

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ADMINISTRAÇÃO
DEPARTAMENTO DE CIÊNCIAS ADMINISTRATIVAS
TÓPICOS ESPECIAIS EM MARKETING II - INTRODUÇÃO A DATA SCIENCE- ADM01007

Trabalho em Grupo – Saúde

NAIOME ALVES PEREIRA

Porto Alegre, 2023

Objetivo

Estudar a área da saúde, para identificar os municípios brasileiros mais adequados para a abertura de novos negócios nessa área.

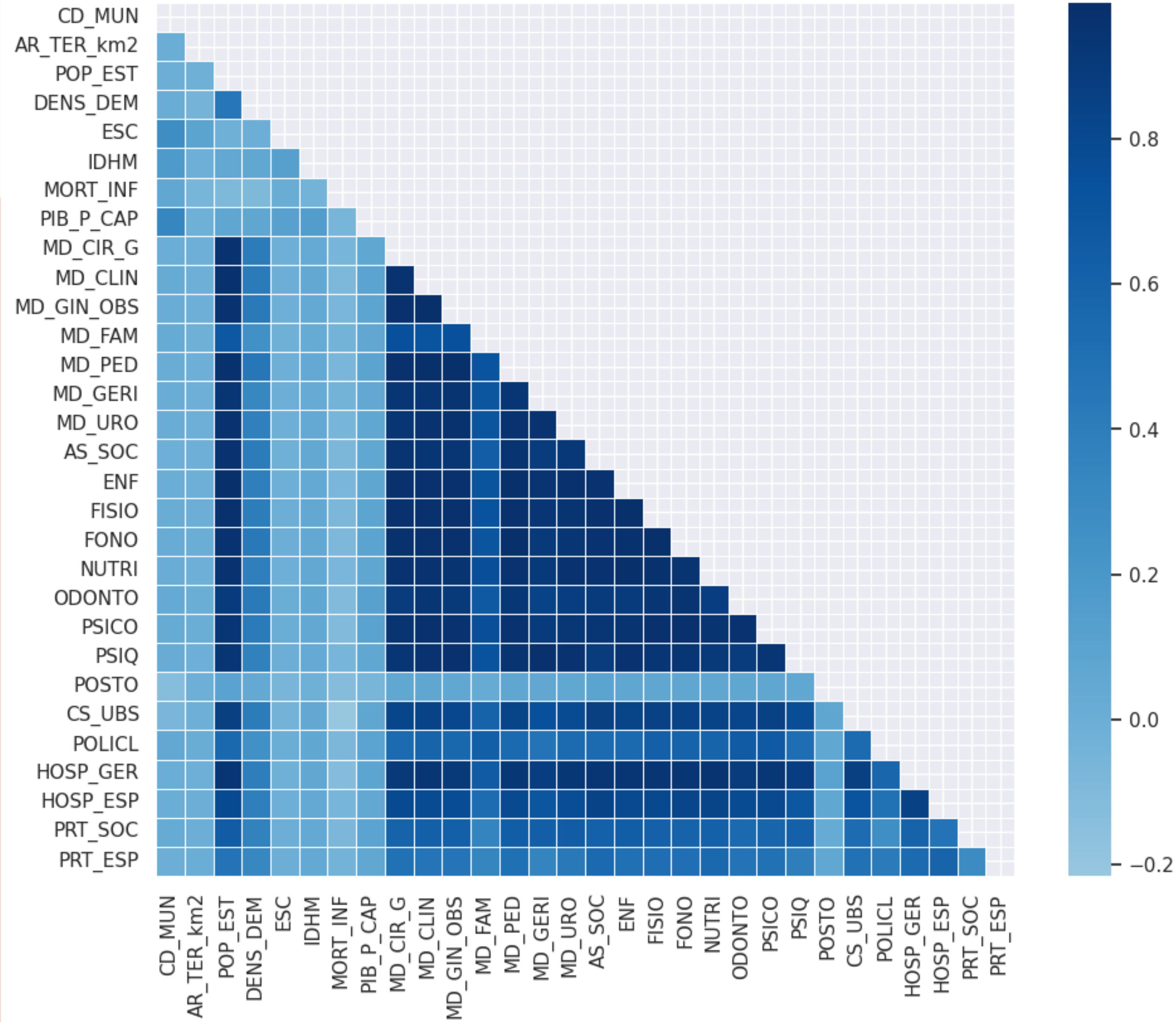
Pré-Processamento

Foi utilizado dados no Tabnet - DATASUS e IBGE com dados sobre população, área territorial, idh, mortalidade infantil, quantidade de estabelecimentos da saúde, e diversos profissionais da saúde por município.

