DAS Departamento de Automação e Sistemas
CTC Centro Tecnológico
UFSC Universidade Federal de Santa Catarina

Avaliação da eficiência de Unidades Básicas de Saúde de Florianópolis usando Análise Envoltória de Dados associado à modelo de predição baseado em floresta aleatória

Relatório submetido à Universidade Federal de Santa Catarina

como requisito para a aprovação da disciplina:

DAS 5511: Projeto de Fim de Curso

Debora Gardinal de Sousa

Florianópolis, Novembro de 2018

Avaliação da eficiência de Unidades Básicas de Saúde de Florianópolis usando Análise Envoltória de Dados associado à modelo de predição baseado em floresta aleatória

Debora Gardinal de Sousa

Esta monografia foi julgada no contexto da disciplina

DAS 5511: Projeto de Fim de Curso

e aprovada na sua forma final pelo

Curso de Engenharia de Controle e Automação

Prof. Carlos Ernani Fries

Errata

0.1 Análise de treinamento supervisionado com a técnica de floresta aleatória

Nesse trabalho, a análise baseada na técnica de floresta aleatória visa relacionar os escores de eficiência a outras variáveis não incluídas na análise envoltória de dados. Dessa forma, pretende-se identificar outras variáveis, além daquelas utilizadas no Modelo 2 DEA, que estejam relacionadas ao desempenho de uma UBS.

Na seleção das variáveis a serem utilizadas nesta etapa, considerou-se, dentre as variáveis levantadas em conjunto com a Secretaria Municipal de Saúde do município de Florianópolis, aquelas variáveis que configuram o processo de transformação de insumos em serviços de saúde pública.

O método de floresta aleatória selecionado para o presente trabalho foi *Gradient Boosting Regressor*. *Boosting* é um processo que combina regras de predição separadas, algumas das quais podem ser bastante fracas, para produzir um classificador combinado mais poderoso [1]. Em 1999, Friedman introduziu o conceito de *gradient boosting*, que combina idéias de *boosting* com árvores de decisão. *Regressor* faz predições numéricas com base em informações sobre uma observação. Em geral, cada observação possui um vetor de variáveis. No contexto deste trabalho, a regressão visa predizer os escores de eficiência VRS, com base nas 34 variáveis selecionadas que configuram o processo de transformação de insumos em serviços de saúde pública [2].

Para contornar possíveis problemas de *overfitting* que comprometem a precisão do modelo treinado, deve-se ajustar progressivamente os hiperparâmetros. O processo de refinamento de um modelo baseado em floresta aleatória consiste em "tentativas e erros", alterando seus hiperparâmetros e observando os resultados alcançados. O método utilizado para selecionar os hiperparâmetros foi o *Grid Search*, que procura o melhor desempenho do modelo sobre o espaço dos hiperparâmetros [3]. Dentre os hiperparâmetros que foram alterados utilizando esse método destacam-se: *learning rate*, número de árvores criadas e máxima profundidade dessas árvores. Após diversas variações de valores dos três hiperparâmetros citados, selecionou-se como 20, o número de árvores criadas, e como 2 níveis a máxima profundidade, o qual limita o número de nós em uma árvore. Para uma análise mais criteriosa, decidiu-se analisar dois diferentes valores para o hiperparâmetro de *learning rate*, sendo eles 0.004 e 0.0004. As Figuras 1 e 2 ilustram as curvas de aprendizado dos dois modelos com *learning rate* diferenciado.

Figura 1 – Curvas de aprendizado com *learning rate* de 0.004

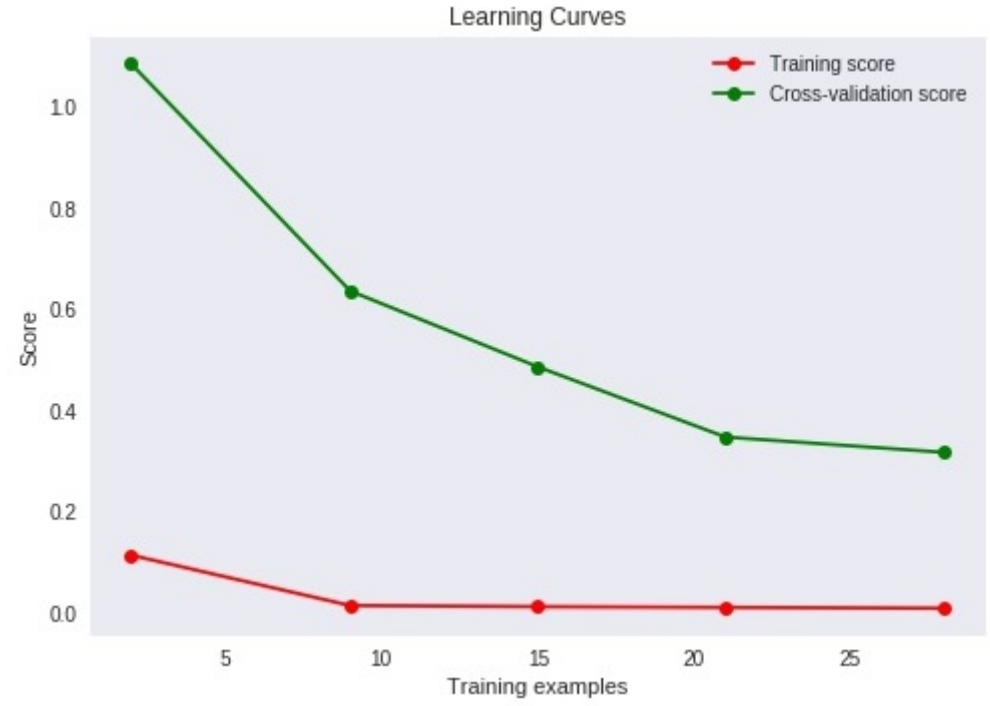


Figura 2 – Curvas de aprendizado com *learning rate* de 0.004

As Figuras 3 e 4 mostram a importância relativa das variáveis determinada pela análise de floresta aleatória. Essas importâncias representam o quanto incluir uma variável específica melhora a predição. Os gráficos das Figuras 3 e 4 mostram que "número de funcionários nutricionistas" é o melhor preditor do escore de eficiência técnica do modelo e

a “área total consultório” se encontra na segunda posição.

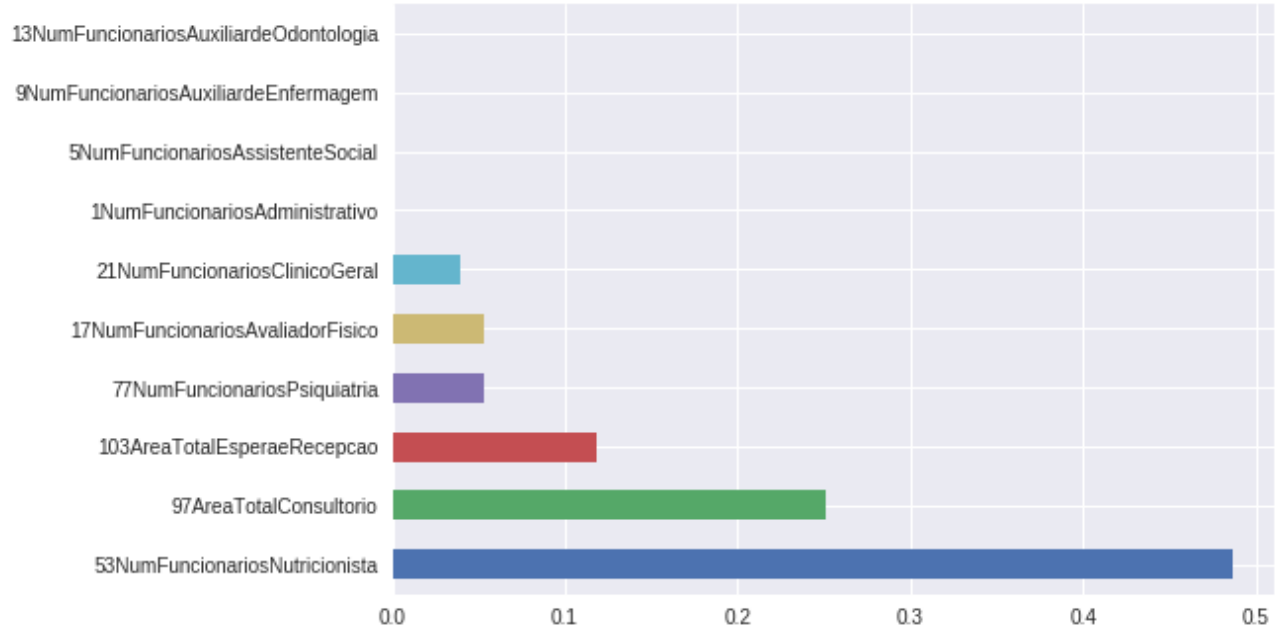


Figura 3 – Variáveis importantes com *learning rate* de 0.004

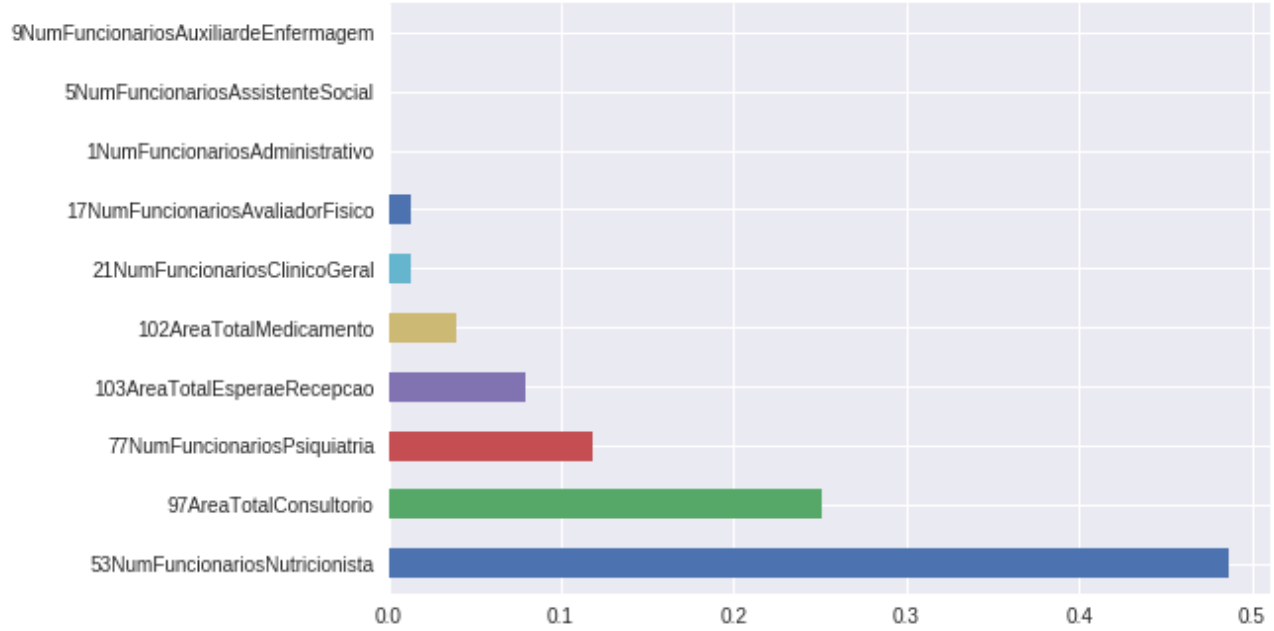


Figura 4 – Variáveis importantes com *learning rate* de 0.0004

Nota-se que mesmo com um valor dez vezes menor para o hiperparâmetro *learning rate*, as variáveis de maior importância no modelo continuam praticamente as mesmas, e as curvas de aprendizado também são bastante similares. Por causa disso, novas formas de avaliação foram incluídas para selecionar o hiperparâmetro *learning rate* mais adequado. Calculou-se, então, o erro quadrático médio, ou MSE (*Mean Squared Error*), que mede a média dos quadrados dos erros entre o valor predito e o valor observado. Quanto menor o valor do erro, melhor o desempenho do modelo calculado. Para o modelo com o hiperparâmetro *learning rate* de 0.004, o valor do MSE foi de 0.0270, enquanto que para o modelo com o hiperparâmetro *learning rate* de 0.0004, esse valor foi de 0.0282. Por conta disso, a análise prosseguiu apenas com a representação da última árvore criada com o modelo que foi configurado com *learning rate* de maior valor, ou seja, 0.004. A representação dessa árvore é apresentada na Figura 5.

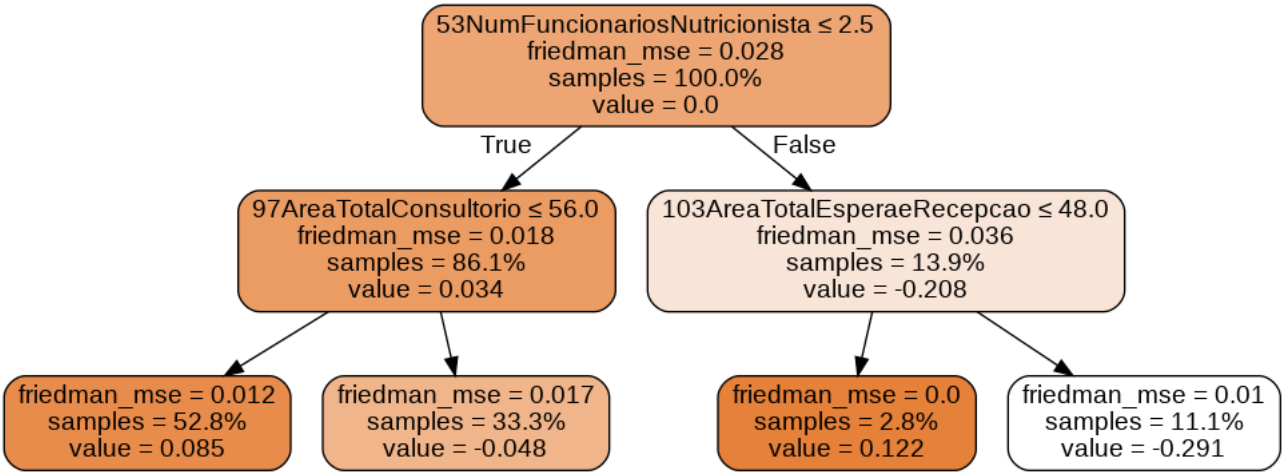


Figura 5 – Representação da árvore resultante da análise

Os nós da árvore são subdivididos, quantas vezes for necessário, até o modelo não identificar maiores ganhos de informação e chegar aos nós terminais, que não possuem bifurcações. Analisando os nós dessa árvore, a primeira linha indica a variável e o valor para dividir esse nó. *Friedman_mse* descreve o valor do erro quadrático médio do nó, enquanto que *samples* descreve o número de *data points* .

O caminho representado por "número de funcionários nutricionistas" menor ou igual a 2.5 e "área total consultório" menor ou igual a 56 representa a maior parcela do conjunto de dados, ou seja, 52.8%.

0.2 Interpretação dos resultados

Maior precisão das predições, para grandes conjuntos de dados, é muitas vezes obtida por modelos complexos difíceis de serem interpretados, como no caso de modelos de aprendizagem *ensemble* [4]. Isso gera tensão entre precisão e interpretabilidade e obriga a escolha entre modelo preciso, porém de difícil interpretação, e modelo simples, como o modelo de regressão logística, que prejudica a precisão, porém é de fácil interpretação [5]. Para se obter o equilíbrio entre precisão e interpretabilidade, vários métodos foram propostos para ajudar os usuários a interpretar as predições de modelos complexos. Um desses métodos é o SHAP (*SHApley Additive exPlanations*), introduzido por Lloyd Shapley em 1953. É uma técnica usada na teoria dos jogos para determinar quanto cada jogador em um jogo colaborativo contribuiu para o seu sucesso. Em outras palavras, cada valor de SHAP mede quanto cada variável contribui, seja positiva ou negativamente, para cada predição (Figura 6) [5].

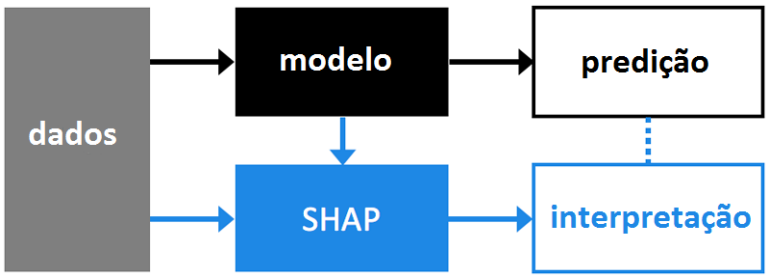


Figura 6 – Representação do método SHAP

De uma maneira simplificada, os valores SHAP calculam a importância de uma variável comparando o que um modelo prevê com e sem essa variável. No entanto, como a ordem em que um modelo vê as variáveis pode afetar suas predições, esse cálculo é feito em todas as ordens possíveis. A biblioteca SHAP do *Python* é uma ferramenta que facilita esse cálculo exaustivo e permite análises mais rigorosas comparadas com métodos de importância de variáveis [6], como mostrado nas Figuras 3 e 4.

Um exemplo de aplicação do método de análise SHAP pode ser visto no desafio do Titanic no site *Kaggle*, um espaço para a prática de data science. O naufrágio do Titanic é um dos mais famosos naufrágios da história. Durante sua viagem inaugural, em 1912, o Titanic afundou depois de colidir com um iceberg, matando 1502 de 2224 passageiros e tripulantes. No desafio proposto pelo site *Kaggle* pretende-se predizer quais passageiros sobreviveram ao naufrágio levando em consideração diversos dados reais disponíveis sobre a tragédia. Dentre os dados disponíveis estão as informações sobre: gênero dos passageiros, idade, o número de irmãos mais cônjuges a bordo do navio, se o passageiro estava na primeira, segunda ou terceira classe do navio, entre outros.

A Figura 7 mostra a interpretação da predição usando a regressão logística. Nota-se que o gênero e a idade, marcados pela cor rosa, têm o impacto mais positivo na sobrevivência, que nesse caso é o que deseja-se prever, enquanto a classe, marcada pela cor azul, tem o impacto mais negativo.

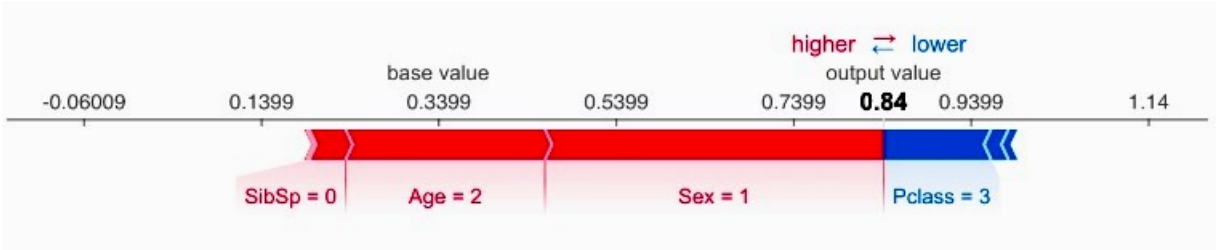


Figura 7 – Decomposição da predição na soma dos efeitos de cada variável no exemplo do desafio do Titanic

A Figura 8 mostra a contribuição de cada variável para empurrar o *output* do modelo do valor base (a saída média do modelo sobre o conjunto de dados de treinamento) para o *output* do modelo. As variáveis que aumentam a predição são mostradas em rosa e seu tamanho visual mostra a magnitude do efeito da variável. As variáveis que diminuem a predição estão em azul [7]. Portanto, percebe-se que as variáveis que contribuem positivamente com o aumento da predição são: em primeiro lugar a "área total consultório" e em segundo, o "número de funcionários nutricionistas". Já neste caso, não há variáveis que diminuam o valor de predição, pois não são indicadas nenhuma variável em azul, como demonstrado na Figura 7.

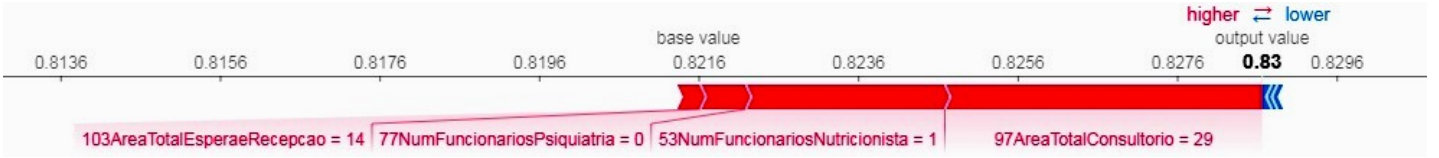


Figura 8 – Decomposição da predição na soma dos efeitos de cada variável

Para entender como uma única variável influencia no *output* do modelo, analisa-se o valor SHAP dessa variável em relação ao valor da variável para todos os exemplos no conjunto de dados. Como os valores SHAP representam a responsabilidade de um variável por uma alteração no *output* do modelo, a Figura 9 representa a alteração do valor SHAP para "área total consultório", conforme a "área total consultório" é alterada. A dispersão vertical em um único valor de "área total consultório" representa efeitos de interação com outros recursos. Para ajudar a revelar essas interações, a biblioteca SHAP do *Python* seleciona automaticamente outra variável para colorir, que neste caso é a variável "número de funcionários nutricionistas". A figura mostra que o maior impacto positivo no valor da predição, ou seja, o *output* que nesse caso são os escores de eficiência VRS, pode ser visto quando a "área total consultório" for menor que 60 e o "número de funcionários nutricionistas" menor que 2.

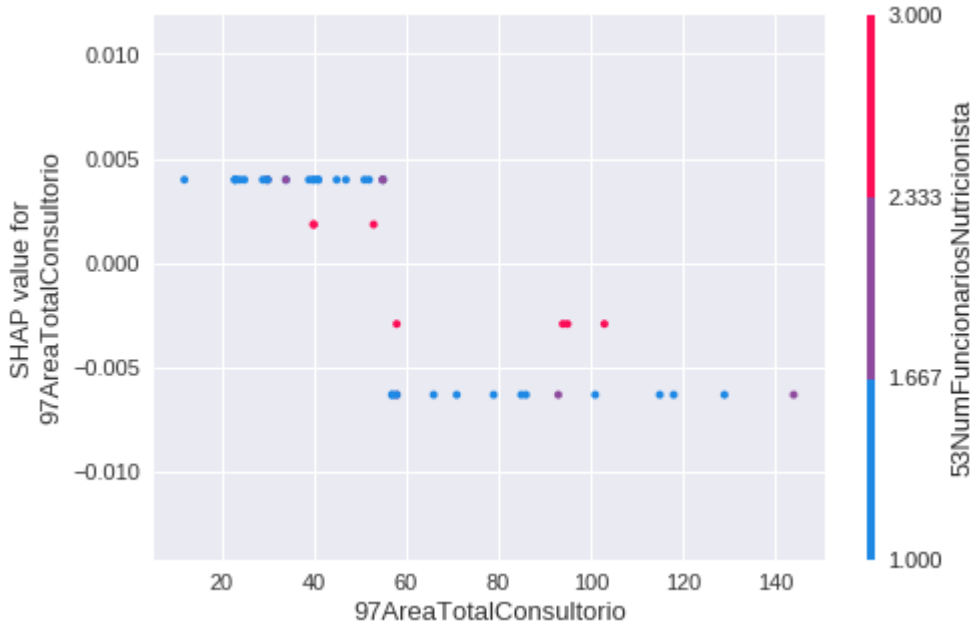


Figura 9 – Valores de SHAP plotados em relação a "área total consultório"colorida por "número de funcionários nutricionistas".

A Figura 10 mostra os valores SHAP de cada variável para cada amostra. Essa análise é feita para se obter uma visão geral de quais variáveis são mais importantes para o modelo. A localização horizontal mostra qual variável em análise, se essa variável possui valor alto (rosa) ou valor baixo (azul) para essa linha do conjunto de dados, enquanto a localização vertical mostra se o efeito desse valor conduz a uma predição do escore de eficiência maior ou menor. Isso revela, por exemplo, que valores baixos no "número de funcionários nutricionistas" aumentam o valor de eficiência predito.

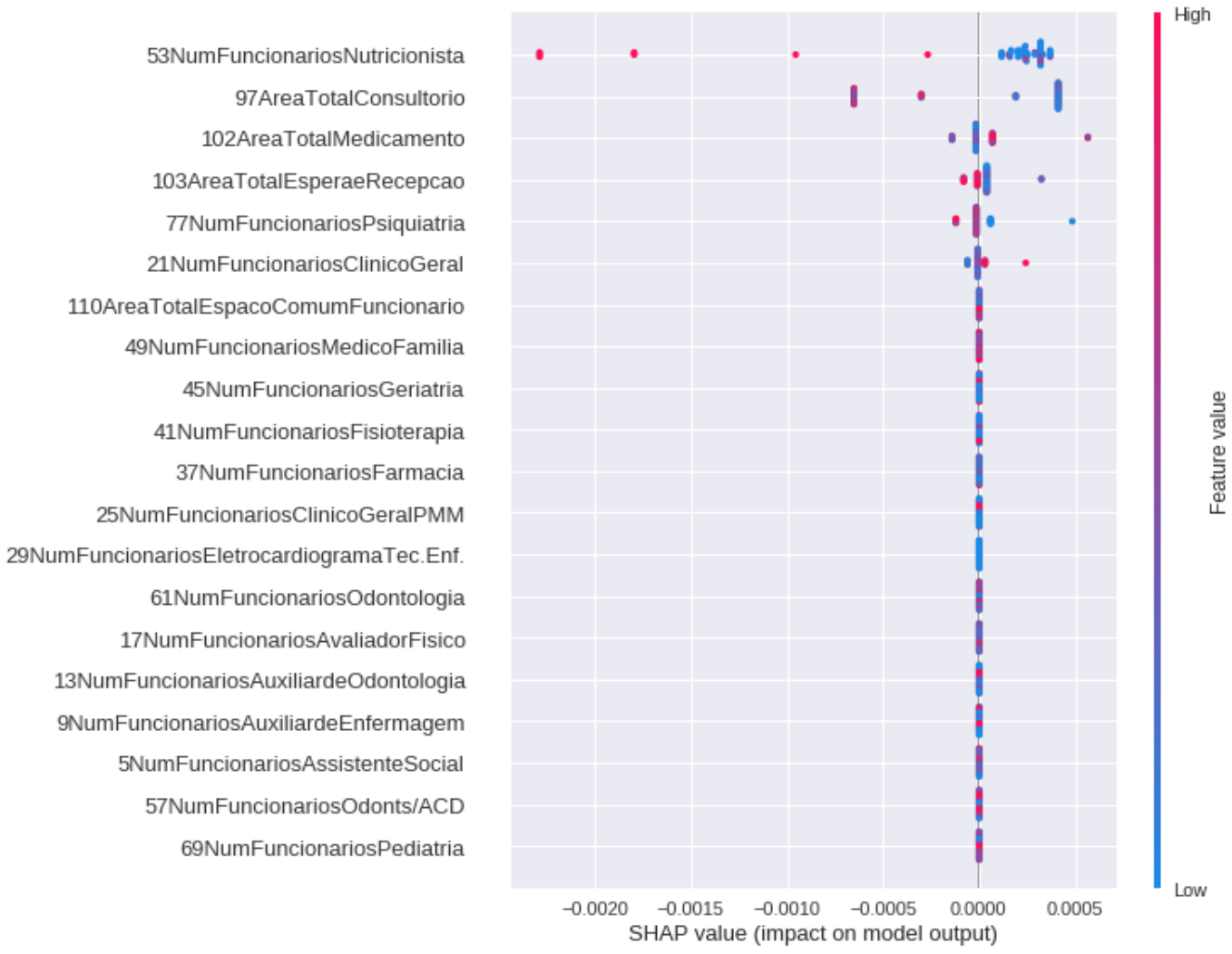


Figura 10 – Efeito, positivo ou negativo, de cada variável na predição

O método SHAP apresenta, de forma análoga aos gráficos das Figuras 45 e 46, a importância relativa das variáveis, mostradas na Figura 11. Este gráfico revela que as variáveis que mais contribuem para a predição são "número de funcionários nutricionistas" e "área total consultório".

