

APLICACIÓN DE MODELOS DE MACHINE LEARNING CON VARIABLES EPIDEMIOLÓGICAS Y METEOROLÓGICAS PARA LA PREDICCIÓN DE CASOS DE LA ENFERMEDAD DEL DENGUE

Gonzalo M. Díaz^{1,4}, Anahí Chiavetta¹, Julieta Siches², Lorena Ferreira¹, Federico Santoro², Dalila Rueda², Silvina Moisés², Cecilia Cappa³, Cecilia González Lebrero², Lucía Maffey², Gladys Fattore², M. Eugenia Utgés³, Mariana Manteca Acosta³
gdiaz@smn.gob.ar

¹Servicio Meteorológico Nacional (SMN)

²Ministerio de Salud de la Nación (MSAL)

³Centro Nacional de Diagnóstico e Investigación en Endemoepidemias (CeNDIE). ANLIS – Malbrán

⁴Departamento de Ciencias de la Atmósfera y los Océanos. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires (DCAO-FCEN-UBA)

Palabras clave: dengue, machine learning, predicción, MSAL, CeNDIE

1. INTRODUCCIÓN

Una de las misiones fundamentales del Servicio Meteorológico Nacional (SMN) consiste en brindar información a las autoridades y la sociedad que permita desarrollar herramientas para prevenir y/o combatir adversidades en las que se encuentren involucrados factores meteorológicos. En el marco de este objetivo, dicho Organismo ha sido convocado por el Ministerio de Salud de la Nación y el CeNDIE – ANLIS Malbrán para llevar adelante un plan de predicción de casos de Dengue; una enfermedad viral que se encuentra en aumento en los últimos años y que puede generar desde síntomas gripales hasta cuadros potencialmente mortales. A partir de mesas de trabajo realizadas entre los tres organismos se analizaron distintos enfoques y se pusieron en común distintos conocimientos en el ámbito meteorológico y salud para una óptima implementación de modelos de predicción de casos de dengue, siendo las variables meteorológicas los predictores más relevantes para explicar el comportamiento del mosquito portador del virus (*Aedes aegypti*). Este trabajo consiste en la aplicación de modelos simples de Machine Learning para la detección de casos de dengue, incorporando tanto datos meteorológicos como epidemiológicos y de vectores, impulsando la sinergia entre las instituciones.

2. DATOS Y METODOLOGÍA

Los datos utilizados para este trabajo fueron datos meteorológicos (precipitación, temperatura mínima, humedad relativa), de balance hídrico (almacenaje de agua en el suelo, evapotranspiración real y potencial) obtenidos del modelo operativo BHOA del SMN (Fernández Long y otros, 2012) y de epidemiología (brindados por el Ministerio de Salud de Nación). Principalmente, los datos relacionados a la meteorología y el BHOA se encontraban en escala diaria y fueron procesados a escala de semana epidemiológica, escala en la cual se registran los casos positivos de dengue. Asimismo, otros parámetros vinculados a la detección de larvas de mosquito por unidad de vivienda (vectores) fueron utilizados como variable predictora de los modelos. Se utilizaron modelos simples de Machine Learning para la detección de casos de dengue: Random Forest (RF), Support Vector Machine (SVM) y Regresión Lineal Múltiple (RLM). Con el fin de lograr modelos de predicción, las variables

predictoras fueron desfasadas dos semanas con la variable respuesta (casos positivos autóctonos de dengue). Los modelos fueron forzados con datos meteorológicos, del BHOA y de vectores, para un período de entrenamiento (Figura 1).

