

# Predicción del picudo del chile en Sinaloa con XGBoost - un modelo de Aprendizaje Automático

Nancy, Villegas-Jiménez<sup>1</sup> ; Martha E. Ramírez-Guzmán<sup>2\*</sup> ; E. Ibarra-Zapata<sup>3</sup> 

<sup>1</sup> Organismo Internacional Regional de sanidad Agropecuaria OIRSA, El Salvador.

<sup>2</sup> Colegio de Postgraduados, Campus Montecillo. Km 36.6, Carretera México-Texcoco. Montecillo, Texcoco, Estado de México. CP: 56264.

<sup>3</sup> Facultad de Agronomía y Veterinaria, Universidad Autónoma de San Luis Potosí, Carretera San Luis - Matuhuala Km. 14.5, Ejido Palma de la Cruz, 78321. Soledad de Graciano Sánchez, S.L.P

\* Autor para correspondencia: martharg@colpos.mx

## Problema

El picudo del chile (*Anthonomus eugenii* (Cano, 1894) [Coleoptera: Curculionidae, nativa de Centroamérica, su hospedero principal es el chile (*Capsicum annuum*). Es una plaga de importancia económica en la producción de chile, pimientos y en general diferentes tipos de ajíes, ya que el daño causado durante la floración y formación del fruto puede reducir significativamente el rendimiento de los cultivos, además de que los frutos invadidos pueden presentar deformaciones, manchas y reducción del tamaño, lo que implica pérdidas económicas significativas para los productores y rechazo en los mercados de exportación. El mecanismo de acción consiste en que la hembra deposita sus huevos en los ovarios de las flores y en la base de los frutos en formación, y cuando las larvas eclosionan, se alimentan del tejido interno ocasionando deformaciones y pudriciones secundarias. La hembra del picudo adulto puede transmitir enfermedades, entre ellas, bacteriosis y virosis principalmente, lo cual ha representado un gran problema económico para los productores de Sinaloa, México en los años 2022 al 2024.

## Solución Planteada

El objetivo de esta investigación fue modelar y predecir el comportamiento diario de la abundancia del picudo del chile, mediante un modelo de aprendizaje automático (Machine Learning en inglés), con el propósito de identificar preventivamente, el período óptimo de aplicación del tratamiento correspondiente para su erradicación o mitigación.

**Cómo citar:** Villegas-Jiménez, N., Ramírez-Guzmán, M. E., & Ibarra-Zapata, E. Predicción del picudo del chile en Sinaloa con XGBoost - un modelo de Aprendizaje Automático. *Agro-Divulgación*, 5(3). <https://doi.org/10.54767/ad.v5i3.488>

**Editores académicos:** Dra. Ma. de Lourdes C. Arévalo Galarza y Dr. Jorge Cadena Iñiguez.

Publicado en línea: Diciembre 2025.

*Agro-Divulgación*, 5(3). Mayo-Junio. 2025. pp: 55-60.

Esta obra está bajo una licencia de Creative Commons Attribution-Non-Commercial 4.0 International



El Aprendizaje Automático, es una técnica que permite a los procesadores computacionales aprender de los datos, es decir, en lugar de ser programadas explícitamente para realizar una tarea, las computadoras usan algoritmos para identificar patrones y hacer predicciones (regresión) o decisiones basadas en esos datos (clasificación). En particular, el Boosting de Gradiente Óptimo (XGBoost: Extreme Gradient Boosting, por sus siglas en inglés), es un algoritmo de aprendizaje automático popular y potente para clasificación y regresión. El Boosting de Gradiente Óptimo (BGO) es conocido por su alta precisión en una variedad de problemas, puede manejar datos mixtos (numéricos y categóricos) y admite varios tipos de pérdida (funciones de pérdida). Estas funciones de pérdida se utilizan para minimizar el error del modelo. Para este caso se utiliza el error cuadrático medio (MSE, Mean Squared Error), el cual se calcula como:

$$MSE = (1/n) * \sum (y_{observado} - y_{predicción})^2$$

donde:  $y_{observado}$  son los valores reales de la variable objetivo y  $y_{predicción}$  son los valores predichos por el modelo y  $n$  es el número de observaciones.

