

UNIVERSIDAD DE COSTA RICA
VICERRECTORÍA DE INVESTIGACIÓN

Reporte de Informe del Proyecto

21/06/2023 - 5:43 pm

Informe de Proyecto

C0145-23 USING MACHINE LEARNING AS A PREDICTIVE TOOL FOR VECTOR-BORNE DISEASES IN COSTA RICA // USO DEL APRENDIZAJE AUTOMÁTICO COMO HERRAMIENTA PREDICTIVA PARA ENFERMEDADES VECTORIALES EN COSTA RICA

INFORME FINAL DE INVESTIGACIÓN 2023

INVESTIGADORES

IDENTIFICACIÓN	NOMBRE	GRADO	TIPO DE PARTICIPACIÓN	VIGENCIA INICIO	VIGENCIA FINAL	HORAS
184001281001	FABIO ARIEL SÁNCHEZ PEÑA	DR.	PRINCIPAL	01/08/2022	15/08/2022	S. C.
184001281001	FABIO ARIEL SÁNCHEZ PEÑA	DR.	PRINCIPAL	02/01/2023	04/03/2023	1/2 T
184001281001	FABIO ARIEL SÁNCHEZ PEÑA	DR.	PRINCIPAL	05/03/2023	31/03/2023	S. C.
184001281001	FABIO ARIEL SÁNCHEZ PEÑA	DR.	PRINCIPAL	12/12/2022	01/01/2023	S. C.
184001281001	FABIO ARIEL SÁNCHEZ PEÑA	DR.	PRINCIPAL	16/08/2022	11/12/2022	1/8 T
184001281001	FABIO ARIEL SÁNCHEZ PEÑA	DR.	PRINCIPAL	28/03/2022	31/07/2022	1/8 T
AQ344108	YURY ELENA GARCÍA PUERTA	DRA.	ASOCIADO	01/04/2022	31/03/2023	S. C.
801070320	SHU WEI CHOU CHEN	DR.	ASOCIADO	01/04/2022	31/03/2023	S. C.

107340170	HUGO HIDALGO LEÓN	DR.	ASOCIADO	01/04/2022	31/03/2023	S. C.
111090348	LUIS ALBERTO BARBOZA CHINCHILLA	DR.	ASOCIADO	01/04/2022	31/03/2023	S. C.
112120635	JUAN GABRIEL CALVO ALPÍZAR	DR.	ASOCIADO	01/04/2022	31/03/2023	S. C.
113310778	PAOLA ANDREA VÁSQUEZ BRENES	BACH.	ASOCIADO	01/04/2022	31/03/2023	S. C.

TESIARIOS

IDENTIFICACIÓN	NOMBRE	TIPO DE PARTICIPACIÓN	GÉNERO	VIGENCIA INICIO	VIGENCIA FINAL
No hay datos disponibles					

ANTECEDENTES

Introduction

Currently, and despite the increasing knowledge acquire throughout history, timely and effective management of vector-borne diseases continues to represent a challenge for public health authorities worldwide. Over the past decades, diseases such as dengue, malaria and more recently chikungunya and zika have spread rapidly in most tropical and subtropical countries around the world, where they inflict substantial health, economic and social repercussions (WHO, 2017).

The emergence and re-emergence of these diseases, as well as the complexity involved in their transmission dynamics have led to the use of a variety of statistical and mathematical models, both internationally (Chowell & Sanchez, 2006; Chowell et al., 2011; Reiner, et al., 2013) and to a lesser extent nationally (Sanchez et al., 2017; Sanchez, Barboza, & Vásquez, 2019). These models, have increased our understanding of the particular characteristics under which an outbreak or epidemic begins and evolves, in addition to serving as an ever-growing tool to guide prevention and control strategies, opening the possibilities to predict with sufficient degree of precision the beginning and possible final size of an epidemic (Hussain-Alkhateeb, et al., 2017; Racloz, et al., 2012).

In this process, the increasing availability of new technologies, such as machine learning models, have come to enter as a technique increasingly used in the modeling of infectious diseases (Jordan & Mitchell, 2015). Through the use of computational algorithms, this method allows more refined prediction models, as it learns from past trends and transmission dynamics. Algorithms under this approach, process data in sufficient quantity and without explicit external instructions (and therefore potentially biased) provided by the researcher (Bishop, 2006).

