PROCESO DE ANALÍTICA DE DATOS APLICADO A LA DESNUTRICIÓN INFANTIL EN NIÑOS DE 0 A 5 AÑOS EN LA CIUDAD DE MEDELLÍN

Álvaro Andrés Loaiza Duque
William Guillermo Moreno Puerta
Juan David Rios Rios

ITM Institución Universitaria
Facultad de Ingeniería
Departamento de Sistemas
Medellín, Colombia
Año 2023

PROCESO DE ANALÍTICA DE DATOS APLICADO A LA DESNUTRICIÓN INFANTIL EN NIÑOS DE 0 A 5 AÑOS EN LA CIUDAD DE MEDELLÍN

Álvaro Andrés Loaiza Duque
William Guillermo Moreno Puerta
Juan David Rios Rios

Trabajo de grado presentado como requisito para optar al título de:

Especialista en Ingeniería de Software

Área: Sistemas e Informática

Director:

Jorge Iván Bedoya Restrepo

Línea de Investigación: Ingeniería de Software y Ciencias de los Datos

ITM Institución Universitaria

Facultad de Ingeniería

Departamento de Sistemas

Medellín, Colombia

Año 2023

Agradecimientos

Queremos expresar nuestro más sincero agradecimiento a nuestros familiares por el invaluable apoyo emocional que nos brindaron durante nuestra especialización, ello, fue fundamental para sobrellevar los desafíos y perseverar en nuestro camino académico.

También deseamos agradecer de manera especial a nuestros estimados docentes y asesor, quienes nos han proporcionado el conocimiento necesario y requerido para la elaboración de nuestro trabajo de grado. Su dedicación y orientación nos han guiado en cada paso del proceso, ayudándonos a alcanzar este importante logro en nuestra formación académica.

Estamos verdaderamente agradecidos por la confianza, el apoyo y la guía que hemos recibido de todas estas personas importantes en nuestro camino. Sin su ayuda, no habríamos podido llegar hasta aquí.

Resumen

El trabajo de grado se encuentra en el marco de la analítica de datos cuyo objetivo principal fue implementar un proceso de analítica de datos para el conocimiento de la desnutrición infantil de niños de 0 a 5 años en la ciudad de Medellín. Esto, con el fin de resolver la siguiente hipótesis: “Con el proceso de analítica de datos se podría obtener conocimiento de la desnutrición infantil de niños de 0 a 5 años en la ciudad de Medellín”. Por ello, para responder, primero se realizó una búsqueda de datos que abarquen la problemática con el fin de obtener un conjunto de datos. Luego, se definieron los requisitos funcionales y arquitectónicos en el cual se concretaron las herramientas para el proceso de limpieza y aplicación de la analítica y paralelamente definir la técnica a aplicar.

Posterior se aplica el paso a paso de limpieza y transformación de los datos y se le hace codificación a través de un lenguaje de programación de la técnica seleccionada y se evalúo la pertinencia del modelo aplicado con el fin de obtener recomendaciones en cara a la desnutrición infantil.

Todo lo anterior se aplicó bajo una metodología aplicada al trabajo de grado fue una adaptación hibrida entre el proceso de ingeniería de software y analítica de datos con CRISP_DM y CATALYST.

Finalmente, los resultados arrojaron unas mejoras para el proceso de analítica a pesar de que fue pertinente y unas recomendaciones en cara al reporte sobre desnutrición en primera infancia realizado por la secretaría de salud municipal.

Palabras Claves: Analítica Datos, Desnutrición Infantil, ETL, Regresión Lineal.

Abstract

The degree work is within the framework of data analytics whose main objective was to implement a data analytics process for the knowledge of child malnutrition in children from 0 to 5 years old in the city of Medellin. This, in order to solve the following hypothesis: "With the process of data analytics it would be possible to obtain knowledge of child malnutrition in children from 0 to 5 years old in the city of Medellin". Therefore, to answer, first a search for data covering the problem was conducted in order to obtain a set of data. Then, the functional and architectural requirements were defined in which the tools for the cleaning process and application of the analytics were specified and, in parallel, the technique to be applied was defined.

Subsequently, the step-by-step data cleaning and transformation process was applied and the selected technique was codified through a programming language and the relevance of the applied model was evaluated in order to obtain recommendations in the face of child malnutrition.

All the above was applied under a methodology applied to the degree work was a hybrid adaptation between the process of software engineering and data analytics with CRISP_DM and CATALYST.

Finally, the results showed some improvements for the analytical process although it was pertinent and some recommendations for the report on malnutrition in early childhood made by the municipal health secretariat.

Keywords: Data Analytics, child malnutrition, ETL, Linear Regression.

Contenido

Lista de figuras

Lista de tablas

Lista de abreviaciones

ICBF	Instituto Colombiano de Bienestar Familiar
RCCV	Red de Ciudades Como Vamos
ETL	Extraction, Transformation, Load
ONG	Organización no gubernamental
UNICEF	Fondo de las Naciones Unidas para la Infancia
KDD	Knowledge Discovery in Databases
CRISP-DM	Cross-Industry Standard Process for Data Mining
SEMMA	Sample, Explore, Modify, Model, Assess
MinTic	Ministerio de Tecnologías de la Información y las Comunicacione
MinSalud	Ministerio de Salud y Protección social
ENDES	Encuesta Demográfica de Salud

Capítulo 1

Introducción

El trabajo de grado se concentra en realizar de acuerdo con un método de análisis de datos, un software que permita identificar las causas o factores principales del problema de la desnutrición de niños de 0 a 5 años en la ciudad de Medellín, abordado desde informes tomados de las entidades tanto públicas como privadas tales como “Medellín Cómo Vamos”, el “Instituto Colombiano De Bienestar Familiar” (ICBF). A la vez, se han realizado análisis de la problemática desde otras disciplinas como la economía, medicina, sociología, enfermería, entre otras a nivel latinoamericano y local. En otros países del primer mundo ya se ha hecho uso de la analítica de datos. Sin embargo, en Colombia y en Medellín, a pesar de tener datos abiertos disponibles, no se ha hecho provecho de la analítica para identificar patrones y generar insumos que permitan a ONG’s y al gobierno formular políticas y planes de acción para atacar la problemática.

Con la implementación del proceso de analítica de datos aplicado en la metadata se podría obtener conocimiento de las principales causas que pueden ocasionar la desnutrición infantil en niños de 0 a 5 años de la ciudad de Medellín.

De acuerdo con los resultados obtenidos se definieron los requisitos funcionales y no funcionales, la caracterización de la metadata, el diseño arquitectónico y lógico del software, el desarrollo del código fuente, la evaluación, los resultados finales de la analítica de datos y su pertinencia. Este proyecto tiende a impactar un conocimiento sobre el método de analítica de datos que, de acuerdo con su pertinencia, puede ser de interés, utilizarse o no, como un insumo para las ONG’s y el gobierno Municipal, Departamental o Nacional y así poder atacar la problemática.

1.1 Justificación

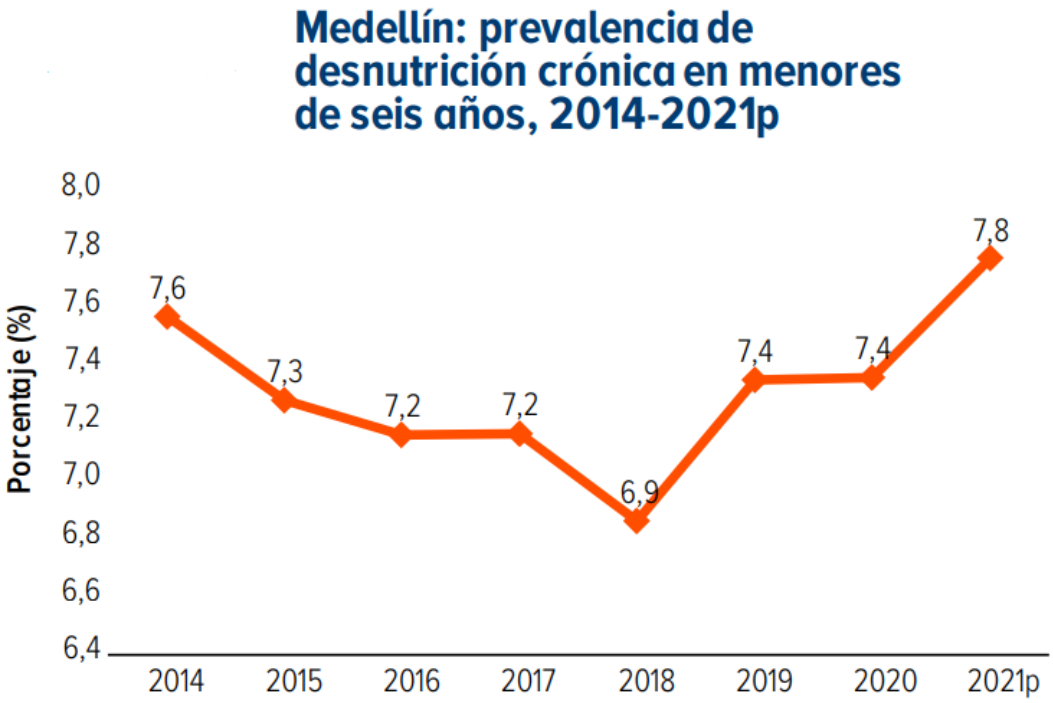
La UNICEF propuso como uno de los ejes centrales la nutrición infantil en los sistemas alimentarios nacionales, ya que, satisfacer las necesidades nutricionales específicas de los niños es crucial para lograr el desarrollo sostenible (UNICEF, 2019). Por otro lado, este mismo informe da a conocer los efectos de la desnutrición los cuales pueden ser:

- Crecimiento deficiente, infección y muerte
- Cognición deficiente, falta de preparación para la escuela, bajo rendimiento académico
- Un reducido potencial de ingresos más tarde
- Enfermedades crónicas en el futuro para el niño o niña

La figura 1, muestra la prevalencia de desnutrición crónica en menores de seis años en la ciudad de Medellín, en el período de 2014 a 2021. En 2014, el porcentaje de desnutrición era del 7.6%. A lo largo de los años, se observó una disminución constante, alcanzando su punto más bajo en 2018, con solo un 6.9% de desnutrición. Durante esos cuatro años, se logró una reducción del 0.7% en el porcentaje de desnutrición. Sin embargo, a partir de ese año, el porcentaje de desnutrición volvió a aumentar de manera constante, llegando al 7.8% en 2021. Esto representa un incremento del 0.8% en comparación con el porcentaje de desnutrición de 2018 y un aumento del 0.1% en comparación con el año 2014. Estos datos demuestran el crecimiento continuo de esta problemática a lo largo de los años.

Figura 1

Medellín: prevalencia de desnutrición crónica en menores de seis años, 2014-2021



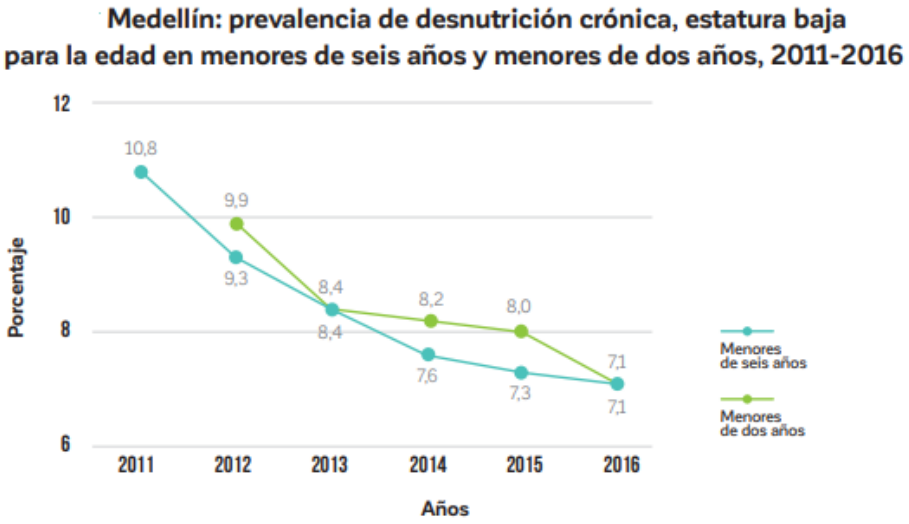
Nota. La figura muestra la prevalencia de desnutrición crónica en menores de seis años, 2014-2021. Fuente: (Medellín Cómo Vamos, 2021)

En la figura 2, se muestra la prevalencia de desnutrición crónica en menores de seis años en la ciudad de Medellín durante el período de 2011 a 2016. Es evidente que hubo una notable disminución en el porcentaje de niños con desnutrición crónica en ese lapso. En 2011, el porcentaje fue del 10.8%, reduciéndose en un 1.5% al año siguiente y alcanzando una disminución del 3.7% en 2016.

Sin embargo, al considerar la información presentada en las figuras 1 y 2, se puede concluir que, en la última década, el porcentaje de desnutrición infantil en la ciudad de Medellín solo ha disminuido en un 3%.

Figura 2

Medellín: prevalencia de desnutrición crónica, estatura baja para la edad en menores de seis años y menores de dos años, 2011-2016



Nota. La figura muestra la prevalencia de desnutrición crónica, estatura baja para la edad en menores de seis años y menores de dos años, 2011-2016 en la ciudad de Medellín. Fuente: (Medellín Cómo Vamos, 2021)

También, según lo consultado en bases de datos referentes como Science Direct, Scopus y Scholar Académico, no hay uso de la analítica de los datos para el tema de desnutrición infantil; sin embargo, en el ámbito internacional se han realizado acercamientos sobre el tema bien sea desde otras disciplinas u otras investigaciones relacionadas. En el ítem de estado del arte, los autores (Vesoulis, 2022) y (Khare, 2017) muestran la ventaja de usar las herramientas como Machine Learning, Inteligencia Artificial, Analítica de Datos para iterar con grandes cantidades de información y procesamiento. Los resultados serán un insumo para que las fundaciones o en el gobierno puedan tomar decisiones o busquen atacar la problemática, inclusive, utilizar la deducción lógica humana que permita entender el comportamiento de la desnutrición a partir de los datos procesados con las tecnologías 4.0.

En el contexto latinoamericano, se evidenció estudios relacionados de métodos de análisis sobre desnutrición infantil desde disciplinas como la economía (Martínez, 2006), medicina (Ayala Gaytán, 2015), enfermería (Tapia, 2021), nutrición (Carreño, 2017), entre otras. Sin embargo, no se hallaron trabajos relacionados desde la disciplina analítica de datos sobre el tema en el Medellín o en todo el país.

1.2 Planteamiento del problema

La desnutrición infantil en niños es un problema que se observa a nivel local, regional, nacional e internacional. La manifestación de este fenómeno puede ser resultado de causas tan amplias como los sistemas sociales, políticos, económicos y culturales de un país, así como de causas más específicas como la ingesta inadecuada de alimentos y enfermedades infecciosas. Los niños en la ciudad de Medellín no están exentos de esta realidad. Estudios realizados entre 2018 y 2021 por entidades tanto públicas como privadas como "Medellín Cómo Vamos" y la UNICEF, revelan que el índice de malnutrición en niños y niñas se debe a diversos factores causales, como el lugar de residencia, el nivel socioeconómico de las familias y el nivel educativo de los padres, entre otros.

HIPÓTESIS

Con el proceso de analítica de datos se podría obtener conocimiento de la desnutrición infantil de niños de 0 a 5 años en la ciudad de Medellín

1.3 Objetivos

General: Implementar un proceso de analítica de datos para el conocimiento de la desnutrición infantil de niños de 0 a 5 años en la ciudad de Medellín.

Específicos:

1. Lograr la Identificación de la metadata disponible que abarque la desnutrición infantil en la ciudad de Medellín en niños de 0 a 5 años para determinar su calidad y clasificación

2. Realizar la definición de los requisitos funcionales y arquitectónicos para el proceso de analítica de datos

3. Garantizar la caracterización los elementos que compone la metadata para definir el tipo de análisis de desnutrición infantil que se deberá abordar

4. Codificar, a través de un lenguaje de programación, el analizador de datos acorde a los requisitos funcionales y arquitectónicos definidos para solución propuesta

5. Evaluar el proceso de analítica con el fin de observar la validez del modelo, para obtener recomendaciones con respecto a la desnutrición infantil en niños de Medellín

1.4 Metodología propuesta

La mayoría de las metodologías para aplicar la ciencia de los datos contemplan que al inicio empiece el entendimiento del negocio, del problema u oportunidad de negocio y luego la fase de muestreo de datos como escenarios de punto de partida del proyecto. La revisión hecha por (Moine y otros, 2011), en la que comparan cuatro metodologías y sus puntos en común se resume en la siguiente tabla 1.

