

# Proyecto Final

## Ciencia de Datos Avanzada

### Instrucciones Generales (Obligatorias)

- Extensión: 5 a 7 páginas (sin contar anexos).
- El documento debe estar **muy bien redactado**, con claridad conceptual, coherencia lógica y precisión técnica.
- No deben existir **supuestos implícitos**. Esto implica que, desde la definición del problema hasta la interpretación de los resultados, toda decisión conceptual, metodológica y técnica debe explicarse y justificarse explícitamente.
- No se debe suponer que el lector conoce el contexto del problema.
- Toda variable debe definirse y justificarse.
- Toda métrica debe motivarse en función del impacto social.
- El Análisis Exploratorio de Datos (EDA) es **obligatorio**.
- La selección de variables y algoritmos es **obligatoria y debe justificarse**.
- Deben discutirse explícitamente las implicaciones éticas.
- Todo uso de herramientas de inteligencia artificial generativa (por ejemplo, ChatGPT u otras similares) debe declararse explícitamente, especificando el tipo de apoyo recibido y el alcance de su utilización.

# **1. Introducción y Justificación del Problema**

## **1.1. Definición del Problema**

- ¿Cuál es el problema social específico?
- ¿En qué contexto ocurre?
- ¿A quién afecta?
- ¿Por qué es relevante?

Evitar generalidades vagas o afirmaciones sin sustento.

## **1.2. Importancia Social**

- Consecuencias de no resolver el problema.
- Impacto potencial del modelo.
- Actores que utilizarán los resultados.

## **1.3. Objetivos**

- Objetivo general.
- Objetivos específicos.
- Decisiones concretas que informará el modelo.

# **2. Estado Actual y Brecha**

- ¿Cómo se aborda actualmente el problema?
- Limitaciones de enfoques existentes.
- Valor agregado del aprendizaje automático.

### **3. Datos y Análisis Exploratorio (Obligatorio)**

#### **3.1. Descripción de los Datos**

- Fuente.
- Periodo temporal.
- Unidad de análisis.
- Número de observaciones.
- Variables disponibles.

#### **3.2. Limpieza y Preparación**

- Tratamiento de valores faltantes.
- Manejo de outliers.
- Transformaciones realizadas.
- Ingeniería de variables.

#### **3.3. Análisis Exploratorio de Datos (EDA)**

Debe incluir:

- Estadísticas descriptivas.
- Distribuciones de variables clave.
- Correlaciones.
- Balance de clases (si aplica).
- Visualizaciones interpretadas.
- Identificación de patrones y posibles sesgos.

Se debe demostrar empíricamente la viabilidad del problema.

## **4. Definición del Problema de Aprendizaje Automático**

### **4.1. Tipo de Problema**

Clasificación, regresión, agrupamiento, etc., con justificación conceptual y empírica.

### **4.2. Variable Objetivo**

- Definición precisa.
- Método de construcción.
- Justificación conceptual.
- Limitaciones.

### **4.3. Variables Predictoras (Obligatorio)**

- Lista explícita de variables utilizadas.
- Justificación teórica.
- Evidencia empírica del EDA.
- Variables descartadas y razones.
- Tratamiento de colinealidad.

## **5. Modelos y Algoritmos (Obligatorio)**

Debe incluir al menos:

- Modelo baseline.
- Modelo intermedio.
- Modelo más flexible (si aplica).

Para cada modelo:

- Descripción técnica.
- Supuestos.

- Ventajas y desventajas.
- Riesgo de sobreajuste.
- Nivel de interpretabilidad.
- Justificación en el contexto social.

## 6. Estrategia de Validación

### 6.1. División de Datos

- Train / Validation / Test.
- Cross-validation.
- Validación temporal (si aplica).

### 6.2. Métricas de Evaluación

Las métricas deben justificarse según el impacto social.

Ejemplos:

- Accuracy.
- Precision / Recall.
- F1-score.
- AUC.
- RMSE / MAE.
- Métricas de equidad.

Debe explicarse qué tipo de error es más costoso socialmente.

## **7. Consideraciones Éticas**

- Privacidad.
- Sesgo algorítmico.
- Discriminación indirecta.
- Transparencia.
- Interpretabilidad.
- Rendición de cuentas (se refiere a la obligación institucional de asumir responsabilidad por las decisiones generadas por el modelo, garantizar su supervisión, permitir auditóreas y establecer mecanismos de corrección ante posibles errores o impactos adversos).

Incluir estrategias de mitigación.

## **8. Limitaciones**

- Limitaciones de datos.
- Restricciones metodológicas.
- Supuestos realizados.
- Problemas de generalización.

## **9. Recomendaciones de Implementación**

- Cómo se usaría el modelo.
- En qué decisiones concretas.
- Propuesta de piloto o evaluación futura.
- Indicadores de impacto.

## **Referencias**

[1] ...

## Rúbrica de Evaluación...

# Rúbrica de Evaluación (100 puntos)

## 1. Claridad y Redacción (15 puntos)

Criterio	Puntos
Redacción clara y formal	5
Ausencia de supuestos implícitos	5
Coherencia argumentativa	5

## 2. Definición del Problema Social (10 puntos)

Problema claramente definido	5
Justificación social sólida	5

## 3. Datos y EDA (20 puntos)

Descripción de datos	5
Limpieza rigurosa	5
Estadísticas descriptivas	5
Visualizaciones interpretadas	5

**Nota:** Sin EDA completo, la calificación máxima del proyecto será 50 puntos.

## 4. Selección de Variables (15 puntos)

Lista explícita de variables	5
Justificación teórica	5
Justificación empírica	5

## 5. Modelos y Algoritmos (15 puntos)

Modelo baseline	5
Modelos adicionales adecuados	5
Justificación técnica	5

## 6. Validación y Métricas (10 puntos)

Estrategia de validación correcta	5
Métricas alineadas al impacto social	5

## **7. Ética y Equidad (10 puntos)**

---

Identificación de riesgos	5
Estrategias de mitigación	5

---

## **8. Coherencia General e Impacto (5 puntos)**

---

Consistencia interna del proyecto	3
Relevancia práctica	2

---

## **Checklist para Evaluación Rápida**

Problema social claramente definido.

Justificación empírica del problema.

EDA completo.

Variables explícitamente listadas y justificadas.

Label claramente definido.

Al menos tres modelos implementados.

Métricas justificadas.

Discusión de errores socialmente costosos.

Análisis ético sólido.

Redacción clara y sin supuestos implícitos.