Given these characteristics, several studies have evidenced its usefulness and predictive capacity, using for this purpose, a diversity of variables (Carvajal, et al., 2018; Guo, et al., 2017; Ong, et al., 2018; Stoleran, Maia, & Kutz, 2018; Lureano-Rosario, et al., 2018). In Costa Rica, the use of these techniques is still in its early stages, however, the study conducted with dengue and climatic variables in five cantons of the country has shown promising results (Vásquez, et al., 2019).

To venture into this area, and to include a greater diversity of variables, both environmental, social and economic, this research proposal aims to unite the efforts of researchers from the University of Costa Rica and from El Estado de la Nación. This joint work will allow the integration of information available at the national level, as well as the analysis of the data that through the identification of patterns, can demonstrate its usefulness as a predictive tool, and in turn, serve as a guide for the design of strategies for vector control based on information management, which facilitates decision-making by the various public health authorities in the country.

Machine learning

Machine Learning models represent an extensive set of algorithms with a high statistical component that allow solving different tasks, for example (Hastie, Tibshirani, & Friedman, 2009)

a) Predict the value or status of a dependent variable with respect to the behavior of auxiliary or covariate variables. The association between the variables is not approached from a classical linear approach, as is done in normalized or generalized linear models, but the degree of association is highly flexible and allows capturing relationships between variables with greater complexity. These models are known as Supervised Learning models and their greatest exponents are generalized linear models, discriminant models, additive models, decision trees, neural networks, vector support machines, among others.

b) Find patterns or groupings in spaces of observations or variables. The models used for these tasks strongly use the notion of distance between observations, in order to find accumulations or clusters (Non-Supervised Learning Models). The most commonly used models, in this case are k-means, self-organizing maps, Principal Component Analysis, Multidimensional Scaling (MS), among others.

Machine learning models are designed under the premise that they must be adjusted through a learning process on a set of test or training data. The predictive capacity is corroborated through usual goodness of fit measures, such as the standardized mean square error, which are calculated on a smaller data set than the one used in the training process (test data). The selection of the training and test sets can be repeated a certain number of times so that random results are generated in the estimation results. This process is called cross-validation and also allows the measure of goodness of fit to be verified against the expected results of its empirical distribution.

Vector-borne diseases in Costa Rica

Regarding prevention, control, and management of vector-borne diseases, Costa Rica has come a long way. From the achievements made in the early and mid-twentieth century, with the implementation of multiple public health policies, the subsequent re-infestation of diseases such as dengue and malaria, in the 1990s, to the most recent emerging viruses such as Chikungunya and Zika, vector-borne diseases have accompanied Costa Rican society, and contributed in shaping the history, knowledge and activities of public health in the country.

The most notable beginnings of the fight against these diseases at the national level occurred since the establishment in 1927 of the Ministry of Health (Ministerio de Salud, 2017). This organism arises in an epidemiological context in which the environmental, economic and social conditions that characterized the country in those years, as well as the few public health activities that were implemented during the nineteenth century, led to communicable diseases such as malaria to occupy the first positions as cause of mortality (OPS, 2003). In this context, various control strategies were born, among which stand out the first anti-malaria control programs in 1937 (Ministerio de Salud, 1938), the campaign against the *Aedes aegypti* mosquito, in 1949, promoted by the Pan American Organization of Health (PAHO), which together with the beginning of vaccination led to the eradication of yellow fever in 1954 (OPS, 2003) and the declaration of the country as free of the vector mosquito in 1961 (Morice, Marín, & Ávila, 2010), as well as the subsequent organization of the Malaria Eradication Program in 1950, which, for the first time, established the mandatory legal and regulatory provisions in the reporting and treatment of cases in the country (Ministerio de Salud, 2017).

Another of the milestones in the fight against vector-borne diseases was the creation of the Costa Rican Social Security Fund (CCSS) in 1941, the emergence of social insurance in the country, and the subsequent universalization in health care, allowed a better diagnosis and treatment of diseases (OCDE, 2017), this, together with the issuance of the Constitutive Law of the Costa Rican Institute of Aqueducts and Sewers (AyA) in 1961, which came to solve the crisis of access to potable water, sanitation and disposal of excreta that was lived in those years (AyA, 2016), led to a significant decrease in the incidence and mortality of infectious and vector-borne diseases, which, by 1970, gave the first places as causes of death to chronic conditions (OPS, 2003).

