

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

Arlette Stefanny Portilla Cartuche¹, Milton Ricardo Palacios Morocho¹, Conde-Zhingre Lorena Elizabeth¹

¹Universidad 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 Cartuche¹, Milton Ricardo Palacios Morocho¹, Conde-Zhingre Lorena Elizabeth¹

¹Universidad 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

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. Por lo cual, representa un desafío relevante debido a la alta variabilidad biológica y estacional que caracteriza a estos fenómenos. En empresas florícolas especializadas, la dependencia de registros manuales y de la experiencia empírica ha limitado la planificación productiva, incrementando la incertidumbre operativa y las pérdidas económicas, particularmente en los contextos empresariales donde no se implementan sistemas de análisis predictivo basados en el aprovechamiento sistemático de los datos disponibles [1].

En la última década, la aplicación de técnicas de inteligencia artificial y aprendizaje automático se ha consolidado como una alternativa eficaz para la automatización y optimización de procesos fenológicos complejos. Los modelos predictivos permiten anticipar eventos críticos mediante el análisis de múltiples variables temporales y operativas, reduciendo así la incertidumbre inherente a los sistemas biológicos [2].

No obstante, es importante señalar que el modelo desarrollado depende de datos históricos provenientes de una

sola empresa, por lo que su capacidad de generalización puede verse limitada al aplicarse en contextos distintos o en especies no incluidas en el entrenamiento. A pesar de ello, el sistema es apto para su implementación en otras especies dentro de Ecuagenera, siempre que presenten características similares a las consideradas, contribuyendo así a la reducción de la incertidumbre y a la optimización de los procesos productivos.

Por lo cual, se evaluaron comparativamente algoritmos de aprendizaje supervisado basados en ensambles y boosting, específicamente los algoritmos de Random Forest, XGBoost y LightGBM. Los modelos fueron analizados mediante métricas estándar (MAE, RMSE y R^2), seleccionándose a él algoritmo Random Forest como el más eficiente para la predicción del tiempo de germinación y al algoritmo XGBoost para el proceso de polinización, en concordancia con recomendaciones metodológicas para sistemas biológicos complejos [4].

2. FUNDAMENTACIÓN TEÓRICA

2.1 Inteligencia artificial y agricultura de precisión

La agricultura de precisión se ha consolidado en los últimos años como un enfoque estratégico que integra tecnologías digitales avanzadas para optimizar el uso de recursos, mejorar la productividad y reducir pérdidas económicas. Dentro de este paradigma, la inteligencia artificial (IA) desempeña un rol fundamental al permitir el análisis automatizado de grandes volúmenes de datos heterogéneos, provenientes de registros históricos, sensores ambientales y sistemas de información productiva [3].

Diversos estudios recientes coinciden en que la aplicación de técnicas de aprendizaje automático en la agricultura de precisión ha permitido superar las limitaciones de los métodos tradicionales de análisis, caracterizados por su dependencia de supuestos lineales y su baja capacidad para manejar la complejidad de los sistemas [4]. En este contexto, los modelos basados en árboles de decisión y algoritmos de ensamble se han posicionado como herramientas eficaces para la predicción de eventos agrícolas y fenológicos.

La literatura científica señala que la agricultura de precisión apoyada en IA no solo contribuye a mejorar la eficiencia productiva, sino que también facilita la toma de decisiones estratégicas mediante la generación de escenarios predictivos confiables. Esto resulta especialmente relevante en empresas agrícolas especializadas, donde los errores de

planificación pueden traducirse en pérdidas económicas significativas [4].

Desde la perspectiva de las Tecnologías de la Información y Comunicación (TICs), la inteligencia artificial actúa como un componente integrador que conecta la recolección de datos, su procesamiento analítico y la visualización de resultados, permitiendo la automatización progresiva de procesos productivos complejos. Esta convergencia tecnológica constituye uno de los pilares de la transformación digital del sector agroindustrial en la última década [6].

2.2 Modelos predictivos en procesos biológicos complejos
Los procesos biológicos complejos se caracterizan por su alta variabilidad, no linealidad y dependencia de múltiples factores interrelacionados, lo que representa un desafío significativo para su modelado y predicción. En este contexto, los modelos predictivos basados en aprendizaje automático han demostrado una notable capacidad para capturar patrones ocultos y relaciones complejas que no pueden ser descritas adecuadamente mediante enfoques estadísticos clásicos [6].