Para el análisis de series de tiempo, cuando no se cuenta con variables potenciales predictoras, una estrategia para identificar patrones en la serie es considerar variables de tiempo individuales como día, mes, año, etc.

Estadísticas descriptivas de la incidencia diaria del Picudo del Chile del 10 de agosto de 2022 al 22 de mayo de 2024 (Cuadro 1), muestran que el 80% del total, se presenta en los municipios de Mazatlán, Concordia, Elota, Rosario, Escuinapa y Navolato, ubicados en

**Cuadro 1.** Estadísticas descriptivas de la incidencia del Picudo del Chile del 10 de agosto de 2022 al 22 de mayo de 2024.

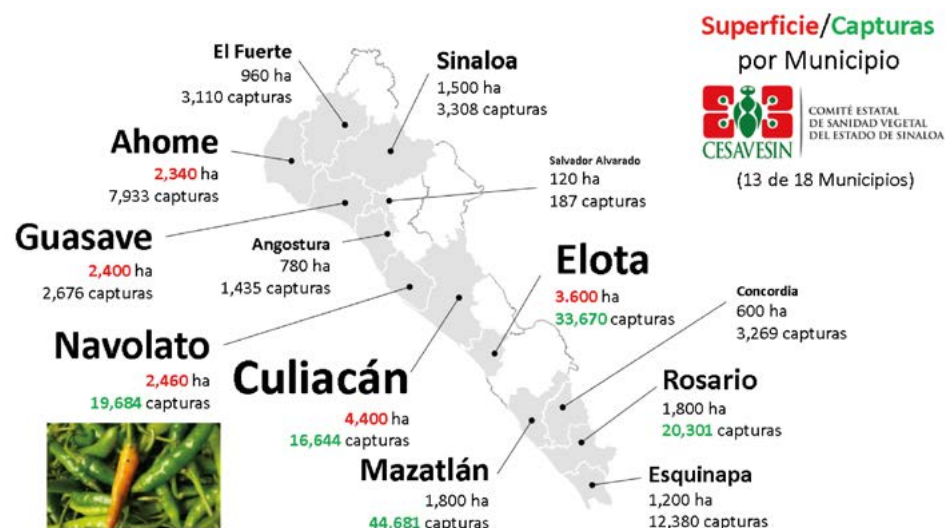
	Municipio	Mean	SD	Median	min	max	Acumulado
1	Mazatlán	98.76	161.66	24	0	1,914	22.1%
2	Concordia	71.59	185.80	8	0	2,164	27.7%
3	Elota	59.59	125.07	12	0	2,343	57.0%
4	Rosario	42.26	70.99	9	0	480	67.9%
5	Escuinapa	38.28	53.81	8	0	325	74.2%
6	Navolato	27.91	69.25	4	0	1,066	82.0%
7	Culiacán	20.83	52.88	3	0	1,015	91.0%
8	Guasave	14.65	32.71	2	0	156	95.3%
9	Ahome	11.28	25.30	5	0	355	97.9%
10	Mocorito	9.59	37.12	1	0	338	98.0%
11	El Fuerte	9.04	20.66	3	0	372	98.9%
12	Sinaloa	7.60	10.98	5	0	175	99.8%
13	Salvador Alvarado	5.94	8.01	2	0	29	99.8%
14	Angostura	5.69	11.66	2	0	79	100.0%

la costa de Sinaloa y alrededor de Mazatlán, cuyo puerto es caracterizado por atender un gran volumen de operaciones de carga, pesca y turismo.