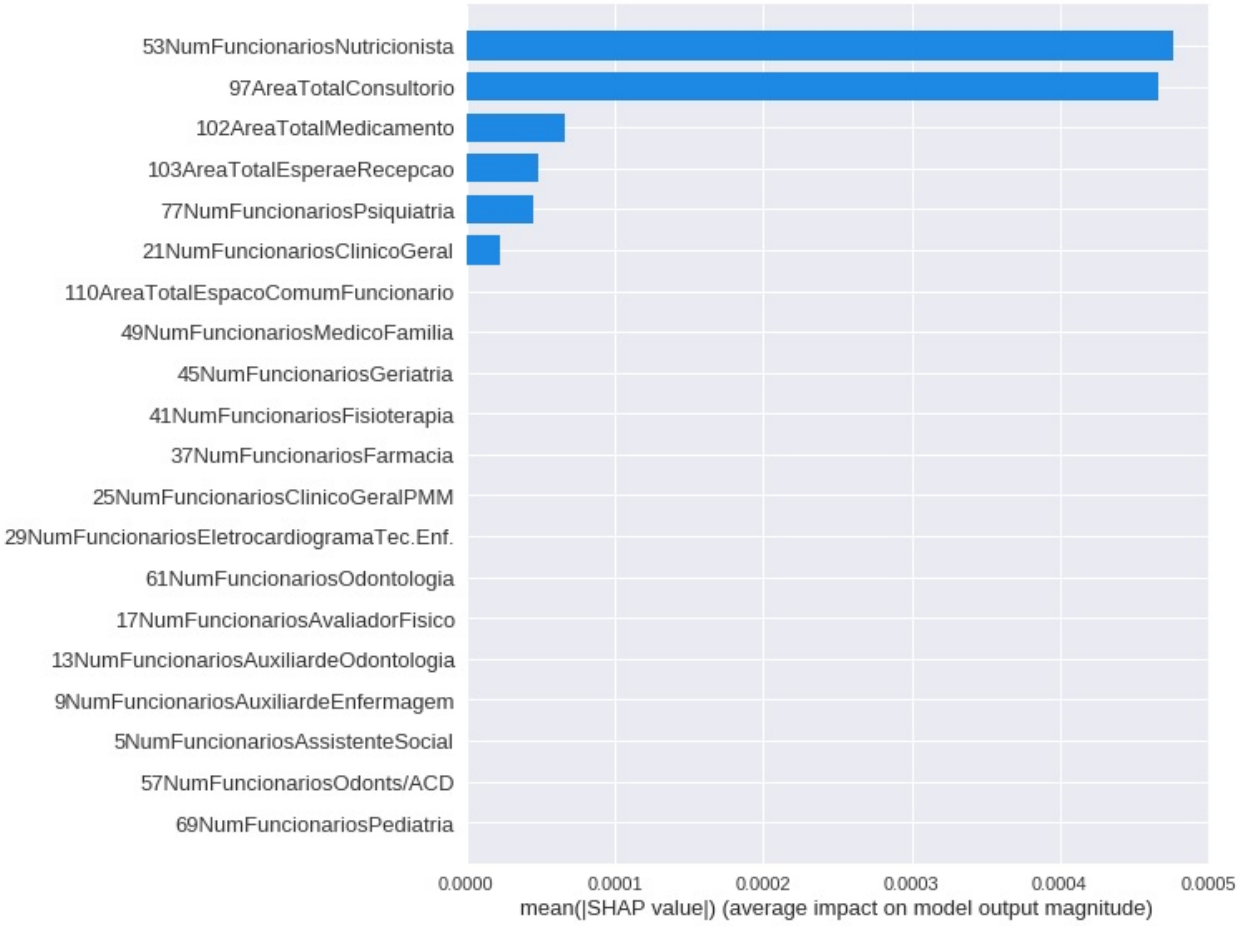


Figura 11 – Importância de cada variável utilizando valores SHAP

0.3 Considerações Finais

A obtenção dos escores de eficiência pela aplicação do modelo DEA VRS orientado para o *input* nos dados das Unidades Básicas de Saúde do município de Florianópolis possibilitou identificar unidades eficientes e ineficientes, além de projetar aquelas com ineficiência na fronteira eficiente de produção. O cálculo dos escores CRS, IRS e DRS foi essencial na classificação das UBS's quanto à escala de operação, encontrando 6 unidades em escala ótima de operação, 9 unidades acima da escala ou subutilizadas e 31 unidades abaixo da escala ótima de operação, ou seja, estão sobrecarregadas.

Apesar de terem sido calculados os valores projetados para as unidades ineficientes, a mudança de todas as áreas de uma UBS e o remanejamento de funcionários do Sistema de Saúde do município não é viável financeiramente e operacionalmente, principalmente a curto

prazo. Com isso, realizou-se um segundo estágio de análise de treinamento supervisionado com a técnica de floresta aleatória, que teve como objetivo analisar outras variáveis além das utilizadas no modelo DEA.

A utilização da técnica de floresta aleatória relacionando os escores de eficiência VRS, encontrados pela Análise Envoltória de Dados, a outras variáveis não utilizadas anteriormente mostrou-se uma técnica de difícil interpretação, sendo assim, necessária uma análise adicional para aprimorar a interpretação dos resultados. O método de análise escolhido foi SHAP, uma técnica unificada para interpretar as predições de qualquer modelo de aprendizado de máquina. De uma maneira bem simples, esse método calcula a predição do modelo sem uma determinada variável, em seguida, calcula a predição do modelo com essa determinada variável, e por fim calcula a diferença entre os dois valores, levantando em consideração todas as possíveis ordens de variáveis.

A Secretaria Municipal da Saúde, que convive diariamente com as dificuldades enfrentadas pela Atenção Básica no município pode, a partir deste trabalho, avaliar as informações encontradas com viés direcionado à necessidade de cada UBS e à facilidade de implantação de cada mudança e então definir quais medidas gerenciais podem efetivamente ser realizadas a fim de aumentar a eficiência do sistema.

0.4 Conclusões e Perspectivas

O trabalho desenvolvido teve como objetivo avaliar a eficiência das Unidades Básicas de Saúde do município de Florianópolis e identificar possíveis medidas gerenciais, que possam auxiliar na tomada de decisão da Secretaria Municipal da Saúde, para a melhoria do desempenho das UBS's do município.

Visto que o modelo DEA, para ser discriminatório, exige um número reduzido de variáveis em relação ao número de DMU's, a primeira etapa da análise consistiu em uma redução de dimensionalidade dos dados. As 110 variáveis, levantadas em conjunto com a Secretaria Municipal da Saúde, passaram por diversas etapas eliminatórias, com o objetivo de se utilizar apenas aquelas que tivessem maior ganho de informação para o modelo. Foi incluído um segundo estágio de análise, considerando a análise de treinamento supervisionado utilizando a técnica de Florestas Aleatórias, a fim de não descartar o possível impacto de outras variáveis no desempenho das UBS's. A técnica de floresta aleatória relaciona os escores de eficiência VRS com as variáveis de *input*, exceto aquelas que já tinham sido analisadas no modelo DEA.

No primeiro estágio do DEA foi possível identificar quais UBS's são tecnicamente eficientes e quais não o são. Para as DMU's que se mostraram em situação crítica, foram sugeridas medidas relacionadas às características utilizadas no modelo DEA. Com a análise de treinamento supervisionado utilizando a técnica de floresta aleatória, pretendia-se

identificar regras de decisão que levam as UBS's aos escores mais baixos de eficiência e regras que conduzem à eficiência máxima. Visto a complexidade em se analisar os resultados gerados a partir da utilização dessa técnica, uma análise adicional, com a técnica SHAP, foi incluída neste projeto. A aplicação dessa técnica mostrou-se uma ferramenta facilitadora de interpretação das predições.

Apesar de Florianópolis ser a capital brasileira mais bem avaliada na questão Atenção à Saúde Primária, segundo o Ministério da Saúde, ela pode ser sujeita à análise de eficiência, com vistas a projetar suas operações na fronteira máxima de eficiência. As análises feitas e as possíveis medidas gerenciais apresentadas podem auxiliar a SMS na tomada de decisões em busca pelo aumento da eficiência dos serviços prestados pelas UBS's de Florianópolis. Com o desafio de otimizar os recursos disponíveis, aumentando o nível de atendimento ao público, é recomendado a realocação de recursos, de forma a projetar as UBS's ineficientes na fronteira de eficiência de escala ditada pelo conjunto de UBS's eficientes. Os resultados apresentados devem ser analisados por profissionais da área da saúde que possuem contato direto com a Atenção Básica, pois estes são capazes de priorizar as medidas gerenciais sugeridas.

A exclusão das unidades, que não possuíam dados relacionados a infraestrutura e/ou dados do quadro de funcionários, foi uma das limitações encontradas no presente trabalho, e pode ser abordada em trabalhos futuros. Além disso, outra limitação observada foi na quantidade de dados levantados em conjunto com a Secretaria Municipal da Saúde de Florianópolis. Por se tratar de um pequeno conjunto de dados, a análise utilizando a técnica de floresta aleatória ficou comprometida. Por fim, uma análise interessante a ser incluída em trabalhos futuros seria o nível de escolaridade dos funcionários das UBS's. Essa discriminação feita no quadro de funcionários seria para analisar se profissionais mais experientes contribuem para o aumento da eficiência das unidades, visto que, apenas uma alteração nos recursos não seja suficiente para uma melhoria na eficiência, se os funcionários não estiverem capacitados para atender à demanda.

A Análise Envoltória de Dados combinada à técnica de floresta aletória pode servir como suporte à tomada de decisões de diversos órgãos públicos e privados ao encontrar escores de eficiência e relacioná-los a outras variáveis que podem ser contínuas ou categóricas. No caso da Secretaria Municipal da Saúde, o trabalho auxilia os tomadores de decisão ao alocar ou realocar recursos e assim melhorar a Atenção Básica da Saúde do município.

Banca Examinadora:

Prof. Carlos Ernani Fries
Orientador na Universidade

Prof. Julio Elias Normey Rico
Responsável pela disciplina

Prof. Ricardo Faria Giglio, Avaliador

Bruno Rodrigues Battistotti, Debatedor

Nathaniel Salvador de Oliveira, Debatedor

Resumo

O investimento na saúde pública no Brasil tem sofrido cortes de recursos devido a oscilações da economia e dificuldades financeiras do governo federal. Isso, além de afetar na qualidade dos serviços prestados e dificultar a manutenção dos mesmos, também faz com que as secretarias municipais tenham que reduzir a oferta de serviços de saúde pública. Dentre os serviços prestados na área da saúde pública, têm-se as Unidades Básicas de Saúde, que funcionam como porta de entrada dos cidadãos no sistema. O presente trabalho analisa, sob a ótica da Análise Envoltória de Dados e técnicas de Apredizado de Máquina, as eficiências das Unidades Básicas de Saúde no município de Florianópolis - SC, o qual é considerado referência em Atenção Primária no Brasil pelo Ministério da Saúde, em 2014. Os dados utilizados nesse trabalho são referentes ao ano de 2017 e foram fornecidos pela Secretaria Municipal de Saúde de Florianópolis. Para selecionar as variáveis *input* e *output*, usadas no modelo DEA, é aplicado um processo sequencial com modelos estatísticos de redução de dimensionalidade seguido de um processo seletivo iterativo. O conceito de entropia da informação de Shannon é aplicado como medida da quantidade de informações úteis fornecida pelo conjunto de dados. Os escores de eficiência de cada unidade é calculado por meio de modelos DEA com retornos variáveis de escala e orientado a *inputs*. No segundo estágio da análise, com vistas a se levantar relações adicionais entre as variáveis *inputs*, que não foram incluídas no modelo DEA, e os escores de eficiência técnica apurados pelo DEA, aplicou-se o modelo de predições baseado na análise de treinamento supervisionado com técnica de floresta aleatória. Os resultados identificaram as unidades que, devido ao porte inadequado, não estão ajustadas às demandas da população. Esses resultados podem servir de suporte à decisão para a Secretaria Municipal de Saúde melhorar a eficiência do conjunto de Unidades Básicas de Saúde, projetando-as para o nível ótimo de produtividade identificado na amostra, com a alocação adequada de recursos.

Palavras-chave: Análise Envoltória de Dados. Eficiência. Atenção Básica à Saúde. Floresta Aleatória.

Abstract

Public health investment in Brazil has suffered cuts in its resources due to fluctuations in the economy and financial difficulties of the federal government. In addition to affecting the quality of services provided and making it more difficult to maintain them, Municipal Health Administration also have to reduce the supply of public health services. Among the services provided in the area of public health, there are the Basic Health Units, which act as a gateway for citizens to enter the system. The present work analyzes, from the perspective of Data Envelopment Analysis and Machine Learning techniques, the efficiencies of the Basic Health Units in the city of Florianópolis - SC, which is considered a reference in Primary Care in Brazil by the Ministry of Health in 2014. The data used in this study refer to the year 2017 and were provided by the Municipal Health Department of Florianópolis. To select the *input* and *output* variables used in the DEA model, a sequential process is applied with statistical models of dimensionality reduction followed by an iterative selective process. Shannon's information entropy concept is applied as a measure of the amount of useful information provided by the dataset. The efficiency scores of each unit are calculated by means of DEA models with variable returns of scale and input oriented. In the second stage of the analysis, in order to establish additional relationships between the *inputs* variables, which were not included in the DEA model, and the technical efficiency scores ascertained by the DEA modelo, the random forest-based prediction model was applied. The results identified the units that, due to inadequate size, are not adjusted to the demands of the population. These results can support the decision for the Municipal Health Department to improve the efficiency of the set of Basic Health Units, projecting them to the optimal level of productivity identified in the sample, with adequate allocation of resources.

Keywords: Data Envelopment Analysis. Efficiency. Basic Health Care. Random Forest.

Lista de ilustrações

Lista de tabelas

Lista de abreviaturas e siglas

ACD - Auxiliar de Consultório Dentário

ACP - Análise Componentes Principais

ACS - Agente Comunitário de Saúde

CAB - Caderno de Atenção Básica

CRS - Modelo DEA com retornos constantes de escala

DEA - *Data Envelopment Analysis* (Análise Envoltória de Dados)

DMU - *Decision Making Unit* (Unidades de Decisão)

DRS - *Decreasing Return-to-Scale* (Retorno Decrescente de Escala)

ESC - Medida de Eficiência de Escala

ESF - Estratégia de Saúde da Família

eSF - Equipe de Saúde da Família

IRS - *Increasing Return-to-Scale* (Retorno Crescente de Escala)

NASF - Núcleo de Apoio à Saúde da Família

PAB - Protocolos da Atenção Básica

PMAQ - Programa Nacional de Melhoria do Acesso e da Qualidade da Atenção Básica

PNAB - Política Nacional de Atenção Básica

PR - Produtividade

SMS - Secretaria Municipal da Saúde

SUS - Sistema Único de Saúde

UBS - Unidade Básica de Saúde

VRS - Modelo DEA com retornos variáveis de escala

Sumário

1 Introdução

1.1 Importância

Crise econômica e política são grandes agentes da redução da qualidade de vida dos cidadãos: o Ministério da Saúde, com um corte na verba de investimento de R$ 179 milhões em 2018, tem que contornar os possíveis efeitos que essa redução pode causar na saúde pública. Pode-se destacar a diminuição dos insumos assegurados às Unidades Básicas de Saúde (UBS's) dos munícipios do país, uma vez que, segundo o Ministério da Saúde, a União é o principal financiador das redes de atenção à saúde pública. Garantir à rede de saúde pública a infraestrutura necessária às demandas da população é um desafio crescente no Brasil, único país do mundo com mais de 100 milhões de habitantes com um sistema de saúde público e gratuito. A infraestrutura de atenção básica em Florianópolis é composta por 50 UBS's e a partir do momento em que elas passam a receber um montante menor de recursos em determinado período de sua atividade, pode-se supor que o nível de serviços ligados à saúde da comunidade possa decair.

1.2 Objetivos do trabalho

A saúde é um direito básico garantido a todo cidadão do país pela Constituição Federal do Brasil. Nesse contexto, a UBS funciona como porta de entrada preferencial do Sistema Único de Saúde (SUS) e desempenha papel central na garantia de acesso à população à saúde de qualidade. A relevância desse estudo está efetivada pelo montante de recursos financeiros direcionados para saúde pública no país e sua racionalização pode implicar em volumes consideráveis de recursos economizados, que poderão ser usados para melhorar o serviço ou até mesmo expandí-lo.

O objetivo geral deste trabalho está em identificar as UBS's eficientes e ineficientes da Secretaria Municipal de Florianópolis, o que causa sua ineficiência e calcular seu grau de ineficiência por meio da Análise Envoltória de Dados (DEA). Para atingir esses objetivos, os seguintes objetivos espefícicos devem ser atendidos:

- Analisar e consolidar dados fornecidos pela Secretaria Municipal de Florianópolis
- Identificar variáveis de *inputs* (entrada) e *outputs* (saída) que melhor representam o processo de transformação do serviço de saúde oferecido pelas UBS's
- Apurar escores de eficiência técnica, de escala e gerencial, utilizando modelos matemáticos DEA.

- Identificar relações causais que explicam os baixos escores de eficiência daquelas unidades ineficientes.

1.3 Limitações

Uma das limitações encontradas foi em relação a disponibilidade dos dados para todas as Unidades Básicas de Saúde do município. Devido a falta de informações de quatro unidades, a análise e a comparação completa dos resultados das UBS's da cidade de Florianópolis foram impossibilitadas.

Como os dados advindos da Secretaria Municipal da Saúde são atualizados de forma manual, pelos atendentes das unidades, sua qualidade pode conter alguns desvios que não podem ser previstos pelo trabalho.

Os modelos de programação matemática e de técnicas estatísticas utilizados possuem suas próprias limitações. O uso de modelos de Análise Envoltória de Dados para o cálculo de eficiências, por sua vez, está submetido às suposições gerais destes modelos de programação linear. Como, por exemplo, a proporcionalidade das relações *input-output* no modelo com retornos constantes e a existência de uma fronteira de produção eficiente.

1.4 Estrutura

Este trabalho está dividido em seis capítulos, nos quais são apresentadas todas as etapas, que vão desde as pesquisas iniciais sobre o assunto até o desenvolvimento e conclusões tiradas desse trabalho. O primeiro capítulo traz o contexto do problema, bem como a sua importância e os objetivos do trabalho.

No segundo capítulo, é realizada revisão de literaturas sobre diversas formas de aplicações de DEA na área da saúde a fim de identificar práticas comuns, limitações, combinações de modelos diferentes e utilizá-los como forma de aprendizagem para melhor entendimento do assunto.

No terceiro capítulo são feitas as definições das técnicas utilizadas para a redução de dimensionalidade, bem como as técnicas de mensuração de produtividade e eficiência de processos produtivos, além de apresentar os modelos DEA e os conceitos de Entropia da Informação, árvores de decisão e floresta aleatória. Por fim, são abordadas questões em relação ao Sistema Único de Saúde (SUS), suas políticas e diretrizes para a atenção básica pública do Brasil, desenvolvidas pelo Ministério da Saúde.

O quarto capítulo detalha o procedimento metodológico do trabalho, que inicia com a obtenção e o tratamento dos dados e finaliza com a análise feita para relacionar as variáveis utilizadas na técnica de floresta aleatória, com os escores de eficiência encontrados

no modelo DEA.

O quinto capítulo apresentada os resultados do estudo e as proposições de medidas gerenciais com foco nas Unidades Básicas de Saúde identificadas como críticas pelos escores de eficiência.

Por último, o sexto capítulo apresenta as conclusões gerais do trabalho, levantando sugestões para futuros trabalhos, bem como pontos relevantes sobre as etapas de seu desenvolvimento.

2 Revisão de literatura

Esse capítulo trata de uma retrospectiva de trabalhos e publicações já realizados e que estão associados à problemática desse projeto. A pesquisa de conteúdo concentrou-se em plataformas para acesso de publicações e revistas científicas e através do Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pois integra o acesso a diversas bases de dados de todas as áreas do conhecimento. Termos relacionados com "Healthcare System", "Efficiency" e "Data Envelopment Analysis" foram consideradas para compor essa procura nos portais de publicações científicas. A pesquisa no Portal da Capes foi realizada com todas as bases de dados disponíveis para acesso e considerou-se como resultado apenas periódicos revisados por pares.

2.1 Análise Envoltória de Dados

O modelo seminal DEA, desenvolvido por Charnes, Cooper e Rhodes [8] e publicado em 1978 é um método matemático não-paramétrico adequado para avaliar a eficiência relativa de um conjunto homogêneo de unidades operacionais em um sistema de produção com múltiplas entradas e múltiplas saídas (como é o caso dos hospitais públicos) [9] [10].

A família de modelos DEA também é empregada rotineiramente em áreas que vão desde a avaliação de organizações públicas, como instituições educacionais e órgãos governamentais, até organizações privadas, como bancos e prestadores de serviços. Além disso, a agricultura, o setor bancário, a cadeia de suprimentos, o transporte e a política pública são os cinco principais campos de aplicação da DEA, com o maior número de artigos publicados em 2015 e 2016 [11].

Usando técnicas de programação linear, DEA identifica uma fronteira de eficiência que será utilizada como referência para medir a eficiência de cada unidade de análise, geralmente conhecida como Unidade de Tomada de Decisão (*Decision Making Unit - DMU*) [12]. A priori, a técnica DEA não precisa definir uma forma funcional para essa fronteira, que é uma de suas principais vantagens. Essa ferramenta apura escores de eficiência como a razão entre uma soma ponderada de produtos e uma soma ponderada de insumos, sendo que o conjunto de pesos encontrados garantem o resultado mais favorável para cada DMU que está em avaliação. Ressalta-se que deve ser aplicado para cada DMU do conjunto das que estão sendo avaliadas. De acordo com os escores apurados, a técnica fornece uma classificação das unidades em dois grupos: as DMU's eficientes, cujo escore é igual a 1, e as ineficientes, cujo escore é menor que 1 [10].

Entre seus benefícios pode-se mencionar: a capacidade de usar várias variáveis de

entradas (*inputs*) e variáveis de saídas (*outputs*) com diferentes pesos atribuídos a cada variável; o potencial para identificar fontes de ineficiência para *inputs* e *outputs* relacionados às entidades em estudo; e sua capacidade de identificar unidades de "benchmark" dentre as classificadas como eficientes [13] apud [14]. No entanto, essas forças também dão origem a algumas limitações importantes. Por exemplo, quando o número de variáveis *input* e *output* é grande em relação ao número de DMU's, os escores de eficiência tendem a se tornar inflados, ou seja, um grande número de DMU's tende a eficiência máxima (em outras palavras, igual a 1) [15].

O modelo DEA é orientado por entrada ou saída. Um modelo DEA orientado para saída é canalizado para maximizar as saídas obtidas pelas DMU's, mantendo as entradas constantes, enquanto os modelos orientados à entrada concentram-se na redução dos *inputs* para processar a quantidade determinada de *outputs* [12]. Segundo a literatura, a grande maioria dos estudos utiliza orientação por entrada assumindo que sob o ponto de vista gerencial, concentra-se no controle de insumos do que no aumento da demanda por cuidados de saúde [15].

Qualquer que seja a estrutura de aplicação, o primeiro passo para análise de eficiência baseada na DEA deve envolver a identificação de *inputs* e *outputs* relevantes a serem consideradas no modelo [10].

2.2 Avaliação da eficiência em sistemas públicos de saúde

A definição de palavras-chaves para a pesquisa bibliométrica, e a combinação entre elas, faz com que a quantidade de informação encontrada seja reduzida e filtrada de acordo com o objetivo do trabalho. Os resultados da pesquisa podem ser vistos na Tabela 1. Ao unir as palavras-chaves vê-se restringir cada vez mais o número de publicações. Após o processo de filtragem, cinco textos foram escolhidos de acordo com seus resumos para serem lidos e utilizados como base inicial no desenvolvimento deste trabalho. Dado que todos os trabalhos selecionados apresentam análise de eficiência na área da saúde, foi destacado as principais características de cada texto, descrevendo a problemática e as conclusões obtidas.

Um estudo do sistema regional de saúde da Espanha foi desenvolvido por Marianela Carrillo *et al*. [10] aplicando DEA para identificar as regiões espanholas que estão utilizando seus insumos de cuidados da saúde de forma mais eficiente e assim ranqueá-las como um meio de examinar as instituições públicas e melhorar a qualidade e eficiência no setor. O método usando envoltória foi escolhido por sua capacidade de lidar com múltiplas entradas e saídas e sua facilidade computacional de uso. Esse trabalho concluiu que as regiões podem desempenhar de forma eficiente ou ineficiente em qualquer nível e qualidade dos recursos de saúde. Isso reforça a importância da realização da análise de eficiência se o

Tabela 1 – Resultados das pesquisas com palavras-chaves

Termo	Número de publicações
Efficiency	5.870.000
Healthcare System	1.050.000
Data Envelopment Analysis	100.000
Healthcare System e *Efficiency*	131.000
Efficiency e *Data Envelopment Analysis*	82.600
Healthcare System e *Data Envelopment Analysis*	1.150
Healthcare System e *Data Envelopment Analysis* e *Efficiency*	1.100

objetivo for o desenvolvimento de políticas e gerenciamento eficaz do sistema de saúde. O ranking criado serve de base para melhoria e incentivo à alocação eficiente de recursos públicos.

Asandului *et al*. [12] desenvolveu estudo com o método DEA para analisar a eficiência de sistemas públicos de saúde na Europa. Este estudo considerou sistemas de saúde de 30 países europes, levando em consideração dados estatísticos de 2010. O estudo revelou que a grande maioria dos sistemas públicos dos países na amostra são ineficientes. Os resultados mostram que os recursos, mesmo limitados, são utilizados por alguns países com eficiência. Em particular, o sistema de saúde romeno, que foi considerado eficiente mesmo com a maior taxa de mortalidade infantil na Europa e um dos menores números de médicos por 10.000 habitantes. Esse resultado indica que sistemas de saúde com poucos recursos têm seus planos de operação projetados na fronteira de eficiência, são similares aos resultados do estudo feito por Marianela Carrillo *et al*. [10].

Estudo de análise de eficiência que utiliza a abordagem DEA em hospitais gregos foi desenvolvido por Kounetas *et al*. [15]. Os autores analisaram o desempenho de 114 hospitais usando diferentes combinações insumo-produto. Os resultados revelaram que mais de 80% dos hospitais examinados apresentam eficiência técnica menor que 0.8. Os resultados mostraram que a taxa de ocupação dos leitos afetou tanto a eficiência técnica quanto a eficiência de escala de maneira negativa. Além disso, o estudo mostrou que a adoção de equipamentos avançados melhora as eficiências técnica e de escala.

Estudo realizado no Brasil [16], publicado em 2006, avaliou o desempenho de 31 hospitais pertencentes a universidades federais brasileiras através de modelos DEA. Nesse estudo utilizou-se uma ferramenta de avaliação de desempenho com visualização 3D, que facilita a análise exploratória, escolha de variáveis e melhor compreensão dos resultados. A modelagem gerou um algoritmo de apoio à decisão orçamentária, que conduz a recomendações sobre a distribuição dos recursos públicos baseada em qualidade/eficiência, além de indicar mudanças necessárias para unidades ineficientes.

Outro estudo mais recente, similar ao realizado em 2006, propôs um modelo

multidimensional também baseado em modelos DEA para investigar e comparar a eficiência dos hospitais públicos brasileiros, em geral [9]. Foram coletados dados de 21 hospitais, sendo que 12 deles tiveram performance ineficiente. Os resultados desse estudo mostram que o sistema de saúde brasileiro tem potencial para ser aprimorado.

A proposta desse estudo visa contribuir nessa direção, analisando as UBS's sob a tutela da Secretaria Municipal da Saúde de Florianópolis.

2.3 Seleção de variáveis *input* e *output* em estudos utilizados como referência

A determinação das variáveis de *input* e *output* é fator crítico para avaliar a eficiência segundo modelos DEA. Para o tema considerado no trabalho, o levantamento bibliométrico permitiu selecionar estudos de referência que mostram as variáveis *input* e *output* consideradas nos mesmos.

A Tabela 2 resume os *inputs* e *outputs* utilizados nos trabalhos levantados durante a pesquisa bibliométrica. Observa-se que aquelas variáveis que descrevem o número de funcionários na equipe médica e o número de leitos, foram utilizadas, de modo geral, como variáveis de *input* nos modelos DEA.

O panorama geral da tabela mostra que questões relacionadas à infraestrutura de setores de saúde são variáveis relevantes para *input*, assim como, tempo médio de internação de pacientes. Além disso, gastos em saúde - como por exemplo, farmácia, equipe médica e manutenção - e gastos em compra complementar de serviços - como serviços médicos e laboratoriais - são, da mesma forma, listados como *input*.

Parâmetros que avaliam taxas de expectativa de vida e mortalidade infantil, os números de dias de tratamento, exames médicos, atendimentos ambulatoriais, internações, emergências não tratadas e cirurgias são considerados como variáveis de *output* nas referências bibliométricas.

Os trabalhos descritos acima referem-se a hospitais, o que difere do presente trabalho, que trata de UBS's. Apenas o último referencial teórico da Tabela 2 aborda o mesmo tema, e nota-se que uma das diferenças é que UBS's não dispõem de internação. Outra diferença é que UBS's não possuem despesas diretas (gastos), nelas se transfere bens. Portanto, o que interessa é a entrada de modo físico.

Uma forma de associar e comparar os inputs levantados para hospitais com os inputs de UBS's seria:

- número de leitos = área da UBS
- custo = número de servidores

Para atender a demanda de saúde da população, as UBS utilizam um conjunto de insumos (*inputs*) que são transformados em serviços oferecidos à comunidade. A eficiência de uma unidade será traduzida pela transformação desse conjunto de variáveis *input*, que no caso mensuram o número de funcionários e profissionais da UBS, bem como a população cadastrada na unidade, em produtos e serviços.

