**DESNUTRICION INFANTIL**

**DESCRIPCION DEL PROBLEMA**

La desnutrición infantil es una enfermedad que aparece como resultado del consumo insuficiente de alimentos en cantidad y calidad suficiente, de la aparición repetitiva de enfermedades infecciosas y de factores sociales que afectan el estado nutricional de la niñez.

**ETAPA 1. ENTENDIMIENTO DE LOS DATOS**

GENERACIÓN DEL CONJUNTO DE DATOS

El dataset de análisis es un conjunto de datos con observaciones obtenidas del diagnóstico médico realizado a niños menores de 10 años en condición de desnutrición, estos diagnóstico de enfermedades está clasificado y codificado a nivel internacional y se conoce como CI-10 (1), el número 10 corresponde a la versión actual.

En resumen cada vez que un médico escribe en una historia clínica de atención a un paciente debe asignar un código a los diagnósticos que elabora al atender un paciente. Habitualmente esta asignación se hace sin conocer el propósito de esa codificación. En este contexto se presentan las clasificaciones internacionales, enfatizando en aquellas que permiten codificar las enfermedades. Se describe la lógica interna, utilización internacional, ventajas e implementación en Colombia de esas clasificaciones sobre enfermedades.

El propósito principal de este dataset es predecir la desnutrición infantil para tratar de mejorar el estado de salud con el fin de recuperarlos de esta enfermedad. Permite el análisis exhaustivo de datos relevantes de la población infantil y ayuda en los métodos de tratamiento y seguimiento para mejorar el estado de salud de la población infantil con desnutrición.

DESCRIPCIÓN CAMPOS DEL DATASET

ID. Variable numérica. Hace referencia a identificación del registro y es un número consecutivo.

DEPARTAMENTO. Variable categórica y hace referencia al Departamento de Colombia donde residen los pacientes, la muestra del dataset hace referencia a los Departamentos de Arauca, Guajira y Santander.

TIPO DE PUNTO. Variable categórica nominal y hace referencia a la atención en salud en zonas rurales o zonas urbanas.

DISCAPACIDAD. Variable categórica nominal, en la discapacidad se identifica los siguiente tipos de salida: Sin Discapacidad, Física, Intelectual, Multiple, Sensorial.

CODIGO UNICO. Variable categórica de tipo alfanumérico compuesta de 20 caracteres que consta de las tres letras iniciales de los nombres y apellidos de los pacientes junto con la fecha de nacimiento concatenados. Para efectos del presente ejercicio y por políticas de protección de datos no se relacionan nombres ni datos de identificación personales, esta columna será reemplazado por ID.

EDAD EN AÑOS. Variable numérica continua que hace referencia a la edad del paciente expresada en número de años, en un rango de 0 a 10 años.

GENERO. Variable categórica nominal, el género hace referencia al sexo del paciente Masculino o Femenino.

DIAGNÓSTICO. Variable categórica nominal, el diagnóstico hace referencia a la enfermedad diagnosticada por el médico nutricionista, para efectos del presente análisis se tuvieron en cuenta las siguientes:

CODIFICACIÓN ENFERMEDADES RELACIONADAS CON DESNUTRICIÓN

|  |  |
| --- | --- |
| CODIGO | DESCRIPCION |
| E441 | DESNUTRICION PROTEICOCALORICA LEVE |
| E440 | DESNUTRICION PROTEICOCALORICA MODERADA |
| E46X | DESNUTRICION PROTEICOCALORICA, NO ESPECIFICADA |
| E568 | DEFICIENCIA DE OTRAS VITAMINAS |
| R639 | BAJO PESO |
| E45X | RETARDO DEL DESARROLLO DEBIDO A DESNUTRICION PROTEICOCALORICA |
| E41X | MARASMO NUTRICIONAL |
| E40X | KWASHIORKOR |
| E343 | ENANISMO, NO CLASIFICADO EN OTRA PARTE |

PESO\_KG. Variable numérica, hace referencia a el peso del paciente expresado en unidades Kilogramos.

TALLA\_CM. Variable numérica, hace referencia a la altura del paciente o talla, este dato es expresado en unidades centímetros.

NOTA. SE IDENTIFICARON 9 VARIABLES Y 834 REGISTROS, LAS OBSERVACIONES FUERON REALIZADAS EN PACIENTES MENORES DE 10 AÑOS EN UN PERIODO DE UN AÑO. (Ver archivo ¨Metadata\_Des\_Men10A¨ para mayor detalle).

**ETAPA 2: PREPARACIÓN DE LOS DATOS**

En esta etapa preparamos los datos para el modelo de regresión logística.

 Procedemos a codificar las variables categóricas con números para darle una categoría ordinal, los campos del dataset que consideraremos son los siguientes:

TIPO DE PUNTO. Campo que transformamos a tipo numérico, codificando en 0: “zonas rurales” o 1: “zonas urbanas”.

EDAD EN AÑOS. Campo de tipo numérico que hace referencia a la edad del paciente expresada en número de años.

GENERO. Campo que transformamos a tipo numérico, el género hace referencia al sexo del paciente 1: “Sexo masculino” o 2: “Sexo femenino”.

DISCAPACIDAD. Campo que transformamos a tipo categorica numérica, la discapacidad se clasifica en las siguientes 0:”Sin Discapacidad”, 1:”Física, Intelectual, Multiple, ó Sensorial”.

DIAGNÓSTICO. Campo de texto que transformamos a tipo variable categórica numérica esto sería la codificación de nuestra variable objetivo.

El diagnóstico hace referencia a la enfermedad diagnosticada por el médico nutricionista, para efectos del presente análisis se tuvieron en cuenta las siguientes:

0:” Desnutrición proteicocalorico leve”;

1:”Desnutrición proteicocalorico moderada”;

2:”Desnutrición proteicocalorico severo”

**ETAPA 3. MODELADO DE LOS DATOS**

(Ver archivo ¨Modelo\_Des\_men10A¨ para mayor detalle).