

## Carrera de Ingeniería en Tecnologías de la Información

Artículo Científico para la obtención del título de  
Ingeniería en Tecnologías de la Información

**AUTOR:** Arlette Stefanny Portilla Cartuche

**TUTOR:** Mgs. Ing. Milton Ricardo Palacios Morocho

Implementación de Modelos Predictivos para la Gestión del  
Proceso de Polinización y Germinación de Semillas en la  
Empresa Ecuagenera

## CERTIFICACIÓN DE AUTORÍA

Yo, Arlette Stefanny Portilla Cartuche declaro bajo juramento que el trabajo aquí descrito es de mi autoría; que no ha sido presentado anteriormente para ningún grado o calificación profesional y que se ha consultado la bibliografía detallada.

Cedo mis derechos de propiedad intelectual a la Universidad Internacional del Ecuador, para que sea publicado y divulgado en internet, según lo establecido en la Ley de Propiedad Intelectual, Reglamento y Leyes.



---

FIRMA AUTOR

## **APROBACIÓN DEL TUTOR**

Yo, Milton Ricardo Palacios Morocho certifico que conozco a la autora del presente trabajo de titulación “Implementación de Modelos Predictivos para la Gestión del Proceso de Polinización y Germinación de Semillas en la Empresa Ecuagenera”, Arlette Stefanny Portilla Cartuche siendo la responsable exclusiva tanto de su originalidad y autenticidad, como de su contenido.



.....

Mgs. Ing. Milton Ricardo Palacios Morocho  
**DIRECTOR DEL TRABAJO DE TITULACIÓN**

# *Predictive Models for the Management of the Pollination and Germination Process of Orchids Seeds*

Arlette Stefanny Portilla Cartuche<sup>1</sup>, Milton Ricardo Palacios Morocho<sup>1</sup>, Conde-Zhingre Lorena Elizabeth<sup>1</sup>

<sup>1</sup>Universidad Internacional del Ecuador (UIDE), Facultad de Ingenierías Digitales y Tecnologías Emergentes, Loja, Ecuador  
arportillaca@uide.edu.ec, mipalaciosmo@uide.edu.ec, locondezh@uide.edu.ec

*Abstract– The study develops and implements a predictive system based on machine learning techniques for managing orchid pollination and germination processes at the Ecuadorian company Ecuagenera. These processes exhibit high biological and seasonal variability, which has historically forced the company to rely on manual records and empirical experience, generating operational uncertainty and production losses. The research adopts a quantitative and experimental approach, using historical data collected in the laboratory and applying feature engineering techniques to incorporate temporal, biological, and operational variables. Different supervised learning algorithms were evaluated, including Random Forest, XGBoost, and LightGBM, selecting the models with the best predictive performance according to metrics such as MAE, RMSE, and coefficient of determination ( $R^2$ ). The results show that Random Forest offers a high level of accuracy in predicting germination time, while XGBoost performs better in predicting the times associated with the pollination and germination process.*

*Keywords: Predictive models; machine learning; germination; pollination; orchids; precision agriculture; plant phenology.*

# Modelos Predictivos Para la Gestión Del Proceso De polinización y Germinación de Semillas De Orquídeas.

Arlette Stefanny Portilla Cartuche, Milton Ricardo Palacios Morocho<sup>1</sup>, Conde-Zhingre Lorena Elizabeth<sup>1</sup>

<sup>1</sup>Universidad Internacional del Ecuador (UIDE), Facultad de Ingenierías Digitales y Tecnologías Emergentes, Loja, Ecuador  
arportillaca@uide.edu.ec, mipalaciosmo@uide.edu.ec, locondzh@uide.edu.ec

**Resumen**– En el estudio se desarrolla e implementa un sistema predictivo basado en técnicas de aprendizaje automático, para la gestión de los procesos de polinización y germinación de orquídeas en la empresa ecuatoriana Ecuagenera. Estos procesos presentan una alta variabilidad biológica y estacional, lo que históricamente ha obligado a la empresa a depender de registros manuales y de la experiencia empírica, generando incertidumbre operativa y pérdidas productivas. La investigación adopta un enfoque cuantitativo y experimental, utilizando datos históricos recolectados en laboratorio y aplicando técnicas de ingeniería de características para incorporar variables temporales, biológicas y operativas. Se evaluaron distintos algoritmos de aprendizaje supervisado, entre ellos Random Forest, XGBoost y LightGBM, seleccionándose los modelos con mejor desempeño predictivo según métricas como MAE, RMSE y coeficiente de determinación ( $R^2$ ). Los resultados evidencian que Random Forest ofrece un alto nivel de precisión en la predicción del tiempo de germinación, mientras que XGBoost presenta un desempeño superior en la predicción de los tiempos asociados al proceso de polinización y germinación.

**Palabras clave:** Modelos predictivos; aprendizaje automático; germinación; polinización; orquídeas; agricultura de precisión; fenología vegetal.

## 1. INTRODUCCIÓN

En el presente estudio, se verificará la gestión de los procesos de polinización y germinación de orquídeas, en conjunto con la empresa Ecuagenera, que representa un reto importante debido a la alta variabilidad biológica y estacional de este proceso. En las empresas florícolas especializadas, la dependencia de registros manuales y de la experiencia empírica ha limitado la planificación productiva, lo que ha intensificado la incertidumbre operativa y, sobre todo, las pérdidas económicas, sobre todo en los contextos empresariales donde no se dispone de sistemas de análisis predictivo basado en la sistematización de los datos existentes [1].