Como resultado da utilização desses insumos, as unidades de saúde produzem serviços, ou *outputs*, que foram classificados como total consultas médicas realizadas, total procedimentos odontológicos realizados, total atendimento de enfermagem realizados e total exames realizados, conforme apresentado por Cachuba, L.M. (2016) [17]

Tabela 2 – Síntese de *inputs* e *outputs* encontrados nas publicações científicas

Fonte	Inputs	Outputs
DEA-Like Efficiency Ranking of Regional Health Systems in Spain - Carrillo, M. and Jorge, J.M., 2017. [10]	1. Profissionais de saúde na atenção primária por mil habitantes 2. Leitos hospitalares por mil habitantes 3. Unidades de ressonância magnética por 100 mil habitantes 4. Gasto em saúde per capita 5. Consumo de tabaco, porcentagem da população	1. Expectativa de vida saudável ao nascer (anos) 2. Taxa de sobrevivência infantil 3. Percentagem da população com autopercepção positiva do estado de saúde
The efficiency of healthcare systems in Europe: a Data Envelopment Analysis Approach - Asandului, L. and Roman, M. and Fatulescu, P., 2014. [12]	1. Número de médicos 2. Número de leitos 3. Gastos com saúde pública como porcentagem do PIB	1. Expectativa de vida ao nascer 2. Expectativa de vida ajustada à saúde 3. Taxa de mortalidade infantil
How efficient are Greek hospitals? A case study using a double bootstrap DEA approach - Kounetas, K. and Papathanassopoulos, F., 2013 [15]	1. Número de leitos 2. Número de médicos 3. Número de pessoal aliado (ex: enfermeiras)	1. Número de dias de tratamento dos pacientes 2. Número de dias de tratamento em departamentos de pacientes externos 3. Número total de cirurgias 4. Número de exames médicos totais realizados em cada hospital
A Multiple Stage Approach for Performance Improvement of Primary Healthcare Practice - Ramírez-Valdivia, M.T. and Maturana, S. and Salvo-Garrido, S., 2011. [14]	1. Custo anual da equipe médica 2. Custo anual do serviço geral (gestão e manutenção) 3. Custo anual de farmácia	1. Número anual de consultas médicas 2. Número anual de visitas de check-up médico

Fonte	Inputs	Outputs
A model for multidimensional efficiency analysis of public hospital management - Soares, A.B. and Pereira, A.A. Milagre, S., 2017. [9]	1. Número de leitos 2. Número de médicos 3. Número de funcionários não médicos 4. Receita anual 5. Tempo médio de internação do paciente	1. Número de atendimentos ambulatoriais 2. Número de internações 3. Número de cirurgias 4. Número de exames
Análisis de la eficiencia técnica en los hospitales del Sistema Nacional de Salud español - Pérez-Romero, C. and Ortega-Díaz, M.I. and Ocaña-Riola, R. and Martín-Martín, J.J., 2016. [18]	1. Leitos instalados 2. Equipe em tempo integral (diferenciação entre corpo facultativo, outro pessoal de saúde e pessoal não-sanitário) 3. Gastos com compras e serviços externos adquiridos	1. Consultas ambulatoriais 2. Emergências não tratadas 3. Principais procedimentos de cirurgia ambulatorial
O uso da Análise Envoltória de Dados (DEA) para avaliação de hospitais universitários brasileiros - Lins, M.E. and Lobo, M.S.C and Silva, A.C.M. and Fiszman, R. and Ribeiro, V.J.P., 2006. [16]	1. Número de médicos 2. Número de funcionários não médicos 3. Receita média mensal proveniente do SUS 4. Número total de docentes 5. Número de docentes com doutorado	1. Relação internações/leito (mensal) 2. Relação cirurgias/sala (mensal) 3. Relação consultas ambulatoriais/sala 4. Número de alunos de medicina (graduação) 5. Número de residentes médicos 6. Número de mestrandos e doutorandos 7. Número de programas de pós-graduação e medicina
Uma análise da eficiência da oferta de serviços de saúde pública na região de Curitiba por meio de Análise Envoltória de Dados - Cachuba, L.M., 2016. [17]	1. Total de funcionários da área administrativa 2. Total de profissionais nível superior de Enfermagem 3. Total de profissionais da área médica 4. Total de profissionais da área odontológica 5. Total de profissionais de outras áreas (Biólogos, Nutricionistas, etc) 6. Técnicos especializados para apoio 7. Agente comunitário de saúde 8. Total de funcionários da unidade 9. População cadastrada na unidade	1. Total consultas médicas realizadas 2. Total procedimentos odontológicos realizados 3. Total atendimento de enfermagem realizados 4. Total exames realizados

3 Fundamentação Teórica

Neste capítulo são apresentados os temas e os conceitos utilizados para o desenvolvimento do trabalho. Serão abordados os conceitos de eficiência em sistemas produtivos e seus modelos de medição, como a Análise Envoltória de Dados tradicional e também aliada a modelos de Regressão. Posteriormente serão abordados árvores de decisão e floresta aleatória.

3.1 Redução de dimensionalidade

Os conjuntos de dados atuais são altamente suscetíveis a ruídos, dados ausentes e inconsistentes devido a erros humanos, falhas mecânicas e ao seu tamanho tipicamente grande. A etapa de redução de dimensionalidade consiste em mapear os dados para um espaço dimensional inferior, de modo que a variação não informativa nos dados seja descartada, ou de tal forma que um subespaço no qual os dados residem seja detectado [19].

3.1.1 Correlação

Correlação é um termo amplamente usado nas estatísticas e significa relação mútua/analogia. Ela geralmente descreve o efeito que dois ou mais fenômenos ocorrem juntos e, portanto, estão ligados. No entanto, a correlação não implica causalidade. Os coeficientes de correlação são calculados para medir a associação linear entre duas variáveis (X e Y). Uma correlação expressa a força da ligação ou co-ocorrência entre as variáveis em um único valor entre -1 e $+1$, e é representada tipicamente pela letra r [20]. O coeficiente de correlação linear é dado pela seguinte fórmula:

$$r = \frac{\sum_i (x_i - \overline{x_i})(y_i - \overline{y_i})}{\sum_i \sqrt{(x_i - \overline{x_i})^2}\sqrt{(y_i - \overline{y_i})^2}} \tag{3.1}$$

onde $\overline{x_i}$ é a média de X, e $\overline{y_i}$ é a média de Y.

Um valor r positivo expressa uma relação positiva entre as duas variáveis (quanto maior X, maior Y) enquanto um valor r negativo indica uma relação negativa (quanto maior X, menor Y). Um coeficiente de correlação de zero indica que não há relação entre as variáveis, considerando o conjunto de dados analisado.

3.1.2 Análise Componentes Principais

A Análise de Componentes Principais (ACP) é uma técnica estatística multivariada de redução de dados usada para identificar um pequeno conjunto de variáveis que

representam grande parte da variância observada nas variáveis originais [21].

O objetivo da ACP é reduzir a dimensionalidade de um conjunto de dados no qual variáveis podem estar correlacionadas. Essa redução de dimensionalidade é obtida pela transformação dos dados em um novo conjunto de variáveis de referência não correlacionadas chamado de componentes principais (CP) [22].

Para atingir esse objetivo, a ACP calcula as CP's como combinações lineares das variáveis originais. A primeira componente principal é posicionada na direção de maior variabilidade dos dados. A segunda componente é computada sob a restrição de ser ortogonal à primeira componente, e as demais componentes são calculadas sequencialmente de forma similar. Os valores dessas novas variáveis para as observações são chamados de escores de fatores, e eles podem ser interpretados geometricamente como as projeções das observações sobre as componentes principais [23].

As componentes principais são classificadas de tal forma que cada uma representa uma porcentagem progressivamente menor de variância observada no conjunto de dados. Se quase toda a variabilidade entre as amostras puder ser explicada por um pequeno número de CP's, então as relações entre as amostras multivariadas podem ser avaliadas por simples inspeção de um gráfico bidimensional, como pode ser visto na Figura 12 [22].

Figura 12 – Representação da ACP com duas componentes principais [24]

3.2 Produtividade e eficiência

O conceito de produtividade é definido por Lovell [25] como a relação entre insumos (*input*) e a produção obtida (*output*) de uma unidade de produção. Ela pode ser representada pela seguinte fórmula:

$$\text{PR} = \frac{\text{Produção}}{\text{Insumos}} \tag{3.2}$$

Essa relação, porém pode ser definida como a proporção de produção total em relação às proporções de insumos utilizados em uma análise multidimensional, com diversos *inputs* e *outputs*.

$$\text{PR} = \frac{\sum_{j=1}^{m} p_j y_j}{\sum_{i=1}^{n} q_i x_i} \tag{3.3}$$

Sendo y_j a quantidade de j-ésimo produto, com j = 1,2,...,m e x_i indica a quantidade do i-ésimo insumo utilizado no processo de transformação, com i= 1,2,...,n. Os pesos p_j e q_i são coeficientes técnicos atribuídos pela tecnologia utilizada no processo de transformação de insumos em produtos.

Com o conceito de produtividade definido, Ferreira e Gomes [26] afirmam que a eficiência técnica é, consequentemente, dada pela relação entre a produtividade ótima (produção com menor utilização possível de insumos) e a produtividade observada, considerando o mesmo valor para os insumos:

$$\text{Eficiência Técnica} = \frac{\text{PR observada}}{\text{PR ótima}} \tag{3.4}$$

Uma forma de se medir o desempenho produtivo dos serviços de saúde é utilizando técnicas de medição de fronteira de eficiência. A Figura 13 auxilia no entendimento desse conceito, e considera um processo de produção simples em que um único input x é utilizado para produzir um único output y.

A técnica de fronteira de eficiência usa uma fronteira de possibilidade de produção para mapear um local de combinações de *outputs* potencialmente tecnicamente eficientes que uma DMU é capaz de produzir em um determinado momento. Uma DMU é tecnicamente ineficiente, quando ela falha em alcançar uma combinação de produção em sua fronteira de produção e se encontra abaixo dessa fronteira [27].

A unidade produtiva que opera no ponto C é ineficiente porque produz menor quantidade de *outputs* do que a unidade no ponto B, embora ambas operem com a mesma quantidade de *inputs*, ou também porque a unidade C produz a mesma quantidade de *outputs* que a unidade operando no ponto A, porém utilizando mais *inputs* [28].

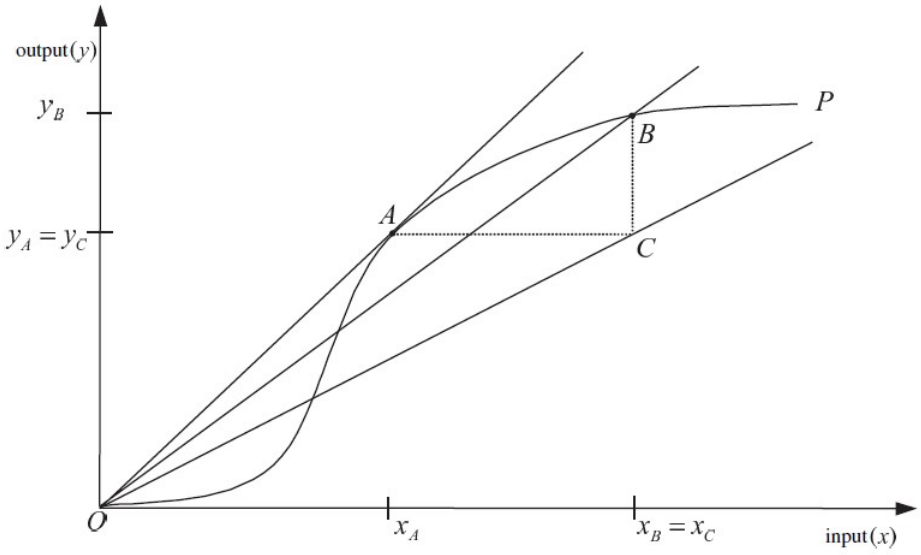


Figura 13 – Eficiência técnica e produtividade [28]

Na Figura 13 pode-se notar também que o ponto A, dentre todos os pontos tecnicamente eficientes, possui a combinação de *outputs* e *inputs* que resulta na maior produtividade possível com a tecnologia disponível. Apenas com melhorias na tecnologia de produção que poder-se-ia obter um aumento nesta produtividade máxima. Os demais pontos na fronteira de produção OP têm a possibilidade de melhorar sua produtividade com a alteração de *inputs* e *outputs* de forma a atingir os padrões de eficiência alcançados pelo ponto A.

No caso da unidade C, essa projeção para a fronteira de eficiência pode ser feita de maneiras diferentes dependendo da capacidade gerencial da mesma. Se for possível aumentar os *outputs*, a projeção é chamada de orientada a *output* e a unidade C é projetada para o ponto B. Porém, caso seja possível apenas agir na diminuição dos *inputs*, a projeção passa a ser chamada de orientada a *inputs* e a unidade C é projetada para o ponto A [29].

3.3 Seleção variáveis de *input* e *output* com a Análise Envoltória de Dados

A definição de uma DMU permite flexibilidade do seu uso sob ampla gama de possíveis aplicações. Genericamente, DMU é considerada a entidade responsável pela conversão dos *inputs* em *outputs* e cujo desempenho deve ser avaliado [13]. Neste trabalho, as DMU's foram definidas como as UBS's vinculadas à Secretaria Municipal de Saúde do município de Florianópolis.

A seleção das variáveis de *input* e *output* é desafio presente em toda a aplicação DEA, e deve representar, de forma fidedigna, o processo de transformação a ser avaliado. Elas serão utilizadas para calcular os escores de eficiência das DMU's.

Essa etapa de escolha das variáveis é crítica, e muitas vezes depende de especialistas no assunto tratado para auxiliarem na retirada de variáveis que sejam redundantes ou tenham informações conflitantes. Isso serve para que apenas informações importantes para a avaliação de eficiência do sistema sejam mantidas.

Em geral, um número reduzido de variáveis compõe o conjunto de *inputs* e *outputs* que será aplicado nos modelos DEA. Essas variáveis foram selecionadas utilizando a técnica estatística ACP, e com elas explica-se a maior parte da variância dos dados originais, ou seja, elas possuem o maior poder explicativo.

Esse trabalho visa facilitar e automatizar esse processo de escolha aplicando cálculo exaustivo dos escores de eficiência com as combinações possíveis de *inputs* e alguns *outputs* pré-selecionados pela ACP. Esse cálculo exaustivo resulta em vários modelos. Utiliza-se um critério de tomada de decisão para selecionar os modelos mais discriminantes, que neste trabalho foi a entropia definida por Shannon.

3.3.1 Número de inputs e outputs

O poder discriminatório da DEA pode diminuir caso sejam selecionados um grande número de *inputs* e *outputs* comparado ao número de DMU's do projeto em estudo. Uma regra geral sugerida é que o número de DMU's seja pelo menos o dobro do número de *inputs* e *outputs* combinados. Banker *et al.* [30], por outro lado, afirmam que o número de DMUs deve ser pelo menos três vezes o número de *inputs* e *outputs* combinados. Essas proporções entre o número de DMU's e o número de *inputs* e *outputs* trata apenas de regra imposta por conveniência, sem nenhuma base estatística [31].

3.4 Modelos DEA

O modelo básico DEA considera eficiência como a soma ponderada dos *outputs* dividida pela soma ponderada dos *inputs*, sem conhecimento a priori dos pesos destes fatores. Cada DMU estabelece seu plano de produção, ou seja, aquele conjunto de pesos para os *inputs* e *outputs* que ela considera apropriada para maximizar sua produtividade [13].

Inicialmente, o modelo proposto por Charnes et al. [8], designado por CCR, foi desenhado para uma análise com retornos constantes de escala (*CRS - Constant Return of Scale*). Posteriormente em 1984, foi estendido por Banker, Charnes e Cooper para incluir retornos variáveis de escala (*VRS - Variable Return of Scale*) e passou a ser chamado de BCC. Assim, os modelos básicos de DEA são conhecidos como CRS (ou CCR) e VRS (ou BCC) [32].

A DEA então pode ser vista como uma metodologia de avaliação de múltiplos critérios onde as DMU's são alternativas, e entradas e saídas da DEA são dois conjuntos

de critérios de desempenho. Na DEA, uma forma de modelar estes múltiplos critérios é, por exemplo, utilizando o modelo CRS [8] apud [31] no qual para dada DMU "0"de um conjunto de n unidades de produção pretende-se determinar o conjunto de pesos que maximiza a relação entre seus *outputs* e *inputs*, ou seja:

Modelo CRS:

máx e_{j_o} sujeito a $e_j \leq 1$

onde

$$e_j = \frac{\sum_{r=1}^{s} u_r Y_{rj}}{\sum_{i=1}^{m} v_i X_{ij}}; \quad j = 1, ..., n; \quad r = 1, ..., s; \quad i = 1, ..., m. \tag{3.5}$$

e X_{ij} e Y_{rj} são todos positivos e representam respectivamente as inputs e as outputs da DEA, e v_i e u_r são ≥ 0 e representam os pesos das variáveis.

O denominador da função objetivo do problema de otimização acima pode ser limitado a 1 de forma que o modelo 3.5 pode ser transformado em um Problema de Programação Linear(PPL). Para resolver a formulação anterior, procede-se a uma linearização equivalente, dada a seguir:

Modelo CRS orientado a input:

$$\text{Máx} \quad \sum_{r=1}^{s} u_r Y_{rj_o} \tag{3.6}$$

sujeito a

$$\sum_{i=1}^{m} v_i X_{ij_o} = 1 \tag{3.7}$$

$$\sum_{r=1}^{s} u_r Y_{rj} \leq \sum_{i=1}^{m} v_i X_{ij} \quad j = 1, ..., n; \tag{3.8}$$

$$r = 1, ..., s; \quad i = 1, ..., m; \quad v_i, u_r \geq 0 \tag{3.9}$$

Esse tipo de modelo considera o menor consumo possível de input para um dado nível de produção de output. As DMUs com e_j=1 estão operando com planos de produção na fronteira de eficiência. Valores de $e_j <$ 1 indicam que as DMUs estão operando fora dessa fronteira e são ineficientes quando comparadas com as primeiras.

O CRS orientado a output também pode ser analisado, quando o número combinado de insumos e produtos for menor que o número de DMUs observadas. Neste caso, sua formulação é:

Modelo CRS orientado a output:

$$\text{Mín} \quad \theta_{CRS} \tag{3.10}$$

sujeito a

$$\sum_{j=1}^{n} \lambda_j X_{ij} \leq \theta_{CRS} X_{ij_o} \quad i = 1, ..., m; \tag{3.11}$$

$$\sum_{j=1}^{n} \lambda_j Y_{rj} \geq Y_{ro} \quad r = 1, ..., s; \tag{3.12}$$

$$\lambda_j \geq 0 \quad j = 1, ..., n; \tag{3.13}$$

Tanto o modelo orientado pela entrada 3.6 quanto pela saída 3.10 devem ter resultados equivalentes, e ambos pressupõem que as unidades avaliadas operam com retornos constantes de escala. Essa análise pode ser enriquecida com o Modelo VRS, que considera retornos variáveis de escala.

A solução do modelo 3.10 provê o escalar θ^*_{CRS} , o qual corresponde ao menor multiplicador das quantidades de insumos da DMU o, que projeta esta unidade na fronteira de eficiência pelos decréscimos nos valores das quantidades de insumo [26], e mantendo constante a proporção de insumos empregada por ela. Na Figura 14 é mostrada a projeção do plano de produção da DMU o sobre a fronteira gerada pelo modelo CRS considerando um único insumo e a projeção de X^o em θ^*_{CRS} X^o, no espaço gerado pelo conjunto de dois insumos (X_1 e X_2). O índice de eficiência obtido com o modelo CRS é chamado por Cooper et al. (2007) de "eficiência técnica global" (*global technical efficiency*).

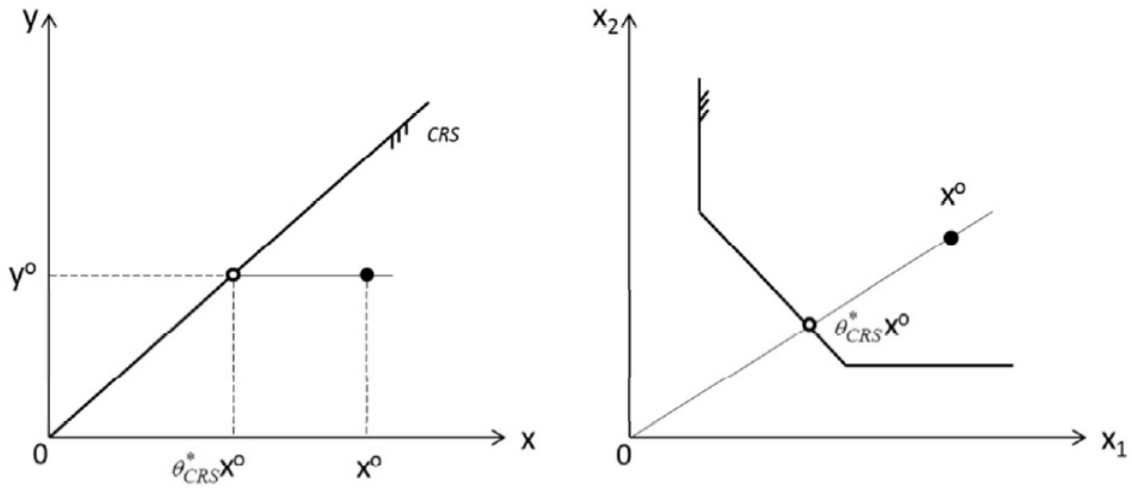


Figura 14 – Medida de eficiência técnica, modelo CRS, *input* orientado. [33]

A adição da restrição de convexidade $\sum_{j=1}^{n} \lambda_j = 1$ ao modelo CRS orientado a output restringe a região de soluções viáveis do modelo CRS às combinações convexas

geradas pelos planos de produção das DMUs observadas. Este é o modelo VRS, que melhor representa a realidade de sistemas de produção. Sua eficiência também é denominada por Cooper et al. (2007) de "eficiência técnica pura local" (*local pure technical efficiency*) e a formulação do modelo é dada por:

Modelo VRS:

$$\text{Mín} \quad \theta_{VRS} \tag{3.14}$$

sujeito a

$$\sum_{j=1}^{n} \lambda_j X_{ij} \leq \theta_{VRS} X_{ij_o} \quad i = 1, ..., m; \tag{3.15}$$

$$\sum_{j=1}^{n} \lambda_j Y_{rj} \geq Y_{ro} \quad r = 1, ..., s; \tag{3.16}$$

$$\sum_{j=1}^{n} \lambda_j = 1 \tag{3.17}$$

$$\lambda_j \geq 0 \quad j = 1, ..., n; \tag{3.18}$$

A Figura 15 ilustra a medida de eficiência técnica sob a condição de retornos variáveis de escala para um processo de transformação de um insumo em um produto. A fronteira de eficiência é definida pelos planos de produção (X^h,Y^h), (X^i,Y^i) e (X^j,Y^j). Para a DMU o, a eficiência técnica relativa ao modelo VRS, θ^*_{VRS}, com orientação para *input*, é definida pela relação das distâncias entre os pontos (0,Y^o) e (θ^*_{VRS} X^o,Y^o) e entre os pontos (0, Y^o) e (X^o,Y^o). [13]

O ponto (θ^*_{VRS} X^o,Y^o) representa a máxima retração possível de insumo que pode ser aplicada à DMU o, tal que esta seja eficiente tecnicamente sob condições de retornos variáveis de escala.

A combinação das medidas de eficiência técnica obtidos com os modelos CRS e VRS pode indicar a existência de ineficiência de escala na operação das DMUs, definida por Cooper, Seiford e Tone (2007) como sendo a relação entre as respectivas eficiências técnicas CRS e VRS:

$$ESC = \frac{\theta^*_{CRS}}{\theta^*_{VRS}} \tag{3.19}$$

Como $\theta^*_{CRS} \leq \theta^*_{VRS}$, qualquer DMU com $\theta^*_{CRS} = 1$ opera na escala mais produtiva possível e, portanto, a eficiência de escala é ESC = 1. Na Figura 16 as DMUs i e k apresentam $\theta^*_{CRS} < \theta^*_{VRS}$ e, portanto, têm eficiências de escala idênticas às respectivas eficiências técnicas CRS.

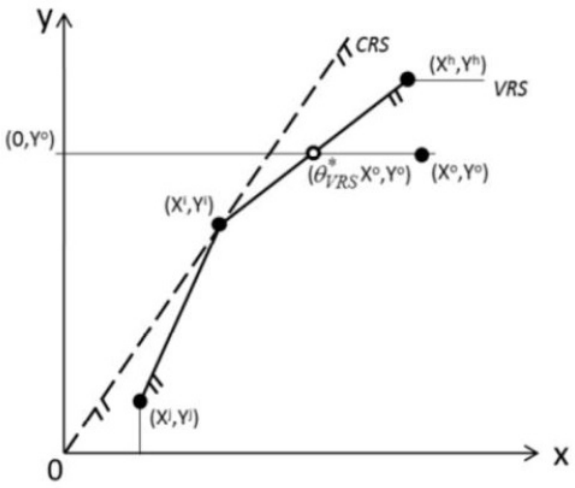


Figura 15 – Medida de eficiência técnica, modelo VRS, *input* orientado. [13]

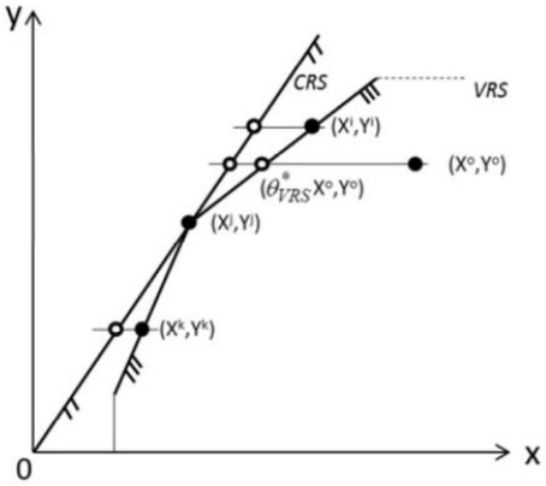


Figura 16 – Medida de eficiência de escala, *input* orientado. [33]

Extensões da abordagem DEA permitem determinar em que sentido será o retorno de escala, crescente ou decrescente. Retornos crescentes de escala (*IRS - Increasing Returns of Scale*) decorrem da operação com produtividade média crescente, que geralmente tende a atingir um patamar de máxima produtividade média e, a partir da qual, os retornos passam a ser decrescentes com o incremento da escala das operações.

O modelo IRS é formulado substituindo a equação 3.17 por $\sum_{j=1}^{n} \lambda_j \geq 1$ [26], resultando no modelo abaixo.

Modelo IRS:

$$\text{Mín} \quad \theta_{IRS} \tag{3.20}$$

sujeito a

$$\sum_{j=1}^{n} \lambda_j X_{ij} \leq \theta_{IRS} X_{ij_o} \quad i = 1, ..., m; \tag{3.21}$$

$$\sum_{j=1}^{n} \lambda_j Y_{rj} \geq Y_{ro} \quad r = 1, ..., s; \tag{3.22}$$

$$\sum_{j=1}^{n} \lambda_j \geq 1 \quad j = 1, ..., n; \tag{3.23}$$

Já o modelo DRS, com retornos não crescentes de escala são modelados substituindo novamente a equação 3.17 pela restrição $\sum_{j=1}^{n} \lambda_j \leq 1$, resultando na seguinte formulação:

Modelo DRS:

$$\text{Mín} \quad \theta_{DRS} \tag{3.24}$$

sujeito a

$$\sum_{j=1}^{n} \lambda_j X_{ij} \leq \theta_{IRS} X_{ij_o} \quad i = 1, ..., m; \tag{3.25}$$

$$\sum_{j=1}^{n} \lambda_j Y_{rj} \geq Y_{ro} \quad r = 1, ..., s; \tag{3.26}$$

$$\sum_{j=1}^{n} \lambda_j \leq 1 \quad j = 1, ..., n; \tag{3.27}$$

A Figura 17 mostra as fronteiras de eficiência e conjuntos de planos de produção possíveis para os modelos IRS e DRS, respectivamente.

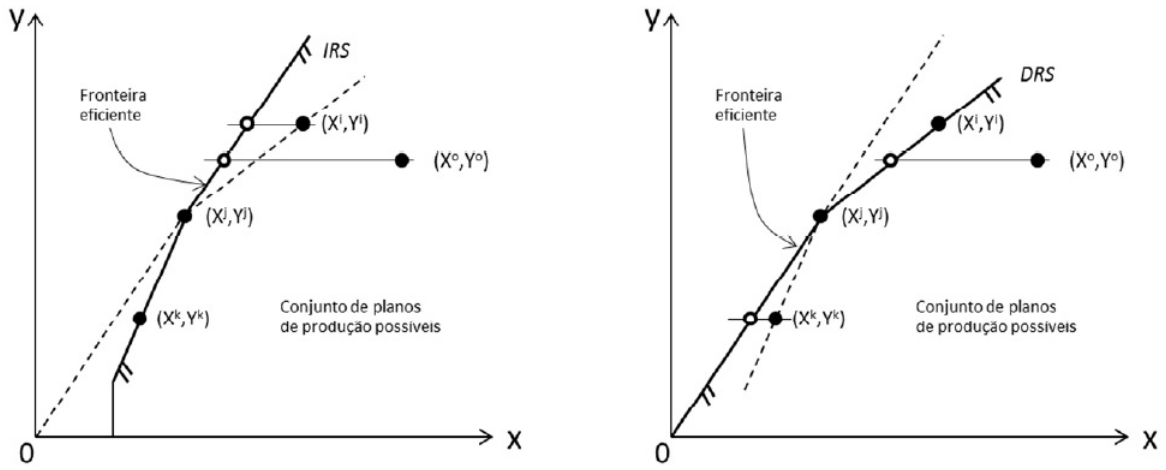


Figura 17 – Medida de eficiência para retornos crescentes (IRS) e decrescente (DRS) de escala, *input* orientados [34].

Para determinar a natureza da escala de uma DMU qualquer, comparam-se as eficiências obtidas com os modelos VRS, IRS e DRS. Se os coeficientes de eficiência técnica

VRS e IRS forem iguais, então a DMU opera numa escala com retornos crescentes. Caso sejam distintos, então a DMU opera numa escala com retornos decrescentes.

Uma análise similar pode ser feita partindo da comparação de coeficientes de eficiência técnica VRS e DRS. Se forem os mesmos, então a DMU em questão apresenta retornos decrescentes de escala. Caso os coeficientes sejam diferentes, então a DMU opera com retornos crescentes de escala. Nesse caso, medidas gerenciais que indiquem aumento do porte da DMU, com o incremento nos inputs nas mesmas proporções utilizadas, devem permitir aumento de sua eficiência técnica CRS.

3.5 Entropia de Shannon

A entropia da informação foi introduzida por Shannon em 1948, e mede a incerteza associada a uma variável aleatória no campo da teoria da informação. A idéia de usar a entropia de Shannon como coeficiente de importância foi proposta primeiramente por Zeleny (1982) em análise de decisão de múltiplos critérios, e tem sido amplamente aplicada para medir incerteza ou importância em muitos outros campos científicos, como por exemplo, matemática, ciências sociais, química e pesquisa operacional [35]. A função de Shannon é baseada no conceito de que o ganho de informação de um evento está inversamente relacionado à sua probabilidade de ocorrência [36].