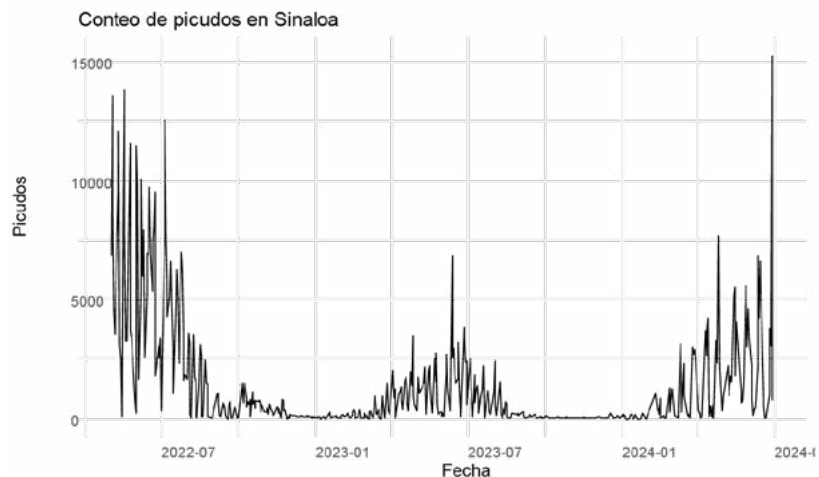
Por otro lado, los municipios muestreados de Sinaloa, que tuvieron menor incidencia del picudo fueron: Salvador, Alvarado y Sinaloa, los cuales se encuentran al norte del estado y no tienen costa (Figura 1).

Los conteos diarios proporcionados por el Comité Estatal de Sanidad Vegetal de Sinaloa (CESAVESIN), fueron registrados de lunes a viernes. La presencia del insecto es estacional, en donde la mayor abundancia se presentó en primavera y verano (Figura 2).

El modelo de Boosting de Gradiente Óptimo aplicado consideró el 80% de los datos como datos de entrenamiento (10 de agosto de 2022 a 29 de diciembre de 2023) y el



**Figura 1.** Capturas registradas mediante sistema de trapeo dirigido a machos del picudo del chile (*Anthonomus eugenii* (Cano, 1894)) entre el 10 de agosto de 2022 al 22 de mayo de 2024, por municipio en Sinaloa, México.



**Figura 2.** Abundancia del Picudo del Chile en el estado de Sinaloa, México del 10 de agosto de 2022 al 22 de mayo de 2024.

resto de prueba (2 de enero de 2024 al 28 de junio de 2024) (Figura 3). Los valores de los hiperparámetros (Cuadro 2) que modelaron mejor al comportamiento de la abundancia del Picudo del Chile, tuvieron los propósitos de evitar el sobreajuste, ahorrar tiempo de procesamiento computacional y mejorar la generalización del modelo.

El modelo ajustado, con los hiperparámetros seleccionado (Cuadro 2), producen predicciones tanto de los datos de entrenamiento, como los de prueba (Figura 4).

Las variables que mejor explicaron al modelo fueron los primeros 4 días, así como un comportamiento mensual (cada 25 días), de acuerdo con la Gráfica de Importancia de Variables (Figura 5). Las variables de menor importancia fueron el día juliano, el día de la semana y el mes. Con el modelo ajustado, se produjeron pronósticos para el horizonte de tiempo del 29 de junio de 2024 al 25 de diciembre de 2024 (Figura 6).

