

Anticipando el Futuro: propuesta para un caso real

Ing. Roberto Asplanato

Agosto / 2025.

*Lo simple es intrínsecamente complejo y
lo complejo es explícitamente simple.*

¿Qué más bella sabiduría prueba la Creación?

– el autor –



Resumen

Este artículo propone un modelo predictivo para gestionar el sistema educativo público. Usando el "Efecto Látigo" de la logística como analogía, explica cómo los errores en la previsión de la matrícula estudiantil crecen y causan problemas de recursos. La solución es un modelo matemático de "Balance de Flujo" que sigue la trayectoria completa de los estudiantes (ingresos, abandonos y egresos) entre los niveles primario, secundario y terciario. El objetivo es crear una herramienta de soporte a la decisión (DSS¹) para optimizar recursos y mejorar la gobernanza educativa a través de decisiones informadas y basadas en evidencia.

Palabras Clave

Ciencia de Datos, Sistema Educativo, Modelo Predictivo, Efecto Látigo, Balance de Flujo, DSS (Sistema de Soporte a la Decisión).

Abstract

This article proposes a predictive model for managing the public education system. Using the "Bullwhip Effect" from logistics as an analogy, it explains how errors in student enrolment forecasting can grow and cause resource problems. The solution is a mathematical "Flow Balance" model that tracks the complete trajectory of students (intake, dropout, and graduation) across the primary, secondary, and tertiary levels. The goal is to create a decision support system (DSS¹) to optimize resources and improve educational governance through informed, evidence-based decisions.

Keywords

Data Science, Education System, Predictive Model, Bullwhip Effect, Flow Balance, DSS (Decision Support System).

Introducción

En el panorama global actual, ninguna organización puede operar de forma aislada, desconectada del ecosistema que integra y, aún menos, prescindir de la tecnología de la información. Esta última es la base para la gestión optimizada de procesos, para la automatización de flujos de trabajo, finalmente para la administración de la infraestructura y de los datos corporativos. Esta realidad impacta directamente en la estructura organizacional y en la cultura laboral.

Nuestro país evidencia un grado creciente de madurez tecnológica en comparación con años anteriores a pesar de las consabidas carencias en recursos humanos profesionales y en competencias digitales de algunos de sus cuadros. Con todo, el Estado y en particular las direcciones que rigen los subsistemas educativos, han incrementado sus capacidades digitales. Actualmente, los organismos de la educación avanzan en un proceso de sistematización que abarca áreas clave. Estos sistemas tienden naturalmente hacia la integración y la transversalidad, buscando automatizar procesos de punta a punta y compartir información de manera fluida. Aunque con distintos niveles de avance, este proceso constituye un cambio organizacional apalancado por la tecnología, es decir, una auténtica transformación digital del sector educativo.

Planteamiento del problema

A pesar de los avances, el sistema educativo enfrenta un desafío crucial: la carencia de información consolidada y oportuna para una planificación estratégica eficaz. Esta brecha afecta directamente la capacidad de prever con precisión la matrícula estudiantil futura (a corto, mediano y largo plazo), un pilar fundamental para la correcta asignación de recursos. La matrícula, compuesta por todos los estudiantes formalmente inscritos, es la razón de ser del sistema educativo. El éxito del proceso de enseñanza-aprendizaje depende de una planificación que los ponga en el centro, y la eficacia del sistema se mide, en última instancia, por sus trayectorias y resultados.

Propuesta de la solución

Para resolver este problema, se propone un modelo conceptual para el desarrollo de una herramienta de soporte a la decisión (DSS¹). El objetivo es apoyar una gobernanza integral y basada en datos del sistema educativo. Esta herramienta facilitará la integración de los subsistemas, la alineación de políticas educativas y la optimización de la gestión de recursos. El foco principal será el seguimiento de la trayectoria educativa del estudiante, aprovechando la madurez organizacional y las capacidades tecnológicas ya instaladas en el país.

Es importante aclarar que **"Integrar los diferentes subsistemas"** en cuanto a la gestión de la matrícula, **no significa necesariamente una integración organizacional, ni siquiera a nivel informático** (debido al nivel de complejidad, distintos niveles de madurez organizacional y costo que ello conlleva); basta con la **alineación y sincronización de los procesos** de cada subsistema que interactúan con el modelo e implantar una estandarización consensuada. Se podrá avanzar en el mediano y largo plazo en una integración estructural si se lo requiere, lo cual tampoco garantiza una mejora u optimización real de la problemática en cuestión.

¿Qué aspectos no se incluyen en la propuesta?

- Cambio organizacional y variables de entorno.
- Gestión de interesados (políticos, sectoriales, sociedad, etc).
- Marco legal y regulatorio.
- Aspectos vinculados a la planificación financiera.
- Indicadores de gestión tales como, Indicador Clave de Rendimiento (KPI), Retorno sobre la inversión (ROI), entre otros.
- Propiedad intelectual.
- Protección de datos personales.
- Ciberseguridad.
- Todo otro factor no desarrollado expresamente en el trabajo.

Situación del sistema educativo nacional

El actual sistema educativo nacional se encuentra en un proceso de reforma que impacta los niveles inicial, primario, secundario y técnico, así como la formación docente, la educación digital y la evaluación de los aprendizajes. Esta transformación se articula principalmente a través de la Transformación Curricular Integral (TCI²) y el nuevo Marco Curricular Nacional (MCN³). Estos documentos incorporan conceptos clave como la Educación Básica Integrada (EBI⁴) y la Educación Media Superior (EMS⁵). No obstante, el diseño y los detalles de esta reforma por razones de enfoque no influyen en la fase que aborda este estudio.

Enfoque sistémico de la propuesta

En este trabajo se plantea una propuesta holística o sistémica cuyo objetivo no es profundizar en los componentes de cada subsistema, en la complejidad de sus interrelaciones, ni en reglas organizacionales específicas. **El foco se sitúa en la totalidad del sistema educativo** (en todos sus niveles al mismo tiempo), priorizando las entradas y salidas de los estudiantes (es decir, su ingreso y egreso al sistema). Desde esta perspectiva global, es posible lograr una optimización real del sistema. Esto contrasta con los enfoques que resuelven los componentes individuales por separado (por ejemplo, los métodos Top-down⁶ y Bottom-up⁷), donde la optimización de una parte no garantiza necesariamente la optimización del sistema completo por tratarse de sistemas interdependientes.