Shannon definiu a entropia para um sistema de n-estados como sendo [37]:

$$H(A) = -\sum_{i=1}^{n} p_i log_2 p_i \tag{3.28}$$

onde p_i é a probabilidade de ocorrência de um evento i e

$$\sum_{i=1}^{n} p_i = 1; \quad 0 \leq p_i \leq 1. \tag{3.29}$$

Os valores de entropia não podem ser negativos, por se tratar de um cálculo de probabilidades, e o uso da base 2 no logaritmo é para medir o conteúdo da informação em bits. Se o resultado medido for zero, isso significa que a saída é certa de ocorrer e não há variabilidade nos dados. Quanto maior o valor de entropia, segundo Shannon, maior a incerteza de ocorrência do evento, portanto maior o poder discriminatório.

3.6 Árvores de decisão

O problema de aprendizado supervisionado é encontrar uma aproximação para uma função desconhecida, dado um conjunto de dados. Para resolver este problema, vários

métodos foram apresentados na literatura, sendo que um dos métodos mais representativos é o de Árvores de Decisão [38].

No contexto de problemas de aprendizagem supervisionada, deve-se distinguir entre problemas de classificação e de regressão. No primeiro caso, a variável de destino (variável dependente) toma valores em um conjunto finito e predefinido de valores não ordenados, e o objetivo usual é minimizar uma função de perda 0-1. No segundo caso, a variável de destino é ordenada e recebe valores em um subconjunto de $\mathbb{R}$. O objetivo usual é minimizar uma função de perda de erro quadrada [39].

Em geral, uma árvore de decisão é montada a partir de uma pergunta e classifica baseado na resposta. Elas são populares porque representam as informações de uma maneira intuitiva e de fácil visualização, como pode ser visto na Figura 18.

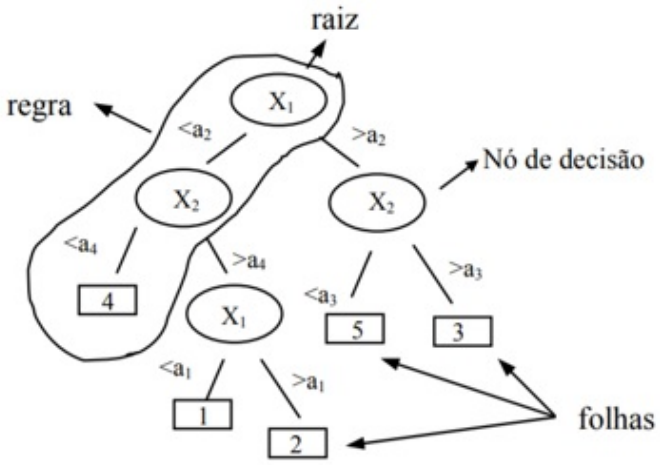


Figura 18 – Exemplo árvore de decisão [39].

Numa árvore de decisão, cada nó de decisão contém um teste para algum atributo e cada ramo descendente corresponde a um possível valor deste atributo. O conjunto de ramos é distinto, cada folha está associada a uma classe e, cada percurso da árvore, da raiz à folha corresponde uma regra de classificação. Para decidir qual atributo será escolhido para realizar as partições da árvore de decisão, analisa-se o critério de ganho de informação. O atributo que possuir o maior de ganho de informação será selecionado, e um novo processo de partição se inicia a partir dele. Nos casos em que a árvore é usada para classificação, o critério de partição mais conhecido é baseado no Índice Gini, e nos casos em que a árvore é usada para regressão, o critério de partição é baseado no conceito de Entropia da Informação [39].

O treinamento de um modelo pode ser estabelecido como a divisão dos dados em treinamento (*training*) e teste (*test*). A Figura 19 ilustra esse processo de divisão dos dados disponíveis. Essa divisão tem como objetivo verificar a acurácia do modelo treinado por meio dos dados de teste. Os dados de teste não são implementados no conjunto de

treinamento, porém possuem vetores de saída conhecidos e podem ser usados para verificar a capacidade de generalização do modelo treinado [40].

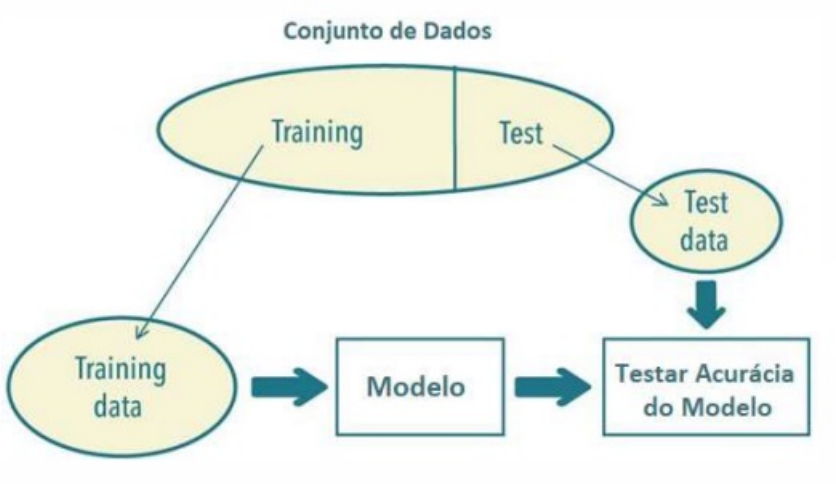


Figura 19 – Esquema do processo de divisão do conjunto de dados para treinamento e teste do modelo [40].

Avaliar a acurácia consiste em verificar quão bem o modelo é capaz de generalizar seus dados de treinamento. No entanto, as árvores de decisão são sensíveis a pequenas perturbações no conjunto de dados de treinamento, e foram identificadas como aprendizes instáveis, propensas a *overfitting*. *Overfitting* significa que a árvore classifica e prediz, de maneira satisfatória, com os dados usados para criá-la, mas não classifica e prediz, de forma similarmente satisfatória, um novo conjunto de dados. Ou seja, ela não possui poder de generalização [40].

A Figura 20 ilustra as curvas de erro para os dados de treino e de teste de um modelo qualquer, apontando as regiões de *overfitting* e *underfitting*. No caso de *underfitting*, as taxas de erro de predição é elevada tanto para os dados de treino como de teste. Já para o caso de *overfitting*, apenas o erro de predição do teste é elevado. Pode-se concluir que com o aumento da complexidade do modelo, os erros de predição também variam.

Para melhorar o desempenho na acurácia das árvores de decisão, o conceito de floresta aleatória foi introduzido por Breiman em 2001.

3.7 Floresta aleatória

A floresta aleatória combina a simplicidade de árvores de decisão com flexibilidade, resultando em uma melhora na precisão. Ela é um método de aprendizagem de máquina versátil, capaz de executar tarefas de classificação e regressão [1].

A floresta aleatória utiliza um tipo de método de aprendizado de *ensemble*, que envolve agrupar modelos preditivos de modo a melhorar a precisão e a estabilidade do modelo. Diversos métodos de classificação *ensemble* foram propostos nos últimos anos,

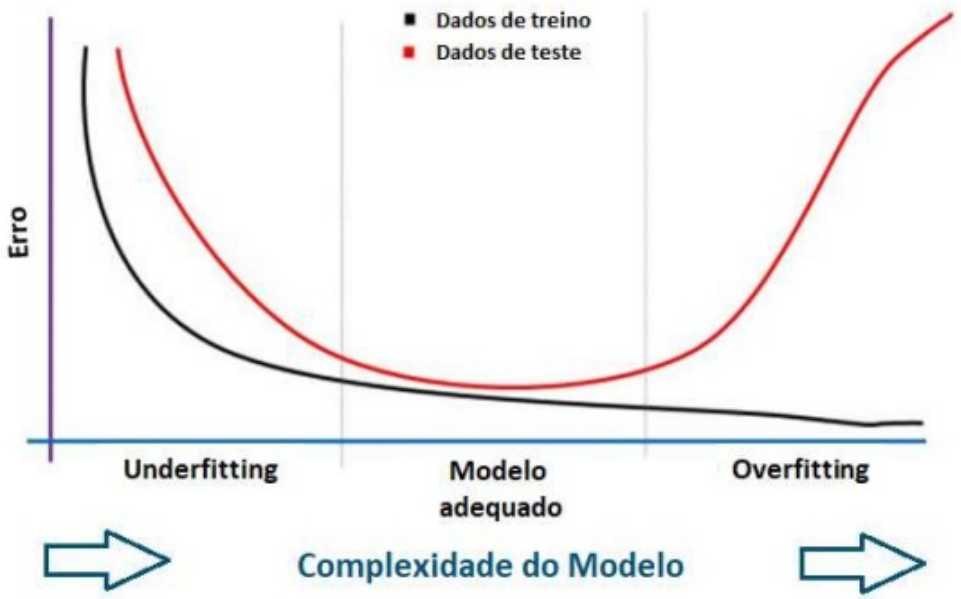


Figura 20 – Curvas de erro de treino e de teste considerando a complexidade de um modelo [40].

porém os métodos mais amplamente utilizados são *bagging* e *boosting*. O *bagging* utiliza o bootstrap no treinamento, e isso reduz a variância da classificação. Em contraste, o *boosting* usa o re-treinamento iterativo, onde as amostras classificadas incorretamente recebem um aumento de peso à medida que as iterações progridem. Consequentemente, o *boosting* se torna mais exigente computacionalmente e é mais lento do que o *bagging* [41].

Uma das vantagens em se aplicar floresta aleatória é sua eficácia em estimar os dados faltantes e manter a precisão mesmo quando grande parte dos dados está faltando. Além disso, possibilita a utilização de grande volume de dados, podendo assim, empregar milhares de variáveis, e identificar as mais significativas (as que possuem alto grau de importância). Dessa forma, a floresta aleatória é também considerada um método de redução de dimensionalidade.

3.7.1 *Bagging* e *Boosting*

Bagging, segundo o próprio criador Breiman, é um método para gerar várias versões de um preditor e usá-las para obter um preditor agregado [42]. Já o termo *boosting* refere-se a um grupo de algoritmos que utilizam médias ponderadas para tornar resultados de aprendizagem fraca em aprendizagem mais forte. Um aprendiz fraco garante um pouco melhor do que adivinhar aleatoriamente, em contraste, um aprendiz forte é um classificador que é arbitrariamente bem correlacionado com a classificação verdadeira. Quando executa cada modelo, ele rastreia quais amostras de dados são mais bem-sucedidas e quais não são [43].

O método *boosting* também requer *bootstrap*, porém é usado de forma contrária ao

bagging, já que aumenta os pesos de amostras dos dados. Isso permite que o modelo ou algoritmo obtenha um melhor entendimento das variâncias e variáveis que existem na reamostragem. O aumento dos pesos de amostras dos dados, utilizado pelo *boosting* faz com que algumas amostras sejam executadas com mais frequência do que outras. Os conjuntos de dados com baixos valores de precisão recebem pesos maiores, pois são considerados os de maior complexidade e exigem mais iterações para treinar adequadamente o modelo.

Diferentemente do *bagging*, em que cada modelo é executado independentemente e no final, as saídas são agregadas sem preferência por nenhum modelo (Figura 21), no *boosting* cada modelo executado determina os recursos nos quais o próximo modelo se concentrará, atuando assim, de forma seriada (Figura 22).

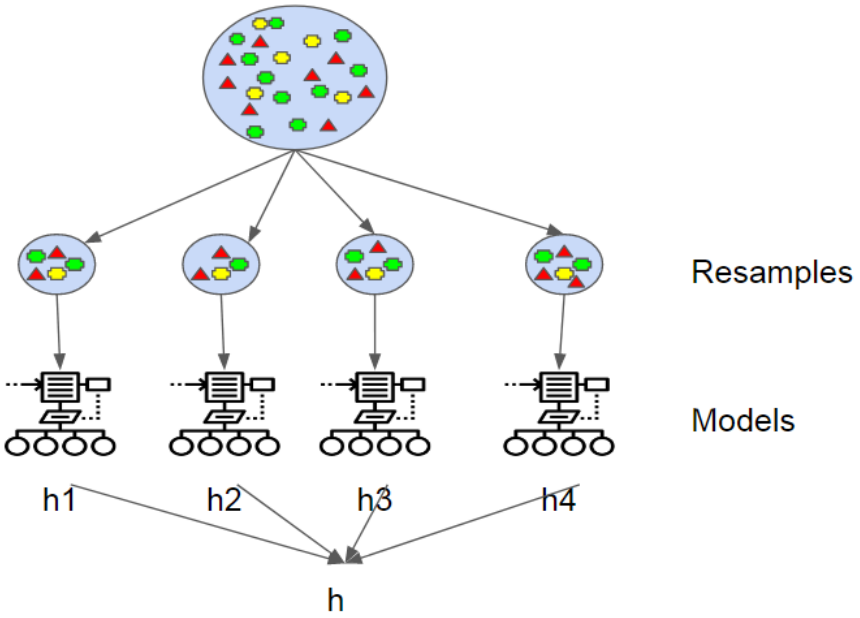


Figura 21 – Exemplo de *bagging* [1].

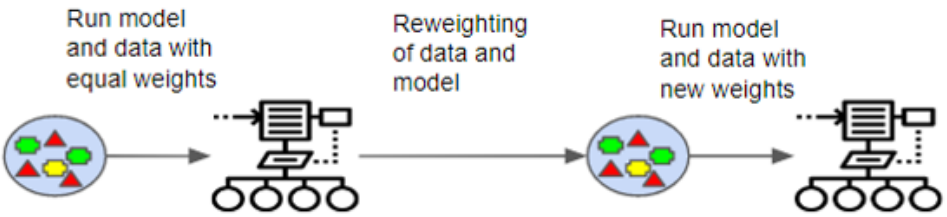


Figura 22 – Exemplo de *boosting* [1].

3.8 Sistema Único de Saúde

O Sistema Único de Saúde (SUS), lançado com a promulgação da Constituição Federal em 1988, tem como objetivos identificar e divulgar fatores condicionantes e determinantes da saúde, formular políticas de saúde e oferecer atendimento às pessoas por meio de ações assistenciais e atividades preventivas [44].

Baseado na missão em que órgãos e instituições públicas federais, estaduais e municipais devem dar assistência à população, o SUS integraliza, unifica e formaliza um conjunto de ações e serviços de saúde como dever do estado e direito de todo cidadão brasileiro conforme regulado pela Lei nº 8.080/1990. As políticas e diretrizes que regem o SUS são baseadas no modelo da promoção, proteção e recuperação da saúde (Conselho Nacional de Secretários de Saúde, 2003).

A direção do SUS deve ser única, sendo exercida em cada esfera de governo pelos seguintes órgãos: a) no âmbito da União, pelo Ministério da Saúde; b) no âmbito dos estados e do Distrito Federal, pela respectiva Secretaria de Saúde ou órgão equivalente; e c) no âmbito dos municípios, pela respectiva Secretaria de Saúde ou órgão equivalente [44].

A distribuição e definição do atendimento dos serviços de saúde é dividida em atenção primária, definida como Atenção Básica de Saúde (ABS ou somente AB), que compõe atenção básica e de baixa complexidade, atenção secundária envolvendo especialidades e média complexidade, e atenção terciária abrangendo atendimento hospitalar e alta complexidade. A ABS, representada pelas UBS's, prioriza a promoção da saúde e a prevenção de doenças [17], e é, preferencialmente, o portal de entrada do cidadão no Sistema de Saúde. A partir do primeiro atendimento, o paciente pode ser encaminhado a outros serviços de maior complexidade da saúde pública como hospitais e clínicas especializadas.

A seguir são detalhados a Política Nacional de Atenção Básica (PNAB), o Programa Nacional de Melhoria do Acesso e da Qualidade da Atenção Básica (PMAQ-AB) e o Caderno de Acolhimento à demanda espontânea, a fim de complementar o entendimento do Sistema de Saúde Pública no Brasil.

3.8.1 Política Nacional de Atenção Básica – PNAB

A Atenção Básica, por ser a principal porta de entrada e centro de comunicação com a Rede de Atenção à Saúde, ocorre no local mais próximo da vida das pessoas. Por isso, é fundamental que ela se oriente pelos princípios da universalidade, da acessibilidade, do vínculo, da continuidade do cuidado, da integralidade da atenção, da responsabilização, da humanização, da equidade e da participação social [45].

A Política Nacional de Atenção Básica (PNAB) determina que as Unidades Básicas de Saúde (UBS) sejam construídas de acordo com as normas sanitárias, tendo como

referência o manual de infraestrutura do Departamento de Atenção Básica. É recomendado que as UBS disponibilizem de consultório médico/enfermagem, consultório odontológico e consultório com sanitário, sala multiprofissional de acolhimento à demanda espontânea, sala de administração e gerência, e sala de atividades coletivas para os profissionais da atenção básica, área de recepção, local para arquivos e registros, sala de procedimentos, sala de vacinas, área de dispensação de medicamentos e sala de armazenagem de medicamentos (quando há dispensação na UBS), sala de inalação coletiva, sala de procedimentos, sala de coleta, sala de curativos e sala de observação [45].

A PNAB tem na Estratégia de Saúde da Família (ESF) sua estratégia prioritária para expansão, qualificação e consolidação da Atenção Básica. De acordo com a ESF, estabelecida em 1994, cada Unidade da Atenção Básica deve possuir, dependendo do número de cidadãos cadastrados na Unidade, uma ou mais equipes de Saúde da Família (eSF). As equipes da Saúde da Família são equipes multiprofissionais que atendem serviços assistenciais básicos limitados a uma determinada região de abrangência e estão inseridas na AB atendendo demandas locais da população (Secretaria Municipal de Saúde, 2014). Cada equipe é composta por, no mínimo, médico generalista ou especialista em Saúde da Família ou médico de Família e Comunidade, enfermeiro generalista ou especialista em Saúde da Família, auxiliar ou técnico de enfermagem e agentes comunitários de saúde, podendo acrescentar a esta composição, como parte da equipe multiprofissional, os profissionais de saúde bucal: cirurgião-dentista generalista ou especialista em Saúde da Família, auxiliar e/ou técnico em saúde bucal. Além disso, cada equipe de Saúde da Família deve ser responsável por, no máximo, 4.000 pessoas, sendo a média recomendada de 3.000 pessoas [45].

Compondo a Atenção Básica, há também os Núcleos de Apoio à Saúde da Família (NASF) que foram criados com o objetivo de ampliar a abrangência e o escopo das ações da atenção básica, bem como sua resolubilidade. São constituídos por equipes compostas por profissionais de diferentes áreas de conhecimento, que devem atuar de maneira integrada e apoiando os profissionais das equipes de Saúde da Família, das equipes de atenção básica e Academia da Saúde, atuando diretamente no apoio matricial às equipes das unidades nas quais o NASF está vinculado e no território das mesmas.

Os NASF devem buscar contribuir para a integralidade do cuidado aos usuários do SUS principalmente por intermédio da ampliação da clínica, auxiliando no aumento da capacidade de análise e de intervenção sobre problemas e necessidades de saúde, tanto em termos clínicos quanto sanitários. São exemplos de ações de apoio desenvolvidas pelos profissionais dos NASF: discussão de casos, atendimento conjunto ou não, interconsulta, construção conjunta de projetos terapêuticos, educação permanente, intervenções no território e na saúde de grupos populacionais e da coletividade, ações intersetoriais, ações de prevenção e promoção da saúde, discussão do processo de trabalho das equipes, entre

outros [45].

3.8.2 Programa Nacional de Melhoria do Acesso e da Qualidade da Atenção Básica (PMAQ-AB)

O Programa Nacional de Melhoria do Acesso e da Qualidade da Atenção Básica (PMAQ-AB) é uma das propostas apresentadas pelo Ministério da Saúde, em 2011, com a finalidade de garantir um padrão de qualidade nacional, regional e local de forma comparável às ações da Atenção Básica em Saúde [46].

A preocupação com a qualidade e eficiência da saúde pública tem respaldo em iniciativas como a formulação da Autoavaliação para Melhoria do Acesso e da Qualidade da Atenção Básica (AMAQ-AB) do Ministério da Saúde de 2012. A AMAQ-AB, inserido no Programa PMAQ como parte da estratégia "Saúde Mais Perto de Você", tem como finalidade ampliar a capacidade de gestão, visando a análise do acesso e da qualidade das ações de saúde no âmbito da AB nas três esferas de governo.

O programa é norteado por sete diretrizes, entre elas, a transparência em todas as suas etapas, permitindo o contínuo acompanhamento de suas ações e resultados pela sociedade. Além disso, estimular o processo contínuo e progressivo de melhoramento dos padrões e indicadores de 'acesso e de qualidade' que envolva a gestão, o processo de trabalho e os resultados alcançados pelas equipes de saúde da Atenção Básica, entre outros.

O PMAQ-AB está organizado em quatro fases, que se complementam, formando um ciclo contínuo de melhoria do acesso e da qualidade da Atenção Básica [46]:

1. Adesão da equipe ao PMAQ-AB;

2. Desenvolvimento de ações com foco na autoavaliação, no monitoramento de indicadores, na educação permanente e no apoio institucional;

3. Avaliação externa das equipes de saúde;

4. Recontratualização das equipes para o próximo ciclo.

3.8.3 Caderno de acolhimento à demanda espontânea

O caderno de acolhimento à demanda espontânea é uma iniciativa do Ministério da Saúde, com o objetivo de padronizar as ações da Atenção Básica no território nacional. O caderno está dividido em dois volumes, no qual o volume I trata do acolhimento contextualizado na gestão do processo de trabalho em saúde na atenção básica, tocando em aspectos centrais à sua implementação no cotidiano dos serviços. E o volume II, como desdobramento do primeiro, apresenta ofertas de abordagem de situações comuns no

acolhimento à demanda espontânea, utilizando-se do saber clínico, epidemiológico e da subjetividade, por meio do olhar para riscos e vulnerabilidades [47].

As equipes de atenção básica estão fortemente expostas à dinâmica cotidiana da vida das pessoas nos territórios. Nesse sentido, a capacidade de acolhida e escuta das equipes aos pedidos, demandas, necessidades e manifestações dos usuários é um elemento-chave. O caderno deve ser encarado como uma ferramenta de auxílio, oferecida pelo Ministério da Saúde, para a construção partilhada e cotidiana de modos de cuidar e gerir.

O caderno também mostra um fluxo padrão dos usuários nas UBS. O foco do fluxograma não é definir a ordem ou o local onde cada ação deve ser realizada, mas sim sinteticamente supor que usuários com atividades agendadas, como por exemplo consultas, devem ser recebidos e devidamente direcionados, a fim de evitar esperas desnecessárias com potencial de confusão na recepção. Observa também que os trabalhadores encarregados de escutar demandas que surgem espontaneamente devem ter capacidade de analisá-las, identificando riscos e analisando vulnerabilidade, clareza das ofertas de cuidado existentes na UBS, possibilidade de diálogo com outros colegas, algum grau de resolutividade e respaldo para acionar as ofertas de cuidado em tempos e modos que considerem a necessidade dos usuários. E por fim, reconhece que situações imprevistas são inerentes à vida e requerem certa organização da unidade e do processo de trabalho da equipe, tanto para compreendê-las quanto para intervir sobre elas [47].

Ressalta-se que o objetivo dos Cadernos de Atenção Básica não é burocratizar o acolhimento e o fluxo do usuário na unidade, senão ampliar a resolutividade e a capacidade de cuidado da equipe. Na atenção básica, os usuários geralmente moram perto das UBS, e que o efetivo trabalho em equipe produz relações solidárias e complementares entre os profissionais (enriquecendo-os individualmente e ao conjunto da equipe), gerando, assim, mais segurança e proteção para os usuários.

4 Procedimentos Metodológicos

Neste capítulo, são apresentados os procedimentos utilizados e as metodologias nas quais foram embasados para o desenvolvimento do trabalho.

4.1 Roteiro Metodológico

As etapas do projeto, que consiste em analisar a eficiência das unidades básicas de saúde do município de Florianópolis, estão citadas a seguir:

1. Obtenção e tratamento dos dados;

2. Caracterização da Atenção Básica do município;

3. Seleção de *inputs* e *outputs* para o modelo DEA;

4. Aplicação do modelo DEA nos dados levantados;

5. Avaliação dos escores de eficiência encontrados;

6. Relacionar as variáveis utilizadas na técnica de Floresta Aleatória com escores de eficiência encontrados no modelo DEA.

A primeira etapa pode ser caracterizada como uma pesquisa exploratória (Gil, 2008), que visa construir uma visão geral de um determinado fenômeno. Esse tipo de pesquisa é comumente utilizada como etapa inicial de um estudo mais amplo, como é o caso deste trabalho.

A segunda etapa é caracterizada como de natureza descritiva, pois procura descrever e caracterizar o nível da Atenção Básica no município e outros fatores qualitativos contextualizando o meio no qual o trabalho está inserido.

As etapas seguintes são caracterizadas como pesquisas explicativas, pois identificam fatores que alteram as eficiências das UBS's do município. Na visão de Gil (2008), esse tipo de pesquisa é a mais complexa, por causa do crescente risco de se cometer erros.

Para melhor visualização e entendimento do roteiro metodológico, elaborou-se a Figura 23, que resume as etapas realizadas no desenvolvimento do trabalho.

4.2 Coleta de dados

A Secretaria Municipal da Saúde do município de Florianópolis possui um sistema integrado e único de informações, que contém todas as entradas e saídas diárias de cada UBS. Essa integração dos dados facilitou o processo de coleta para o trabalho.

Figura 23 – Etapas do desenvolvimento

Para definir quais dados a serem selecionados, dentre todas as informações disponíveis das UBS's de Florianópolis, foram consultados o Caderno de Atenção Básica sobre Atendimento à Demanda Espontânea e a Política Nacional de Atenção Básica. Foram também utilizados os estudos apresentados no Capítulo 2, pois eles mostraram que as variáveis utilizadas em seus estudos de eficiência tiveram grande relação com a eficiência da DMU analisada e representaram de forma satisfatória o seu desempenho.

Tendo como base os trabalhos estudados no levantamento bibliométrico, e o conjunto de informações disponíveis no sistema da SMS do município de Florianópolis, foram recolhidos os seguintes dados relativos a cada uma das 50 UBS's da cidade, referente ao ano de 2017.

- Quadro de funcionários com especialidades no ano estudado;
- Plantas baixas com as medidas dos ambientes;
- Consumo de materiais e remédios;
- Número de atendimentos em todos os setores;
- Tempo de atendimento em horas de cada setor da UBS;
- Número de dias com atendimento em cada setor da UBS.

Através da uso da ferramenta Tabela Dinâmica no software *Excel*, organizou-se para cada UBS: as áreas de cada ambiente, o número de atendentes de cada especilidade, bem como o número de horas e dias com atendimento de cada especialidade, e por fim, a soma total de atendimentos no ano.

Em relação à infraestrutura das UBS's, pode-se separar em setores de: consultório, odontolgia, curativos, imunização, medicamento, espera e recepção, administração, enfermagem/triagem/nebulização, auditório, depósito material de enfermagem, esterilização, banheiro e espaço comum funcionário. Já em relação ao quadro de funcionários, levantou-se as seguintes especialidades: administrativo, assistente social, auxiliar de enfermagem, auxiliar de odontologia, avaliador físico, clínico geral, clínico geral PMM, enfermagem, farmácia, fisioterapia, técnico em eletrocardiograma, geriatria, médico família, nutricionista, odontologia s/ACD, odontolgia, ortopedia, pediatria, psicologia, psiquiatria, técnico higiene dental e técnico em laboratório.

Para as UBS's que não possuíam informação em algum setor de infraestrutura ou especialidade, foi atríbuido o valor zero à respectiva variável faltante. Porém, se a UBS não apresentava nenhuma informação de infraestrutura e/ou especialidade, essa unidade foi retirada da análise. As unidades do Continente e do Canto da Lagoa não dispunham de informações referentes às especilidades, e a unidade da Fazenda do Rio Tavares não dispunha de suas áreas. Já a unidade do Centro, que é integrada a uma policlínica, não possuia informações confiáveis de área, pois não era clara a distinção entre as áreas da UBS e da policlínica na planta baixa. Por causa disso, este trabalho retirou as quatro UBS's da análise.

4.3 Caracterização da saúde no município

As informações dos recursos utilizados e a saída, que está na forma de atendimento à população, de cada Unidade Básica de Saúde foram coletados diretamente na SMS do município de Florianópolis, que gerencia o sistema vigente em todas as unidades. Esse

sistema reúne todas as informações, eliminando a necessidade de aplicação de questionários ou outras formas de coleta de dados.

A distribuição e a organização das equipes de Saúde da Família são outros tipos de dados que podem ser úteis para a compreensão do funcionamento das UBS's. Os dados das regiões abrangidas por cada UBS e eSF são disponibilizados pelo município, em uma plataforma online. Além disso, o site da Prefeitura Municipal contém informações sobre políticas vigentes e indicadores de saúde medidos no município. Com isso, as caraterísticas de cada UBS são coletadas por meio do sistema integrado da SMS do município de Florianópolis, e as informações relacionadas à organização da saúde no município são retiradas online.

4.4 Seleção de variáveis *input* e *output*

A obtenção de 110 variáveis para um modelo com 46 DMU's (UBS's) exige que a etapa de seleção de variáveis *input* e *output*, a serem utilizadas, posteriormente, nos modelos DEA, seja criteriosa. Conforme citado no Capítulo 3.3.1, é desejável que o número de DMU's exceda o número de variáveis *input* e *output* combinados. Por causa disso, técnicas de redução de dimensionalidade foram utilizadas para se obter um modelo reduzido, porém fidedigno ao modelo original.

Essa etapa de seleção de variáveis de *input* e *output* foi feita com a composição de três técnicas, na qual cada uma delas serviu para refinar o processo. As três técnicas utilizadas podem ser vistas na Figura 24.