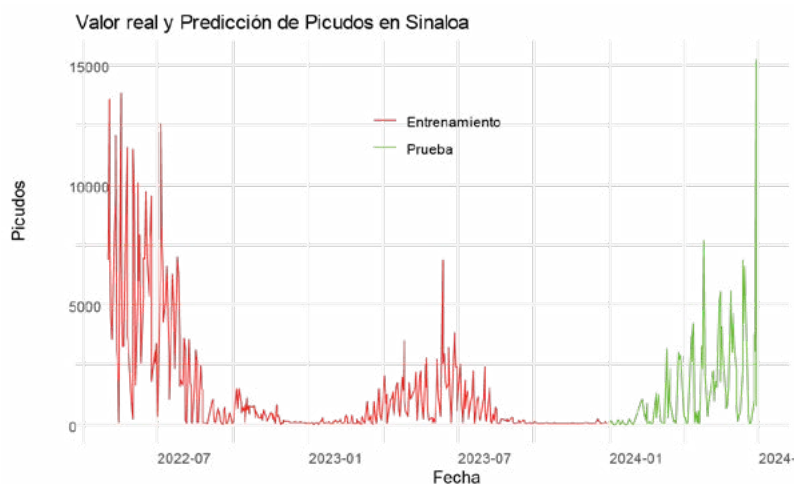
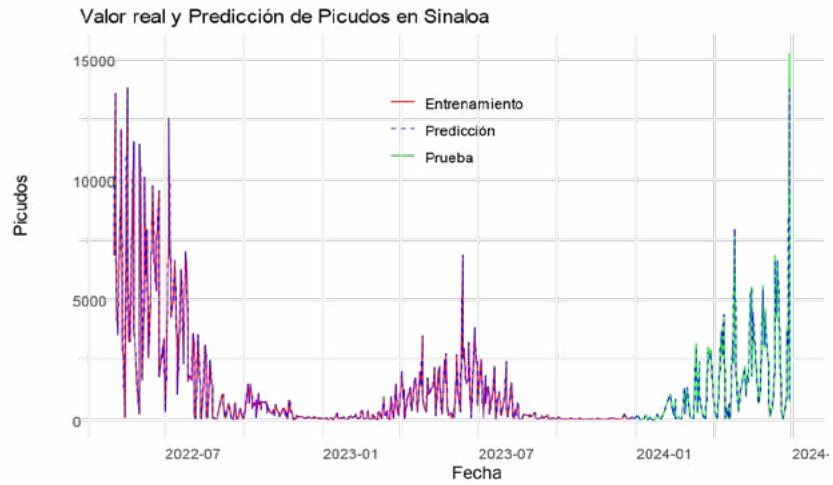


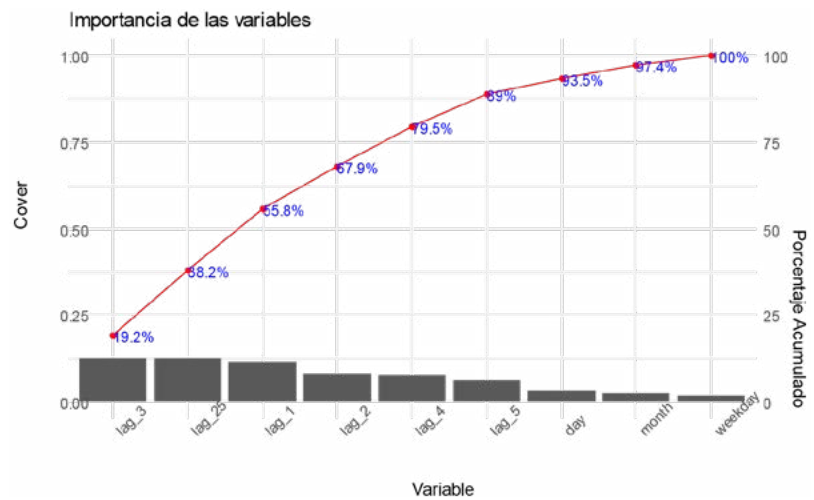
Figura 3. Datos de entrenamiento y de prueba para ajustar un modelo de Boosting de Gradiente Óptimo.

Cuadro 2. Mejores valores de los hiperparámetros del modelo (XGBoost).

