

## Validación de modelos de predicción trimestrales del stock de pasturas consociadas utilizando imágenes satelitales

Juan Ignacio Malcolm



**DOCENTE TUTOR:**

Ing. Agr. Dra. Leticia Ithurrart

**DOCENTES CONSEJEROS:**

Ing. Agr. Mgtr. Romina Andrioli

Ing. Agr. Dr. Guillermo Chantre

**CONSULTORES EXTERNOS:**

Ing. Agr. Dra. Valentina Astiz

Ing. Agr. Magtr. Pascual Ciccioli



## Resumen

Con el objetivo de culminar mi formación profesional realicé una Práctica Profesional Supervisada (PPS), la cual me brindó la oportunidad de aplicar los conocimientos adquiridos a lo largo de mi carrera universitaria y, al mismo tiempo, de adquirir nuevas herramientas y experiencias.

Esta práctica tuvo lugar en el establecimiento "La Carola", situado en las cercanías del partido de General Lamadrid, en la provincia de Buenos Aires, Argentina. Dicho establecimiento pertenece a "Administración de Campos La Colina".

El trabajo se organizó en trimestres, los cuales fueron agrupados conforme a las estaciones del año: septiembre-octubre-noviembre (SON), diciembre-enero-febrero (DEF), marzo-abril-mayo (MAM) y junio-julio-agosto (JJA). De estos trimestres, los dos últimos fueron realizados por mí. La práctica se desarrolló entre el 28 de marzo y el 24 de agosto de 2024.

El trabajo consistió en asistir al establecimiento una vez al mes, portando un mapa que indicaba el stock de pasto, obtenido a partir de imágenes satelitales. En primer lugar, se tomaron los datos de los modelos de predicción del stock de pasto desarrollados y validados para la zona de General Belgrano. Posteriormente, se midió el stock de pasto disponible en las pasturas en ese momento, utilizando un pasturómetro de disco. Esto nos permitió contrastar las predicciones teóricas con las mediciones reales obtenidas en el campo. Luego, con el apoyo de inteligencia artificial, se corrigieron los modelos de predicción del stock de pasto desarrollados para la zona de General Belgrano, utilizando los datos obtenidos del sudoeste de la provincia de Buenos Aires. Este proceso garantizó la fiabilidad y la efectividad de los modelos en condiciones reales, lo que resulta esencial para ajustar y mejorar su precisión, incrementando así su capacidad predictiva a través de un enfoque basado en datos concretos y análisis continuos.

A través de esta Práctica Profesional Supervisada (PPS), logré avances significativos tanto en el ámbito técnico como en la aplicación práctica de modelos predictivos para la gestión de recursos naturales.

A lo largo de esta experiencia, adquirí valiosas competencias que previamente no había tenido la oportunidad de desarrollar. Comprendí la importancia de integrar tecnologías emergentes, como imágenes satelitales e inteligencia artificial, en la gestión de recursos naturales. Estas herramientas no solo optimizan la precisión de las predicciones, sino que también mejoran y facilitan los procesos de recolección y análisis de datos. Además, la práctica me permitió enfrentar los desafíos y las dificultades asociadas con el ajuste de modelos predictivos, tomando en cuenta las características del terreno y las variaciones climáticas de la zona. Este proceso resaltó la necesidad de basar las decisiones en datos reales y ajustados a las particularidades del entorno, reforzando así la importancia de una gestión precisa y adaptada a las condiciones locales.

## Agradecimientos

Quiero comenzar expresando mi más profundo agradecimiento a todos aquellos que han sido fundamentales en este proceso tan significativo de mi vida. Sin duda, no habría llegado al final de esta carrera sin el apoyo incondicional de mis padres, de mis familiares, de mis amigos, de la universidad, de mis consejeros externos y de todas las personas que han estado a mi lado en este recorrido.

A mis padres, quienes, con su amor incondicional, sacrificio y apoyo constante me han dado la fuerza para superar cada desafío. Gracias por creer en mí, por ser mi fuente inagotable de inspiración y por enseñarme que los sueños se alcanzan con esfuerzo, pasión y perseverancia. Sin ustedes, este logro no habría sido posible.

A mis familiares, por su cariño, comprensión y paciencia. Cada palabra de aliento, cada gesto de apoyo, ha sido un pilar fundamental en este camino tan exigente. Gracias por estar siempre ahí, en cada momento de duda y de alegría, brindándome su apoyo incondicional.

A la Universidad Nacional del Sur y a todos sus integrantes por brindarme la oportunidad de crecer, aprender y desarrollarme en un ambiente tan enriquecedor. Gracias por fomentar en mí la curiosidad, la disciplina y la responsabilidad, cualidades que me acompañarán a lo largo de toda mi vida profesional.

A mis docentes tutores, Leticia, Romina y Guillermo, por su buena onda en todo momento y por ayudarme en la finalización del Trabajo de Intensificación, y a cada uno de los docentes que han sido parte de este proceso.

A mis amigos, por hacer que este recorrido haya sido mucho más fácil y llevadero. Este logro es tan suyo como mío, porque sin su apoyo, su alegría y su confianza, no habría sido posible alcanzar esta meta. Este trabajo es un reflejo de todo lo que me han dado, y siempre llevaré en mi corazón cada gesto de generosidad y cuidado que me han brindado.

A Adm. De campos “La colina”, gracias por todo el conocimiento que me brindaron y por las experiencias inolvidables que viví. Ustedes fueron el primer paso en este camino profesional que elegí y por ello siempre estaré agradecido.

A Héctor “Juanito” Armendano, por su valiosa colaboración para conseguir el pasturómetro y facilitar las tablas de disponibilidad de forraje.

Y, para finalizar, mi más sincero agradecimiento a Valentina y a Pascual, mis consultores externos, quienes estuvieron a la par de este trabajo desde el primer día, ayudándome en cada paso, tanto en la parte práctica como teórica, brindándome su tiempo y su sabiduría de forma admirable en todo momento.

Gracias infinitas a todos.

## Índice

<b>Resumen .....</b>	<b>1</b>
<b>Agradecimientos.....</b>	<b>2</b>
<b>Introducción .....</b>	<b>4</b>
<b>Materiales y Métodos .....</b>	<b>6</b>
1) Toma de imágenes satelitales del área de estudio en General Lamadrid y cálculo de índices de vegetación .....	6
a) Índice de Vegetación de Diferencia Normalizada (NDVI):.....	7
b) Índice de Vegetación Ajustado al Suelo (SAVI):.....	8
c) Polarización VV (Vertical-Vertical) y VH (Vertical-Horizontal) en sistemas de radar de apertura sintética (SAR): .....	8
2) Vinculación de la información de General Lamadrid con la de General Belgrano en modelos predictivos de aprendizaje supervisado.....	10
a) Modelos de predicción de aprendizaje supervisado: .....	12
b) Validación a campo a través de mediciones realizadas con pasturómetro de disco: .....	16
3) Evaluación del desempeño de los modelos de predicción.....	22
<b>Resultados.....</b>	<b>24</b>
1) Análisis trimestral del desempeño de cada modelo de predicción .....	24
a) Marzo-abril-mayo .....	24
b) Junio-julio-agosto.....	27
2) Análisis mensual del desempeño del modelo Random Forest .....	29
<b>Discusión .....</b>	<b>31</b>
<b>Conclusión .....</b>	<b>33</b>
<b>Bibliografía .....</b>	<b>34</b>

## Introducción

El avance tecnológico ha transformado el sector agropecuario en las últimas décadas, facilitando la optimización de procesos productivos y la mejora en la toma de decisiones. Una de las innovaciones más significativas ha sido la incorporación de imágenes satelitales, las cuales han revolucionado la manera en que se gestionan y optimizan tanto la producción agrícola como ganadera (Marini *et al.*, 2025).

Las imágenes satelitales permiten monitorear grandes extensiones de terreno con una precisión y nivel de detalle sin precedentes. Desde la detección de cambios en la vegetación hasta el seguimiento de las condiciones del suelo, estas herramientas brindan una visión integral y continua de los campos, lo que resulta esencial para la implementación de la agricultura de precisión. Gracias a este enfoque, las empresas agropecuarias pueden tomar decisiones informadas, optimizar el uso de insumos y mejorar la eficiencia de sus operaciones (Segarra *et al.*, 2020).

En el ámbito de la gestión del pasto, las imágenes satelitales se han convertido en una herramienta indispensable para evaluar la oferta forrajera en tiempo real, aspecto crucial para la planificación del pastoreo y la gestión sostenible de los recursos forrajeros (Segarra *et al.*, 2020).

El empleo de esta herramienta consta de un paso previo ineludible: la cuantificación de la relación existente entre la información provista por la imagen satelital y la disponibilidad de forraje en el campo.

Por un lado, las imágenes satelitales proporcionan información georreferenciada sobre distintos índices de vegetación, que son calculados a partir de los valores de radiación reflejada por la superficie terrestre en sus diversas bandas de longitud de onda.

Por otro lado, el stock de pasto indica la disponibilidad de forraje expresada en base a materia seca ( $\text{kg MS.ha}^{-1}$ ). Cuando esta es obtenida a partir de un muestreo georreferenciado, se puede establecer una relación entre los índices de vegetación y la oferta forrajera.

Una vez establecida esta relación, sería posible conocer el stock de pasto en otras superficies ganaderas, exclusivamente mediante el análisis de las imágenes satelitales. No obstante, la relación establecida para un recurso forrajero en circunstancias edafoclimáticas específicas podría no coincidir con la del recurso objeto de nuestra nueva evaluación. Es por esto que resulta esencial la validación a campo para cada recurso, región y situación particular.

En el marco de mi Trabajo de Intensificación de la carrera de Ingeniería Agronómica, me incorporé a la línea de investigación que venían desarrollando la Ing. Agr. (Dra.) Valentina Astiz y el Ing. Agr. (M.Sc.) Pascual Ciccioli. Este equipo contaba con un modelo previamente ajustado, que demostraba un desempeño satisfactorio en la zona de General Belgrano. Sin embargo, al aplicarlo en pasturas ubicadas en el partido de General Lamadrid, se observaron limitaciones en su precisión y un aumento significativo del error.

A partir de esta problemática, surgió el objetivo central de mi trabajo: contribuir con información local y específica que permitiera mejorar la calibración y el ajuste del modelo en la región de General Lamadrid, con el fin de incrementar su exactitud y reducir los márgenes de error detectados. Esta contribución busca fortalecer las herramientas de manejo y toma de decisiones agronómicas en dicha zona, promoviendo una mejor adaptación de los modelos a las condiciones locales.

## Materiales y Métodos

Con el fin de cumplir con el objetivo planteado se llevaron a cabo una serie de pasos metodológicos cuidadosamente planificados:

- 1- Toma de imágenes satelitales del área de estudio en General Lamadrid y cálculo de índices de vegetación;
- 2- Vinculación de la información de General Lamadrid con la de General Belgrano en modelos predictivos de aprendizaje supervisado;
- 3- Evaluación del desempeño de los modelos de predicción.

### 1) Toma de imágenes satelitales del área de estudio en General Lamadrid y cálculo de índices de vegetación

Existen diversas plataformas que facilitan el procesamiento y la generación de información a partir de imágenes satelitales. Una de las más destacadas es Google Earth Engine (GEE), la cual es una herramienta basada en la nube que permite realizar análisis geoespaciales mediante el procesamiento de datos satelitales y su almacenamiento en la infraestructura de Google. Cuenta con una extensa base de datos que incluye imágenes de satélites como Landsat, Sentinel y MODIS, las cuales se actualizan constantemente, proporcionando un catálogo global de datos geoespaciales accesibles sin necesidad de descargar las imágenes (Tabla 1) (Perilla y Mas, 2020).

Tabla 1. Comparación entre diferentes satélites (Quintero y Sienna, 2021).

Satélite	MODIS	LANDSAT - 8	SENTINEL - 2
Resolución Pixel	250 m	30 m	10 y 20 m
Cobertura Imagen	2.200 km	185 km	290 km
Resolución Temporal	Composiciones de 8-10 días	16 días	10 días
Bandas	36 bandas (incluye NDVI, EVI)	11 bandas	12 bandas

De estos, el satélite Sentinel-2 destaca por su capacidad de proporcionar imágenes de alta resolución en diversas bandas espectrales. Esta misión de observación terrestre, parte del programa Copernicus de la Agencia Espacial Europea, y cuenta con dos satélites gemelos (Sentinel-2A y Sentinel-2B) que cubren la Tierra a una altitud de aproximadamente 786 km. Los satélites capturan imágenes de manera periódica con un ciclo de repetición de 5 días, lo que les permite monitorear grandes áreas de manera continua.