En la última década, por otro lado, la aplicación de técnicas de inteligencia artificial y aprendizaje automático se ha establecido como una alternativa viable para la automatización y optimización de procesos fenológicos complejos, considerando la posibilidad de realizar modelos predictivos de eventos a partir de un análisis de diferentes variables temporales y operativas y disminuyendo de esta manera la incertidumbre que se presenta en todos los sistemas biológicos [2]. En equis sentido, la aplicación del presente tipo de técnicas en el proceso de polinización y la germinación de orquídeas es

además una oportunidad estratégica de gran relevancia, dado que se podría mejorar la planificación de la producción en empresas florícolas especializadas, tal como Ecuagenera.

Así, además, se evaluaron comparativamente algoritmos de aprendizaje supervisado basados en ensambles y boosting (entendiendo como tales el algoritmo Random Forest, XGBoost y LightGBM); así, se analizó cada uno de estos modelos de acuerdo con métricas estándar (MAE, RMSE y  $R^2$ ), para seleccionar a él algoritmo Random Forest como el más eficiente para la predicción del tiempo de germinación; mientras que el algoritmo XGBoost resultó ser el algoritmo más eficiente para el proceso de polinización; hallazgos que se encuentran en concordancia con lo que se expuso previamente con respecto a recomendaciones de métodos para sistemas biológicos complejos [4].

## 2. FUNDAMENTACIÓN TEÓRICA

### 2.1 Inteligencia artificial y agricultura de precisión

La agricultura de precisión ha sido muy útil en los últimos años, dado que junto con la innovación ofrece un enfoque que aúna tecnologías digitales de punta para hacer un uso óptimo de los recursos, aumentar la productividad y disminuir la pérdidas de tipo económico. La inteligencia artificial (IA) tiene un rol central, puesto que permite realizar un análisis automatizado de grandes volúmenes de datos de tipos heterogéneos, a partir de las bases de datos del pasado, de los sensores ambientales y de los sistemas de información productiva [3].

Distintos estudios recientes afirman que la incorporación de técnicas de aprendizaje automático en la agricultura de precisión posibilita el salto de los métodos tradicionales de análisis, caracterizados por fundamentarse sobre supuestos lineales y baja capacidad para el tratamiento de la complejidad de los sistemas [4]. Los métodos basados en árboles de decisión y algoritmos de ensamble se han especializado en el cálculo de predicciones en eventos agrícolas y fenológicos.

La literatura científica apunta a que la agricultura de precisión basada en IA no solo permite obtener mejoras en la eficiencia productiva, sino que también cabe la posibilidad de difundir la toma de decisiones estratégicas gracias a una generación de posibles escenarios predictivos fiables. Las consecuencias del error de planificación pueden traducirse en

pérdidas económicas considerables dentro de este tipo de empresa agrícola especializada [4].

Desde el lado de las TICs, la inteligencia artificial actúa como un componente integrador que conecta la captura de datos, el análisis de los mismos, y la visualización de los resultados, permitiendo la automatización gradual de los diferentes procesos productivos más complejos de realizar. Esta confluencia tecnológica sería uno de los pilares importantes de la transformación digital del sector agroindustrial en la última década [6].

### 2.2 Modelos predictivos en procesos biológicos complejos

La modelización y predicción de los procesos biológicos complejos es difícil debido a su gran variabilidad, a su naturaleza no lineal y a la existencia de múltiples factores interrelacionados. En este contexto, los modelos predictivos basados en aprendizaje automático han mostrado una interesante capacidad para aprender patrones ocultos y relaciones complejas que no pueden ser adecuadamente descritas mediante modelos estadísticos clásicos [6].

Random Forest y XGBoost son, entre los muchísimos algoritmos existentes para modelar un proceso biológico, los más usados, ambos perteneciendo a la familia de modelos de ensamble. En Random Forest se construyen algunos árboles de decisión entrenados sobre subconjuntos de los datos aleatorios, lo que permite la reducción de la varianza del modelo y el aumento en su contenido de generalización [7]. Tal y como se observa en la Fig. 1, los algoritmos evaluados presentan diferencias en cuanto a complejidad, interpretabilidad y desempeño predictivo.

Algoritmo	Casodeusotípico	Fortaleza principal	Limitación
RandomForest	Datos heterogéneos y ruidosos	Alta estabilidad y reducción del sobreajuste	Menor sensibilidad a patrones muy complejos
XGBoost	Predicción de procesos complejos y no lineales	Alta precisión y control del sobreajuste	Mayor costo computacional
LightGBM	Grandes volúmenes de datos	Alta eficiencia en tiempo y memoria	Riesgo de sobreajuste sin ajuste fino

Fig. 1. Comparación de algoritmos de aprendizaje automático evaluados (elaboración propia).

En lo correspondiente, XGBoost se basa en un enfoque de gradient boosting, en el cual los modelos se entrenan de forma secuencial, corrigiendo los errores del modelo anterior. De tal modo, la técnica ha mostrado ser muy útil para la predicción de procesos con elevada no linealidad y muchas interacciones entre variables como, por ejemplo, los eventos fenológicos y reproductivos de las plantas [8].