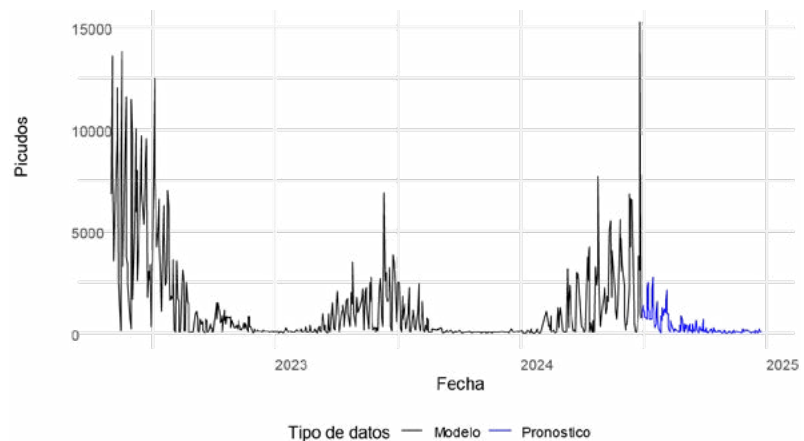
Concepto	Valor (Rango posible)	Características del modelo de entrenamiento de cada árbol para evitar sobre ajuste
<b>Métrica de evaluación</b>		
Objetivo	-----	Minimizar el error cuadrado medio
Métrica	-----	Raíz del error cuadrado medio
Máxima profundidad	5	Relaciones entre características hasta 5 niveles
<b>Arquitectura del árbol</b>		
Gamma	0.5 (0-0.5)	Controla el umbral para expandir un nodo
Alpha	1 (0-10)	Penaliza variables con coeficientes grandes con L1: Lasso
Lambda	1 (0-10)	Penaliza variables con coeficientes grandes con L1: Ridge
<b>Muestreo y Evaluación</b>		
Submuestra	0.8	El modelo utiliza el 80% de los datos para entrenamiento
Columnas por árbol	0.8	El modelo utiliza el 80% de las variables para entrenamiento
Paro temprano	5	Número máximo de rondas consecutivas para el aprendizaje



**Figura 4.** Datos de entrenamiento, prueba y pronóstico del modelo de Boosting de Gradiente Óptimo.



**Figura 5.** Gráfica de la importancia de variables del modelo de Boosting de Gradiente Óptimo.



**Figura 6.** Pronóstico de la abundancia del picudo del 29 de junio de 2024 al 25 de diciembre de 2024 con el modelo de Boosting de Gradiente Óptimo, ajustado del 2 de mayo de 2022 al 29 de diciembre de 2023.

De acuerdo con los resultados del modelo de aprendizaje automático de Boosting de Gradiente Óptimo, la mayor incidencia apareció en los meses de primavera y verano y sobre todo en las comunidades con costa y más cercanas a Nayarit.

### Innovación, Impactos e Indicadores

Nivel de Innovación	Descripción	Transferido	Impacto		Indicador General de Políticas Públicas	Indicadores Específicos	Subindicador
			Sector	Ámbito			
Incremental	Busca mejorar los sistemas que ya existen haciéndolos mejores, más rápidos, más baratos, etc.	Asociaciones de Productores Productores independientes Comunidades Agrarias	Primario: Agricultura, Ganadería, Pesca, Explotación forestal, Minería	Social Económico Ambiental Conocimiento	Ciencia y Tecnología Económico Educación Responsabilidad Ambiental	Competitividad Recursos Humanos	Número de tesis Número de egresados (Lic. M.C., D.C.) Número de publicaciones Número de familias beneficiadas Transferencias tecnológicas Desarrollo de productos y servicios para la sociedad Exportación incremento (%)
Innovación sostenible	Desarrollo de productos y procesos que contribuyen al desarrollo sostenible		Procesos de Investigación, Desarrollo e Innovación (I+D+i)				
Innovación frugal	Hacer más con menos. Idear estrategias de bajo costo para sortear las complejidades institucionales o limitaciones de recursos, conseguir innovar, desarrollar y entregar productos y servicios a los usuarios de bajos ingresos con poco poder adquisitivo						

