

# Digital NAO- Backlog

## Reto 5

**Desarrollo de algoritmos en modelos  
con Machine Learning**

Luis Alejandro Azúa Urrutia

NAO ID: 3165

18/02/2026

**AI Engineer Core**

- 1 **Directora de Comunicación Social**

**Como** directora de comunicacion **quiero un analisis claro de la calidad y funcionalidad de los datos del INEGI 2020** **Para** proporcionar una base solida en el entrenamiento de un modelo.

  - Desarrollar script para hacer un análisis exploratorio bivariado, multivariado de los datos del proyecto para uso posterior.
  - Identificar variables clave
  - Escribir reporte con hallazgos del analisis exploratorio.
- 2 **Lizbeth Torres, Analista de Datos**

**Como** experta en investigación clínica **Puedo** utilizar metodos de limpieza de datos y transformacion a formatos numericos **Para** que el modelo de regresión logística pueda procesarlos correctamente.

  - Script en python para transformar datos, variables numericas y categoricas de manera especifica.
- 3 **Ingenieria en Machine Learning**

**Como** Lizbeth, experta en ML **Puedo** supervisar el entrenamiento del prototipo base del modelo de regresión logisitica. **Para** obtener una aproximación de los resultados de una implementación de esta tecnologia en el problema actual.

  - Desarrollar script de entrenamiento para modelo de regresion logistica con sklearn que tome como variable objetivo a HeartDisease y lo demas como predictores.
- 4 **Empleo de tecnicas de regularización**

**Como** parte del equipo de Machine learning **Puedo** emplear mis conocimientos en tecnicas de regularización y balanceo de clases para evitar overfitting y minimizar los falsos negativos.

  - Evaluar metricas precision, recall, accuracy y f1-score en conjuntos de entrenamiento y prueba con modelo de sprint 2.
  - Evaluar metricas precision, recall, accuracy y f1-score en conjuntos de entrenamiento y prueba con modelo de sprint 3 balancing.
- 5 **Presentar resultados**

**Como** directora de comunicación **quiero** organizar los resultados y presentarlos por medio de un reporte final ante la secretaria de salud, de manera clara y comprensible.

  - Presentación ejecutiva de analisis de resultados obtenidos en scripts contrastados con los resultados de sprint 2.



	Requerimientos	Etapas	Estima...	Entregables	Estado	+
1	- Crear código para hacer un análisis exploratorio bivariado, multivariado de los datos del proyecto para uso posterior.	Etapas 1	4 horas	eda_heart_disease.ipynb	Done	
2	- Escribir reporte con hallazgos del analisis exploratorio	Etapas 1	1 hora	report_eda_heart_disease.pdf	Done	
3	- Script en python para transformar datos, variables numericas y categoricas de manera especifica.	Etapas 1	3 hora	processing_heart_disease.py processing_heart_disease.csv	Done	
4	- Crear repositorio de Github	Etapas 1	15 minutos	Repositorio GitHub Reto5.pdf	Done	
5	- Desarrollar script de entrenamiento para modelo de regresion logistica con sklearn que tome como variable objetivo a HeartDisease y lo demas como predictores.	Etapas 2	4 horas	training_model_heart_disease.py	In-progress	
6	- Evaluar metricas precision, recall, accuracy y f1-score en conjuntos de entrenamiento y prueba con modelo de sprint 2.	Etapas 2	2 horas	training_model_heart_disease.py	To do	
7	- Script de codigo para entrenar modelo de regresion logistica, ademas, ejecutar calibracion de hyper parametro C. Considerar validacion cruzada con esquema kFold de 5 hojas.	Etapas 3	por definir	regularization_model_heart_disease.py	To do	
8	- Evaluar metricas precision, recall, accuracy y f1-score en conjuntos de entrenamiento y prueba con modelo de sprint 3.	Etapas 3	2 horas	regularization_model_heart_disease.py	To do	
9	- Crear guion de python que cargue datos, procese, entrene modelo de regresion logistica. Implementar 2 tecnicas para desbalance.	Etapas 3	por definir	balacing_model_heart_disease.py	To do	
10	- Evaluar metricas precision, recall, accuracy y f1-score en conjuntos de entrenamiento y prueba con modelo de sprint 3 balancing.	Etapas 3	2 horas	regularization_model_heart_disease.py	To do	
11	- Presentación ejecutiva de analisis de resultados obtenidos en scripts contrastados con los resultados de sprint 2.	Etapas 3	por definir	results_models_heart_disease.pdf	To do	