Tabla 1

Fases del proceso de minería de datos en cada modelo

FASES	KDD	CRISP-DM	SEMMA	CATALYST
Análisis y comprensión del negocio	X	X		X
Selección y Preparación de datos	X	X	X	X
Modelado	X	X	X	X
Evaluación	X	X	X	X
Implementación	X	X		X

Nota. Comparaciones de algunos modelos y sus etapas. Fuente (Moine y otros, 2011)

Se puede evidenciar que en todos los modelos se contempla la selección y preparación de los datos, posteriormente aplicar las técnicas y luego evaluar sus resultados, cuya retroalimentación permitirá conocer el desempeño del modelo o su utilidad que aporta al dominio del problema. En la última etapa de implementación, por lo general son análisis postmortem cuyo fin es valorar los resultados para mejoras en proyectos futuros.

Por lado, según el estudio llevado a cabo por (Moine J. , 2013), donde se compara los modelos de minería de datos de KDD, CRISP–DM, SEMMA y CATALYST, centrándose en las fases generales de selección y preparación de los datos; modelado; evaluación e implementación (ver figura 3). Se concluye que el enfoque de los procesos de minería de datos de KDD y SEMMA están más cercanos a ser un modelo que una metodología, ya que, según (Pressman, 2010) un modelo de procesos define qué hacer y también define cómo hacerlo.

Figura 3

Correspondencia entre las fases de cada metodología.

Fases	KDD	CRISP – DM	SEMMA	CATALYST
Análisis y comprensión del negocio	Comprensión del dominio de aplicación	Comprensión del negocio		Modelado del negocio
Selección y preparación de los datos	Crear el conjunto de datos	Entendimiento de los datos	Muestreo	Preparación de los datos
			Comprensión	
	Limpieza y pre-procesamiento de los datos	Preparación de los datos	Modificación	
	Reducción y proyección de los datos			
Modelado	Determinar la tarea de minería	Modelado	Modelado	Selección de herramientas y modelado inicial
	Determinar el algoritmo de minería			
	Minería de datos			
Evaluación	Interpretación	Evaluación	Valoración	Refinamiento del modelo
Implementación	Utilización del nuevo conocimiento	Despliegue		Comunicación

Nota. La figura muestra la Correspondencia entre las fases de los modelos y metodologías KDD, CRISP-DM, SEMMA y CATALYST. Fuente: (Moine J. , 2013)

Con lo anterior, este proyecto se llevó a cabo a través de una metodología personalizada que se adapte a cada uno de los objetivos específicos establecidos. Se tomarán como referencia las etapas y actividades de las metodologías CRISP-DM y CATALYST que sean más adecuadas para el proyecto según sus necesidades.

1.4.1 Plan de trabajo

Luego de reconocer las metodologías planteadas en el ítem anterior y teniendo en cuenta que el análisis y entendimiento del problema ya fue superada en la etapa de la presentación de la propuesta, la metodología del presente trabajo de grado se dividió en cinco fases que corresponden a: fase de exploración; fase de requisitos funcionales y arquitectónicos; fase de preparación de los datos; fase de

modelado y finalmente la fase de evaluación. En la figura 4 se muestra cada una de las fases.

Figura 4

Fases de la metodología propuesta

Nota. La figura muestra las fases de la metodología propuesta para la solución de los objetivos específicos. Fuente: Elaboración propia.

Fase 1: Exploración

En la fase de exploración, se lleva a cabo la recolección de datos necesarios y se realiza un proceso de familiarización con dichos datos. Durante esta etapa, pueden surgir las primeras hipótesis sobre posible información oculta que pueda existir y que sea relevante o no para el proyecto. Las actividades para realizar en esta fase son las siguientes:

Actividad 1.1 Recolectar los datos iniciales y describir los datos: En esta actividad, se lleva a cabo la búsqueda de datos necesarios para el proyecto en diversos portales, tanto de entidades gubernamentales como no gubernamentales. Se describe detalladamente el proceso seguido para obtener el conjunto de datos,

incluyendo la lista de bases de datos obtenidas, su formato, el tipo de variables presentes en cada una de ellas, así como la cantidad de registros recopilados en cada base de datos.

Actividad 1.2 Explorar los datos y verificar la calidad de los datos: En esta actividad, se realiza un análisis exhaustivo de los datos recopilados, centrándose en la distribución de estos, el comportamiento de las variables más relevantes, la consistencia en formato y estructura, y la verificación de la calidad de los datos, asegurando que estén libres de errores y ruidos. El objetivo principal es garantizar que los datos sean adecuados y de alta calidad para el análisis y entrenamiento del modelo de analítica de datos que se va a desarrollar.

Fase 2: Requisitos funcionales y arquitectónicos

La fase de requisitos funcionales y arquitectónicos se enfoca en comprender las necesidades para el proceso de analítica de datos y establecer una base sólida para el diseño y desarrollo de este, definiendo los objetivos y funcionalidades clave, así como la estructura general del sistema. Las actividades para realizar en esta fase son las siguientes:

Actividad 2.1 Requisitos funcionales: En esta actividad, se definen los diferentes requisitos funcionales del sistema y sus criterios de aceptación. Los requisitos funcionales son credos como historias de usuario.

Actividad 2.2 Diagrama de Clases: En esta actividad, se realiza la creación del diagrama de clases. Dicho diagrama se realiza posterior a la compresión de la distribución y la estructura de los datos previamente obtenidos en la fase de exploración.

Actividad 2.3 Modelo Entidad Relación: En esta actividad, se realiza la creación del modelo entidad relación. Dicho modelo se realiza posterior a la compresión de la distribución y la estructura de los datos previamente obtenidos en la fase de exploración.

Actividad 2.4 Selección de plataforma para el proceso de ETL: En esta actividad, se lleva a cabo un análisis exhaustivo de las distintas herramientas disponibles en el mercado para la generación del proceso ETL. Se investigan y evalúan las características, funcionalidades y capacidades de cada una de estas herramientas. Posteriormente, se realiza una selección cuidadosa de aquella herramienta que mejor se ajuste a las necesidades y requerimientos específicos del proyecto.

Actividad 2.5 Proceso ETL para el conjunto de datos: En esta actividad, se realiza el proceso de extracción, transformación y carga de información (ETL). La base para la creación de la ETL son los datos organizados y definidos en el modelo entidad relacional.

Actividad 2.6 Selección de técnica para el proceso de analítica de datos: En esta actividad, se lleva a cabo un estudio de las diversas técnicas de analítica de datos definidas en el marco teórico del capítulo 2. Se analizan detalladamente las características de cada una de estas técnicas, considerando aspectos como su aplicabilidad, precisión, velocidad de procesamiento y capacidad para extraer información relevante. Posteriormente, se realiza una cuidadosa selección de la técnica más adecuada, teniendo en cuenta las necesidades específicas del proyecto.

Actividad 2.7 Selección de herramienta para el proceso de analítica de datos: Se define el framework, el lenguaje de programación y la tecnología de acuerdo con las necesidades del proyecto.

Fase 3: Preparación de los datos

La fase de preparación de datos engloba las actividades necesarias para procesar los datos y construir el conjunto final de datos sobre el cual se aplicarán las técnicas de minería. Durante esta etapa, se realizan diversas tareas para garantizar que los datos estén en condiciones óptimas y sean adecuados para el análisis. Las actividades para realizar en esta fase son las siguientes:

Actividad 3.1 Proceso de extracción, transformación y carga (ETL): En esta etapa, se lleva a cabo la selección de los datos que serán utilizados en el análisis, tanto en términos de filas como de columnas. Se realiza un proceso de identificación y extracción de los datos relevantes para el estudio en cuestión. Posteriormente, se procede a realizar la limpieza de los datos, con el objetivo de mejorar su calidad y eliminar posibles errores, duplicados o valores atípicos que puedan afectar los resultados del análisis.

Una vez completada la limpieza, se lleva a cabo la construcción de nuevos datos derivados de los datos disponibles. Esto implica la creación de variables adicionales o el cálculo de medidas específicas que sean relevantes para el análisis y que proporcionen información adicional.

Puede ser necesario integrar datos provenientes de diferentes tablas o registros para realizar un análisis completo. En esta etapa, se realiza la integración de los datos, asegurando la coherencia y la consistencia de la información combinada.

Por último, se realiza el formateo de los datos, adecuando su estructura y presentación según las necesidades del análisis y las herramientas utilizadas. Esto puede incluir la normalización de formatos, la conversión de tipos de datos y la estandarización de unidades de medida, entre otros aspectos, para asegurar la consistencia y facilitar la interpretación de los datos.

Fase 4: Modelado

En esta fase, se aplican una variedad de técnicas y algoritmos de minería de datos al conjunto de datos con el fin de descubrir información oculta y patrones implícitos en ellos. Estas técnicas y algoritmos están diseñados para analizar los datos de manera sistemática y revelar conocimientos valiosos que no son fácilmente visibles a simple vista. Las actividades para realizar en esta fase son las siguientes:

Actividad 4.1 Visualización y análisis de los datos para la construcción del modelo: En esta actividad, se realiza el análisis de las diferentes variables con el objetivo de examinar su distribución y determinar cuáles son las variables que pueden generar ruido para el modelo a construir.

Actividad 4.2 Distribución y correlación de los datos para la construcción del modelo: En esta actividad, se lleva a cabo el análisis de distribución y correlación de los datos con el objetivo de identificar aquellos que son más relevantes para su interpretación dentro del modelo de analítica de datos seleccionado.

Fase 5: Evaluación

En esta fase se realiza el análisis de los patrones obtenidos en relación con los objetivos establecidos y al modelo de analítica de datos construido. Es importante resaltar que esta fase no solo se centra en la interpretación de los patrones obtenidos, sino también en la evaluación de la eficacia y utilidad del modelo de analítica de datos construido. Se examina si el modelo cumple con los objetivos establecidos, si es capaz de proporcionar insights relevantes y si se ajusta adecuadamente a los datos analizados. Las actividades para realizar en esta fase son las siguientes:

Actividad 5.1 Evaluación del modelo de analítica de datos: En esta actividad, se procede a la evaluación del modelo de analítica de datos seleccionado, utilizando como parámetros de entrada las variables previamente identificadas como relevantes para la interpretación del modelo.

Durante esta actividad, se deben de realizar pruebas exhaustivas del modelo utilizando diferentes combinaciones de las variables de interés. El objetivo principal de esta evaluación es determinar la efectividad y la capacidad predictiva del modelo al ingresar las variables seleccionadas como entrada. Con el fin de lograr analizar el rendimiento del modelo en términos de su capacidad para generar resultados coherentes y útiles.

Actividad 5.2 Evaluación del del modelo de analítica de datos haciendo uso de Otros criterios disponibles: En esta actividad, se deben aplicar criterios adicionales para evaluar el desempeño del modelo más allá de los parámetros iniciales. Estos criterios pueden incluir métricas de rendimiento, como precisión, exactitud, sensibilidad y especificidad, entre otros.

Capítulo 2
Desnutrición y Analítica De Datos: Marco Teórico y Estado del Arte.

En este capítulo se proporciona el marco teórico y estado del arte de los temas abordados. En la primera sesión se brinda el concepto de desnutrición y sus tipos. Adicional, se detallan las manifestaciones clínicas que se presentan en los niños durante la primera infancia. Así mismo, se da la noción de ciencia de los datos y algunas técnicas que permitirán aplicar análisis a una metadata disponible. Finalmente, se muestra algunas aproximaciones de análisis de desnutrición desde otras disciplinas y se exploran herramientas similares en el tema de salud.

2. 1 Marco teórico

2.1.1 Desnutrición

De acuerdo con el Ministerio de Salud (Minsalud, 2016), define la desnutrición como enfermedad de origen social, como la última expresión de situación de inseguridad alimentaria y nutricional de una población, que afecta principalmente a los niños y a las niñas. Se caracteriza por un deterioro de la composición corporal y alteración sistemática de las funciones orgánicas y psicosociales que en algunos casos son irreversibles.

Entre las principales causas de la desnutrición (Minsalud, 2016), se encuentran: el consumo insuficiente en cantidad y calidad de alimentos, por ejemplo, la ausencia o inadecuada lactancia materna y las malas prácticas en alimentación complementaria. También las enfermedades infecciosas y otros factores de riesgo como el bajo peso materno y el bajo peso y talla al nacer. Así mismo, están

relacionados con la desnutrición los determinantes sociales como; el bajo nivel educativo de los padres, los limitados ingresos económicos de la familia, las condiciones insalubres de la vivienda, las necesidades básicas insatisfechas, el hacinamiento, el bajo acceso a agua apta para consumo humano, el maltrato, el abandono entre otros.

2.1.1.1 Tipos de desnutrición

a) Peso Bajo para la Talla o Desnutrición Aguda: Está asociada a una pérdida de peso reciente y acelerada u otro tipo de incapacidad para ganar peso dada en la mayoría de los casos, por un bajo consumo de alimentos o la presencia de enfermedades infecciosas.

b) Retraso del Crecimiento: También conocida como desnutrición crónica está asociado a problemas prolongados y persistentes (de larga duración) que afectan negativamente el crecimiento infantil.

c) Deficiencias de Micronutrientes: Se producen cuando las personas no tienen acceso a alimentos fuentes de éstos tales como frutas, verduras, carnes y alimentos fortificados; en general, se debe a su alto costo o no están disponibles a nivel local. Las deficiencias de micronutrientes aumentan el riesgo de presentar enfermedades infecciosas y de morir por diarrea, sarampión, malaria y neumonía; las cuáles a su vez, son las 10 principales causas de morbilidad a nivel mundial. (Organization, 2007)

2.1.1.2 Manifestaciones clínicas de la desnutrición aguda severa (Minsalud, 2016)

a) Marasmo: Se caracteriza por atrofia severa de la masa grasa y muscular, los cuales el cuerpo ha utilizado como fuente de energía, dejando “los huesos forrados en la piel”.

b) Kwashiorkor: Se caracteriza esencialmente por el edema bilateral (que suele comenzar en los pies y piernas), disminución del peso corporal que

se encuentra enmascarado por el edema, y puede estar acompañado de erupciones en la piel y/o cambios en el color del pelo (de color grisáceo o rojizo) los cuales están asociados a deficiencias nutricionales específicas.

c) Marasmo – kwashiorkor: Caracterizado por una combinación de emaciación grave y edema bilateral. Esta es una forma muy severa de desnutrición aguda

De acuerdo con el trabajo “Factores asociados a la desnutrición en niños menores de 5 años” (Escobar, 2014), define los tipos de desnutrición cómo:

a) Desnutrición Aguda: Se manifiesta por bajo peso en relación con la talla del individuo, el cual se origina por una situación reciente de falta de alimentos o una enfermedad que haya producido una pérdida rápida de peso. Este tipo de desnutrición es recuperable, sin embargo, de no ser atendida oportunamente pone en alto riesgo la vida del individuo

b) Desnutrición Crónica o Retardo del Crecimiento: Se manifiesta por una baja talla de acuerdo con la edad del individuo, a consecuencia de enfermedades recurrentes y/o una ingesta alimentaria deficiente y prolongada. Este tipo de desnutrición disminuye permanentemente las capacidades físicas, mentales y productivas del individuo, cuando ocurre entre la gestación y los treinta y seis (36) meses.

c) Retardo en Niños Menores de 5 Años: La evaluación del crecimiento físico de los niños menores de 5 años, con edades comprendidas entre los 0 años a 4 años 11 meses, mediante el indicador talla para la edad ha permitido contar con información primaria sobre el estado nutricional de la población. Los censos de talla en niños menores de 5 años se convierten en instrumentos adicionales para dar respuesta a las necesidades de información a mediano plazo, respecto a la cuantificación de los logros en

desarrollo humano como resultado de acciones sociales asociadas, sostenidas y de cambios económicos.