En investigaciones recientes se ha observado que una buena selección de las variables predictoras y un tratamiento correcto de las variables de carácter temporal, resultan

determinantes para que el modelo predictivo sea de carácter satisfactorio en sistemas biológicos.

El uso de variables cíclicas y temporales permite representar de una manera significativa los ciclos naturales, mejorando de forma importante el rendimiento del modelo [9]. En este sentido se incluyeron además variables temporales o procesos de germinación y polinización, fundamentales estos en el mes de polinización, el día del año, la semana del año, el trimestre y el año de registro, año. Aunado a esto, para captar la naturaleza cíclica de los ciclos fenológicos.

Se diseñaron variables temporales cíclicas, por medio de transformaciones trigonométricas de seno y coseno del día y del mes del año que permite que la periodicidad de las variables temporales pueda ser modelada correctamente al contribuir de forma importante a evitar interrumpir artificialmente los ciclos que van uno detrás de otro, tanto la eficacia predictiva de los modelos.

De acuerdo con las lecturas en el campo aplicado, los modelos predictivos se han utilizado de manera exitosa en la predicción de germinación, crecimiento y eventos reproductivos en diferentes tipos de cultivo, los cuales atestiguan el potencial de reducir la incertidumbre y contribuir a la planificación de operaciones en entornos productivos reales [10].

### 2.3 Germinación y polinización de orquídeas como problema fenológico

La polinización y la germinación de orquídeas son procesos fenológicos complejos que dependen de una combinación especial de factores genéticos, medioambientales y temporales. La germinación de semillas de orquídeas es diferente de los cultivos agrícolas y, en comparación con otros cultivos, las orquídeas exhiben ciclos de desarrollo muy especializados y esto aumentaría la dificultad para predecir a qué tiempo corresponde cada fase del ciclo reproductivo [9].

Se ha escrito en estudios diversos, que la germinación de las semillas de orquídeas se puede ver afectada por circunstancias estacionales, condiciones micro climáticas y las especificidades de cada género, de esta manera genera una dispersión elevada en los tiempos para desarrollarse. Todo esto convierte a la germinación en un problema fenológico idóneo para aplicar modelos predictivos basados en inteligencia artificial [10].

La polinización es un proceso que se complejiza con la interacción de los factores intrínsecos de la planta, así como también las variaciones externas, es decir, los polinizadores e incluso los recursos ambientales. Recientemente, se ha demostrado que el uso de modelos de machine learning hace un mejor trabajo al capturar estas interacciones para poder predecir

de una manera más fiable el tiempo y las probabilidades que se obtienen del éxito reproductivo [10].

En este sentido, los géneros *Lepanthes* y *Dracula* son ejemplos de casos relevantes dada su diferencia fenológica, en donde *Lepanthes* tiene un desarrollo fenológico relativamente más plano y estable y en donde el género *Dracula* es más sensible a las variaciones ambientales, lo que se asocia a una mayor disgregación en los tiempos del desarrollo fenológico. Esta disgregación refuerza la justificación de plantear modelos predictivos diferentes, y a su vez, la aplicación de técnicas avanzadas para su análisis [2].

En términos aplicados, plantear la germinación y polinización de orquídeas como un problema fenológico permite conseguir dos cosas: mejorar la planificación de las producciones y reducir pérdidas económicas asociadas a la incertidumbre y a un manejo empírico de los mismos. En Ecuagenera, la modelización predictiva de estos eventos representaba un hito hacia la automatización, la operatividad y la sostenibilidad.

### 3. CONTEXTO INSTITUCIONAL Y DIAGNÓSTICO DEL PROBLEMA EN ECUAGENERA

Ecuagenera, se trata de una entidad ecuatoriana dedicada a la producción, a la preservación y a la comercialización de orquídeas, suficientemente conocida a nivel internacional y en el país, ya que cuenta con una diversidad genética variada, además de ofrecer un importante desarrollo dentro de la producción ornamental. Su actividad productiva destaca por la utilización de especies muy sensibles a las condiciones ambientales, con largos ciclos biológicos, así como por su acentuada variabilidad fenotípica y fenológica, lo que contribuye a la complejidad de la planificación de las operaciones.

Desde un punto de vista organizacional, Ecuagenera © opera con un modo de producción donde la toma de decisiones ha estado históricamente basada, en la experiencia acumulada por el propio personal técnico y en los registros manuales de los procesos de germinación y polinización. Este modelo ha sido útil para el crecimiento inicial que ha tenido la empresa, pero muestra limitaciones significativas ante los nuevos desafíos que presentan los mercados actuales, que demandan eficiencia, trazabilidad y capacidad de anticipación [9].

La literatura más reciente indica que las empresas agroindustriales especializadas en este tipo de producción y que no incorporan herramientas digitales avanzadas, tienden a presentar altos riesgos operativos, especialmente en contextos de alta variabilidad biológica propia de las especies ornamentales [5].

#### 3.1 Descripción del sistema tradicional de gestión de información

Previo a la implementación del sistema predictivo propuesto en este estudio, Ecuagenera utilizaba un sistema tradicional de gestión de información caracterizado por la recolección manual de datos y el almacenamiento en formatos no estructurados.

