

A relação entre Mobilidade Social, Medidas Restritivas e Casos de COVID-19 no Estado de Alagoas: Um Estudo de Caso com Uso de Ciência de Dados

Lucas A. Lisboa¹, Lucas M. A. Assunção¹, João Messias Lima Pereira¹

¹Instituto de Computação – Universidade Federal de Alagoas (UFAL)
Maceió – AL – Brasil

{lal, lmaa, jmp}@ic.ufal.br

Abstract. *The COVID-19 pandemic has significantly impacted all countries in the world, generating a great concern for state managers to avoid a high number of confirmed cases and, consequently, deaths. For this, several measures were adopted, aiming to limit the social mobility of individuals. In this way, the present work aims to evaluate which types of mobilities most influence the number of cases and which decrees in the state of Alagoas proved to be more effective. Based on the use of the Google COVID-19 Community and Bing COVID-19 Data databases, the Random Forest Classifier algorithm was used, obtaining strong indications of the need for attention regarding public transport, classroom classes and restaurants.*

Resumo. *A pandemia da COVID-19 impactou de maneira significativa todos os países no mundo, gerando uma grande preocupação dos gestores estatais em evitar um alto número de casos confirmados e, consequentemente, de mortes. Para isso, diversas medidas foram adotadas, visando limitar a mobilidade social dos indivíduos. Dessa forma, o presente trabalho visa avaliar quais os tipos de mobilidades que mais influenciam no número de casos e quais os decretos no estado de Alagoas se mostraram mais efetivos. Partindo do uso das bases de dados Google COVID-19 Community e Bing COVID-19 Data, foi utilizado o algoritmo Random Florest Classifier, obtendo fortes indícios da necessidade de atenção quanto ao transporte público, aulas presenciais e restaurantes.*

1. Aplicação

Os anos de 2020 e 2021 foram marcados pela pandemia da COVID-19, alterando significativamente a dinâmica das relações sociais e econômicas em todo o mundo, tendo ocasionado mais de 370 milhões de casos, com cerca de 5,66 milhões de mortes¹ [Microsoft 2021]. Nessa perspectiva, uma preocupação das gestões estatais consiste em evitar o maior número possível de casos e, consequentemente, de mortes. Para diminuir tais valores, diversos tipos de medidas de restrições de mobilidade social foram adotadas, dentre elas: fechamento de estabelecimentos comerciais, diminuição da taxa de ocupação máxima, redução do horário de atendimento, adoção de trabalho remoto, entre outros. Dessa forma, o presente trabalho tem o intuito de avaliar a efetividade das medidas adotadas no estado de Alagoas, mediante os decretos estaduais publicados no período entre

¹Dados do dia 31/01/2021

08/04/2020 e 08/01/2022. Além disso, foram avaliados quais pontos de mobilidade mais influenciam nas taxas de casos confirmados e mortes.

Tal estudo tem como trabalho correlato o artigo de [Sun et al. 2021] o qual avaliou as medidas restritivas adotadas nos Estados Unidos. Este tipo de estudo contribui em esquematizar políticas públicas perante epidemias, pois é possível mensurar quais as medidas são mais efetivas no combate desse tipo de calamidade. Outro aspecto relevante é observar quais locais e atividades influenciam mais no aumento do número de casos e no número de mortes

2. Experimentos

2.1. Base de Dados

Para realizar este trabalho, foram utilizadas duas principais bases: Google COVID-19 Community [Ritchie et al. 2020] e a Bing COVID-19 Data [Microsoft 2021]. A primeira consiste numa base de dados acerca da Mobilidade Social disponibilizada pela Google, a qual registra os valores de deslocamento dos usuários Android ao longo do tempo. São utilizados seis parâmetros: varejo e lazer; mercados e farmácias; parques; estações de transporte público; locais de trabalho e áreas residenciais. Os valores são descritos como porcentagem, podendo ser positiva ou negativa. Tal porcentagem usa como valor de referência a mediana de mobilidade registrada entre os dias 03/01/2020 e 06/02/2020. Dessa forma, quanto maior o valor, maior é a média geral das pessoas que frequentam aquele determinado ambiente, apenas a áreas residenciais que representam o tempo gasto pelas pessoas. Já a segunda consiste numa base de dados acerca do número de casos confirmados e do número de mortes disponibilizada pela Microsoft.