Paralelamente, otra definición importante a conocer en esta propuesta es el análisis de datos: ciencia de la informática que corresponde a analizar datos sin procesar para sacar conclusiones sobre esa información. Muchas de las técnicas y procesos de análisis de datos se han automatizado en procesos mecánicos y algoritmos que funcionan con datos sin procesar para el consumo humano (Frankenfield J. , 2022).

2.1.2 Analítica de datos

Minería de datos: Es el conjunto de técnicas y tecnologías las cuales permiten examinar grandes bases de datos de manera automática, con el objetivo de encontrar patrones similares que ayuden a comprender el comportamiento de estos. La minería de datos tiene como objetivo comprender las grandes cantidades datos, para que estos puedan ser utilizados como insumo en cuanto a la generación de conclusiones o resultados que contribuyan a la mejora de procesos, lo cual hace que se puedan alcanzar objetivos definidos en la materia (Vallejo Ballesteros y otros, 2018).

Algunos modelos de la Minería de Datos (Perez Lopez & Santin Gonzalez , 2007)**:**

a) KDD: Proceso de obtención de conocimiento, se inicia con la recopilación e integración de la información a partir de datos iniciales, la fase inicial es la más importante ya que de ella depende la sucesión correcta de las siguientes, la información obtenida se encuentra en diferentes fuentes, sean internas y externas. Consta de siete etapas:

i. Selección.
ii. Exploración
iii. Limpieza.
iv. Transformación.
v. Minería de datos.
vi. Evaluación.
vii. Difusión.

b) CRISP-DM: Cross-Industry Standard Process for Data Mining. Esta metodología plantea el ciclo de vida de un proyecto de minería de datos, que contiene seis etapas, comienzan en una buena comprensión y conocimiento del negocio y la necesidad del proyecto y concluye con el despliegue de la solución que cumple la necesidad específica del negocio. Estas etapas son secuenciales por naturaleza, pero con un gran enfoque en la realimentación. Consta de seis etapas:
   i. Comprensión del negocio
   ii. Comprensión de los datos
   iii. Preparación de los datos
   iv. Modelado
   v. Evaluación
   vi. Despliegue.

c) SEMMA: Proceso de seleccionar, explorar, modificar, modelizar, y valorar grandes cantidades de datos con la tarea de descubrir, clasificar o asociar patrones desconocidos los cuales pueden utilizarse como comparación respecto a modelos similares o de iguales funciones. Consta de cinco etapas:
   i. Selección.
   ii. Exploración.
   iii. Modificación.
   iv. Modelización

v. Valoración.

d) CATALYST: Conformado por dos modelos, modelo de negocio, guía para desarrollo y construcción de modelo para resolver el problema u oportunidad de negocio (MII) y modelo de explotación de información, guía para la ejecución y realización de modelos de minería de datos basados en el modelo (MIII).

MII (Modelo de negocio) consta de cinco etapas:

i. Dato
ii. Oportunidad.
iii. Prospectiva
iv. Definido
v. Estratégico

MIII (Modelo de explotación de información) consta de cinco etapas:

i. Preparación.
ii. Selección.
iii. Refinar
iv. Implementar
v. Comunicación.

2.1.2.1 Tipos de Análisis de Datos (Frankenfield J. , 2022):

a) Análisis Descriptivo: Examina los datos y analiza los acontecimientos pasados para saber cómo abordar el futuro. La analítica descriptiva examina el rendimiento pasado y entiende ese rendimiento al extraer datos históricos para buscar las razones detrás del éxito o el fracaso del pasado, ya sea a través de medidas (estimadores), gráficas o tablas en donde se pueda apreciar claramente el comportamiento y las tendencias de la información recopilada. Casi todos los informes de gestión, tales como ventas, marketing, operaciones y finanzas, utilizan este tipo de análisis

post-mortem. Lo cual nos permite afirmar que la razón principal de este análisis es tratar de responder a la pregunta ¿Qué está pasando?

b) Análisis Diagnóstico: Es la determinación de los datos necesarios y los métodos útiles para recolectar algún tipo de información dentro de la empresa. La recolección y el análisis de datos es una de las actividades más complejas del desarrollo organizacional. Incluye técnicas y métodos para describir el sistema organizacional y las relaciones entre sus elementos o subsistemas, así como los modos de identificar problemas y temas importantes.

c) Análisis Predictivo: Es una parte de la analítica avanzada que se usa para hacer predicciones sobre sucesos futuros desconocidos. Utiliza diversas técnicas de la minería de datos para reunir toda la información tecnológica, la gestión y el proceso de construcción empresarial para elaborar predicciones de cara al futuro.

d) Análisis Prescriptivo: Sintetiza automáticamente grandes datos, ciencias matemáticas, reglas de negocio y machine learning para hacer predicciones y luego sugiere opciones de decisión para aprovechar las predicciones.

Hay varios métodos analíticos y técnicas (Frankenfield J. , 2022) diferentes que los analistas de datos pueden usar para procesar datos y extraer información. Algunos de los métodos más populares se enumeran a continuación.

a) Análisis de Regresión: Implica analizar la relación entre las variables dependientes para determinar cómo un cambio en una puede afectar el cambio en otra.

b) El Análisis Factorial: Implica tomar un gran conjunto de datos y reducirlo a un conjunto de datos más pequeño. El objetivo de esta maniobra es intentar descubrir tendencias ocultas que de otro modo habrían sido más difíciles de ver.

c) El Análisis de Cohortes: Es el proceso de dividir un conjunto de datos en grupos de datos similares, a menudo divididos en un grupo demográfico de clientes. Esto permite a los analistas de datos y otros usuarios de análisis de datos profundizar en los números relacionados con un subconjunto específico de datos.

d) Las Simulaciones de Monte Carlo: Modelan la probabilidad de que sucedan diferentes resultados. A menudo utilizadas para la mitigación de riesgos y la prevención de pérdidas, estas simulaciones incorporan múltiples valores y variables y, a menudo, tienen mayores capacidades de pronóstico que otros enfoques de análisis de datos.

e) El Análisis de Series Temporales: Realiza un seguimiento de los datos a lo largo del tiempo y solidifica la relación entre el valor de un punto de datos y la ocurrencia del punto de datos. Esta técnica de análisis de datos generalmente se usa para detectar tendencias cíclicas o para proyectar pronósticos financieros.

2.1.2.2 Técnicas predictivas de analítica de datos (Perez Lopez & Santin Gonzalez , 2007).

a) Regresión lineal: Es una técnica de análisis de datos que predice la estimación de datos desconocidos mediante el uso de otra estimación de datos relacionados. formar matemáticamente la variable desconocida o dependiente y la variable conocida o independiente como una ecuación lineal. La regresión lineal, se ajusta una línea recta a los datos revelados para estimar o determinar los valores de la variable dependiente en servicio

de los valores de las variables independientes. La línea recta se determina a través de un proceso de minimización de errores, utilizando el método de los mínimos cuadrados. La técnica de regresión lineal es relativamente simple y proporciona un método matemático y estadístico fácil de interpretar para generar predicciones.

b) Regresión logística: Es una técnica de análisis de datos que utiliza la matemática para encontrar las relaciones entre dos factores de datos. Después utiliza la relación para predecir el valor de uno de esos factores apoyándose en el otro, la predicción tiene un número fijo de resultados, como un sí o un no. La regresión logística es una técnica destacable en el área de la inteligencia artificial y el machine learning. Los modelos de machine learning creados por medio de la regresión logística ayudan a las organizaciones a obtener información procesable de acuerdo de sus datos empresariales. Las mismas organizaciones usan esta información para el análisis predictivo con el fin de reducir costos en su operación, aumentar la eficiencia y escalar más rápido.

c) El clustering: Es una técnica de machine learning fundamentado en el análisis estadístico que se emplea para examinar los datos en entornos Big Data. El clustering consiste en agrupar ítems en grupos con características similares que se conocen como clústeres, el objetivo es reconocer patrones, aunque también se utiliza en tareas de segmentación. Los clústeres están formados por una recopilación de objetos o datos similares, pero con apariencia que los diferencien de otros objetos de datos que forman parte de un clúster independiente. el clustering se puede aplicar prácticamente en todos los sectores.

d) Reglas de Asociación: Es un conjunto de técnicas que posibilitan establecer relaciones con la finalidad de descubrir eventos que aporten valor dentro de las variables que facilitan los datos que son muy grandes.

Las reglas de asociación tienen como misión encontrar los elementos y producir las normas debidas, acatando una clasificación para predecir cualquier atributo o combinación de los mismos atributos. Tanto en la minería de datos como el aprendizaje automático, las reglas de asociación se utilizan para analizar la información dentro de un determinado conjunto de datos. Esta técnica utiliza diferentes algoritmos que genera y testean distintas pautas.

2.1.3 Ingeniería de software

La ingeniería de software es una rama de la ingeniería que comprende los aspectos de la realización del software, desde la etapa inicial de la definición del sistema, hasta el mantenimiento de este. La ingeniería de software es prácticamente la aplicación del conocimiento científico, de acuerdo con el diseño y posterior construcción del producto.

a) Análisis: Fase de obtener las necesidades del usuario que requiere el sistema y las cuales deben cumplir y satisfacer por medio del funcionamiento.

b) Diseño: Fase de la elaboración del esquema en el cual se contempla lo necesario para que el sistema funcione de acuerdo a lo especificado, adicional se debe organizar su construcción para obtener una mejor optimización de los recursos.

c) Codificación: Fase de la producción y la materialización, en la cual se construirán los elementos mediante las herramientas y lenguajes de programación necesarias para el correcto funcionamiento del producto o resultado.

d) Pruebas: Fase de verificación y validación del funcionamiento del sistema, en esta etapa se utilizan diferentes tipos de métodos que buscan evaluar cada uno de los componentes desarrollados y realizar su completa ejecución.

e) Mantenimiento: Fase de atención la cual se pueden realizar cambios, corregir errores no detectados en las fases anteriores, introducción de mejoras y evolución, todo esto como soporte del producto.

2.2 Estado del arte

Según lo consultado en bases de datos tales como Science Direct, Scopus y Scholar Académico como no hay uso de la analítica de los datos para el tema de desnutrición infantil; sin embargo, en el ámbito internacional se han realizado acercamientos sobre el tema bien sea desde otras disciplinas u otras investigaciones relacionadas. A continuación, mencionamos los más relevantes:

Improving child health through Big Data and data science (Vesoulis, 2022)**:** Propone la recolección de diferentes fuentes de datos y abarca seis características denominadas 6V para abarcar la big data: Volumen, Valor, Veracidad, Variabilidad, Velocidad, y Variedad y unirlas con unas propiedades cuantitativas clasificadas como eje de los datos de la salud: Tamaño de muestra, fenotipo, seguimiento longitudinal, interacción entre sujetos, heterogeneidad y diversidad, estandarización de datos y conexión entre los datos. Finalmente, muestra algunos softwares de analítica de datos que se han implementado tales como PEDSNet PhysioNet y Genomic Information Commons (GIC).

Investigation of Nutritional Status, Children based on Machine Learning Techniques using Indian Demographic and Health Survey Data (Khare, 2017)**:** La desnutrición es la principal causa de mortalidad infantil entre los países en desarrollo tales como EEUU, incluida la India. Este estudio diseña un modelo de

predicción para la desnutrición basado en el enfoque de aprendizaje automático, utilizando las características disponibles en el conjunto de datos de la Encuesta Demográfica y de Salud de la India (IDHS). Posteriormente, se realizó el análisis en dos fases: la primera, en el cual se realizó preselección de datos usando aprendizaje automático. En la segunda, se realizó una regresión logística nominal para identificar las probabilidades a partir de las características tales como historial de nacimiento, embarazo, parto, lactancia, antropometría e identificación del hogar. El documento contribuye a explorar las posibilidades de utilizar la inteligencia artificial para identificar posibles correlatos de desnutrición. Las funciones utilizadas combinan y agrupan diferentes condiciones poblacionales de los niños para tratar de identificar las causas más importantes de la desnutrición. La combinación de la inteligencia artificial y humana, ambas son útiles para llegar a la selección de características importantes y decisiones políticas en cuanto al tratamiento de esta problemática.

Desnutrición infantil en menores de cinco años en Perú: tendencias y factores determinantes (Sobrino y otros, 2014)**:** Analiza indicadores nutricionales de menores de 5 años de la Encuesta Demográfica y de Salud Familiar ENDES del año 2011 y su evolución a partir de datos del mismo de los años 2000, 2005 y 2008 mediante datos como talla-edad-peso. Este articulo encontró que factores como el sexo, edad, zona de residencia, escolaridad de los padres pueden ser consecuencias de la desnutrición infantil y no solo como un problema de carácter alimentario.

La primera infancia de Medellín está en riesgo (Medellín Cómo Vamos, 2021)**:** Este informe de la alianza interinstitucional privada ‘Medellín Como Vamos’ dio a conocer dos datos relevantes acerca de la desnutrición infantil: Primero, cada vez menos bebés son alimentados con leche materna en sus primeros seis meses de vida. En 2021 el 7,8% de los niños que asistieron al control de crecimiento y desarrollo tenía desnutrición crónica, la más alta en ocho años. Esto podría ser un indicio de una problemática aún más grave que lo que nos muestran los datos. Y

Segundo, En 2021 se registró la mayor proporción de bajo peso al nacer en los últimos ocho años.

ENSIN: Encuesta Nacional de Situación Nutricional 2015 (Minsalud, 2016)**:** Esta encuesta halló que en desnutrición crónica (retraso en talla para la edad) en un 10.8% ubicándose por encima del promedio en Latinoamérica (9,9%) demostrando que no alcanzó los Objetivos del Milenio en un 8%. Igualmente, encontró mayor prevalencia a la desnutrición aguda en un 2,3% a diferencia del 2010 que estaba en un 0.9%. Finalmente, el 24.7% de los niños menores de 5 años presentaron anemia. Esta encuesta no se realizó en el año 2020 debido a la pandemia.

Como se puede evidenciar en los anteriores documentos, se ha realizado análisis de datos sobre desnutrición infantil en países fuera de Latinoamérica. En el ámbito local, se han realizado análisis e informes desde otras disciplinas como la medicina, sociología, trabajo social, entre otros. Sin embargo, no se ha aprovechado el uso de la analítica y ciencia de los datos para lograr conocimiento e información que sirva de insumo para las organizaciones gubernamentales realizar políticas públicas que permitan intervenir la problemática alrededor de la desnutrición infantil habiendo un portal de datos abiertos en Colombia con datos disponibles sobre el tema.

Capítulo 3
Identificación de metadata sobre desnutrición infantil en niños de 0 a 5 años en Medellín

En este capítulo se explora a través de portales de datos abiertos dispuestos por el gobierno y las organizaciones no gubernamentales, se plantea un flujo de búsqueda y al mismo tiempo, se solicitan fuentes a los entes independientes en los que se basaron para el estudio de la problemática. Finalmente, se hace una identificación de las principales características que deben cumplir la metadata para que sea adecuada y tomar la decisión de cuál será el conjunto de datos candidato para el proceso de analítica de datos.

3.1 Fase de Exploración

Aprovechando la iniciativa del gobierno de Colombia, a través de la ley 1712 de 2014 sobre Transparencia y Acceso a la Información Pública Nacional de brindar datos abiertos de cada una de las entidades públicas o privadas que cumplen funciones públicas y disponen a cualquier ciudadano de forma libre y sin restricciones datos abiertos en formatos que permiten su uso y reutilización bajo licencia abierta y sin restricciones legales para su aprovechamiento (MinTic, 2023). A nivel nacional se encuentra el portal https://www.datos.gov.co/ y en Medellín se tiene http://medata.gov.co/.

Igualmente se encuentran iniciativas no gubernamentales que realizan análisis sociales, entre las que se encuentra la red de ciudades como vamos (RCCV) nació con el propósito de generar información confiable, imparcial y comparable en torno a temas de calidad de vida urbana y participación ciudadana (Red de

Ciudades Cómo Vamos, 2023) y entre las ciudades se encuentran Medellín. Ellos publican a través de este portal https://redcomovamos.org/.

Otra entidad que se consideró dentro de los datos abiertos para consulta fue el Instituto Colombiano de Bienestar Familiar (ICBF), pero después de revisar su política de publicación de datos, se evidenció que transmiten a través de portal de datos abiertos de Colombia.

3.2 Búsqueda de información sobre desnutrición infantil

A través de los buscadores propios de cada portal de Datos Abiertos Colombia y Medellín Data, se ingresó la búsqueda con las siguientes palabras claves: “Desnutrición Infantil” o “Desnutrición”. En la figura 5 se relaciona la estrategia que se planteó para búsqueda de datos.