Análise Exploratória e Descritiva

- Dataframe com 5570 linhas e 31 colunas
- Gráfico Person
- Gráfico de Contagem

Gráfico Person



Análise Inferencial

A partir da Análise Exploratória, decidiu-se verificar a relação dos profissionais da saúde disponíveis nos municípios com a quantidade de estabelecimentos de saúde ofertado pelo SUS, pois os estabelecimentos de saúde não apresentam uma correlação com quantidade de profissionais ou a estimativa populacional

Análise Inferencial

- Criação da coluna QTD_ESTAB
- Criação da coluna QTD_ESP
- Definindo a variável Y - TAXA_ESTAB

Modelos

Para descobrirmos quais são as possíveis variáveis que nos de entendimento da distribuição de estabelecimentos de saúde em relação aos profissionais disponíveis, utilizaremos modelos de predição por aprendizado de máquina.

- Decision Tree
- Random Forest
- MLP

Conclusão

Visto que nos três modelos testados (dois baseados em árvore e um baseado em rede neural) tiveram resultados similares e esses modelos obtiveram ótimos resultados de previsões (aproximadamente 99% de acerto) decidimos por não testar outros modelos, pois dificilmente teríamos um resultado melhor. Nesse conjunto de dados com o ajuste dos hiperparâmetros, assim como o escalonamento feito, no qual já deixa todas as variáveis com pesos semelhantes, não viu-se a necessidade de utilizar regularização nos modelos.

Conclusão

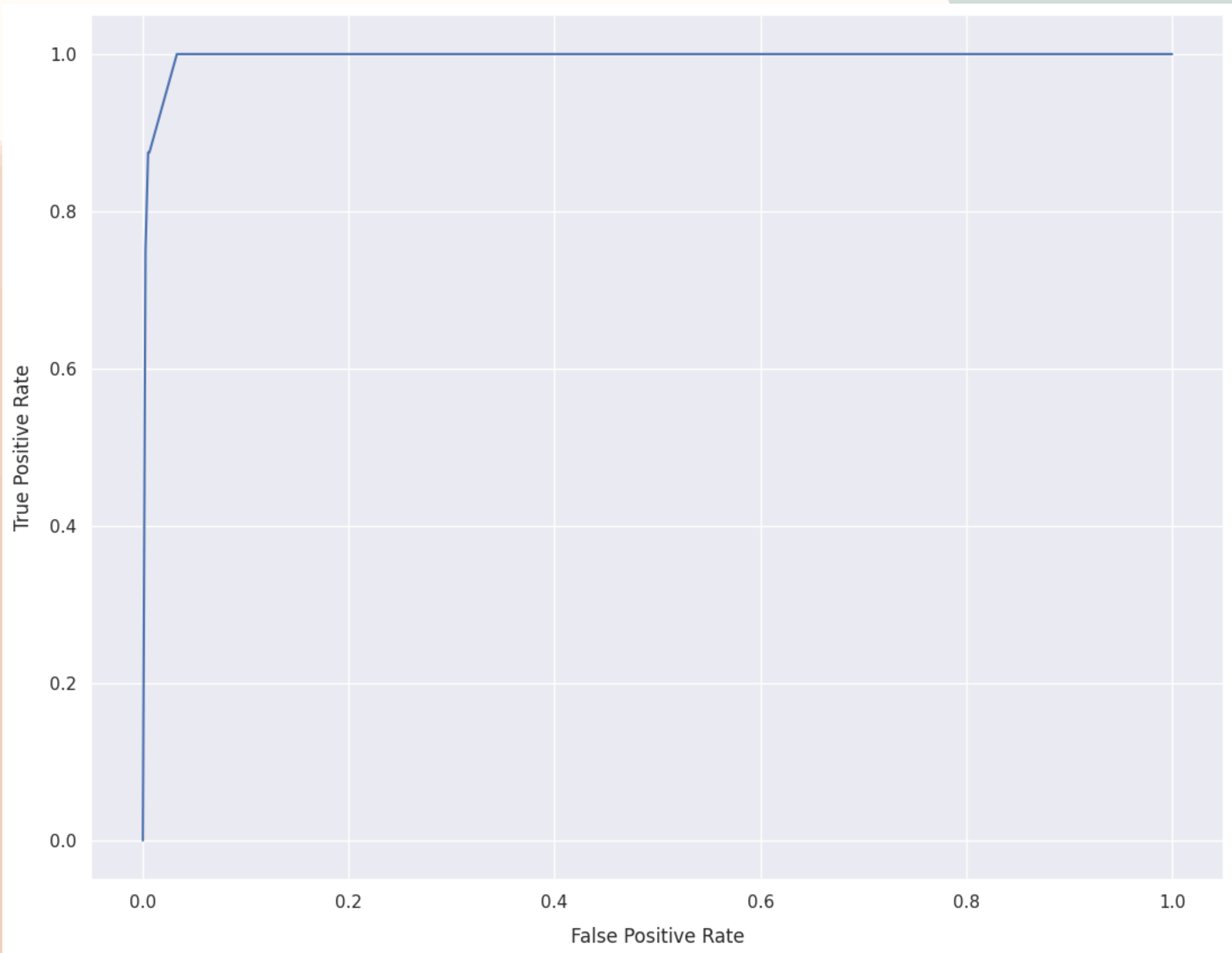
Para a escolha do melhor modelo foi utilizada as seguintes métricas:

- Acurácia,
- Area Under the Curve (AUC),
- f1-score, recall, precision.