Entre los algoritmos más utilizados para el modelado de procesos biológicos se encuentran Random Forest y XGBoost, ambos pertenecientes a la familia de modelos de ensamble. Random Forest se basa en la construcción de múltiples árboles de decisión entrenados sobre subconjuntos aleatorios de datos, lo que permite reducir la varianza del modelo y mejorar su capacidad de generalización [7]. Como se observa en la Fig. 1, los algoritmos evaluados presentan diferencias significativas en cuanto a complejidad, interpretabilidad y desempeño predictivo.

Algoritmo	Caso de uso típico	Fortaleza principal	Limitación
Random Forest	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 conceptual de los algoritmos de aprendizaje automático evaluados, basada en análisis teórico y experimental (elaboración propia).

Por su parte, XGBoost utiliza un enfoque de gradient boosting, donde los modelos se entrenan de manera secuencial, corrigiendo los errores del modelo anterior. Este algoritmo incorpora mecanismos de regularización que ayudan a prevenir el sobreajuste, así como optimizaciones computacionales que mejoran su eficiencia y escalabilidad, siendo especialmente adecuado para datasets estructurados con alta complejidad. Esta estrategia ha demostrado ser particularmente eficaz en la predicción de procesos altamente no lineales y con interacciones complejas entre variables, como los eventos fenológicos y reproductivos en plantas [8].

Desde un punto de vista de la crítica del uso de los modelos predictivos en los procesos biológicos, podemos considerar tres aspectos tóxicos a la hora de su uso del mismo: un aspecto técnico, un aspecto operativo y un aspecto económico. En cuanto al aspecto técnico, hay ciertos riesgos asociados a su uso, el riesgo de sobreajustar los modelos cuando tenemos pocos datos, la sensibilidad del modelo a la calidad de los datos históricos, y la dificultad de generalizar los resultados para condiciones no observadas.

En el aspecto operativo, un uso inadecuado de las predicciones puede dar lugar a decisiones erróneas que pueden causar problemas a la planificación del proceso productivo, sobre todo si no se tiene en cuenta el margen de errores asociados al modelo. Finalmente, el aspecto económico, o el coste computacional del entrenamiento y actualización de los modelos, también es el principal punto a considerar para comenzar a utilizarlos en los entornos productivos.

Los modelos de ensamble como los Random Forest y los XGBoost pueden requerir tiempos de procesamiento altos, la propuesta que se presenta en el estudio ofrece que la forma de entrenarlos es con una máquina convencional, sin la necesidad de disponer de máquinas especializadas como las GPU, lo que hace que bajar las barreras de entrada para adoptar aquellas prácticas sea mucho más sencilla para entornos con recursos limitados de producción. Los tiempos de entrenamiento, por otro lado, están en línea con los veterinarios para poder reentrenarse periódicamente, lo que aumenta la sostenibilidad operativa del mismo.

Investigaciones recientes destacan que la selección adecuada de variables predictoras, así como el tratamiento correcto de las variables temporales, resulta determinante para el éxito de los modelos predictivos en sistemas biológicos.

2.3 Germinación y polinización de orquídeas como problema fenológico

La germinación y la polinización de orquídeas constituyen procesos fenológicos complejos que dependen de una combinación de factores genéticos, ambientales y temporales. A diferencia de otros cultivos agrícolas, las orquídeas presentan ciclos de desarrollo altamente especializados, lo que incrementa la dificultad de predecir con precisión los tiempos asociados a cada etapa del proceso reproductivo [9].

Diversos estudios han señalado que la germinación de semillas de orquídeas puede verse influenciada por variaciones estacionales, condiciones micro climáticas y características específicas de cada género, lo que genera una alta dispersión en los tiempos de desarrollo. Esta variabilidad convierte a la germinación en un problema fenológico ideal para la aplicación de modelos predictivos basados en inteligencia artificial [10].

En el caso de la polinización, la complejidad se incrementa debido a la interacción entre factores internos de la planta y condiciones externas, como la disponibilidad de polinizadores y las condiciones ambientales. Investigaciones recientes han demostrado que los modelos de aprendizaje automático pueden capturar estas interacciones de manera más efectiva que los métodos tradicionales, permitiendo predecir con mayor precisión los tiempos y probabilidades de éxito del proceso reproductivo [10].

