



## **1er Examen Parcial de Inteligencia Artificial**

**Importante:** Lea atentamente todo el enunciado antes de comenzar a responder.

Como entrega del examen, deberá entregar en papel con todas las respuestas a las preguntas teóricas y un cuaderno de Google Colab (formato ipynb), donde estarán incluidos todos los detalles para entender el estudio que ha realizado.

Cuando se pide resolver el ejercicio en un cuaderno de Google Colab, muestre mediante gráficos y explicaciones acordes, las conclusiones a las que ha arribado luego del aprendizaje. Documente cada uno de los pasos del análisis para demostrar su entendimiento sobre el tema.

### **Condiciones de aprobación**

Para aprobar este examen, el estudiante deberá demostrar un conocimiento adecuado y equilibrado en **cada una de las siguientes áreas temáticas**. La calificación final no se promedia globalmente: se requiere una nota mínima del **60% en cada área** para considerar el examen aprobado.

Las áreas evaluadas son:

- Machine Learning
- Introducción a la IA y Algoritmos Genéticos

### **Parte Teórica**

#### **1. Preguntas a Desarrollar**

- Explique con sus propias palabras qué representan el sesgo y la varianza en el contexto de aprendizaje supervisado.
- Defina overfitting y underfitting, indicando cómo se manifiestan en el error de entrenamiento y validación. Mencione al menos dos estrategias concretas para mitigar cada uno de estos problemas.
- ¿Por qué la inteligencia artificial se volvió un fenómeno explosivo recién en los últimos años? Mencione y explique con claridad tres factores clave que contribuyeron a este auge.
- Describa en detalle una aplicación real y contemporánea que pueda ser resuelta mediante la utilización de Algoritmos Genéticos. Ilustre cómo se implementaría cada una de las etapas principales del algoritmo en este caso de uso.

#### **2. Verdadero/Falso – Justifique su respuesta**

- El ajuste de hiperparámetros como el grado de un polinomio o la profundidad máxima de un árbol no afecta la generalización del modelo.
- En K-fold Cross Validation, una vez terminadas las iteraciones, uno de los K modelos es utilizado para hacer predicciones sobre nuevos datos.
- La diferencia entre los AG tradicionales y los de evolución continua radica en que, en la evolución continua, se evalúa toda la población al finalizar un número continuo de generaciones.



d) Los algoritmos genéticos resuelven problemas que tienen una única solución óptima y bien definida, siempre garantizando su hallazgo.

**3. Multiple Choice:** Marcar las opciones correctas. **Cada pregunta** puede tener **una o más respuestas correctas**. Cada opción (A, B, C o D) es **completamente verdadera o completamente falsa**. No hay medias verdades.

Para que la pregunta sea considerada correcta es necesario identificar correctamente **TODAS** las respuestas correctas y no seleccionar opciones incorrectas.

**3a. Sobre Regularización en Modelos de Machine Learning:**

- A) Ridge Regression penaliza la suma de los cuadrados de los coeficientes.
- B) Lasso puede llevar a que algunos coeficientes sean exactamente cero.
- C) La regularización incrementa automáticamente el poder predictivo de un modelo en el training set.
- D) Si  $\lambda$  es cero, el modelo regularizado es equivalente al modelo sin regularizar.

**3b. Sobre Árboles de Decisión:**

- A) En clasificación, los árboles buscan maximizar la pureza de los nodos hijos.
- B) Una alta profundidad de árbol tiende a reducir el sesgo pero aumentar la varianza.
- C) La entropía mide cuán desordenadas o inciertas son las clases dentro de un nodo.
- D) Un árbol muy profundo tiende a generalizar mejor que un Random Forest.

**3c. Comparación: Regresión Lineal vs Regresión Logística:**

- A) La Regresión Lineal puede predecir valores fuera del rango [0,1].
- B) La Regresión Logística usa la función logística (sigmoide) para acotar su salida entre 0 y 1.
- C) Ambas minimizan la misma función de coste (MSE).
- D) La Regresión Logística es más adecuada que la Regresión Lineal para problemas de clasificación binaria.