El sensor a bordo del satélite Sentinel-2 es el MSI (Multispectral Instrument), que captura imágenes en 12 bandas espectrales diferentes (Tabla 2). Estas bandas cubren desde el espectro visible hasta el infrarrojo cercano y de onda corta, lo que permite analizar fenómenos naturales de la superficie terrestre de forma detallada (Sentinel-2, 2017).

Tabla 2. Bandas espectrales del satélite Sentinel-2 con sus respectivas longitudes de onda (nm) y resolución espacial (m) (Quintero y Sierra, 2021).

Bandas Sentinel-2	Sensor	Longitud de Onda Central (µm)	Resolución (m)
Banda 1 – Costero / aerosol	VNIR	0.443	60
Banda 2 – Azul		0.49	10
Banda 3 – Verde		0.56	10
Banda 4 – Rojo		0.665	10
Banda 5 – Visible e Infrarrojo Cercano (VNIR)		0.705	20
Banda 6 – Visible e Infrarrojo Cercano (VNIR)		0.74	20
Banda 7 – Visible e Infrarrojo Cercano (VNIR)		0.783	20
Banda 8 – Visible e Infrarrojo Cercano (VNIR)		0.842	10
Banda 8A – Visible e Infrarrojo Cercano (VNIR)		0.865	20
Banda 9 – Vapor de Agua		0.945	60
Banda 10 – Cirrus	SWIR	1.375	60
Banda 11 – Onda Corta Infrarroja (SWIR)		1.61	20
Banda 12 – Onda Corta Infrarroja (SWIR)		2.19	20

El análisis de la superficie terrestre se realiza a partir del estudio de índices espectrales relacionados con el aspecto a estudiar, en nuestro caso: la disponibilidad de pasto. Por esto, a partir de las imágenes del satélite Sentinel-2, se calcularon los siguientes índices espectrales o de vegetación:

a) Índice de Vegetación de Diferencia Normalizada (NDVI):

El NDVI es un indicador de la biomasa fotosintéticamente activa presente. Este índice se calcula a partir de dos bandas: la banda roja visible (R) (Banda 4 en Sentinel-2) y la banda del infrarrojo cercano (NIR) (Banda 8 en Sentinel-2). La ecuación para calcularlo es:

$$NDVI = \frac{NIR - R}{NIR + R}$$

El NDVI varía en un rango de -1 a 1. Valores superiores a 0,1 indican la presencia de vegetación, y a medida que el valor se acerca a 1, mejores son las condiciones de cobertura y/o estado de salud de la vegetación (Figura 1).

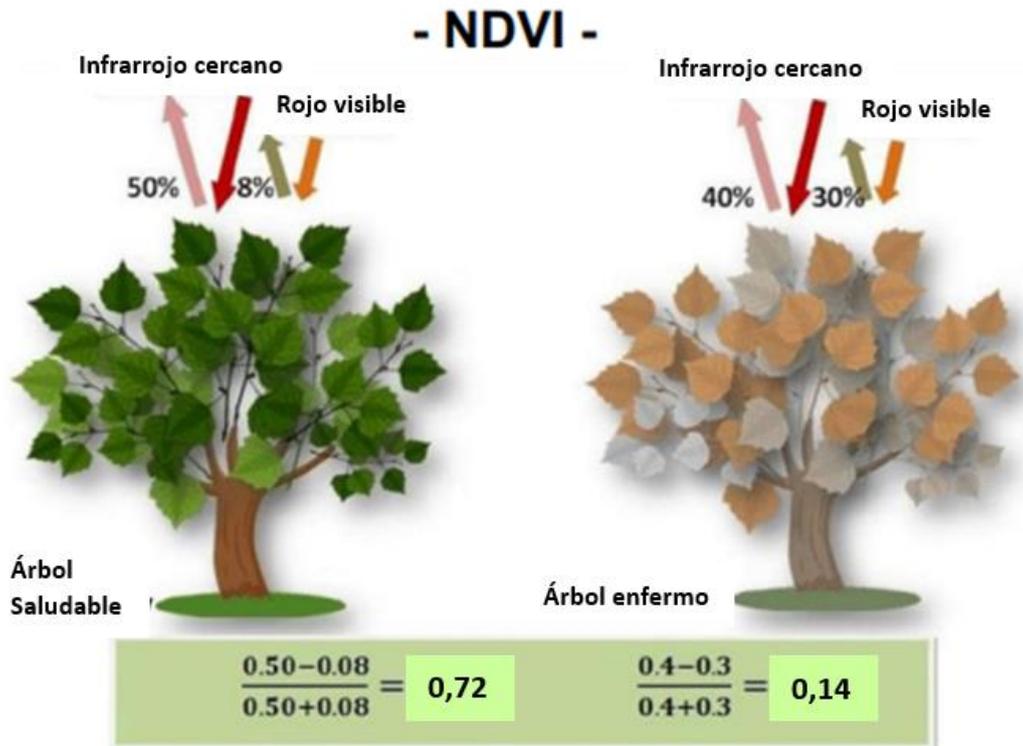


Figura 1. Diagrama de la absorción/reflexión de la energía en las porciones IR Cercano y Rojo del espectro electromagnético (Quintero y Sienna, 2021).

b) Índice de Vegetación Ajustado al Suelo (SAVI):

El SAVI es similar al NDVI, pero incluye un factor de corrección (L) para reducir la influencia del suelo en áreas con baja cobertura vegetal. La ecuación para su cálculo es:

$$SAVI = \frac{NIR - R}{NIR + R + L} \times (1 + L)$$

El factor de corrección L varía entre -1 y +1. Para su cálculo en áreas con vegetación densa, se utiliza L=0, mientras que en zonas de vegetación escasa se emplea L=1, o se ajusta a un valor medio de 0,5 para su uso general.

c) Polarización VV (Vertical-Vertical) y VH (Vertical-Horizontal) en sistemas de radar de apertura sintética (SAR):

En SAR, las polarizaciones hacen referencia a la orientación del campo electromagnético de la señal transmitida y recibida por el sensor. Esta puede ser horizontal o vertical, y ser transmitida y recibida en la misma o distintas direcciones (Figura 2). El uso del Radar SAR presenta ciertas ventajas como su funcionalidad en cualquier clima y momento del día, ya que no depende de luz solar. Además, puede penetrar nubes, humo e incluso parte de la vegetación para revelar detalles del terreno.

Así como también ofrece detalles precisos en áreas grandes y remotas (Copernicus, 2018).

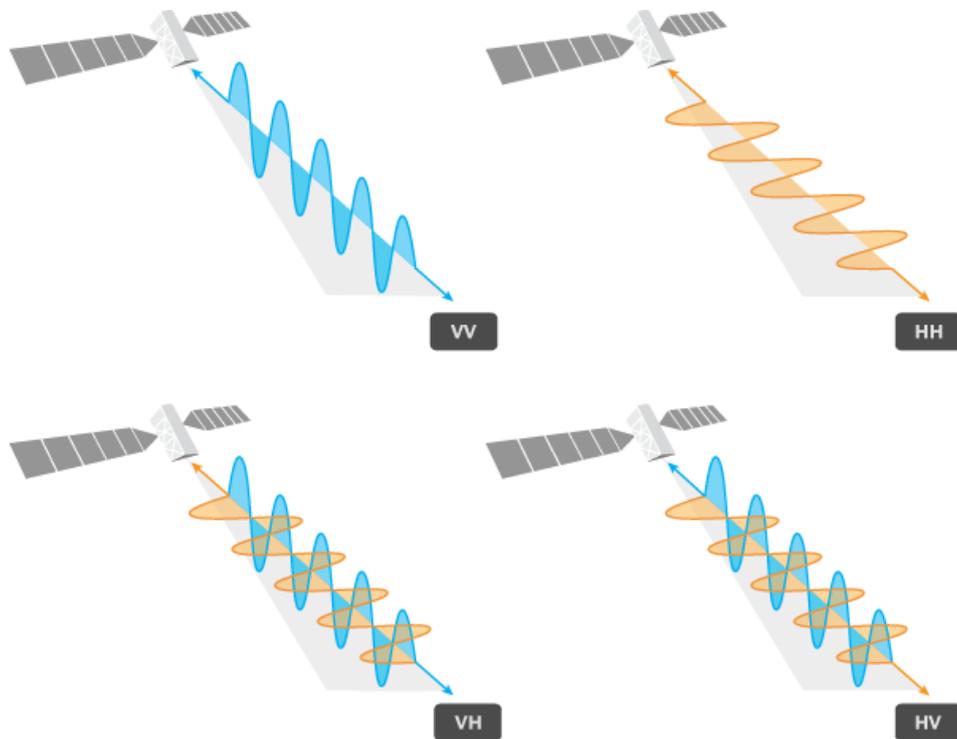


Figura 2. Transmisión y recepción de ondas polarizadas (ArcGIS Pro, 2023).

Las distintas combinaciones de transmisión-recepción permiten capturar diferentes características de la superficie terrestre, proporcionando información valiosa en una variedad de aplicaciones. En nuestro caso, las polarizaciones de interés fueron:

*Polarización VV (Vertical-Vertical):*

- Reflexión especular: es más sensible a superficies lisas como cuerpos de agua (ríos, lagos, océanos) o suelos húmedos.
- Detección de estructuras lineales: útil para identificar caminos, canales y otras características alineadas con la dirección de transmisión.
- Vegetación densa: puede proporcionar información sobre la biomasa de vegetación cerrada o cultivos de alta densidad.

### *Polarización VH (Vertical-Horizontal):*

- Dispersión volumétrica: captura mejor la interacción de las ondas con vegetación de menor densidad, como pastizales o arbustos, donde hay más dispersión entre ramas, hojas y suelo.
- Textura del terreno: más efectivo para detectar diferencias en la rugosidad del suelo o superficies heterogéneas.
- Contraste en áreas agrícolas: permite distinguir entre diferentes tipos de cultivos y sus estados de desarrollo.

La utilización de ambos modos (VV y VH) en un análisis combinado proporciona un conjunto de datos más completo:

- Mapeo de uso del suelo: diferencia entre áreas urbanas, agrícolas, forestales y cuerpos de agua.
- Monitoreo de cultivos: identifica tipos de vegetación y detecta estrés hídrico o estados fenológicos.
- Análisis de inundaciones: VV detecta agua superficial, mientras que VH ayuda a identificar vegetación inundada.
- Seguimiento de deforestación: permite detectar cambios en la cobertura forestal y degradación del suelo.

## **2) Vinculación de la información de General Lamadrid con la de General Belgrano en modelos predictivos de aprendizaje supervisado**

Los índices de vegetación y los valores de stock de pasto de General Belgrano habían permitido el desarrollo de un modelo de predicción de disponibilidad forrajera de desempeño satisfactorio en la zona de General Belgrano. A este conjunto de datos se fue incorporando trimestralmente nueva información del mismo tipo, aunque procedente de General Lamadrid, en un proceso iterativo de aprendizaje supervisado con el objetivo de mejorar gradualmente la capacidad predictiva del modelo, es decir, predecir acertadamente el stock de pasto a partir de índices de vegetación.

A grandes rasgos, la generación de un modelo de predicción de aprendizaje supervisado consta de tres etapas: procesamiento, entrenamiento, validación y prueba. A cada una de ellas se asigna una parte de todos los datos recolectados; usualmente, se asignan el 70% de los datos a la etapa de entrenamiento, 10-20% a la etapa de validación y 10-20% a la etapa de prueba (Hastie *et al.*, 2009).

La etapa de entrenamiento consiste en buscar el patrón que define la relación entre las variables de entrada y de salida; en nuestro caso, la relación entre los índices de vegetación y el stock de pasto, respectivamente. Este patrón, o relación matemática

entre variables, se busca por distintos modelos. Luego, la etapa de validación consiste en entregar a los algoritmos una nueva porción de los datos completos (entradas y salidas) para realizar un ajuste más fino de sus parámetros. Y finalmente, en la etapa de prueba, se verifica cuán bien funciona su capacidad predictiva empleando la última porción de los datos. En este último paso, se entrega al modelo solo la variable de entrada (índice de vegetación) y se recibe un valor de salida (disponibilidad de pasto). Se compara, entonces, el valor predicho con el valor real y se analiza el desempeño de cada modelo con el objeto de elegir y emplear en el futuro aquel que haya mostrado mejor capacidad predictiva.