En este contexto, los géneros *Lepanthes* y *Dracula* representan casos de estudio particularmente relevantes debido a sus diferencias fenológicas. Mientras *Lepanthes* presenta patrones de desarrollo relativamente más estables, *Dracula* exhibe una mayor sensibilidad a las variaciones ambientales, lo que se traduce en una mayor dispersión temporal. Esta diferencia refuerza la necesidad de modelos predictivos diferenciados y justifica la aplicación de técnicas avanzadas de aprendizaje automático para su análisis [2].

Desde una perspectiva aplicada, el abordaje de la germinación y polinización de orquídeas como problemas fenológicos permite no solo mejorar la planificación productiva, sino también reducir pérdidas económicas asociadas a la incertidumbre y la gestión empírica de los procesos. En empresas como Ecuagenera, la modelización predictiva de estos eventos representa un paso fundamental hacia la automatización, la eficiencia operativa y la sostenibilidad a largo plazo.

3. CONTEXTO INSTITUCIONAL Y DIAGNÓSTICO DEL PROBLEMA EN ECUAGENERA

Ecuagenera es una empresa ecuatoriana especializada en la producción, conservación y comercialización de orquídeas, reconocida a nivel nacional e internacional por su diversidad genética y su aporte al sector ornamental. Su actividad productiva se caracteriza por el manejo de especies altamente sensibles a las condiciones ambientales, con ciclos biológicos largos y una marcada variabilidad fenológica, lo que incrementa la complejidad de la planificación operativa.

Desde el punto de vista organizacional, Ecuagenera opera bajo un esquema productivo donde la toma de decisiones históricamente ha estado basada en la experiencia acumulada del personal técnico y en registros manuales de los procesos de germinación y polinización. Este modelo, si bien permitió el crecimiento inicial de la empresa, presenta limitaciones importantes frente a los actuales desafíos del mercado, que demandan eficiencia, trazabilidad y capacidad de anticipación [9].

La literatura reciente indica que las empresas agroindustriales especializadas que no incorporan herramientas digitales avanzadas tienden a enfrentar mayores riesgos operativos, especialmente en contextos de alta variabilidad biológica, como el cultivo de 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 centró en la recolección, depuración y estructuración de datos históricos provenientes de los registros productivos de la empresa Ecuagenera. Estos datos corresponden a eventos reales de polinización y germinación de semillas de orquídeas, registrados a lo largo de varios ciclos productivos, bajo condiciones controladas de laboratorio.

De acuerdo con la literatura reciente, la calidad y estructuración de los datos constituye un factor crítico en el desempeño de los modelos predictivos, especialmente en sistemas biológicos complejos [5]. Por esta razón, se realizó un proceso exhaustivo de limpieza de datos, eliminando registros

inconsistentes, valores atípicos no representativos y datos incompletos.

Los datos originales, que previamente se encontraban dispersos en registros manuales y archivos no estandarizados, fueron transformados en una base de datos estructurada, permitiendo su posterior análisis computacional. Este proceso de digitalización representa un cambio sustancial frente al sistema tradicional utilizado por la empresa, alineándose con 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 de ingeniería de características (feature engineering), con el objetivo de mejorar la capacidad predictiva de los modelos. Este proceso incluyó la transformación de variables temporales en representaciones cíclicas, utilizando funciones seno y coseno, con el fin de capturar la continuidad estacional de los procesos biológicos.

Estudios recientes destacan que la ingeniería de características temporales es determinante para el éxito de los modelos predictivos en procesos fenológicos, ya que permite representar adecuadamente los ciclos naturales de las plantas [9]. Asimismo, las variables categóricas fueron codificadas mediante técnicas de encoding adecuadas para modelos basados en árboles, evitando la pérdida de información, el conjunto final de características incluyó un total de 17 variables predictoras, seleccionadas por su relevancia biológica y operativa.