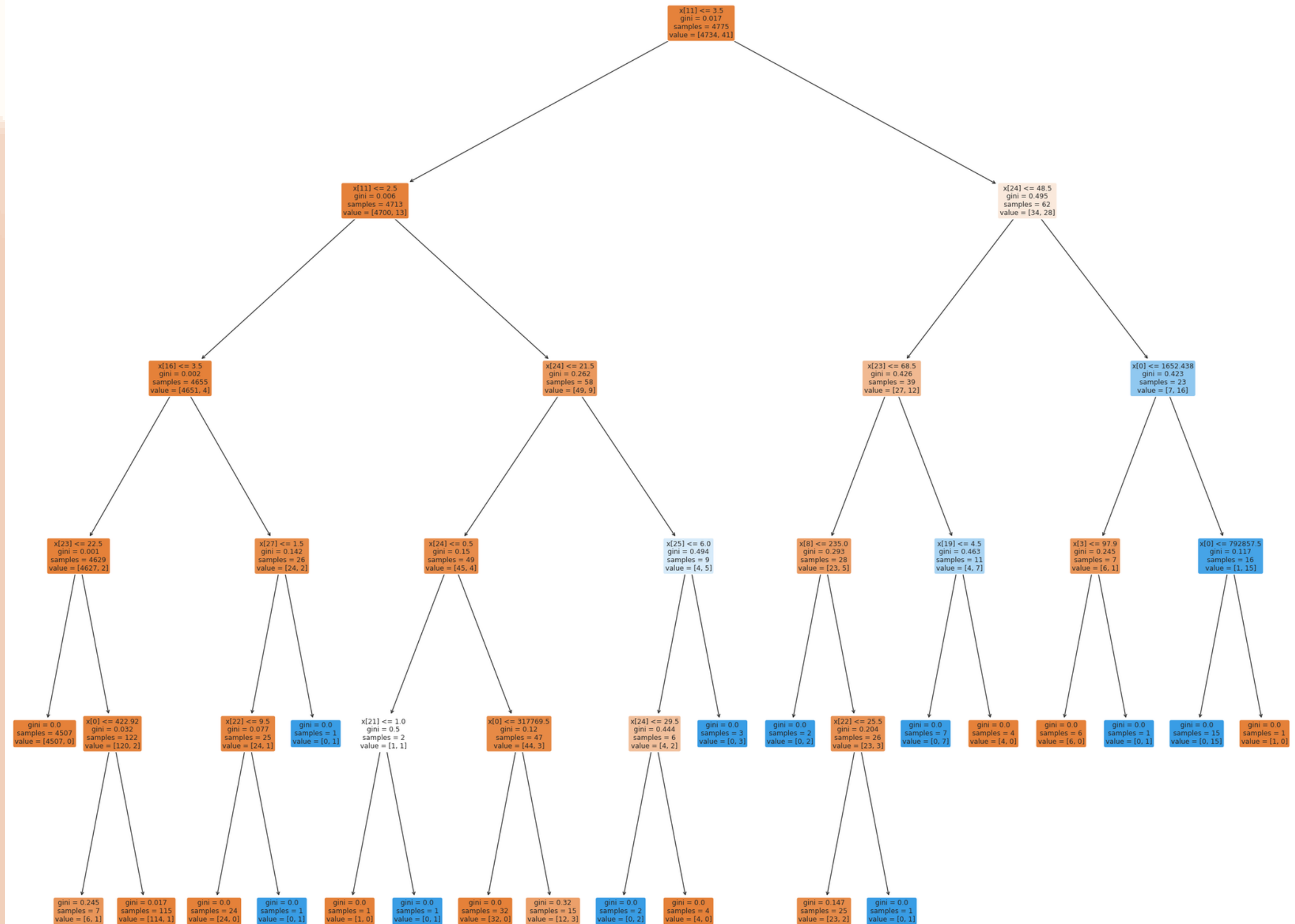
Decision Tree

- Reamostragem por Cross Validation
- Acurácia média obtida foi de 99.14% com um desvio padrão de 0.16%
- Acurácia de treinamento 99,85% e Acurácia de teste 99,49%
- Hiperparâmetro max depth (5)

AUC



Decision Tree



Realizado o feature importance nos três modelos a variável Policlínica (POLICL) apareceu como uma das principais. E a árvore de decisão nos apresentou a mesma variável como de maior importância global para a variável resposta. Sendo assim, nos a utilizamos como medida de qual cidades seriam ideais abrir nosso negócio.

Estimativa de negócio

Em 2 cidades (Jangada-MT e Morro do Pilar-MG) não possuem nenhum tipo de estabelecimento hospitalar público. E em 427 municípios possuem apenas uma UBS.

Esses dados nos mostram que há uma estimativa de no mínimo 429 cidades que não possui médicos especialista para sua população. Nos apresentando um potencial de mercado para a abertura de clínicas multidisciplinares, que contaria com diversos profissionais da saúde e com um preço da consulta de valor competitivo. Além de ser possível fazer uma parceria com o governo, para auxiliar a aliviar a longa fila do SUS desses municípios.