# Digital NAO- Roadmap

## Reto 5

**Desarrollo de algoritmos en modelos  
con Machine Learning**

Luis Alejandro Azúa Urrutia

NAO ID: 3165

18/02/2026

**AI Engineer Core**

## Roadmap Reto 5

### Modelo Aprendizaje profundo para Chatbot

## Objetivos:

- Transformar y analizar los datos numéricos mediante un modelo estadístico. Basandose en datos del Censo Nacional de Poblacion y Vivienda 2020 de INEGI
- Dividir los datos y construir un modelo de regresión lineal utilizando Machine Learning para predecir la presencia de enfermedades cardiovasculares a través del efecto de diversas variables en la incidencia de enfermedades cardiovasculares.
- Diseñar un modelo de Machine Learning para analizar y predecir factores asociados a enfermedades cardiovasculares, con el fin de mejorar la efectividad de campañas de prevención.
- Generar datos rigurosos y un modelo de machine learning robusto.

# Planificación de sprint

Objetivos individuales a completar por sprint.

## **Sprint 1**

Elaborar el Backlog y Roadmap del proyecto; desarrollar un análisis exploratorio de datos, transformaciones de datos y un reporte de hallazgos.

## **Sprint 2**

Entrenar un modelo inicial de regresión logística, evaluando sus métricas y reportando los resultados.

## **Sprint 3**

Implementar técnicas avanzadas de regularización y manejo de desbalance en los datos para mejorar el modelo.

## **Final**

Integrar en una presentación de análisis y resultados en formato PDF y una videografía en formato MP4 los entregables de los tres Sprint.

≡ Entregables

📖 Requerimientos

🕒 Prioridad

+

1	<b>eda_heart_disease.ipynb.</b>	- Código para análisis exploratorio.	Alta
2	<b>report_eda_heart_disease.pdf</b>	- Reporte con hallazgos del analisis exploratorio	Alta
3	<b>processing_heart_disease.py</b> <b>processing_heart_disease.csv</b>	- Script en python para transformar datos, variables numericas y categoricas de manera especifica.	Media
4	<b>Repositorio GitHub Reto5.pdf</b>	- Liga a repositorio de Github	Alta
5	<b>training_model_heart_disease.py</b>	- Script de entrenamiento para modelo de regresion logistica con sklearn.	Media
6	<b>regularization_model_heart_disease.py</b>	- Script de codigo para entrenar modelo de regresion logistica, con calibracion de hyper parametro C.	Baja
7	<b>balacing_model_heart_disease.py</b>	- Guion de python que entrene modelo de regresion logistica. Implemtandoi 2 tecnicas para desbalance.	Baja
8	<b>results_models_heart_disease.pdf</b>	- Presentación ejecutiva de analisis de resultados obtenidos.	Baja

+



## Requerimientos

## Etapas

## Estima...



1	- Crear código para hacer un análisis exploratorio bivariado, multivariado de los datos del proyecto para uso posterior.	Etapas 1	4 horas
2	- Escribir reporte con hallazgos del analisis exploratorio	Etapas 1	1 hora
3	- Script en python para transformar datos, variables numericas y categoricas de manera especifica.	Etapas 1	3 hora
4	- Crear repositorio de Github	Etapas 1	15 minutos
5	- Desarrollar script de entrenamiento para modelo de regresion logistica con sklearn que tome como variable objetivo a HeartDisease y lo demas como predictores.	Etapas 2	4 horas
6	- Evaluar metricas precision, recall, accuracy y f1-score en conjuntos de entrenamiento y prueba con modelo de sprint 2.	Etapas 2	2 horas
7	- Script de codigo para entrenar modelo de regresion logistica, ademas, ejecutar calibracion de hyper parametro C. Considerar validacion cruzada con esquema kFold de 5 hojas.	Etapas 3	por definir
8	- Evaluar metricas precision, recall, accuracy y f1-score en conjuntos de entrenamiento y prueba con modelo de sprint 3.	Etapas 3	2 horas
9	- Crear guion de python que cargue datos, procese, entrene modelo de regresion logistica. Implemtar 2 tecnicas para desbalance.	Etapas 3	por definir
10	- Evaluar metricas precision, recall, accuracy y f1-score en conjuntos de entrenamiento y prueba con modelo de sprint 3 balancing.	Etapas 3	2 horas
11	- Presentación ejecutiva de analisis de resultados obtenidos en scripts contrastados con los resultados de sprint 2.	Etapas 3	por definir