Figura 24 – Processo de seleção de variáveis *input* e *output* para o modelo DEA

4.5 Aplicação da Análise Envoltória de Dados

Após aplicar a Análise de Envelopamento de Dados e calcular os escores de eficiência VRS, CRS, DRS e IRS, é possível obter o escore de eficiência de escala (ESC) de cada DMU. A eficiência de escala é obtida através da decomposição da eficiência CRS nas componentes de eficiência de escala (ESC) e eficiência VRS. Ela mede a razão entre a quantidade de inputs mais apropriada para a escala de produção e a quantidade de inputs efetivamente usada pela unidade sob dado nível de produção. Quando a escala é pequena, os benefícios da economia de escala não estão sendo usufruídos pela unidade. Por outro lado, quando a escala é grande resulta na inabilidade de usar adequadamente os recursos.

Se as eficiências VRS e CRS são iguais, a DMU em questão opera em escala ótima e deve ser utilizada como *benchmark* pelas demais.

A natureza da escala de uma DMU qualquer é obtida com a comparação entre os escores de eficiência dos modelos VRS, IRS e DRS (FRIES, 2003). Se os coeficientes de eficiência técnica VRS e IRS forem iguais, a DMU opera numa escala com retornos crescentes, portanto abaixo da escala ótima de operação. Caso a eficiência VRS seja igual à DRS, a DMU encontra-se em escala de operação com retornos decrescentes, estando assim, acima da escala ótima. A partir das medidas apresentadas para a avaliação dos escores de eficiência, as DMU's podem ser classificadas em: "Ótimo", "Acima" ou "Abaixo" da escala de operação. Com isso, propõe-se medidas gerenciais que otimizem a eficiência das DMU's dadas como ineficientes pelo modelo DEA.

4.6 Avaliação 2º estágio com a utilização da técnica de floresta aleatória

A partir dos resultados do modelo DEA, são analisadas as influências de outras variáveis não utilizadas nas eficiências das UBS's. Apesar da seleção de *inputs* e *outputs* ter sido criteriosa, não é possível analisar em um modelo DEA todos os fatores que influenciam o desempenho das DMU's. Sendo assim, tentativas devem ser feitas para incluir aquelas que fazem sentido prático para o cenário sob investigação. Em um 2º estágio deste trabalho, utilizou-se a análise de treinamento supervisionado com a técnica de floresta aleatória, com o auxílio da linguagem Python, para assegurar a inclusão dessas variáveis.

Na análise de treinamento supervisionado com a técnica de floresta aleatória, utilizou-se o resultado calculado de eficiência VRS como variável dependente, com o objetivo de analisar as características que influenciam na eficiência de uma UBS, mas também aquelas que influenciam a adequação ao porte da mesma.

Como variáveis independentes, são utilizados todos os *inputs* recebidos pela Secretaria Municipal da Saúde do município de Florianópolis, exceto aqueles que já tinham sido incluídos no modelo DEA. Assim, todas as variáveis restantes podem ser comparadas aos resultados a fim de identificar qualquer relação que possam ter com o desempenho das unidades.

O fato de não se utilizar *outputs* na análise de treinamento supervisionado com a técnica de floresta aleatória deve-se à dificuldade, para a SMS, de agir sobre essas variáveis. Portanto, mesmo que valores de *outputs* mostrem relações com os escores de eficiência, a informação não é base para medidas gerenciais que possam ser realizadas e efetivas na melhoria da eficiência das unidades.

5 Desenvolvimento e Resultados

Neste capítulo são descritas todas as etapas realizadas para o desenvolvimento do trabalho de acordo com o roteiro apresentado no Capítulo 4. A apresentação dos resultados é detalhada desde o processo de seleção de variáveis, até a obtenção dos escores de eficiência DEA e comparação desses escores com as variáveis *inputs*, que não foram consideradas no modelo DEA, por meio da análise de treinamento supervisionado com a técnica de floresta aleatória. Ao final, são analisados os resultados das DMU's utilizadas.

5.1 Caracterização da saúde no município

Florianópolis é a capital brasileira mais bem avaliada na questão Atenção à Saúde Primária, segundo o Ministério da Saúde. 93% das equipes de Saúde da Família do município tiveram avaliação ótimo, muito bom e bom. Já a média das outras capitais é de 43%. Esses dados foram divulgados pelo Programa Nacional de Melhoria do Acesso e da Qualidade da Atenção Básica (PMAQ) [48].

Sobre a organização das Unidades Básicas de Saúde, de acordo com a Secretaria Municipal da Saúde, as 50 unidades são distribuídas contemplando a orientação do Ministério da Saúde que estabelece o número de até 4000 habitantes por equipe.

5.2 Seleção de variáveis *input* e *output*

Como apresentado no Capítulo 4.4, o procedimento para seleção das variáveis *input* e *output*, a serem utilizadas nos modelos DEA, foi feito por um processo constituído de três etapas de redução de dimensionalidade, as quais são detalhadas nos tópicos subsequentes. Essas etapas de redução de variáveis foram feitas porque para o modelo DEA ser discriminatório é necessário um número reduzido de variáveis em relação ao número de DMUs.

5.2.1 Análise de Correlação

Após a coleta dos dados na SMS do município de Florianópolis, sua organização em uma tabela dinâmica e divisão por setores, as 110 variáveis levantadas foram submetidas à análise de correlação. Nesse momento, não foi feita a distinção entre as variáveis que representavam *inputs* e aquelas que representavam *outputs*.

A análise de correlação foi implementada em linguagem Python, e retorna uma matriz simétrica quadrada de 110x110, na qual a diagonal principal assume valor 1. Essa

matriz mostra os índices de correlação entre as variáveis e, se esse índice for igual ou superior a 0.85, foi considerado que essas variáveis possuem uma forte correlação entre elas.

Dessa forma, o programa analisa de maneira automática todas as linhas e colunas da matriz de correlação, e exclui as variáveis que são fortemente correlacionadas. Notou-se que, para cada especialidade, os dias com atendimento revelaram trazer informações semelhantes aos dados de horas com atendimento e aos dados de número de atendimentos. Essa relação entre dias, horas e números já era esperada, e a maioria dos dados com "Dias com Atendimento" e "Horas com Atendimento" foram excluídos. A Figura 25 mostra as variáveis excluídas utilizando a análise de correlação.

No final desse processo, das 110 variáveis iniciais, apenas 63 foram mantidas e submetidas à análise de componentes principais.

5.2.2 Análise de Componentes Principais

O uso da análise de componentes principais viabiliza estimar quantos componentes são necessários para descrever os dados. Isso pode ser determinado observando-se a taxa de variação cumulativa explicada como uma função do número de componentes mostrada na Figura 26.

A curva quantifica o quanto da variância total de 63 dimensões está contida nas primeiras componentes. Nota-se que as 10 primeiras contêm pouco mais de 71% da variabilidade observada nos dados. Apesar desse valor não ser 100%, os acréscimos de variância tornam-se, à medida que aumenta o número de componentes, menos significativos. Portanto, a prioridade foi obter um menor número de componentes.

O cálculo da ACP retorna uma matriz de autovalores, que indicam o grau de importância da variável analisada naquela componente principal. Portanto, foram selecionadas apenas as variáveis que possuiam um alto autovalor em alguma componente principal. Assim, têm-se as informações mais significativas de cada componente. O valor de corte para o autovalor foi, arbitrariamente escolhido, ≥ 0.194, resultando em 29 variáveis, ilustradas na Figura 27.

5.2.3 Método iterativo do cálculo de entropia

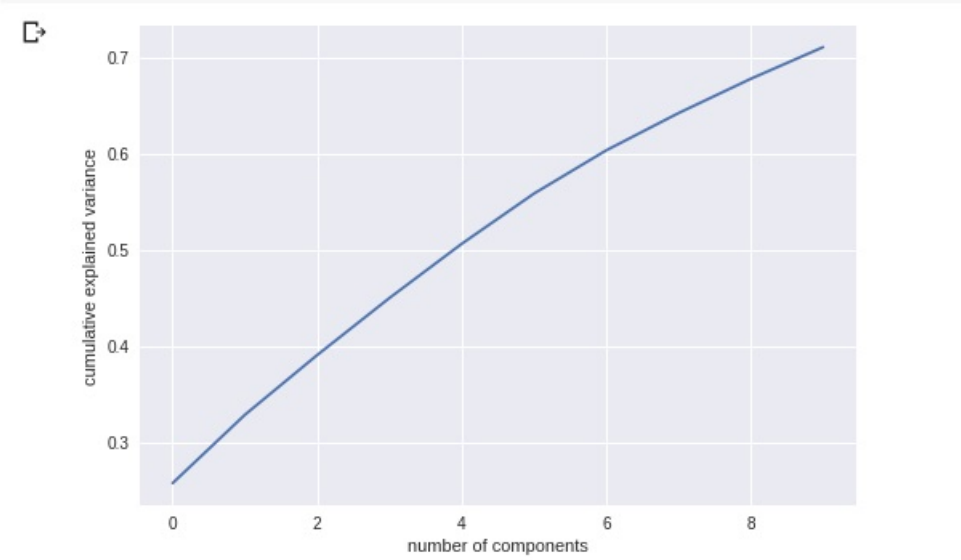
Jenkins e Anderson (2003) citam que quanto maior o número de *inputs* e *outputs* em um modelo DEA, maior é a dimensionalidade do espaço de programação linear e menos discriminante é a análise, resultando em uma grande quantidade de DMUs com escore de eficiência máxima. Por esse motivo, foi desenvolvido, em linguagem Phyton, um programa que combina variáveis de *input*, com combinações de 2 a 8, e aplica o modelo DEA com retorno variável de escala (VRS), orientado para o *input*.

```
['2 - Atendimentos - Administrativo',
 '3 - Dias Atendimento - Administrativo',
 '4 - Horas Atendimento - Administrativo',
 '7 - Dias Atendimento - Assistente Social',
 '8 - Horas Atendimento - Assistente Social',
 '11 - Dias Atendimento - Auxiliar de Enfermagem',
 '12 - Horas Atendimento - Auxiliar de Enfermagem',
 '15 - Dias Atendimento - Auxiliar de Odontologia',
 '16 - Horas Atendimento - Auxiliar de Odontologia',
 '19 - Dias Atendimento - Avaliador Fisico',
 '20 - Horas Atendimento - Avaliador Fisico',
 '23 - Dias Atendimento - Clinico Geral',
 '24 - Horas Atendimento - Clinico Geral',
 '27 - Dias Atendimento - Clinico Geral PMM',
 '28 - Horas Atendimento - Clinico Geral PMM',
 '30 - Atendimentos - Eletrocardiograma Tec.Enf.',
 '31 - Dias Atendimento - Eletrocardiograma Tec.Enf.',
 '32 - Horas Atendimento - Eletrocardiograma Tec.Enf.',
 '35 - Dias Atendimento - Enfermagem',
 '36 - Horas Atendimento - Enfermagem',
 '95 - Total Dias Atendimento',
 '96 - Total Horas Atendimento',
 '39 - Dias Atendimento - Farmacia',
 '40 - Horas Atendimento - Farmacia',
 '89 - Num Funcionarios - Tecnico em Laboratorio',
 '90 - Atendimentos - Tecnico em Laboratorio',
 '91 - Dias Atendimento - Tecnico em Laboratorio',
 '92 - Horas Atendimento - Tecnico em Laboratorio',
 '102 - Area Total Medicamento ',
- '44 - Horas Atendimento - Fisioterapia',
 '47 - Dias Atendimento - Geriatria',
 '48 - Horas Atendimento - Geriatria',
 '51 - Dias Atendimento - Medico Familia',
 '52 - Horas Atendimento - Medico Familia',
 '56 - Horas Atendimento - Nutricionista',
 '59 - Dias Atendimento - Odont s/ACD',
 '60 - Horas Atendimento - Odont s/ACD',
 '63 - Dias Atendimento - Odontologia',
 '64 - Horas Atendimento - Odontologia',
 '67 - Dias Atendimento - Ortopedia',
 '68 - Horas Atendimento - Ortopedia',
 '71 - Dias Atendimento - Pediatria',
 '72 - Horas Atendimento - Pediatria',
 '76 - Horas Atendimento - Psicologia',
 '79 - Dias Atendimento - Psiquiatria',
 '80 - Horas Atendimento - Psiquiatria',
 '83 - Dias Atendimento - Tec Higiene Dental',
 '84 - Horas Atendimento - Tec Higiene Dental',
 '87 - Dias Atendimento - Tecnico de Enfermagem',
 '88 - Horas Atendimento - Tecnico de Enfermagem']
```

Figura 25 – Variáveis excluídas na análise de correlação

Primeiramente, foi necessário definir quais das 29 variáveis selecionadas pela análise ACP representavam *inputs* e quais eram *outputs*. Como *inputs*, foram utilizadas aquelas que possuíam informações sobre o número de funcionários de alguma especialidade e também sobre as áreas dos ambientes especificados, resultando em 19 *inputs*. Para os *outputs*, foram considerados os dados relativos aos atendimentos, sendo então 10 *outputs* selecionados.

Para que esta etapa utilizasse apenas variáveis relevantes para o modelo, outra




```
array([25.74, 32.85, 39.06, 44.99, 50.63, 55.85, 60.35, 64.22, 67.79,
       71.08])
```

Figura 26 – Variância acumulada explicada pela ACP

```
['10 - Atendimentos - Auxiliar de Enfermagem',
 '100 - Area Total Curativos',
 '101 - Area Total Imunizacao',
 '103 - Area Total Espera e Recepcao',
 '106 - Area Total Auditorio',
 '108 - Area Total Esterilizacao',
 '14 - Atendimentos - Auxiliar de Odontologia',
 '18 - Atendimentos - Avaliador Fisico',
 '25 - Num Funcionarios - Clinico Geral PMM',
 '26 - Atendimentos - Clinico Geral PMM',
 '29 - Num Funcionarios - Eletrocardiograma Tec.Enf.',
 '33 - Num Funcionarios - Enfermagem',
 '34 - Atendimentos - Enfermagem',
 '41 - Num Funcionarios - Fisioterapia',
 '45 - Num Funcionarios - Geriatria',
 '50 - Atendimentos - Medico Familia',
 '53 - Num Funcionarios - Nutricionista',
 '57 - Num Funcionarios - Odont s/ACD',
 '58 - Atendimentos - Odont s/ACD',
 '61 - Num Funcionarios - Odontologia',
 '62 - Atendimentos - Odontologia',
 '65 - Num Funcionarios - Ortopedia',
 '69 - Num Funcionarios - Pediatria',
 '74 - Atendimentos - Psicologia',
 '81 - Num Funcionarios - Tec Higiene Dental',
 '93 - Total Num Funcionarios',
 '94 - Total Atendimentos',
 '97 - Area Total Consultorio',
 '98 - Area Total Odontologia']
```

Figura 27 – Variáveis selecionadas pela ACP

redução foi feita analisando os dados de cada variável nas DMU's com o objetivo de evitar a utilização de alguma característica presente em menos de metade das Unidades Básicas de Saúde (Figura 28). Com isso, das 29 variáveis restantes da ACP, foram excluídas aquelas

que apareciam em menos de 50% das DMU's, como mostrado na Figura 29. Ao final, restaram apenas 19 variáveis, sendo elas, 12 *inputs* e 7 *outputs*.

108 - Area Total Esterilizacao	100.000000
53 - Num Funcionarios - Nutricionista	100.000000
98 - Area Total Odontologia	100.000000
97 - Area Total Consultorio	100.000000
94 - Total Atendimentos	100.000000
93 - Total Num Funcionarios	100.000000
74 - Atendimentos - Psicologia	100.000000
33 - Num Funcionarios - Enfermagem	100.000000
34 - Atendimentos - Enfermagem	100.000000
41 - Num Funcionarios - Fisioterapia	100.000000
101 - Area Total Imunizacao	97.826087
103 - Area Total Espera e Recepcao	95.652174
100 - Area Total Curativos	93.478261
18 - Atendimentos - Avaliador Fisico	93.478261
61 - Num Funcionarios - Odontologia	91.304348
62 - Atendimentos - Odontologia	91.304348
69 - Num Funcionarios - Pediatria	91.304348
50 - Atendimentos - Medico Familia	89.130435
14 - Atendimentos - Auxiliar de Odontologia	50.000000
45 - Num Funcionarios - Geriatria	41.304348
58 - Atendimentos - Odont s/ACD	39.130435
57 - Num Funcionarios - Odont s/ACD	39.130435
26 - Atendimentos - Clinico Geral PMM	36.956522
25 - Num Funcionarios - Clinico Geral PMM	36.956522
106 - Area Total Auditorio	34.782609
10 - Atendimentos - Auxiliar de Enfermagem	30.434783
81 - Num Funcionarios - Tec Higiene Dental	15.217391
65 - Num Funcionarios - Ortopedia	4.347826
29 - Num Funcionarios - Eletrocardiograma Tec.Enf.	2.173913

Figura 28 – Porcentagem da frequência de cada variável nas DMU's

108 - Area Total Esterilizacao	100.000000
41 - Num Funcionarios - Fisioterapia	100.000000
98 - Area Total Odontologia	100.000000
97 - Area Total Consultorio	100.000000
94 - Total Atendimentos	100.000000
93 - Total Num Funcionarios	100.000000
74 - Atendimentos - Psicologia	100.000000
33 - Num Funcionarios - Enfermagem	100.000000
53 - Num Funcionarios - Nutricionista	100.000000
34 - Atendimentos - Enfermagem	100.000000
101 - Area Total Imunizacao	97.826087
103 - Area Total Espera e Recepcao	95.652174
100 - Area Total Curativos	93.478261
18 - Atendimentos - Avaliador Fisico	93.478261
69 - Num Funcionarios - Pediatria	91.304348
61 - Num Funcionarios - Odontologia	91.304348
62 - Atendimentos - Odontologia	91.304348
50 - Atendimentos - Medico Familia	89.130435
14 - Atendimentos - Auxiliar de Odontologia	50.000000

Figura 29 – Variáveis presentes em mais de 50% do conjunto de DMU's

Para os *outputs*, entretanto, não seria possível utilizar as sete variáveis mantidas até esta etapa da redução de dimensionalidade. Com isso, devido às exigências da PNAB, que reconhece como obrigatória em uma UBS a presença de profissionais como médicos

de família e enfermeiros, foram rodados dois modelos. O Modelo 1 considerando o "Total de Atendimentos", visto que representa a soma dos atendimentos da DMU e Modelo 2 com dois *outputs*, sendo eles, "Atendimentos – Enfermagem" e "Atendimentos – Médico Família". Com isso, é possível analisar qual dos *outputs* resulta em um modelo com maior discriminação.

Para cada um dos modelos, diferenciados pelos *outputs* utilizados, foram processados 3.784 modelos DEA. Obteve-se então os escores de eficiência de todas as 46 DMU's, além da entropia de Shannon de cada DEA, que será utilizada na escolha dos modelos a serem analisados na seleção de variáveis. De acordo com Shannon [], os modelos que apresentam maior entropia, são os que carregam maior incerteza e, portanto, maior discriminação entre as UBS's. Com os resultados do modelo, foi analisado o número de DMU's consideradas eficientes pelo número de variáveis combinadas (Figuras 30 e 31).

combinações eficientes	2	3	4	5	6	7	8	Total eficientes
3	4	0	0	0	0	0	0	4.0
4	6	1	0	0	0	0	0	7.0
5	10	6	0	0	0	0	0	16.0
6	8	14	2	0	0	0	0	24.0
7	4	15	8	0	0	0	0	27.0
8	3	9	14	2	0	0	0	28.0
9	1	18	17	6	0	0	0	42.0
10	2	8	13	8	1	0	0	32.0
11	4	17	23	11	1	0	0	56.0
12	2	9	26	20	5	0	0	62.0
13	1	12	28	23	8	0	0	72.0
14	0	7	30	35	12	2	0	86.0
15	0	2	25	34	16	3	0	80.0
16	0	0	8	41	29	3	0	81.0
17	0	2	7	25	31	12	0	77.0
18	0	0	6	23	33	14	2	78.0
19	0	0	2	11	32	19	4	68.0
20	0	0	1	10	21	25	5	62.0
21	0	0	0	2	14	22	11	49.0
22	0	0	0	1	5	12	11	29.0
23	0	0	0	0	2	7	9	18.0
24	0	0	0	0	0	1	3	4.0
28	5	6	4	1	0	0	0	16.0
29	3	13	16	11	3	0	0	46.0
30	1	15	35	32	17	4	0	104.0
31	1	7	30	40	21	7	1	107.0
32	9	35	82	130	108	39	6	409.0
33	0	4	28	85	137	107	35	396.0
34	1	10	41	95	147	153	94	541.0
35	0	0	4	26	71	110	104	315.0
37	1	9	34	67	76	51	19	257.0
38	0	0	2	16	43	56	39	156.0
39	0	1	9	37	91	145	152	435.0
Total combinações	66	220	495	792	924	792	495	3784.0

Figura 30 – Número de conjuntos de variáveis *input* e *output* gerados por número de DMU's com eficiência iguais a um e por número de variáveis combinadas no Modelo 1

combinações eficientes	2	3	4	5	6	7	8	Total eficientes
5	2	0	0	0	0	0	0	2.0
6	4	0	0	0	0	0	0	4.0
7	3	2	0	0	0	0	0	5.0
8	5	1	0	0	0	0	0	6.0
9	3	5	1	0	0	0	0	9.0
10	9	10	2	0	0	0	0	21.0
11	2	4	1	0	0	0	0	7.0
12	2	11	6	1	0	0	0	20.0
13	6	14	8	0	0	0	0	28.0
14	4	11	12	4	0	0	0	31.0
15	3	10	13	3	0	0	0	29.0
16	0	15	21	8	0	0	0	44.0
17	0	8	27	18	4	0	0	57.0
18	1	7	13	15	3	0	0	39.0
19	1	9	18	21	7	0	0	56.0
20	0	4	17	23	14	1	0	59.0
21	0	4	20	23	14	5	0	66.0
22	0	2	14	22	18	4	0	60.0
23	0	3	14	18	14	7	0	56.0
24	0	0	9	24	19	11	4	67.0
25	0	0	8	24	19	6	0	57.0
26	0	0	4	16	20	11	4	55.0
27	0	0	1	18	24	10	3	56.0
28	5	4	2	8	23	16	2	60.0
29	2	10	10	6	15	18	9	70.0
30	0	5	12	13	12	11	3	56.0
31	2	7	10	10	11	14	8	62.0
32	5	17	36	37	20	10	9	134.0
33	4	21	37	41	34	17	6	160.0
34	1	12	50	58	30	8	1	160.0
35	1	7	40	105	104	55	16	328.0
36	0	5	17	56	100	72	24	274.0
37	1	8	29	52	72	78	47	287.0
38	0	3	25	71	98	79	43	319.0
39	0	1	16	66	131	142	85	441.0
40	0	0	2	25	82	130	118	357.0
41	0	0	0	6	32	69	79	186.0
42	0	0	0	0	4	18	34	56.0
Total combinações	66	220	495	792	924	792	495	3784.0

Figura 31 – Número de conjuntos de variáveis *input* e *output* gerados por número de DMU's com eficiência iguais a um e por número de variáveis combinadas no Modelo 2

O próximo passo foi analisar, em cada grupo de modelos, separado pelo número de variáveis e pelo número de DMU's eficientes, quais apresentaram a maior entropia em relação às demais. Pode-se esperar que as combinações de variáveis *input* e *output* com alto número de DMU's eficientes apresentem baixa entropia, pois a probabilidade de ocorrer um escore de eficiência igual a um é alta, mostrando que a combinação de variáveis *input* e *output* é pouco discriminatória. Por esse motivo, foram analisadas as entropias máximas para cada situação de número de variáveis combinadas e número de DMU's eficientes, como mostram as Figuras 32 e 33. Como esperado, as maiores entropias foram encontradas nas combinações de variáveis *input* e *output* com menor número de DMU's com eficiências iguais a um.

combinações eficientes	2	3	4	5	6	7	8
3	2.57797	0.0	0.0	0.0	0.0	0.0	0.0
4	3.06375	2.29037	0.0	0.0	0.0	0.0	0.0
5	3.00512	2.6777	0.0	0.0	0.0	0.0	0.0
6	3.08268	3.03401	2.36395	0.0	0.0	0.0	0.0
7	2.88836	3.03759	2.61783	0.0	0.0	0.0	0.0
8	2.84368	3.03794	2.6295	2.32521	0.0	0.0	0.0
9	2.36703	2.8975	2.54095	2.42153	0.0	0.0	0.0
10	2.80687	2.81782	2.69719	2.1933	2.11565	0.0	0.0
11	2.75066	2.85563	2.68597	2.19503	1.85062	0.0	0.0
12	2.74066	2.65877	2.73394	2.40708	2.09484	0.0	0.0
13	2.55581	2.65871	2.6329	2.55119	2.25375	0.0	0.0
14	0.0	2.75	2.68926	2.46907	2.191	2.04376	0.0
15	0.0	2.50869	2.59951	2.35455	2.1047	1.79605	0.0
16	0.0	0.0	2.38219	2.38058	2.20827	1.80405	0.0
17	0.0	2.40174	2.34411	2.31436	2.21562	2.04597	0.0
18	0.0	0.0	2.23065	2.31977	2.18667	2.04211	1.54723
19	0.0	0.0	2.23317	2.21277	2.10814	2.02552	1.82029
20	0.0	0.0	2.14166	2.04377	2.02672	1.88563	1.68137
21	0.0	0.0	0.0	1.99019	1.96473	1.84945	1.73551
22	0.0	0.0	0.0	1.94582	1.91517	1.81448	1.5527
23	0.0	0.0	0.0	0.0	1.77593	1.69438	1.59318
24	0.0	0.0	0.0	0.0	0.0	1.52356	1.43717
28	1.58429	1.72693	1.62325	1.49194	0.0	0.0	0.0
29	1.67161	1.78527	1.75244	1.6056	1.42777	0.0	0.0
30	1.55995	1.78527	1.76668	1.63879	1.56129	1.32056	0.0
31	1.47771	1.61965	1.71726	1.66465	1.55063	1.41009	1.16965
32	1.61107	1.70813	1.63759	1.56759	1.46205	1.33914	1.18921
33	0.0	1.54052	1.61107	1.54052	1.48063	1.3502	1.25661
34	1.29258	1.4425	1.48598	1.50239	1.40968	1.38045	1.31207
35	0.0	0.0	1.35998	1.35998	1.31075	1.30009	1.16965
37	1.01259	1.05607	1.20291	1.20291	1.14303	1.08314	0.997994
38	0.0	0.0	1.04147	1.10136	1.08495	1.05788	0.970926
39	0.0	0.868547	0.868547	0.912025	0.912025	0.912025	0.912025

Figura 32 – Entropia máxima por número de DMU's com eficiências iguais a 1 e por número de variáveis combinadas no Modelo 1

combinações eficientes	2	3	4	5	6	7	8
5	2.41739	0.0	0.0	0.0	0.0	0.0	0.0
6	2.9076	0.0	0.0	0.0	0.0	0.0	0.0
7	2.82096	2.59735	0.0	0.0	0.0	0.0	0.0
8	2.96839	2.63123	0.0	0.0	0.0	0.0	0.0
9	2.85525	2.55877	2.0634	0.0	0.0	0.0	0.0
10	2.95325	2.66461	2.33022	0.0	0.0	0.0	0.0
11	2.45461	2.80037	2.33351	0.0	0.0	0.0	0.0
12	2.80731	2.76286	2.41657	1.73551	0.0	0.0	0.0
13	2.74588	2.68401	2.51347	0.0	0.0	0.0	0.0
14	2.77895	2.64037	2.41288	2.0634	0.0	0.0	0.0
15	2.70042	2.51667	2.32728	1.92523	0.0	0.0	0.0
16	0.0	2.5317	2.45414	2.0356	0.0	0.0	0.0
17	0.0	2.58093	2.4044	2.15652	1.65906	0.0	0.0
18	2.31578	2.31487	2.35042	2.23962	1.82514	0.0	0.0
19	2.21318	2.48177	2.19601	2.09818	1.70179	0.0	0.0
20	0.0	2.46122	2.15055	2.08274	1.93008	1.60492	0.0
21	0.0	2.37277	2.32223	2.01921	1.79681	1.54571	0.0
22	0.0	2.23626	2.33146	2.02769	1.84029	1.51771	0.0
23	0.0	2.03528	2.14082	2.10366	1.75735	1.65398	0.0
24	0.0	0.0	2.1468	2.04548	1.79169	1.51289	1.42103
25	0.0	0.0	2.01266	2.08474	1.83171	1.58343	0.0
26	0.0	0.0	1.74384	1.77884	1.68322	1.51556	1.15107
27	0.0	0.0	1.7068	1.76583	1.76093	1.57545	1.41074
28	1.86095	1.86887	1.58683	1.65737	1.66549	1.41782	1.09958
29	1.71726	1.77285	1.72937	1.47216	1.50745	1.49194	1.36085
30	0.0	1.75244	1.69831	1.55912	1.41009	1.30455	1.27173
31	1.41782	1.49704	1.61172	1.43085	1.37196	1.30415	1.18921
32	1.5777	1.58618	1.63759	1.48282	1.33914	1.2402	1.13819
33	1.48063	1.52194	1.54901	1.55118	1.36085	1.2402	1.12971
34	1.42392	1.4425	1.48598	1.39903	1.23803	0.952702	0.556663
35	1.19455	1.38704	1.40346	1.3165	1.28151	1.16743	1.09689
36	0.0	1.26014	1.24639	1.24373	1.22732	1.16743	1.0783
37	1.01259	1.24639	1.24639	1.14303	1.09955	1.09955	1.08314
38	0.0	0.981583	1.10136	1.10136	1.05788	0.981583	0.955504
39	0.0	0.825069	0.998982	1.04246	0.998982	0.955504	0.895615
40	0.0	0.0	0.792431	0.85232	0.85232	0.85232	0.808842
41	0.0	0.0	0.0	0.661396	0.704874	0.748353	0.704874
42	0.0	0.0	0.0	0.0	0.556663	0.600142	0.600142

Figura 33 – Entropia máxima por número de DMU's com eficiências iguais a 1 e por número de variáveis combinadas no Modelo 2

Por se ter grande número de combinação de variáveis *input* e *output* com alta entropia, é possível encontrar diversos combinações com boa qualidade quanto à discriminação dos dados. Por conta disso, foi feita análise de distribuição de frequência para cada modelo, como mostram as Figuras 34 e 35, com o objetivo de reduzir esse número de combinações de variáveis *input* e *output*. Os dados são divididos em intervalos de classes, denominados *bins*, em que cada intervalo mostra a frequência em que cada valor diferente em um conjunto de dados ocorre. Para determinar o limite dos intervalos, utiliza-se os valores de entropia máxima e mínima calculada em cada modelo, e cada intervalo possui a mesma amplitude. Se a distribuição dos dados acontece nos últimos intervalos, maior é a frequência de valores altos de entropia na combinação de variáveis *input* e *output* analisada. Portanto, a seleção levou em consideração os conjuntos que possuíam maior número de combinações de variáveis e maiores valores atribuídos no último intervalo da distribuição (maiores entropias).

		bins1	bins2	bins3	bins4	bins5	bins6	bins7	bins8	bins9
combinações	eficientes									
2	**3**	0	0	0	0	0	0	2	2	0
	4	0	0	0	0	1	1	1	2	1
	5	0	0	0	0	0	0	5	1	4
	6	0	0	0	0	0	0	0	2	6
	7	0	0	0	0	0	0	2	1	1
	8	0	0	0	0	0	0	0	1	2
	9	0	0	0	0	0	0	1	0	0
	10	0	0	0	0	0	0	0	1	1
	11	0	0	0	0	0	0	1	3	0
	12	0	0	0	0	0	0	1	1	0
	13	0	0	0	0	0	0	0	1	0
3	**4**	0	0	0	0	0	0	1	0	0
	5	0	0	0	0	0	0	3	3	0
	6	0	0	0	0	0	1	6	5	2
	7	0	0	0	0	0	0	7	5	3
	8	0	0	0	0	0	1	4	3	1
	9	0	0	0	0	0	1	5	9	3
	10	0	0	0	0	0	0	4	2	2
	11	0	0	0	0	0	0	6	8	3
	12	0	0	0	0	0	0	3	6	0
	13	0	0	0	0	0	1	6	5	0
	14	0	0	0	0	0	0	3	4	0
	15	0	0	0	0	0	0	1	1	0
	17	0	0	0	0	0	0	2	0	0
4	**6**	0	0	0	0	0	0	2	0	0
	7	0	0	0	0	0	2	4	2	0
	8	0	0	0	0	0	3	8	3	0
	9	0	0	0	0	1	6	8	2	0
	10	0	0	0	0	0	3	9	1	0
	11	0	0	0	0	0	4	13	6	0
...	**...**	...	...	...	...	...	...	...	...	...