Como ya fue mencionado, este proceso se realizó de manera iterativa en nuestro trabajo. Es decir, en un primer paso, se utilizó la información proveniente de General Belgrano para entrenar cada modelo, mientras que los datos obtenidos durante el primer mes en General Lamadrid se emplearon exclusivamente para su validación y prueba. Este enfoque permitió verificar el desempeño inicial de cada modelo en una zona diferente a aquella en la que había sido calibrado.

En los meses siguientes, el proceso de ajuste se realizó de forma progresiva para cada modelo. Por ejemplo, en el segundo mes, la etapa de entrenamiento de los modelos empleó los datos de General Belgrano; la etapa de validación empleó los datos de General Lamadrid del primer mes; y la etapa de prueba empleó los datos del segundo mes en General Lamadrid.

Esta dinámica se replicó de manera continua a lo largo de cuatro trimestres (septiembre-octubre-noviembre -SON-, diciembre-enero-febrero -DEF-, marzo-abril-mayo -MAM- y junio-julio-agosto -JJA-), permitiendo un aprendizaje iterativo y una mejora gradual en la exactitud de las predicciones de cada modelo. La plataforma utilizada para la ejecución de los códigos de aprendizaje automático fue Google Colab, un entorno basado en la nube diseñado para la ejecución de código Python. Está basado en Jupyter Notebooks, lo que permite escribir y ejecutar códigos, visualizar los resultados y documentar el trabajo de manera integrada en un único archivo (Hostgator, 2024). A continuación, se describen en más detalle los pasos operativos hasta aquí mencionados.

a) Modelos de predicción de aprendizaje supervisado:

La relación entre los índices de vegetación y el stock de pasto se modelizó por cuatro vías distintas: regresión lineal simple, regresión polinómica de segundo orden, árbol de decisiones y Random Forest. A grandes rasgos, la regresión lineal y la regresión polinómica son simples y fáciles de interpretar, pero pueden no ser adecuadas para datos complejos. Los árboles de decisión son más flexibles, pero son propensos al sobreajuste<sup>1</sup> de los datos. Random Forest, al ser un conjunto de árboles de decisión, es más robusto y preciso, especialmente con datos complejos, pero su interpretación es más difícil. El desempeño de estos dependerá de las características de los datos y del problema a resolver (Hastie *et al.*, 2009).

**Regresión lineal.** Es un modelo estadístico que se utiliza para predecir el valor de una variable dependiente (y: disponibilidad de pasto) a partir de una variable independiente (x: índice de vegetación) mediante una línea recta de ecuación  $y = ax + b$  (Figura 3). En esencia, busca la línea recta que mejor se ajusta a un conjunto de datos (x e y conocidos); es decir, busca la recta que minimiza la distancia de los puntos a ella. Esto permite entender y predecir cómo cambia una variable en función de la otra minimizando el error de predicción. Puede ser simple, con una única variable independiente (Figura 3), o múltiple, con más de una variable independiente tal como fue nuestro caso al tomar múltiples índices de vegetación (Master's in Data Science, 2025).

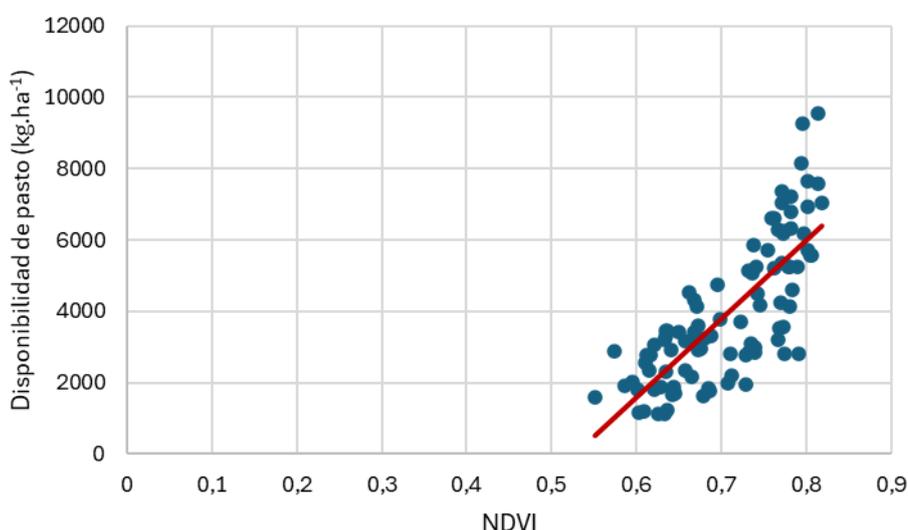


Figura 3. Ejemplo de regresión lineal simple para datos de disponibilidad de pasto en función del NDVI.

---

<sup>1</sup> El sobreajuste en aprendizaje automático se refiere a un modelo que se ajusta demasiado a los datos de entrenamiento, capturando ruido y patrones irrelevantes en lugar de las relaciones subyacentes generales. Esto resulta en un excelente rendimiento en los datos de entrenamiento, pero un pobre desempeño en datos nuevos o no vistos.

**Regresión polinómica de segundo orden.** En el contexto del aprendizaje supervisado, los modelos polinómicos se emplean para capturar relaciones no lineales entre las variables de entrada y la variable objetivo. Este enfoque es particularmente útil en situaciones donde los datos muestran curvaturas o patrones que no pueden ser representados mediante una regresión lineal. Se puede ajustar el grado del polinomio en función de la complejidad del comportamiento observado en los datos.

No obstante, esta técnica también presenta ciertas limitaciones. Si se selecciona un grado de polinomio demasiado alto, el modelo puede ajustarse excesivamente a los datos de entrenamiento, lo que compromete su capacidad de generalización y lo vuelve propenso al sobreajuste. Asimismo, la regresión polinómica tiende a ser más sensible a la presencia de valores atípicos en comparación con la regresión lineal simple. En nuestro caso, se utilizó una regresión polinómica de segundo orden; es decir, de ecuación  $y = ax^2 + bx + c$  (Figura 4). Al igual que la regresión lineal, puede constar de una única variable independiente (Figura 4), o múltiples variables independientes tal como fue en nuestro estudio (Hastie *et al.*, 2009).

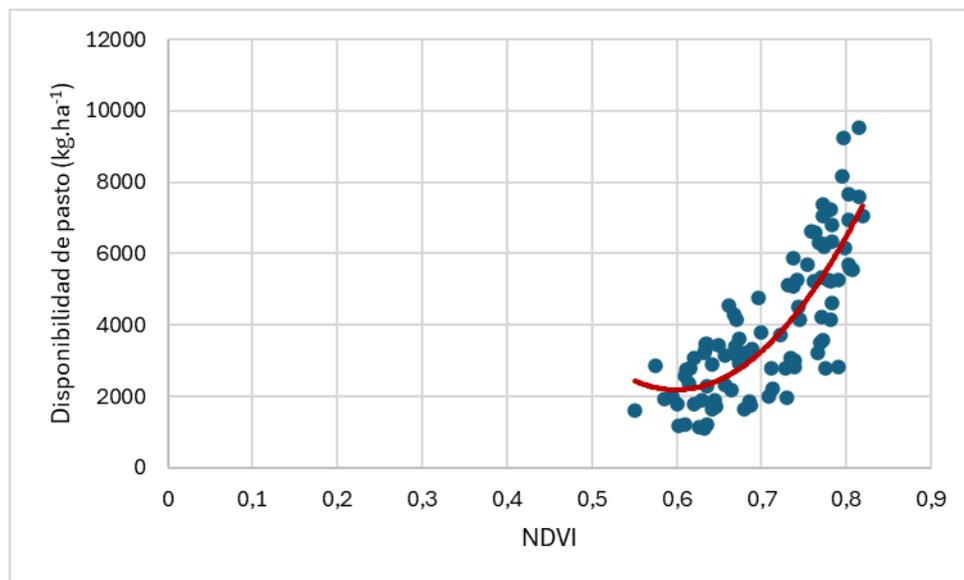


Figura 4. Ejemplo de regresión polinómica de segundo orden para datos de disponibilidad de pasto en función del NDVI.

**Árbol de decisión.** Se trata de un algoritmo que utiliza una estructura similar a un árbol para clasificar en categorías o predecir valores continuos (Figura 5). El árbol comienza con un nodo raíz, que representa la característica o las características más importantes para la clasificación o predicción. A partir de la raíz, el árbol se ramifica en función de las posibles categorías o valores de la/s característica/s. Cada nodo intermedio representa una nueva subdivisión en categorías o rango de valores continuos. Los nodos finales u hojas representan la categoría o predicción numérica final. Tal como se mencionó en los modelos de regresión lineal y polinómica, los árboles de decisión pueden construirse a partir de una única variable independiente (Figura 5) o múltiples variables independientes como en el presente trabajo (Blanco y Caparrini, 2017).

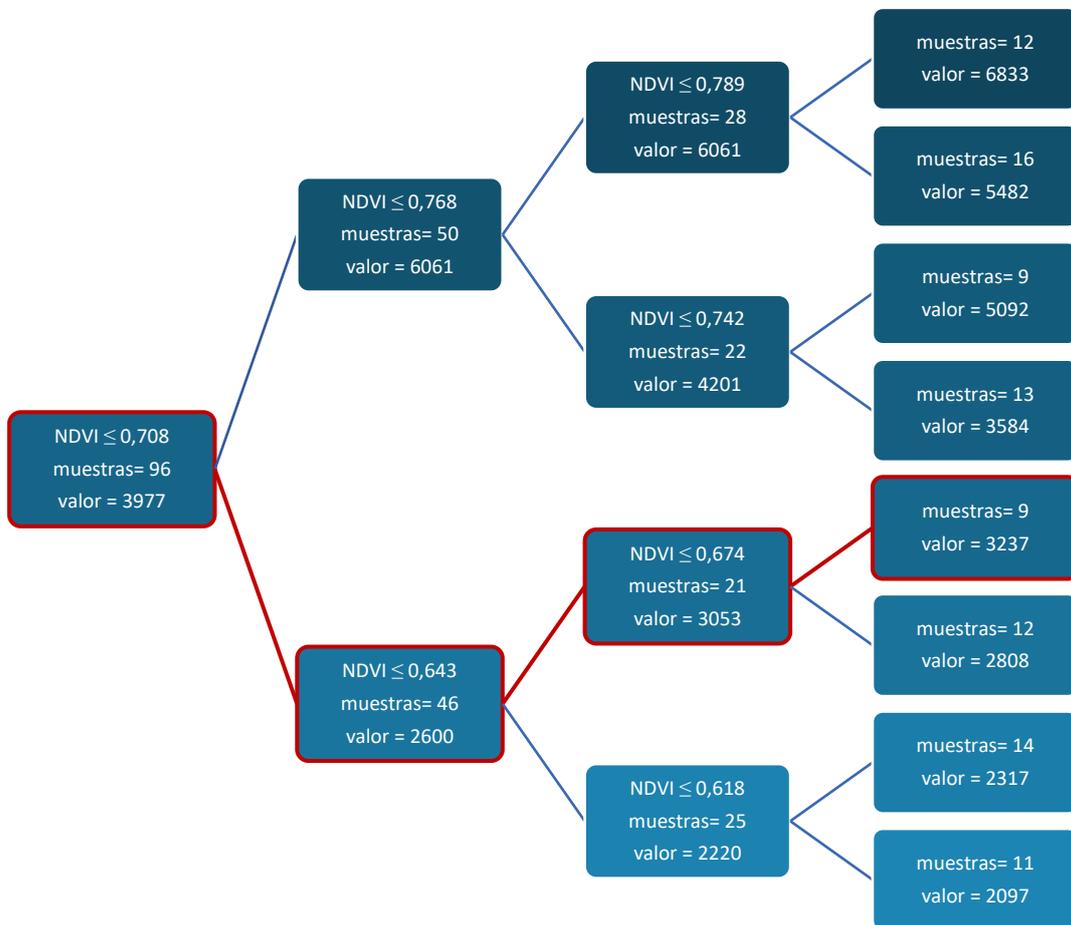


Figura 5. Ejemplo de árbol de decisión para predecir valores de disponibilidad de pasto a partir de datos de NDVI. Las bifurcaciones hacia abajo muestran que se cumple la condición declarada en el nodo del cual parten, mientras que lo contrario ocurre cuando la bifurcación se dirige hacia arriba. El recorrido marcado en rojo marca la predicción de 3237 kg MS.ha<sup>-1</sup> a partir de un valor de NDVI ≤ 0,0708 y > 0,674.