Comportamiento de un sistema interdependiente

Descripción

El "Efecto Látigo" es un fenómeno descrito por Jay Forrester⁸ y sigue siendo valioso hoy en día por razones prácticas y estratégicas. En el mismo, se demuestra que una pequeña variación en la demanda del consumidor final se amplifica a medida que se transmite hacia atrás en una cadena logística, generando distorsiones crecientes para el resto de los actores (proveedores, fabricantes, distribuidores y comerciantes). Esta distorsión tiene similitudes con otros sistemas de pronóstico de demanda utilizados, por ejemplo, en la industria y el comercio, y también puede aplicarse al modelo del sistema educativo propuesto porque ambos sistemas, la educación y la cadena logística, son flujos interdependientes. En el modelo educativo, la entrada principal son los estudiantes. Si la estimación inicial del número de estudiantes ("la demanda") es imprecisa, incluso un error mínimo se amplifica a medida que se propaga a través de los diferentes niveles del sistema (primario, secundario, terciario). Esto genera distorsiones y desajustes, un comportamiento análogo al "efecto látigo" en la cadena logística.

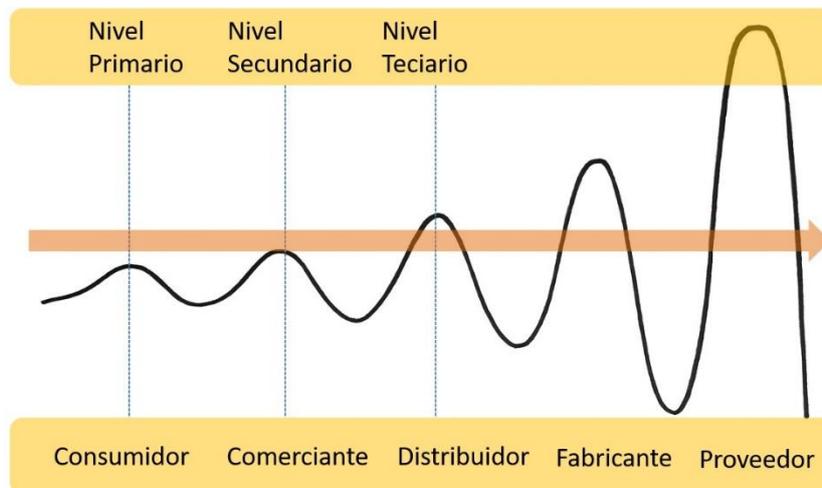


Fig.1 – Efecto látigo en sistemas diferentes.

A continuación, se hace una comparativa.

Aspecto	Cadena Logística	Sistema Educativo (Modelo propuesto)
Flujo	Bienes y productos	Estudiantes
Entrada inicial	Pedido del cliente final (demanda)	Estimación de la matrícula
Variación	Pequeña variación en la demanda del consumidor final	Error mínimo en la estimación de ingreso estudiantil
Amplificación	Se amplifica a medida que se transmite hacia atrás en la cadena logística	Cualquier error se amplifica en el flujo siguiente de actividades
Decisiones	Los actores toman decisiones independientes y reaccionan a su propia percepción de la demanda de inventario	Cada subsistema tiende a tomar decisiones independientes, reaccionando a su propia percepción del ingreso esperado
Consecuencias	Distorsiones crecientes para los actores, excesos de inventario, incremento de los costos y del tiempo	Asignación excesiva o insuficiente de recursos, duplicación de recursos, descoordinaciones y errores, infraestructura inadecuada (sobrante o deficitaria), retrasos, uso de cronogramas desactualizados y sobrecostos operativos
Impacto Final	Ineficiencias en la gestión de la cadena logística en general con pérdidas económicas	Impacto negativo en el proceso de enseñanza-aprendizaje y en la calidad educativa (escasez de docentes, aulas no preparadas, cursos suspendidos, etc)

Tabla 1 – Comparativa de sistemas.

Acciones para mitigar el Efecto Látigo en el Sistema educativo

Basado en las lecciones aprendidas se pueden resumir.

Recomendación	Acción de mitigación en la Educación
Transparencia y acceso a datos en tiempo real	Asegurar la disponibilidad compartida de datos para el acceso en tiempo real por parte de los subsistemas previendo la existencia de controles, sincronización y consolidación de la información
Visión global del proceso por todos los subsistemas	Modelo unificado de subsistemas y visualización completa de trayectorias educativas a nivel global
Monitoreo continuo y puntos de control	Implementación de alertas tempranas por abandono, retraso académico, saturación de cupos, etc
Pronósticos colaborativos y precisos	Uso de modelos estadísticos, de IA ⁹ e híbridos (series de tiempo, regresión, redes neuronales) para estimar demanda real de estudiantes, por asignatura, centro, localidad, ciudad, región y subsistema, entre otros
Minimizar retrasos en decisiones	Integrar a todos los actores a los sistemas de gestión asignando roles claros con el fin de potenciar la coordinación y optimizar los tiempos de respuesta y toma de decisiones (por ejemplo, en inscripciones, asignaciones y calendarización)

Tabla 2 – Acciones para mitigar el Efecto Látigo en el Sistema educativo.

Corolario

En esencia, la aplicación de esta lección del "Efecto Látigo" permite comprender cómo las variaciones multicausales en la matrícula pueden generar distorsiones operativas y organizativas asociadas a pérdidas y retardos, y cómo mitigar estas ineficiencias para optimizar el uso de recursos con el fin de mejorar la calidad educativa.

Modelo del sistema educativo

Descripción a alto nivel

Desde un punto de vista holístico formal, el sistema educativo público podría representarse como una función con entradas y salidas:

$$f(e_i, s_j)$$

Referencias:

f : función / e, s : variables (estudiantes que ingresan o egresan del sistema) / $i: 0..n, j: 0..m / n \in \mathbb{N}, m \in \mathbb{N}$

Donde una cantidad de entradas $\{e_i: i = 1, 2, \dots, n\} \subset \mathbb{N}$ y salidas $\{s_j: j = 1, 2, \dots, m\} \subset \mathbb{N}$ se pueden definir mediante una función $f: \mathbb{N}^n \times \mathbb{N}^m \rightarrow \mathbb{N}$, siendo \mathbb{N} el conjunto de los números naturales.