➤ Modelos predictivos implementados

Los modelos predictivos se desarrollaron, entrenaron y validaron a través del uso de una infraestructura híbrida, es decir, la combinación de recursos locales y en la nube. A nivel local, se utilizó un ordenador adoptando la arquitectura Intel Core i7 de 12ª generación con 8GB de memoria RAM, hardware suficiente para realizar tareas de preprocesamiento y realizar pruebas con datos estimados.

De igual manera, se usó la plataforma Google colab permitiendo optimizar el tiempo de entrenamiento y facilitar la gestión del código a través de un entorno accesible, escalable y basado en la nube, evitando la necesidad de contar con una infraestructura especializada como una GPU. De este modo, se garantizaba que los experimentos fueran reproducibles y podía llegarse a la conclusión de que el sistema era viable en entornos con recursos tecnológicos limitados.

➤ 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. Asimismo, este algoritmo destaca por su eficiencia computacional y su capacidad de regularización, lo que permite obtener modelos robustos incluso en presencia de datos con alta variabilidad, como es el caso de los procesos fenológicos en orquídeas.

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, Ω 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 ensemble:

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

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 η 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 , γ es el parámetro de penalización por complejidad del árbol, λ corresponde a la regularización L2 y α a la regularización L1.

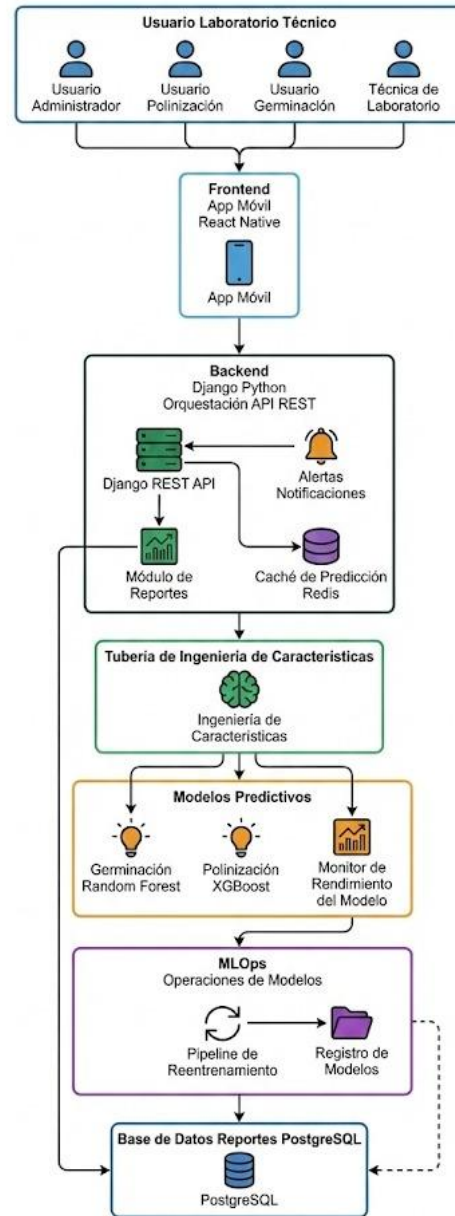
Esta configuración permitió obtener un modelo estable y preciso para la predicción de la maduración post polinización, ajustándose a la variabilidad temporal y biológica observada en los datos de Ecuagenera.

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 cercanía de los puntos a la línea de predicción perfecta refleja una baja dispersión y una buena capacidad de generalización del modelo.

Estos resultados son consistentes con estudios recientes que reportan valores similares de precisión en la predicción de procesos fenológicos utilizando modelos de ensamble [10].

➤ Germinación – *Dracula*

Para el género *Dracula*, el modelo mostró un R^2 de 0.885 y un MAE de 6.17 días. Aunque el coeficiente de determinación es elevado, se observa una mayor dispersión en los valores predichos, lo que refleja la mayor variabilidad fenológica característica de este género.

Este comportamiento ha sido documentado en la literatura, donde se señala que especies con alta sensibilidad ambiental presentan mayores errores locales, incluso cuando el ajuste global del modelo es adecuado [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 aplicado a *Lepanthes* alcanzó un R^2 de 0.961 y un MAE de 2.60 días, evidenciando un desempeño excepcional. La alta concentración de puntos alrededor de la línea de predicción perfecta indica una capacidad predictiva sobresaliente, lo que posiciona a este modelo como una herramienta confiable para la planificación operativa.

