

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

Historias de usuario

Requerimientos

+

1	<u>DIRECTORA DE COMUNICACIÓN SOCIAL</u> Como directora de comunicación quiero un análisis claro de la calidad y funcionalidad de los datos del INEGI 2020 Para proporcionar una base sólida 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 análisis exploratorio.
2	<u>LIZBETH TORRES, ANALISTA DE DATOS</u> Como experta en investigación clínica Puedo utilizar métodos de limpieza de datos y transformación a formatos numéricos Para que el modelo de regresión logística pueda procesarlos correctamente.	- Script en python para transformar datos, variables numéricas y categóricas de manera específica.
3	<u>INGENIERIA EN MACHINE LEARNING</u> Como Lizbeth, experta en ML Puedo supervisar el entrenamiento del prototipo base del modelo de regresión logística. Para obtener una aproximación de los resultados de una implementación de esta tecnología en el problema actual.	- Desarrollar script de entrenamiento para modelo de regresión logística con sklearn que tome como variable objetivo a HeartDisease y lo demás como predictores.
4	<u>EMPLEO DE TÉCNICAS DE REGULARIZACIÓN</u> Como parte del equipo de Machine learning Puedo emplear mis conocimientos en técnicas de regularización y balanceo de clases para evitar overfitting y minimizar los falsos negativos.	- Evaluar métricas precision, recall, accuracy y f1-score en conjuntos de entrenamiento y prueba con modelo de sprint 2. - Evaluar métricas precision, recall, accuracy y f1-score en conjuntos de entrenamiento y prueba con modelo de sprint 3 balancing.
5	<u>PRESENTAR RESULTADOS</u> Como directora de comunicación quiero organizar los resultados y presentarlos por medio de un reporte final ante la secretaría de salud, de manera clara y comprensible.	- Presentación ejecutiva de análisis 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.	Etapa 1	4 horas	eda_heart_disease.ipynb	Done
2	- Escribir reporte con hallazgos del análisis exploratorio	Etapa 1	1 hora	report_eda_heart_disease.pdf	Done
3	- Script en python para transformar datos, variables numéricas y categóricas de manera específica.	Etapa 1	3 hora	processing_heart_disease.py processing_heart_disease.csv	Done
4	- Crear repositorio de Github	Etapa 1	15 minutos	Repositorio GitHub Reto5.pdf	Done
5	- Desarrollar script de entrenamiento para modelo de regresión logística con sklearn que tome como variable objetivo a HeartDisease y lo demás como predictores.	Etapa 2	4 horas	training_model_heart_disease.py	In-progress
6	- Evaluar métricas precision, recall, accuracy y f1-score en conjuntos de entrenamiento y prueba con modelo de sprint 2.	Etapa 2	2 horas	training_model_heart_disease.py	To do
7	- Script de código para entrenar modelo de regresión logística; además, ejecutar calibración de hyper parámetro C. Considerar validación cruzada con esquema kFold de 5 hojas.	Etapa 3	por definir	regularization_model_heart_disease.py	To do
8	- Evaluar métricas precision, recall, accuracy y f1-score en conjuntos de entrenamiento y prueba con modelo de sprint 3.	Etapa 3	2 horas	regularization_model_heart_disease.py	To do
9	- Crear guion de python que cargue datos, procese, entrene modelo de regresión logística. Implementar 2 técnicas para desbalance.	Etapa 3	por definir	balancing_model_heart_disease.py	To do
10	- Evaluar métricas precision, recall, accuracy y f1-score en conjuntos de entrenamiento y prueba con modelo de sprint 3 balancing.	Etapa 3	2 horas	regularization_model_heart_disease.py	To do
11	- Presentación ejecutiva de análisis de resultados obtenidos en scripts contrastados con los resultados de sprint 2.	Etapa 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	eda_heart_disease.ipynb.	- Código para análisis exploratorio.	Alta
2	report_eda_heart_disease.pdf	- Reporte con hallazgos del análisis exploratorio	Alta
3	processing_heart_disease.py processing_heart_disease.csv	- Script en python para transformar datos, variables numéricas y categóricas de manera específica.	Media
4	Repositorio GitHub Reto5.pdf	- Liga a repositorio de Github	Alta
5	training_model_heart_disease.py	- Script de entrenamiento para modelo de regresión logística con sklearn.	Media
6	regularization_model_heart_disease.py	- Script de código para entrenar modelo de regresión logística, con calibración de hyper parámetro C.	Baja
7	balancing_model_heart_disease.py	- Guion de python que entrena modelo de regresión logística. Implementando 2 técnicas para desbalance.	Baja
8	results_models_heart_disease.pdf	- Presentación ejecutiva de análisis 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.	Etapa 1	4 horas	
2	- Escribir reporte con hallazgos del análisis exploratorio	Etapa 1	1 hora	
3	- Script en python para transformar datos, variables numéricas y categóricas de manera específica.	Etapa 1	3 horas	
4	- Crear repositorio de Github	Etapa 1	15 minutos	
5	- Desarrollar script de entrenamiento para modelo de regresión logística con sklearn que tome como variable objetivo a HeartDisease y lo demás como predictores.	Etapa 2	4 horas	
6	- Evaluar métricas precision, recall, accuracy y f1-score en conjuntos de entrenamiento y prueba con modelo de sprint 2.	Etapa 2	2 horas	
7	- Script de código para entrenar modelo de regresión logística, además, ejecutar calibración de hyper parámetro C. Considerar validación cruzada con esquema kFold de 5 hojas.	Etapa 3	por definir	
8	- Evaluar métricas precision, recall, accuracy y f1-score en conjuntos de entrenamiento y prueba con modelo de sprint 3.	Etapa 3	2 horas	
9	- Crear guion de python que cargue datos, procese, entrene modelo de regresión logística. Implementar 2 técnicas para desbalanceo.	Etapa 3	por definir	
10	- Evaluar métricas precision, recall, accuracy y f1-score en conjuntos de entrenamiento y prueba con modelo de sprint 3 balancing.	Etapa 3	2 horas	
11	- Presentación ejecutiva de análisis de resultados obtenidos en scripts contrastados con los resultados de sprint 2.	Etapa 3	por definir	
+				