

Proyección del índice de vegetación de diferencia normalizada satelital de cebada en crecimiento con series temporales e inteligencia artificial

Autor:

Dr. Ing. Agr. Adrián Lapaz Olveira

Director:

Esp. Lic. Maria Carina Roldán (FIUBA)

Codirector:

Esp. Ing. Ariadna Garmendia (FIUBA)

${\rm \acute{I}ndice}$

1. Des	scripción técnica-conceptual del proyecto a realizar	5
1.1	Antecedentes	5
1.2	Metodología	6
	1.2.1 Datos agrometeorológicos	8
	1.2.2 Datos satelitales	8
	1.2.3 Función β	9
	1.2.4 NDVI potencial	9
	v	10
		11
1.3	Referencias	13
2. Ide	ntificación y análisis de los interesados	15
3. Pro	pósito del proyecto	15
4. Alc	ance del proyecto	16
5. Sup	ouestos del proyecto	17
6. Rec	querimientos	17
7. His	torias de usuarios ($Product\ backlog$)	18
8. Ent	regables principales del proyecto	18
9. Des	glose del trabajo en tareas	19
10. Di	agrama de Activity On Node	21
11. Di	agrama de Gantt	22
12. Pr	resupuesto detallado del proyecto	22
13. Ge	estión de riesgos	22
14. Ge	estión de la calidad	25
15. Pr	rocesos de cierre	27



Registros de cambios

Revisión	Detalles de los cambios realizados	Fecha
0	Creación del documento	4 de marzo de 2025
1	Se completa hasta el punto 5 inclusive	20 de marzo de 2025
2	Se completa hasta el punto 5 inclusive	28 de marzo de 2025
3	Se completa hasta el punto 12 inclusive	1 de abril de 2025



Acta de constitución del proyecto

Buenos Aires, 4 de marzo de 2025

Por medio de la presente se acuerda con el Dr. Ing. Agr. Adrián Lapaz Olveira que su Trabajo Final de la Carrera de Especialización en Inteligencia Artificial se titulará "Proyección del índice de vegetación de diferencia normalizada satelital de cebada en crecimiento con series temporales e inteligencia artificial" y consistirá en la predicción temprana del índice de vegetación de diferencia normalizada máximo de los cultivos de cebada, a través del uso series temporales de datos Sentinel-2 y modelos de inteligencia artificial para las campañas 2023-2025. El trabajo tendrá un presupuesto preliminar estimado de 600 horas y está financiado por un proyecto de articulación academia - sector Productivo de ANNI (INIA - Ambev, UY), con fecha de inicio el 4 de marzo de 2025 y fecha de presentación pública en agosto de 2025.

Se adjunta a esta acta la planificación inicial.

Dr. Ing. Ariel Lutenberg Director posgrado FIUBA Ph.D. Ing. Agr. Andrés Berger INIA

Esp. Lic. Maria Carina Roldán Director del Trabajo Final



1. Descripción técnica-conceptual del proyecto a realizar

1.1. Antecedentes

La producción de cebada (Hordeum vulgare L.) con destino industrial tiene como objetivo maximizar el rendimiento con una concentración de proteína en grano del 9,5 al 12,5 % (MacLeod s.f.; Hu et al. 2021; Reussi Calvo et al. 2022). Para alcanzar este objetivo, es esencial realizar un diagnóstico temprano y preciso de los nutrientes para garantizar un crecimiento sin restricciones a lo largo de todo el ciclo del cultivo, sin sub o sobrefertilización (Baethgen y Christianson 1995; Araújo et al. 2022). No obstante, la principal dificultad para alcanzar este objetivo es la falta de precisión en el diagnóstico actual, ya que, al estar basado en determinaciones a la siembra, emergencia y/o estadíos iniciales del cultivo, este no contempla cómo los factores climáticos afectan el crecimiento de las plantas, y en consecuencia, cambian el rendimiento potencial y la demanda nutricional futura (Pettersson y Eckersten 2007; Cammarano et al. 2024; Halstead et al. 2022). Por lo tanto, el diagnóstico nutricional actual de cebada demanda herramientas complementarias que permitan proyectar el crecimiento potencial del cultivo (Bertholdsson 1999; Reussi Calvo et al. 2020; Berger et al. 2024).

Una de las herramientas calibradas para proyectar el crecimiento de los cultivos es la función β (Yin et al. 2003; Beck et al. 2006). Al modelar la acumulación de biomasa según los días desde la siembra (DDS), sus parámetros son biológicamente interpretables, lo que facilita su adopción en modelos de simulación de cultivos (Reussi Calvo et al. 2022). Sin embargo, su parametrización demanda determinaciones de biomasa durante el desarrollo fenológico, lo cual representa una desventaja para su adopción por parte de los productores y asesores (Bending et al. 2014). La determinación de la biomasa consiste en colectar plantas a campo para luego secarlas en laboratorio; por consiguiente, su procedimiento demanda mucho tiempo, es costoso y tiene limitada representatividad a escala de chacra (Berger et al. 2013). Por lo tanto, el uso de indicadores relacionados con la biomasa permitiría proyectar el crecimiento del cultivo a través del ajuste de una función β y, en consecuencia, predecir la demanda nutricional futura (Yoosefzadeh-Najafabadi et al. 2021; Paruelo et al. 2024; Guszcza 2008).

Para llevar a cabo este ajuste de la función β y estimar sus parámetros a partir de observaciones iniciales de NDVI, es especialmente adecuado utilizar la inferencia bayesiana (Yin et al. 2003; Paruelo et al. 2024). Bajo este enfoque, se parte de un conocimiento inicial acerca de los parámetros representado por una distribución previa (prior) y se actualiza dicha información conforme se obtienen más mediciones de NDVI a lo largo del ciclo de cultivo (Guszcza 2008; Bürkner 2018; Martin 2024). De esta manera, la distribución posterior de los parámetros de



la función β (como los descritos originalmente por Yin et al., 2003) reduce su dispersión, lo que aumenta la precisión en la proyección del NDVI. Este proceso no solo aprovecha los datos iniciales para predecir el crecimiento potencial, sino que también permite incorporar ajustes a medida que surgen nuevas observaciones de NDVI (Guszcza 2008; Bürkner 2018; Martin 2024). Por lo tanto, la precisión del enfoque bayesiano para predecir el NDVI potencial del cultivo aumenta con el ciclo fenológico, y en efecto, también la precisión del diagnóstico nutricional (Paruelo et al. 2024; Guszcza 2008; Bürkner 2018; Martin 2024).

