

A UCS É
PRA VOCÊ
QUE CRIA O
FUTURO.



XXIX Encontro de Jovens Pesquisadores
e XI Mostra Acadêmica de Inovação e Tecnologia

De 5 a 7/10

Local: UCS - Cidade Universitária,
Caxias do Sul

www.jovenspesquisadores.com.br



APLICAÇÃO DE TÉCNICAS DE CADEIAS DE MARKOV PARA MAXIMIZAR A ACURACIDADE DE MÉTODO DE MÉDIAS PARA PREVISÃO DE COVID-19 NO RS

PROJETO: OTIMIZAQUANT

Autores: Andressa Daiane Ferrazza (PIBIC-CNPq),
Leandro Luís Corso (Orientador)



PIBIC-CNPq

INTRODUÇÃO / OBJETIVO

O Brasil é um dos países mais impactados pela *Coronavirus Disease 2019* (COVID-19), ocupando o segundo lugar na classificação dos países com maior número total de casos e de mortes pela doença. No país, o Rio Grande do Sul (RS) encontra-se na 4ª e 5ª posição em relação ao maior número de casos e óbitos por COVID-19, respectivamente. Ademais, devido à imprevisibilidade do comportamento da nova doença na população, os modelos matemáticos, especialmente Cadeias de Markov (CM), estão gradativamente sendo mais aplicados na área da saúde como auxílio para o controle e previsão de dados associados à nova infecção.

O modelo de CM consiste num processo estatístico, no qual busca a previsão do próximo estado levando em consideração unicamente o estado atual, sem interferência de estados anteriores que possam levar a vieses. Todavia, para a aplicação desse método, é necessário uma base de dados prévia sólida a fim de obter previsões precisas. Assim, constatou-se que a falta de uniformidade na liberação dos dados pelo sistema de saúde gera grandes oscilações na coleta total de dados, prejudicando posteriormente o aprendizado de máquina para a previsão de dados futuros. Logo, o estudo tem por objetivo utilizar o modelo de CM para avaliar o período mais estável para o cálculo da previsão de casos e de óbitos por COVID-19, também feito pelo mesmo modelo matemático de Cadeias de Markov, sendo possível prever as novas taxas da infecção de forma contínua e acurada.

METODOLOGIA

Foi realizada uma análise das variações do número de casos e de óbitos por COVID-19 no RS entre janeiro e junho de 2021. Os dados foram fundamentados pelo *Brasil.io Datasets*, com respaldo do Instituto Brasileiro de Geografia e Estatística (IBGE), totalizando 160 dados. A metodologia completa das CM foram aplicadas três vezes com diferentes intervalos de dias (3, 5 e 7 dias, nomeados respectivamente MM3, MM5 e MM7, onde MM significa “média móvel”) sobre os dados prévios. Posteriormente, foram avaliados os resultados comparativamente de forma quantitativa e gráfica.

RESULTADOS E DISCUSSÃO

Primeiramente, foram realizados as sequências de cálculos de cada período (3, 5, e 7 dias), dividindo as taxas de variações dos casos de COVID em faixas de análise, denominadas A, B, C, D, E e F (Tabela 1), a fim de quantificar as variações de casos. Posteriormente, a análise comparativa entre as CM feitas com 3, 5 e 7 dias de intervalo (Gráfico 1)

RESULTADOS E DISCUSSÃO

entre a atualização com os novos dados demonstraram que o modelo matemático é mais eficiente para realizar previsões quando utilizado o intervalo de 7 dias (MM7). Este obteve 160 acertos dos 160 dados (100%) analisados ao avaliar-se o total de acertos das previsões do sentido da variação. Os intervalos de 3 e 5 dias (MM3 e MM5), entretanto, obtiveram menos acertos: 133 (83%) e 68 (43%), respectivamente. Ademais, estudado de forma gráfica, a MM7 dias apresentou mais suavidade na transição das taxas diárias de casos e óbitos por COVID-19 e, conseqüentemente, menos oscilações bruscas em extremidades em comparação com os outros intervalos. Em relação ao período de 7 dias apresentar melhores resultados, pode-se atribuir esse resultado ao delongamento da liberação de todos os dados de casos e óbitos por COVID-19 pelo sistema público de saúde.

Taxas diárias de casos COVID-19 por Faixas						
	A	B	C	D	E	F
Mínima	-infinito	>-10%	>-5%	>0%	>5%	>10%
Máxima	-10%	-5%	0%	5%	10%	+infinito

Tabela 1 – Taxas diárias de casos COVID-19 segmentadas em faixas denominadas A, B, C, D, E e F, as quais cada faixa corresponde a um intervalo percentual de casos, compreendido entre os valores percentuais citados em mínima e máxima.

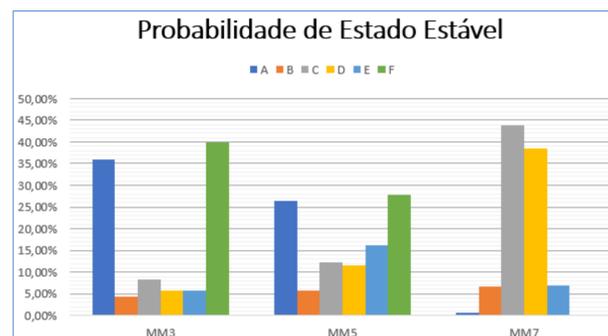


Gráfico 1 – Probabilidade de Estado Estável de cada intervalo avaliado (MM3, MM5, MM7), de acordo com as taxas diárias de casos COVID-19 determinados por faixas A, B, C, D, E, F.

CONCLUSÕES

Conclui-se que no período de 7 dias todos os dados prévios da doença são totalmente enviados ao sistema, permitindo que o modelo matemático interprete os dados de forma correta para posteriormente gerar uma previsão acurada dos próximos dias. Desse modo, foi necessário utilizar a própria abordagem matemática de Cadeias de Markov, utilizada para a previsão dos dados da infecção, para também aperfeiçoar a acurácia dos mesmos. Os resultados finais demonstraram substancialmente eficazes para gerar previsões extremamente confiáveis do curso da COVID-19 no RS, sendo um potencial modelo para assessorar o sistema de saúde estadual.

REFERÊNCIAS BIBLIOGRÁFICAS

- COVID-19 United States Cases by County - Johns Hopkins Coronavirus Resource Center [Internet]. Johns Hopkins Coronavirus Resource Center. 2021 [cited 10 July 2021]. Available from: <https://coronavirus.jhu.edu/us-map>
- Coronavírus Brasil [Internet]. Covid.saude.gov.br. 2021 [cited 10 July 2021]. Available from: <https://covid.saude.gov.br/>
- Rio Grande do Sul | Cidades e Estados | IBGE [Internet]. Ibge.gov.br. 2021 [cited 15 July 2021]. Available from: <https://www.ibge.gov.br/cidades-e-estados/rs.html>
- Grigoletti, PS. Cadeias de Markov. 19.10 (2011): 2014.