However, multiple factors such as economic crises, natural disasters, as the earthquake in Limón, the urbanization processes, increased population mobilization, migratory patterns, lack of monitoring and resources for vector control and surveillance programs, the emergence of vector resistance to insecticides, as well as, changes in environmental and climatic factors, played a fundamental role so that in the early 1990s, several of these diseases and vectors, considered as controlled and many of them eradicated from the country, became relevant again, contributing also to the introduction of new viruses, such as chikungunya in 2014 and Zika in 2016.

This scenario has promoted that public health authorities, researchers and other actors in the health sector to search for new tools to better understand and address this emerging problem, as well as the search for models that allow the prediction of possible outbreaks. and epidemics of these diseases nationwide.

Dengue, Zika and Chikungunya in Costa Rica

Worldwide, *Aedes aegypti* and *Aedes albopictus* are two of the most important species of mosquitoes in which refers to disease transmission and medical relevance. These arthropods are considered the natural vectors of the dengue virus, Zika and chikungunya, with *Ae. Aegypti* as the main vector in urban areas and *Ae. Albopictus* as the second vector in importance. In general, these mosquitoes inhabit areas below 2,000 meters above sea level, although they have also been recorded at heights of around 2,400 meters above sea level (Quintero, Osorio, & Martínez, 2010).

These vectors are characterized by presenting a practically imperceptible sting, in which during each feeding period it can bite

several people, especially if it is disturbed before reaching satiety (Wilder & Schwartz, 2005). The periods in which the bites intensify are in hours of low solar intensity, usually at the beginning of the morning and at sunset. When conditions are favorable, the mosquito does not usually travel long distances from oviposition sites, and has an approximate flight range of 200 meters, which suggests that it is people instead of mosquitoes, the main form of dissemination of viruses within and between different communities (Harrington, et al., 2005).

In Costa Rica, the success in the fight against *Ae. aegypti*, in the 1950s, led to the declaration of the country as free of the vector by 1961. This result was achieved after a series of strategies and actions promoted by the Pan American Health Organization (PAHO), in coordination with the Insect Control Program of the Ministry of Health. However, the absence of dengue cases and the displacement of resources from this surveillance system towards malaria control weakened the campaign. This lack of continuity in active surveillance caused that ten years later, the country again presented positive locations for the presence of the mosquito. After that time, a new eradication campaign was established, however, by 1992 the vector was already in almost all of the national territory. By September 1993, the first cases of dengue were reported on the Pacific coast (Morice, Marín, & Ávila, 2010). Extending the case report in the months and years after most of the country (Mena, et al., 2011).

From 1993 to 2018, more than 376,000 suspected and confirmed cases of the disease have been reported in the country, with coastal region being the most affected historically. During these years, the transmission of the disease in the country, has been characterized by epidemic peaks that are cyclic, occurring every 2 to 5 years, and seasonal within the same year. Being the epidemic year with the highest number of cases 2013, with a total of 49,993 reported cases by health centers around the country, followed by 2005 with a total of 37,798 cases and 2010 with a total of 31,484 cases (Ministerio de Salud, 2019).

As of 2014, the first cases of chikungunya were reported in the country, in these 4 years, more than 9,000 suspected and confirmed cases have been reported, with a reduction in the number of patients reported in the last two years. In February 2016, the first two indigenous cases of Zika virus were reported in Nosara, Guanacaste. From that year on, the virus spread rapidly throughout national territory, causing more than 10,000 cases affecting, mainly, the provinces of Guanacaste, Puntarenas and Limón (Ministerio de Salud, 2019). A total of six newborns with microcephaly have been confirmed with the virus, however, the true number of affected could be greater, this according to the report of the Reference Center of Congenital Diseases of the country, in which it is reported that since the introduction of the Zika virus in Costa Rica, the cases of microcephaly doubled with respect to the baseline (2011-2015), in the first year of the outbreak; and exceeded the quadruple in 2017, year in which most of the children from infected mothers in 2016 were born. In all provinces, the reported cases of microcephaly increased, although the most affected were Guanacaste, Puntarenas and Limón, sites most infested by the vector and most affected by the outbreak of dengue and Zika (INCIENSA, 2018).

Malaria in Costa Rica

Malaria is a vector-borne disease caused by parasites of the genus *Plasmodium* and transmitted to humans by the bite of infected female mosquitoes of the *Anopheles* genus, *Anopheles albimanus* being the main vector in the country. There are also four different species of plasmodium that infect humans *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale* and *Plasmodium malariae*, of which, *Plasmodium vivax* is the one with the highest circulation in the country, followed by *Plasmodium falciparum*, the most dangerous species responsible for the majority of deaths caused by the disease.