En la actualidad, la inteligencia artificial es utilizada exitosamente para monitorear diversos parámetros biofísicos y químicos en cultivos y pasturas (Yoosefzadeh-Najafabadi et al. 2021; Farbo et al., 2025). Estudios previos han utilizado eficazmente datos históricos de NDVI, temperaturas y precipitación para predecir valores futuros del índice (Cavalli et al., 2023). En esta línea de investigación, un estudio reportó que las Redes Neuronales Artificiales (ANN) predicen con precisión el NDVI futuro de maíz (Zea mays L.) en el corto plazo (5–15 días) (Farbo et al., 2025). En este estudio, las ANN fueron entrenadas con el NDVI, índice de diferencia de agua normalizado (NDWI), precipitaciones acumuladas (GPM) y grados días acumulados (AGDD). Cabe mencionar que, además de estas variables agroclimáticas, existen otras que determinan el crecimiento de la cebada, como el coeficiente fototermal (Q) y fotoperíodo (P) (Hu et al., 2003; Reussi Calvo et al., 2022). También son determinantes las variables de manejo como el cultivar, fecha de siembra y dosis de nitrógeno (N) (Baethgen et al., 1995; Araújo et al., 2022). Por otro lado, en pasturas de campo natural se ha demostrado que el NDVI puede predecirse en el largo plazo (1-3 meses) (Paruelo et al., 2024). Esto se ha logrado a través del desarrollo de modelos híbridos, que proyectan la producción primaria neta con modelos paramétricos sigmoideos, similares a la función β (Yin et al. 2003), y usan algoritmos de ANN para estimar los errores de esta proyección (Martin 2024). Por lo tanto, los modelos híbridos que estimen mediante ANN los errores de la proyección bayesiana de la función β incrementarán la exactitud predictiva del NDVI potencial y optimizarán el diagnóstico nutricional del cultivo (Paruelo et al. 2024; Martin, 2024).

A pesar de que se ha demostrado que el NDVI es un fiel indicador del crecimiento de las plantas, su proyección durante el ciclo del cultivo aún no ha sido investigada en profundidad, lo cual limita la precisión del diagnóstico temprano. Esta falta de investigación representa una oportunidad para mejorar la capacidad de predecir el desarrollo de los cultivos y optimizar su manejo. Por lo tanto, este trabajo tiene como objetivo i) predecir el NDVI potencial del cultivo a partir de observaciones realizadas desde la siembra hasta los 45, 60 y 75 DDS, y ii) evaluar si las variables agroclimáticas y de manejo contribuyen a mejorar la precisión en la estimación del índice (Bayes vs modelo híbrido).

1.2. Metodología

La figura 1 representa el flujo de trabajo desarrollado para predecir el NDVI potencial en lotes comerciales de cebada maltera durante las campañas 2023 y 2024. El procedimiento se inicia con la recopilación de información agronómica y espacial, representada mediante polígonos georreferenciados (formato JSON), que incluyen la fecha de siembra, el cultivar seleccionado y la dosis de N aplicada (kg ha⁻¹). Esta información sirve como base para extraer, de forma concurrente, datos meteorológicos y satelitales.

El conjunto de datos se dividirá de acuerdo con su objetivo analítico: los registros correspondientes a la campaña 2023 se destinarán a la calibración del modelo, mientras que los de la campaña 2024 se utilizarán para su validación. A su vez, cada conjunto se truncará en tres momentos clave del ciclo del cultivo: 45, 60 y 75 días después de la siembra (DDS), incorporando también



variables previas al evento de siembra (hasta -5 DDS). De este modo, se simularán diferentes escenarios de disponibilidad temporal de información para evaluar la capacidad predictiva del modelo.

La primera etapa del modelado consistirá en el ajuste de una función de crecimiento β a las observaciones de NDVI en función de DDS. Esta función permite capturar la dinámica estacional del desarrollo del cultivo mediante parámetros con interpretación biológica como w_{max} (ver sección 1.2.3.). A partir de este ajuste se obtendrán las predicciones iniciales del NDVI potencial para cada horizonte temporal.

A continuación, se calcularán los residuales como la diferencia entre los valores observados y los estimados por el modelo sigmoideo. Estos residuales se utilizarán como variable respuesta para entrenar una ANN, alimentada con variables agroclimáticas y de manejo agrícola. La ANN se ajustará mediante validación cruzada, utilizando un $75\,\%$ de los datos para entrenamiento y el $25\,\%$ restante para validación, y se aplicarán técnicas de regularización para reducir el riesgo de sobreajuste.

Finalmente, se integrarán el modelo paramétrico y la ANN en un esquema de modelado híbrido. Esta estrategia busca aprovechar la capacidad explicativa de la curva de crecimiento, basada en los principios de la dinámica vegetal, junto con la flexibilidad de las ANN para capturar respuestas no lineales del cultivo frente a condiciones ambientales variables. El modelo híbrido será evaluado utilizando los datos de la campaña 2024 en los tres horizontes definidos, y su precisión se medirá a partir de la diferencia entre el NDVI potencial proyectado y el observado.

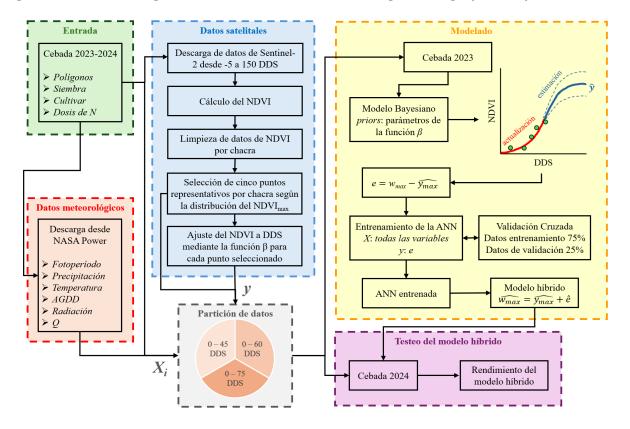


Figura 1. Diagrama del flujo de trabajo. N: nitrógeno; Q: coeficiente fototermal; DDS: días desde la siembra; NDVI: índice de vegetación de diferencia normalizada; NDVI $_{máx}$: valor máximo del NDVI; priors: distribuciones a priori; $w_{máx}$: parámetro de la función β que representa la máxima acumulación de biomasa; $\hat{y}_{máx}$: valor máximo estimado mediante proyección bayesiana; e: error; ANN: red neuronal artificial.