En total se logró tres fuentes de datos que son:

1. Mortalidad Anual por desnutrición en menores de cinco años Antioquia desde 2005
2. Desnutrición aguda en menores de 5 años Medellín
3. Alertas poblacionales Medellín

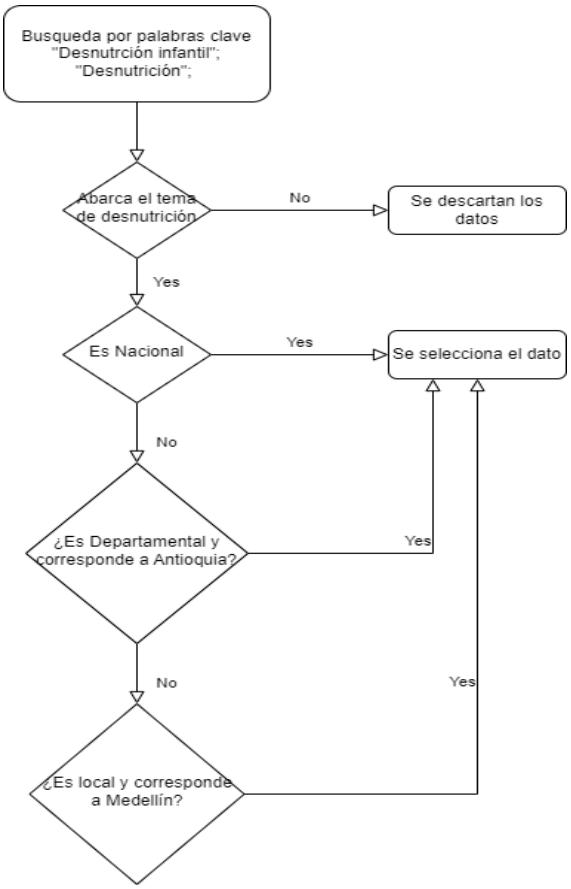
Para el primer grupo de datos, al filtrar solo en la ciudad de Medellín, en la figura 6 se evidencia que solo se cuenta con 17 registros, que corresponden desde el año 2005 hasta el año 2021 y relaciona la cantidad de niños menores de 5 años que murieron por causa de la desnutrición.

Para el segundo grupo de datos se encontró que se cuentan con 2802 registros con información segregada por edad, barrio, comuna, régimen de seguridad social, peso y talla tanto al nacimiento como en el momento de la consulta, entre

otros datos como se muestra en la figura 7 lo que permite inferir que es una buena fuente de datos candidata para ingresar a la analítica de datos.

Figura 5

Diagrama de estrategia de búsqueda de datos.



Nota. La figura muestra el flujo de la estrategia para la recolección de los datos.

Fuente: Elaboración propia.

Para el segundo grupo de datos se encontró que se cuentan con 2802 registros con información segregada por edad, barrio, comuna, régimen de seguridad social, peso y talla tanto al nacimiento como en el momento de la consulta, entre otros datos como se muestra en la figura 6 lo que permite inferir que es una buena fuente de datos candidata para ingresar a la analítica de datos.

Figura 6

*Mor*talidad anual por desnutrición en menores de 5 años en Antioquia

NombreMunicipio	CodigoMunicipio		NombreRegion	Codi	Año	CausaMortalidad	TipoPoblacionObjetivo	Num	NumeroCasos
Medellín	5001	PO	VALLE DE ABURRA	1	2005	Desnutrición	menores de 5 años	148171	5
Medellín	5001	PO	VALLE DE ABURRA	1	2006	Desnutrición	menores de 5 años	146942	8
Medellín	5001	PO	VALLE DE ABURRA	1	2007	Desnutrición	menores de 5 años	144497	2
Medellín	5001	PO	VALLE DE ABURRA	1	2008	Desnutrición	menores de 5 años	142359	3
Medellín	5001	PO	VALLE DE ABURRA	1	2009	Desnutrición	menores de 5 años	140706	1
Medellín	5001	PO	VALLE DE ABURRA	1	2010	Desnutrición	menores de 5 años	139798	2
Medellín	5001	PO	VALLE DE ABURRA	1	2011	Desnutrición	menores de 5 años	139853	2
Medellín	5001	PO	VALLE DE ABURRA	1	2012	Desnutrición	menores de 5 años	140656	2
Medellín	5001	PO	VALLE DE ABURRA	1	2013	Desnutrición	menores de 5 años	141137	0
Medellín	5001	PO	VALLE DE ABURRA	1	2014	Desnutrición	menores de 5 años	141162	0
Medellín	5001	PO	VALLE DE ABURRA	1	2015	Desnutrición	menores de 5 años	141005	0
Medellín	5001	PO	VALLE DE ABURRA	1	2016	Desnutrición	menores de 5 años	141160	1
Medellín	5001	PO	VALLE DE ABURRA	1	2017	Desnutrición	menores de 5 años	141859	0
Medellín	5001	PO	VALLE DE ABURRA	1	2018	Desnutrición	menores de 5 años	143827	0
Medellín	5001	PO	VALLE DE ABURRA	1	2019	Desnutrición	menores de 5 años	146601	0
Medellín	5001	PO	VALLE DE ABURRA	1	2020	Desnutrición	menores de 5 años	148533	0
Medellín	5001	PO	VALLE DE ABURRA	1	2021	Desnutrición	menores de 5 años	148663	0

Nota. La figura muestra los registros de mortalidad anual por desnutrición en menores de 5 años en Antioquia. Fuente: Secretaría de Salud y Protección Social de Antioquia

Figura 7

Desnutrición aguda en menores de 5 años

id	semana	edad_	uni_med_	sexo_	nombre_barrio	comuna	tipo_ss_	cod_ase_	fec_con_	ini_sin_	tip_cas_	pac_hos_	peso_nac	talla_nac
1	19	1	1	M	Picacho	Doce de Octubre	S	CCF002	30/03/2016	30/03/2016	4	2	2800	49
2	19	3	1	M	Picachito	Doce de Octubre	S	CCF002	31/03/2016	31/03/2016	4	2	2600	48
3	27	2	1	M	Santo Domingo Savio No.2	Popular	S	CCF002	22/06/2016	1/01/1900	4	2	2930	48
4	27	1	1	M	LLanaditas	Villa Hermosa	C	EPS010	5/07/2016	5/06/2016	4	2	2900	52
5	37	9	2	M	El Nogal-Los Almendros	Belen	C	EPSS16	4/08/2016	4/08/2016	4	1	3700	48
6	38	9	2	F	Moravia	Aranjuez	S	CCF002	19/09/2016	7/09/2016	4	2	2570	49
7	43	1	1	F	Mirador del Doce	Doce de Octubre	S	CCF002	26/10/2016	2/06/2016	4	2	3300	47
8	49	6	2	M	Kennedy	Doce de Octubre	S	CCF002	5/12/2016	29/08/2016	4	2	3100	50
9	50	6	2	M	Brasilia	Aranjuez	P	RES004	16/12/2016	1/01/1900	4	2	3490	50
10	50	2	1	M	Santa Elena Sector Central	Corregimiento De Santa Elena	C	EPS037	9/08/2016	18/07/2016	4	2	900	33
11	52	1	1	M	Brasilia	Aranjuez	S	CCF002	23/12/2016	1/01/1900	4	2	1400	41
12	52	8	2	F	El Chagualo	La Candelaria	C	EPS002	13/12/2016	19/04/2016	4	1	2490	0
13	11	10	2	M	Villa Hermosa	Villa Hermosa	C	EPS016	15/03/2016	1/01/1900	4	2	2630	46
14	14	4	1	M	Florencia	Castilla	C	EPS016	4/04/2016	1/01/1900	4	2	2500	47
15	14	3	1	F	SIN INFORMACION	SIN INFORMACION	S	CCF002	4/04/2016	4/04/2015	4	2	2600	48
16	15	2	1	F	Miranda	Aranjuez	C	EPS010	14/04/2016	14/04/2016	4	2	2190	49
17	17	1	1	F	Belencito	San Javier	S	CCF002	23/04/2016	1/04/2016	4	1	900	32
18	17	4	1	M	Villa Hermosa	Villa Hermosa	S	CCF002	29/04/2016	29/04/2016	4	2	3300	54.3

Nota. La figura muestra los registros de desnutrición aguda en menores de 5 años en la ciudad de Medellín. Fuente: Secretaría de Salud Medellín

Finalmente, en el caso específico de Alertas Poblacionales Medellín (consultar tabla 2), se llevó a cabo un análisis que reveló la presencia de alertas relacionadas con desnutrición y malnutrición. Sin embargo, se encontró que solamente existen dos registros en esta categoría, lo cual lleva a la conclusión de que no es una fuente de datos adecuada para realizar análisis de manera efectiva.

Tabla 2

Alertas Poblacionales

EDAD_NNA	GENERO_NNA	TIPO_ALERTA_TEMPRANA	BARRIO_CASO	RIESGO	COMUNA_CASO	FECHA_CREACION_ALERTA	QUIEN_REPORTA
16	2	3	Versalles # 2	1	3	8/10/2018	1
18	2	5	Santo Domingo Savio # 1	1	1	4/09/2018	1

Nota. Registros de alertas poblacionales. Fuente: Secretaría de la Juventud Medellín

En la tabla 3, relacionamos el resumen de los datos encontrados:

Tabla 3

Resumen de cantidad de registros y columnas en datos abiertos

Fuente de datos	Cantidad de registros	Cantidad de columnas significativas
Mortalidad Anual por desnutrición en menores de cinco años Antioquia desde 2005	17	3
Desnutrición aguda en menores de 5 años	2802	23
Alertas poblacionales	2	4

Nota. Resumen de cantidad de registros y columnas en datos abiertos. . Fuente: Elaboración propia

3.3 Datos de las organizaciones no gubernamentales

Entre las opciones de las organizaciones no gubernamentales, se encuentra Medellín Cómo Vamos, y en su último informe (Medellín Cómo Vamos, 2021) hallaron que el año 2021 se alcanzó la mayor proporción en los últimos años. Igualmente, otros datos como la proporción del peso al nacer, lactancia materna exclusiva, prevalencia de desnutrición entre otros. Lo que se puede inferir que tienen datos importantes. Por ello, se contacta mediante correo electrónico solicitando los datos en el que se basaron en el informe y su respuesta fue afirmativa.

Al explorar los datos se encontró que cuenta con 10 registros y con campos tales como comuna, año, talla, entre otros. Estos datos visibles son resumidos por año desde el 2011 al 2020. Además, no proporciona el dato de manera independiente por cada menor de edad como se puede ver en la figura 8. Por lo tanto, se puede concluir que no es una data adecuada para aplicar analítica de datos ya que cuenta con pocos registros y además los datos no son independientes sino resumidos por año.

Figura 8

Nutrición Medellín

	munici	cod_zc	zona	cod_cc	comun	ano	tasa_m	num_r	desnu	num_c	desnu	num_c	desnu	num_c	por_de	num_r	por_de	num_c	por_de	num_c
05001	MEDELLIN	Total	Total	Total	Tota	2011	0	0	4.08439	2509	10.80597	6638	4.299272	2641	10.36092	5638	23.2317	14271	16.66477	10237
05001	MEDELLIN	Total	Total	Total	Total	2012	1.367905	2	3.133706	2947	9.317114	8762	4.036494	3796	10.70149	8723	23.16731	21787	16.81376	15812
05001	MEDELLIN	Total	Total	Total	Total	2013	0	0	2.515487	2278	8.39011	7598	3.609801	3269	10.23772	8062	23.58794	21361	16.57041	15006
05001	MEDELLIN	Total	Total	Total	Total	2014	0	0	1.938206	1796	7.567206	7012	3.335744	3091	10.06598	8116	22.80198	21129	15.80027	14641
05001	MEDELLIN	Total	Total	Total	Total	2015	0	0	1.919001	1494	7.280387	5668	3.27155	2547	10.2904	7080	22.34339	17395	15.76176	12271
05001	MEDELLIN	Total	Total	Total	Total	2016	0.684294	1	1.636041	1372	7.160659	6005	3.298315	2766	9.427505	7906	22.47886	18851	15.21446	12759
05001	MEDELLIN	Total	Total	Total	Total	2017	0	0	1.115056	1231	7.163173	7908	2.708382	2990	8.839834	9759	23.90895	26395	15.31097	16903
05001	MEDELLIN	Total	Total	Total	Total	2018	0	0	1.129297	1293	6.862248	7857	2.428906	2781	8.628249	9879	24.2908	27812	15.24944	17460
05001	MEDELLIN	Total	Total	Total	Total	2019	0	0	1.078258	1245	7.352075	8489	2.483025	2867	8.522137	9840	25.01299	28881	15.3589	17734
05001	MEDELLIN	Total	Total	Total	Total	2020	0	0	1.101366	874	7.357982	5839	2.421997	1922	8.468169	6720	24.75301	19643	15.15953	12030

Nota. La figura muestra los registros de desnutrición en menores de 5 años en la ciudad de Medellín. Fuente: (Medellín Cómo Vamos, 2021)

3.4 Selección de Metadata

Los datos deben cumplir con ciertas características que permitan el análisis y entrenamiento del modelo, entre ellas consideran (Fan & Bifet, 2013):

- Deben tener suficiente volumen de datos para capturar y permitir el aprendizaje, entre más datos se tengan, mejor el modelo
- Los datos deben tener variabilidad de datos en el fenómeno que se esté estudiando, con casos posibles para que el modelo aprenda de patrones
- Los datos deben estar libre de errores, sin ruido y con calidad
- Los datos deben contener variables relevantes para el problema en estudio, incluso suficientes características para la extracción de las más útiles
- Los datos deben ser consistentes en formato y estructura

A continuación, la tabla 4 se resumen todos los datos obtenidos:

Tabla 4

Resumen de datos obtenidos con cantidad de registros y columnas

Fuente de datos	Cantidad de registros	Cantidad de columnas significativas
Mortalidad Anual por desnutrición en menores de cinco años Antioquia desde 2005	17	3
Desnutrición aguda en menores de 5 años	2802	23
Alertas poblacionales	2	4
Medellín Como vamos	10	15

Nota. Resumen de cantidad de registros y columnas de los diferentes conjuntos de datos obtenidos. Fuente: Elaboración propia

Con lo anterior se puede concluir que la mejor fuente de datos es la segunda (Desnutrición infantil aguda), ya que tiene una cantidad de registros suficientes y una cantidad de segregación de información tales como el peso, talla, régimen de salud, barrio/comuna, entre otros, lo que permite deducir que es significativa para realizar un análisis y definir su estructura final. Además, cuenta con buena variedad de datos, volumen y clasificados.

3.5 Conclusiones del capítulo

En este capítulo se realizó la exploración con dos fuentes, una de ellas a través del portal de datos abiertos de Colombia y solicitud formal de la fuente de datos a Medellín Cómo Vamos que permitió conocer algunos aspectos importantes de la problemática. Este paso permitió conseguir cuatro grupos de datos que y a través de unas características de los datos tales como variedad, variabilidad y veracidad permitió tomar una decisión de cuál es la mejor metadata que será objeto de estudio para el proceso de analítica de datos.

Capítulo 4

Requisitos funcionales y arquitectónicos

Este capítulo presenta los requisitos funcionales para realizar el proceso de analítica de datos al conjunto de datos seleccionado, se reconoce cómo están conformados los datos a través de un diagrama de clases y modelo Entidad-Relación. Igualmente, se presenta la propuesta del flujo ETL que se realizará a los datos con el fin de unificar, limpiar y generar valores nuevos. Finalmente, se presentará las decisiones de herramientas y tecnologías para dar cumplimiento al paso a paso de análisis de datos.