Figura 34 – Distribuição de frequência para os diversos *bins* dos valores de entropia, exemplarmente mostrado para as combinações 2, 3 e 4 do Modelo 1

		bins1	bins2	bins3	bins4	bins5	bins6	bins7	bins8	bins9
combinações	eficientes									
2	**5**	0	0	0	0	0	0	1	1	0
	6	0	0	0	0	0	0	0	2	2
	7	0	0	0	0	0	0	1	1	1
	8	0	0	0	0	0	0	1	2	2
	9	0	0	0	0	0	0	0	0	3
	10	0	0	0	0	0	0	3	2	4
	11	0	0	0	0	0	0	1	1	0
	12	0	0	0	0	0	0	0	0	2
	13	0	0	0	0	0	0	2	1	3
	14	0	0	0	0	0	0	1	2	1
	15	0	0	0	0	0	0	0	2	1
	18	0	0	0	0	0	0	1	0	0
	19	0	0	0	0	0	0	1	0	0
	28	0	0	0	0	4	1	0	0	0
3	**7**	0	0	0	0	0	0	1	1	0
	8	0	0	0	0	0	0	0	1	0
	9	0	0	0	0	0	1	2	2	0
	10	0	0	0	0	0	0	5	5	0
	11	0	0	0	0	0	0	1	2	1
	12	0	0	0	0	0	2	5	3	1
	13	0	0	0	0	0	2	7	4	1
	14	0	0	0	0	0	1	6	4	0
	15	0	0	0	0	0	4	3	3	0
	16	0	0	0	0	0	3	7	5	0
	17	0	0	0	0	0	2	1	5	0
	18	0	0	0	0	0	3	4	0	0
	19	0	0	0	0	0	1	5	3	0
	20	0	0	0	0	0	1	2	1	0
	21	0	0	0	0	0	0	4	0	0
	22	0	0	0	0	0	1	1	0	0
...	**...**	...	...	...	...	...	...	...	...	...

Figura 35 – Distribuição de frequência para os diversos *bins* dos valores de entropia, exemplarmente mostrado para as combinações 2 e 3 do Modelo 2

Ao final desse processo, para o Modelo 1 foram selecionadas 143 combinações de variáveis *input* e *output*, e para o Modelo 2, apenas 42. Para a escolha final das variáveis *input* e *output* a serem utilizadas na Análise Envoltória de Dados, foi observado como positivo a utilização de variáveis com dimensões diferentes. Isso significa, que combinações que continham apenas dados de infraestrutura, ou apenas dados do quadro de funcionários fossem descartadas. Com isso, foram escolhidos três diferentes conjuntos de *input* e *output*, que representam as Unidades Básicas de Saúde do município de Florianópolis, vistas nas Figuras 36, 37 e 38.

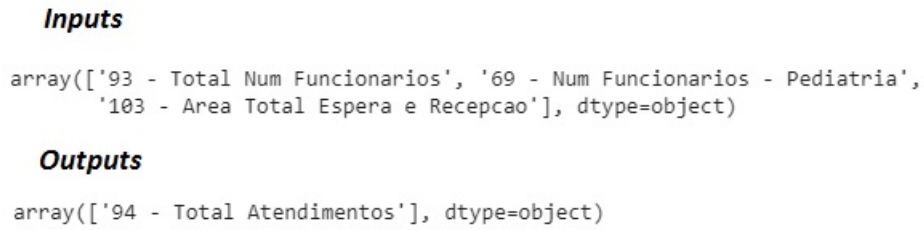


Figura 36 – Combinação selecionada com 3 *inputs* e 1 *output*

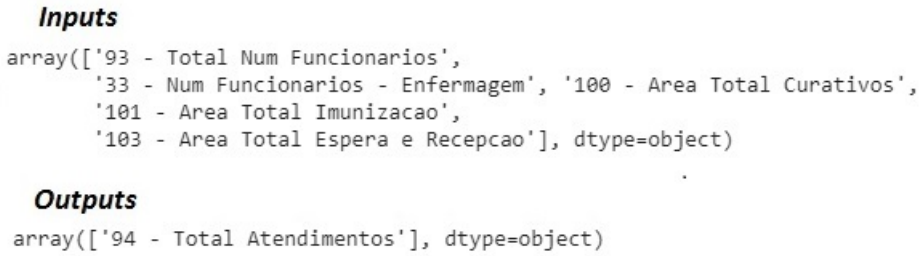


Figura 37 – Combinação selecionada com 5 *inputs* e 1 *output*

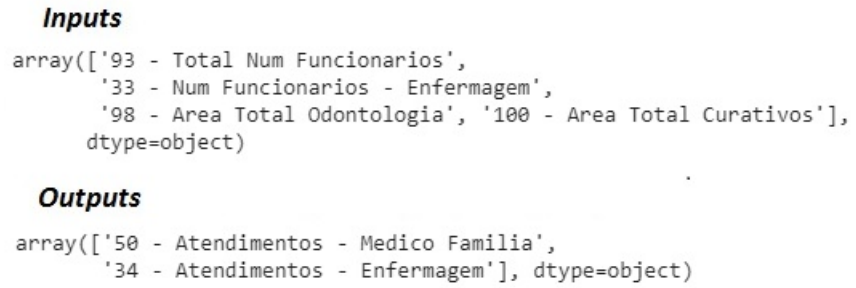


Figura 38 – Combinação selecionada com 4 *inputs* e 2 *outputs*

Com as variáveis de *inputs* e *outputs* definidas, segue-se para a etapa de aplicação dos modelos DEA e análises das eficiências apuradas.

5.3 Análise dos escores de eficiência

Para cada DMU (UBS), a partir da combinação de variáveis *inputs* e *outputs*, foram calculados os escores de eficiência VRS, CRS, IRS e DRS para cada um dos três modelos selecionados, como mostram as Figuras 39, 40 e 41. A eficiência de escala (ESC), que identifica se a UBS está em escala ótima de operação, foi obtida pela relação entre as eficiências dos modelos CRS e VRS. As unidades que não se encontram na escala ótima, podem ou estar acima ou abaixo desta situação, considerando tanto a magnitude de *inputs* quanto de *outputs*. Essa medida é feita pela comparação das eficiências VRS com DRS e IRS, sendo que se a VRS é igual à DRS, assume-se que a DMU opera acima da escala ótima e caso possua valor igual à IRS, opera abaixo da escala ótima, o que significa que está subdimensionada.

	VRS	CRS	DRS	IRS	ESC	Situação
UBS ABRAÃO	0.671	0.478	0.478	0.671	0.713	Abaixo
UBS AGRONÔMICA	0.666	0.600	0.600	0.666	0.900	Abaixo
UBS ALTO RIBEIRÃO	0.752	0.318	0.318	0.752	0.422	Abaixo
UBS ARMAÇÃO	0.596	0.420	0.420	0.596	0.704	Abaixo
UBS BALNEÁRIO	1.000	1.000	1.000	1.000	1.000	Ótima
UBS BARRA DA LAGOA	0.753	0.524	0.524	0.753	0.697	Abaixo
UBS CACHOEIRA BOM JESUS	1.000	1.000	1.000	1.000	1.000	Ótima
UBS CAIEIRA BARRA DO SUL	0.969	0.233	0.233	0.969	0.241	Abaixo
UBS CAMPECHE	0.813	0.611	0.611	0.813	0.751	Abaixo
UBS CANASVIEIRAS	1.000	1.000	1.000	1.000	1.000	Ótima
UBS CAPOEIRAS	0.695	0.481	0.481	0.695	0.693	Abaixo
UBS CARIANOS	0.532	0.364	0.364	0.532	0.685	Abaixo
UBS COLONINHA	0.697	0.591	0.591	0.697	0.848	Abaixo
UBS COQUEIROS	0.854	0.694	0.694	0.854	0.812	Abaixo
UBS CÓRREGO GRANDE	0.814	0.564	0.564	0.814	0.693	Abaixo
UBS COSTA DA LAGOA	1.000	0.273	0.273	1.000	0.273	Abaixo
UBS COSTEIRA DO PIRAJUBAÉ	0.741	0.679	0.679	0.741	0.916	Abaixo
UBS ESTREITO	1.000	0.679	1.000	0.679	0.679	Acima
UBS INGLESES	0.902	0.888	0.902	0.888	0.984	Acima
UBS ITACORUBI	0.748	0.649	0.649	0.748	0.867	Abaixo
UBS JARDIM ATLÂNTICO	0.920	0.406	0.406	0.920	0.442	Abaixo
UBS JOÃO PAULO	0.731	0.561	0.561	0.731	0.768	Abaixo
UBS JURERE	0.787	0.408	0.408	0.787	0.518	Abaixo
UBS LAGOA DA CONCEIÇÃO	0.519	0.412	0.412	0.519	0.793	Abaixo
UBS MONTE CRISTO	0.930	0.841	0.930	0.841	0.904	Acima
UBS MONTE SERRAT	0.804	0.656	0.656	0.804	0.817	Abaixo
UBS MORRO DAS PEDRAS	0.610	0.334	0.334	0.610	0.547	Abaixo
UBS NOVO CONTINENTE	0.691	0.557	0.557	0.691	0.805	Abaixo
UBS PANTANAL	0.679	0.373	0.373	0.679	0.549	Abaixo
UBS PANTANO DO SUL	0.824	0.465	0.465	0.824	0.564	Abaixo
UBS PONTA DAS CANAS	0.870	0.488	0.488	0.870	0.561	Abaixo
UBS PRAINHA	0.563	0.471	0.471	0.563	0.836	Abaixo
UBS RATONES	0.589	0.244	0.244	0.589	0.413	Abaixo
UBS RIBEIRÃO DA ILHA	0.676	0.238	0.238	0.676	0.352	Abaixo
UBS RIO TAVARES	0.394	0.292	0.292	0.394	0.740	Abaixo
UBS RIO VERMELHO	0.861	0.788	0.861	0.788	0.915	Acima
UBS SACO DOS LIMÕES	0.749	0.547	0.547	0.749	0.730	Abaixo
UBS SACO GRANDE	1.000	1.000	1.000	1.000	1.000	Ótima
UBS SANTINHO	1.000	0.758	0.758	1.000	0.758	Abaixo
UBS SANTO ANTONIO LISBOA	0.574	0.419	0.419	0.574	0.730	Abaixo
UBS SAPE	0.582	0.361	0.361	0.582	0.620	Abaixo
UBS TAPERA	0.740	0.729	0.729	0.740	0.985	Abaixo
UBS TRINDADE	0.607	0.533	0.533	0.607	0.878	Abaixo
UBS VARGEM GRANDE	0.535	0.284	0.284	0.535	0.530	Abaixo
UBS VARGEM PEQUENA	0.793	0.227	0.227	0.793	0.286	Abaixo
UBS VILA APARECIDA	0.982	0.387	0.387	0.982	0.394	Abaixo

Figura 39 – Escores de eficiências das UBS's utilizando a combinação de variáveis com 3 *inputs* e 1 *output*

	VRS	CRS	DRS	IRS	ESC	Situação
UBS ABRAÃO	0.734	0.562	0.562	0.734	0.766	Abaixo
UBS AGRONÔMICA	0.661	0.600	0.600	0.661	0.908	Abaixo
UBS ALTO RIBEIRÃO	0.746	0.318	0.318	0.746	0.426	Abaixo
UBS ARMAÇÃO	0.576	0.430	0.430	0.576	0.746	Abaixo
UBS BALNEÁRIO	1.000	1.000	1.000	1.000	1.000	Ótima
UBS BARRA DA LAGOA	0.759	0.565	0.565	0.759	0.744	Abaixo
UBS CACHOEIRA BOM JESUS	1.000	1.000	1.000	1.000	1.000	Ótima
UBS CAIEIRA BARRA DO SUL	1.000	0.258	0.258	1.000	0.258	Abaixo
UBS CAMPECHE	0.812	0.765	0.765	0.812	0.943	Abaixo
UBS CANASVIEIRAS	1.000	1.000	1.000	1.000	1.000	Ótima
UBS CAPOEIRAS	0.675	0.609	0.609	0.675	0.902	Abaixo
UBS CARIANOS	0.509	0.471	0.471	0.509	0.926	Abaixo
UBS COLONINHA	0.667	0.581	0.581	0.667	0.872	Abaixo
UBS COQUEIROS	0.879	0.875	0.875	0.879	0.996	Abaixo
UBS CÓRREGO GRANDE	0.828	0.564	0.564	0.828	0.680	Abaixo
UBS COSTA DA LAGOA	1.000	0.273	0.273	1.000	0.273	Abaixo
UBS COSTEIRA DO PIRAJUBAÉ	0.947	0.924	0.947	0.924	0.976	Acima
UBS ESTREITO	0.598	0.529	0.529	0.598	0.884	Abaixo
UBS INGLESES	0.878	0.872	0.872	0.878	0.994	Abaixo
UBS ITACORUBI	0.750	0.649	0.649	0.750	0.865	Abaixo
UBS JARDIM ATLÂNTICO	0.728	0.313	0.313	0.728	0.430	Abaixo
UBS JOÃO PAULO	0.732	0.561	0.561	0.732	0.766	Abaixo
UBS JURERE	1.000	0.561	0.561	1.000	0.561	Abaixo
UBS LAGOA DA CONCEIÇÃO	0.586	0.474	0.474	0.586	0.809	Abaixo
UBS MONTE CRISTO	0.799	0.798	0.798	0.799	0.999	Abaixo
UBS MONTE SERRAT	0.804	0.660	0.660	0.804	0.821	Abaixo
UBS MORRO DAS PEDRAS	1.000	0.616	0.616	1.000	0.616	Abaixo
UBS NOVO CONTINENTE	0.686	0.557	0.557	0.686	0.811	Abaixo
UBS PANTANAL	0.747	0.403	0.403	0.747	0.539	Abaixo
UBS PANTANO DO SUL	0.868	0.519	0.519	0.868	0.598	Abaixo
UBS PONTA DAS CANAS	0.926	0.488	0.488	0.926	0.527	Abaixo
UBS PRAINHA	0.578	0.493	0.493	0.578	0.853	Abaixo
UBS RATONES	1.000	0.366	0.366	1.000	0.366	Abaixo
UBS RIBEIRÃO DA ILHA	1.000	0.348	0.348	1.000	0.348	Abaixo
UBS RIO TAVARES	0.492	0.325	0.325	0.492	0.661	Abaixo
UBS RIO VERMELHO	0.714	0.713	0.714	0.713	0.999	Acima
UBS SACO DOS LIMÕES	0.683	0.543	0.543	0.683	0.794	Abaixo
UBS SACO GRANDE	1.000	1.000	1.000	1.000	1.000	Ótima
UBS SANTINHO	1.000	1.000	1.000	1.000	1.000	Ótima
UBS SANTO ANTONIO LISBOA	0.596	0.468	0.468	0.596	0.786	Abaixo
UBS SAPE	0.657	0.416	0.416	0.657	0.633	Abaixo
UBS TAPERA	0.736	0.734	0.734	0.736	0.997	Abaixo
UBS TRINDADE	0.607	0.533	0.533	0.607	0.878	Abaixo
UBS VARGEM GRANDE	0.640	0.284	0.284	0.640	0.443	Abaixo
UBS VARGEM PEQUENA	1.000	0.279	0.279	1.000	0.279	Abaixo
UBS VILA APARECIDA	1.000	0.913	0.913	1.000	0.913	Abaixo

Figura 40 – Escores de eficiências das UBS's utilizando a combinação de variáveis com 5 *inputs* e 1 *output*

	VRS	CRS	DRS	IRS	ESC	Situação
UBS ABRAÃO	0.644	0.533	0.533	0.644	0.827	Abaixo
UBS AGRONÔMICA	0.834	0.828	0.834	0.828	0.993	Acima
UBS ALTO RIBEIRÃO	0.728	0.107	0.107	0.728	0.147	Abaixo
UBS ARMAÇÃO	0.708	0.701	0.708	0.701	0.991	Acima
UBS BALNEÁRIO	1.000	0.809	0.809	1.000	0.809	Abaixo
UBS BARRA DA LAGOA	0.673	0.494	0.494	0.673	0.734	Abaixo
UBS CACHOEIRA BOM JESUS	0.809	0.752	0.752	0.809	0.929	Abaixo
UBS CAIEIRA BARRA DO SUL	1.000	0.209	0.209	1.000	0.209	Abaixo
UBS CAMPECHE	0.865	0.843	0.865	0.843	0.974	Acima
UBS CANASVIEIRAS	1.000	1.000	1.000	1.000	1.000	Ótima
UBS CAPOEIRAS	0.667	0.479	0.479	0.667	0.718	Abaixo
UBS CARIANOS	0.534	0.487	0.487	0.534	0.912	Abaixo
UBS COLONINHA	0.667	0.387	0.387	0.667	0.581	Abaixo
UBS COQUEIROS	0.904	0.803	0.904	0.803	0.889	Acima
UBS CÓRREGO GRANDE	0.992	0.585	0.585	0.992	0.590	Abaixo
UBS COSTA DA LAGOA	1.000	0.061	0.061	1.000	0.061	Abaixo
UBS COSTEIRA DO PIRAJUBAÉ	1.000	1.000	1.000	1.000	1.000	Ótima
UBS ESTREITO	0.595	0.578	0.595	0.578	0.971	Acima
UBS INGLESES	1.000	0.835	1.000	0.835	0.835	Acima
UBS ITACORUBI	1.000	1.000	1.000	1.000	1.000	Ótima
UBS JARDIM ATLÂNTICO	0.632	0.192	0.192	0.632	0.304	Abaixo
UBS JOÃO PAULO	0.752	0.611	0.611	0.752	0.812	Abaixo
UBS JURERE	1.000	0.450	0.450	1.000	0.450	Abaixo
UBS LAGOA DA CONCEIÇÃO	0.733	0.688	0.688	0.733	0.938	Abaixo
UBS MONTE CRISTO	0.953	0.843	0.953	0.843	0.884	Acima
UBS MONTE SERRAT	0.984	0.909	0.909	0.984	0.923	Abaixo
UBS MORRO DAS PEDRAS	1.000	0.634	0.634	1.000	0.634	Abaixo
UBS NOVO CONTINENTE	0.689	0.580	0.580	0.689	0.843	Abaixo
UBS PANTANAL	0.772	0.772	0.772	0.772	1.000	Ótima
UBS PANTANO DO SUL	0.795	0.481	0.481	0.795	0.605	Abaixo
UBS PONTA DAS CANAS	0.814	0.528	0.528	0.814	0.649	Abaixo
UBS PRAINHA	0.516	0.470	0.470	0.516	0.909	Abaixo
UBS RATONES	1.000	0.495	0.495	1.000	0.495	Abaixo
UBS RIBEIRÃO DA ILHA	1.000	0.222	0.222	1.000	0.222	Abaixo
UBS RIO TAVARES	0.349	0.242	0.242	0.349	0.692	Abaixo
UBS RIO VERMELHO	0.804	0.770	0.804	0.770	0.958	Acima
UBS SACO DOS LIMÕES	0.644	0.604	0.604	0.644	0.939	Abaixo
UBS SACO GRANDE	1.000	0.990	1.000	0.990	0.990	Acima
UBS SANTINHO	1.000	1.000	1.000	1.000	1.000	Ótima
UBS SANTO ANTONIO LISBOA	0.521	0.497	0.497	0.521	0.953	Abaixo
UBS SAPE	0.861	0.408	0.408	0.861	0.474	Abaixo
UBS TAPERA	1.000	1.000	1.000	1.000	1.000	Ótima
UBS TRINDADE	0.881	0.878	0.878	0.881	0.997	Abaixo
UBS VARGEM GRANDE	0.807	0.489	0.489	0.807	0.606	Abaixo
UBS VARGEM PEQUENA	1.000	0.336	0.336	1.000	0.336	Abaixo
UBS VILA APARECIDA	1.000	0.308	0.308	1.000	0.308	Abaixo

Figura 41 – Escores de eficiências das UBS's utilizando a combinação de variáveis com 4 *inputs* e 2 *outputs*

Após o cálculo de eficiências, optou-se por continuar a análise apenas com o modelo de 4 *inputs* e 2 *outputs*. Pode-se observar, pelos resultados mostrados na Figura 41, que apenas 6 UBS's operam em escala ótima, 9 se encontram acima da escala ótima e 31 unidades estão em situação de operação abaixo da ótima. Os Centros de Saúde identificados com status "Ótimo" devem ser utilizados como *benchmark* pela Secretaria Municipal da Saúde de Florianópolis, a fim de ajustar a escala de operação das outras UBS's. As unidades 3, 8, 16, 21, 34 e 46 apresentaram os menores escores de eficiência de escala do grupo, demonstrando que o porte atual destas não está adequado às demandas da região. Para essas unidades, podem-se realizar análises adicionais no intuito de identificar medidas que aumentem a eficiência de escala das mesmas.

O escore de eficiência técnica VRS foi encontrado a partir de um modelo DEA orientado a *inputs*, o qual considera que as DMU's ineficientes poderiam reduzir a quantidade de *inputs* para se tornarem mais eficientes, mantendo o mesmo nível de produção de *outputs*. Por definição, a multiplicação do escore obtido para cada DMU pela magnitude dos *inputs* utilizados resultaria em DMU's eficientes. Para obter os respectivos valores de *input* projetados na fronteira de eficiência, multiplica-se cada valor de *input* pelo escore de eficiência, como é mostrado na Figura 42.

	VRS	93 - Total Num Funcionarios	33 - Num Funcionarios - Enfermagem	98 - Area Total Odontologia	100 - Area Total Curativos	Valor ideal 93	Valor ideal 33	Valor ideal 98	Valor ideal 100	Redução 93 em %	Redução 33 em %	Redução 98 em %	Redução 100 em %
UBS ABRAÃO	0.643814	31	4	37	22	20.0	3.0	24.0	14.0	35.0	25.0	35.0	36.0
UBS AGRONÔMICA	0.833740	46	9	36	9	38.0	8.0	30.0	8.0	17.0	11.0	17.0	11.0
UBS ALTO RIBEIRÃO	0.727867	22	4	13	9	16.0	3.0	9.0	7.0	27.0	25.0	31.0	22.0
UBS ARMAÇÃO	0.707944	35	6	20	18	25.0	4.0	14.0	13.0	29.0	33.0	30.0	28.0
UBS BALNEÁRIO	1.000000	23	2	18	9	23.0	2.0	18.0	9.0	0.0	0.0	0.0	0.0
UBS BARRA DA LAGOA	0.672983	28	4	24	11	19.0	3.0	16.0	7.0	32.0	25.0	33.0	36.0
UBS CACHOEIRA BOM JESUS	0.808812	24	5	23	9	19.0	4.0	19.0	7.0	21.0	20.0	17.0	22.0
UBS CAIEIRA BARRA DO SUL	1.000000	15	2	12	12	15.0	2.0	12.0	12.0	0.0	0.0	0.0	0.0
UBS CAMPECHE	0.865423	27	3	20	10	23.0	3.0	17.0	9.0	15.0	0.0	15.0	10.0
UBS CANASVIEIRAS	1.000000	36	6	12	9	36.0	6.0	12.0	9.0	0.0	0.0	0.0	0.0
UBS CAPOEIRAS	0.666667	29	3	39	8	19.0	2.0	26.0	5.0	34.0	33.0	33.0	38.0
UBS CARIANOS	0.533567	39	4	20	18	21.0	2.0	11.0	10.0	46.0	50.0	45.0	44.0
UBS COLONINHA	0.666667	33	3	20	18	22.0	2.0	13.0	12.0	33.0	33.0	35.0	33.0
UBS COQUEIROS	0.903825	29	3	36	11	26.0	3.0	33.0	10.0	10.0	0.0	8.0	9.0
UBS CÓRREGO GRANDE	0.991908	25	5	10	14	25.0	5.0	10.0	14.0	0.0	0.0	0.0	0.0
UBS COSTA DA LAGOA	1.000000	14	4	15	12	14.0	4.0	15.0	12.0	0.0	0.0	0.0	0.0
UBS COSTEIRA DO PIRAJUBAÉ	1.000000	39	4	21	9	39.0	4.0	21.0	9.0	0.0	0.0	0.0	0.0
UBS ESTREITO	0.594823	47	11	30	11	28.0	7.0	18.0	7.0	40.0	36.0	40.0	36.0
UBS INGLESES	1.000000	55	11	32	17	55.0	11.0	32.0	17.0	0.0	0.0	0.0	0.0
UBS ITACORUBI	1.000000	38	8	18	9	38.0	8.0	18.0	9.0	0.0	0.0	0.0	0.0
UBS JARDIM ATLÂNTICO	0.631668	25	4	48	11	16.0	3.0	30.0	7.0	36.0	25.0	38.0	36.0
UBS JOÃO PAULO	0.751682	29	10	36	11	22.0	8.0	27.0	8.0	24.0	20.0	25.0	27.0
UBS JURERE	1.000000	22	2	9	12	22.0	2.0	9.0	12.0	0.0	0.0	0.0	0.0
UBS LAGOA DA CONCEIÇÃO	0.733338	45	6	15	9	33.0	4.0	11.0	7.0	27.0	33.0	27.0	22.0
UBS MONTE CRISTO	0.953248	51	10	36	24	49.0	10.0	34.0	23.0	4.0	0.0	6.0	4.0

Figura 42 – Valores projetados para a fronteira de eficiência

	VRS	93 - Total Num Funcionarios	33 - Num Funcionarios - Enfermagem	98 - Area Total Odontologia	100 - Area Total Curativos		Valor ideal 93	Valor ideal 33	Valor ideal 98	Valor ideal 100	Redução 93 em %	Redução 33 em %	Redução 98 em %	Redução 100 em %
UBS MONTE SERRAT	0.984337	29	5	12	7	29.0	5.0	12.0	7.0	0.0	0.0	0.0	0.0	
UBS MORRO DAS PEDRAS	1.000000	30	2	12	9	30.0	2.0	12.0	9.0	0.0	0.0	0.0	0.0	
UBS NOVO CONTINENTE	0.688697	34	8	37	11	23.0	6.0	25.0	8.0	32.0	25.0	32.0	27.0	
UBS PANTANAL	0.772482	26	4	62	10	20.0	3.0	48.0	8.0	23.0	25.0	23.0	20.0	
UBS PANTANO DO SUL	0.795245	22	3	15	15	17.0	2.0	12.0	12.0	23.0	33.0	20.0	20.0	
UBS PONTA DAS CANAS	0.814167	22	4	14	8	18.0	3.0	11.0	7.0	18.0	25.0	21.0	12.0	
UBS PRAINHA	0.516417	45	7	44	9	23.0	4.0	23.0	5.0	49.0	43.0	48.0	44.0	
UBS RATONES	1.000000	27	2	10	9	27.0	2.0	10.0	9.0	0.0	0.0	0.0	0.0	
UBS RIBEIRÃO DA ILHA	1.000000	25	2	15	9	25.0	2.0	15.0	9.0	0.0	0.0	0.0	0.0	
UBS RIO TAVARES	0.349121	52	7	46	11	18.0	2.0	16.0	4.0	65.0	71.0	65.0	64.0	
UBS RIO VERMELHO	0.803974	54	14	38	11	43.0	11.0	31.0	9.0	20.0	21.0	18.0	18	
UBS SACO DOS LIMÕES	0.643549	38	7	17	9	24.0	5.0	11.0	6.0	37.0	29.0	35.0	33	
UBS SACO GRANDE	1.000000	67	12	55	0	67.0	12.0	55.0	0.0	0.0	0.0	0.0	0.0	
UBS SANTINHO	1.000000	20	2	11	0	20.0	2.0	11.0	0.0	0.0	0.0	0.0	0.0	
UBS SANTO ANTONIO LISBOA	0.521088	38	5	36	11	20.0	3.0	19.0	6.0	47.0	40.0	47.0	45	
UBS SAPE	0.860557	33	4	11	10	28.0	3.0	9.0	9.0	15.0	25.0	18.0	10	
UBS TAPERA	1.000000	45	7	20	10	45.0	7.0	20.0	10.0	0.0	0.0	0.0	0	
UBS TRINDADE	0.880753	49	11	20	18	43.0	10.0	18.0	16.0	12.0	9.0	10.0	11	
UBS VARGEM GRANDE	0.807401	34	6	12	11	27.0	5.0	10.0	9.0	21.0	17.0	17.0	18	
UBS VARGEM PEQUENA	1.000000	20	2	12	9	20.0	2.0	12.0	9.0	0.0	0.0	0.0	0	
UBS VILA APARECIDA	1.000000	16	2	9	0	16.0	2.0	9.0	0.0	0.0	0.0	0.0	0.0	

Figura 42 – Valores projetados para a fronteira de eficiência (Continuação)

As alterações que se referem à diminuição do tamanho da área de odontologia ou curativos não podem ser consideradas, visto que são pontos fixos já construídos e sua redução envolveria gastos adicionais com obra. Com isso, a melhor solução seria aproveitar da melhor forma os espaços existentes oferecendo outros tipos de serviços de atenção básica ou até algum serviço especializado de saúde.

