ARTIGO SIMPLES (RESUMO) - SISTEMAS DE MONITORAMENTO E ALERTA CLIMÁTICOS

AVALIAÇÃO DA IMPORTÂNCIA DA INCLUSÃO DE VARIÁVEIS METEOROLÓGICAS EM MODELOS DE PREVISÃO DE HOSPITALIZAÇÃO POR DOENÇAS RESPIRATÓRIAS – PROJETO CAIPORA

Gregori De Arruda Moreira (gregori.moreira@ifsp.edu.br)

Alexandre Cacheffo (cachefo@ufu.br)

Fábio Juliano Da Silva Lopes (fjslopes@unifesp.br)

Izabel Andrade (sandradeizabel@gmail.com)

Laura Silva Pelicer (laura.pelicer@usp.br)

Marlon Miguel Cedeno Puig (marlonsamba83@yahoo.es)

Samara Carbone (samara.carbone@ufu.br)

In the last decade, the use of techniques based on artificial intelligence to solve problems in the most varied areas of knowledge has been increasing. Many solutions based on Machine Leaning or Deep Learning have been quite successful, and in the area of Biometeorology and Public Health it has been no different. However, one of the most complex points during the process of developing these models is the choice of input variables, either due to the limitation of access to certain data, or due to the complexity that they can add to the model. This paper will present a comparison between two Deep Learning models, used to predict respiratory diseases CIDs [J00, J01, J02, J03, J04, J05, J06, J11, J18, J20, J21, J22, 23, J30, J32, J40, J44, J45, J46, J47] in the city of

São Paulo, one of which is univariate, having as input only past values, while the other will be a multivariate model, which have as input, in addition to past values, information associated with pollutant concentration and meteorological variables. This comparison seeks to demonstrate the degree of importance of the use of meteorological data and pollutant concentration in the construction of predictive models in the area of public health.

The project involves two stages: first, data analysis and creation of a predictive model with Artificial Neural Networks (ANN) for the city of São Paulo. This model will use as input air pollution data, surface meteorological variables, social indicators from IBGE and information on hospitalizations and deaths provided by DATASUS. From there, a Recursive Neural Network (RNN) will be built so that it is possible to predict the variation of the diseases analyzed. The second stage will be the replication of the methodology in other Brazilian capitals using Transfer Learning. This technique consists of using the ANN structure initially created and training only the external layers. This will make it possible to create an RNN even for capitals with little data.

The univariate model presents satisfactory results when compared with the mean value of the series, with a Mean Absolute Error (MAE) of approximately 35. However, when performing a more detailed analysis, it is possible to observe that during the winter months, the period in which hospitalizations peak, the univariate model is unable to correctly predict the peaks. On the other hand, the multivariate model is able to predict the peaks of the winter months satisfactorily, and the periods in which there are few variations, resulting in an MAE of 15.

In the context of public health, the results of the CAIPORA project allow decision-makers to forecast the number of users of the public health system, especially during periods when the number of cases tends to increase abruptly due to weather conditions. Such forecasts will enable advance organization, so that the health system is already prepared for a possible high demand of users.

Palavras-chave: projeto caipora; deep learning; saúde humana; poluição atmosférica.