4.1 Requisitos funcionales

En la tabla 5 se detallan los requisitos funcionales para realizar el proceso de analítica de datos, especificado a través de historias de usuario **(Mountain Goat Software, 2023)**

Tabla 5

Requisitos funcionales

ID	Descripción Requisito	Criterios de Aceptación
RF01	Yo como científico de datos requiero recopilar datos para crear un modelo de analítica de datos	*Los datos deben tener una estructura tabular *Pueden estar en formato CSV, SQL, XSLX, TXT, JSON, XML, TSV
RF02	Yo como científico de datos requiero realizar limpieza de datos para garantizar calidad y coherencia en el análisis de datos	*No debe haber campos vacíos *Eliminar los datos que generan ruido
RF03	Yo como científico de datos requiero preprocesar los datos con el fin de eliminar datos alfabéticos y permitan el análisis en el modelo de analítica de datos	*La tabla final no debe contener datos alfabéticos *En caso de ser variables categóricas, reemplazar una tabla que entre 0 y 1 donde pregunte una categoría
RF04	Yo como científico de datos requiero integrar los datos con otras fuentes de datos para mejorar la analítica de datos	*Los datos se deben integrar en una sola fuente de datos

RF05	Yo como científico de datos requiero utilizar una técnica de análisis de datos para realizar una predicción, clasificación o agrupamiento	*Seleccionar una técnica que permita aplicar análisis de datos acorde al conjunto de datos seleccionado
RF06	Yo como científico de datos requiero visualizar los datos para aplicar criterios de selección, análisis de datos y resultados	*Seleccionar librerías o frameworks que permitan la visualización de datos
RF07	Yo como científico de datos requiero evaluar el proceso de análisis de datos para identificar el rendimiento, predicción/diagnóstico	*Selección de criterios que permitan evaluar el modelo si genera valor y es acorde a sus resultados

Nota. Requisitos funcionales del sistema. Fuente: Elaboración propia

4.2 Diagrama de Clases

En la figura 9 se presenta cómo están conformados el conjunto de datos seleccionado en el capítulo tres con el fin de reconocer las diferentes variables y cómo están segregadas. Este reconocimiento nos permitirá una mejor exploración de datos, la dependencia o independencia de las variables.

4.3 Modelo Entidad Relación

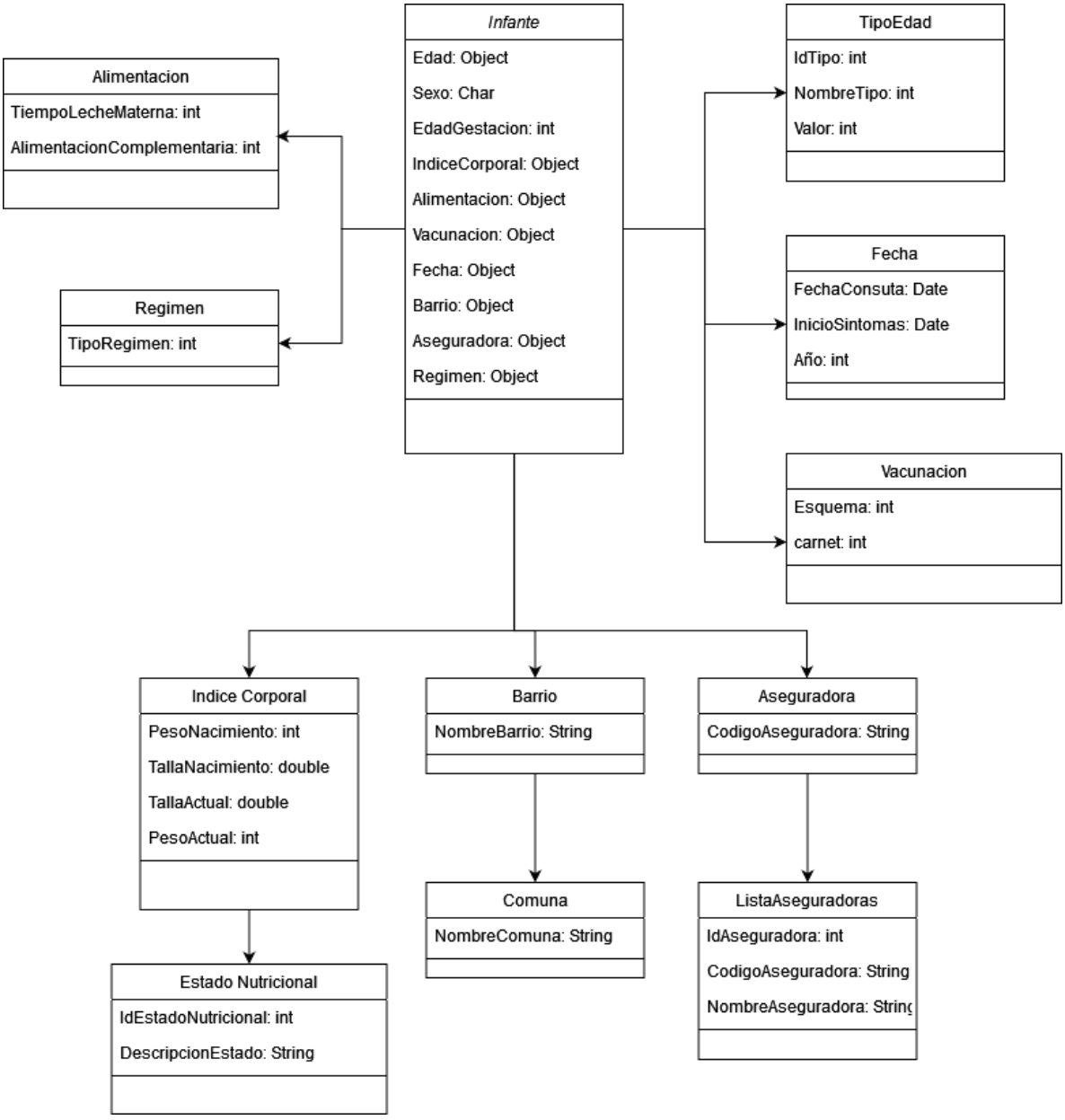
En la figura 10 se presenta el modelo Entidad-Relación del conjunto de datos seleccionado en el capítulo tres. Este, nos permitirá tomar una decisión de cómo se realizará el proceso de ETL, que datos se van a unificar y cuáles variables se transformarán de valores de cadena a numéricos.

4.4 Proceso ETL para el conjunto de datos

En la figura 11 se muestra la aplicación del proceso ETL, con el fin de llevar a cabo la extracción de datos a la herramienta, transformar los datos de cadena a numéricos y unificar el tipo de edad, las aseguradoras y el estado nutricional ya que en el conjunto de datos está sin especificar si el dato la edad está representada por días, semanas o meses, no lista a qué aseguradora pertenece y no establece el estado nutricional del infante respectivamente. Adicional, se eliminan los campos vacíos que afectarán el proceso de analítica de datos.

Figura 9

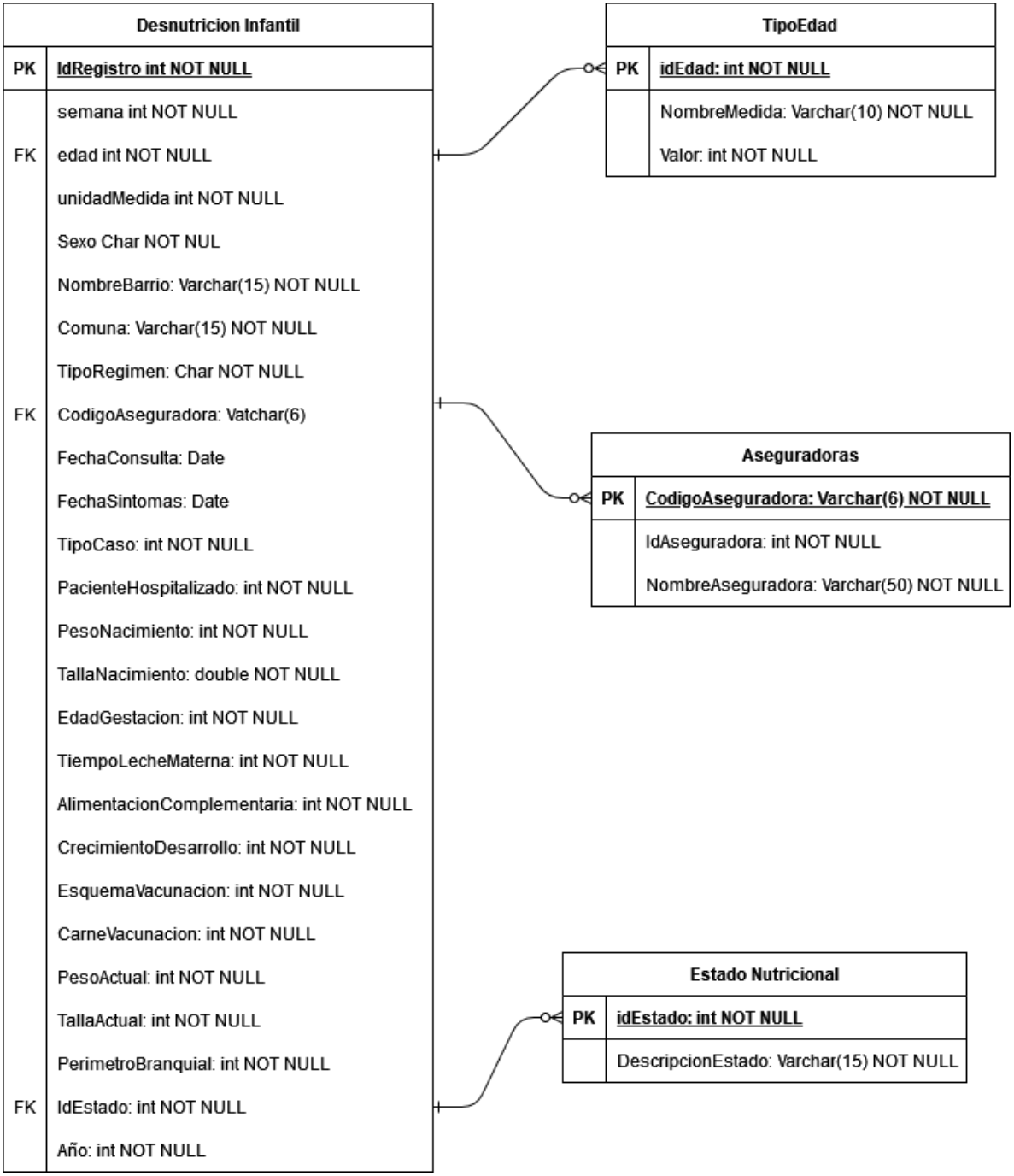
Diagrama de Clases Desnutrición Infantil Aguda



Nota. La figura muestra el diagrama de clases desnutrición infantil aguda. Fuente: Elaboración Propia

Figura 10

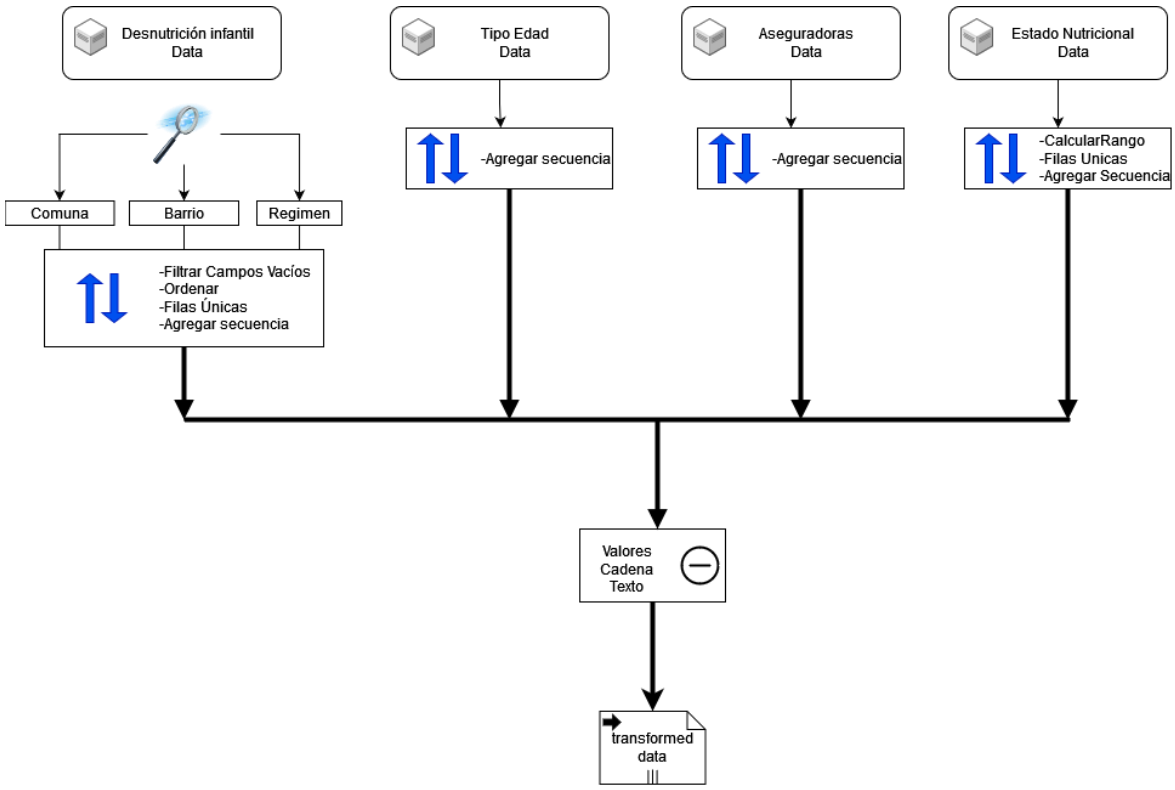
Diagrama Entidad Relación Desnutrición Infantil Aguda



Nota. La figura muestra el diagrama entidad relación desnutrición infantil aguda.
Fuente: Elaboración Propia

Figura 11

Diagrama Proceso ETL para Desnutrición Infantil Aguda



Nota. La figura muestra el diagrama proceso ETL para desnutrición infantil aguda. Fuente: Elaboración Propia

4.5 Selección de plataforma para el proceso de ETL

Para el proceso de ETL se requiere seleccionar una plataforma que permita transformar algunos datos, eliminar datos que generan ruido y campos vacíos. De igual manera, se requiere una herramienta que sea acorde con la naturaleza del proyecto, el costo, la facilidad de uso, el soporte o comunidad detrás de ella y en especial, que no exige altos recursos de hardware.

Inicialmente se cuenta con 2802 registros, lo que se deduce que no es necesario una herramienta potente, por temas de costos se optará por el open source y

finalmente cuente con una curva de aprendizaje fácil, intuitiva y con buen soporte de consulta. En la tabla 6 se resumen algunas herramientas candidatas para el proceso de ETL y si cumple o no con las características y la tabla 7 presenta los requisitos de hardware de Pentaho y Talend.

Tabla 6

Resumen de criterios de plataformas para el proceso ETL y la Tabla # los requerimientos del sistema

Plataforma	Código Abierto	Facilidad de Uso	Soporte/ Documentación	Curva de Aprendizaje baja	Recursos bajos	Calificación Garnet
Pandas	**X**		**X**			**N/A**
Pentaho	**X**	**X**	**X**	**X**	**X**	**4.1**
Talend	**X**	**X**	**X**	**X**	**X**	**4.1**
AWS Glue		**X**	**X**		**N/A**	**4.2**
PowerCenter		**X**	**X**			**4.4**

Nota. La tabla muestra si se cumplen los criterios Fuente: Elaboración Propia

En este orden de ideas observemos las ventajas y desventajas entre Pentaho y Talend según **(Ruiz Borja, 2018)** tomando características tales como costo, facilidad, visualización, velocidad y ejecución como se expone en la tabla 8.

Tabla 7

Requisitos de Hardware

Proceso/Plataforma	Talend	Pentaho
RAM	3GB-4GB	2GB
CPU	2GHz	DUAL-CORE en adelante
Lenguajes	Java	Java

Nota. Requisitos de hardware para las herramientas Pentaho y Talend. Fuente: Elaboración Propia

Tabla 8

Ventajas y desventajas de Talend y Pentaho.

Plataforma	Ventajas	Desventajas
Talend	• GUI Intuitiva • Más rápido en proceso de transformación y carga • Mejor Rendimiento • Mejor para extracción de datos en nube	• Más complejidad en la visualización de datos, ya que es por consola • Presenta un riesgo alto para manipulación de datos • requiere configuración específica y manual, con conocimiento previo de los datos a utilizar
Pentaho	• Mejor en extracción de datos on-premise • GUI Sencilla • Visualización de datos sencilla • Más rápido en extracción	• Menor capacidad que Talend en el proceso de ETL • Comunidad menor que a Talend

Nota. La tabla muestra las ventajas y desventajas Pentaho y Talend. Fuente: Elaboración propia

En conclusión, Pentaho y Talend presenta características muy similares. Igualmente, **(Ruiz Borja, 2018)** concluye que todo depende del contexto, la herramienta que se esté acostumbrado a usar y no deja una recomendación única, sino por parte de los usuarios. Por ende, se hace una exploración grupal internamente y por entrenamiento en ETL se opta por Pentaho para realizar el proceso de ETL.