**Random Forest.** Se basa en un conjunto de árboles de decisión, donde cada uno se entrena con una submuestra aleatoria de los datos y un subconjunto aleatorio de características. Se utiliza tanto para clasificar en categorías como para predecir valores continuos. La predicción final se obtiene combinando las predicciones de todos los árboles; es decir tomando la categoría mayormente votada o calculando el promedio de los valores obtenidos, según se trate de un conjunto de categorías discretas o de datos continuos (Figura 6). En la figura se muestran cuatro árboles de decisión, cada uno con una variable predictora, pero podrían desarrollarse con múltiples variables predictoras por árbol (Brital, 2021).

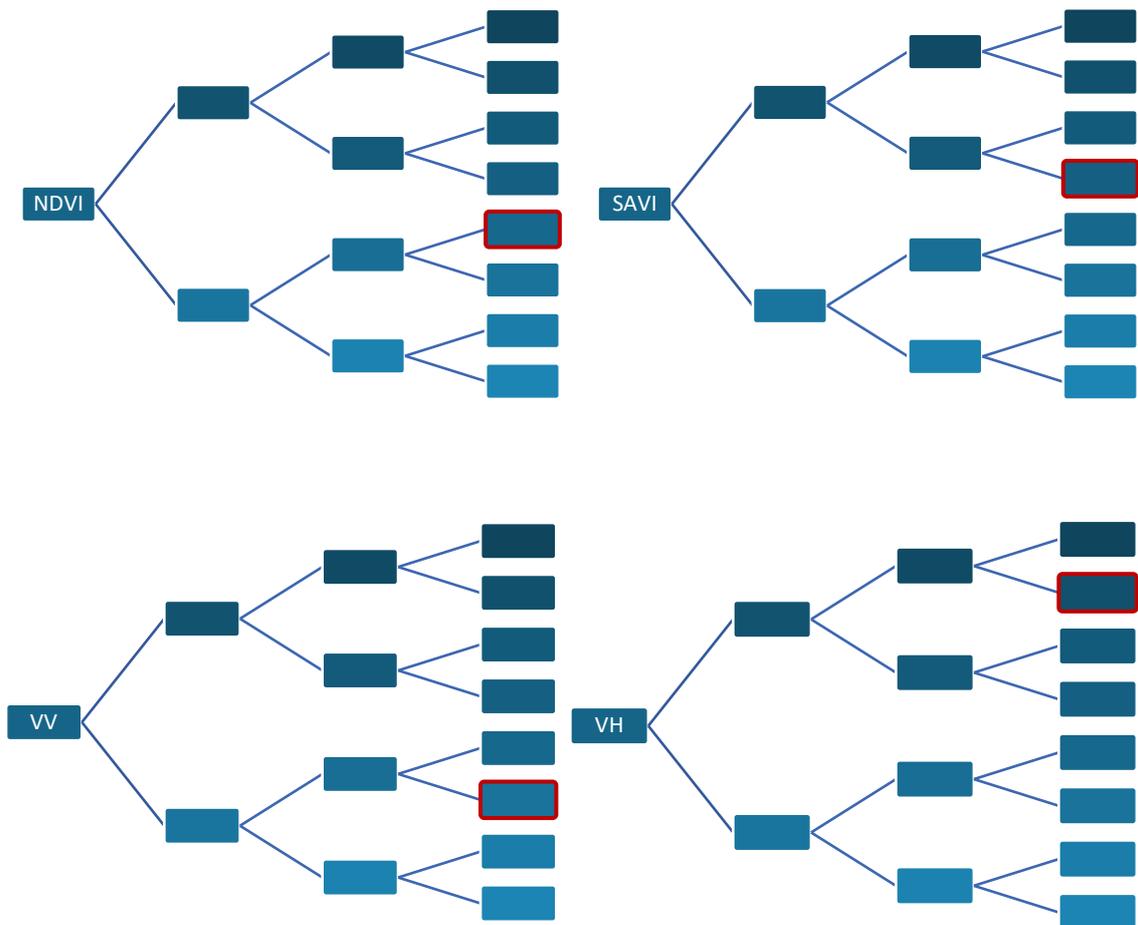


Figura 6. Ejemplo de aplicación de Random Forest para predecir valores de disponibilidad de pasto a partir de datos de NDVI, SAVI, polarización VV y polarización VH. Los recuadros marcados en rojo representan el final del recorrido para cada árbol de decisión individual. Estos valores se promediarán para obtener la predicción final.

b) Validación a campo a través de mediciones realizadas con pasturómetro de disco:

La validación de los modelos de predicción de stock de pasto se llevó a cabo en el establecimiento "La Carola" (Figura 7), ubicado en las cercanías del partido de General Lamadrid, provincia de Buenos Aires, Argentina. Las coordenadas geográficas del lugar son 37°18'58.82"S (Latitud) y 61°29'56.62"O (Longitud).

El área de estudio estuvo compuesta por 8 lotes que, en conjunto, abarcaron una superficie total de 633 ha. En estos lotes se encontraba implantada una pastura consociada compuesta por Alfalfa (*Medicago sativa*), Pasto ovilla (*Dactylis glomerata*) y trébol blanco (*Trifolium repens*), en proporciones de siembra de 8, 4 y 1 kg.ha<sup>-1</sup>, respectivamente (Figura 8). Las pasturas evaluadas tenían entre 2 y 4 años de antigüedad. El manejo de la defoliación consistía en un sistema de pastoreo rotativo a lo largo de todo el año. Cabe destacar que la alfalfa utilizada era de grupo de reposo invernal (GRI) 6, lo que implica una latencia fisiológica de aproximadamente tres meses durante el invierno. Este período de menor actividad de la alfalfa era compensado por el crecimiento activo del pasto ovilla, que aseguraba una cobertura forrajera más estable en los meses de baja temperatura.

La metodología de muestreo a campo consistía en posicionarse en puntos previamente georreferenciados mediante la aplicación Avenza Maps<sup>2</sup> (Figura 9) y efectuar 20 mediciones de la altura del pasto acumulado utilizando un pasturómetro de disco (Figura 10). Posteriormente, se sumaban los valores obtenidos y el resultado se ingresaba en una tabla, previamente calibrada para este protocolo de medición y misma consociación de cultivos, la cual permitía estimar la biomasa de pasto disponible en ese sector (Tablas 3 y 4). El valor resultante era registrado directamente en la misma aplicación, vinculando así la información productiva con su ubicación geoespacial.

Este procedimiento se repitió en 20 puntos por lote, aplicándose de manera sistemática en los 8 lotes evaluados. En total, se realizaron aproximadamente 3200 mediciones individuales, que representaron 160 puntos georreferenciados distribuidos en toda el área de estudio (Figura 11).

---

<sup>2</sup> Avenza Maps es una aplicación de navegación y mapeo que permite a los usuarios visualizar mapas georreferenciados en dispositivos móviles sin necesidad de conexión a Internet dado que el GPS del dispositivo puede mostrar la ubicación exacta del usuario sobre el mapa (Sistemas Avenza Inc., 2024).

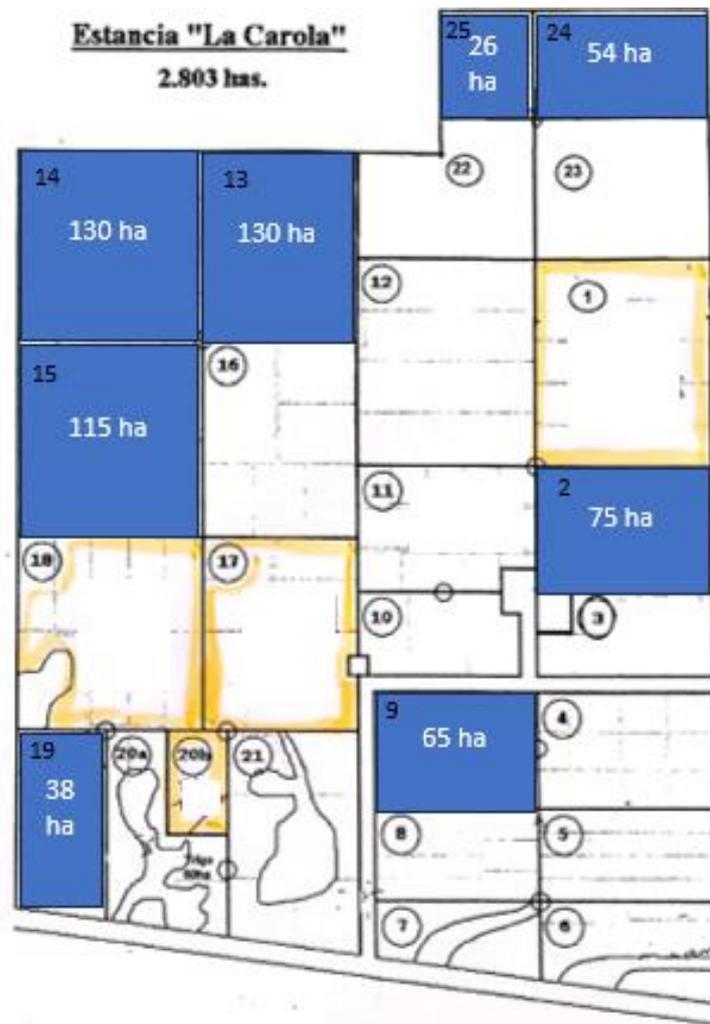


Figura 7. Mapa del Establecimiento La Carola. En azul se indican los lotes empleados para el estudio.



Figura 8. Imagen de la pastura consociada.

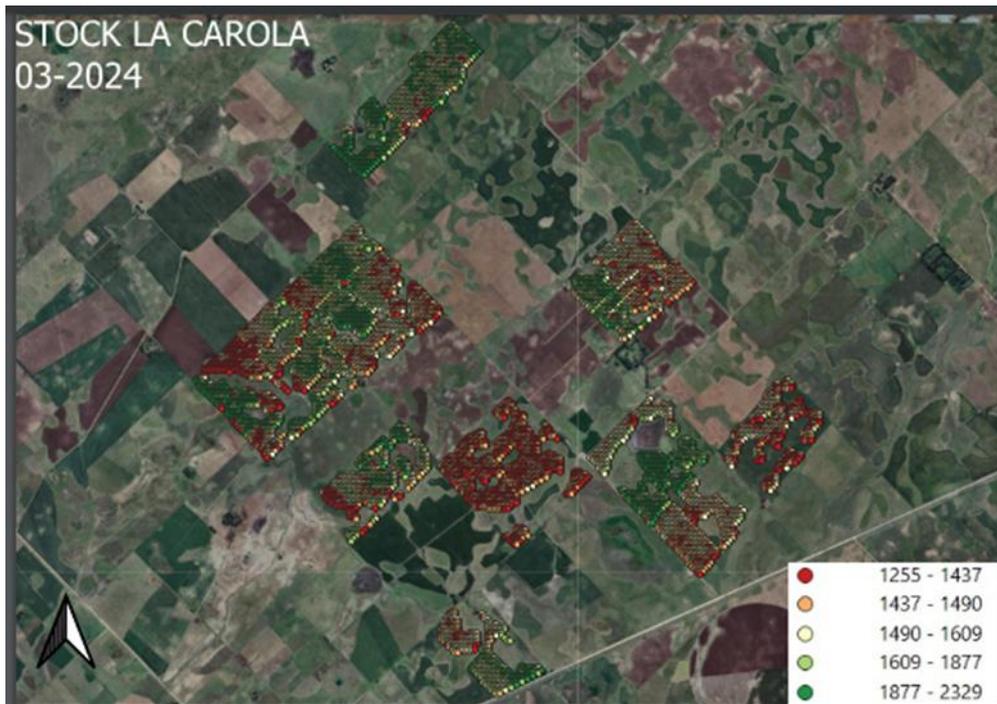


Figura 9. Mapa de stock cargado en Avenza Maps. Cada color corresponde a un rango de stock de pasto predicho por uno de los modelos.

