

Inteligencia Artificial Predictiva Departamento Nacional de Planeación (DNP)

Juan Pablo Ángel Quitian

Juan Sebastián Calderón Martínez

Joel Sebastián Bueno Medina

Manuel José Rivera Guzmán

Fundamentos de Inteligencia Artificial NRC: 81405

Corporación Universitaria Minuto de Dios, Facultad de Ingeniería, Ingeniería de Sistemas

18 de septiembre

Universidad Minuto de Dios

2025

Contenido

Descripción del problema elegido	3
Justificación	4
Diccionario de datos	5
DATASET	6

Descripción del problema elegido

Utilizamos el conjunto de datos Seguimiento de Proyectos del Departamento Nacional de Planeación (DNP) de Colombia, que contiene información sobre el avance físico, avance financiero, sector, entidad responsable y presupuesto de proyectos de inversión pública en Colombia.

El problema que buscamos resolver es la detección temprana de proyectos con alto riesgo de incumplimiento. Actualmente, identificar estos casos requiere revisiones manuales, lo que hace el proceso lento y poco escalable. Muchos proyectos solo se identifican como problemáticos cuando ya es demasiado tarde para intervenir de manera efectiva.

Nuestro objetivo es desarrollar un modelo predictivo de clasificación binaria que permita etiquetar automáticamente cada proyecto en una de dos categorías:

- **0: “Bajo riesgo”**: proyectos que avanzan según lo planeado.
- **1: “Alto riesgo”**: proyectos con señales de alerta, como avance físico bajo con alto gasto financiero, o falta total de avance.

Este modelo servirá como base para un sistema de alertas tempranas que permita a las entidades públicas actuar a tiempo.

Justificación

Elegimos este dataset y este problema por las siguientes razones:

1. **Impacto social:** Los proyectos de inversión pública afectan directamente a la ciudadanía. Contar con la capacidad de predecir su desempeño no solo ayuda a optimizar recursos, también mejora la transparencia y la rendición de cuentas.
2. **Calidad de los datos:** El dataset es oficial, abierto y contiene variables clave (avance físico, avance financiero, sector, entidad) que son ideales para entrenar un modelo predictivo.
3. **Alineación con el curso:**
 - Problema de aprendizaje supervisado (clasificación binaria).
 - Aplicación de técnicas como preprocesamiento como limpieza, normalización y feature engineering (por ejemplo, calcular la brecha entre avance físico y financiero).
 - Uso de algoritmos como Regresión Logística, Naive Bayes, Árboles de Decisión o Random Forest.
 - Evaluación del modelo con métricas claras: accuracy, precisión, recall, F1-score y matriz de confusión.

4. **Potencial para un bot predictivo:** El modelo entrenado puede integrarse en una herramienta automática (Inteligencia Artificial predictiva) que:
- Clasifique nuevos proyectos en tiempo real.
 - Capacidad de generación de reportes periódicos de proyectos en riesgo.
 - Permita a los gestores públicos enfocarse en donde más se necesita.
5. **Escalabilidad y ética:** Una solución así es escalable a nivel nacional y abre debates importantes sobre ética en IA: equidad, transparencia y mitigación de sesgos en los datos públicos.

Diccionario de datos

DICCIONARIO DE DATOS BASE DE DATOS DNP					
Campo	Tipo de dato	Longitud	Nulo	Clave	Descripción
Bpin	INT		NO	PK	Código BPIN del proyecto
AvanceFisico	DOUBLE		NO		Indicador que mide el avance real del proyecto con respecto a unas metas
AvanceFinanciero	DOUBLE		NO		Indicador que mide la ejecución del presupuesto del proyecto, comparando costos reales con los planificados
NombreProyecto	VARCHAR	255	NO		Nombre del proyecto
Horizonte	VARCHAR	30	NO		Recibe el valor de dos fechas una estipulando el inicio y la otra la fecha de finalización estimada del proyecto, el formato esperado es ("YYYY - YYYY")
SectorProyecto	VARCHAR	255	NO		Nombre del sector en el que esta enfocado el proyecto
EntidadResponsable	VARCHAR	255	NO		Entidad responsable de ejecutar el proyecto

DATASET

url: https://www.datos.gov.co/Econom-a-y-Finanzas/DNP-SeguimientoProyecto/7mxfbp6x/data_preview

El archivo del dataset fue entregado junto con este archivo, se realizo una limpieza para que los datos fueran legibles.