Quanto às mudanças com funcionários, considerando que muitas DMU's se encontram abaixo da escala ótima de operação, pode ser feito o remanejamento de enfermeiros, por exemplo, daquelas DMU's que se mostraram acima da escala para aquelas que se mostraram abaixo da escala ótima.

A análise dos escores de eficiência VRS, CRS, IRS e DRS utilizando as variáveis selecionadas permitiu o cálculo do valor do escorde de eficiência de escala (ESC), concluindo quais unidades se encontram acima ou abaixo da escala ótima de operação. Com isso, seria possível sugerir medidas gerenciais que podem projetar as unidades ineficientes na fronteira eficiente. Como o modelo DEA não aceita um grande número de variáveis, foram utilizadas apenas aquelas que mostraram maior relevância de acordo com os métodos de seleção de *inputs* e *outputs* aplicados. Porém, não se pode garantir a inclusão todas as variáveis que influenciam no desempenho das UBS's. Sendo assim, tentativas devem ser feitas para incluir aquelas que fazem sentido prático para o cenário sob investigação. Portanto, um segundo estágio deste trabalho, irá relacionar a eficiência técnica VRS com as variáveis *input*, não utilizadas no modelo DEA, por meio de análise baseada na técnica de floresta aleatória.

5.4 Análise de treinamento supervisionado com a técnica de floresta aleatória

A análise baseada na técnica de floresta aleatória visa relacionar os escores de eficiência a outras variáveis, com o objetivo de encontrar padrões nas características das UBS's que indiquem valores semelhantes de escores de eficiência. Dessa forma, pretende-se identificar outras variáveis, além daquelas utilizadas no Modelo 2 DEA, que estejam relacionadas ao desempenho de uma UBS.

Na seleção das variáveis a serem utilizadas nesta etapa, considerou-se, dentre as variáveis levantadas em conjunto com a Secretaria Municipal de Saúde do município de Florianópolis, apenas aquelas que representam *input*. Desse conjunto de *inputs*, foram excluídas da análise as quatro variáveis, que foram utilizadas como *input* no Modelo 2 do DEA, restando assim, 34 variáveis para serem relacionadas aos escores de eficiência VRS.

O método de floresta aleatória selecionado para o presente trabalho foi *Gradient Boosting Regressor*. *Boosting* é um processo que combina regras de predição separadas,

algumas das quais podem ser bastante fracas, para produzir um classificador combinado mais poderoso [1]. Em 1999, Friedman introduziu o conceito de *gradient boosting*, que combina idéias de *boosting* com árvores de decisão. *Regressor* faz predições numéricas com base em informações sobre uma observação. Em geral, cada observação possui um vetor de variáveis. No contexto do deste trabalho, a regressão visa predizer os escores de eficiência VRS, com base nos 34 *inputs*, selecionados para essa etapa, e nos escores de eficiência VRS calculados no Modelo 2 DEA. A acurácia da predição pode ser verificada por meio de um gráfico que ilustra o comportamento do erro durante o processo de treinamento [2].

Para contornar possíveis problemas de *overfitting*, que comprometem a acurácia do modelo treinado, deve-se ajustar os hiperparâmetros pouco a pouco. O método utilizado para selecionar os hiperparâmetros foi o *Grid Search*, que procura o melhor desempenho do modelo sobre o espaço dos hiperparâmetros [3]. Dentre os parâmetros, que foram alterados utilizando esse método, destaca-se: *learning rate*, quantidade de árvores criadas e máxima profundidade dessas árvores. O processo de refinamento de um modelo baseado em floresta aleatória consiste em "tentativas e erros", alterando seus parâmetros e observando os resultados alcançados. Após diversas variações de valores dos três hiperparâmetros citados, selecionou-se como máxima profundidade o valor 2, e como número de árvores criadas o valor 20. O valor de máxima profundidade limita o número de nós em uma árvore. Para uma análise mais criteriosa, decidiu-se analisar dois diferentes valores para o parâmetro de *learning rate*, sendo eles 0.004 e 0.0004. As Figuras 43 e 44 ilustram as curvas de aprendizado dos dois modelos.

A fim de quantificar a utilidade das variáveis na técnica de floresta aleatória, pode-se observar as importâncias relativas das variáveis, como mostram as Figuras 45 e 46. Essas importâncias representam o quanto incluir uma variável específica melhora a predição. E analisando a Figura 45, o gráfico de barras mostra que número de funcionários nutricionista é o melhor preditor do escore de eficiência técnica do modelo. A área total de consultório se encontra na segunda posição e assim por diante.

Figura 43 – Curvas de aprendizado com *learning rate* de 0.004

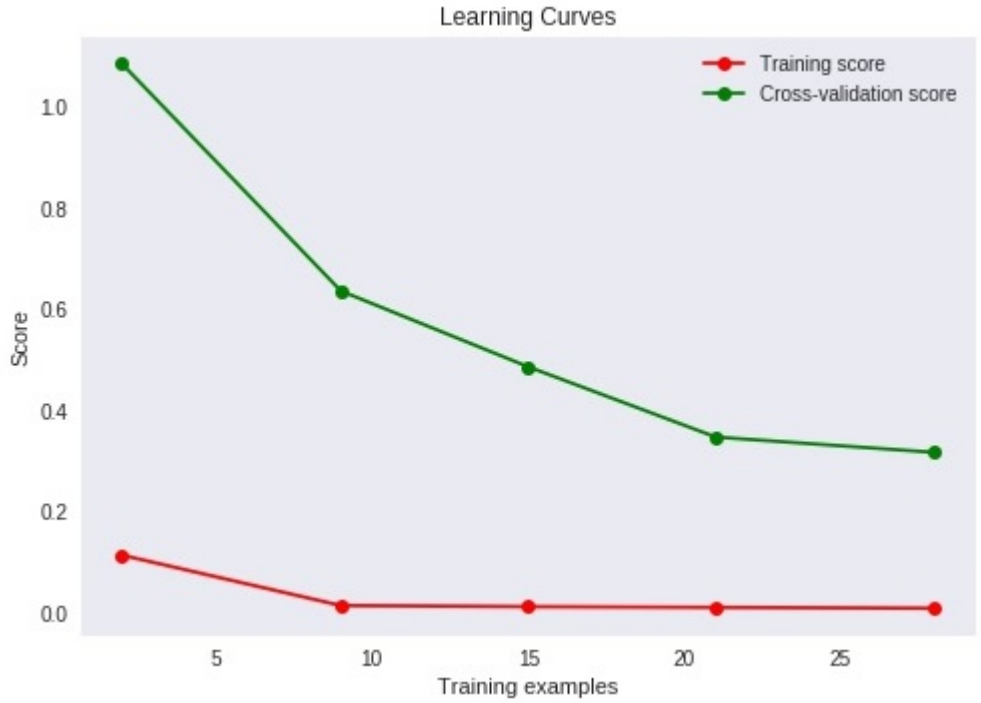


Figura 44 – Curvas de aprendizado com *learning rate* de 0.004

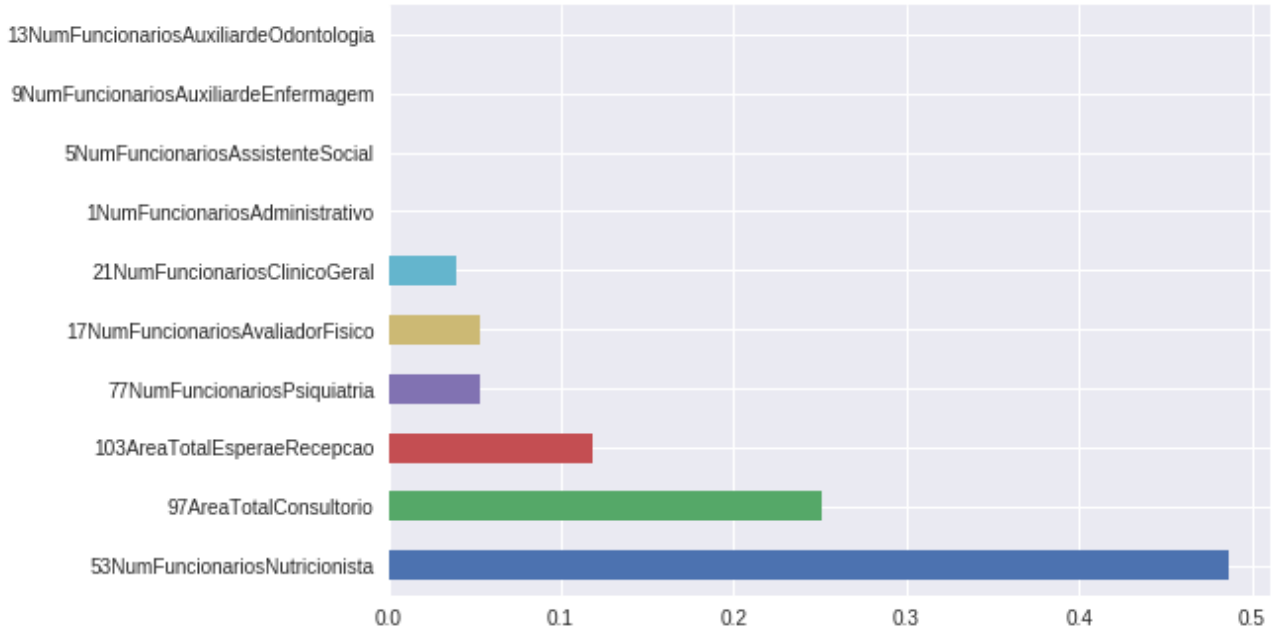


Figura 45 – Variáveis importantes com *learning rate* de 0.004

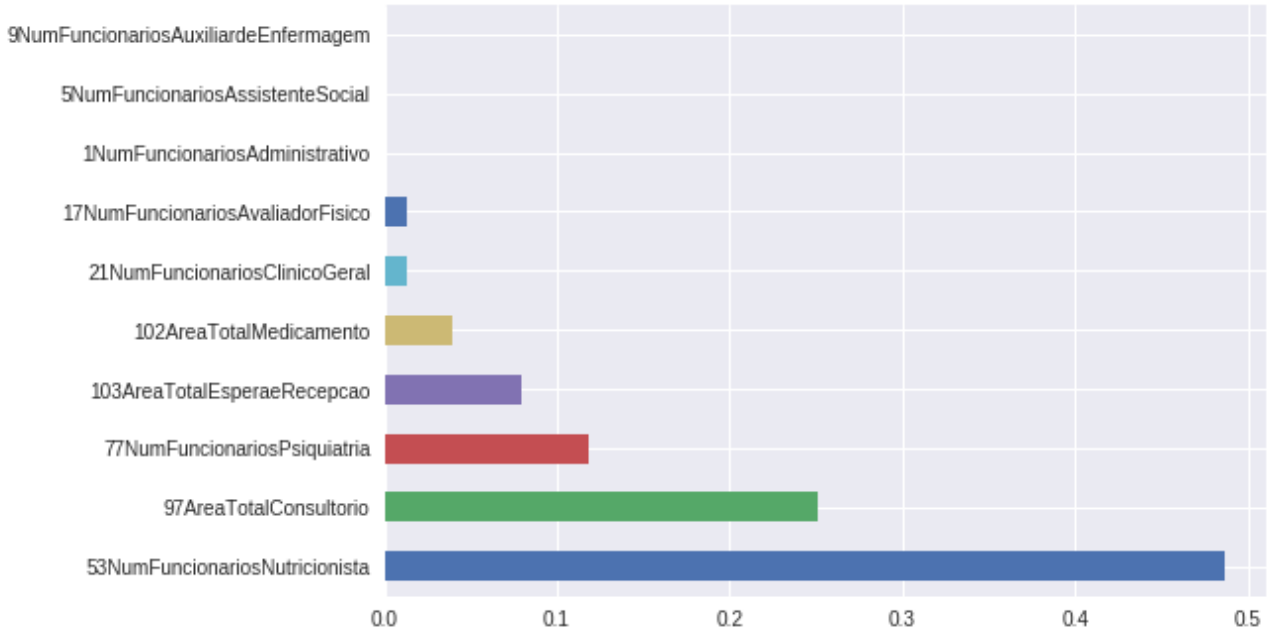


Figura 46 – Variáveis importantes com *learning rate* de 0.0004

Nota-se que mesmo com um valor dez vezes menor de *learning rate*, as variáveis de maior importância no modelo continuam praticamente as mesmas, e as curvas de aprendizado também são bastante similares. Por causa disso, novas formas de avaliação foram incluídas para permitir selecionar o *learning rate* mais adequado. Calculou-se, então, o erro quadrático médio, ou MSE (*Mean Squared Error*), que mede a média dos quadrados do erro, e também, a acurácia da predição de cada um dos modelos. Quanto menor o valor do primeiro, e maior o do segundo, melhor o desempenho do modelo calculado. Para o modelo com *learning rate* de 0.004, o valor do MSE foi de 0.0270 e a acurácia de 81.35%, já para o modelo com *learning rate* de 0.0004, o valor do MSE foi de 0.0282 e a acurácia de 80.85%. Por conta disso, a análise prosseguiu apenas com a representação da última árvore criada com o modelo que continha o maior valor de *learning rate*. A representação dessa árvore é apresentada na Figura 47.

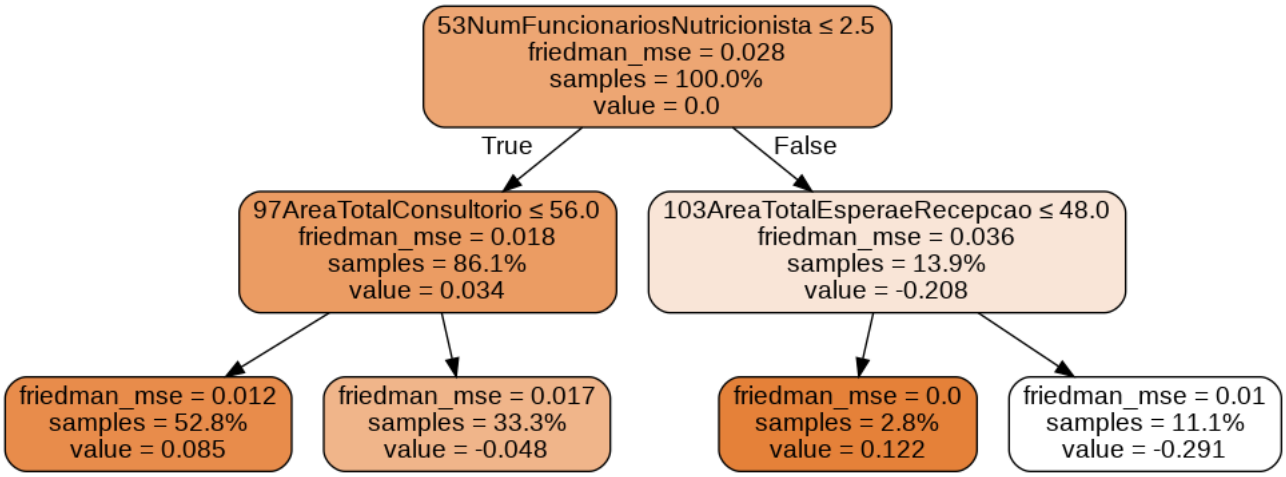


Figura 47 – Representação da árvore resultante da análise

Os nós da árvore são subdivididos, quantas vezes for necessário, até o modelo não identificar maiores ganhos de informação e chegar aos nós terminais, que não possuem bifurcações. Analisando os nós dessa árvore, a primeira linha indica a variável e o valor para dividir esse nó. *Friedman mse* descreve o valor do erro quadrático médio do nó, *samples*o número de *data points* e por fim, *value* a predição para todos os *data points*.

O caminho representado por número de funcionários nutricionista maior que 2.5 e área total espera e recepção menor ou igual à 48 leva ao maior escore de eficiência técnica da árvore, no valor de 0.122. Esse valor é extremamente baixo em relação aos escores de eficiência calculados pelo modelo DEA. Portanto, a análise de eficiência utilizando a técnica de floresta aleatória não se mostrou apropriada para o conjunto de dados deste trabalho.

5.5 Considerações Finais

A obtenção dos escores de eficiência pela aplicação do modelo DEA VRS orientado para o *input* nos dados das Unidades Básicas de Saúde do município de Florianópolis possibilitou identificar unidades eficientes e ineficientes, além de projetar aquelas com ineficiência na fronteira eficiente de produção. O cálculo dos escores CRS, IRS e DRS foi essencial na classificação das UBSs quanto à escala de operação, encontrando 6 unidades em escala ótima de operação, 9 unidades acima da escala ou subutilizadas e 31 unidades abaixo da escala ótima de operação, ou seja, estão sobrecarregadas.

Apesar de terem sido calculados os valores projetados para as unidades ineficientes, a mudança de todas as áreas de uma UBS e o remanejamento de funcionários do Sistema de Saúde do município não é viável financeiramente e operacionalmente, principalmente a curto prazo. Com isso, realizou-se um segundo estágio de análise de treinamento supervisionado com a técnica de floresta aleatória, que teve como objetivo analisar outras variáveis além das utilizadas no modelo DEA.

A utilização da técnica de floresta aleatória relacionando os escores de eficiência VRS, encontrados pela Análise Envoltória de Dados, a outras variáveis não utilizadas anteriormente não se mostrou uma técnica ideal de análise, visto a quantidade dos dados levantados em conjunto com a Secretaria Municipal da Saúde.

A SMS, que convive diariamente com as dificuldades enfrentadas pela Atenção Básica no município pode, a partir deste trabalho, avaliar as informações encontradas com viés direcionado à necessidade de cada UBS e à facilidade de implantação de cada mudança e então definir quais medidas gerenciais podem efetivamente ser realizadas a fim de aumentar a eficiência do sistema.

6 Conclusões e Perspectivas

O trabalho desenvolvido teve como objetivo avaliar a eficiência das Unidades Básicas de Saúde do município de Florianópolis e identificar possíveis medidas gerenciais, que possam auxiliar na tomada de decisão da Secretaria Municipal da Saúde, para a melhoria do desempenho das UBS's do município.

Visto que o modelo DEA, para ser discriminatório, exige um número reduzido de variáveis em relação ao número de DMU's, a primeira etapa da análise consistiu em uma redução de dimensionalidade dos dados. As 110 variáveis, levantadas em conjunto com a Secretaria Municipal da Saúde, passaram por diversas etapas eliminatórias, com o objetivo de se utilizar apenas aquelas que tivessem maior ganho de informação para o modelo. Foi incluído um segundo estágio de análise, considerando a análise de treinamento supervisionado utilizando a técnica de Florestas Aleatórias, a fim de não descartar o possível impacto de outras variáveis no desempenho das UBS's. A técnica de floresta aleatória relaciona os escores de eficiência VRS com as variáveis de *input*, exceto aquelas que já tinham sido analisadas no modelo DEA.

No primeiro estágio do DEA foi possível identificar quais UBS's são tecnicamente eficientes e quais não o são. Para as DMU's que se mostraram em situação crítica, foram sugeridas medidas relacionadas às características utilizadas no modelo DEA. Com a análise de treinamento supervisionado utilizando a técnica de floresta aleatória, pretendia-se identificar regras de decisão que levam as UBS's aos escores mais baixos de eficiência e regras que as levam à eficiência máxima.

Apesar de Florianópolis ser a capital brasileira mais bem avaliada na questão Atenção à Saúde Primária, segundo o Ministério da Saúde, ela pode ser sujeita à análise de eficiência, com vistas a projetar suas operações na fronteira máxima de eficiência. As análises feitas e as possíveis medidas gerenciais apresentadas podem auxiliar a SMS na tomada de decisões em busca pelo aumento da eficiência dos serviços prestados pelas UBS's de Florianópolis. Com o desafio de otimizar os recursos disponíveis, aumentando o nível de atendimento ao público, é recomendado a realocação de recursos, de forma a projetar as UBS's ineficientes na fronteira de eficiência de escala ditada pelo conjunto de UBS's eficientes. Os resultados apresentados devem ser analisados por profissionais da área da saúde que possuem contato direto com a Atenção Básica, pois estes são capazes de priorizar as medidas gerenciais sugeridas.

A exclusão das unidades, que não possuíam dados relacionados a infraestrutura e/ou dados do quadro de funcionários, foi uma das limitações encontradas no presente trabalho, e pode ser abordada em trabalhos futuros. Além disso, outra limitação observada

foi na quantidade de dados levantados em conjunto com a Secretaria Municipal da Saúde de Florianópolis. Por se tratar de um pequeno conjunto de dados, a análise utilizando a técnica de floresta aleatória ficou bastante comprometida. Por fim, uma análise interessante a ser incluída em trabalhos futuros seria o nível de escolaridade dos funcionários das UBS's. Essa discriminação feita no quadro de funcionários seria para analisar se profissionais mais experientes contribuem para o aumento da eficiência das unidades, visto que, apenas uma alteração nos recursos não seja suficiente para uma melhoria na eficiência, se os funcionários não estiverem capacitados para atender à demanda.

A Análise Envoltória de Dados combinada à técnica de floresta aletória pode servir como suporte à tomada de decisões de diversos órgãos públicos e privados ao encontrar escores de eficiência e relacioná-los a outras variáveis que podem ser contínuas ou categóricas. No caso da Secretaria Municipal da Saúde, o trabalho auxilia os tomadores de decisão ao alocar ou realocar recursos e assim melhorar a Atenção Básica da Saúde do município.

Referências

1 ROGOJAN, B. *Boosting and Bagging: How To Develop A Robust Machine Learning Algorithm*. 2017. Disponível em: <https://hackernoon.com/how-to-develop-a-robust-algorithm-c38e08f32201>. Citado 5 vezes nas páginas 3, 21, 53, 55 e 90.

2 ATKINSON, E. J. et al. Assessing fracture risk using gradient boosting machine (GBM) models. *Journal of Bone and Mineral Research*, v. 27, n. 6, p. 1397–1404, 2012. ISSN 08840431. Citado 2 vezes nas páginas 3 e 90.

3 PARR, T.; HOWARD, J. *Gradient boosting: Distance to target*. Disponível em: <https://explained.ai/gradient-boosting/L2-loss.html>. Citado 2 vezes nas páginas 3 e 90.

4 M. Lundberg, S.; LEE, S.-I. *A Unified Approach to Interpreting Model Predictions*. Long Beach, CA, USA: 31st Conference on Neural Information Processing Systems, 2017. Citado na página 7.

5 KHALID, B. *SHAP Values : The efficient way of interpreting your model*. 2018. Disponível em: <https://medium.com/datadriveninvestor/shap-values-the-efficient-way-of-interpreting-your-model-7de632ed7d2d>. Citado na página 7.

6 TSENG, G. *Interpreting complex models with SHAP values*. 2018. Disponível em: <https://medium.com/@gabrieltseng/interpreting-complex-models-with-shap-values-1c187db6ec83>. Citado na página 7.

7 M. Lundberg, S. *SHAP (SHapley Additive exPlanations)*. 2018. Disponível em: <https://github.com/slundberg/shap>. Citado na página 8.

8 CHARNES, a.; COOPER, W. W.; RHODES, E. Measuring the efficiency of decision making units. *European Journal of Operational Research*, v. 2, n. 6, p. 429–444, 1978. ISSN 03772217. Citado 3 vezes nas páginas 33, 45 e 46.

9 SOARES a.B.; PEREIRA a.a.; MILAGRE, S. A model for multidimensional efficiency analysis of public hospital management. *Research on Biomedical Engineering*, v. 33, n. 4, p. 352–361, 2017. ISSN 24464740. Citado 3 vezes nas páginas 33, 36 e 39.

10 CARRILLO, M.; JORGE, J. M. DEA-Like Efficiency Ranking of Regional Health Systems in Spain. *Social Indicators Research*, v. 133, n. 3, p. 1133–1149, 2017. ISSN 15730921. Citado 4 vezes nas páginas 33, 34, 35 e 38.

11 AHN, H. et al. Recent developments on the use of DEA in the public sector. *Socio-Economic Planning Sciences*, v. 61, p. 1–3, 2018. ISSN 00380121. Citado na página 33.

12 ASANDULUI, L.; ROMAN, M.; FATULESCU, P. The Efficiency of Healthcare Systems in Europe: A Data Envelopment Analysis Approach. *Procedia Economics and*

Finance, Elsevier B.V., v. 10, n. 14, p. 261–268, 2014. ISSN 22125671. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S2212567114003013>. Citado 4 vezes nas páginas 33, 34, 35 e 38.

13 COOPER, W.; SEIFORD, L.; TONE, K. *Data envelopment analysis: a comprehensive text with models, applications, references and DEA-Solver Software*. 2. ed. [S.l.]: Springer US, 2007. XXXVIII, 492 p. ISBN 978-0-387-45281-4. Citado 6 vezes nas páginas 21, 34, 44, 45, 48 e 49.

14 RAMÍREZ-VALDIVIA, M. T.; MATURANA, S.; SALVO-GARRIDO, S. A multiple stage approach for performance improvement of primary healthcare practice. *Journal of Medical Systems*, v. 35, n. 5, p. 1015–1028, 2011. ISSN 01485598. Citado 2 vezes nas páginas 34 e 38.

15 KOUNETAS, K.; PAPATHANASSOPOULOS, F. How efficient are Greek hospitals? A case study using a double bootstrap DEA approach. *European Journal of Health Economics*, v. 14, n. 6, p. 979–994, 2013. ISSN 16187598. Citado 3 vezes nas páginas 34, 35 e 38.

16 LINS, M. E. et al. O uso da Análise Envoltória de Dados (DEA) para avaliação de hospitais universitários brasileiros. *Ciência & Saúde Coletiva*, v. 12, n. 4, p. 985–998, 2007. ISSN 1413-8123. Disponível em: <http://www.scielo.br/scielo.php?script=sci{_}arttext{&}pid=S1413-81232007000400020{&}lng=p>. Citado 2 vezes nas páginas 35 e 39.

17 CACHUBA, L. M. Uma análise da eficiência da oferta de serviços de saúde pública na região de curitiba por meio de análise envoltória de dados. 2016. Citado 3 vezes nas páginas 37, 39 e 56.

18 PéREZ-ROMERO, C. et al. Análisis de la eficiencia técnica en los hospitales del Sistema Nacional de Salud español. *Gaceta Sanitaria*, v. 31, n. 2, p. 108–115, 2017. ISSN 0213-9111. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0213911116302588>. Citado na página 39.

19 BURGES, C. Dimension reduction: A guided tour. *Foundations and Trends in Machine Learning*, Now Publishers, Inc., v. 2, n. 4, p. 275–365, 2010. Citado na página 41.

20 YU, L.; LIU, H. Feature selection for high-dimensional data: A fast correlation-based filter solution. In: *Proceedings of the 20th international conference on machine learning*. [S.l.: s.n.], 2003. p. 856–863. Citado na página 41.

21 BELCH, B.; HUANG, C. *Multivariate Statistical Methods for Business and Economics*. Englewood Cliffs, New Jersey: Prentice-Hall, 1974. Citado na página 42.

22 JOHNSON, G. W. et al. Introduction to Environmental Forensics. p. Pages 207–272, 2007. Citado na página 42.

23 ABDI, H.; J. Lynne, W. Principal component analysis. v. 2, p. 1–27, 2010. ISSN 01377183. Citado na página 42.

24 VANDERPLAS, J. *In Depth: Principal Component Analysis*. Disponível em: <https://jakevdp.github.io/PythonDataScienceHandbook/05.09-principal-component-analysis.html>. Citado 2 vezes nas páginas 21 e 42.

25 FRIED, H.; LOVELL, C. a. K.; SCHMIDT, S. Production Frontiers and Productive Efficiency:The Measurement of Productive Efficiency. Vi, n. January 1993, 1993. Citado na página 43.

26 FERREIRA, C. M. de C.; GOMES, A. P. *Introdução à análise envoltória de dados: teoria, modelos e aplicações.* Viçosa, MG: UFV, 2009.: UFV, 2009. 389 p. p. Disponível em: <https://books.google.com.br/books?id=G7neZwEACAAJ>. Citado 3 vezes nas páginas 43, 47 e 49.

27 WORTHINGTON, A. C. Frontier efficiency measurement in health care: A review of empirical techniques and selected applications. *Medical Care Research and Review*, v. 61, n. 2, p. 135–170, 2004. ISSN 10775587. Citado na página 43.