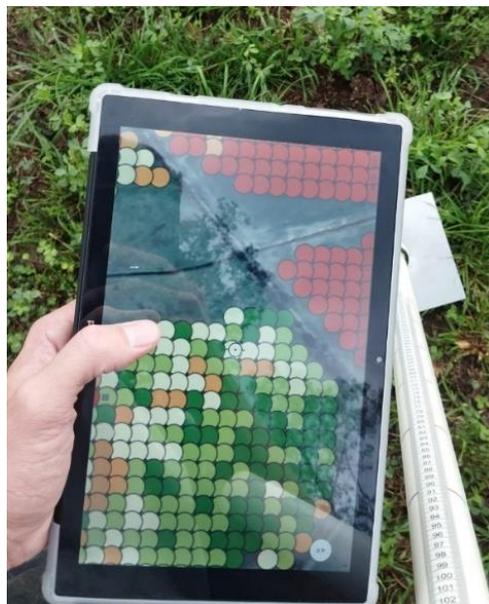


Figura 10. Imagen del pasturómetro midiendo la altura (cm) del pasto acumulado.

Tabla 3. Tabla utilizada cuando los animales aún no habían ingresado a la pastura. En la primera columna se muestra la suma de las 20 mediciones y en la segunda la disponibilidad de forraje en kg MS.ha<sup>-1</sup> (Rubio et al., 2001).

Método: Disponibilidad INICIAL							
Tomar al menos 20 mediciones por potrero, NO MENOS.							
Sumar las alturas de las 20 mediciones y comparar con el valor aproximado de disponibilidad indicado.							
Suma de 20 alturas	Disponibilidad inicial	Suma de 20 alturas	Disponibilidad inicial	Suma de 20 alturas	Disponibilidad inicial	Suma de 20 alturas	Disponibilidad inicial
10	907	110	1597	210	2287	310	2977
15	942	115	1632	215	2321	315	3011
20	976	120	1666	220	2356	320	3046
25	1011	125	1701	225	2390	325	3080
30	1045	130	1735	230	2425	330	3115
35	1080	135	1770	235	2459	335	3149
40	1114	140	1804	240	2494	340	3184
45	1149	145	1833	245	2528	345	3218
50	1183	150	1873	250	2563	350	3253
55	1216	155	1908	255	2597	355	3287
60	1252	160	1942	260	2632	360	3322
65	1287	165	1977	265	2666	365	3356
70	1321	170	2011	270	2701	370	3391
75	1356	175	2046	275	2735	375	3425
80	1390	180	2080	280	2770	380	3459
85	1425	185	2115	285	2804	385	3494
90	1459	190	2149	290	2839	390	3528
95	1494	195	2183	295	2873	395	3563
100	1528	200	2218	300	2908	400	3597
105	1563	205	2252	305	2942	405	3632

Tabla 4. Tabla utilizada cuando los animales habían ingresado a la pastura. En la primera columna se muestra la suma de las 20 mediciones y en la segunda la disponibilidad de forraje en kg MS.ha<sup>-1</sup> (Rubio et al., 2001).

Método: Disponibilidad INICIAL							
Tomar al menos 20 mediciones por potrero, NO MENOS.							
Sumar las alturas de las 20 mediciones y comparar con el valor aproximado de disponibilidad indicado.							
Suma de 20 alturas	Disponibilidad inicial	Suma de 20 alturas	Disponibilidad inicial	Suma de 20 alturas	Disponibilidad inicial	Suma de 20 alturas	Disponibilidad inicial
10	1253	110	1597	210	1942	310	2286
15	1270	115	1614	215	1959	315	2303
20	1287	120	1632	220	1976	320	2321
25	1304	125	1649	225	1993	325	2338
30	1322	130	1666	230	2011	330	2355
35	1339	135	1683	235	2028	335	2372
40	1356	140	1701	240	2045	340	2390
45	1373	145	1718	245	2062	345	2407
50	1390	150	1735	250	2080	350	2424
55	1408	155	1752	255	2097	355	2441
60	1425	160	1769	260	2114	360	2459
65	1442	165	1787	265	2131	365	2476
70	1459	170	1804	270	2148	370	2493
75	1477	175	1821	275	2166	375	2510
80	1494	180	1838	280	2183	380	2527
85	1511	185	1856	285	2200	385	2545
90	1528	190	1873	290	2217	390	2563
95	1545	195	1890	295	2235	395	2579
100	1563	200	1907	300	2252	400	2596
105	1580	205	1924	305	2269	405	2614

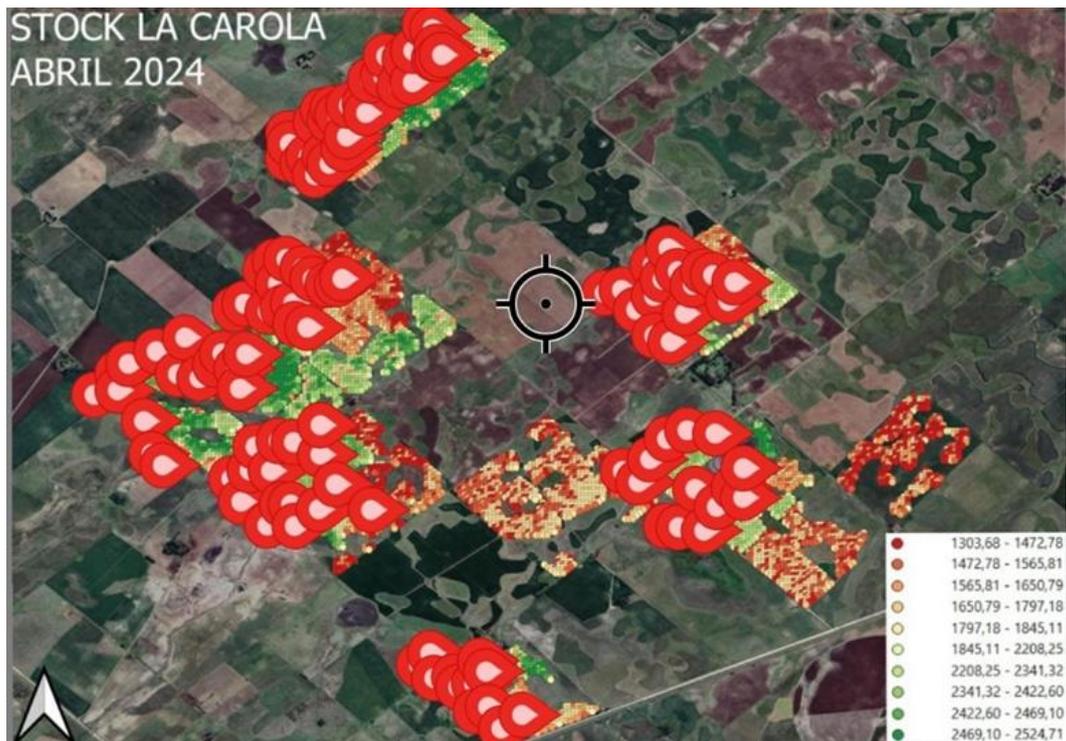


Figura 11. Mapa obtenido en Avenza Maps luego de cargar los 20 puntos por lote (con sus 20 mediciones por punto) en 8 lotes.

Finalizado el relevamiento en campo, todos los datos recolectados fueron exportados en formato CSV, para el posterior procesamiento (Figura 12).

1	Title,"Date Created", "Latitude", "Longitude", "Northing", "Easting", "Description", "Elevation"
2	Lote 19,"2024-05-31T08:35:39-03:00",-37.34377829,-61.51782242,0,0,"1114","0.000000"
3	Lote 19,"2024-05-31T08:37:58-03:00",-37.3435924765351,-61.5188415796052,0,0,"2390","0.000000"
4	Lote 19,"2024-05-31T08:41:50-03:00",-37.34261524,-61.52029475,0,0,"1356","0.000000"
5	Lote 19,"2024-05-31T08:44:10-03:00",-37.3419476051974,-61.5209049052386,0,0,"2011","0.000000"
6	Lote 19,"2024-05-31T08:49:26-03:00",-37.34287465,-61.5161199,0,0,"2287","0.000000"
7	Lote 19,"2024-05-31T08:51:35-03:00",-37.34226206,-61.51582778,0,0,"2321","0.000000"
8	Lote 19,"2024-05-31T08:54:36-03:00",-37.34253614,-61.51467814,0,0,"1735","0.000000"
9	Lote 19,"2024-05-31T08:57:33-03:00",-37.34397511,-61.51663508,0,0,"1425","0.000000"
10	Lote 19,"2024-05-31T08:59:57-03:00",-37.34550329,-61.51586444,0,0,"1632","0.000000"
11	Lote 19,"2024-05-31T09:03:08-03:00",-37.34710215,-61.51708,0,0,"2287","0.000000"
12	Lote 19,"2024-05-31T09:05:19-03:00",-37.34732282,-61.51571786,0,0,"2356","0.000000"
13	Lote 19,"2024-05-31T09:08:13-03:00",-37.34733281,-61.5135371,0,0,"2252","0.000000"
14	Lote 19,"2024-05-31T09:10:30-03:00",-37.34710203,-61.51221134,0,0,"2287","0.000000"
15	Lote 19,"2024-05-31T09:12:36-03:00",-37.34678613,-61.51086527,0,0,"2149","0.000000"
16	Lote 19,"2024-05-31T09:33:04-03:00",-37.32507937,-61.50214341,0,0,"2218","0.000000"

Figura 12. Imagen del archivo con formato CSV con las mediciones georeferenciadas.

Es relevante señalar en este punto las diferencias climáticas de los ambientes involucrados en este estudio. A partir del análisis de los gráficos de precipitaciones y temperaturas anuales (Figura 13 y Figura 14), se pueden observar diferencias significativas en los climas de General Belgrano y General Lamadrid.

General Belgrano presenta un régimen de lluvias más alto y regular a lo largo del año. Las precipitaciones mensuales rara vez bajan de los 40 mm, incluso en los meses de invierno. En cambio, General Lamadrid muestra un patrón más seco y variable, con mínimos cercanos a los 20 mm durante el invierno. Esta diferencia sugiere que el clima de General Belgrano es más húmedo y estable, lo que puede traducirse en mejores condiciones para los cultivos sin necesidad de riego suplementario, mientras que General Lamadrid podría presentar limitantes hídricas en ciertas épocas del año.

En cuanto a las temperaturas, ambas localidades presentan veranos calurosos, con máximas que superan los 30 °C. No obstante, General Lamadrid experimenta inviernos más fríos, con temperaturas mínimas que se acercan a los 0 °C, mientras que en General Belgrano los valores mínimos invernales son algo más altos. Esto indica que General Lamadrid tiene un clima más continental, con mayor amplitud térmica anual, lo cual puede implicar mayor riesgo de heladas y una ventana de crecimiento del forraje más corta que en General Belgrano.

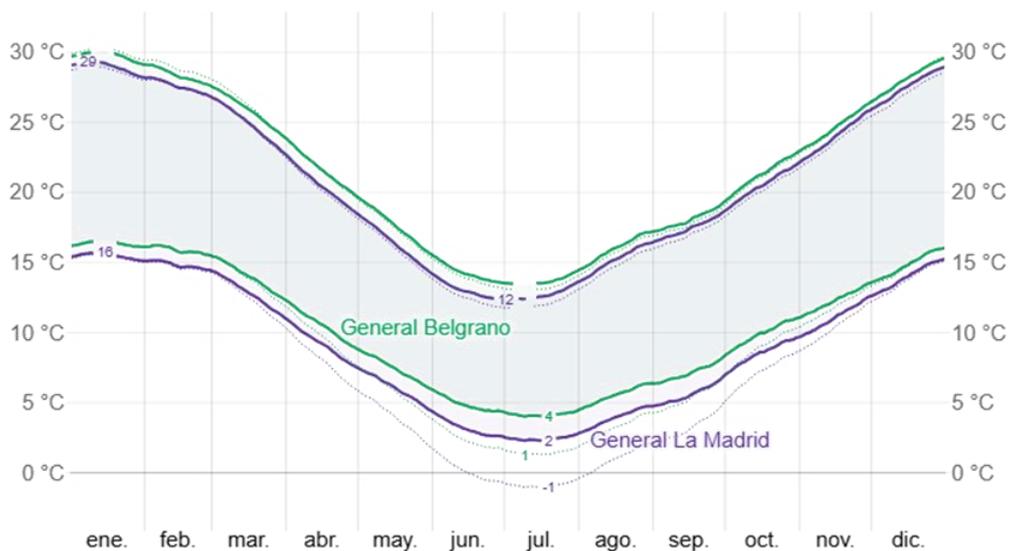


Figura 13. Temperatura media mensual (1973-2025) para las localidades de General Belgrano y General Lamadrid (Weather Spark, 2016a y b).