Among the symptoms of malaria fever, headaches and vomiting, which usually appear 10 to 15 days after the mosquito bite, are the most common. If untreated, malaria can endanger the patient's life in a short time, as it alters the blood supply to vital organs.

Female *Anopheles* mosquitoes, are characterized by presenting nocturnal or crepuscular habits, being therefore more active at dusk and dawn. Females can fly up to 3 kilometers and bite both inside and outside the houses. Malaria vectors associate their presence with characteristics of the landscape, so maintaining natural areas prevents their propagation. The behavior and distribution of these vectors is also closely related to climatic conditions, especially with temperature and precipitation, deforestation has also been related to alterations in the dynamics of disease transmission.

In Costa Rica, thanks to the actions implemented in 1957 by the National Program for the Eradication of this disease, the spread and involvement of malaria had been reduced to a large extent in the national territory affected by the parasite. This program used as main measure the spray with the DDT insecticide. However, despite these efforts, by 1991 the disease reemerged in rural regions of the Atlantic coast and Huetar Norte (OPS, 2003). In 1991, a total of 1,142 cases are reported, however, by 1992 this figure would rise to 6,951 cases, the year in which the largest number of patients affected in the country would be presented. Since then, malaria continued to be a public health problem showing significant fluctuations. Multiple efforts led by the Ministry of Health achieved that as of 1999, the incidence of the disease will decrease, to achieve in 2013 the non-report of indigenous cases. Unfortunately, after three years, in 2016, the first indigenous cases are reported again, a situation that stopped the process of declaring the country as free of the disease. The local transmission of the disease has slowly resumed its upward trend, 4 cases in 2016, and 12 in 2017. In 2018, a total of 108 cases were reported, of which 86.1% of patients presented in the Northern Region of the country (Ministerio de Salud, 2019).

Bibliography

- AyA. (2016). Plan Estratégico Institucional 2016-2020. San José.
- Bishop, C. (2006). Pattern Recognition and Machine Learning. New York Springer.
- Carvajal, T., Viacrusis, K., Hernández, L., Ho, H. A., & Watanabe. (2018). Machine learning methods reveal the temporal pattern of dengue incidence using meteorological factors in metropolitan Manila, Philippines. *BMC Infectious Diseases*, 3-15.
- Chowell, G., & Sanchez, F. (2006). Climate-based descriptive models of Dengue fever The 2002 epidemic in Colima, Mexico. *Journal of Environmental Health*, 40-55.
- Chowell, G., Cazelles, B., Broutin, H., & Munayco, C. (2011). The influence of geographic and climate factors on the timing of dengue epidemics in Perú, 1994-2008. *BMC Infectious Diseases*, 2-14.
- Guo, P., Zhang, Q., Xia, J., Zhang, Q., Luo, G., Li, Z., et al. (2017). Developing a dengue forecast model using machine learning A case study in China. *Plos Neglected Tropical Diseases*, 1-22.
- Harrington, L., Scott, T. L., Coleman, R., Costero, A., Clark, G., Jones, J., et al. (2005). Dispersal of the dengue vector *Aedes aegypti* within and between rural communities. *The American Journal of tropical medicine and hygiene*, 209-220.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning Data mining, inference and prediction. Springer.
- Hussain-Alkhateeb, L., Kroeger, A., Oliaro, P., Rocklov, J., Sewe, M., Tejada, G., et al. (2017). Early warning and response system (EWARS) for dengue outbreaks Recent advancements towards widespread applications in critical settings. *PLOS One*, 1-14.
- INCIENSA (2018), Informe de vigilancia de Arbovirus basada en laboratorio. Available from <https://bit.ly/2klQATO>. Accessed Jan 23, 2019.
- Jordan, M., & Mitchell, T. (2015). Machine learning Trends, perspectives, and prospects. *Science*, 255-260.
- Lureano-Rosario, A., Duncan, A., Méndez, P., García, J., S., G., Farfan, J., et al. (2018). Application of Artificial Neural Networks for Dengue fever outbreak predictions in the northwest coast of Yucatan, Mexico and San Juan, Puerto Rico. *Tropical Medicine and Infectious Disease*, 1-16.
- Mena, N., Troyo, A., Bonilla-Carrión, R., & Calderón-Arguedas, Ó. (2011). Factores asociados con la incidencia de dengue en Costa Rica. *Revista Panamericana de Salud Pública*, 234-242.
- Ministerio de Salud. (1938). Memoria de la Secretaría de Salubridad Pública y Protección Social del año Administrativo 1937. San José Ministerio de Salud.
- Ministerio de Salud. (2017). 90 años trabajando por la salud y el bienestar de Costa Rica. San José Ministerio de Salud.
- Ministerio de Salud (2019) Ministerio de Salud Análisis de Situación de Salud, Available from <https://www.ministeriodesalud.go.cr/index.php/vigilancia-de-la-salud/analisis-de-situacion-de-salud>. Accessed June 15, 2019.
- Morice, A., Marín, R., & Ávila, M. (2010). El dengue en Costa Rica evolución histórica, situación actual y desafíos. *La Salud Pública en Costa Rica. Estado actual, retos y perspectivas*, 197-217
- OCDE. (2017). Estudios de la OCDE sobre los sistemas de salud Costa Rica. San José Ministerio de Salud.
- Ong, J., Liu, X., Rajarethinam, J., Kok, S., Liang, S., Tang, C., et al. (2018). Mapping dengue risk in Singapore using Random Forest. *Plos Neglected Tropical Diseases*, 1-12.
- OPS. (2003). 100 años de salud en Costa Rica. San José Organización Panamericana de la Salud.
- Quintero, D., Osorio, J., & Martínez, M. (2010). Competencia vectorial consideraciones entomológicas y su influencia sobre la epidemiología del Dengue. *IATERIA*, 146-156.
- Racloz, V., Ramsey, R., Tong, S., & Hu, W. (2012). Surveillance of Dengue Fever Virus A Review of Epidemiological Models and Early Warning Systems. *PLOS Neglected Tropical Diseases*, 1-9.
- Reiner, J. R., Perkins, T., C.M., B., Niu, T., Chaves, L., Ellis, A., et al. (2013). A systematic review of mathematical models of