1.2.1. Datos agrometeorológicos

Dado el amplio alcance geográfico y la necesidad de contar con datos bianuales, los datos agrometeorológicos serán obtenidos a través de la API RESTful de NASA POWER, utilizando el cliente PyNASApower (en lenguaje Python, versión 3.13). El punto de consulta será el centroide de cada lote. Las variables diarias que se recopilarán incluirán precipitación (PP), temperatura (T), fotoperíodo (P) y radiación (R). A partir de estas variables, se calcularán para cada día (i-ésimo) hasta el último día disponible (n), la precipitación global acumulada (GPM), los grados día de crecimiento acumulados (AGDD) y el coeficiente fototermal (Q).

$$GPM = \sum_{i=1}^{n} PP_i \tag{1}$$

$$AGDD = \sum_{i=1}^{n} GDD_i \tag{2}$$

$$Q_i = \frac{R_i}{T_i - T_{\text{base}}} \tag{3}$$

1.2.2. Datos satelitales

Para procesar eficientemente los datos de reflectancia superficial corregidos geométrica y radiométricamente en una amplia área geográfica durante las campañas de cebada 2023–2024, se accederá al conjunto de datos Harmonized Sentinel-2 MSI Level-2A, disponible a través del Earth Engine Data Catalog, utilizando la API de Python de Earth Engine (versión 3.13 de Python). Este conjunto de datos ofrecerá una resolución espacial de 10 metros de pixel, una resolución radiométrica de 12 bits y una precisión posicional de aproximadamente 3 metros. Dentro de cada lote, se calculará el NDVI como la diferencia normalizada entre la reflectancia (ρ) en el infrarrojo cercano (833 nm, \pm 106 nm, pixel 10 m) y en el rojo (665 nm, \pm 31 nm, pixel 10 m) (ESA, 2025).

$$NDVI = \frac{\rho_{\text{infrarrojo}} - \rho_{\text{rojo}}}{\rho_{\text{infrarrojo}} + \rho_{\text{rojo}}}$$
(4)

Una vez obtenida la serie temporal de NDVI para cada píxel, se calcularán los valores mínimos y máximos de NDVI correspondientes a cada píxel del lote durante el período de crecimiento del cultivo. Este procedimiento permitirá filtrar aquellos píxeles que no representen al cultivo de interés. Para ello, se eliminarán los píxeles cuyos valores máximos de NDVI no superen el umbral de 0,5, como ocurre en superficies de suelo desnudo, donde los valores suelen ser consistentemente bajos. También se descartarán aquellos píxeles cuyos valores mínimos no sean inferiores a 0,4, ya que en coberturas permanentes, como montes o vegetación densa, el NDVI tiende a mantenerse por encima de ese valor. Luego, se excluirán los registros que presenten una disminución superior al 10 % respecto al valor registrado en la fecha anterior. Este tipo de caída abrupta podría estar relacionada con condiciones atmosféricas adversas, como nubosidad, o con eventos antrópicos puntuales, como el pastoreo o la recolección de plantas para heno. Una vez aplicados estos filtros, se obtendrá una serie de valores de NDVI depurados, adecuada para su posterior análisis estadístico.



1.2.3. Función β

Dado que el comportamiento del NDVI a lo largo del ciclo de crecimiento del cultivo presenta una dinámica similar a la de las plantas, se ajustará la función β dado que modela con precisión la acumulación de biomasa en función de los DDS (Yin et al., 2003). A diferencia de otros modelos sigmoides, esta función incorpora parámetros biológicamente interpretables, como se muestra en la figura 2. Además, su estructura matemática permite modelar de manera continua la evolución de la biomasa sin generar discontinuidades en la transición entre fases de crecimiento, lo que proporcionará estimaciones más realistas y ajustadas a la naturaleza determinante del crecimiento de estos cultivos. Por lo tanto, la parametrización de la dinámica del NDVI mediante la función β ofrece una estrategia promisoria para la proyección temprana de la biomasa acumulada en cebada.

La significancia de estos parámetros radica en su capacidad para representar con alta fidelidad la trayectoria sigmoidal inherente al proceso de crecimiento. Concretamente, posibilitan la discriminación cuantitativa de la fase inicial de desarrollo, caracterizada por una tasa de incremento gradual; el período intermedio de expansión acelerada, donde la acumulación de biomasa exhibe su máxima velocidad; y la fase final de atenuación, marcada por una progresiva estabilización en los valores de la variable de crecimiento.

La ecuación para calcular estos parámetros mediante mínimos cuadrados no lineales (nls) es la siguiente

$$w = w_b + (w_{\text{máx}} - w_b) \left(1 + \frac{t_e - t}{t_e - t_m} \right) \left(\frac{t - t_b}{t_e - t_b} \right)^{\frac{t_m - t_b}{t_e - t_m}}$$
(5)

sujeto a la condición $t_b \leq t_m < t_e$.

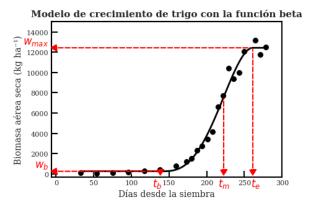


Figura 2. Dinámica temporal observada (puntos) y ajustada mediante la función de crecimiento beta (curva) de la biomasa aérea seca en trigo de invierno (datos de Gregory et al., 1978 citado por Yin et al., 2003). Los parámetros de la función beta son los siguientes, w_b : biomasa inicial; $w_{\text{máx}}$: biomasa máxima alcanzada; t_b : tiempo de inicio del crecimiento significativo; t_m : punto de inflexión (tasa de crecimiento máxima); t_e : tiempo en el que finaliza el crecimiento.

1.2.4. NDVI potencial

Dada la estrecha relación entre la dinámica del crecimiento del cultivo y la evolución del NDVI, se considerará al NDVI potencial como la variable más adecuada para realizar predicciones



relevantes en el diagnóstico nutricional de la cebada. Este análisis se basará en series temporales de datos NDVI que abarcarán la totalidad del ciclo fenológico del cultivo durante las campañas 2023 y 2024.

Una vez obtenida la serie temporal de NDVI para cada píxel, se calculará el valor máximo de NDVI correspondiente. Esto permitirá generar un mapa de NDVI máximo con una resolución espacial de 10 metros. A partir de dicho mapa, se analizará la distribución de los valores con el objetivo de identificar las cinco ubicaciones más representativas dentro de cada lote. Estas corresponderán al valor mínimo, al valor máximo y a los percentiles 25, 50 y 75 de la distribución del NDVI máximo.