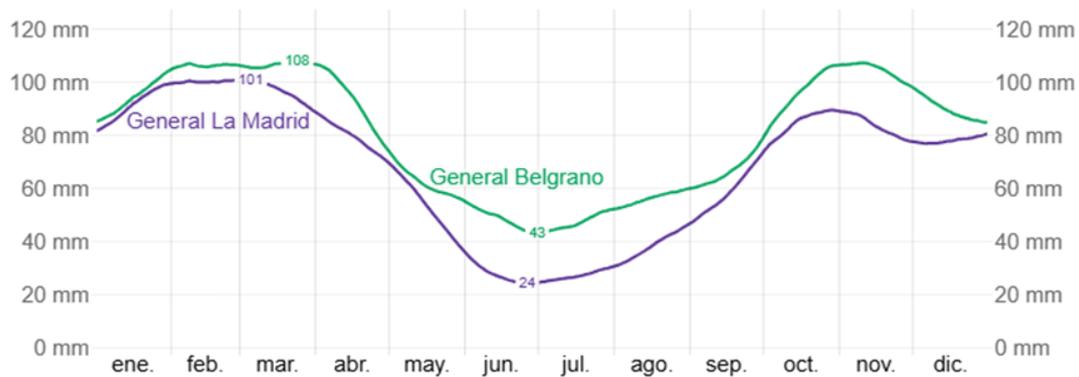


Figura 14. Precipitación media mensual (1973-2025) para las localidades de General Belgrano y General Lamadrid (Weather Spark, 2016a y b).

### 3) Evaluación del desempeño de los modelos de predicción

Con el propósito de evaluar el desempeño de cada uno de los cuatro modelos desarrollados (regresión lineal, regresión polinómica de segundo orden, árbol de decisión y Random Forest) se compararon los valores predichos por ellos con los observados en el campo. Para esto los datos recolectados fueron cargados en la plataforma Google Earth Engine (GEE).

Luego de desarrollar y ajustar cada uno de los modelos se utilizó el RMSE (raíz del error cuadrático medio) como principal criterio para evaluar su rendimiento y determinar cuál presentaba la mejor capacidad de predicción.

El RMSE es una medida de la diferencia entre los valores predichos por un modelo y los valores observados, y se calcula como la raíz cuadrada de la media de los cuadrados de las diferencias. Cuanto menor sea su valor, mejor será la capacidad predictiva del modelo. Este se calcula tanto para los datos de entrenamiento (RMSE train) como para los datos de prueba (RMSE test).

El RMSE en el conjunto de entrenamiento refleja cuán bien el modelo ajusta los datos con los que fue entrenado:

- RMSE bajo en train: indica que el modelo ha aprendido bien los patrones en los datos de entrenamiento. Sin embargo, un RMSE demasiado bajo (cercano a 0) podría ser señal de sobreajuste (overfitting), especialmente si el RMSE en el conjunto de test es mucho mayor.
- RMSE alto en train: sugiere que el modelo no está capturando bien los patrones en los datos de entrenamiento, lo que podría indicar subajuste (underfitting). Esto puede ocurrir si el modelo es demasiado simple (por ejemplo, una regresión lineal para datos no lineales) o si los indicadores espectrales (NDVI, SAVI, VH, VV) no son suficientemente informativos.

El RMSE en el conjunto de prueba mide la capacidad del modelo para generalizar a datos nuevos, no vistos durante el entrenamiento:

- RMSE bajo en test: indica que el modelo generaliza bien y puede predecir el stock de pasto con precisión en nuevas observaciones.
- RMSE alto en test: sugiere que el modelo no generaliza bien, lo que podría deberse a sobreajuste (el modelo memorizó los datos de entrenamiento, pero no aprendió patrones generales) o a una mala selección de características.

La relación entre los valores de RMSE en train y test es clave para evaluar el rendimiento del modelo:

- $RMSE_{train} \approx RMSE_{test}$ : si los valores son similares y bajos, el modelo tiene un buen equilibrio entre ajuste y generalización. Es el escenario ideal.
- $RMSE_{train} < RMSE_{test}$ : es común que el RMSE en train sea menor que en test, ya que el modelo está optimizado para los datos de entrenamiento. Sin embargo:
  - Si la diferencia es pequeña, el modelo generaliza bien.
  - Si la diferencia es grande, es una señal de sobreajuste. El modelo se ajustó demasiado a los datos de entrenamiento y no puede predecir bien en datos nuevos.
- $RMSE_{train} > RMSE_{test}$ : esto es raro, pero podría indicar un problema con la división de los datos (por ejemplo, el conjunto de prueba contiene casos más fáciles de predecir) o un error en el proceso de validación.

## Resultados

Como se mencionó anteriormente, los datos fueron agrupados en trimestres según las estaciones del año, de los cuales participé en los dos últimos: marzo-abril-mayo (MAM) y junio-julio-agosto (JJA). A continuación, se presenta (1) el análisis del desempeño de cada modelo en cada uno de estos dos trimestres y (2) análisis mes a mes del modelo que mostró el mejor desempeño en el análisis trimestral (Random Forest).

### 1) Análisis trimestral del desempeño de cada modelo de predicción

#### a) Marzo-abril-mayo

En la Figura 15 se observa el estado de la pastura en cada uno de los meses de este trimestre. En la Tabla 5 se presentan los resultados obtenidos para los modelos evaluados, comparando sus desempeños en términos de RMSE (raíz del error cuadrático medio) tanto para los datos de entrenamiento como los de prueba.

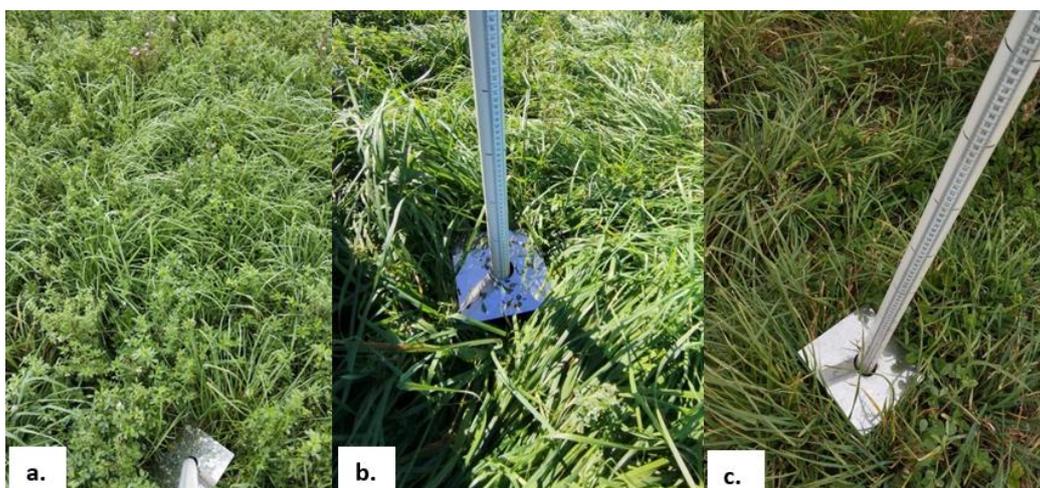


Figura 15. Fotografía de la pastura en el mes de marzo (a), abril (b) y mayo (c).

Tabla 5. Evaluación del desempeño de distintos modelos de aprendizaje supervisado en el trimestre marzo-abril-mayo.

MODELOS	VARIABLES	RMSE TRAIN (kg MS.ha <sup>-1</sup> )	RMSE TEST (kg MS.ha <sup>-1</sup> )
Árbol de decisión	mes, NDVI, SAVI, VH y VV	200	695
Regresión lineal		547	615
Regresión polinómica		497	651
Random Forest		108	310

El árbol de decisión presentó un RMSE de 200 kg MS.ha<sup>-1</sup> en entrenamiento y 695 kg MS.ha<sup>-1</sup> en prueba. Esto indica un sobreajuste en el modelo, lo que es común en los árboles de decisión debido a su alta capacidad de aprender patrones específicos, pero a su vez su falta de generalización (Breiman *et al.*, 1984).

Regresión lineal obtuvo un RMSE de 547 kg MS.ha<sup>-1</sup> en entrenamiento y 615 kg MS.ha<sup>-1</sup> en prueba, lo que sugiere que este modelo, aunque más simple y fácil de interpretar, no es tan efectivo para capturar la complejidad de la relación entre las variables.

El modelo polinómico de segundo orden presentó un RMSE de 497 kg MS.ha<sup>-1</sup> en entrenamiento y 651 kg MS.ha<sup>-1</sup> en prueba. Aunque mejor que la regresión lineal, la regresión polinómica puede ser sensible a sobreajustes cuando se elige un grado u orden de polinomio muy alto, lo que podría haber ocurrido en este caso (Hastie *et al.*, 2009).

El modelo Random Forest resultó ser el más preciso, con un RMSE de 108 kg MS.ha<sup>-1</sup> en entrenamiento y 310 kg MS.ha<sup>-1</sup> en prueba. Esto indica que, al ser un conjunto de árboles que promedia sus predicciones, Random Forest reduce el riesgo de sobreajuste y mejora la capacidad de generalización (Breiman, 2001). Este modelo ha demostrado ser el más adecuado para la estimación del stock de pasto en base a los índices de vegetación.

Los gráficos de distribución de los errores y los gráficos de dispersión de valores predichos versus reales (Figura 16) reflejan lo antedicho. A izquierda, el árbol de decisión muestra una altísima densidad de errores muy pequeños durante el entrenamiento y un amplio rango de tamaños de errores a baja densidad durante la prueba del modelo con datos desconocidos. Esto muestra gráficamente el síntoma de sobreajuste típico de este modelo de predicción. A su derecha, el gráfico de dispersión de valores predichos versus reales no muestra ningún signo de acercamiento a la relación 1:1 esperada. En las regresiones, tanto lineal como polinómica, se observa la similitud en los errores de entrenamiento y prueba en los gráficos de distribución de errores. A su lado, los gráficos de dispersión de valores predichos versus reales muestran nubes de puntos más densas, probablemente agrupados a lo largo de una recta, pero alejados de la relación 1:1 buscada. Por último, en el gráfico de distribución de errores del modelo Random Forest se ve reflejada su capacidad predictiva frente a datos desconocidos (dado que el rango de valores en el eje x es más acotado que en los otros modelos) y se puede observar también su menor tendencia al sobreajuste cuando se lo compara con el árbol de decisión. El gráfico de dispersión de valores predichos versus reales del modelo Random Forest fue el que mejor desempeño tuvo frente a la relación 1:1 esperada.