En resumen, se tiene:

- e_i Entrada o ingreso: Cantidad de estudiantes que ingresan al sistema / $i: 0..n, n \in \mathbb{N}$
- s_j Salida o egreso: Cantidad de estudiantes que egresan del sistema / $j: 0..m, m \in \mathbb{N}$

El modelo representa el sistema educativo como un flujo secuencial a través de tres niveles principales: Primario, Secundario y Terciario. La lógica central es que los estudiantes que egresan exitosamente de un nivel constituyen la entrada principal del siguiente.

- P: Nivel primario (educación Inicial y Primaria).
- S: Nivel secundario (educación Secundaria y Formación técnico profesional).
- T: Nivel terciario (educación Universitaria, Terciaria y Formación en educación).

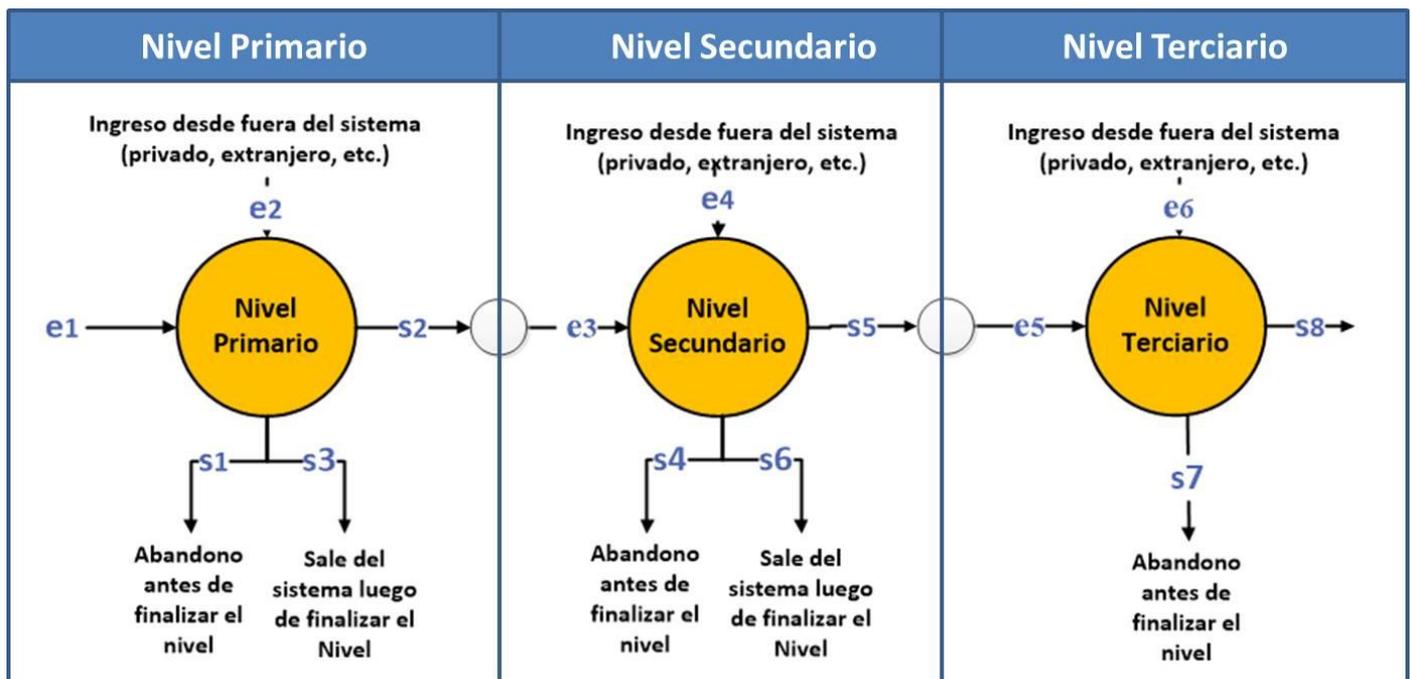


Fig.2 – Modelo del Sistema Educativo.

Comportamiento

1. Nivel Primario (P):

- Entradas: Recibe estudiantes por dos vías:
 - e₁: Ingreso regular al sistema (ej. niños en edad de inicio escolar).
 - e₂: Ingreso excepcional (ej. transferencias desde el extranjero, sistema privado).
- Salidas: Tiene tres posibles salidas:
 - s₁: Abandono antes de finalizar el nivel.
 - s₂: Egreso exitoso. Este flujo se convierte en la entrada e₃ del Nivel Secundario.
 - s₃: Salida del sistema después de finalizar el nivel (no continúan estudiando).

2. Nivel Secundario (S):

- Entradas: Recibe estudiantes por dos vías:
 - e₃: Ingreso de estudiantes que finalizaron exitosamente el Nivel Primario.
 - e₄: Ingreso excepcional (ej. transferencias, reingresos).
- Salidas: Tiene tres posibles salidas:
 - s₄: Abandono antes de finalizar el nivel.
 - s₅: Egreso exitoso. Este flujo se convierte en la entrada e₅ del Nivel Terciario.
 - s₆: Salida del sistema después de finalizar el nivel (no continúan estudiando).

3. Nivel Terciario (T):

- Entradas: Recibe estudiantes por dos vías:
 - e₅: Ingreso de estudiantes que finalizaron exitosamente el Nivel Secundario.
 - e₆: Ingreso excepcional (ej. transferencias, reingresos).
- Salidas: Tiene dos posibles salidas:
 - s₇: Abandono antes de finalizar el nivel.
 - s₈: Egreso final y exitoso del sistema educativo.

Desarrollo

El modelo se formaliza a través de **ecuaciones de balance de flujo** para cada nivel y para el sistema en su totalidad. El "Balance de Flujo por Nivel" es una ecuación que establece que, para cualquier nivel educativo (Primario, Secundario o Terciario), la cantidad total de estudiantes que ingresan debe ser exactamente igual a la cantidad total de estudiantes que salen de ese mismo nivel. Es un principio de conservación aplicado al modelo educativo. Imagina que cada nivel es un contenedor. El balance de flujo simplemente afirma que no se pueden "crear" ni "destruir" estudiantes dentro del contenedor. Todos los que entran deben eventualmente salir por alguna de las vías definidas (abandono, egreso exitoso o salida del sistema).

Balance de Flujo por Nivel

Estas ecuaciones describen cómo se distribuyen los estudiantes dentro de cada nivel.