Este enfoque permitirá capturar de manera robusta la variabilidad y los contrastes espaciales del NDVI máximo a escala de lote. Posteriormente, sobre estos cinco puntos seleccionados en cada lote, se trabajará con la serie temporal de NDVI en función de los DDS, con el propósito de estimar, mediante mínimos cuadrados lineales, los parámetros de la función β . En este contexto, el parámetro $w_{\rm max}$ se interpretará como el NDVI potencial de cada píxel seleccionado.

1.2.5. Modelo Bayesiano

Para estimar el NDVI potencial a partir de observaciones parciales de la serie temporal del cultivo, se implementará un modelo jerárquico bayesiano no lineal, estructurado en tres niveles: observacional, de proceso y de parámetros. Este modelo se basa en la función de crecimiento tipo β , utilizada previamente en ecología y ciencia actuarial para representar trayectorias sigmoideas (Clark, 2003; Guszcza, 2008).

Nivel observacional. El valor de NDVI observado en el tiempo t para el píxel i se modela como una variable aleatoria con error gaussiano:

$$NDVI_{i,t} \sim \mathcal{N}(\hat{w}_{i,t}, \sigma^2)$$
 (6)

donde $\hat{w}_{i,t}$ representa el valor estimado por la función de crecimiento beta.

Nivel de proceso. La evolución temporal de $\hat{w}_{i,t}$ se describe mediante la función β :

$$\hat{w}_{i,t} = w_b + (w_{\text{máx},i} - w_b) \left(1 + \frac{t_e - t}{t_e - t_m} \right) \left(\frac{t - t_b}{t_e - t_b} \right)^{\frac{t_m - t_b}{t_e - t_m}}$$
(7)

sujeta a la condición asintótica que refleja la estabilización del dosel:

$$w(t) = \begin{cases} \hat{w}_{i,t} & \text{si } t \le t_e \\ w_{\text{máx}} & \text{si } t > t_e \end{cases}$$
 (8)

Esta formulación permite capturar las fases inicial, intermedia y final del crecimiento con parámetros interpretables biológicamente: t_b (inicio del crecimiento activo), t_m (punto de máxima tasa de crecimiento) y t_e (fin del crecimiento activo). El valor inicial de NDVI se fija en



 $w_b = 0.1$, estimado empíricamente a partir de valores de NDVI de suelo desnudo o emergencia temprana.

Nivel de parámetros. Las distribuciones a priori de los parámetros fueron derivadas empíricamente a partir de los ajustes de la función β sobre los datos completos de NDVI correspondientes a la campaña 2023. Esta estrategia permite definir *priors* informativos basados en la variabilidad real observada en lotes representativos, evitando tanto supuestos arbitrarios como la necesidad de elicitar conocimiento experto.

Las distribuciones ajustadas para cada parámetro fueron:

$$\begin{split} w_{\text{máx},i} &\sim \mathcal{N}(\mu_{w_{\text{máx}}}, \sigma_{w_{\text{máx}}}) \\ t_b &\sim \mathcal{N}(\mu_{t_b}, \sigma_{t_b}) \\ t_m &\sim \mathcal{N}(\mu_{t_m}, \sigma_{t_m}) \\ t_e &\sim \mathcal{N}(\mu_{t_e}, \sigma_{t_e}) \\ \log(\sigma) &\sim \mathcal{N}(\log(\hat{\sigma}), 1) \end{split}$$

donde cada par (μ, σ) se obtiene a partir de los estimadores de máxima verosimilitud obtenidos por mínimos cuadrados no lineales sobre la campaña calibrada (2023). Este enfoque ofrece dos ventajas clave: permite capturar la estructura real de crecimiento del NDVI bajo condiciones locales, y garantiza una coherencia probabilística entre el conjunto calibrado y el conjunto proyectado (2024). El uso de priors basados en datos previos ha sido documentado extensamente como práctica robusta en modelos bayesianos jerárquicos (Martin, 2024; Schneider et al., 2006).

Implementación computacional. El modelo se implementará en Python utilizando la librería 'PyMC' (v4.1). La inferencia se realizará mediante el algoritmo Hamiltonian Monte Carlo (NUTS), el cual permite una exploración eficiente del espacio posterior incluso en modelos no lineales de alta dimensión (Carpenter et al., 2017). Los criterios de convergencia incluirán:

- $\hat{R} < 1.05$ para todos los parámetros
- Tamaño efectivo de muestra > 1000
- Inspección visual de las trazas MCMC y densidades posteriores

Este enfoque ha sido validado en estudios previos de modelado jerárquico de crecimiento en ecología vegetal (Schneider et al., 2006) y en reservorios de seguros (Guszcza, 2008), y constituye un marco robusto para la proyección del NDVI potencial en series temporales truncadas.

1.2.6. Modelo híbrido con redes neuronales artificiales

Con el objetivo de incrementar la precisión en la predicción del NDVI potencial, se adoptará una estrategia de modelado híbrido que combina el ajuste paramétrico bayesiano de la función β con ANN. Este enfoque permite modelar de forma explícita la trayectoria sigmoidea del crecimiento del cultivo —mediante un modelo con parámetros biológicamente interpretables—



y, simultáneamente, corregir de manera flexible las no linealidades residuales que este modelo no captura. Tal como fue demostrado por Paruelo et al. (2024), este tipo de modelos híbridos mejora sustancialmente la capacidad predictiva en sistemas vegetales de alta variabilidad ambiental.

La ANN se entrenará sobre los residuales del modelo bayesiano, definidos como la diferencia entre el NDVI observado y el proyectado por la función β . Esta red tendrá como insumos un conjunto multivariado de predictores agroclimáticos y de manejo agrícola, entre los cuales se incluyen:

- Precipitación diaria (PP), en mm día⁻¹.
- Temperatura del aire (T), en °C.
- Fotoperíodo (P), en horas (h).
- Radiación (R), en MJ m⁻² día⁻¹.
- Precipitación global acumulada (GPM), en mm.
- Grados día de crecimiento acumulados (AGDD), en °C día.
- Coeficiente fototermal (Q), en MJ m⁻² día⁻¹ °C⁻¹.
- Cultivar (variable categórica).
- Fecha de siembra, en día del año (DOY).
- Dosis de nitrógeno (N), en kg ha⁻¹.

La arquitectura base consistirá en un perceptrón multicapa (MLP) con función de activación ReLU y optimización mediante retropropagación estocástica (SGD). Para evitar el sobreajuste, se incorporarán técnicas de regularización como dropout (p=0.3) y penalización L2. El conjunto de datos será particionado utilizando validación cruzada estratificada (k-fold=4), reservando el 25 % para evaluación en cada iteración.