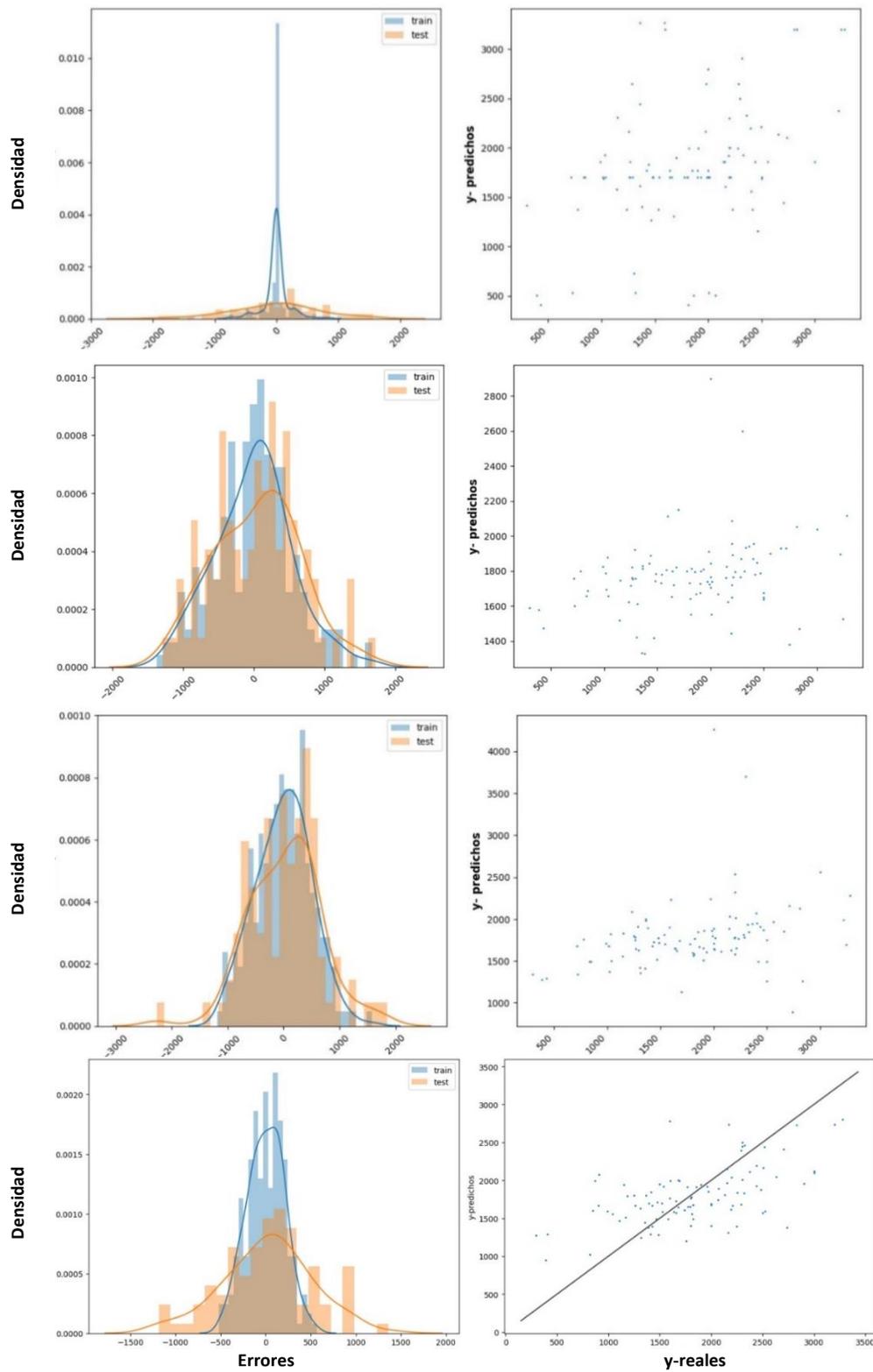


Figura 16. Gráficos de distribución de los errores (izquierda) y gráficos de dispersión de “y-reales” vs “y-predichos” (derecha) para los modelos árbol de decisión, regresión lineal, regresión polinómica de segundo grado y Random Forest (de arriba hacia abajo) para el trimestre marzo-abril-mayo.

b) Junio-julio-agosto

En la Figura 17 se observa el estado de la pastura en cada uno de los meses de este trimestre. En la Tabla 6 se presentan los resultados obtenidos para los modelos evaluados, comparando sus desempeños en términos de RMSE (raíz del error cuadrático medio) tanto para los datos de entrenamiento como los de prueba.

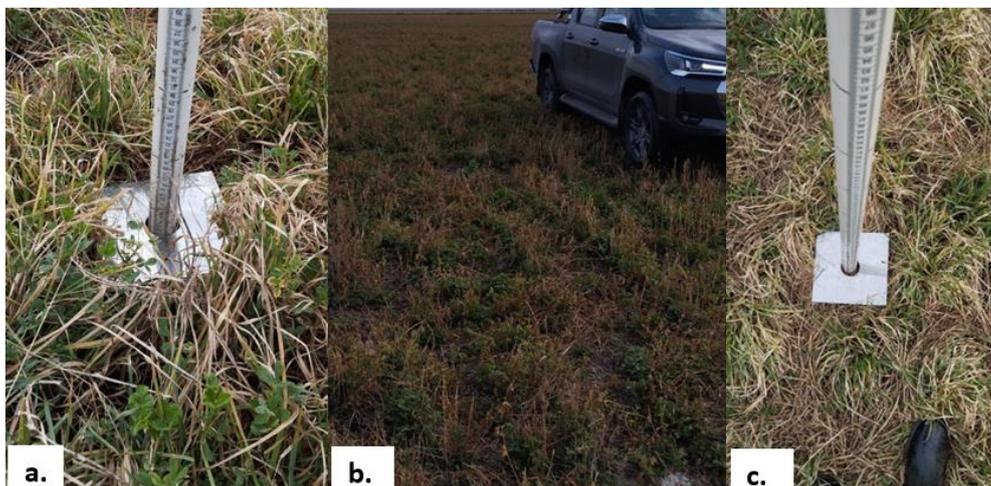


Figura 17. Fotografía de la pastura en los meses de junio (a), julio (b) y agosto (c).

Tabla 6. Evaluación del desempeño de distintos modelos de aprendizaje supervisado en el trimestre junio-julio agosto.

MODELOS	VARIABLES	RMSE TRAIN (kg MS.ha <sup>-1</sup> )	RMSE TEST (kg MS.ha <sup>-1</sup> )
Árbol de decisión	mes, NDVI, SAVI, VH y VV	186	644
Regresión lineal		476	453
Regresión polinómica		456	473
Random Forest		180	488

El árbol de decisión presentó un RMSE de 186 kg MS.ha<sup>-1</sup> en entrenamiento y 644 kg MS.ha<sup>-1</sup> en prueba, por lo que se puede decir que este modelo sigue teniendo dificultades para generalizar. La regresión lineal obtuvo un RMSE de 476 kg MS.ha<sup>-1</sup> en entrenamiento y 453 kg MS.ha<sup>-1</sup> en prueba. Aunque la regresión lineal no es tan precisa como el modelo Random Forest, su capacidad para generalizar es mejor que la del árbol de decisión. La regresión polinómica de segundo grado tuvo un RMSE de 456 kg MS.ha<sup>-1</sup> en entrenamiento y 473 kg MS.ha<sup>-1</sup> en prueba. Este modelo, al igual que la regresión lineal, tiene una capacidad predictiva inferior a Random Forest. El modelo Random Forest continuó destacándose con un RMSE de 180 kg MS.ha<sup>-1</sup> en entrenamiento y 488 kg MS.ha<sup>-1</sup> en prueba, confirmando su robustez para este conjunto de datos y su capacidad para manejar complejas relaciones no lineales entre las variables (Liaw y Wiener, 2002).

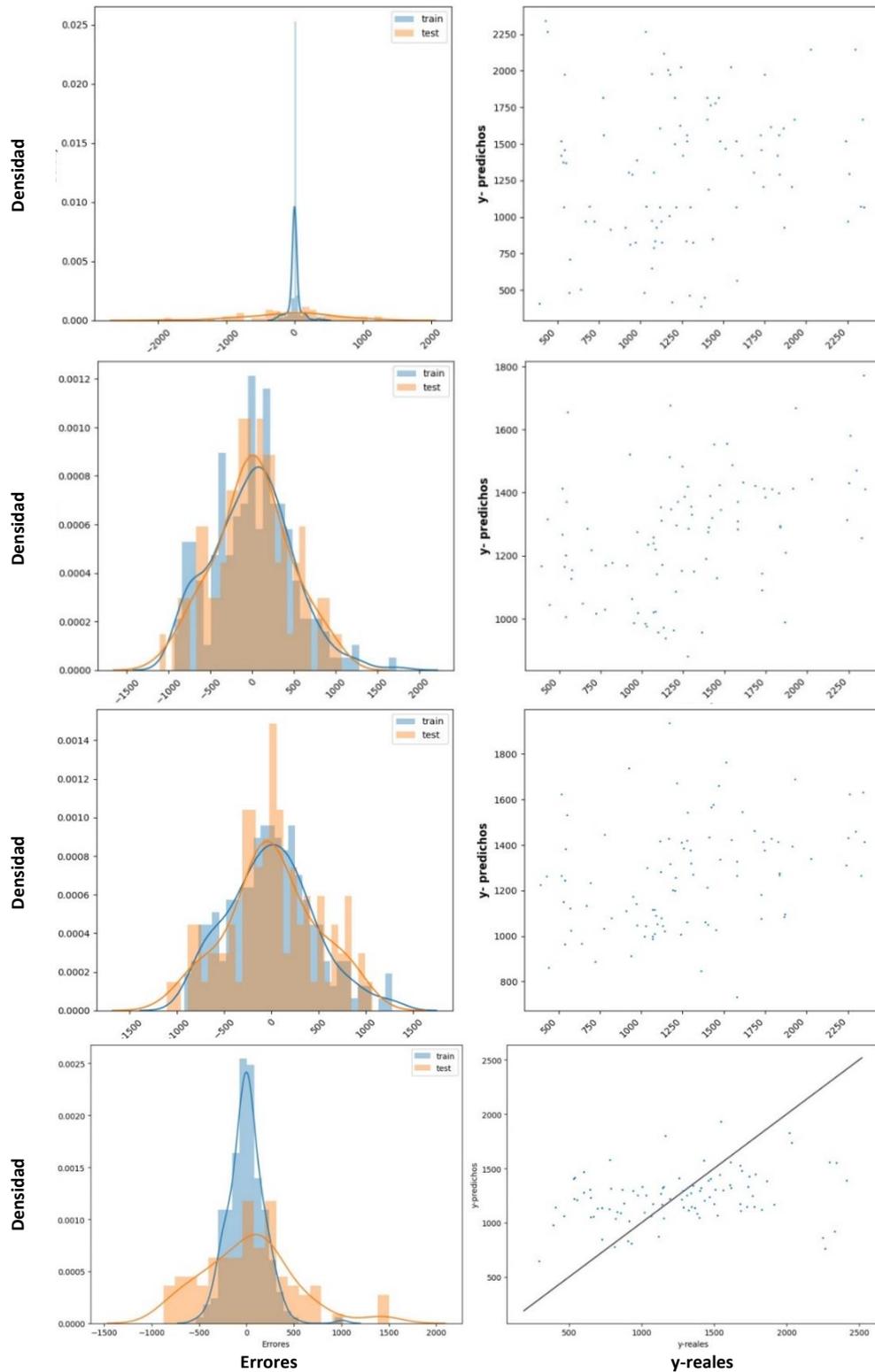


Figura 18. Gráficos de distribución de los errores (izquierda) y gráficos de dispersión de “y-reales” vs “y-predichos” (derecha) para los modelos árbol de decisión, regresión lineal, regresión polinómica de segundo grado y Random Forest (de arriba hacia abajo) para el trimestre junio-julio-agosto.

Los gráficos de distribución de los errores y los gráficos de dispersión de valores predichos versus reales (Figura 18) reflejan lo antedicho. El análisis de estos es análogo al del trimestre anterior, aunque puede observarse una mayor tendencia al sobreajuste tanto en árbol de decisión como en Random Forest y una mayor dispersión de valores predichos versus reales cuando se compara este trimestre con el anterior.

## **2) Análisis mensual del desempeño del modelo Random Forest**

La validación del modelo Random Forest en campo evidenció cambios en su capacidad de predicción del stock de pasto a lo largo del tiempo. Tal como se observa en los gráficos de dispersión de datos predichos vs datos reales, la nube de puntos se hace progresivamente más compacta y en el último mes se aproxima a la esperada recta de relación 1:1, lo que indica un aumento en la capacidad del modelo para capturar la dinámica del sistema. Esta mejora podría estar asociada a una mayor disponibilidad de datos y a un mejor entrenamiento del modelo con el paso del tiempo (Liaw y Wiener, 2002).