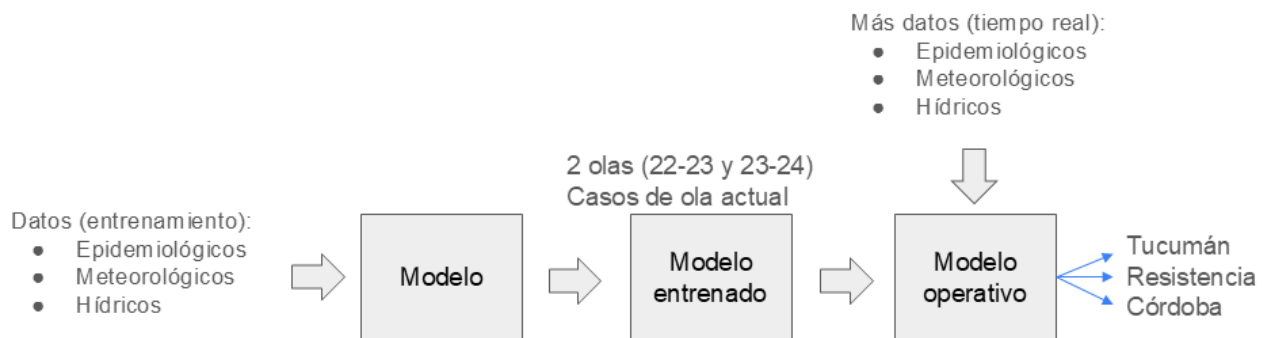


Figura 1: Esquema representativo de la forma de trabajo con los modelos.

Se utilizaron datos epidemiológicos, meteorológicos e hídricos (estos últimos en referencia al BHOA) para forzar los modelos y entrenarlos para la ola o temporada 22-23 (en este entrenamiento, una de las variables predictoras a utilizar también fue la de los casos totales, incluyendo autóctonos e importados). Se realizaron varias pruebas de combinación de parámetros y según la ciudad analizada (San Miguel de Tucumán, Resistencia o Córdoba) se alcanzaron distintos niveles de ajuste. Finalmente, el modelo entrenado se encuentra ejecutándose de forma operativamente experimental en el SMN para ir analizando su performance semana a semana (Figura 1).

3. DISCUSIÓN

La mejor combinación de variables predictoras varía por modelo y según la ciudad analizada. Para cada modelo de RF, SVM y RLM se eligieron dos combinaciones distintas de variables que fueron las mejores obtenidas en función a varias pruebas de combinaciones. La Figura 2 en su panel principal muestra para Córdoba la predicción de cada grupo de modelos y se indica con línea negra la evolución real de casos de dengue para esas semanas epidemiológicas. Para la predicción de la semana 10 y 11 de la temporada 24-25 los mejores modelos que representaban la evolución de casos de dengue eran los asociados a RF (barras celestes y azules en Figura 2). Los modelos RLM también muestran un aumento de casos para esas dos semanas con la salvedad de que el valor absoluto se observa desplazado hacia casos superiores (barras naranjas y rojas en Figura 2). Por último, los modelos asociados a SVM, para esas dos semanas, muestran un pequeño aumento o un decaimiento, manteniendo los casos casi constantes entre semana y semana (barras verdes y violetas en Figura 2). Vale destacar que los modelos deben continuar su mejora y optimización ya que en varios momentos presentan casos de falsas alarmas, como se observa entre la semana 50/2024 y 02/2025, en donde RF registra casos y en las observaciones no se detectan (Figura 2).