Este sistema incluía:

- Registros físicos o archivos digitales aislados.
- Anotaciones manuales de fechas de polinización y germinación.
- Ausencia de una base de datos centralizada.
- Falta de estandarización en los criterios de registro.
- Dependencia del conocimiento empírico del personal técnico.

Este enfoque coincide con lo descrito en estudios recientes sobre gestión tradicional de datos agrícolas, donde se señala que la falta de digitalización limita la capacidad de análisis y dificulta la generación de conocimiento a partir de los datos históricos [3].

#### 3.2 Identificación del problema central

El principal problema identificado en Ecuagenera radica en la incapacidad del sistema tradicional para anticipar con precisión los tiempos de germinación y polinización, lo que genera una cadena de efectos negativos a nivel operativo y económico. La ausencia de modelos predictivos impide planificar de manera eficiente los recursos, incrementando el riesgo de pérdidas por descoordinación, retrasos o decisiones tardías.

Estudios recientes han demostrado que la falta de sistemas predictivos en procesos agrícolas complejos puede incrementar las pérdidas operativas entre un 15% y un 30%, especialmente en cultivos de alto valor y ciclos largos [11]

En el caso específico de Ecuagenera, este problema se agrava debido a:

- La diversidad de especies manejadas.
- Las diferencias fenológicas entre géneros como *Lepanthes* y *Dracula*.
- La variación estacional de los procesos.
- La falta de herramientas analíticas para explotar los datos históricos.

#### 3.3 Impacto económico y operativo del problema

La gestión empírica de la información ha derivado en pérdidas económicas asociadas a la mala planificación de los ciclos productivos. Estas pérdidas se manifiestan en:

- Incremento de costos operativos.
- Uso ineficiente del espacio en laboratorio.
- Retrasos en la disponibilidad de plantas.
- Dificultades para cumplir con compromisos comerciales.

La literatura especializada señala que las empresas que no adoptan soluciones basadas en datos presentan una menor capacidad de adaptación frente a la variabilidad del entorno productivo [4].

#### *Justificación de la automatización mediante TICs*

Frente a este diagnóstico, la automatización de la gestión de datos mediante TICs y modelos predictivos surge como una necesidad estratégica para Ecuagenera. La incorporación de inteligencia artificial permite transformar los datos históricos en información accionable, facilitando la toma de decisiones basadas en evidencia.

Investigaciones recientes destacan que la automatización de procesos agrícolas mediante sistemas inteligentes contribuye significativamente a la reducción de pérdidas, la optimización de recursos y la mejora de la competitividad empresarial [3]

En este sentido, el presente estudio se posiciona como una respuesta tecnológica concreta al problema identificado, proponiendo un sistema predictivo que integra:

- Digitalización estructurada de datos.
- Modelos de aprendizaje automático.
- Análisis temporal avanzado.
- Soporte a la toma de decisiones.

#### *a) Relevancia del diagnóstico para el desarrollo del estudio*

El diagnóstico realizado constituye la base sobre la cual se estructura el desarrollo metodológico del presente trabajo. La identificación clara del problema permite justificar la selección de los modelos predictivos, las variables analizadas y el enfoque metodológico adoptado.

Además, este contexto institucional refuerza el carácter aplicado de la investigación, alineándola con las demandas actuales de transformación digital en el sector agroindustrial, especialmente en la última década [2].

## 4. METODOLOGÍA

### *4.1 Recolección y estructuración de datos*

La fase inicial de la metodología se ubica en la recolección, limpieza y organización de datos de carácter histórico, obtenidos de los registros productivos de la empresa Ecuagenera. Entiéndase que estos datos son equivalentes a lo acontecido en la realidad, en cuanto a los eventos de polinización y de germinación de las semillas de orquídeas, recogidos de entre los ciclos productivos experimentados por la empresa en medio de las condiciones de laboratorio controladas.

En la línea de las afirmaciones realizadas en torno a la literatura reciente, la calidad y la organización de los datos constituyen un factor crítico en el rendimiento de los modelos predictivos, especialmente para aquellos sistemas biológicos complejos [5]; por lo que se llevó a cabo un proceso exhaustivo

de limpieza de datos, siendo retirados registros inconsistentes, valores atípicos poco representativos y datos incompletos.

Los datos originales, es decir, aquellos que se encontraban en registros manuscritos dispersos y a archivos no estandarizados, fueron transformados en una base de datos estructurada, que permitiera su análisis computacional. Este proceso de digitalización, en gran medida podría considerarse un cambio de escala frente al sistema tradicional utilizado por la empresa; de la misma forma, el proceso de digitalización se ubica dentro de los principios de la transformación digital en el sector agroindustrial [3]

### *4.2 Ingeniería de características*

Una vez estructurada la base de datos, se aplicó un proceso

Con la base de datos ya estructurada, se comenzó el proceso de feature engineering (ingeniería de características) con el objetivo de aumentar la capacidad de predicción de los modelos. Este proceso incluía transformar variables temporales en representaciones cíclicas con funciones senos y cosenos, con el objetivo de representar la continuidad estacional de los procesos biológicos.