28 LUPTÁČIK, M. Mathematical optimization and economic analysis, Springer optimization and its applications. *Springer New York Dordrecht Heidelberg London*, v. 36, 2010. Citado 3 vezes nas páginas 21, 43 e 44.

29 COOK, W.; ZHU, J. Data Envelopment Analysis. *Springer US*, p. 401, 2005. Citado na página 44.

30 BANKER, R. D. et al. An introduction to data envelopment analysis with some of its models and their uses. *Research in governmental and nonprofit accounting*, JAI Press, Inc Greenwich, CT, v. 5, p. 125–163, 1989. Citado na página 45.

31 COOK, W. D.; TONE, K.; ZHU, J. Data envelopment analysis: Prior to choosing a model. *Omega (United Kingdom)*, Elsevier, v. 44, p. 1–4, 2014. ISSN 03050483. Disponível em: <http://dx.doi.org/10.1016/j.omega.2013.09.004>. Citado 2 vezes nas páginas 45 e 46.

32 BIAN, Y.; YANG, F. Resource and environment efficiency analysis of provinces in China: A DEA approach based on Shannon's entropy. *Energy Policy*, Elsevier, v. 38, n. 4, p. 1909–1917, 2010. ISSN 03014215. Disponível em: <http://dx.doi.org/10.1016/j.enpol.2009.11.071>. Citado na página 45.

33 FÄRE, R.; GROSSKOPF, S.; LOVELL, C. A. K. Production Frontiers. *The Economic Journal*, Cambridge: University Press, v. 105, n. 430, p. 738, may 1995. ISSN 00130133. Citado 3 vezes nas páginas 21, 47 e 49.

34 FRIES, C. E. Avaliação do Impacto do uso de Tecnologias de Informação e Comunicação na Eficiência de Prestadores de Serviços Logísticos. *Universidade Federal de Santa Catarina*, p. 195, 2013. Citado 2 vezes nas páginas 21 e 50.

35 BIAN, Y.; YANG, F. Resource and environment efficiency analysis of provinces in China: A DEA approach based on Shannon's entropy. *Energy Policy*, Elsevier, v. 38, n. 4, p. 1909–1917, 2010. ISSN 03014215. Disponível em: <http://dx.doi.org/10.1016/j.enpol.2009.11.071>. Citado na página 51.

36 PAL, N. R.; PAL, S. K. Entropy: A New Definition and its Applications. *IEEE Transactions on Systems, Man and Cybernetics*, v. 21, n. 5, p. 1260–1270, 1991. ISSN 21682909. Citado na página 51.

37 SHANNON, C. E. A mathematical theory of communication. *Bell System Technical Journal*, v. 27, n. 3, 1948. Citado na página 51.

38 BREIMAN, L. et al. Classification and regression trees. *Wadsworth International Group*, 1984. Citado na página 52.

39 GAMA, J. a. Functional trees. *Mach. Learn.*, Kluwer Academic Publishers, Hingham, MA, USA, v. 55, n. 3, p. 219–250, jun. 2004. ISSN 0885-6125. Disponível em: <http://dx.doi.org/10.1023/B:MACH.0000027782.67192.13>. Citado 2 vezes nas páginas 21 e 52.

40 Data Science Academy. *Deep Learning Book*. [s.n.], 2008. Disponível em: <http://deeplearningbook.com.br/capitulos/>. Citado 3 vezes nas páginas 21, 53 e 54.

41 GISLASON, P. O.; BENEDIKTSSON, J. A.; SVEINSSON, J. R. Random forests for land cover classification. *Pattern Recognition Letters*, v. 27, n. 4, p. 294–300, 2006. ISSN 01678655. Citado na página 54.

42 BREIMAN, L. Bagging {Predictors}. *Machine Learning*, v. 24, n. 2, p. 123–140, 1996. ISSN 1573-0565. Disponível em: <http://statistics.berkeley.edu/sites/default/files/tech-reports/421.pdf>. Citado na página 54.

43 KEARNS, M. Thoughts on hypothesis boosting. *Unpublished manuscript, Machine Learning class project*, v. 45, p. 105, 1988. Citado na página 54.

44 BRASIL. *Lei 8.080, de 19 de setembro 1990.* Disponível em: <http://www.planalto.gov.br/ccivil_03/Leis/L8080.htm>. Citado na página 56.

45 BRASIL. *Política Nacional de Atenção Básica*. Brasília: Ministério da Saúde, 2012. Disponível em: <http://dab.saude.gov.br/portaldab/biblioteca.php?conteudo=publicacoes/pnab>. Citado 3 vezes nas páginas 56, 57 e 58.

46 BRASIL. *Retratos da Atenção Básica nº 2 - Gestão da Atenção Básica Volume 2 - Insumos e Medicamentos nas Unidades Básicas de Saúde*. Ministério da Saúde, 2015. Disponível em: <http://bvsms.saude.gov.br/bvs/publicacoes/retratos_atencao_basica_gestao_atencao_n2_v2.pdf>. Citado na página 58.

47 BRASIL. *Cadernos de Atenção Básica - Acolhimento à demanda espontânea*. Brasília, DF: Ministério da Saúde, 2013. I. ISBN 9788533418431. Citado na página 59.

48 Secretaria Municipal de Saúde. *Florianópolis tem a melhor saúde primária entre as capitais do país*. 2018. Disponível em: <http://www.pmf.sc.gov.br/entidades/saude/index.php?pagina=notpagina&noti=20162>. Citado na página 67.

Apêndices

APÊNDICE A – Código em Python

```
!pip3 install PyDrive
!pip3 install xlrd
!pip3 install openpyxl
!pip3 install xlsxwriter

import io
import os
import pandas as pd
import numpy as np

from pandas import ExcelWriter
from pandas import ExcelFile
from pydrive.auth import GoogleAuth
from pydrive.drive import GoogleDrive
from google.colab import auth
from oauth2client.client import GoogleCredentials

auth.authenticate_user()
gauth = GoogleAuth()
gauth.credentials = GoogleCredentials.get_application_default()
drive = GoogleDrive(gauth)

!mkdir -p drive
#!google-drive-ocamlfuse drive
from google.colab import drive
drive.mount('/content/gdrive')


###CORRELACAO

dados_secretaria_completo = pd.read_excel('/content/gdrive/My Drive/
PFC segunda tentativa/Final.xlsx', sheetname='Sheet1')

print("Column headings:")
indices = dados_secretaria_completo.columns[1:]

dados_secretaria = dados_secretaria_completo.iloc[:,3:]

###Aplicando correlacao

dados_corr = dados_secretaria.corr()
print (dados_corr.to_string())
```

```
###Elimina diagonal principal da analise

np.fill_diagonal(dados_corr.values, 0)
dados_corr

###Elimina pela Coluna

variaveis = []

b=0
def dropcolumn(b):
    for i in range(0, len(dados_corr)):
        for j in range(0, len(dados_corr)):
            if (dados_corr.iloc[i][j] >= 0.85):
                #print(dados_corr.iloc[i][j])
                variaveis.append(dados_corr.index[j])
                dados_corr.drop(dados_corr.index[j], inplace=True)
                dados_corr.drop(dados_corr.columns[j], axis=1, inplace=True)
                b=1
                break
        if(b==1):
            b=0
            break



for i in range(0, len(dados_corr)):
    b=0
    dropcolumn(b)


###lista variaveis elimininadas

variaveis.remove('94 - Total Atendimentos')
variaveis


###Dados originais com as variaveis da correlacao retiradas

dados_secretaria_completo.drop(variaveis, inplace=True, axis=1)
dados_secretaria_completo


###variaveis mantidas
```

```
indices = dados_secretaria_completo.columns
indices[3:]

###ACP

from sklearn import decomposition
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler
from scipy import stats

%matplotlib inline
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns; sns.set()

dados_secretaria_filtrado = dados_secretaria_completo.iloc[:,3:]

###Normalizacao dados

dados_normalizados = stats.zscore(dados_secretaria_filtrado)

pca = decomposition.PCA(n_components = 10)
pca.fit(dados_normalizados)
dados_pca = pca.transform(dados_normalizados)
print("original shape:   ", dados_secretaria_filtrado.shape)
print("transformed shape:", dados_pca.shape)

###Cumulative Variance explains

var1=np.cumsum(np.round(pca.explained_variance_ratio_, decimals=4)*100)
plt.plot(np.cumsum(pca.explained_variance_ratio_))
plt.xlabel('number of components')
plt.ylabel('cumulative explained variance');

matriz_pca = (pca.components_ )
matriz_pca = pd.DataFrame(matriz_pca,
                          columns=indices[3:],
                          index=('PC1','PC2','PC3','PC4',
                                 'PC5','PC6','PC7','PC8','PC9',  '
                                    PC10'))
print(matriz_pca.to_string())

figura_calor = matriz_pca.transpose()
figura_calor
plt.figure(figsize=(20,6))
sns.heatmap(matriz_pca,cmap='plasma')
```

```
###Variaveis mais relevantes

variaveis_pca = []

for linha in range(0, len(matriz_pca)):
  for coluna in matriz_pca.columns:
     if (matriz_pca.iloc[linha][coluna] >= 0.194):
        variaveis_pca.append(coluna)

variaveis_pca


###valores nao repetidos

my_set = set(variaveis_pca)

###pegando valores mais negativos das componentes principais

ultimas_variaveis = np.array(matriz_pca.idxmin(axis=1).values).tolist()
ultimas_variaveis

###conferindo os que nao existem na primeira lista

adicionar = set(ultimas_variaveis).difference(set(variaveis_pca))
adicionar
my_set.update(adicionar)
my_set

###Variaveis seelecionadas pelo ACP

variaveis_pca = list(my_set)
variaveis_pca.sort()
variaveis_pca

to_drop = ['1 - Num Funcionarios - Administrativo',
 'Num Funcionarios - Auxiliar de Odontologia',
 'Atendimentos - Farmacia',
 'Atendimentos - Geriatria',
 'Atendimentos - Ortopedia',
 'Num Funcionarios - Psiquiatria',
 'Atendimentos - Psiquiatria',
 'Num Funcionarios - Auxiliar de Enfermagem']

for value in to_drop:
  for dado in variaveis_pca:
    if value in dado:
      variaveis_pca.remove(dado)
```

```
###variaveis selecionadas apos ACP e filtragem Statistica

variaveis_pca

###Porcentagem nao-zeros variaveis ACP

dados_secretaria_completo

###analise somente das variaveis selecionadas pelo ACP

dados_sms_filtrados = dados_secretaria_completo[dados_secretaria_completo.
    columns.intersection(variaveis_pca)]

###contagem nao-zeros das variaveis selecionadas pelo ACP

nao_zeros = dados_sms_filtrados.astype(bool).sum(axis=0)
porcentagem = (nao_zeros*100)/46
print(porcentagem.sort_values(ascending=False))

###remover valores menores 50%

variaveis_filtradas = porcentagem[porcentagem >= 50]
variaveis_filtradas.sort_values(ascending=False)

###Selecao inputs e outputs

input = []
for i in variaveis_filtradas.index:
    if 'Atendimentos' in i:
        continue
    input.append(i)

input

output = []
for i in variaveis_filtradas.index:
    if 'Atendimentos' in i:
       output.append(i)

output

###Programa Iterativo

###MODELO 1

import json
```

```
import codecs

modelo1 = pd.read_json('/content/gdrive/My Drive/PFC segunda tentativa/
    model1_num_cols8.json', lines=True)

modelo1['e']

###variaveis inseridas a mais pelo Statistica

to_drop = ['Area Total Medicamento',
           'Area Total Administracao',
           'Num Funcionarios - Avaliador Fisico',
           'Num Funcionarios - Medico Familia',
           'Num Funcionarios - Psicologia']

to_drop

def filtragem(dados):
  for value in to_drop:
    for dado in dados:
      if value in dado:
        dados.clear()
  return dados

modelo1['w'].apply(filtragem)
modelo1


###numero de linhas retiradas

sum(modelo1['w'].str.len() == 0)

modelo1 = modelo1[modelo1.astype(str)['w'] != '[]']
modelo1 = modelo1.reset_index(drop=True)
modelo1

#Create a Pandas Excel writer using XlsxWriter as the engine.
writer = pd.ExcelWriter('modelo1giglio_filtrado.xlsx', engine='xlsxwriter')

# onvert the dataframe to an XlsxWriter Excel object.
modelo1.to_excel(writer, sheet_name='Sheet1')

#Close the Pandas Excel writer and output the Excel file.
writer.save()

resultados = []
```

```
for i in range(0, len(modelo1)):
  resultados.append(sum(value >= 0.9999 for value in modelo1['e'][i]))

modelo1['eficientes'] = resultados

#calcula quantas variaveis combinadas entre si

tamanho = []

for linha in range (0, len(modelo1)):
  tamanho.append(len([i.split('\t', 1) for i in modelo1['w'][linha]]))

modelo1['combinacoes'] = tamanho

#Numero de Ocorrencias - Combinacoes x Eficientes

table1 = pd.pivot_table(modelo1, values='e', index=['eficientes'],
                        columns=['combinacoes'], aggfunc=np.ma.count)

table1.fillna("0")
table1['Total de eficientes'] = table1.sum(axis=1)
table1.loc['Total de combinacoes'] = table1.sum(axis=0)
table1.fillna("0.0")

###MODELO 2

import json
import codecs

modelo2 = pd.read_json('/content/gdrive/My Drive/PFC segunda tentativa/
    model2_num_cols8.json', lines=True)

modelo2

# variaveis inseridas a mais pelo Statistica

to_drop = ['Area Total Medicamento',
           'Area Total Administracao',
           'Num Funcionarios - Avaliador Fisico',
           'Num Funcionarios - Medico Familia',
           'Num Funcionarios - Psicologia']

to_drop

def filtragem(dados):
  for value in to_drop:
    for dado in dados:
```

```
        if value in dado:
          dados.clear()
    return dados

modelo2['w'].apply(filtragem)
modelo2

# numero de linhas retiradas

sum(modelo2['w'].str.len() == 0)
modelo2 = modelo2[modelo2.astype(str)['w'] != '[]']
modelo2

# reset index

modelo2 = modelo2.reset_index(drop=True)
modelo2

resultados = []

for i in range(0, len(modelo2)):
  resultados.append(sum(value >= 0.9999 for value in modelo2['e'][i]))
modelo2['eficientes'] = resultados

# calcula quantas variaveis combinadas entre si

tamanho = []

for linha in range (0, len(modelo2)):
  tamanho.append(len([i.split('\t', 1) for i in modelo2['w'][linha]]))

modelo2['combinacoes'] = tamanho

# Create a Pandas Excel writer using XlsxWriter as the engine.
writer = pd.ExcelWriter('analise_combinacoes_modelo2.xlsx', engine='
    xlsxwriter')

# Convert the dataframe to an XlsxWriter Excel object.
modelo2.to_excel(writer, sheet_name='Sheet1')

# Close the Pandas Excel writer and output the Excel file.
writer.save()

# Combinacoes x Eficientes

table2 = pd.pivot_table(modelo2, values='e', index=['eficientes'],
                        columns=['combinacoes'], aggfunc=np.ma.count)
```

```
table2.fillna("0.0")
table2['Total_eficientes'] = table2.sum(axis=1)
table2.loc['Total_combinacoes'] = table2.sum(axis=0)
table2.fillna("0.0")


###CALCULO PROBABILIDADES

#MODELO 1
valor1 =[]

for linha in range(0,len(modelo1)):
  conta1 = len(list(x for x in modelo1['e'][linha] if x <= 0.1))
  valor1.append(conta1)

divisor = 46
probab1 = [x/divisor for x in valor1]
modelo1['prob_0-0.1'] = probab1

valor2 =[]

for linha in range(0,len(modelo1)):
  conta2 = len(list(x for x in modelo1['e'][linha] if 0.1 < x <= 0.2))
  valor2.append(conta2)

divisor = 46
probab2 = [x/divisor for x in valor2]
modelo1['prob_0.1-0.2'] = probab2

valor3 =[]

for linha in range(0,len(modelo1)):
  conta3 = len(list(x for x in modelo1['e'][linha] if 0.2 < x <= 0.3))
  valor3.append(conta3)

divisor = 46
probab3 = [x/divisor for x in valor3]
modelo1['prob_0.2-0.3'] = probab3

valor4 =[]

for linha in range(0,len(modelo1)):
  conta4 = len(list(x for x in modelo1['e'][linha] if 0.3 < x <= 0.4))
  valor4.append(conta4)

divisor = 46
```

```
probab4 = [x/divisor for x in valor4]
modelo1['prob_0.3-0.4'] = probab4

valor5 =[]

for linha in range(0,len(modelo1)):
  conta5 = len(list(x for x in modelo1['e'][linha] if 0.4 < x <= 0.5))
  valor5.append(conta5)

divisor = 46
probab5 = [x/divisor for x in valor5]
modelo1['prob_0.4-0.5'] = probab5

valor6 =[]

for linha in range(0,len(modelo1)):
  conta6 = len(list(x for x in modelo1['e'][linha] if 0.5 < x <= 0.6))
  valor6.append(conta6)

divisor = 46
probab6 = [x/divisor for x in valor6]
modelo1['prob_0.5-0.6'] = probab6

valor7 =[]

for linha in range(0,len(modelo1)):
  conta7 = len(list(x for x in modelo1['e'][linha] if 0.6 < x <= 0.7))
  valor7.append(conta7)

divisor = 46
probab7 = [x/divisor for x in valor7]
modelo1['prob_0.6-0.7'] = probab7


valor8 =[]

for linha in range(0,len(modelo1)):
  conta8 = len(list(x for x in modelo1['e'][linha] if 0.7 < x <= 0.8))
  valor8.append(conta8)

divisor = 46
probab8 = [x/divisor for x in valor8]
modelo1['prob_0.7-0.8'] = probab8

valor9 =[]

for linha in range(0,len(modelo1)):
```

```
  conta9 = len(list(x for x in modelo1['e'][linha] if 0.8 < x <= 0.9))
  valor9.append(conta9)

divisor = 46
probab9 = [x/divisor for x in valor9]
modelo1['prob_0.8-0.9'] = probab9


valor10 =[]

for linha in range(0,len(modelo1)):
  conta10 = len(list(x for x in modelo1['e'][linha] if 0.9 < x <= 2.0))
  valor10.append(conta10)

divisor = 46
probab10 = [x/divisor for x in valor10]
modelo1['prob_0.9-1.0'] = probab10

####CALCULO ENTROPIA

#MODELO 1

import math
entropia = []

for linha in range(0,len(modelo1)):
  ent = 0.

# Compute entropy

  for i in modelo1.iloc[linha,6:16]:
    if i==0:
      ent = ent+ 0
    else:
      ent -= i * math.log(i, 2)
  entropia.append(ent)

modelo1['entropia'] = entropia
modelo1

# Tabela Minima Entropia - Combinacoes x Eficientes

tablemin = pd.pivot_table(modelo1, values='entropia', index=['eficientes'],
                          columns=['combinacoes'], aggfunc=np.ma.min)

tablemin.fillna("0.0")
```

```
# Tabela Maxima Entropia - Combinacoes x Eficientes

tablemax = pd.pivot_table(modelo1, values='entropia', index=['eficientes'],
                          columns=['combinacoes'], aggfunc=np.ma.max)

tablemax.fillna("0.0")

###DISTRUICAO DE FREQUENCIA

#MODELO 1

# Selecionando apenas altas entropias

modelo1_highentropies  = modelo1[modelo1['eficientes'] < 23]

def MakeList(x):
    T = tuple(x)
    if len(T) > 1:
        return T
    else:
        return T[0]

for_hist = modelo1_highentropies.groupby(['combinacoes','eficientes']).agg
    ({'entropia': MakeList})

histogramas1 = []
for linha in range(0,len(for_hist)):
  histogramas1.append((np.histogram(for_hist.iloc[linha]['entropia'], bins
      =[0.4967,0.7841,1.0714,1.3587,1.6461,1.9334,2.2207,2.5080,2.7954,3.0827])
      )[0])

df = pd.DataFrame(histogramas1, index=for_hist.index, columns = ('bins1','
    bins2','bins3','bins4','bins5','bins6','bins7','bins8','bins9'))


### FILTRANDO DISTRIBUICOES

#MODELO 1
selection1 = modelo1.loc[(modelo1['combinacoes'] == 2) & (modelo1['
    eficientes'] == 6)]

array = [6,7,9,11]
selection2 = modelo1.loc[(modelo1['combinacoes'] == 3) & modelo1['
    eficientes'].isin(array)]

array = [7,8,12]
selection3 = modelo1.loc[(modelo1['combinacoes'] == 4) & modelo1['
```

```
    eficientes'].isin(array)]

selection4 = modelo1.loc[(modelo1['combinacoes'] == 5) & (modelo1['
    eficientes'] == 13)]

result_modelo1 = pd.concat([selection1, selection2, selection3, selection4
    ])

contador = []

for linha in range(0,len(result_modelo1)):
  count = 0
  if ('1-45-47-93 - Total Num Funcionarios' in result_modelo1['w'].iloc[
      linha]
                or '3-15-15-33 - Num Funcionarios - Enfermagem' in
                    result_modelo1['w'].iloc[linha]):
    count += 1
  contador.append(count)

result_modelo1['verificador']=contador

# se verificador =1, mantenho a linha

result_modelo1 = result_modelo1[result_modelo1['verificador']==1]
analysis_1 = result_modelo1[['combinacoes','eficientes','entropia','w']]
analysis_1.set_index('w')


####DEA

!pip3 install pulp statsmodels scikit-learn pandas
!wget https://github.com/jzuccollo/pyDEA/archive/master.zip
!unzip master.zip -d .
!mv pyDEA-master/* .
!python3 setup.py sdist

from pydea import DEAProblem

dados_eficiencias_modelo1 = pd.read_excel('/content/gdrive/My Drive/PFC
    segunda
                tentativa/DEA Escores de Eficiencia.xlsx', sheetname='DEA -
    M1 3x7')
dados_eficiencias_modelo1_2 = pd.read_excel('/content/gdrive/My Drive/PFC
    segunda
                tentativa/DEA Escores de Eficiencia.xlsx', sheetname='DEA -
    M1 5x13')
dados_eficiencias_modelo2 = pd.read_excel('/content/gdrive/My Drive/PFC
```

```
    segunda
␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣tentativa/DEA␣Escores␣de␣Eficiencia.xlsx', sheetname='DEA␣-
    ␣M2␣4x16')

# Modelo 1 com 3 inputs e 7 combinacoes
#inputs = Total Num. Func., Num. Func. Pediatria e Area Total Espera e
    Recepcao
#output = Total Atendimentos

dados_secretaria.columns.values
X37= dados_secretaria[['93␣-␣Total␣Num␣Funcionarios', '69␣-␣Num␣
    Funcionarios␣-␣Pediatria',
         '103␣-␣Area␣Total␣Espera␣e␣Recepcao']]
y37= dados_secretaria[['94␣-␣Total␣Atendimentos']]

dea_prob = DEAProblem(X37, y37, returns='VRS')
results = dea_prob.solve()
VRS_=results['Efficiency']

DEA_eficiencias_modelo1=VRS_.to_frame()
DEA_eficiencias_modelo1.rename(columns={'Efficiency': 'VRS'},inplace=True)



dea_prob = DEAProblem(X37, y37, returns='CRS')
results = dea_prob.solve()
CRS_=results['Efficiency']

DEA_eficiencias_modelo1 = pd.concat([DEA_eficiencias_modelo1, CRS_.to_frame
    ()], axis=1)
DEA_eficiencias_modelo1 = pd.concat([DEA_eficiencias_modelo1,
    dados_eficiencias_modelo1[['DRS','IRS']]], axis=1)

DEA_eficiencias_modelo1['ESC'] = DEA_eficiencias_modelo1['CRS']/
    DEA_eficiencias_modelo1['VRS']
DEA_eficiencias_modelo1 =DEA_eficiencias_modelo1.round(decimals=3)

y=[]

for linha in range(0,len(DEA_eficiencias_modelo1)):

  if DEA_eficiencias_modelo1['ESC'][linha]==1:
    y.append('Otima')
  elif DEA_eficiencias_modelo1['VRS'][linha]==DEA_eficiencias_modelo1['DRS'
      ][linha]:
    y.append('Acima')
  else:
    y.append('Abaixo')
```

```
DEA_eficiencias_modelo1['Situacao']=y
DEA_eficiencias_modelo1

count = 0
for linha in range(0,len(DEA_eficiencias_modelo1)):
  if ('Otima' in DEA_eficiencias_modelo1['Situacao'].iloc[linha]):
    count += 1
count


####GRADIENT BOOSTING REGRESSOR

#retirando os inputs selecionados pelo Modelo 2

#-Total Num. Func., Num. Func. Enfermagem, Area Total Odontologia e Area
    Total Curativos

input_cols = []
for i in dados_secretaria.columns:
    if 'Atendimento' in i:
        continue
    input_cols.append(i)


X = dados_secretaria[input_cols]

variaveis_to_remove = ['93 - Total Num Funcionarios', '33 - Num
    Funcionarios - Enfermagem' ,     '98 - Area Total Odontologia', '100 -
    Area Total Curativos']
X.drop(variaveis_to_remove, axis=1, inplace=True)


X.columns = X.columns.str.replace(' ','')
X.columns = X.columns.str.replace('-','')
X_list = X.columns
X_list

y = DEA_eficiencias_modelo2['VRS']
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
    random_state=42)
# 20% de teste

##Gradient-boosted tree regression
```

```
from sklearn.ensemble import GradientBoostingRegressor
gbt = GradientBoostingRegressor()

## Shuffle Split

from sklearn.metrics import accuracy_score
from sklearn.cross_validation import ShuffleSplit

cv_ = ShuffleSplit(X_train.shape[0], n_iter=20, test_size=0.2, random_state
    =0)

## Grid Search

# testar quais melhores parametros para o meu modelo


param_grid = {

    'n_estimators': [20],
    'learning_rate':
        [0.0004,0.0005,0.0006,0.0007,0.0008,0.0009,0.001,0.002,0.003,0.004],
    'max_depth': [2]
}

from sklearn.model_selection import GridSearchCV

grid = GridSearchCV(estimator=gbt, param_grid=param_grid, n_jobs=-1,
    verbose=True, cv=cv_)
grid.fit(X_train, y_train)

grid.best_params_
grid.best_estimator_


###RESULTS

def plot_learning_curve(estimator, title, X, y, ylim=None, cv=None,
                        n_jobs=1, train_sizes=np.linspace(.1, 1.0, 5)):
    """
    Generate a simple plot of the test and traning learning curve.

    Parameters
    ----------
    estimator : object type that implements the "fit" and "predict" methods
        An object of that type which is cloned for each validation.

    title : string
```

```
        Title for the chart.

    X : array-like, shape (n_samples, n_features)
        Training vector, where n_samples is the number of samples and
        n_features is the number of features.

    y : array-like, shape (n_samples) or (n_samples, n_features), optional
        Target relative to X for classification or regression;
        None for unsupervised learning.

    ylim : tuple, shape (ymin, ymax), optional
        Defines minimum and maximum yvalues plotted.

    cv : integer, cross-validation generator, optional
        If an integer is passed, it is the number of folds (defaults to 3).
        Specific cross-validation objects can be passed, see
        sklearn.cross_validation module for the list of possible objects

    n_jobs : integer, optional
        Number of jobs to run in parallel (default 1).
    """
    plt.figure()
    plt.title(title)
    if ylim is not None:
        plt.ylim(*ylim)
    plt.xlabel("Training examples")
    plt.ylabel("Score")
    train_sizes, train_scores, test_scores = learning_curve(
        estimator, X, y, cv=cv, n_jobs=n_jobs, train_sizes=train_sizes)
    train_scores_mean = np.mean(train_scores, axis=1)
    train_scores_std = np.std(train_scores, axis=1)
    test_scores_mean = np.mean(test_scores, axis=1)*(-1)
    test_scores_std = np.std(test_scores, axis=1)*(-1)
    plt.grid()

    plt.plot(train_sizes, train_scores_mean, color="r",
             label="Training score")
    plt.plot(train_sizes, test_scores_mean, color="g",
             label="Cross-validation score")

    plt.legend(loc="best")
    return plt

from sklearn.learning_curve import learning_curve
title = 'Learning Curves' %grid.best_estimator_
gbt = GradientBoostingRegressor(alpha=0.9, criterion='friedman_mse', init=
    None,
```

```
                learning_rate=0.004, loss='ls', max_depth=2,
                max_features=None, max_leaf_nodes=None,
                min_impurity_decrease=0.0, min_impurity_split=None,
                min_samples_leaf=1, min_samples_split=2,
                min_weight_fraction_leaf=0.0, n_estimators=20, presort='auto',
                random_state=None, subsample=1.0, verbose=0, warm_start=False)
plot_learning_curve(gbt, title, X_train, y_train, cv=cv_)
plt.show()

gbt.fit(X_train, y_train)
gbt.score(X_train, y_train)

# make predictions for test data
predictions = gbt.predict(X_test)

# Calculate the absolute errors
errors = abs(predictions - y_test)


# Calculate mean absolute percentage error (MAPE)
mape = 100 * (errors / y_test)

# Calculate and display accuracy
accuracy = 100 - np.mean(mape)
print('Accuracy:', round(accuracy, 2), '%.')

feat_importances = pd.Series(gbt.feature_importances_, index=X_list)
feat_importances.nlargest(10).plot(kind='barh')


#instalar a biblioteca graphviz
!apt-get install graphviz -y

# Import tools needed for visualization
from sklearn import tree
from sklearn.tree import export_graphviz
#Modules to display decision tree
from IPython.display import Image

!pip3 install pydotplus
import pydotplus

sub_tree = gbt.estimators_[-1, 0]

dot_data = tree.export_graphviz(
    sub_tree,
    out_file=None, filled=True,
```

```
    rounded=True,
    feature_names = X_list,
    special_characters=True,
    proportion=True,
)
graph = pydotplus.graph_from_dot_data(dot_data)
Image(graph.create_png())
```