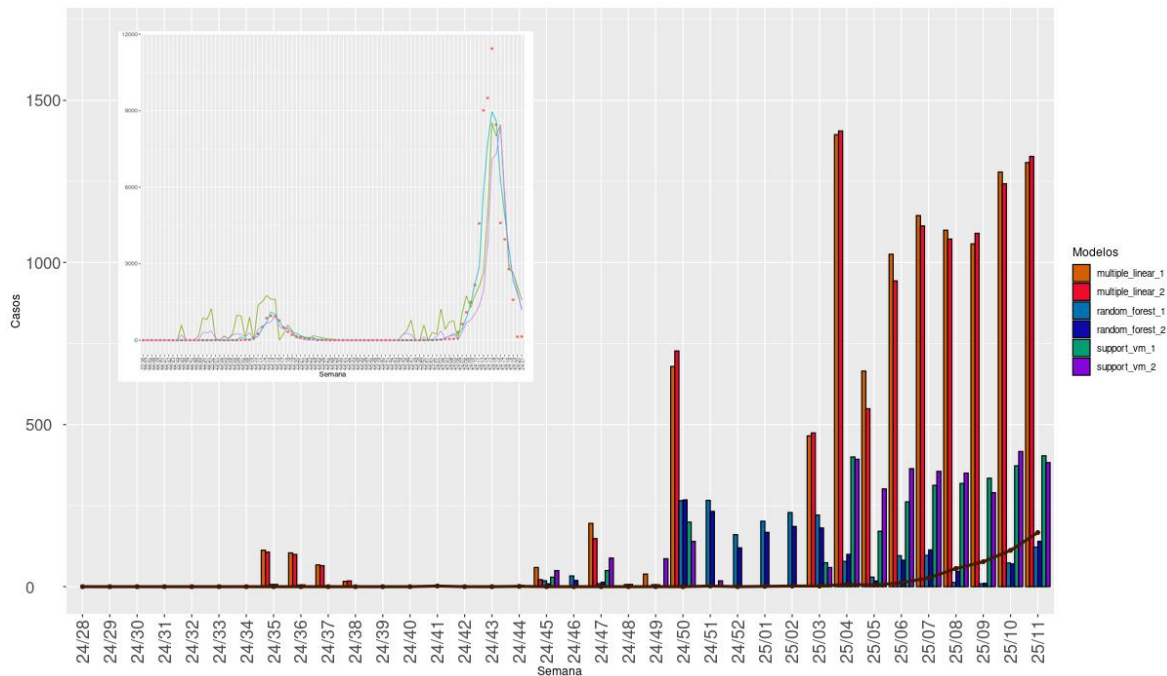


Figura 2: Panel superior, período de entrenamiento para los modelos RF (línea celeste), SVM (línea violeta) y RLM (línea verde). Panel principal, período de la temporada actual para los dos tipos de: modelo RF (barras celestes y azules), modelo SVM (barras verdes y violetas) y modelo RLM (barras naranjas y rojas). En línea negra la cantidad real de casos autóctonos de dengue. Ambas figuras representan el caso de Córdoba.

Los modelos fueron entrenados con las dos temporadas anteriores, siendo ambas temporadas muy distintas en comportamiento (Figura 2 panel superior izquierdo). Considerar estas dos temporadas como entrenamiento, tenía como fin que los modelos pudieran aprender situaciones tan distintas, y así obtener una mejor performance para el período actual. Sin embargo, vale destacar que la temporada actual, también tuvo un comportamiento muy diferente de las anteriores y esto dificultó que los modelos puedan detectar comportamientos específicos en la presente temporada.

Finalmente, la implementación de modelos de Machine Learning tiene como objetivo contribuir al fortalecimiento del sistema de salud mediante el desarrollo de herramientas aplicables a distintas proyecciones de riesgo o favorabilidad de brotes de dengue, con un nivel de desagregación jurisdiccional a escala nacional. Se espera que los resultados obtenidos permitan la generación de una capa SIG (Sistema de Información Geográfica), integrable al análisis espacial y gestión del riesgo, brindando soporte a la toma de decisiones sanitarias informadas.

4. AGRADECIMIENTOS

Se agradece a la Bach. Eugenia Bontempi del Servicio Meteorológico Nacional por el procesamiento realizado de los datos del modelo BHOA operativo del SMN para su aplicación a los modelos de Machine Learning.

5. REFERENCIAS

Fernández Long M.E., Spescha L., Barnatán I. y Murphy G., 2012: Modelo de balance hidrológico operativo para el agro (BHOA). *Agronomía y Ambiente*, vol. 32, num. 1-2.