En efecto, estudios recientes han mostrado que el feature engineering temporal es clave en el éxito de modelos predictivos de procesos fenológicos, dado que permite que se representen de manera adecuada los ciclos de la naturaleza en las plantas. [9].

Del mismo modo, las variables categóricas fueron codificadas con técnicas de encoding adecuadas a modelos basados en árboles, de modo que se evitara la pérdida de información relevante. El conjunto de características finales estuvo conformado por un total de 17 variables predictoras, seleccionadas por su biología y operatividad.

### *Modelos predictivos implementados*

#### *➤ Modelo predictivo de polinización basado en XGBoost*

Para la predicción de los tiempos de maduración posteriores a la polinización, se implementó un modelo de regresión basado en XGBoost (Extreme Gradient Boosting), seleccionado por su capacidad para modelar relaciones no lineales complejas y su alto desempeño en problemas de regresión temporal asociados a procesos biológicos.

XGBoost optimiza una función objetivo que combina una función de pérdida y un término de regularización, definida como:

$$\mathcal{L} = \sum_{i=1}^n L(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (1)$$

donde  $L$  es la función de pérdida,  $\Omega$  es el término de regularización y  $K$  es el número de árboles del modelo.

Para problemas de regresión, la función de pérdida utilizada corresponde al error cuadrático medio, expresado como:

$$L(y_i, \hat{y}_i) = (y_i - \hat{y}_i)^2 \quad (2)$$

La predicción final del modelo se obtiene como la suma de las predicciones de todos los árboles del ensamble:

$$\hat{y} = \sum_{k=1}^K f_k(\mathbf{x}) \quad (3)$$

donde  $\hat{y}$  representa los días de maduración predichos,  $f_k(\mathbf{x})$  es la predicción del árbol  $k$ -ésimo y  $\mathbf{x}$  es el vector de características de entrada.

El modelo se construye de forma aditiva; en cada iteración se añade un nuevo árbol que corrige los errores residuales del modelo anterior:

$$\hat{y}^{(t)} = \hat{y}^{(t-1)} + \eta \cdot f_t(\mathbf{x}) \quad (4)$$

donde  $\eta$  es la tasa de aprendizaje que controla la contribución de cada árbol.

El término de regularización empleado por XGBoost para prevenir el sobreajuste se define como:

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 + \alpha \sum_{j=1}^T |w_j| \quad (5)$$

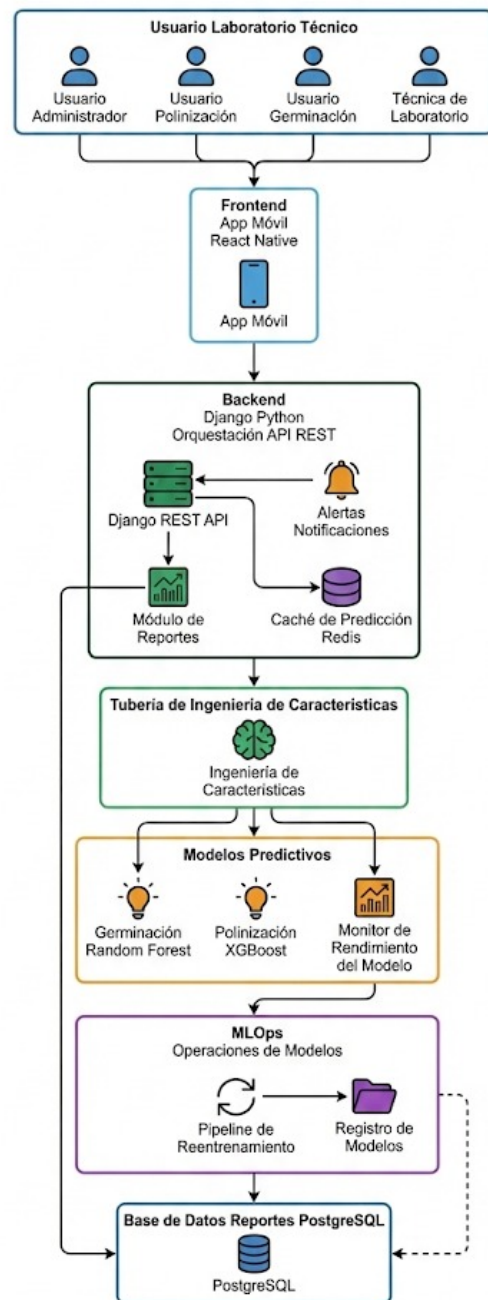
donde  $T$  es el número de hojas del árbol,  $w_j$  es el peso asociado a la hoja  $j$ ,  $\gamma$  es el parámetro de penalización por complejidad del árbol,  $\lambda$  corresponde a la regularización L2 y  $\alpha$  la regularización L1.

Esta configuración permitió obtener un modelo estable y preciso para la predicción de la maduración post polinización, ajustándose a la variabilidad temporal y biológica observada en los datos de Ecuagenera.

#### 4.4 Arquitectura del sistema predictivo PoliGer

Busca integrar los modelos de aprendizaje automático desarrollados en este estudio dentro de una solución tecnológica funcional, escalable y orientada a la automatización de los procesos de polinización y germinación de orquídeas en la empresa Ecuagenera. Esta arquitectura responde directamente a las limitaciones identificadas en el diagnóstico institucional, donde la gestión manual de datos y la ausencia de herramientas analíticas dificultaban la planificación productiva y la toma de decisiones basada en evidencia.