- Ecuación para el Nivel Primario (P):

El total de estudiantes que ingresan a primaria debe ser igual a la suma de los que abandonan, los que egresan exitosamente y los que salen del sistema tras finalizar.

$$e_1 + e_2 = s_1 + s_2 + s_3$$

- Ecuación para el Nivel Secundario (S):

El total de estudiantes que ingresan a secundaria debe ser igual a la suma de los que abandonan, los que egresan exitosamente y los que salen del sistema tras finalizar.

$$e_3 + e_4 = s_4 + s_5 + s_6$$

- Ecuación para el Nivel Terciario (T):

El total de estudiantes que ingresan a terciaria debe ser igual a la suma de los que abandonan y los que egresan exitosamente del sistema.

$$e_5 + e_6 = s_7 + s_8$$

Ecuaciones del Sistema

Estas son las condiciones que conectan los niveles y describen el comportamiento global del sistema.

- Condiciones de Transición:

Garantizan la correcta conexión del flujo entre niveles.

La entrada regular a Secundaria es la salida exitosa de Primaria: $e_3 = s_2$

La entrada regular a Terciario es la salida exitosa de Secundaria: $e_5 = s_5$

- Ecuación de Entradas Totales al Sistema:

Suma de todos los ingresos desde fuera del sistema en cualquier nivel.

$$\text{Entradas_Totales} = e_1 + e_2 + e_4 + e_6$$

- Ecuación de Salidas Totales del Sistema:

Suma de todos los flujos de estudiantes que abandonan o egresan definitivamente del sistema.

$$\text{Salidas_Totales} = s_1 + s_3 + s_4 + s_6 + s_7 + s_8$$

- Ecuación de Balance General del Sistema:

Para que el modelo sea consistente, el total de estudiantes que entran al sistema debe ser igual al total de estudiantes que salen.

$$\text{Entradas_Totales} = \text{Salidas_Totales}$$

O, de forma expandida:

$$e_1 + e_2 + e_4 + e_6 = s_1 + s_3 + s_4 + s_6 + s_7 + s_8$$

Un ejemplo de corrida inicial

Para entender rápidamente el planteo formal, se propone un ejemplo práctico “de corrida inicial y con datos ficticios perfectos”, entendiendo como datos perfectos al concepto de contar con todos los datos sin errores ni retrasos, al momento de ejecución del modelo. Se tiene según figura.

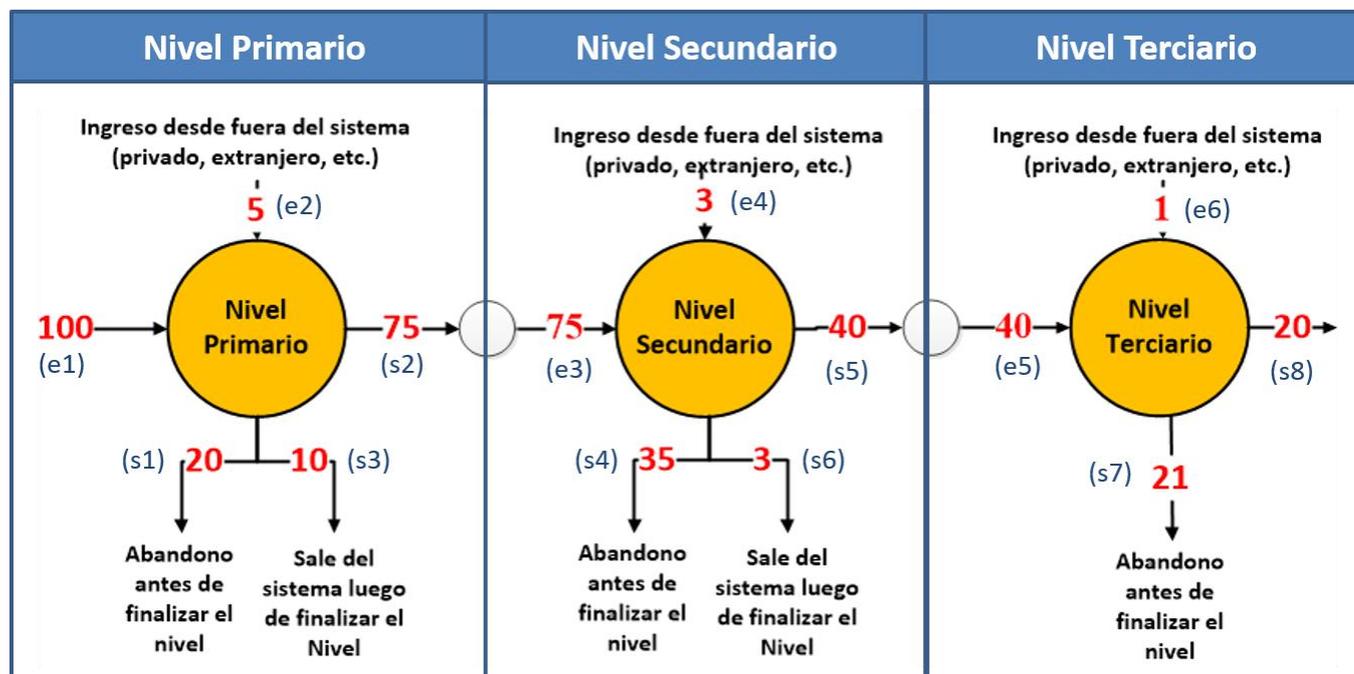


Fig.3 – Ejemplo de corrida inicial.

Los datos del ejemplo de la Fig.3 satisfacen las nuevas ecuaciones de balance de flujo.

- **Balance Primario:** $100 + 5 = 20 + 75 + 10$ ($105 = 105$)
- **Balance Secundario:** $75 + 3 = 35 + 40 + 3$ ($78 = 78$)
- **Balance Terciario:** $40 + 1 = 21 + 20$ ($41 = 41$)

Algunas interpretaciones a modo de ejemplo

- El 19% de los estudiantes que inicia los estudios en Primaria [$20 / (100+5)$] finaliza estudios de Nivel Terciario.
- El 45% [$35 / (75+3)$] de los estudiantes que inicia los estudios de Nivel Secundario lo abandona.
- El 70% [$(20+35+21) / (100+5+3+1)$] de los estudiantes que ingresan al Sistema educativo, lo abandonan sin finalizar el subsistema que están cursando.
- El 63% de los estudiantes [$(20+10+35+3) / (100+5+3)$] no finalizan los estudios en el Nivel Secundario o dicho de otra forma, de los $75+3=78$ que cursaron secundaria, 40 continuaron y $35+3=38$ abandonaron. Por lo que el nivel de deserción es alto. En este último caso, y a modo de ejemplo, una posible solución podría ser desestimular el abandono mediante la aplicación de alertas tempranas junto con otras medidas (apoyos pedagógicos, psicológicos, económicos y sociales).