Adicionalmente, foi criada uma base extra pelo próprios autores, a qual consiste na representação dos decretos emitidos pelo Governo do Estado de Alagoas. Tal base consiste em 21 atributos (ou colunas), sendo eles: Data Inicial do Decreto; Data Final do Decreto; Horário sem Restrição; Restaurantes, bares, lanchonetes e estabelecimentos congêneres; Shows e Música ao Vivo; Lojas do centro; Lojas de rua e galerias; Shoppings; Aula em Instituição Privada (Crianças); Aula em Instituição Privada (Adultos); Aula em Instituição Pública (Crianças); Aula em Instituição Pública (Adultos); Academias; Praias, calçadões, lagoas, rios e clubes; Templos, igrejas e demais instituições religiosas; Salões de beleza e barbearias; Transporte Intermunicipal e Turístico; Eventos sociais, corporativos e celebrações; Jogos do Campeonato de Futebol; Cruzeiros e outras embarcações de passageiros de grande porte; e Cinemas, teatro e museu. Com exceção dos dois primeiros (que são preenchidos com datas), todos os atributos recebem um valor entre 0 e 1, sendo o 0 representando a total restrição e 1 a total liberação, então, por exemplo, um valor de 0,5 representa 50% da capacidade de ocupação. Vale destacar de que, se cidades diferentes tinham níveis de restrições diferentes, o critério adotado era adotar a restrição mais flexível no momento em questão, então, por exemplo, se na capital era permitido 0,5 de capacidade e no interior apenas 0,3, foi optado para entrar na base o valor de 0,5.

O primeiro passo do pré-processamento foi concatenar os três datasets das bases principais (Google e Microsoft): 2020, 2021 e 2022. Com os datasets concatenados, tem-se que ambas as bases têm seus dados descritos dia a dia, possuindo uma coluna para especificar a região descrita. Assim, o segundo passo foi realizar o filtro dos dados para retornar apenas as linhas referentes à Alagoas, eliminando as demais regiões do Globo da

análise. Após isso, foram eliminadas as colunas que não seriam relevantes para o estudo, tais como: Código da Região, Tamanho da Área, Código Id, Código ISO, Código do CENSO.

Nesse momento, havia alguns dias com dados faltantes em determinado atributo. A solução adotada foi preencher as células faltantes com a média das cidades de Alagoas naquele dia em específico (Base do Google). Depois deste ponto, os nomes das regiões e cidades não são mais necessários e são eliminados. O próximo passo foi alinhar o período de ambas as bases, pois cada base iniciava em dias distintos. Dessa forma, ao final, obteve-se o período entre 08/04/2020 e 10/01/2022. Por fim, as duas bases principais foram juntadas.

2.2. Estatística descritiva e inferência

Aplicando a Correlação de Pearson entre os atributos em relação ao número de casos confirmados, obteve-se o seguinte resultado:

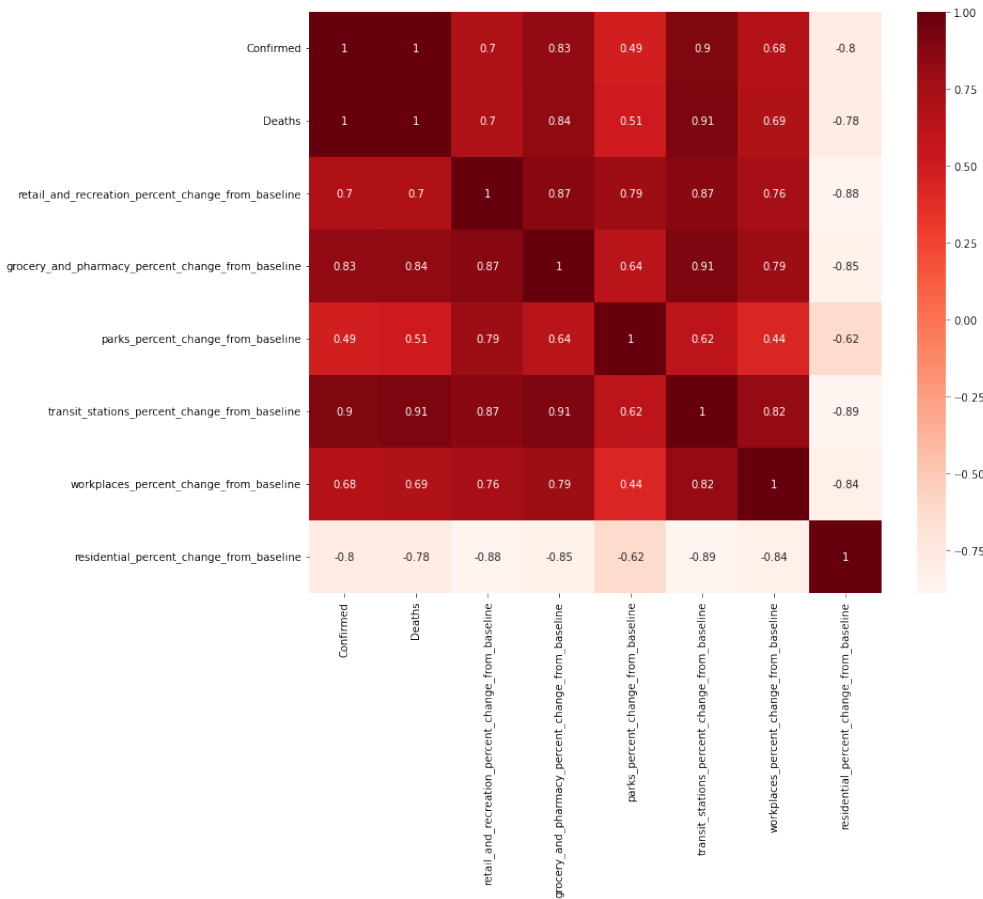


Figura 1. Matriz de Correlação de Pearson: Mobilidade e Casos



Figura 2. Matriz de Correlação de Pearson: Restrições e Casos

As observações iniciais na Figura 1 confirmam a relação intuitiva entre casos confirmados e mortes, isto é, quanto mais casos, mais mortes, com uma relação de 100% de correlação positiva. Já com relação às variáveis de mobilidade, com exceção da variável residencial, todas tiveram uma correlação positiva, enquanto a residencial apresenta uma correlação negativa, ou seja, quanto mais pessoas passando mais tempo em casa, menos casos confirmados. Já na Figura 2, tem-se uma forte correlação com o número de casos ambientes / atividades como aulas presenciais, shoppings, lojas, shows, bares, restaurantes, cinema, museu e teatro, enquanto a restrição de horário não parece ser muito significativo (as demais medidas restritivas obtiveram uma correlação média).

Já aplicando os gráficos de séries temporais para cada um dos atributos, tem-se a seguinte distribuição:

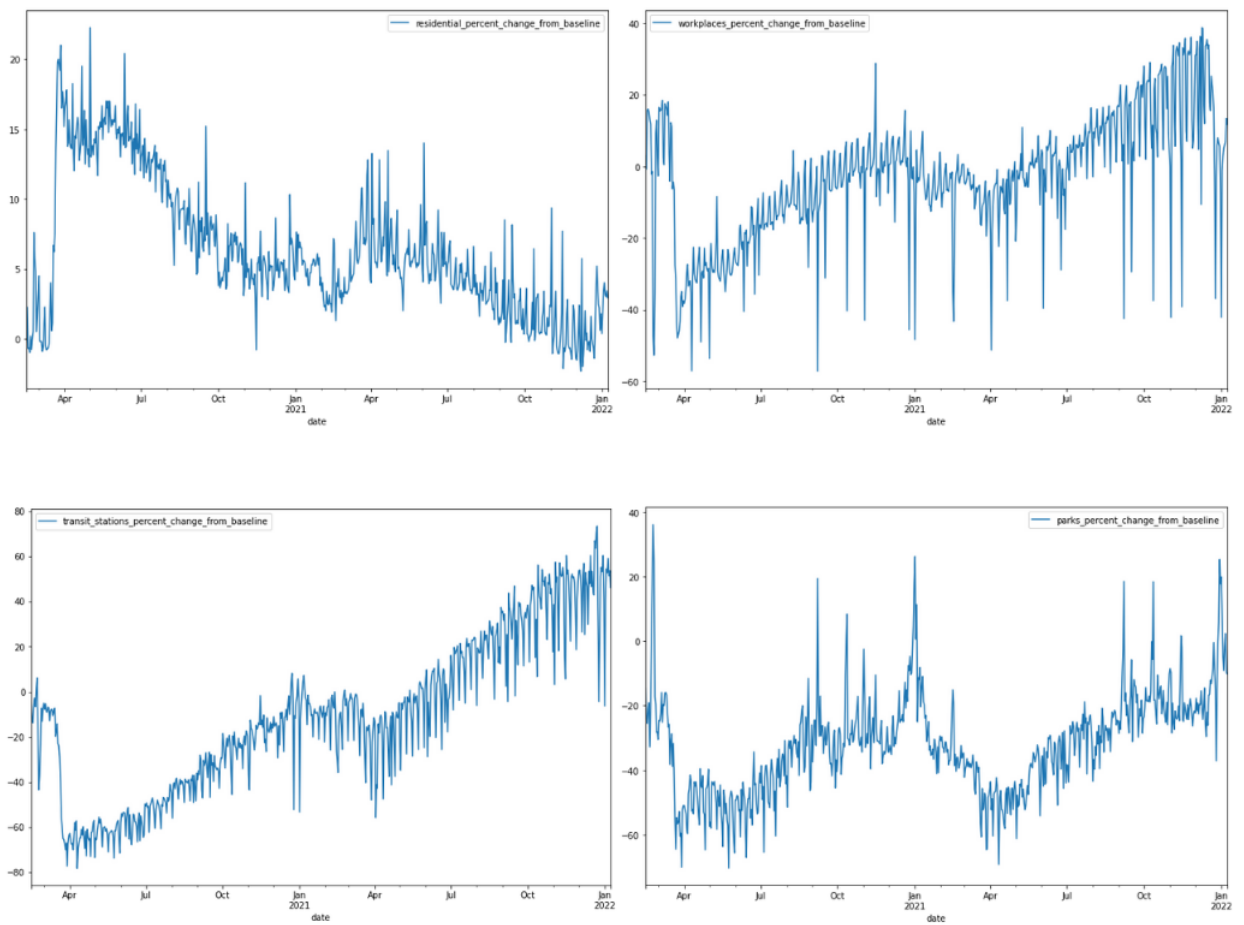


Figura 3. Mobilidade x Tempo

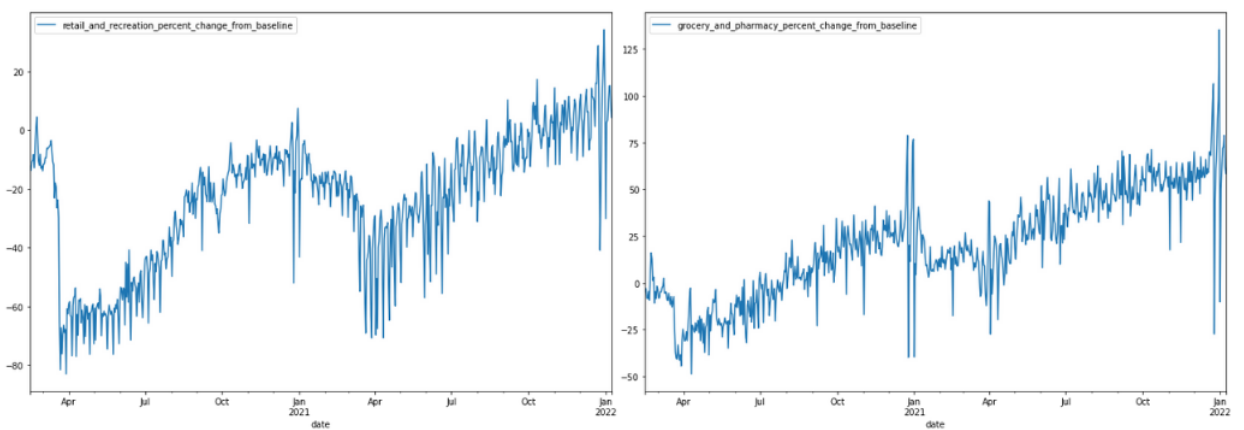


Figura 4. Mobilidade x Tempo

Para a manipulação do dados e plotagem dos gráficos nesta etapa, foram utilizadas as bibliotecas Pandas, Datetime, Seaborn e Matplotlib, todas para a linguagem Python.