La arquitectura integra una aplicación móvil para el registro de datos de laboratorio, un backend basado en Django con API REST, modelos predictivos especializados (Random Forest para germinación y XGBoost para polinización), un pipeline de ingeniería de características y un enfoque MLOps para el monitoreo, reentrenamiento y versionado de modelos, se presenta en la Fig. 2,



La Fig. 2 La arquitectura del sistema PoliGer,

## 5. RESULTADOS

Se implementa los resultados obtenidos de los modelos de predicción en germinación y polinización de las especies *Lepanthes* y *Dracula*.

### 5.1 Resultados del modelo de germinación

Los resultados obtenidos para el proceso de germinación evidencian una alta capacidad predictiva de los modelos

implementados, especialmente cuando se consideran las diferencias fenológicas entre los géneros analizados.

➤ *Germinación – Lepanthes*

En el caso del género *Lepanthes*, el modelo Random Forest alcanzó un coeficiente de determinación ( $R^2$ ) de 0.835 y un error absoluto medio (MAE) de 2.56 días, lo que indica una alta precisión en la predicción del tiempo de germinación. La proximidad de los puntos al horizonte de predicción perfecta demuestra una escasa disparidad y, por lo tanto, una buena generalización que puede hacer el modelo.

Estos resultados son de acuerdo con trabajos recientes que emergen con resultados como el que estamos tratando de predecir procesos fenológicos a través de modelos de ensamble [10].

➤ *Germinación – Dracula*

Para el grupo taxonómico *Dracula* se obtuvo un  $R^2=0.885$  y un  $MAE=6.17$  días. Aunque se muestra un buen coeficiente de determinación, se apreció una mayor dispersión en los valores predichos, lo cual refleja la mayor variabilidad fenológica de este grupo, tal como sugieren los resultados mostrados en la sección anterior.

Este comportamiento ha sido recogido entre otros en la literatura científica, donde se indica que especies con alta sensibilidad ambiental presentan mayores errores locales incluso teniendo un buen ajuste global del modelo [10].

5.2 Resultados del modelo de polinización

El proceso de polinización presentó una mayor complejidad, lo que justificó el uso de modelos más robustos como XGBoost para capturar interacciones no lineales entre variables.

➤ *Polinización – Lepanthes*

El modelo XGBoost para *Lepanthes* logró un  $R^2$  de 0.961 y un MAE de 2.60 días, evidenciando un modo de funcionamiento excepcional. La elevada densidad de los puntos que se ubican alrededor de la línea de predicción perfecta manifiesta una capacidad predictiva destacada y convierte el modelo en una herramienta útil para la planificación operativa.

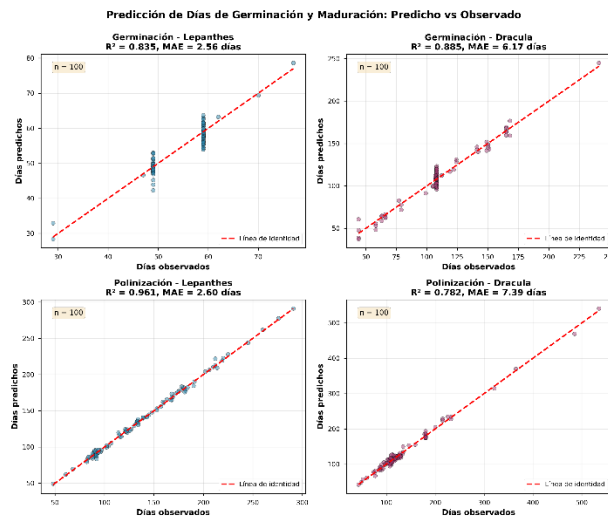
Nuestros resultados son congruentes con otros resultados similares obtenidos en otros trabajos recientes que han valorado la predicción de los eventos reproductivos vegetales a través de gradient [10].

➤ *Polinización – Dracula*

Para el caso de *Dracula*, el modelo presentó valores  $R^2$  de 0.782 y MAE de 7.39 días. El desempeño corresponde a un valor menor al obtenido para *Lepanthes* y concretamente a un valor no muy favorable desde un punto de vista práctico pero que, igualmente, permite anticipar algunas tendencias generales del proceso.

La literatura reconoce que, en sistemas biológicos altamente variables, valores de  $R^2$  cercanos a 0.75 son

considerados aceptables para aplicaciones prácticas [6]. El desempeño del modelo predictivo de germinación se evalúa mediante la comparación entre los valores observados y los valores predichos, como se muestra en la Fig. 3.



En la Fig. 3 Valores observados y los predichos para las especies *Lepanthes* y *Dracula*,

La comparación entre los valores observados y los valores predichos por los modelos de aprendizaje automático para los procesos de germinación y polinización en las especies *Lepanthes* y *Dracula*. La línea diagonal roja representa la predicción perfecta. Los valores del coeficiente de determinación ( $R^2$ ) y del error absoluto medio (MAE) evidencian un desempeño superior del modelo en la germinación y polinización de *Lepanthes*, así como una mayor variabilidad en las predicciones correspondientes a *Dracula*, asociada a su comportamiento fenológico más complejo.