En paralelo, se evaluará una arquitectura alternativa de mayor capacidad secuencial: redes neuronales recurrentes del tipo LSTM bidireccionales. Este diseño resulta especialmente útil para modelar secuencias temporales con irregularidades en la adquisición de imágenes satelitales, como nubes o lagunas temporales, y ha demostrado alto desempeño en la predicción de NDVI a corto y mediano plazo en maíz (Farbo et al., 2024).

El entrenamiento de las redes se implementará en PyTorch (versión 2.x) bajo un paradigma imperativo y dinámico que facilita la depuración y trazabilidad del modelo (Paszke et al., 2019). Las métricas de desempeño incluirán el error cuadrático medio (RMSE), el coeficiente de determinación (R²) y el error absoluto medio (MAE), tanto para los datos de calibración como de validación.

Finalmente, la predicción del NDVI potencial para cada unidad espacial y temporal se calculará mediante la siguiente expresión funcional:

$$\hat{y}_{\text{híbrido}} = \hat{w}_{\text{máx}} + \hat{\varepsilon} \tag{9}$$

donde $\hat{w}_{\text{máx}}$ corresponde al valor estimado mediante el modelo bayesiano de la función β y $\hat{\varepsilon}$ es la corrección estimada por la red neuronal artificial. De este modo, el modelo híbrido



ofrece una combinación sinérgica de interpretabilidad biológica y capacidad de generalización no paramétrica, orientada a mejorar la precisión del diagnóstico nutricional temprano del cultivo.

1.3. Referencias

- 1. Araújo, Bruno O. N., Cáceres, Cintia, Barreto, Lorena y Farías, Pablo. (2022). Crecimiento y rendimiento fisiológico de plantas de cebada producidas bajo manejo de nitrógeno. *Ingeniería e Investigación*, 42(2), e200.
- 2. Baethgen, Walter E. y Christianson, Charles B. (1995). Nitrogen fertilizer effects on growth and yield of barley. *Field Crops Research*, 43(1), 87–99.
- 3. Beck, Pieter S. A., Atzberger, Clement, Høgda, Kjell A., Johansen, Bernt y Skidmore, Andrew K. (2006). Improved monitoring of vegetation dynamics at very high latitudes: A new method using MODIS NDVI. Remote Sensing of Environment, 100(3), 321–334.
- 4. Bending, Niall A., Atkinson, Peter M. y Jarman, Richard J. (2014). Estimating biomass of barley using crop surface models (CSMs) derived from UAV-based RGB imaging. *International Journal of Remote Sensing*, 35(9), 3446–3469.
- 5. Berger, Andrés, Gaso, Diego, Ciganda, Víctor y Otero, Alberto. (2013). Evaluación de la dinámica temporal de índices espectrales y su relación con variables biofísicas en trigo para estimación de rendimiento. X Simposio Brasileiro de Sensoriamento Remoto.
- 6. Berger, Andrés G., Repetto, Daniel, Mazzilli, Sebastián y Fernández, Gerardo. (2024). OPTIFERT-N: Nueva herramienta para optimizar la fertilización nitrogenada en trigo. Revista INIA, 77, 10–19.
- 7. Bertholdsson, Nils O. (1999). Characterization of malting barley cultivars with more or less stable grain protein content under varying environmental conditions. *European Journal of Agronomy*, 10(1), 1–8.
- 8. Bürkner, Paul-Christian. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, 10(1), 395–411.
- 9. Cammarano, Davide, Paruelo, José M., Reussi Calvo, Natalia I. y Sadras, Víctor O. (2024). Impact of nitrogen and water on barley grain yield and malting quality. *Journal of Soil Science and Plant Nutrition*, 24, 6718–6730.
- 10. Campbell, J.B., Wynne, R.H. y Thomas, V.A. (2022). Introduction to remote sensing (6.^a ed.). The Guilford Press, New York, USA. https://doi.org/10.1117/3.673407.ch1
- 11. Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1), 1–32. https://doi.org/10.18637/jss.v076.i01
- 12. Cavalli, S., Penzotti, G., Amoretti, M. y Caselli, S. (2023). A Machine Learning Approach for NDVI Forecasting based on Sentinel-2 Data. En 16th International Conference on Software Technologies, 473–480.
- 13. Clark, D. R. (2003). LDF curve-fitting and stochastic reserving: a maximum likelihood approach. En *CAS Forum* (Vol. 3, No. 4, pp. 41–92).



- 14. ESA. (2025). Sentinel Online ESA [sitio web]. Recuperado de https://sentinel.esa.int/web/sentinel/home [Consulta: 04 de abril de 2025].
- 15. Farbo, A., Sarvia, F., De Petris, S., Basile, V. y Borgogno-Mondino, E. (2024). Forecasting corn NDVI through AI-based approaches using sentinel 2 image time series. *ISPRS Journal of Photogrammetry and Remote Sensing*, 211, 244–261.
- 16. Guszcza, James. (2008). Hierarchical growth curve models for loss reserving. Casualty Actuarial Society Forum, Fall, 103–151.
- 17. Halstead, Morgan, Tavarez, Rosa, Thomas, Charles y Smith, Daniel. (2022). Barley grain protein is influenced by genotype, environment, and nitrogen management. *Crop Science*, 63(1), 115–127.
- 18. Hu, Yuntao, Barmeier, Gerhard y Schmidhalter, Urs. (2021). Genetic variation in grain yield and quality traits of spring malting barley. *Agronomy*, 11(6), 1177.
- Lapaz Olveira, A.M., Saínz Rozas, H.R., Castro-Franco, M., Carciochi, W.D., Nieto, L., Balzarini, M., Ciampitti, I. y Reussi Calvo, N.I. (2023). Monitoring Corn Nitrogen Concentration from Radar (C-SAR), Optical, and Sensor Satellite Data Fusion. Remote Sensing, 15. https://doi.org/10.3390/rs15030824
- 20. MacLeod, Aaron. (s.f.). *Understanding Malting Barley Quality*. Center for Craft Food and Beverage, Hartwick College. Recuperado de https://www.canr.msu.edu/uploads/234/78941/Understanding_Malting_Barley_Quality_-_Aaron_MacLeod.pdf [Consulta: 04 de abril de 2025].
- 21. Martín, Osvaldo. (2024). Bayesian Analysis with Python: A Practical Guide to Probabilistic Programming. Packt Publishing.
- 22. Paruelo, José M., Piñeiro, Germán, Grau, Ricardo y Alcaraz-Segura, David. (2024). Hybrid modeling for grassland productivity prediction: Combining mechanistic models and machine learning. *Ecological Indicators*, 155, 111372.
- 23. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., & Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. arXiv preprint arXiv:1912.01703.
- 24. Pettersson, Carl G. y Eckersten, Henrik. (2007). Prediction of grain protein in spring malting barley. *European Journal of Agronomy*, 27, 205–214.
- 25. Reussi Calvo, Natalia I., Carciochi, Walter D. y Prystupa, Pablo. (2022). Economic optimum nitrogen rate analysis for feed and malting barley. *Crop Science*, 62(4), 1997–2010.
- 26. Reussi Calvo, Natalia I., Sadras, Víctor O., Cerrudo, Aníbal, Monzón, Juan P. y Abbate, Pedro E. (2020). Canopy indices: A model to estimate the nitrogen rate for barley and wheat. *Journal of Soil Science and Plant Nutrition*, 20, 535–548.
- 27. Reussi Calvo, Natalia I., Sadras, Víctor O., Cerrudo, Aníbal, Monzón, Juan P. y Abbate, Pedro E. (2022). Economic optimum nitrogen rate analysis for feed and malting barley. *Crop Science*, 62(2), 937–952.
- 28. Schneider, M. K., Law, R., & Illian, J. B. (2006). Quantification of neighbourhood-dependent plant growth by Bayesian hierarchical modelling. *Journal of Ecology*, 310–321.



