUS**. 2007; 248p.

FLORIANÓPOLIS. Sala de Situação da Atenção Primária. Secretaria Municipal de Saúde de Florianópolis. 2018a

_____. Sala de Situação da Regulação. Secretaria Municipal de Saúde de Florianópolis. 2018b

GONÇALVES, Cláudia Ângela et al. Estratégias para o enfrentamento do absenteísmo em consultas odontológicas nas Unidades de Saúde da Família de um município de grande porte: uma pesquisa-ação. **Ciência & Saúde Coletiva**, v. 20, p. 449-460, 2015.

PIOLA, Sérgio Francisco et al. Estruturas de financiamento e gasto do sistema público de saúde: **A saúde no Brasil em 2030-prospecção estratégica do sistema de saúde brasileiro: estrutura do financiamento e do gasto setorial** [online]. Rio de Janeiro: Fundação Oswaldo Cruz, v. 4, p. 19-70, 2013.

R – A language and environment for statistical computing R Foundation Statistical Computing, 2018. Disponível em: <<https://www.r-project.org>>. Acessado em: 01/10/2018.

RStudio: integrated development for R

RStudio, Inc, 2018. Disponível em: <<http://www.rstudio.com/>>. Acessado em: 01/10/2018.

SALDIVA, Paulo Hilário Nascimento; VERAS, Mariana. Gastos públicos com saúde: breve histórico, situação atual e perspectivas futuras. **Estudos Avançados**, v. 32, n. 92, p. 47-61, 2018.

Apêndice A – Resultados banco OdontoEndodontia

CONFUSION MATRIX

		Referência	
		Sim	Não
Predição	Sim	157	260
	Não	418	830

DETALHES

Sensitivity 0.761	Specificity 0.273	Precision 0.665	Recall 0.761	F1 0.71
	Accuracy 0.593		Kappa 0.037	

Apêndice B - Resultados banco OdontoRadiologica

CONFUSION MATRIX

		Referência	
		Sim	Não
Predição	Sim	857	1211
	Não	1171	2119

DETALHES

Sensitivity 0.636	Specificity 0.423	Precision 0.644	Recall 0.636	F1 0.64
	Accuracy 0.555		Kappa 0.059	

Apêndice C - Resultados banco OdontoPerio

CONFUSION MATRIX

		Referência	
		Sim	Não
Predição	Sim	87	237
	Não	270	961

DETALHES

Sensitivity 0.802	Specificity 0.244	Precision 0.781	Recall 0.802	F1 0.791
	Accuracy 0.674		Kappa 0.047	