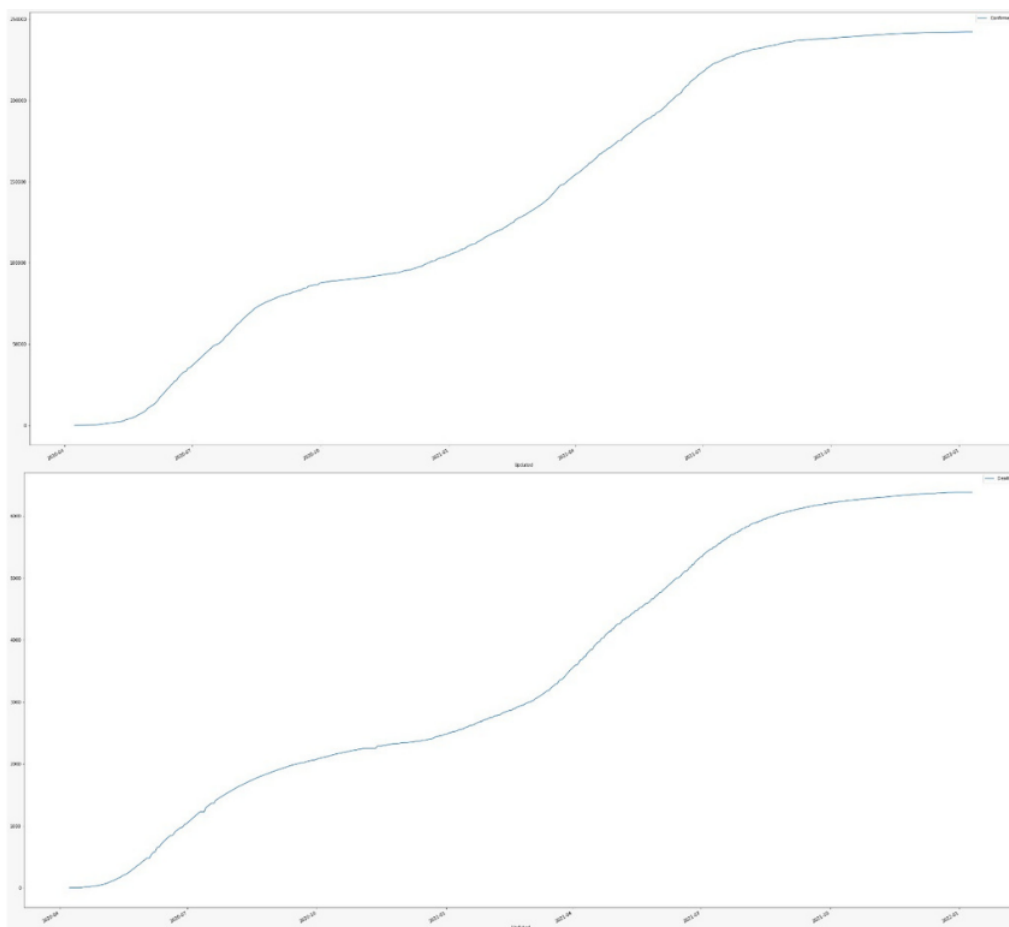


Figura 5. Casos Confirmados e Mortes x Tempo

2.3. Métodos avaliados

O método de Aprendizado de Máquina adotado foi o Random Florest Classifier, o qual gerou diversas árvores de classificação e, ao final do processo, pegou a média de cada feature, gerando um ranking de importância de cada atributo analisado. No caso em questão, foi utilizado para obter a relação de medidas mais importantes e significativas adotadas no período. Para isso, foi utilizado a biblioteca Scikit-learn (Python) com a configuração *default*, com exceção da entropia, a qual utilizou o padrão *gini*, e o parâmetro *random_state*, que foi utilizado o valor 13.

2.4. Métricas de avaliação

Como métricas para validação, foram adotadas F1-Score e Precisão, pois julgou-se mais relevante para o gestor estatal que lidaria com as respostas de tal estudo que as respostas fossem mais precisas e acertadas do que cobrir todas as variáveis, tendo em vista que uma decisão errada implica em um número significativo de vidas perdidas, além dos impactos econômicos acarretados no processo.