- 29. Yoosefzadeh-Najafabadi, Majid, Tulpan, Dan y Eskandari, Mostafa. (2021). Using hybrid artificial intelligence and evolutionary optimization algorithms for estimating soybean yield and fresh biomass using hyperspectral vegetation indices. *Remote Sensing*, 13(13), 2555.
- 30. Yin, Xinyou, Goudriaan, Jan, Lantinga, Egbert A., Vos, Jan y Spiertz, Hendrik J. (2003). A flexible sigmoid function of determinate growth. *Annals of Botany*, 91(3), 361–371.

2. Identificación y análisis de los interesados

Rol Nombre y Apellido Organización Puesto Cliente Ph.D. Ing. Agr. Andrés **INIA** Investigador Berger INIA Impulsor Ph.D. Ing. Agr. Andrés Investigador Berger Responsable Dr. Ing. Agr. Adrián **FIUBA** Alumno Lapaz Olveira Colaboradores PhD. MSc. Ing. Agr. **INIA** Investigador INIA José Paruelo Investigador Dr. Ing. Agr. Sebastian Mazzilli Orientadores Esp. Lic. Maria Carina **FIUBA** Director delTrabajo Roldán **FIUBA** Final Esp. Ing. Ariadna Gar-Co-Director del Trabamendia io Final

Cuadro 1. Identificación de los interesados

El cliente del proyecto será quien supervise la validez científica de la metodología y los resultados, ya que fue quien impulsó la idea original y cuenta con amplia experiencia en el manejo agronómico de cereales de invierno, simulación de cultivos y monitoreo satelital. A partir de esta base, el responsable se encargará de llevar adelante la investigación, cumpliendo con los requerimientos curriculares del proyecto, lo cual implicará redactar la tesis, desarrollar los códigos en Python y plantear e implementar los modelos de inteligencia artificial más adecuados. En paralelo, los colaboradores que son especialistas en cultivos y teledetección, acompañarán el proceso brindando seguimiento técnico y retroalimentación constante. Asimismo, los orientadores tendrán un rol clave en guiar al responsable tanto en la orientación académica como en el desarrollo técnico del proyecto, especialmente en lo referido a programación y teledetección, considerando que ni el cliente, ni los colaboradores, ni el responsable cuentan son profesionales del áerea informática. Por lo tanto, la complementariedad de conocimientos y el compromiso de los interesado asegura la viabilidad técnica y académica del proyecto.

3. Propósito del proyecto

Este proyecto busca desarrollar una herramienta complementaría al diagnóstico nutricional actual de cebada que permita estimar indicadores, como el NDVI, del crecimiento potencial de los cultivos. Esto permitiría a los productores desarrollar estrategias de manejos anticipadas,



reducir pérdidas en el rendimiento potencial del cultivo y alcanzar una concentración de proteína en grano óptima para la calidad maltera.

4. Alcance del proyecto

En este proyecto se incluye información sobre los sistemas productivos de cebada con destino industrial para la empresa AmBev. Dada las limitaciones en tiempo para realizar el proyecto dentro del contexto académico de la especialización, no se incluyen información de la producción de cebada 2025 y mediciones de biomasa realizadas en 2024 y 2025.

El proyecto incluye:

- Información de la producción de cebada 2023 y 2024 para miles de chacras.
 - Polígono de la chacra.
 - Fecha de siembra.
 - Cultivar.
 - Aplicación de fertilizantes.
- Datos agroclimáticos por lote.
 - Precipitación.
 - Temperatura diaria.
 - Fotoperíodo.
 - Grados días.
 - Radiación.
 - Coeficiente fototermal.
- Datos satelitales para cinco puntos representativos por chacra.
 - NDVI.
 - Parámetros de la cruva β .

El presente proyecto no incluye información sobre el corriente año (2025) y 60 observaciones realizadas en chacras que produjeron cebada en el 2024 y otras mediciones que se realizarán en el 2025, porque la dimensión del proyecto está acotada a la aplicación de la inteligencia artificial en datos de teledetección. No obstante, podrían incluirse los datos 2024 para ver si el NDVI estimado se correlaciona con las variables observadas que son:

- Observaciones en espigazón.
 - Biomasa aérea seca y fresca.
 - Concentración y acumulación de nitrógeno en la biomasa.
 - Índice de nutrición nitrogenada (INN).
 - Índice de diagnóstico hídrico (WDI).
- Observados en cosecha.



- Rendimiento en grano.
- Concentración y acumulación de nitrógeno en grano.
- Proteína en grano.
- Índice de cosecha.
- Calidad de grano.

5. Supuestos del proyecto

Para el desarrollo del presente proyecto se supone que:

- Se contará de la asesoría de expertos en la temática de agronomía, computación y percepción remota.
- El tiempo requerido para el entrenamiento de modelos será adecuado para cumplir con los requisitos curriculares de la especialización.
- La financiación del proyecto ya está cubierta por las instituciones del estudiante.
- Los resultados del proyecto fomentarán el uso de herramientas emergentes de teledetección en los sistemas productivos anuales.
- Este trabajo impulsará futuras investigaciones en otros cultivos.

Dado que la metodología es clara y sus resultados replicables, este trabajo impulsará futuras investigaciones similares.