El desempeño del modelo predictivo de polinización fue evaluado mediante métricas estadísticas estándar para los géneros *Lepanthes* y *Dracula*, por un mayor valor de  $R^2$  y menores errores MAE y RMSE en comparación con *Dracula*.. como se presenta en la Tabla I.

Métrica	<i>Lepanthes</i>	<i>Dracula</i>
$R^2$ Score	0.9612	0.7820
MAE (días)	2.60	7.39
RMSE (días)	10.80	22.74
Error relativo (%)	4.0 %	6.0 %
Registros totales	1,845	987
Registros de entrenamiento	1,476	789

Registros de prueba	369	198
Especies únicas	217	183
Tiempo medio de maduración	60–65 días	120–125 días

Tabla I. Métricas de desempeño predictivo de polinización.

## 6. ESTUDIOS RELACIONADOS

### 6.1 *Discusión general de los resultados obtenidos*

Los resultados alcanzados en el presente estudio confirman que los modelos predictivos basados en aprendizaje automático constituyen una herramienta eficaz para la gestión de procesos biológicos complejos, como la germinación y la polinización de semillas de orquídeas. La alta capacidad explicativa evidenciada por los valores de  $R^2$ , particularmente en el género *Lepanthes*, demuestra que los modelos implementados lograron capturar patrones relevantes asociados a la variabilidad temporal y fenológica.

Este comportamiento concuerda con lo reportado por estudios recientes, donde se señala que los algoritmos de ensamble presentan un desempeño superior en contextos agrícolas debido a su capacidad para manejar relaciones no lineales y reducir la varianza del modelo [4].

Asimismo, los resultados evidencian que la precisión del modelo depende directamente de la estabilidad biológica de la especie analizada. En este sentido, *Lepanthes* mostró una mayor predictibilidad en comparación con *Dracula*, lo cual refuerza la necesidad de enfoques diferenciados por género y especie, tal como lo sugieren investigaciones recientes en modelado fenológico [10].

### 6.2 *Comparación con estudios previos sobre germinación vegetal*

Al comparar los resultados del modelo de germinación con investigaciones similares, se observa que los valores de  $R^2$  obtenidos se encuentran dentro del rango considerado alto en estudios de predicción fenológica [10]. reportaron valores de  $R^2$  entre 0.78 y 0.88 al aplicar Random Forest en la predicción de etapas de desarrollo vegetal, resultados comparables a los alcanzados en este trabajo

En el caso específico de la germinación de orquídeas, la literatura es limitada; sin embargo, [10] destacan que la alta variabilidad intrínseca de estas especies representa un desafío para la modelización, lo que hace particularmente relevantes los resultados obtenidos en el presente estudio

La reducción del error absoluto medio lograda mediante la incorporación de variables temporales y cíclicas coincide con lo señalado por (Li et al., 2023)[9]. Quienes enfatizan que la correcta representación del tiempo es un factor crítico para mejorar la precisión de los modelos predictivos en sistemas biológicos

### ➤ *Comparación con estudios sobre polinización y eventos reproductivos*

En relación con el proceso de polinización, el desempeño del modelo XGBoost aplicado a *Lepanthes* supera los valores reportados en varios estudios recientes sobre predicción de eventos reproductivos vegetales [10]. Se reportaron valores de  $R^2$  cercanos a 0.90 al utilizar modelos de gradient boosting, mientras que el presente estudio alcanzó un  $R^2$  de 0.961

En el caso de *Dracula*, los valores obtenidos, aunque menores, se mantienen dentro de los márgenes aceptables para aplicaciones prácticas [6].

sostienen que en sistemas biológicos altamente variables, valores de  $R^2$  superiores a 0.75 son considerados adecuados para la toma de decisiones operativas

Esta comparación refuerza la validez del enfoque metodológico adoptado y confirma que las diferencias observadas no responden a limitaciones del modelo, sino a la complejidad inherente de los procesos biológicos analizados.

### ➤ *Aporte del estudio frente al estado del arte*

A diferencia de otros trabajos centrados en cultivos agrícolas tradicionales, el presente estudio aborda un contexto poco explorado: la gestión predictiva de procesos fenológicos en orquídeas. Este enfoque contribuye a ampliar el estado del arte al demostrar que las técnicas de aprendizaje automático pueden ser aplicadas con éxito en sistemas productivos altamente especializados.

Además, el uso de datos reales provenientes de una empresa productiva fortalece el carácter aplicado de la investigación, alineándose con la tendencia actual hacia estudios basados en casos reales (real-world data) [2].

## 7. IMPACTO EN ECUAGENERA

### 7.1 *Impacto tecnológico*

La implementación del sistema predictivo en este estudio representa un avance significativo para la empresa Ecuagenera, en su proceso de transformación digital. La automatización de la gestión de datos permitió pasar de un sistema manual y reactivo a una plataforma basada en análisis predictivo, alineada con los principios de la agricultura inteligente y la industria 4.0 [10].

Desde el punto de vista tecnológico, el sistema facilita:

- La digitalización estructurada de datos históricos.
- La centralización de la información.
- La generación automática de predicciones.
- La reducción de errores humanos en el registro y análisis de datos.

