



Simpósio de Integração Acadêmica

“Bicentenário da Independência: 200 anos de ciência, tecnologia e inovação no Brasil e 96 anos de contribuição da UFV”

SIA UFV 2022



Análise Descritiva de Dados Digitais Públicos da Saúde nos Cenários Social e Regional

Palavras-Chave: Mineração de Dados, Clusterização, Banco de Dados Geográficos, Epidemiologia, Leptospirose

Autor: Ricardo Bortolucci Patriani de Carvalho, estudante de Sistemas de Informação UFV-CRP (ricardo.patriani@ufv.br)

Professora Orientadora: Adriana Zanella Martinhago (adriana.martinhago@ufv.br)

Introdução

Esta pesquisa representa o Trabalho de Conclusão de Curso trazendo a Mineração de Dados (MD) como abordagem computacional para estudo epidemiológico do estado de São Paulo perante a Leptospirose. A epidemiologia é a investigação de casos de uma certa doença a fim de encontrar causas e impactos na sociedade (BRITO, 2012). Para essa busca em meio a tantos dados, a MD se destaca por seus algoritmos que podem lidar com alto volume de registros multivalorados. Dentre as diversas áreas que ela abrange foi selecionada a abordagem descritiva, utilizando o processo KDD (*Knowledge Discovery in Databases*) com foco em clusterização (agrupamento) de dados. Após a validação e extração de conhecimento dos grupos formados, pretende-se gerar mapas por meio de Banco de Dados Geográficos para facilitar a visualização e interpretação dos resultados por humanos.

Objetivos

Analisar os casos de Leptospirose por municípios do estado de São Paulo em âmbito social e enriquecê-los com referenciamento geográfico para produção de paisagem digital. A finalidade da pesquisa é: facilitar estratégias de gestão de riscos e controle de reincidência, além de fomentar a exploração da transparência governamental.

Material e Métodos

Como citado anteriormente a pesquisa é lastreada pelo KDD, representado pela Figura 1 abaixo :



Figura 1 - Processo descrito pelo KDD com uso do agrupamento de dados (XU; WUNSCH, 2005 - adaptada)

Nesse estudo do caso de São Paulo, foram utilizadas as bases de dados do DATASUS e do IBGE, ambas as bases relativas à série histórica de 2007 a 2020, para se obter respectivamente

informações sobre a leptospirose e dos municípios. Já as etapas intermediárias do KDD são focadas em aumentar o desempenho dos algoritmos descritivos utilizados (no caso são: K-means e clusterização hierárquica), em que o pré-processamento combinou diversas formas de normalização dos dados para que os algoritmos fossem bem explorados e avaliados de acordo com uma métrica, validação. Por fim, os resultados são submetidos a métodos estatísticos para se finalizar, ou não, o processo (KDD permite ciclos de etapas). Desde o pré-processamento até a interpretação dos dados está sendo utilizado a linguagem Python e sua biblioteca scikit-learn e no próximo processo, criação de mapas, será necessário um ambiente com: Sistema de Gerenciamento de Dados (SGBD) com extensão espacial PostGis e um Sistema de Informação Geográfica (SIG), o QGIS.

Resultados Esperados e Discussão

A pesquisa já retornou agrupamentos de municípios que estão sendo analisados por meio da estatística descritiva. Com isso espera-se encontrar os melhores agrupamentos e qual informação eles estão comunicando, uma relação generalizável entre IDHM e número total de casos, por exemplo.

Conclusões

O processo geral desta pesquisa consiste em se aproveitar da captura do mundo real em digital para descoberta de informações relevantes para saúde pública. Observa-se também que para aproximar os resultados do mundo real, novamente, deseja-se formar uma paisagem digital com mapas que contextualizam os grupos de municípios em dispersão geográfica.

Bibliografia

BRITO, M. A. de. *Bonita r, beaglehole r, kjellstrom t. epidemiologia básica. são paulo:Grupo editorial nacional; 2010. Ciência & Saúde Coletiva, Associação Brasileira de Pós-Graduação em Saúde Coletiva, v. 17, n. 6, p. 1657–1658, 2012.*

XU, R.; WUNSCH, D. Survey of clustering algorithms. *IEEE Transactions on neural networks, IEEE, v. 16, n. 3, p. 645–678, 2005.*