6. Requerimientos

- 1. Requerimientos funcionales:
 - 1.1. El modelo de predicción debe poder estimar el NDVI potencial en lotes de producción comercial de cebada.
 - 1.2. La estimación del NDVI potencial debe estar entre 0,4 y 1,0.
 - 1.3. El usuario debe poder estimar el NDVI potencial trazando el poligono del lote y los datos de fecha de siembra, cultivar y fertilización nitrogenada.
- 2. Requerimientos de documentación:
 - 2.1. Requerimiento 1. Los resultados deben ser publicados en revistas de investigación locales e internacionales.
 - 2.2. Requerimiento 2. La tesis debe ser publicada de forma abierta al público general, de modo de poder comparlir los resultados en redes sociales y en correo electrónico entre colegas y asesores del sector.
- 3. Requerimiento de testing durante la campaña 2025. Si bien estarán fuera del período del proyecto, la estructura de código debe permitir la ejecución sobre los datos recolectados durante la campaña de cebeda del 2025.
- 4. Requerimientos de la interfaz deben de ser sencillo, fáciles de manipular. En general, cargar una excel con los datos requeridos de manejo y una columna para la estructura JSON.



7. Historias de usuarios (*Product backlog*)

A continuación se presentan ocho historias de usuario vinculadas a los desafíos actuales en la gestión de nitrógeno en cebada cervecera, con especial énfasis en la aplicación de tecnologías de teledetección, la toma de decisiones agronómicas y la heterogeneidad espacial en el lote. Cada historia considera un enfoque centrado en el usuario, sus necesidades concretas y el contexto de uso de una potencial solución tecnológica.

- 1. "Como asesor de productores que entregan a la maltería Ambev, necesito automatizar y personalizar las recomendaciones de fertilización con nitrógeno en múltiples chacras, porque la gran cantidad de lotes no me permite hacerlo de forma manual."
- 2. "Como profesional que asesora a productores, necesito estimar el crecimiento futuro del cultivo para clasificar zonas aptas para maltería dentro del lote, identificar aquellas que no conviene cosechar por baja calidad, y cuantificar la producción acumulada en cada una."
- 3. "Como investigador agronómico, quiero utilizar herramientas de teledetección para identificar zonas con diferente potencial productivo dentro del lote, con el fin de instalar ensayos representativos, pero no tengo experiencia en sistemas de información geográfica."
- 4. "Como investigadora en fisiología de cultivos, me gustaría contar con soluciones que permitan predecir deficiencias nutricionales futuras en lugar de solo monitorear el estado actual del cultivo, para así evitar pérdidas de rendimiento."
- 5. "Como asesor técnico de chacra, quiero disponer de una estimación anticipada y confiable de la dosis óptima de nitrógeno a aplicar, para asegurar niveles adecuados de proteína sin comprometer el rendimiento."
- 6. "Como productor de cebada, necesito una herramienta que me permita estimar la dosis de nitrógeno sin tener que realizar muestreos en la etapa de encañazón, ya que no dispongo de tiempo ni recursos logísticos para esa tarea."
- 7. "Como asesor técnico, realizo análisis de suelos para definir prácticas de manejo, pero a menudo los resultados no se traducen en buenas decisiones porque no capturan la variabilidad espacial ni los efectos del clima en la demanda de nutrientes del cultivo."
- 8. "Como productor de cebada, me resulta difícil ajustar la dosis de nitrógeno: si aplico en exceso, me rechazan camiones por alta proteína, y si aplico poco, no cubro los costos de producción. Necesito una herramienta diagnóstica más precisa y predictiva."

8. Entregables principales del proyecto

Los entregables del proyecto son:

- Manual de usuario.
- Diagrama de análisis con modelos híbridos.
- Código fuente del análisis.
- Diagrama de estructura de la red neuronal.



- Memoria del trabajo final.
- Resultado publicados en una revista con refarato.

9. Desglose del trabajo en tareas

A continuación, se presenta el desglose del trabajo en tareas (WBS) para cumplir con los requerimientos funcionales y no funcionales definidos en el proyecto. Las actividades están organizadas en grupos de tareas relacionados con los entregables clave del proyecto y alineadas con la metodología propuesta. Ninguna tarea excede las 40 h, en cumplimiento de las recomendaciones.

1. Grupo de tareas 1: Adquisición y procesamiento de datos satelitales y agrometeorológicos (130 h).

- 1.1. Obtención de polígonos de chacras, fechas de siembra, cultivares y dosis de nitrógeno (20 h).
- 1.2. Descarga y procesamiento de datos Sentinel-2 para cálculo de NDVI (30 h).
- 1.3. Cálculo de NDVI máximo y selección de puntos representativos (25 h).
- 1.4. Obtención y procesamiento de datos meteorológicos vía NASA POWER (25 h).
- 1.5. Cálculo de GPM, AGDD, Q y fotoperíodo por lote (30 h).

2. Grupo de tareas 2: Ajuste de la función β y modelado bayesiano (120 h).

- 2.1. Ajuste inicial de la función β a series de NDVI (25 h).
- 2.2. Estimación de parámetros mediante métodos bayesianos (40 h).
- 2.3. Validación de la curva ajustada por lote (30 h).
- 2.4. Identificación de errores sistemáticos (residuales) del modelo β para modelar con la ANN (25 h).

3. Grupo de tareas 3: Diseño e implementación del modelo híbrido con ANN (140 h).

- 3.1. Limpieza y preparación de datos de entrada para ANN (20 h).
- 3.2. Definición y entrenamiento de arquitectura base de ANN (30 h).
- 3.3. Ajuste de hiperparámetros y regularización (25 h).
- 3.4. Validación cruzada y evaluación de métricas de predicción (30 h).
- 3.5. Integración del modelo ANN con la función β para proyección híbrida (35 h).

4. Grupo de tareas 4: Evaluación técnica y análisis con datos 2024 (70 h).

- 4.1. Comparación de desempeño entre modelo bayesiano y modelo híbrido (20 h).
- 4.2. Discusión de resultados y recomendaciones de mejora (30 h).
- 4.3. Presentación interna y retroalimentación de colaboradores técnicos (20 h).

5. Grupo de tareas 5: Desarrollo de la interfaz de usuario y documentación técnica (100 h).

5.1. Diseño de estructura JSON y formato de entrada vía Excel (15 h).



- 5.2. Desarrollo de interfaz para carga de archivos y ejecución del modelo (25 h).
- 5.3. Visualización del NDVI proyectado y parámetros de la curva β (25 h).
- 5.4. Exportación de resultados y generación de informes automáticos (20 h).
- 5.5. Redacción de manual de usuario y guía técnica del sistema (15 h).
- 6. Grupo de tareas 6: Redacción, publicación y defensa del trabajo final (40 h).
 - 6.1. Redacción de memoria del Trabajo Final Integrador (20 h).
 - 6.2. Preparación de artículo científico para revista con referato (10 h).
 - 6.3. Preparación de la defensa y presentación pública (10 h).