**3.d. Al resolver distintos problemas con algoritmos genéticos, es necesario ajustar ciertos componentes del algoritmo. Lo que debe modificarse en comparación con otro problema resuelto con AG es...**

- A) ...la función de aptitud
- B) ...la cantidad de generaciones
- C) ...la representación de los individuos
- D) ...la etapa de mutación

**3.e. Indique cuáles de las siguientes expresiones están directamente relacionadas con el principio de selección natural en los algoritmos genéticos:**

- A) ...método exponencial
- B) ...mating pool
- C) ...aceptar/rechazar (rejection sampling)
- D) ...método del dardo



## **Parte Práctica**

### **1. Predicción de Éxito en Campañas de Crowdfunding**

Formas parte del equipo de analítica en una plataforma de crowdfunding donde emprendedores publican campañas para recaudar fondos. Tu equipo quiere desarrollar un modelo de Machine Learning para predecir si una campaña será exitosa (es decir, si alcanzará su objetivo de recaudación).

El objetivo del negocio es priorizar qué campañas mostrar en el homepage para maximizar la tasa de campañas exitosas. Como sólo pueden destacar 10 campañas en la homepage, se plantea que es preferible evitar destacar campañas que terminan fracasando, aunque eso signifique eventualmente dejar de mostrar algunas que sí hubieran resultado exitosas.

#### **¿Qué debes hacer?**

1. Realizar un breve análisis exploratorio del dataset (gráfico y analítico) y comentar tus conclusiones al respecto.
2. Realizar la limpieza del dataset que consideres necesaria y crear un pipeline para preprocesar correctamente los datos (valores nulos, outliers, codificación, escalado, etc.).
3. Entrenar al menos 3 modelos de clasificación optimizando hiperparámetros y buscando maximizar la métrica que consideres adecuada.
4. Reportar métricas de performance (recall, precisión, accuracy, AUC y matriz de confusión) para cada uno de estos modelos y seleccionar el mejor modelo bajo tu criterio.
5. Justificar y comentar todas tus decisiones y el código que utilices.

### **2. Algoritmos Genéticos - Diseño de Menú para un Restaurante**

Un nuevo restaurante en la zona de microcentro debe armar el menú del almuerzo y cena para los 7 días de la semana. Cuenta con 60 platos disponibles, cada uno con:

- Costo por porción (de 1 a 5)
- Popularidad estimada (de 5 a 10)
- Tipo de plato (italiano, carne, pescado, ensalada, etc.)

El objetivo del restaurante es seleccionar un menú semanal (14 platos, uno por día y turno) que maximice la diferencia entre popularidad y costo, considerando además ciertas penalizaciones por:

- Repetir un mismo plato dentro de la semana, y
- Repetir el mismo tipo de plato en comidas consecutivas (almuerzo y cena del



mismo día o cena de un día y almuerzo del día siguiente).

**¿Qué debes hacer?**

1. Crear, desde cero:
  - a. La función para inicializar la población.
  - b. La función de aptitud (fitness).
  - c. La función de selección de padres (elige un método y justifica por qué lo usaste).
2. Corregir la función de crossover y la función de mutación que se presentan con errores intencionales.
3. Integrar todas las funciones en el loop del algoritmo genético, ejecutarlo e imprimir:
  - a. El mejor fitness de un individuo en cada generación (de forma gráfica y analítica).
  - b. El mejor menú encontrado tras finalizar el loop.
  - c. El valor fitness del mejor menú.
4. Realizar el punto 3 pero aplicando elitismo. Comentar qué diferencias encuentra entre los resultados obtenidos en el punto 3 y el punto 4.
5. Justificar y comentar todas tus decisiones y el código que utilices.