2.5. Métodos de avaliação

Para validar o experimento, foi utilizado o método K-Fold Cross Validation, com o número de partições (k) sendo igual a 5. A escolha desse valor para k se deu pela

quantidade de instâncias para teste, tendo em vista uma quantidade maior de partições representaria um número de baixo de instâncias de teste.

3. Resultados

Executando o algoritmo Random Florest Classifier, obtém-se o seguinte resultado:

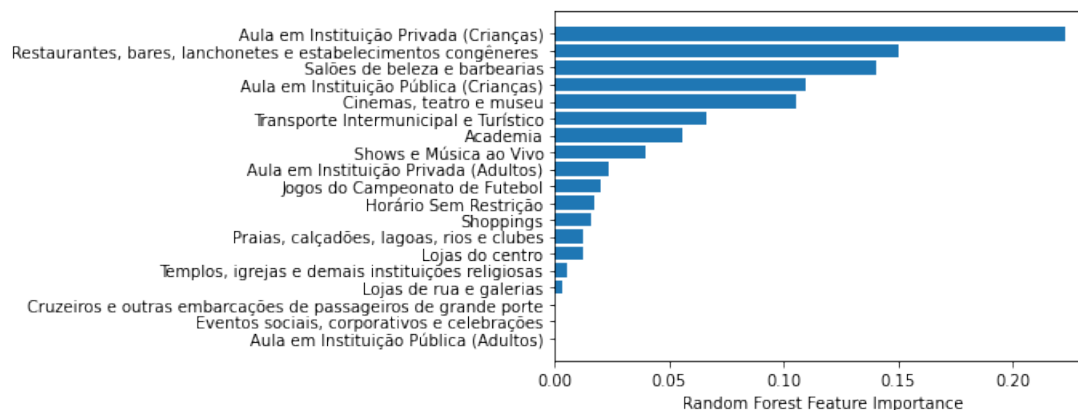


Figura 6. Decretos mais Efetivos

Ao analisar o gráfico, percebe-se a confirmação de alguns fatores já observados na correlação de Pearson, como as aulas presenciais para criança e bares e restaurantes. Por outro lado, também é perceptível a importância outrora com baixa correlação, como salão de beleza e barbearia.

As métricas obtidas foram 0,6 para a precisão e 0,56 para F1-Score. Esses valores podem ser considerados aceitáveis que há uma baixa quantidade de decretos em relação à base de dias presente nas bases do Bing e do Google, servindo como forte indício da efetividade dessas medidas. No entanto, para generalização, é necessário mais testes envolvendo mais decretos, bem como verificar o desempenho com o uso de outro algoritmo.

4. Conclusão

Os resultados obtidos revelam que medidas que limitam o público em ambientes fechados, com bastante fluxo de pessoas e que precisam tirar a máscara (no caso de bares e restaurantes) possui grande importância para evitar o aumento de casos, enquanto medidas como a restrição de horário não possui impactos significativos. Outro ponto a se observar é que, dentre as métricas de mobilidade, o transporte público revelou ter bastante correlação com o aumento de casos. Dessa forma, recomenda-se a adoção de medidas como o aumento da frota, limpeza frequente do ambiente, diminuição da ocupação e adoção do trabalho home-office para contornar esse problema.

Destaca-se como trabalhos futuros a necessidade de avançar nos experimentos, ampliando a base de decretos, em especial com a adição de novos estados na análise. Além disso, realizar experimentos com outros algoritmos de Aprendizagem de Máquina também se mostra importante. Assim, será possível generalizar os resultados obtidos.

Referências

Microsoft (2021). Bing covid-19 data. <https://github.com/microsoft/Bing-COVID-19-Data>.

- Ritchie, H., Mathieu, E., Rodés-Guirao, L., Appel, C., Giattino, C., Ortiz-Ospina, E., Hasell, J., Macdonald, B., Beltekian, D., and Roser, M. (2020). Coronavirus pandemic (covid-19). *Our World in Data*. <https://ourworldindata.org/coronavirus>.
- Sun, J., Kwek, K., Li, M., and Shen, H. (2021). Effects of social mobility and stringency measures on the covid-19 outcomes: Evidence from the united states. *Frontiers in Public Health*, 9.