mosquito-borne pathogen transmission 1970-2010. *Journal of the Royal Society*, 1-13.

Sanchez, F., Barboza, L., Burton, D., & Cintrón-Arias, A. (2017). Comparative analysis of dengue versus chikungunya outbreaks in Costa Rica. *Riserche di Matematica*, 1-14

Sanchez, F., Barboza, L., & Vasquez, P. (2019). Parameter estimates of the 2016-2017 Zika outbreak in Costa Rica An Approximate Bayesian Computation (ABC) approach. *Mathematical Biosciences and Engineering*, 2738-2755.

Stolerman, L., Maia, P., & Kutz, N. (2018). Forecasting dengue fever in Brazil An assessment of climate conditions. *Plos One*, 1-16.

Vásquez, P., Loría, A., Sanchez, F., & Barboza, L. (2019). Climate-Driven Statistical Models as effective predictors of local dengue incidence in Costa Rica A Generalized Additive Model and Random Forest approach. *Special Issue Mathematical Modeling A New Frontier. Revista de Matemática Teoría y Aplicaciones*, 1-21.

World Health Organization. (2017). *Global Vector Control Response 2017-2030*. Geneva World Health Organization.

Wilder, S., & Schwartz, E. (2005). Dengue in travelers. *New England Journal of Medicine*, 924-932.

Anotaciones:

Información no suministrada

JUSTIFICACIÓN

Recently, the increase in data availability in different areas and disciplines has changed the vision of how to describe and predict different events. In this sense, using data that, a priori, has not been related (such as incidence of diseases, demographic, climatic, social data, among others) can give a better understanding of the process of spreading infectious diseases. Likewise, the development of computational tools to process large amounts of data has allowed the development of machine learning algorithms, a trend in recent years, which allows debugging, processing and learning of the data themselves. In this way, the research project aims to develop predictive tools through the use of machine learning algorithms, based on clinical, epidemiological, environmental and social factors, which allow improving the implementation of prevention strategies and resource allocation, as well as contributing with the welfare and health of the Costa Rican population. At the national level, such studies do not exist, so they come to develop an area of relevant knowledge, not only with scientific implications but with immediate contributions to the population.