Esta representación es sencilla si se conocen los valores, ello ocurre con los datos históricos. Pero **el desafío que se pretende resolver es, conocer los datos futuros lo más aproximado posible**, mediante estimaciones a corto, mediano y largo plazo.

Función de Error

Debido a que estadísticamente toda medición o estimación conlleva un componente de error, es necesario incorporar una función de error al modelo:

$$\mathit{error}(e_i, s_j)$$

Es la función que agrupa los errores de medición o estimación. El error se define como el valor real menos el valor calculado por el modelo. Cuanto más pequeño es el error, más preciso es el resultado final de la función. **El objetivo, es minimizar la función $\mathit{error}(e_i, s_j)$ todo lo posible a un costo-beneficio que resulte razonable.**

Nota: Para minimizar el error mediante algoritmos de predicción es posible emplear herramientas estándar como: Error Medio Absoluto (MAE¹⁰), Raíz del Error Cuadrático Medio (RMSE¹¹), Error Porcentual Absoluto Medio (MAPE¹²), entre otras opciones.

Extensión del cálculo

Continuando el análisis, el modelo propuesto puede “refinarse” en fases sucesivas e irse complejizando para aportar más información para la toma de decisiones; por ejemplo, un segundo nivel de detalle podría ser el de incluir un **pronóstico de la demanda de los estudiantes por asignatura**. Ello permitiría, además, establecer el número de grupos de una asignatura que deberá formarse, prever la disponibilidad de docentes y su actualización, y provisión de salones, laboratorios, infraestructura y servicios necesarios; todo lo cual, redundaría en una previsión de gastos y tiempos de preparación, que podrían ser optimizados.

En este último caso, se incorporaría al modelo, el cálculo de la demanda futura de una asignatura dada:

$$d(a_k, t+1) = \mathit{apr}(a_{k-1}, t) + \mathit{rep}(a_k, t) + \mathit{aba}(a_k, t)$$

Referencias:

d: demanda de una asignatura / apr: aprobado / rep: reprobado / aba: abandono / a: asignatura / t: tiempo o período considerado, t+1: tiempo o período siguiente / k: instancia actual o presente, k-1: instancia anterior, k: 0..q, q ∈ N

Donde se tiene:

- $\mathit{apr}(a_{k-1}, t)$: estudiantes que aprobaron la asignatura previa requerida.
- $\mathit{rep}(a_k, t)$: estudiantes que reprobaron la asignatura actual y deben recusarla.
- $\mathit{aba}(a_k, t)$: estudiantes que abandonaron la asignatura y se espera que se reinscriban.

Restricciones: no se refleja aquí la previsión de estudiantes que puedan eventualmente estar en cola de espera por algún motivo (falta de docentes, ausencia de las condiciones para la realización del curso, entre otras).

Condición:

- Representa sólo los estudiantes que permanecen en el sistema o con perspectiva de reincorporarse en el siguiente período.

- Requiere el desarrollo de un grafo o árbol de previaturas de asignaturas y/o reglas de aprobación. Las mismas forman parte de las restricciones del cálculo.

Finalmente, se puede **extender el pronóstico de una asignatura a uno o más grupos de asignaturas** que forman parte de un programa o plan de estudios, también a cursos obligatorios u opcionales, con o sin créditos, entre otras opciones. En cuyo caso, podría representarse mediante la siguiente expresión:

$$\sum_{k=1}^q d(a_k, t+1)$$

Referencias:

d: demanda de una asignatura / *a*: asignatura / *t*: tiempo o período considerado, *t+1*: tiempo o periodo siguiente / *k*: instancia actual o presente / $q \in N$

Donde se tiene:

- $d(a_{k,t+1})$ representa al número de la demanda futura de una asignatura considerada.
- $\sum_{k=1}^q d(a_k, t+1)$ representa la suma de la demanda futura de *q* asignaturas consideradas, por ejemplo, constituyentes de un plan de estudios.

Discusión técnica

Para que el modelo funcione y represente adecuadamente la realidad, es necesario atender algunos puntos clave, entre ellos se encuentran: modelo de pronóstico, calidad de los datos, variables, metodología, algoritmos de cálculo, y la precisión esperada. Aquí haremos mención a algunos de ellos.

Modelos de pronóstico

La elección de los modelos de pronóstico es crítica. Se debe buscar el mejor equilibrio entre precisión y costo computacional. Este campo ha tenido un gran desarrollo con el avance de la IA por lo que, al momento de su aplicación es recomendable pensar en ellos como un "kit de herramientas" que pueden incluir desde modelos tradicionales hasta soluciones de IA y de aplicación tanto individual como combinada o híbrida.

A continuación, se muestra a modo informativo, un resumen.

Categoría Principal	Subcategoría	Modelos Representativos	Algunas aplicaciones en el Contexto Educativo
Modelos Cuantitativos Tradicionales	Series de Tiempo Clásicas	Promedio Móvil, Suavización Exponencial, ARIMA ¹³ , SARIMA ¹⁴	Predicción de matrícula anual o trimestral, demanda de recursos basada en datos históricos
	Causales (Regresionales)	Regresión Lineal Simple y Múltiple, Regresión Logística	Predecir la probabilidad de éxito de un estudiante basándose en variables como su rendimiento previo

			o factores socioeconómicos
Modelos de Aprendizaje Automático (ML¹⁵) y Profundo (DL¹⁶)	Modelos de Aprendizaje Automático Clásico	Random Forest ¹⁷ , XGBoost ¹⁸ , Support Vector Machines (SVM ¹⁹)	Clasificación de estudiantes en riesgo de deserción, predicción de la demanda de asignaturas
	Redes Neuronales (NN²⁰)	Perceptrón Multicapa (MLP ²¹), Redes Neuronales Recurrentes (RNN ²²), Long Short-Term Memory (LSTM ²³), Redes Neuronales Convolucionales (CNN ²⁴), Redes Transformers	Pronóstico de series de tiempo complejas y no lineales (ej. rendimiento estudiantil), predicción de la trayectoria educativa
	Modelos Híbridos y de Conjunto (Ensemble)	Combinación de ARIMA y Red Neuronal, Ensemble ²⁵ de Random Forest y XGBoost	Mejora de la precisión del pronóstico de matrícula al combinar la capacidad de modelar tendencias lineales con la de capturar patrones no lineales
Enfoques Avanzados y Metodológicos	Pronóstico Probabilístico	Modelos Bayesianos ²⁶ , Cuantílicos ²⁷ (ej. regresión cuantílica)	Proporcionar rangos de pronóstico (intervalos de confianza) para la demanda de profesores o aulas, en lugar de un único valor
	Ingeniería de Características²⁸	Creación de variables a partir de datos demográficos, socioeconómicos, de rendimiento y geográficos	Integración de datos externos como tasas de natalidad o empleo para mejorar la precisión de las predicciones de matrícula
	IA Explicable²⁹ (XAI)	Técnicas como SHapley Additive exPlanations (SHAP ³⁰) y Local Interpretable Model-agnostic Explanations (LIME ³¹)	Justificación de las predicciones de riesgo de deserción estudiantil, identificando los factores más influyentes

