

# TANI - Reporte Proyecto Desarrollo

ACH - Grupo 1

## Table of contents

<b>1. Introducción y Contexto</b>	<b>2</b>
1.1. Antecedentes del problema de desnutrición infantil . . . . .	2
1.2. Objetivos del proyecto . . . . .	2
1.3. Alcance y limitaciones . . . . .	2
1.4. Actores y beneficiarios clave (stakeholders) . . . . .	2
<b>2. Recepción y Consolidación de Datos</b>	<b>4</b>
2.1. Fuentes de datos recibidas . . . . .	4
2.2. Desafíos encontrados en los datos originales . . . . .	4
2.3. Proceso de consolidación . . . . .	4
2.4. Resultado de la consolidación . . . . .	5
<b>3. Análisis Exploratorio de Datos (EDA)</b>	<b>5</b>
3.1. Descripción general del dataset . . . . .	5
3.2. Evaluación de calidad post-consolidación . . . . .	5
3.3. Estadísticas descriptivas . . . . .	5
3.4. Visualizaciones iniciales . . . . .	5
3.5. Identificación y tratamiento de outliers . . . . .	5
3.6. Correlaciones y relaciones clave . . . . .	5
3.7. Insights preliminares . . . . .	6
<b>4. Patrones e Insights Relevantes</b>	<b>6</b>
4.1. Variables predictoras fuertes . . . . .	6
4.2. Subpoblaciones en mayor riesgo . . . . .	6
4.3. Hallazgos contextuales relevantes . . . . .	6
<b>5. Modelado Predictivo</b>	<b>6</b>
5.1. Definición del problema de negocio . . . . .	6
5.2. Hipótesis de modelado basadas en el EDA . . . . .	6
5.3. Modelos evaluados y resultados . . . . .	6

5.4. Selección del modelo final y justificación . . . . .	7
5.5. Evaluación de métricas (precisión, recall, AUC, F1-score) . . . . .	7
<b>6. Recomendaciones para la ONG</b>	<b>7</b>
6.1. Uso práctico del modelo en la toma de decisiones . . . . .	7
6.2. Acciones inmediatas basadas en insights . . . . .	7
6.3. Recomendaciones sobre mejora en recolección de datos . . . . .	7
<b>7. Conclusiones y Próximos Pasos</b>	<b>7</b>
7.1. Síntesis del proceso (de datos crudos a modelo accionable) . . . . .	7
7.2. Impacto esperado en la prevención de desnutrición . . . . .	7
7.3. Próximas fases sugeridas (nuevos datos, análisis longitudinal, validación en campo) . . . . .	7

## 1. Introducción y Contexto

La Asociación Taller de los Niños (TANI) es una ONG peruana con más de 45 años de experiencia dedicada a mejorar la calidad de vida de niños y familias en situación de vulnerabilidad, especialmente en los primeros años de vida, a través de un enfoque integral que combina educación, salud, nutrición y fortalecimiento comunitario. El estudio tiene el fin de realizar un análisis predictivo respecto al bajo desarrollo infantil y así fortalecer la toma de decisiones en salud y nutrición.

### 1.1. Antecedentes del problema de desnutrición infantil

Esta primera entrega tiene como objetivo presentar lo hallado en la información brindada por la ONG as is, además de mostrar un análisis univariado comentando los insights relevantes encontrados en la exploración inicial de las variables.