Anotaciones:

Información no suministrada

METODOLOGÍA

The combination of predictive models with epidemiological applications requires the use of data using the number of observed cases of a given disease with at least weekly or monthly frequencies, spatially distributed in administrative units such as cantons or districts. On the side of the explanatory variables of the predictive models, climate variables such as temperature, precipitation, relative humidity, Southern Oscillation Index (ENSO) can be used (see Vásquez, Loría, Sanchez, & Barboza, 2019 as an example). Biological variables such as the vegetation index (EVI and NDVI) can give information about the nature of the hatcheries in a certain area can also be considered. Likewise, social variables such as poverty, education and health indices can be incorporated into the analysis, with the main drawback that the frequency of which they are calculated in most cases differ from the weekly frequency that observed cases are collected.

The collection of epidemiological information will be carried out with data from the Ministry of Health and the Costa Rican Social Security Fund (CCSS), climatological information, such as, temperature, precipitation, humidity, as well as indexes of the El Niño Southern Oscillation (ENSO) will be obtained from the National Oceanic and Atmospheric Administration (NOAA) database. Demographic, social, economic and land use data will be provided by El Programa del Estado de la Nación, which will include

Population density the number of inhabitants per square kilometer (km²) of the district.

Percentage of urban population It is the percentage relationship between the population residing in the urban area, with respect to the total population.

Percentage of children under 12 years old this is the percentage ratio between the population under 12 years old and the total population.

Percentage of older adults this is the percentage ratio between the population over 65 years of age and the total population.

Percentage of population with Unsatisfied Basic Needs (NBI) poverty proxy variable. Shows the percentage of households that have at least one Unsatisfied Basic Need (NBI). The method is to identify households that have some basic deficiency in four dimensions or macro needs access to decent housing, access to a healthy life, access to knowledge and access to other goods and services.

Population with complete secondary school or more it is the percentage relationship between the population with at least one year of secondary school level or more, with respect to the population 18 years of age and over.

Distribution of economic sectors (primary, secondary, tertiary) shows the percentage distribution of the employed population aged 15 years and over according to the branch of activity to which it belongs.

Extension in hectares of agricultural farms (based on the agricultural census) It is the extension of land, totally or partially dedicated to agricultural production for sale or self- consumption, managed by a home, society, company, public institution or another, whose work can be directed or executed directly by one person or with the help of others.

For environmental data, the monitoring study for modeling is proposed based on the combination of multispectral images of the Lansat and MODIS sensors that allow a spatial and temporal vision of the vegetation of the different municipalities the country. With multispectral images, first, it is intended to perform time series analysis where the phenological changes at the local level are considered. It is also intended that this information be integrated into the modeling of vector diseases. Time series information will be done using online processing with "Earth Engine" (Gorelick, et al., 2017). With the "Earth Engine" tool, a series of "routines" generated by the Ecosystem Remote Sensing Laboratory (LabTEC), of INSEFOR at the municipality level will be processed and adapted.

In general terms the methodological process can be schematized as follows

- 1. Collection of information and processing of data tables** As the sources of information come from different institutions it is necessary to join all the information and match the observations on the same temporal and spatial frequency. For example, vegetation index data is calculated by satellites and its spatial frequency is high, but its temporal frequency is biweekly, so it is necessary to interpolate temporarily, and to select or accumulate data geographically.
- 2. Selection of covariates and lags** The predictive capacity of the models depends on the type of structure and degree of association of the dependent variable with respect to covariates. Likewise, measures such as cross correlations of the covariates with respect to them allow inferring and selecting lags that are potentially significant.
- 3. Model selection** Once it has been determined which variables turn out to be potentially predictors of the dependent variable, some type of model that manages to describe the association between the variables must be selected. For this, a very wide range of possibilities are available depending on the nature of the dependent variable, for example if what is sought is to explain categorical variables, classification models or decision tree techniques and all their variants can be used. Counting variables and continuous dependent variables can be modeled through techniques such as Random Forests, generalized additive models, neural networks. If the variables present some degree of hierarchy, Deep Learning type models can be considered more complex associations between the same covariates.
- 4. Calculation of predictive capacity and cross-validation measures** In the comparison process, goodness of fit analysis and predictive capacity measurement of the models, bootstrap-type samples can be used to define cross-validation schemes. This allows to measure the variability of the classic error indicators, such as the mean square error, and thus evaluate the predictive capacity of the models without the need to assume additional hypotheses about the data.