4.6 Selección de técnica para el proceso de analítica de datos

Se delimita el análisis para seleccionar la técnica de analítica de datos para aplicar al conjunto de datos (Desnutrición infantil aguda) de cuatro tipos: regresión lineal, regresión logística, clustering y reglas de asociación. La tabla 9 muestra los requisitos de los datos para poder aplicar las técnicas.

Tabla 9

Relación de requisitos para poder aplicar las técnicas de analítica de datos

Característica	Regresión Lineal	Regresión Logística	Clustering	Reglas de Asociación
Tipo de variable dependiente	Continua	Categórica	No aplica	No aplica
Tipo de variables independientes	Continuas	Continuas o categóricas	Continuas o categóricas	Transacciones o items
Tamaño de la muestra	Grande	Grande	Grande o pequeño	Grande o pequeño
Distribución de los datos	Preferiblemente normal	No aplica	No aplica	No aplica
Escala de las variables	Preferiblemente lineal	No aplica	No aplica	No aplica

Ausencia de valores faltantes	Deseable	Deseable	Deseable	Deseable
Ausencia de valores atípicos	Deseable	Deseable	Deseable	Deseable
Relación lineal entre variables	Deseable	No aplica	No aplica	No aplica
Independencia de los datos	Deseable	Deseable	No aplica	Deseable
Homogeneidad de las varianzas	Deseable	No aplica	No aplica	No aplica
No multicolinealidad entre variables	Deseable	No aplica	No aplica	No aplica
Representatividad de los datos	Deseable	Deseable	Deseable	Deseable

Nota. La tabla muestra los requisitos de entrada de los datos. Fuente: Elaboración propia

Como se puede evidenciar en la tabla 9, para estudiar qué técnica de analítica de datos se requieren explorar, analizar, qué estructura tienen, visualizar correlaciones, la distribución e identificar patrones en los datos. El conjunto de datos cuenta con valores numéricos, algunas variables se pueden convertir en una lista numérica y finalmente, las columnas que cuenta con dos categorías se pueden convertir fácilmente en datos de 1 y 0. Además, por limitaciones del tiempo para el desarrollo del trabajo de grado, se decide indagar de una vez a la primera técnica (regresión lineal) y las demás quedarán en trabajos futuros.

4.7 Selección de herramienta para el proceso de analítica de datos

Uno de los requisitos que solicitaron durante el trabajo de grado fue de que se tratara de no alejar la ingeniería de software y tratar al menos codificar un segmento de código. En este orden de ideas, basado en el documento **(Amat Rodrigo, 2023)** se decide phyton junto con sus librerías Pandas, Numpy, Seaborn, Scikit-learn.

Para el ejercicio de regresión lineal con Python se requiere el uso de Jupyter notebook. Además, para eliminar la necesidad de instalar un ambiente local en el equipo personal (Docker, Anaconda, entre otros) se optó por usar Google Colab, ya que permitirá escribir y ejecutar código phyton desde el navegador, altamente recomendado para aprendizaje automático, análisis de datos y educación, eliminándose la necesidad de descargar e instalar **(Google, 2023)**.

4.8 Conclusiones del capítulo

El alcance de los requisitos funcionales permitirá el paso a paso para cumplir con el proceso de analítica de datos y la parte arquitectónica las herramientas con la cual se realizará el proceso de ETL y la aplicación del proceso de analítica de datos. Es importante tener en cuenta las limitaciones de presupuesto y de hardware con el fin de optimizar la eficiencia de los recursos disponibles y la eficacia de cumplir con los tiempos establecidos del trabajo de grado.

Capítulo 5

Caracterización los elementos que compone la metadata sobre desnutrición Infantil

Este capítulo permite garantizar que el conjunto de datos contenga las características que son necesarias para los algoritmos de analítica de datos, en especial frente a la regresión lineal, que parte de los requisitos es que las variables sean numéricas, por ello, transforma los datos de las comunas, barrios y régimen de salud en datos numéricos asignándoles un valor. De igual manera, en la metadata hay ausencias como los datos de las aseguradoras de salud y en qué tipo de medida se encuentran los fatos. Finalmente, se realiza limpieza de los valores nulos con el fin de que el algoritmo acepte el entrenamiento.

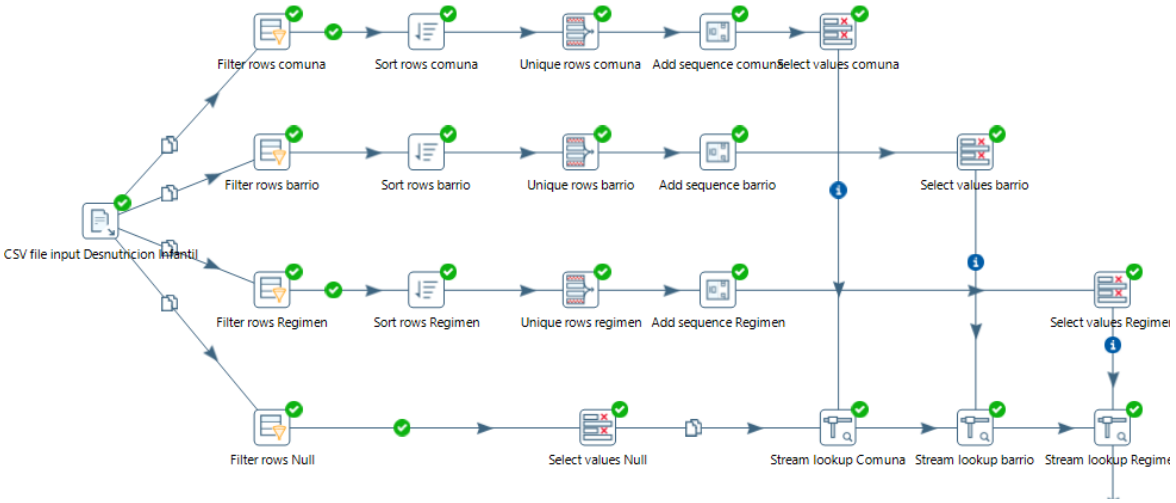
5.1 Proceso de extracción, transformación y carga (ETL)

Luego del análisis exploratorio de los datos de desnutrición infantil aguda en niños de Medellín y en aras de cumplir con los requisitos de la analítica de los datos para la regresión lineal, en el capítulo anterior se dio a conocer el diagrama que da solución (Ver figura 11) al tema de eliminar los datos de cadena y asignarles un valor entero. Igualmente, de eliminar los valores vacíos. La herramienta seleccionada para realizar el proceso de ETL fue Pentaho (ver sección 4.5).

La primera parte es filtrar por barrio, comuna y régimen de salud, eso con el fin de dar valores únicos a las variables. Igualmente, ya no cuenta con campos NULL en las mismas. Con esto, ya tengo valores asignados como se puede ver en la figura 12.

Figura 12

Agregar valores a comuna barrio y régimen de salud.



id_comuna	id_barrio	Id_Regimen
11	213	6
11	212	6
18	252	6
24	169	1
3	82	1

Nota. La figura muestra como Pentaho aplica valores a comunas, barrios y régimen de seguridad social. Fuente Elaboración propia

Ahora, como en el conjunto de datos se desconocen las aseguradoras de salud y el tipo de medida con respecto a la edad, es decir, no se sabe si el infante durante el reporte estaba por días (cuando son los primeros siete días posterior al parto), por semanas (a partir de la primera semana hasta el año) o por años (desde el primer año hasta los cinco años. Se unen los datos como se puede ver en la figura 13.

Finalmente, es desconocido el estado del niño en el nacimiento (Bajo peso, peso esperado o por encima) se le adiciona el cálculo usado tres rangos de valores:

- 1 – 2500: Bajo peso identificado como 1
- 2501 – 4000: Peso esperado identificado como 2

- Encima de 4000: Encima del peso identificado como 3

La figura 14 y muestran el cálculo del peso y los resultados en el conjunto de datos.

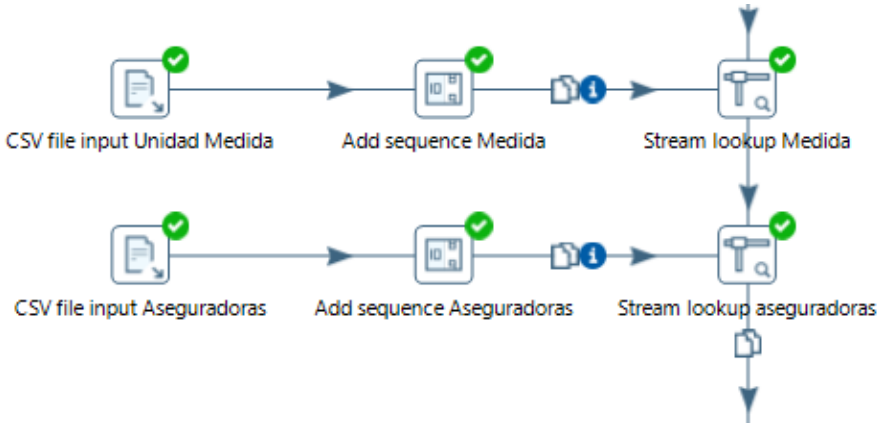
Finalmente, es necesario eliminar todos los campos que no son necesarios (contiene valores de cadenas de texto) y campos con valores NULL.

La figura 15 muestra el proceso de ETL completo.

En la figura 16 se muestra el diagrama Entidad-Relación del conjunto de datos resultante posterior al proceso de ETL.

Figura 13

Adicción de Unidad de Medida y Aseguradoras

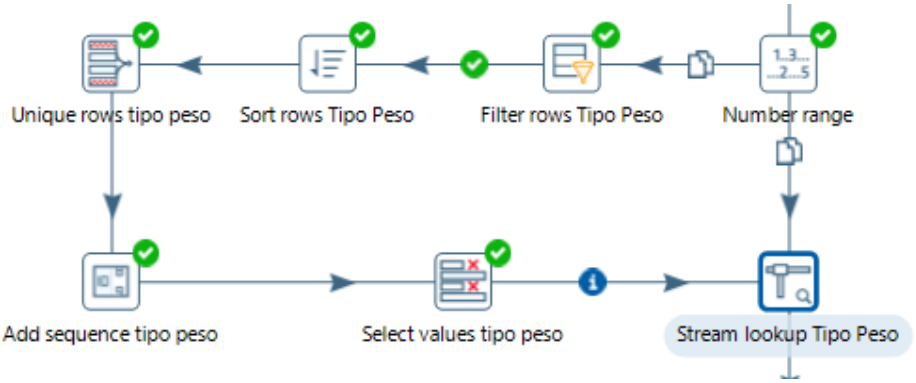


Tipo Medida	Aseguradora	IdAseguradora
Meses	SAVIA SALUD E.P.S.	80
Anios	EPS SURA	37
Anios	SAVIA SALUD E.P.S.	80
Meses	COOMEVA EPS SA	42
Anios	SAVIA SALUD E.P.S.	80
Meses	EPS SURA	37
Anios	NUEVA EPS SA	56
Anios	SAVIA SALUD E.P.S.	80
Meses	FONDO DE PRESTACIONES SOCIALES DEL MAGISTERIO	102

Nota. La figura muestra cómo se unen los datos de Unidad de Medida y listado de aseguradoras al conjunto de datos. Fuente: Elaboración propia

Figura 14

Cálculo del peso y los resultados en el conjunto de datos.

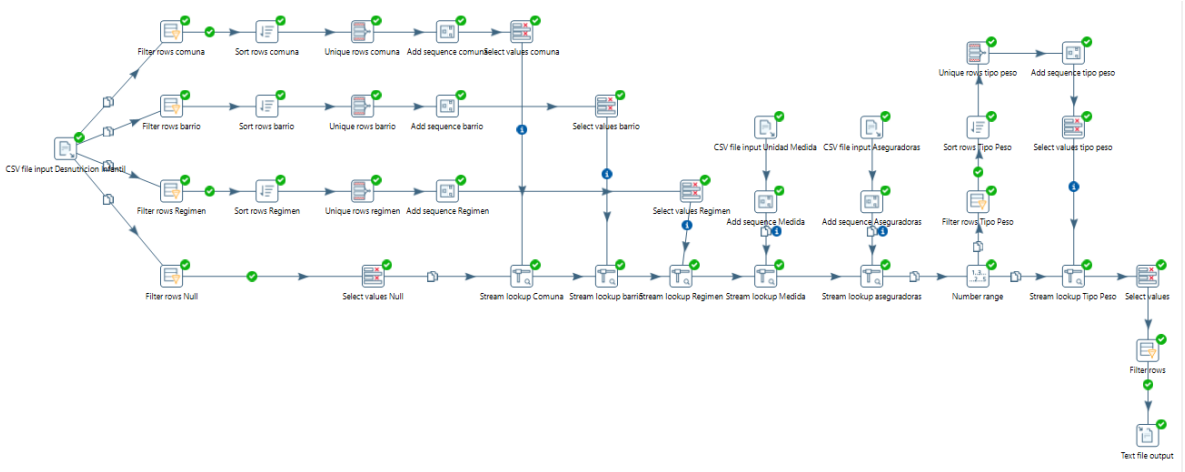


Tipo Peso Nacimiento	Id_Tipo_peso
1. Bajo peso	1
2. Peso Esperado	2
2. Peso Esperado	2
2. Peso Esperado	2
3. Sobrepeso	3
2. Peso Esperado	2

Nota. La figura muestra como asigna valores a los rangos del peso de nacimiento. Fuente: Elaboración propia

Figura 15

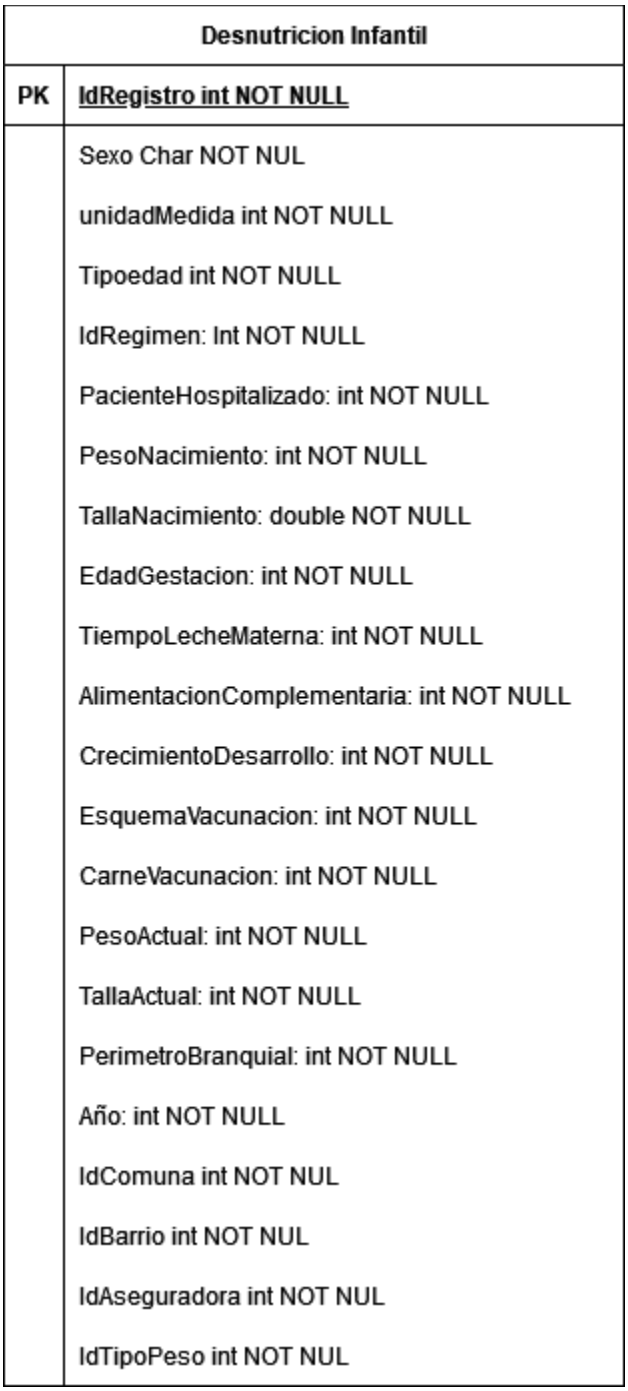
Proceso ETL para desnutrición infantil aguda



Nota. La figura muestra el proceso de ETL completo en Pentaho. Fuente: Elaboración propia

Figura 16

Diagrama Entidad-Relación después del proceso ETL



Nota. La figura muestra el modelo Entidad-Relación resultante después de del proceso de ETL. Fuente: Elaboración propia

5.2 Conclusiones del capítulo

El proceso de ETL permite dar mejor calidad a los datos, transformarlos y estandarizarlos de datos de cadenas de texto a datos numéricos secuenciales. Gracias a este proceso, los datos se encuentran listos para ingresarlos al proceso de análisis y visualizaciones de datos para ver su distribución y correlaciones. Por lo general, en el entorno productivo es muy útil en las grandes organizaciones ya que los datos vienen de múltiples fuentes y organizaciones.