Tabla 3 – Clasificación de Modelos de pronóstico.

VARIABLES DEL MODELO

Las variables a considerar deben ser robustas y representativas. Si bien las mismas pueden variar según ajuste o necesidad, se proponen las siguientes categorías.

- **Indicadores Demográficos:** Grupo etario, sexo, localización geográfica (departamento, ciudad, localidad, barrio).
- **Indicadores de Trayectoria Académica:** Nivel de estudio, eventos (aprobado, reprobado, abandono), plan de estudio, asignatura, modalidad (presencial, a distancia).
- **Datos Institucionales:** Subsistema, centro de estudio, docente a cargo, plan, programa.

Nota: Las categorías de variables quedan abiertas a una fase posterior de definición donde según el modelo y la estrategia seleccionada, es posible incorporar otras categorías, como: variables socioeconómicas (ej. nivel de ingresos del hogar, nivel educativo de los padres) y variables de comportamiento del estudiante (ej. participación en actividades educativas).

GESTIÓN DE DATOS

La calidad del modelo depende directamente de la calidad de los datos. Se requiere:

- **Estandarización:** Unificar criterios operativos, procesos, definiciones y granularidad de los datos.
- **Centralización:** Implementar un repositorio centralizado, como un Data Warehouse³², que funcione como *la principal fuente única de la verdad (Single Source of Truth, SSOT³³) para evitar inconsistencias.*
- **Automatización:** Construir un pipeline³⁴ de datos robusto para los procesos de Extracción, Transformación y Carga (ETL), que alimente el repositorio de forma continua y fiable desde las distintas fuentes de datos.
- **Arquitecturas Modernas:** Aunque se propone el Data Warehouse, se debe considerar su integración con otras arquitecturas modernas como los Data Lakes³⁵ para la exploración y el análisis avanzado de datos no estructurados.

Nota: Si bien existen varias tecnologías de almacenamiento y tratamiento de datos; dicho análisis está fuera del alcance de este trabajo, por lo que, se menciona una de las soluciones tradicionales posibles. En nuestro caso, se plantea definir una tecnología del tipo Data Warehouse; ya que a priori se deben almacenar los diferentes valores que toma una variable a lo largo del tiempo (por ejemplo, que almacene los diferentes centros de estudio de un estudiante en el transcurso de su trayectoria por los tres niveles del sistema educativo, vinculado también a sus actuaciones). Y junto con dicho repositorio, se requiere establecer los procesos asociados mediante los que se accede a diferentes fuentes de datos que lo alimentan (ETL). Finalmente, es necesario en una siguiente etapa definir roles, responsabilidades y políticas sobre los datos, para asegurar su sostenibilidad.

Lógica del Proyecto: Fases de Implementación

Se propone una implementación estructurada en fases, alineada con el ciclo de vida de un proyecto de ciencia de datos.

Fase 1: Definición y Alcance

- **Objetivos:** Definir los objetivos principales y específicos del proyecto.
- **Alcance:** Documentar el alcance, supuestos, restricciones, beneficios e interesados.
- **Análisis Costo-Beneficio:** Evaluar la inversión requerida frente a los ahorros y mejoras esperadas.

Fase 2: Análisis y Preparación de Datos

- **Relevamiento:** Auditar la situación actual, identificar fuentes de datos (internas y externas) y evaluar su calidad.
- **Procesamiento y limpieza de datos:** Los datos se preparan para el análisis. Esto incluye la eliminación de datos duplicados, el manejo de valores faltantes y la corrección de errores.
- **Modelado de Datos:** Análisis de datos y diseño del modelo de datos para el Data Warehouse.
- **Pipeline ETL:** Construir y probar los procesos de extracción, transformación y carga de datos.

Fase 3: Modelado y Simulación

- **Selección de Modelos:** Evaluar y seleccionar los algoritmos de pronóstico más adecuados (regresión, series de tiempo, machine learning, híbridos, etc).
- **Entrenamiento:** Entrenar los modelos utilizando los datos preparados.
- **Calibración:** Ajustar los parámetros de los modelos para optimizar su rendimiento.

Fase 4: Evaluación y Validación

- **Medición de Precisión:** Evaluar el rendimiento de los modelos con datos de validación, utilizando métricas de precisión.
- **Análisis de Resultados:** Comparar los resultados de diferentes modelos y analizar el tratamiento de errores.
- **Generación de Recomendaciones:** Formular sugerencias para la toma de decisiones basadas en los pronósticos (tales como, árbol de decisión).

Fase 5: Implementación y Operación

- **Despliegue:** Poner el modelo en un entorno de producción.
- **Desarrollo de la Herramienta (DSS):** Construir la interfaz de usuario (dashboards, reportes, etc) para visualizar los pronósticos.
- **Capacitación:** Formar a los usuarios finales en el uso de la herramienta.
- **Mantenimiento y Monitoreo:** Establecer un plan de soporte continuo y monitoreo del rendimiento del modelo.

Perspectiva a futuro

- **Profundizar en la Estrategia de Datos:** Analizar los desafíos esperados en la integración de datos y proponer un enfoque gradual, como un proyecto piloto en una región o área específica, para validar el modelo antes de un despliegue nacional.
- **Incorporar Simulación de Políticas:** Implementar la solución DSS en una primera etapa como una herramienta de pronóstico, y en una etapa más avanzada de **simulación estratégica**. Por ejemplo, permitiría a los responsables evaluar el impacto a futuro de una política destinada a reducir las tasas de abandono.
- **Enfatizar el Análisis Causal:** En una fase posterior, el modelo puede emplearse para realizar análisis causales, correlacionando las variaciones en los flujos de estudiantes (ej. aumento del abandono) con factores externos (ej. crisis económicas, cambios curriculares), pasando de la predicción a la comprensión.