Anotaciones:
Información no suministrada

POBLACIÓN DE ESTUDIO

Comunidad matemática local e internacional, profesionales de salud pública y población costarricense.

Anotaciones:
Información no suministrada

CONVENIOS EXTERNOS

NÚMERO DE	ENTIDAD	TIPO DE	NÚMERO DE	MONTO	ENTE
-----------	---------	---------	-----------	-------	------

CONVENIO	COLABORADORA	COLABORACIÓN	CUENTA		ADMINISTRADOR
No hay datos disponibles					

OBJETIVO GENERAL

CREATE PREDICTIVE TOOLS TO DEVELOP MITIGATION STRATEGIES AND OPTIMAL RESOURCE ALLOCATION IN THE FIGHT AGAINST VECTOR-BORNE DISEASES IN COSTA RICA BY COMBINING TECHNIQUES IN MATHEMATICAL MODELING AND MACHINE LEARNING.

Anotaciones del objetivo general:

OBJETIVOS ESPECÍFICOS

<p>Objetivo: Identify the main climate, socioeconomic, demographic and epidemiological variables related to the spread of the prevailing vector-borne diseases in Costa Rica.</p> <p>Meta 1: Create a database with the variables associated with the transmission of vector-borne pathogens in Costa Rica.</p> <p>Actividades realizadas y resultados</p>
<p>Objetivo: Fit mathematical models with machine learning algorithms to predict prevalent vector-borne disease outbreaks in Costa Rica.</p> <p>Meta 1: Early detection of vector-borne disease outbreaks.</p> <p>Actividades realizadas y resultados</p>
<p>Objetivo: Proposal of public health strategies based on model prediction.</p> <p>Meta 1: Implementation of early detection model for the development of public health strategies.</p> <p>Actividades realizadas y resultados</p>

Nuevos objetivos propuestos

CRONOGRAMA

NÚMERO DE TAREA	NOMBRE	FECHA DE INICIO	FECHA FINAL
1	Reuniones con funcionarios del Ministerio de Salud	01/04/2022	30/06/2022
2	Formulación de estrategias de salud pública	01/07/2022	31/10/2022
3	Escritura de artículos científicos	01/11/2022	28/02/2023
4	Informe final	01/03/2023	31/03/2023

Anexos de Renovación/Ampliación

Información no suministrada

Presupuesto: No

Ampliación: No

Documentos no disponibles.

TRANSFERENCIA DE LOS RESULTADOS DE LA INVESTIGACIÓN

Artículo en revista

<https://doi.org/10.1371/journal.pntd.0011047>

Artículo en revista

<https://doi.org/10.26633/RPSP.2022.113>

Artículo en revista

<https://doi.org/10.3934/mbe.2023024>

Artículo en revista

<https://doi.org/10.1016/j.epidem.2022.100577>

Artículo en revista

<https://doi.org/10.1038/s41598-022-06236-1>

Artículo en revista

<https://doi.org/10.3390/life11121336>

Artículo en revista

<https://doi.org/10.3390/epidemiologia2030022>

PUBLICACIONES



TÍTULO	TIPO	AUTOR	FECHA
No hay datos disponibles			

DIFICULTADES ENCONTRADAS

Anotaciones:

EJECUCIÓN PRESUPUESTARIA

EQ	PARTIDA	ASIGNADO	AUMENTOS Y AMPLIACIONES	DISMINUCIONES Y DEDUCCIONES	EGRESOS	DISPONIBLE
0	No hay datos	0.00	0.00	0.00	0.00	0.00

EJECUCIÓN HISTÓRICA PRESUPUESTARIA

PARTIDA	ASIGNADO	AMPLIACIONES	DISMINUCIONES	EGRESOS

No hay datos disponibles

¿Contó con financiamiento externo?

No

¿Contó con exoneración del fondo de desarrollo institucional (FDI)?

No

Indique los beneficios obtenidos del plan de inversión

CONCLUSIONES

Anotaciones:

A pesar de las dificultades encontradas por el tiempo de pandemia se lograron los objetivos del proyecto. Además, se logró dar apoyo a las autoridades del país durante el periodo difícil de pandemia siendo el equipo de investigación oficial que le proveía insumos de los modelos matemáticos al Ministerio de Salud y autoridades. De igual forma, se publicaron varios artículos en revistas internacionales de alto prestigio.

REFERENCIAS BIBLIOGRÁFICAS



ANEXOS Y APÉNDICES