Sin embargo, durante el desarrollo de este capítulo se evidenció que los datos sobre desnutrición infantil aguda se encontraron en formatos distintos en valores reales (algunos con comas y otros en puntos), lo que hubo que hacer la corrección manual de traspasar los valores de comas a puntos para que las librerías de phyton que se utilizarán en el desarrollo del siguiente capítulo sean capaz de aceptar y procesarlos.

Capítulo 6

Codificación del proceso de analítica de datos

En este capítulo se realiza el proceso de construcción del modelo de regresión lineal definido en el capítulo cuatro. Inicialmente se visualizará y comprenderá cómo están distribuido y correlacionado los datos, para finalmente decidir con cuáles variables son relevantes y con ello realizar la codificación del modelo a través de phyton y scikit-learn.

6.1 Un primer paso: visualizar y analizar los datos

Se requiere analizar las distintas variables con el fin de observar su distribución y visualizar si contiene datos que generan ruido. Para ello, se utilizó pandas como librería de análisis y ciencia de datos. También, seaborn para visualización de datos.

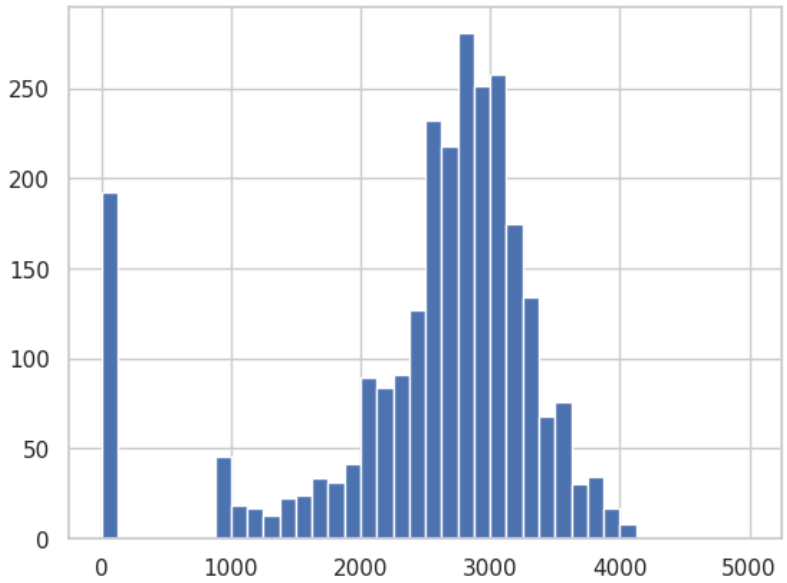
Se mira cómo están distribuidas una a una de las variables a través de un histograma y se encontró que datos como peso al nacer, talla al nacer, edad de gestación y talla actual presentaron datos con valores atípicos como se pueden ver en la figura 17, 18, 19 y 20.

Para visualizar los datos se realizó el siguiente código:

```
print(df.shape)
df.peso_nac.hist(bins = 40)

print(df.shape)
df.peso_nac.hist(bins = 40)

print(df.shape)
df.talla_nac.hist(bins = 40)
```

Figura 17

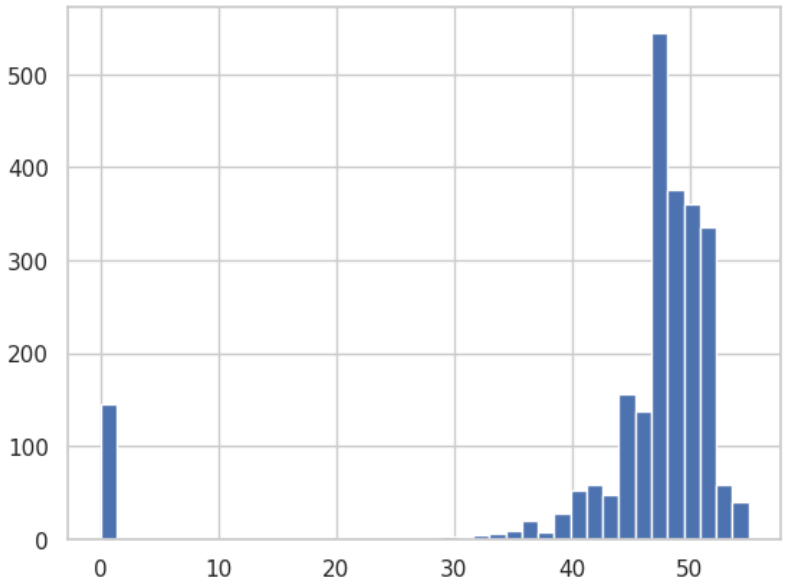
Distribución de los datos de peso al nacer



Nota. La figura muestra los valores atípicos de los datos al nacer. Fuente: Elaboración propia

Figura 18

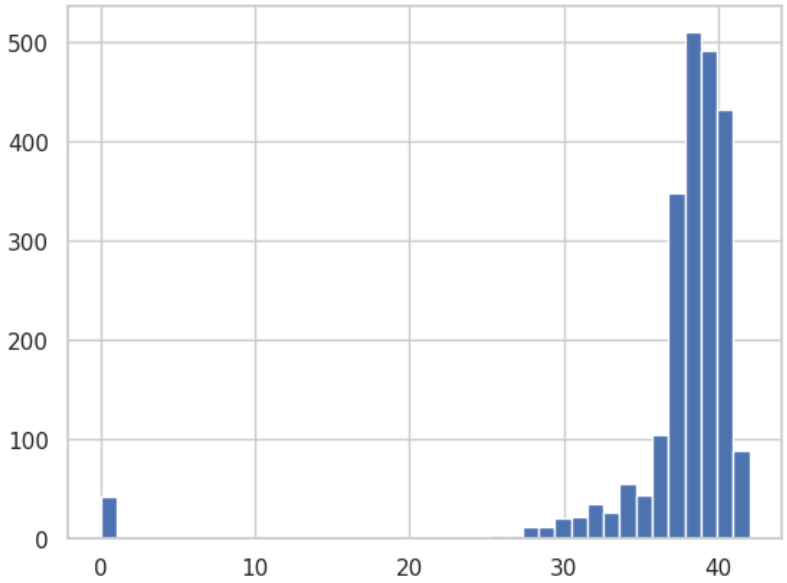
Distribución de los datos de talla al nacer



Nota. La figura muestra valores atípicos. Fuente: Elaboración propia

Figura 19

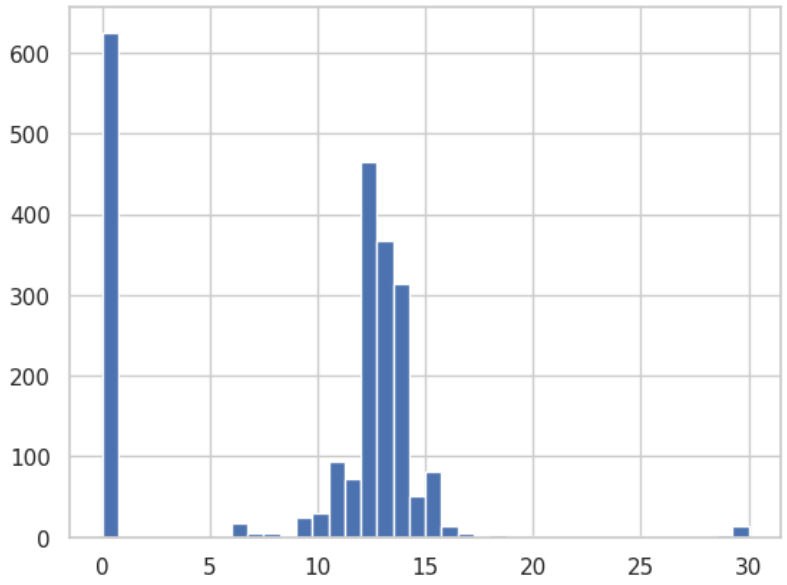
Distribución de los datos de semana de gestación



Nota. La figura muestra valores atípicos. Fuente: Elaboración propia

Figura 20

Distribución de los datos de perímetro branquial



Nota. La figura muestra valores atípicos. Fuente: Elaboración propia

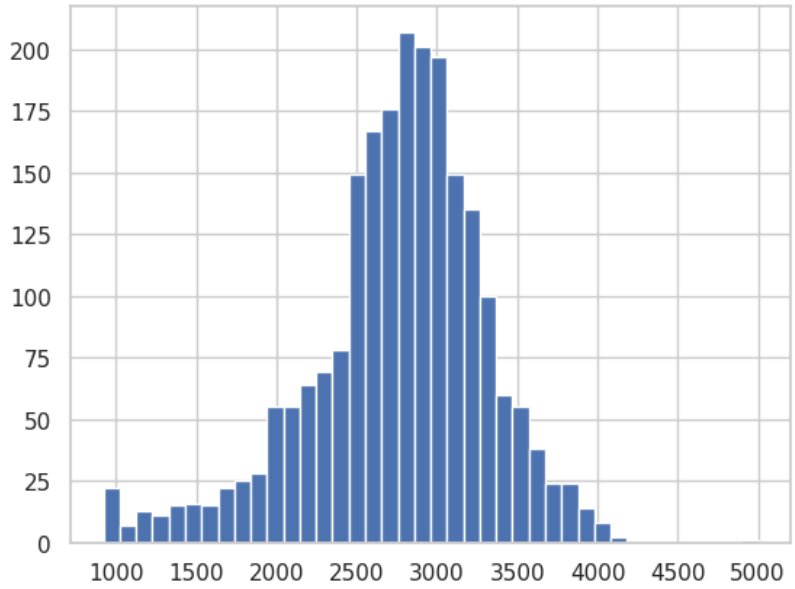
Evidentemente fue necesario realizar limpieza de los datos para eliminar datos atípicos y pueden dañar el entrenamiento y predicción de la regresión lineal. Para ello, se ingresó el siguiente código:

```
df[df.peso_nac<900]
df = df[df.peso_nac>900]

df[df.talla_nac<20]
df = df[df.talla_nac>20]

df[df.edad_ges<20]
df = df[df.edad_ges>20]
```

En las figuras 21, 22 y 23 Se visualizan los datos que ya no presentan datos atípicos.

Figura 21

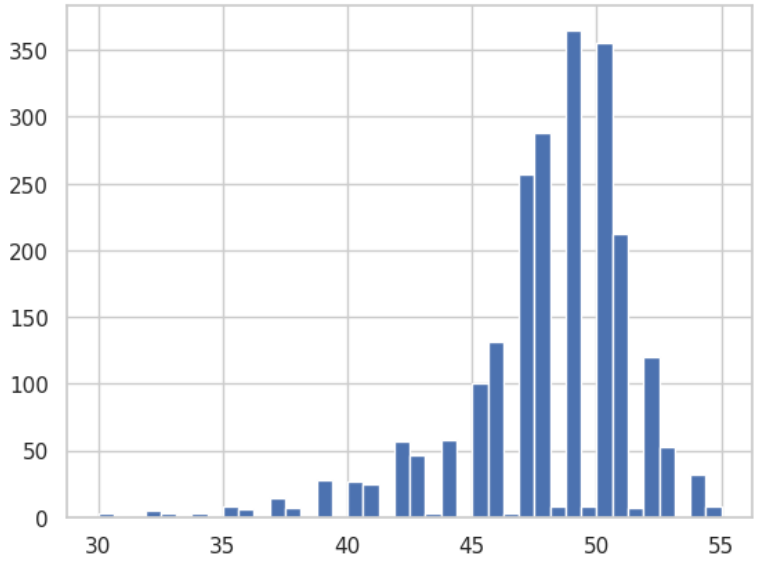
Distribución de los datos de peso al nacer sin valores atípicos



Nota. La figura muestra la limpieza de datos sin valores atípicos. Fuente: Elaboración propia

Figura 22

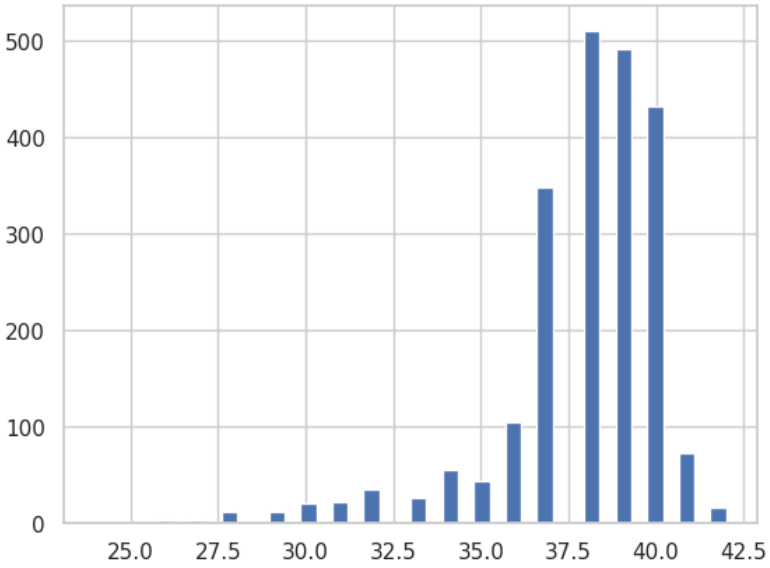
Distribución de los datos de talla al nacer sin valores atípicos



Nota. La figura muestra la limpieza de datos sin valores atípicos. Fuente: Elaboración propia

Figura 23

Distribución de los datos de semana de gestación sin valores atípicos



Nota. La figura muestra la limpieza de datos sin valores atípicos. Fuente: Elaboración propia

6.2 Distribución y Correlación de los Datos

Cabe recordar que en los resultados del capítulo anterior los datos: “'sexo, 'pac_hos' y 'carne_vac' son variables categóricas y dentro de la regresión lineal se recomienda transformarlos, para ello se utilizará la función get_dummies de pandas.

```
df = pd.get_dummies(df, columns=['sexo_','pac_hos_','carne_vac'],
drop_first=True)
df.head()
```

Ahora, el siguiente paso es ver cómo se distribuyen y se correlacionan, para ello se inserta el siguiente código cuyos resultados se pueden ver en la figura 23.