10. Diagrama de Activity On Node

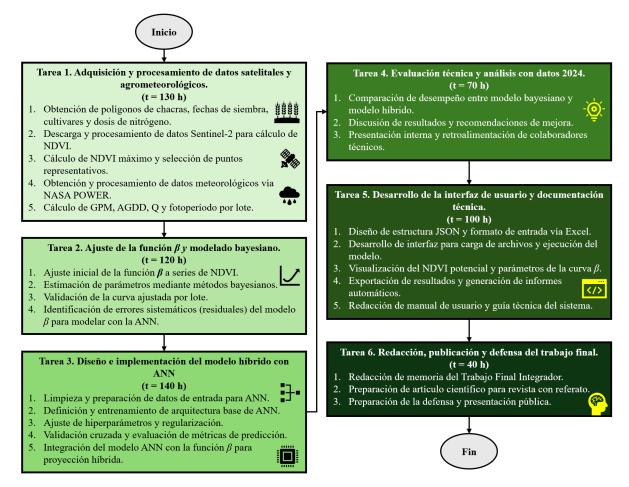


Figura 3. Diagrama de Activity on Node.



11. Diagrama de Gantt

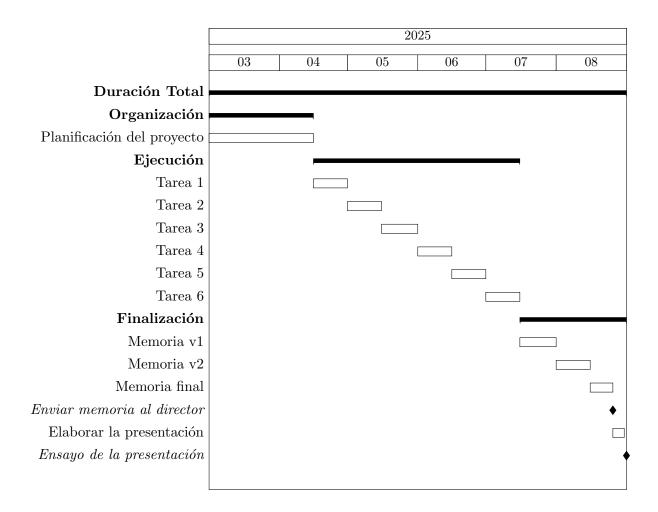


Figura 4. Diagrama de gantt del proyecto (marzo-agosto 2025)

12. Presupuesto detallado del proyecto

No se dispone de información sobre el presupuesto ya que es parte de un proyecto mayor. Además, es información confidencial de las entidades pertinentes. No obstante, el proyecto no corre riesgo por falta de presupuesto.

13. Gestión de riesgos

a) Identificación de los riesgos (al menos cinco) y estimación de sus consecuencias:

Riesgo 1: detallar el riesgo (riesgo es algo que si ocurre altera los planes previstos de forma negativa)

• Severidad (S): mientras más severo, más alto es el número (usar números del 1 al 10). Justificar el motivo por el cual se asigna determinado número de severidad (S).



 Probabilidad de ocurrencia (O): mientras más probable, más alto es el número (usar del 1 al 10).
 Justificar el motivo por el cual se asigna determinado número de (O).

Riesgo 2:

- Severidad (S): X.
 Justificación...
- Ocurrencia (O): Y. Justificación...

Riesgo 3:

- Severidad (S): X.
 Justificación...
- Ocurrencia (O): Y. Justificación...
- b) Tabla de gestión de riesgos: (El RPN se calcula como RPN=SxO)

Riesgo	S	О	RPN	S*	O*	RPN*

Criterio adoptado:

Se tomarán medidas de mitigación en los riesgos cuyos números de RPN sean mayores a...

Nota: los valores marcados con (*) en la tabla corresponden luego de haber aplicado la mitigación.

c) Plan de mitigación de los riesgos que originalmente excedían el RPN máximo establecido:

Riesgo 1: plan de mitigación (si por el RPN fuera necesario elaborar un plan de mitigación). Nueva asignación de S y O, con su respectiva justificación:

- Severidad (S*): mientras más severo, más alto es el número (usar números del 1 al 10). Justificar el motivo por el cual se asigna determinado número de severidad (S).
- Probabilidad de ocurrencia (O*): mientras más probable, más alto es el número (usar del 1 al 10). Justificar el motivo por el cual se asigna determinado número de (O).

Riesgo 2: plan de mitigación (si por el RPN fuera necesario elaborar un plan de mitigación).

Riesgo 3: plan de mitigación (si por el RPN fuera necesario elaborar un plan de mitigación).



14. Gestión de la calidad

Elija al menos diez requerimientos que a su criterio sean los más importantes/críticos/que aportan más valor y para cada uno de ellos indique las acciones de verificación y validación que permitan asegurar su cumplimiento.

- Req #1: copiar acá el requerimiento con su correspondiente número.
 - Verificación para confirmar si se cumplió con lo requerido antes de mostrar el sistema al cliente. Detallar.
 - Validación con el cliente para confirmar que está de acuerdo en que se cumplió con lo requerido. Detallar.

Tener en cuenta que en este contexto se pueden mencionar simulaciones, cálculos, revisión de hojas de datos, consulta con expertos, mediciones, etc.

Las acciones de verificación suelen considerar al entregable como "caja blanca", es decir se conoce en profundidad su funcionamiento interno.

En cambio, las acciones de validación suelen considerar al entregable como "caja negra", es decir, que no se conocen los detalles de su funcionamiento interno.

15. Procesos de cierre

Establecer las pautas de trabajo para realizar una reunión final de evaluación del proyecto, tal que contemple las siguientes actividades:

- Pautas de trabajo que se seguirán para analizar si se respetó el Plan de Proyecto original:
 - Indicar quién se ocupará de hacer esto y cuál será el procedimiento a aplicar.
- Identificación de las técnicas y procedimientos útiles e inútiles que se emplearon, los problemas que surgieron y cómo se solucionaron:
 - Indicar quién se ocupará de hacer esto y cuál será el procedimiento para dejar registro.
- Indicar quién organizará el acto de agradecimiento a todos los interesados, y en especial al equipo de trabajo y colaboradores:
 - Indicar esto y quién financiará los gastos correspondientes.