### 1.2. Objetivos del proyecto

### 1.3. Alcance y limitaciones

### 1.4. Actores y beneficiarios clave (stakeholders)

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 257178 entries, 0 to 257177
Data columns (total 27 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Fecha                 257178 non-null  datetime64[ns]
```

1	Nº_HC	257169	non-null	float64
2	Tipo_Paciente	257178	non-null	object
3	Categoría	257178	non-null	object
4	Sexo	257178	non-null	object
5	Edad	257178	non-null	object
6	Nº_Control	257177	non-null	float64
7	Peso	257171	non-null	object
8	Talla	257160	non-null	object
9	CabPC	214658	non-null	object
10	Diag_Nacimiento	257178	non-null	object
11	P/T	209486	non-null	object
12	T/E	209489	non-null	object
13	P/E	209483	non-null	object
14	Ganancia_Peso_Talla	215157	non-null	object
15	Dx_Nutricional	257177	non-null	object
16	CN-CA	90122	non-null	object
17	Mantiene_Diag_Fav/Desf	68654	non-null	object
18	Recuperado	9226	non-null	object
19	Lactancia	70546	non-null	object
20	Razón	18369	non-null	object
21	ACA	144557	non-null	object
22	Tam_para	5061	non-null	object
23	Tam_graha	4055	non-null	object
24	Tam_hb	13616	non-null	float64
25	Acompaña_control	257178	non-null	object
26	Enfermera	257173	non-null	object

dtypes: datetime64[ns](1), float64(3), object(23)

memory usage: 53.0+ MB

	features	types	non_null_counts
0	Fecha	datetime64[ns]	257178
1	Diag_Nacimiento	object	257178
2	Tipo_Paciente	object	257178
3	Categoría	object	257178
4	Sexo	object	257178
5	Edad	object	257178
6	Acompaña_control	object	257178
7	Nº_Control	float64	257177
8	Dx_Nutricional	object	257177
9	Enfermera	object	257173
10	Peso	object	257171

	features	types	non_null_counts
11	Nº_HC	float64	257169
12	Talla	object	257160
13	Ganancia_Peso_Talla	object	215157
14	CabPC	object	214658
15	T/E	object	209489
16	P/T	object	209486
17	P/E	object	209483
18	ACA	object	144557
19	CN-CA	object	90122
20	Lactancia	object	70546
21	Mantiene_Diag_Fav/Desf	object	68654
22	Razón	object	18369
23	Tam_hb	float64	13616
24	Recuperado	object	9226
25	Tam_para	object	5061
26	Tam_graha	object	4055

Nota: \* Para insertar bloque de código usar Shift+Cmd o Ctrl+I \* Para ejecutar bloque de código: Cmd+shift+enter

## 2. Recepción y Consolidación de Datos

### 2.1. Fuentes de datos recibidas

- Archivos internos de la ONG
- Fuentes externas (ej. OMS, INEI, ministerios)

### 2.2. Desafíos encontrados en los datos originales

- Inconsistencias de formato y nombres de variables
- Valores faltantes y atípicos
- Duplicados y registros incompletos

### 2.3. Proceso de consolidación

- Estandarización de variables y unidades
- Integración en un dataset maestro
- Documentación de reglas de negocio aplicadas

## **2.4. Resultado de la consolidación**

- Métricas de completitud
- Descripción del dataset final

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

### **3.1. Descripción general del dataset**

- Número de registros y variables
- Tipos de variables (categóricas, numéricas)

### **3.2. Evaluación de calidad post-consolidación**

- Completitud
- Consistencia
- Identificación de sesgos

### **3.3. Estadísticas descriptivas**

- Distribución de edad
- Distribución de peso y talla
- Indicadores nutricionales básicos

### **3.4. Visualizaciones iniciales**

- Histogramas y boxplots
- Comparaciones por género, edad, región

### **3.5. Identificación y tratamiento de outliers**

- Casos extremos detectados
- Criterios de imputación o exclusión

### **3.6. Correlaciones y relaciones clave**

- Matriz de correlación
- Variables asociadas a desnutrición

### **3.7. Insights preliminares**

- Patrones por subgrupos
- Posibles hipótesis sobre causas de riesgo

## **4. Patrones e Insights Relevantes**

### **4.1. Variables predictoras fuertes**

### **4.2. Subpoblaciones en mayor riesgo**

### **4.3. Hallazgos contextuales relevantes**

## **5. Modelado Predictivo**

### **5.1. Definición del problema de negocio**

### **5.2. Hipótesis de modelado basadas en el EDA**

### **5.3. Modelos evaluados y resultados**

- Regresión logística
- Random Forest
- XGBoost (u otro)

**5.4. Selección del modelo final y justificación**

**5.5. Evaluación de métricas (precisión, recall, AUC, F1-score)**

## **6. Recomendaciones para la ONG**

**6.1. Uso práctico del modelo en la toma de decisiones**

**6.2. Acciones inmediatas basadas en insights**

**6.3. Recomendaciones sobre mejora en recolección de datos**

## **7. Conclusiones y Próximos Pasos**

**7.1. Síntesis del proceso (de datos crudos a modelo accionable)**

**7.2. Impacto esperado en la prevención de desnutrición**

**7.3. Próximas fases sugeridas (nuevos datos, análisis longitudinal, validación en campo)**