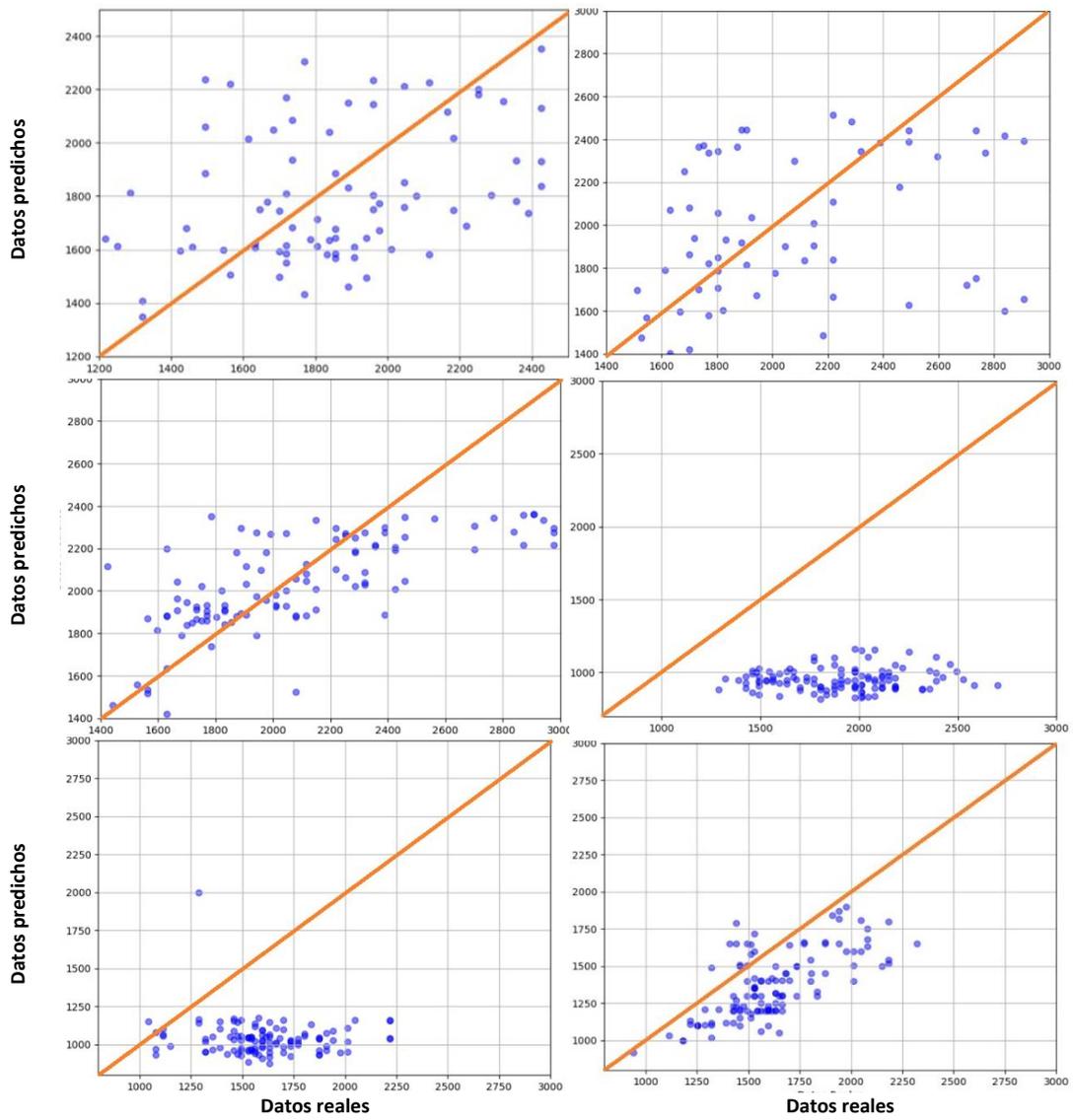


Figura 19. Gráficos de dispersión de datos reales vs datos predichos para el modelo Random Forest para los seis meses evaluados (de izquierda a derecha y de arriba hacia abajo: marzo, abril, mayo, junio, julio y agosto).

## Discusión

La aplicación de modelos de predicción en prácticas agrícolas de precisión podría mejorar significativamente la toma de decisiones relacionadas con el manejo de pasturas y la optimización de recursos en sistemas ganaderos. Los índices de vegetación como NDVI, SAVI, VV y VH demostraron ser herramientas robustas para el monitoreo del estado de la vegetación, ofreciendo una alternativa eficiente y no invasiva para medir el stock de pasto (Reinermann *et al.*, 2021). En nuestro estudio, los modelos más simples como el árbol de decisión y la regresión lineal y polinómica mostraron limitaciones en cuanto a la capacidad de generalización, lo que subraya la ventaja de técnicas más avanzadas como Random Forest (Breiman, 2001).

Si bien aún queda un amplio margen para mejorar y se requieren más datos para fortalecer las predicciones, se observó una tendencia positiva en el desempeño del modelo Random Forest. A medida que se incorporaron nuevos datos mes a mes, la precisión del modelo fue mejorando progresivamente, lo que sugiere que su capacidad para capturar relaciones complejas entre variables se incrementa con el tiempo. Esta evolución gradual refuerza el potencial de Random Forest como una herramienta valiosa para la predicción del stock de pasto en contextos agrícolas dinámicos.

La baja capacidad predictiva que mostró durante los primeros meses de trabajo podría deberse al fuerte peso que aún representaban los datos de entrenamiento de General Belgrano. Una de las causas de esta disparidad podría ser la diferencia en las características climáticas entre esa zona y la de General Lamadrid. Por un lado, la cantidad y distribución de las precipitaciones difieren notablemente, siendo más elevadas en General Belgrano, cuyos valores se mantienen sistemáticamente por encima de los de General Lamadrid durante todo el año (Weather Spark, 2016a y b). Esta mayor disponibilidad de agua favorece un crecimiento más sostenido de las pasturas, resultando en una mayor producción de materia seca y, por lo tanto, en datos de rendimiento más altos y estables.

Por otro lado, otro factor que podría explicar la variación en los resultados es la diferencia en las temperaturas medias mínimas y máximas. General Lamadrid presenta inviernos más fríos y veranos menos calurosos, lo que puede restringir notablemente el crecimiento de las pasturas durante los meses fríos. Además, aunque los veranos no son necesariamente más calurosos que en otras zonas, la combinación de temperaturas elevadas y posibles déficits hídricos puede generar estrés térmico e hídrico en la pastura. Por el contrario, el clima más templado y húmedo de General Belgrano, con inviernos más suaves y precipitaciones más regulares permite un desarrollo vegetativo más sostenido y menos afectado por condiciones climáticas extremas (Weather Spark, 2016a y b). Esto favorece una mayor productividad de las pasturas durante todo el año.

En conjunto, estas condiciones contribuyen a generar diferencias en la tasa de crecimiento de las pasturas y, por ende, ocasionan diferencias en las variables estructurales de la misma. Esto último impacta sobre la relación existente entre disponibilidad de pasto y altura registrada por pasturómetro de disco. Puede sugerirse, entonces, que determinar la relación existente entre disponibilidad de pasto y altura de pasturómetro para cada región y momento del año en particular contribuiría enormemente a mejorar la capacidad predictiva de este modelo.

## Conclusión

En base a los resultados obtenidos se puede concluir que el presente trabajo cumplió con los objetivos planteados inicialmente, demostrando que los algoritmos de aprendizaje supervisado son herramientas efectivas para predecir el stock de pasturas consociadas.

Respecto al ajuste y mejora de los modelos se constató que, a pesar de algunas diferencias entre los datos de entrenamiento y prueba, las técnicas avanzadas como Random Forest superaron las limitaciones de modelos más simples, como la regresión lineal, la regresión polinómica y el árbol de decisión. Estos ajustes fueron claves para optimizar los resultados de predicción.

La comparación entre predicciones y mediciones reales evidenció una baja capacidad predictiva en el inicio los dos trimestres analizados. No obstante, frente al entrenamiento progresivo con información local, el modelo Random Forest mostró una clara tendencia positiva en su precisión. También se identificaron variabilidades asociadas a factores climáticos locales que sugieren áreas de mejora.

Para finalizar dicho trabajo, resta avanzar en la promoción de la adopción de estos modelos entre los productores locales, lo que representaría un paso clave para mejorar la gestión del pasto y la toma de decisiones en la producción ganadera de la región. Asimismo, es fundamental continuar con la recolección sistemática de datos de stock de pasto, lo que permitirá seguir ajustando y afinando la precisión de estos modelos de predicción, contribuyendo así a la reducción progresiva del error. La implementación de estas herramientas no solo permitirá optimizar el uso de los recursos naturales, sino que también contribuirá a la sostenibilidad y eficiencia de los sistemas productivos, promoviendo un manejo más preciso y adaptado a las condiciones específicas de cada zona. De este modo, se cerrará un ciclo que no solo pondrá a disposición de los productores tecnologías avanzadas, sino que también los fortalecerá para tomar decisiones estratégicas y fundamentadas en el manejo de las pasturas.

## Bibliografía

- ArcGIS Pro. 2023. <https://pro.arcgis.com/es/pro-app/latest/help/analysis/image-analyst/introduction-to-synthetic-aperture-radar.htm>
- Blanco, P.A. y Caparrini, F.S. 2017. Inducción de árboles de decisión. [https://arxiv.org/abs/1708.05563?utm\\_source](https://arxiv.org/abs/1708.05563?utm_source)
- Breiman, L., Friedman, J., Olshen, R.A., & Stone, C.J. (1984). Classification and Regression Trees (1st ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781315139470>
- Breiman, L. 2001. Random forests. *Machine learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>
- Brital, A. 2021. Algoritmo de bosque aleatorio explicado. <https://anasbrital98.github.io/blog/2021/Random-Forest/>
- Copernicus. 2018. Fundamentos de teledetección radar. <https://www.ign.es/web/resources/docs/IGNCnig/actividades/OBS/Programa Marco Copernicus User Uptake/5 Fundamentos de teledeteccion radar.pdf>
- Hastie, T., Tibshirani, R. y Friedman, J. 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. Springer. <https://doi.org/10.1007/978-0-387-84858-7>
- Hostgator. 2024. Google Colab. <https://www.hostgator.mx/blog/google-colab/>
- Liaw, A., y Wiener, M. 2001. Classification and regression by RandomForest. *R news*, 2(3), 18-22. [https://www.researchgate.net/publication/228451484\\_Classification\\_and\\_Regression\\_by\\_RandomForest](https://www.researchgate.net/publication/228451484_Classification_and_Regression_by_RandomForest)
- Marini, F., D'Amico, M.B., Calandrini, G.L., Renzi, J.P., y Chantre, G.R. 2025. Utilización de índices de vegetación satelitales para predecir niveles de rendimiento de *Vicia villosa* Roth. *SADIO Electronic Journal of Informatics and Operations Research*, 24(2), e076. <https://doi.org/10.24215/15146774e076>
- Mastersindatascience.org. 2025. Introduction to Machine Learning Algorithms. <https://www.mastersindatascience.org/learning/machine-learning-algorithms/>
- Perilla, G.A. y Mas, J.F. 2020. Google Earth Engine – GEE: una poderosa herramienta que vincula el potencial de los datos masivos y la eficacia del procesamiento en la nube. *Investigaciones geográficas*, (101), e59929. <https://doi.org/10.14350/rig.59929>
- Reinermann, S.; Asam, S. y Kuenzer, C. 2020. Remote Sensing of Grassland Production and Management—A Review. *Remote Sensing*, 12(12), 1949. <https://doi.org/10.3390/rs12121949>
- Rubio, R., Dalla Valle, D., Calafatic, G., Inza, C. y Wade, M. 2001. Evaluación de biomasa y crecimiento de pasturas en el SE de la prov. de Bs. As. Trabajo financiado por M.S.G.P de Magnasco a iniciativa de la comisión de lechería de la AACREA.
- Segarra, J.; Buchailot, M.L.; Araus, J.L.; Kefauver, S.C. 2020. Remote Sensing for Precision Agriculture: Sentinel-2 Improved Features and Applications. *Agronomy*, 10, 641. <https://doi.org/10.3390/agronomy10050641>
- Sentinel-2. 2017. Tasking and catalog. Collections. <https://docs.up42.com/data/datasets/sentinel-2>
- Quintero, J. y Sienna, I. 2021. Módulo GeoAnalytics. Diplomatura en Agroanalytics. Universidad Austral.
- Weather Spark. 2016a. El clima y el tiempo promedio en todo el año en General Lamadrid. <https://es.weatherspark.com/y/28565/Clima-promedio-en-General-La-Madrid-Argentina-durante-todo-el-a%C3%B1o>
- Weather Spark. 2016b. El clima y el tiempo promedio en todo el año en General Belgrano. <https://es.weatherspark.com/y/28956/Clima-promedio-en-General-Belgrano-Argentina-durante-todo-el-a%C3%B1o>