Tabla 1. Métricas de desempeño predictivo de polinización

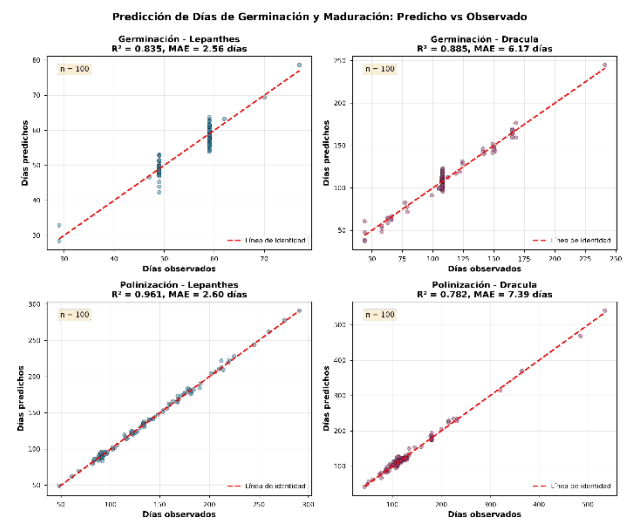
Métrica	<i>Lepanthes</i>	<i>Dracula</i>
R^2 Score	0.9612	0.7820
MAE (días)	2.60	7.39
RMSE (días)	10.80	22.74
Error relativo (%)	4.0 %	6.0 %
Registros totales	1,845	987
Registros de entrenamiento	1,476	789
Registros de prueba	369	198
Especies únicas	217	183
Tiempo medio de maduración	60–65 días	120–125 días

Resultados similares han sido reportados en investigaciones recientes sobre predicción de eventos reproductivos vegetales mediante gradient [10].

➤ Polinización – *Dracula*

En el caso de *Dracula*, el modelo obtuvo un R^2 de 0.782 y un MAE de 7.39 días. Aunque el desempeño es inferior al observado en *Lepanthes*, los resultados siguen siendo relevantes desde un punto de vista operativo, ya que permiten anticipar 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, 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.

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].

➤ 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 en el proceso de transformación digital de Ecuagenera. 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 del sistema se refleja principalmente en la reducción de pérdidas asociadas a una planificación ineficiente. La posibilidad de tomar decisiones basadas en predicciones confiables reduce el riesgo de fallos operativos y optimiza la asignación de recursos.

La literatura señala que las empresas que incorporan analítica avanzada y modelos predictivos logran una disminución significativa de costos operativos y una mayor estabilidad financiera [3]

8. CONCLUSIONES

El presente estudio demuestra que la aplicación de modelos predictivos basados en inteligencia artificial constituye una solución efectiva para la gestión de los procesos de germinación y polinización de semillas de orquídeas en la empresa Ecuagenera. Los resultados obtenidos evidencian que los algoritmos Random Forest y XGBoost presentan un alto nivel de precisión y estabilidad, especialmente cuando se aplican a especies con patrones fenológicos relativamente estables.

Asimismo, la investigación confirma que la automatización de la gestión de datos mediante TICs permite transformar procesos empíricos en sistemas basados en evidencia, contribuyendo a la reducción de pérdidas económicas y a la mejora de la eficiencia operativa. La comparación con estudios previos respalda la validez del enfoque adoptado y posiciona este trabajo como un aporte relevante al estado del arte.

A pesar de ello, implementar estos modelos conlleva riesgos por errores en dichas predicciones, que pueden influir en la toma de decisiones productivas si no son controlados. Por lo tanto, la utilización de los modelos debe necesariamente validarse continuamente y con control técnico.

Por otra parte, los costes computacionales asociados al entrenamiento resultan mínimos, dado el hecho de que no precisan de plena infraestructura, por lo que favorecen los entornos con un escaso soporte de recursos. La principal limitación de los modelos radica en que se entrenaron en los datos de una sola empresa y para géneros concretos, lo que

impide su generalización a otros ámbitos sin un proceso de adaptación.

Finalmente, se concluye que la integración de inteligencia artificial, análisis de datos y conocimiento biológico representa una estrategia clave para la sostenibilidad y competitividad de empresas agroindustriales especializadas, especialmente en el contexto actual de transformación digital entre la última década [2]

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.