Conclusión

El país posee el conocimiento, los profesionales y los recursos tecnológicos para implementar un proyecto de esta magnitud. La madurez organizacional de los actores involucrados ofrece una oportunidad que hace pocos años era impensable para su implementación. Se estima que la inversión se vería compensada –al menos en parte– por los ahorros generados a través de una gestión optimizada.

Este proyecto es una **oportunidad** ideal para la colaboración entre el Sistema educativo público, la Academia (universidades públicas y privadas), la Industria tecnológica y el Gobierno como principal sponsor del proyecto. Por supuesto, no reemplaza el compromiso y el afecto que le pongan sus diferentes actores –ya sean docentes y otros colaboradores del sistema educativo, las familias, y los niños y jóvenes que se benefician de él–. Más bien, los potencia.

Finalmente, este trabajo tiene como propósito presentar un **modelo conceptual** con el fin de promover la reflexión para el **desarrollo de una herramienta** de alto impacto, creada con talento y recursos nacionales; que fortalezca la **gobernanza educativa y su sistema**, pero ante todo, que beneficie directamente a nuestros niños y jóvenes, que son el futuro del Uruguay.

Referencias

Figuras: CC BY-SA 4.0, via Wikimedia Commons – 2025 – Autor: Roberto Asplanato.

- Portada -Gob basada en Datos – URL: commons.wikimedia.org/wiki/File:Portada_-_Gob_basada_en_Datos.jpg
- Fig.1 – Efecto látigo en sistemas diferentes. – URL: [commons.wikimedia.org/wiki/File:Curva_-_Analogia_\(Edu_vs_Log\)_Efecto_Latigo.jpg](https://commons.wikimedia.org/wiki/File:Curva_-_Analogia_(Edu_vs_Log)_Efecto_Latigo.jpg)
- Fig.2 – Modelo del Sistema Educativo. – URL: commons.wikimedia.org/wiki/File:Proceso_-_1.jpg
- Fig.3 – Ejemplo de corrida inicial. – URL: commons.wikimedia.org/wiki/File:Proc_-_2.png

Tablas:

- Tabla 1 – Comparativa de sistemas.
- Tabla 2 – Acciones para mitigar el Efecto Látigo en el Sistema educativo.
- Tabla 3 – Clasificación de Modelos de pronóstico.

¹ DSS: Sistema de Soporte a la Decisión (Decision Support System, en inglés) es una herramienta informática que recopila, analiza y visualiza datos para ayudar a los directivos a tomar decisiones estratégicas informadas. No es solo un tablero de control que muestra datos, se trata de una herramienta que reúne las siguientes características principales: Enfoque en la Decisión Futura (conocer los datos futuros lo más aproximado posible), Integra Datos con Modelos predictivos, Simulación y Análisis "What-If" (análisis de escenarios posibles basados en evidencia temprana) e Interactividad para No Expertos (permite a los responsables de políticas educativas interactuar con los datos y los modelos sin necesidad de ser expertos en ciencia de datos).

² TCI: recuperado de: www.anep.edu.uy/documentos-curriculares

³ MCN: recuperado de: www.anep.edu.uy/documentos-curriculares

⁴ EBI: recuperado de: www.anep.edu.uy/documentos-curriculares

⁵ EMS: recuperado de: www.anep.edu.uy/documentos-curriculares

⁶ Top-down: parten de lo general a lo específico, definiendo la visión global y descomponiéndola en tareas más pequeñas

⁷ Bottom-up: parten de lo específico para construir hacia un todo, integrando soluciones de menor nivel hasta alcanzar un objetivo general

⁸ Jay Forrester: fue un ingeniero y científico estadounidense del MIT que acuñó el término "Efecto Látigo" y fue el primero en estudiarlo en su libro (Industrial Dynamics, 1961)

⁹ IA: La inteligencia artificial es una rama de la informática que busca crear sistemas y máquinas capaces de simular capacidades humanas como el aprendizaje, el razonamiento, la percepción, la toma de decisiones y la resolución de problemas

¹⁰ Error Medio Absoluto: El MAE (Mean Absolute Error, en inglés) calcula el promedio de los errores absolutos entre los valores predichos y los valores reales

¹¹ Raíz del Error Cuadrático Medio: El RMSE (Root Mean Square Error, en inglés) es la raíz cuadrada del promedio de los errores al cuadrado. Penaliza más los errores grandes porque eleva al cuadrado la diferencia entre el valor real y el predicho

¹² Error Porcentual Absoluto Medio: El MAPE (Mean Absolute Percentage Error, en inglés) expresa el error de predicción en términos de porcentaje, lo que facilita su interpretación y comparación entre diferentes conjuntos de datos o modelos

¹³ ARIMA: Media móvil integrada autorregresiva (Autoregressive integrated moving average, en inglés) es una técnica para el análisis de series temporales y para pronosticar posibles valores futuros de una serie temporal, recuperado de: www.ibm.com/mx-es/think/topics/arima-model

¹⁴ SARIMA: Promedio móvil integrado autorregresivo (Seasonal autoregressive integrated moving average, en inglés) es un modelo econométrico avanzado para predecir series de tiempo que exhiben patrones estacionales, como ciclos diarios, semanales o anuales, recuperado de: https://www.ibm.com/es-es/think/topics/arima-model?mhsrc=ibmsearch_a&mhq=arima

¹⁵ ML: El Aprendizaje de Máquina (Machine Learning, en inglés) es una rama de la inteligencia artificial (IA) que permite a las computadoras aprender de los datos y mejorar su rendimiento sin ser programadas explícitamente para cada tarea, recuperado de: www.ibm.com/think/topics/machine-learning

¹⁶ DL: El Aprendizaje Profundo (Deep Learning, en inglés) es una rama del aprendizaje automático que utiliza redes neuronales artificiales con múltiples capas para analizar datos y aprender patrones complejos, de manera similar a como lo hace el cerebro humano. Es una rama de la inteligencia artificial (IA) centrada en permitir que las computadoras y las máquinas imiten la forma en que aprenden los humanos, realicen tareas de forma autónoma y mejoren su rendimiento y precisión a través de la experiencia y la exposición a más datos, recuperado de: www.ibm.com/think/topics/deep-learning