### 7.2 *Impacto productivo*

A nivel productivo, la capacidad de anticipar los tiempos de germinación y polinización permite mejorar la planificación

de los ciclos productivos, optimizar el uso del espacio en laboratorio y reducir retrasos en la disponibilidad de plantas.

Estudios recientes indican que la adopción de sistemas predictivos en empresas agrícolas puede mejorar la eficiencia operativa entre un 20% y un 35% [5]. En este contexto, los resultados obtenidos posicionan a Ecuagenera en una situación favorable para incrementar su competitividad.

### 7.3 Impacto económico

El impacto económico de un sistema se percibe en la disminución de pérdidas por la realización de una planificación inefectiva; es decir, la capacidad de tomar decisiones a partir de predicciones objetivas permite no sólo reducir los riesgos de que ocurran anomalías operativas, sino también optimizar la utilización de recursos.

La literatura indica que las organizaciones que incluyen analítica de datos / avanzada y modelos predictivos pueden reducir de forma considerable los costes operativos y alcanzar un mayor equilibrio financiero [3]

## 8. CONCLUSIONES

La investigación también pone de manifiesto que la automatización del manejo de datos incorporando TICs permite, entre otros, transformar procesos empíricos en sistemas que se basen en la evidencia, permitiendo la reducción de pérdidas económicas y a la misma vez mejorar la eficiencia operacional. Comparar los resultados frente a trabajos previos avala el enfoque utilizado y llega a situar a este trabajo como un referente en el mismo estado del arte.

Finalmente, se concluye que la incorporación de inteligencia artificial, análisis de datos y conocimiento biológico representa una estrategia importante para la sostenibilidad y competitividad de empresas agroindustriales, y mucho más en el contexto de transformación digital que se vive en la última década [2] de este trabajo.

## 9. REFERENCIAS BIBLIOGRÁFICAS

- [1] O. Fabian, J. Arley, W. Emilio y P. Andrea, "Inteligencia artificial en la agricultura de precisión: Tendencias y direcciones futuras," RIVAR (Santiago), vol. 12, no. 37, pp. 251–267, 2025.
- [2] R. Zhang, X. Wu, J. Li, P. Zhao, Q. Zhang, L. Wuri, D. Zhang, Z. Zhang y L. Yang, "A bibliometric review of deep learning in crop monitoring: trends, challenges, and future perspectives," *Frontiers in Artificial Intelligence*, vol. 8, 2025.
- [3] Food and Agriculture Organization of the United Nations, "e-Agriculture," FAO, 2025. [En línea]. Disponible en: <https://www.fao.org/e-agriculture/>
- [4] C. K. Ojeda Riaños, C. A. Torres, J. C. Zapata Calero, J. P. Romero-Leiton e I. F. Benavides, "A machine learning approach to map the potential agroecological complexity in an indigenous community of Colombia," *Journal of Environmental Management*, vol. 370, p. 122655, 2024.
- [5] N. Machuca, M. R. Navia Mendoza y L. Cedeño Valarezo, "Evaluación de la eficacia de modelos de machine learning para la toma de decisiones basado en datos meteorológicos en la producción agrícola," *Ecuadorian Science Journal*, vol. 9, no. 2, pp. 14–20, 2025, doi: 10.46480/esj.9.2.232.
- [6] J. Botero-Valencia, V. García-Pineda, A. Valencia-Arias, J. Valencia, E. Reyes-Vera, M. Mejía-Herrera y R. Hernández-García, "Machine learning

in sustainable agriculture: Systematic review and research perspectives," *Agriculture*, vol. 15, no. 4, p. 377, 2025.

[7] O. A. Hernández, J. E. Campos y E. Sandoval, "Phenological modeling of flowering and fruiting processes using machine learning techniques," *Computers and Electronics in Agriculture*, vol. 198, p. 107019, 2022.

[8] T. Chen y C. Guestrin, "XGBoost: A scalable tree boosting system," en *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016.

[9] Y. Li, H. Zhang y J. Wang, "Machine learning approaches for predicting biological growth processes," *Biosystems Engineering*, vol. 226, pp. 45–57, 2023.

[10] Z. Zhang, X. Liu y Y. Wang, "Ensemble learning for biological process prediction," *IEEE Access*, vol. 12, pp. 33421–33435, 2024.

[11] Y. Chabalala, E. Adam y M. Kganyago, "Phenological prediction using machine learning methods," *CABI Agriculture and Bioscience*, vol. 4, p. 51, 2023.

## ACUERDO DE CONFIDENCIALIDAD

La Biblioteca de la Universidad Internacional del Ecuador se compromete a:

1. No divulgar, utilizar ni revelar a otros la **información confidencial** obtenida en el presente trabajo, ya sea intencionalmente o por falta de cuidado en su manejo, en forma personal o bien a través de sus empleados.
2. Manejar la **información confidencial** de la misma manera en que se maneja la información propia de carácter confidencial, la cual en ninguna circunstancia podrá estar por debajo de los estándares aceptables de debida diligencia y prudencia.



---

Arlette Stefanny Portilla Cartuche  
(Estudiante)

Gabriela Fernández

---

Gabriela Fernández  
Gestora Cultural