¹⁷ Random Forest: Es un algoritmo de aprendizaje automático de uso común que combina el resultado de múltiples árboles de decisión para llegar a un resultado único, recuperado de: www.ibm.com/mx-es/think/topics/random-forest

¹⁸ XGBoost: Potenciación Extrema de Gradiente (eXtreme Gradient Boosting, en inglés) es un método de aprendizaje automático supervisado para clasificación y regresión y se utiliza en la herramienta de entrenamiento en general, recuperado de: www.ibm.com/mx-es/think/topics/xgboost

¹⁹ SVM: Máquinas de Vectores de Soporte (Support Vector Machines, en inglés) se emplean comúnmente en el procesamiento del lenguaje natural (NLP) para tareas como análisis de opiniones, detección de spam y modelado de temas (aprendizaje automático, clasificación tanto lineales como no lineales), recuperado de: www.ibm.com/mx-es/think/topics/support-vector-machine

²⁰ NN: Redes Neuronales (Neural Networks, en inglés) son un tipo de algoritmo de aprendizaje automático inspirado en el cerebro humano. Están compuestas por nodos interconectados, llamados neuronas, que procesan y transmiten información. Se utilizan para tareas complejas como el reconocimiento de imágenes y el procesamiento de lenguaje natural, recuperado de: www.ibm.com/es-es/think/topics/neural-networks

- ²¹ MLP: Perceptrón Multicapa (Multilayer Perceptron, en inglés) es un tipo de red neuronal artificial de propagación hacia adelante que se utiliza para el aprendizaje supervisado, capaz de aprender relaciones no lineales complejas entre datos mediante el entrenamiento, recuperado de: www.ibm.com/docs/es/spss-statistics/31.0.0?topic=networks-multilayer-perceptron
- ²² RNN: Las Redes Neuronales Recurrentes (Recurrent Neural Networks, en inglés) son modelos de aprendizaje profundo que procesan datos secuenciales, como texto o series temporales, utilizando una "memoria" interna que les permite tener en cuenta las entradas previas al generar una salida, recuperado de: www.ibm.com/es-es/think/topics/recurrent-neural-networks
- ²³ LSTM: Memoria a Largo Corto Plazo (Long Short-Term Memory, en inglés), es un tipo de arquitectura de red neuronal recurrente (RNN) diseñada para manejar secuencias de datos, especialmente aquellas con dependencias a largo plazo, recuperado de: www.ibm.com/es-es/think/topics/recurrent-neural-networks
- ²⁴ CNN: Las Redes Neuronales Convolucionales (Convolutional Neural Networks, en inglés) son un tipo de red neuronal artificial especializada en aprendizaje profundo para procesar datos como imágenes, audio y señales, recuperado de: www.ibm.com/es-es/think/topics/convolutional-neural-networks
- ²⁵ Ensemble: Aprendizaje en conjunto (Ensemble Learning, en inglés) se refiere a combinar múltiples modelos de aprendizaje automático para crear un modelo más robusto, preciso y fiable que cualquiera de los modelos individuales, recuperado de: www.ibm.com/es-es/think/topics/ensemble-learning
- ²⁶ Modelos Bayesianos: son marcos estadísticos que utilizan el Teorema de Bayes para actualizar las creencias sobre las hipótesis a medida que se dispone de nueva evidencia, resultando en el cálculo de una distribución posterior para los parámetros del modelo
- ²⁷ Cuantílicos: La regresión cuantil modela la relación entre un conjunto de variables predictoras (independientes) y percentiles específicos (o "cuantiles") de una variable de destino (dependiente), recuperado de: www.ibm.com/docs/es/spss-statistics/31.0.0?topic=regression-quantile
- ²⁸ Ingeniería de Características: (Feature Engineering, en inglés) es el proceso de transformar datos brutos en información útil y relevante para los modelos de aprendizaje automático, mediante la selección, transformación y creación de características (atributos o variables) que mejoren la precisión y eficiencia del modelo, recuperado de: www.ibm.com/mx-es/think/topics/feature-engineering
- ²⁹ IA Explicable: La Inteligencia Artificial Explicable (XAI o Explainable AI, en inglés) es un conjunto de métodos y procesos que permiten a los humanos comprender y confiar en las decisiones tomadas por los algoritmos de aprendizaje automático. Su objetivo es convertir los modelos de IA "caja negra" en sistemas transparentes, explicando cómo y por qué se llegan a un resultado determinado, facilitando así su adopción, la mitigación de sesgos, la mejora de la gobernanza y el desarrollo responsable de la IA, recuperado de: www.ibm.com/es-es/think/topics/explainable-ai
- ³⁰ SHAP: (Shapley Additive Explanations, en inglés) es un marco teórico y una biblioteca de Python que utiliza la teoría de juegos para explicar las predicciones de cualquier modelo de aprendizaje automático.
- ³¹ LIME: Explicaciones Locales Interpretables y Agnósticas al Modelo (Local Interpretable Model-Agnostic Explanations, en inglés), es un algoritmo que puede explicar las predicciones de cualquier clasificador de una manera fiel, aproximándolo localmente con un modelo interpretable. Modifica una sola muestra de datos ajustando los valores de la variable estudiada y observa el impacto resultante en la salida, recuperado de: www.analyticsvidhya.com/blog/2022/07/everything-you-need-to-know-about-lime
- ³² Data Warehouse: Almacén de Datos es un repositorio centralizado de datos, estructurado y optimizado, diseñado para el análisis y la elaboración de informes
- ³³ SSOT: Fuente única de la verdad (Single Source of Truth, en inglés) es el principio de que todos los datos de una organización deben residir en un único repositorio centralizado, garantizando la consistencia y la fiabilidad de la información.
- ³⁴ Pipeline de datos: Es un proceso automatizado que mueve datos desde múltiples fuentes, los convierte en un formato útil y estandarizado, y los almacena en un repositorio central
- ³⁵ Data Lakes: El Lagos de datos es un gran repositorio de datos que contiene datos sin procesar, o "brutos", en su formato original. Es una arquitectura de datos moderna.

Esta obra está licenciada bajo la Licencia Creative Commons Atribución-CompartirIgual 4.0 Internacional. Para ver una copia de esta licencia, visite <http://creativecommons.org/licenses/by-sa/4.0/> o envíe una carta a Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.



Autor: Roberto Asplanato | rasplanato@hotmail.com | Fecha: 22/08/2025 | Montevideo – Uruguay.