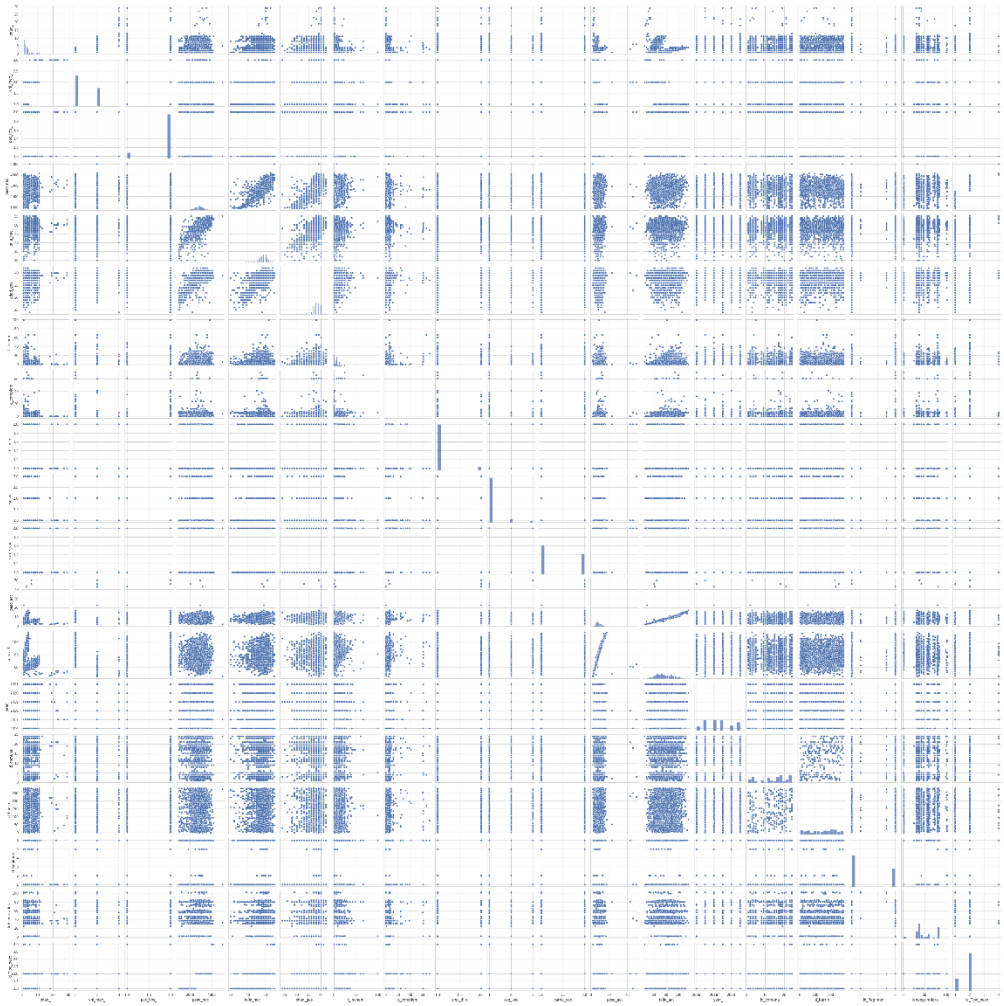
```
import seaborn as sns
import matplotlib.pyplot as plt
sns.set(style='whitegrid', context='notebook')
columns = ['edad_', 'uni_med_', 'pac_hos__2', 'peso_nac', 'talla_nac',
'edad_ges', 't_lechem', 'e_complem', 'crec_dllo', 'esq_vac',
'carne_vac_2', 'peso_act', 'talla_act', 'per_braqu', 'Id_Regimen',
'Id_Tipo_peso', 'sexo__M']
sns.pairplot(df[columns], height=2.5)
plt.show()
```

Evidentemente, al tener bastantes tipos de datos se hace complejo analizarlos, para ello, se recurre a otro método de la librería numpy para visualizarlos con la matriz gaussiana insertando el siguiente código y gráficamente la podemos ver en la figura 24.

```
import numpy as np
numeric_cols = ['edad_', 'uni_med_', 'peso_nac', 'talla_nac',
'edad_ges', 't_lechem', 'e_complem', 'crec_dllo', 'esq_vac',
'peso_act', 'talla_act', 'year_', 'id_comuna', 'id_barrio',
'Id_Regimen', 'IdAseguradora', 'Id_Tipo_peso',
'sexo__M','pac_hos__2','carne_vac_2']
cm = np.corrcoef(df[numeric_cols].values.T)
sns.set(font_scale=0.4)
sns.heatmap(cm, cbar=True,
annot=True,yticklabels=numeric_cols,xticklabels=numeric_cols)
```

Figura 24

Correlación y distribución de los datos



Nota. La figura muestra la distribución y correlación entre los valores. Fuente: Elaboración propia

Con la figura 25 se evidencia que los datos 'peso_nac', 'talla_nac', 'edad_ges' presentan una correlación con 'Id_Tipo_peso' debido a que presentan valores más cercanos a 1. Por ende, son candidatos para la codificación de la regresión lineal. Ahora, se generaron los mismos gráficos para observar las correlaciones entre ellos mostrado en las figuras 26 y 27. El código que ayudó fueron los siguientes:

```
import seaborn as sns
```

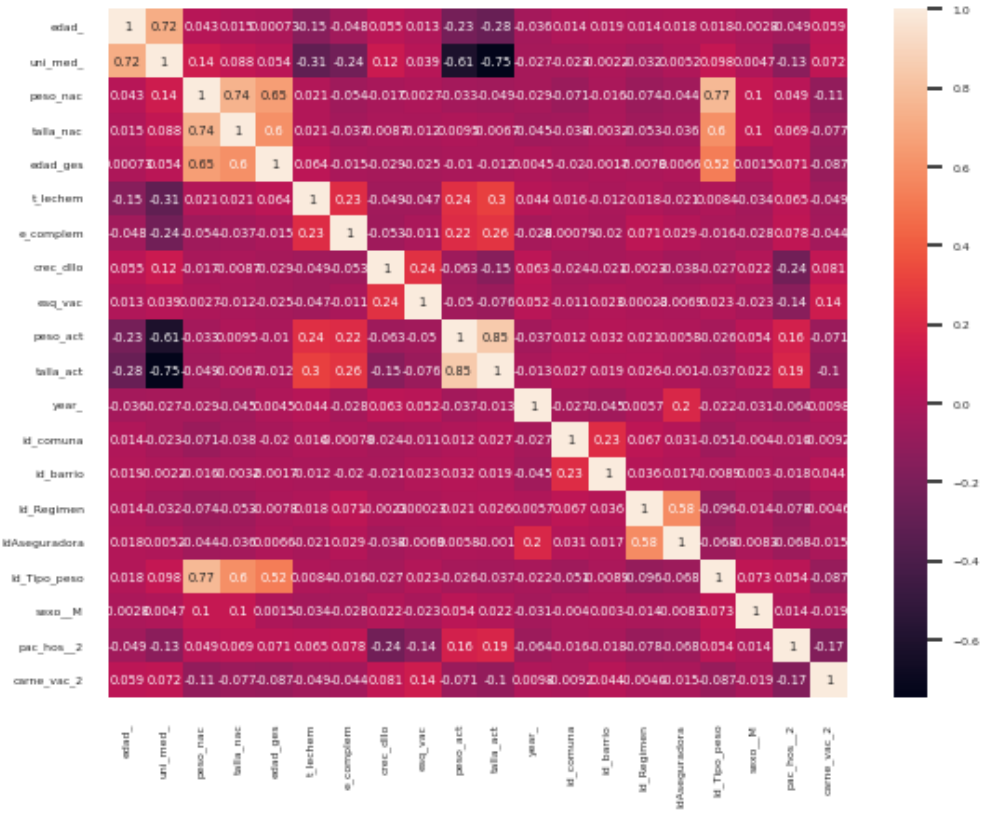
```
import matplotlib.pyplot as plt
sns.set(style='whitegrid', context='notebook')
columnsFilter = ['peso_nac', 'talla_nac', 'edad_ges', 'Id_Tipo_peso']
sns.pairplot(df[columns], height=2.5)
plt.show()
```

Código para generar la Matriz Gaussiana

```
import numpy as np
numeric_cols_filter = ['peso_nac', 'talla_nac', 'edad_ges',
'Id_Tipo_peso']
cm = np.corrcoef(df[numeric_cols_filter].values.T)
sns.set(font_scale=1.5)
sns.heatmap(cm, cbar=True,
annot=True,yticklabels=numeric_cols_filter,xticklabels=numeric_cols_fi
lter)
```

Figura 25

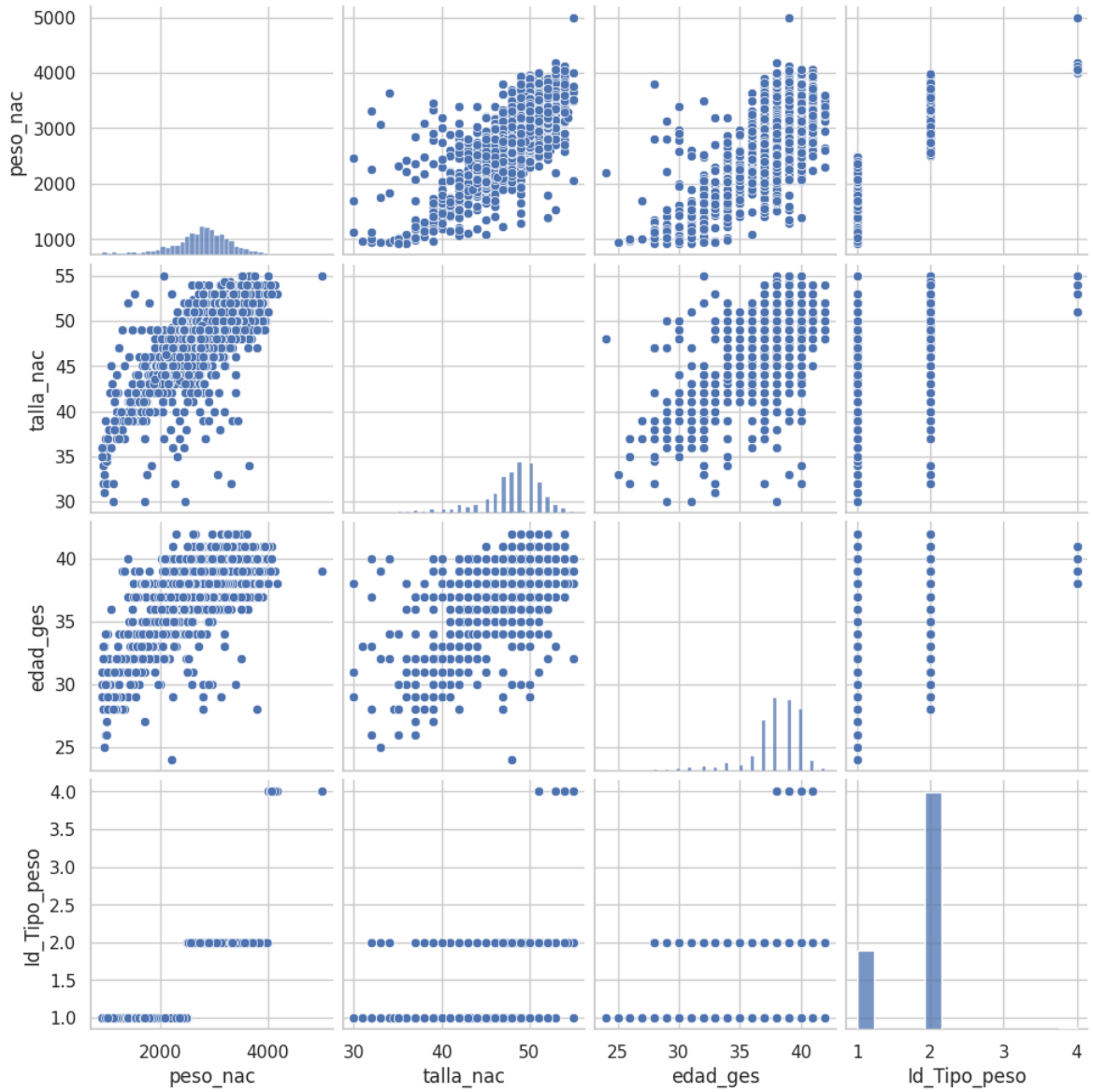
Matriz Gaussiana de distribución de los datos



Nota. La figura muestra en términos de valores la correlación entre los datos.
Fuente: Elaboración propia

Figura 26

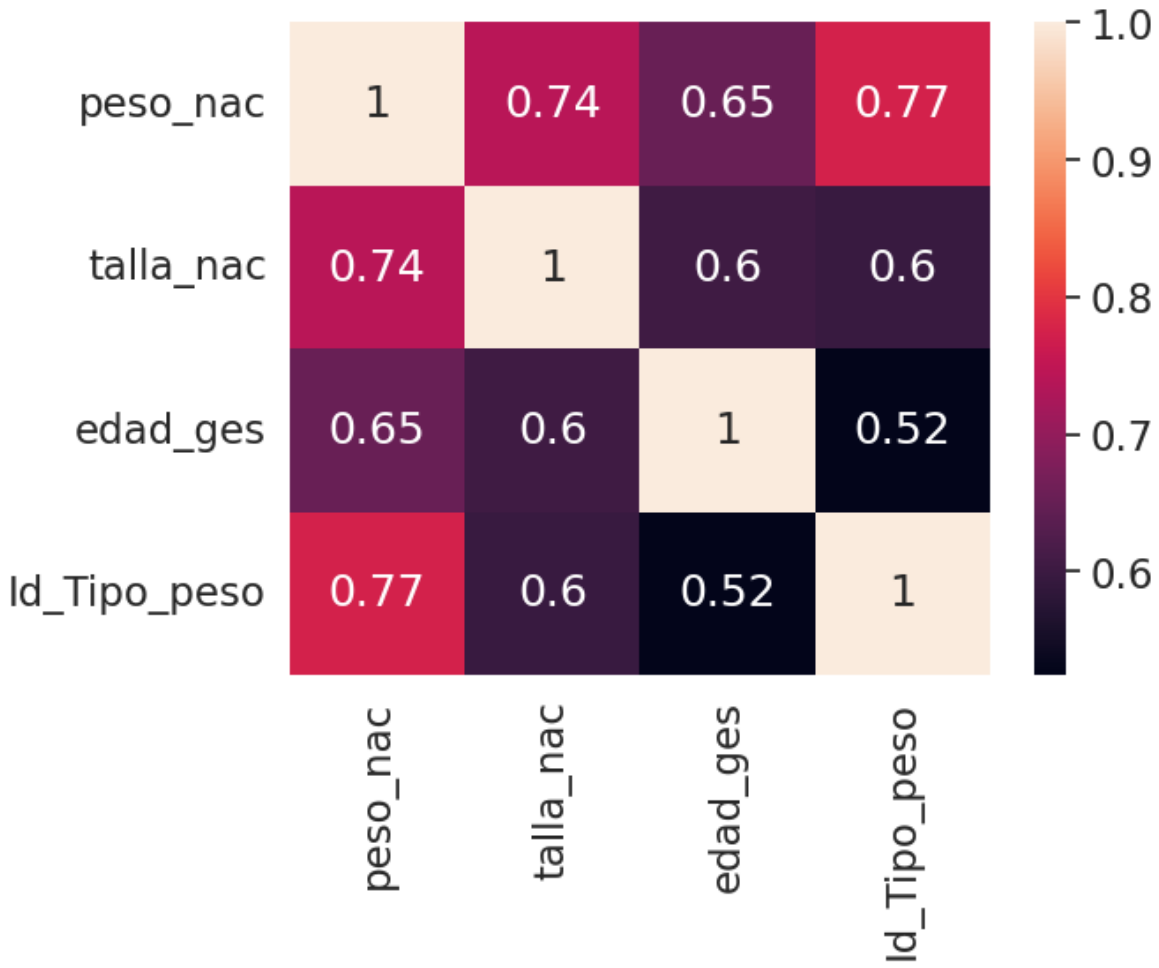
Distribución de los datos de peso, talla y edad de nacimiento



Nota. La figura muestra la correlación entre los datos candidatos. Fuente: Elaboración propia

Figura 27

Matriz gaussiana para los datos de peso, talla y edad de nacimiento



Nota. La figura muestra en términos de valores la correlación entre los datos

Fuente: Elaboración propia

6.3 Creación del modelo de Regresión lineal

Ya con los datos candidatos para la técnica de regresión lineal ('peso_nac', 'talla_nac', 'edad_ges' y 'Id_Tipo_peso') se intentará crear un modelo de regresión usando tres variables: Talla y peso al nacer en el eje X e Id del peso en eje Y. Se descarta edad de gestación debido a su menor valor. Para la creación del modelo se usó scikit-learn que es una librería de phyton para el aprendizaje automático ampliamente utilizada en la ciencia de los datos.

```
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LinearRegression

X = df[['peso_nac', 'talla_nac',]].values
y = df['Id_Tipo_peso'].values.reshape(-1, 1)

sc_x = StandardScaler()
sc_y = StandardScaler()

X_std = sc_x.fit_transform(X)
y_std = sc_y.fit_transform(y)

slr = LinearRegression()
slr.fit(X_std, y_std)
```

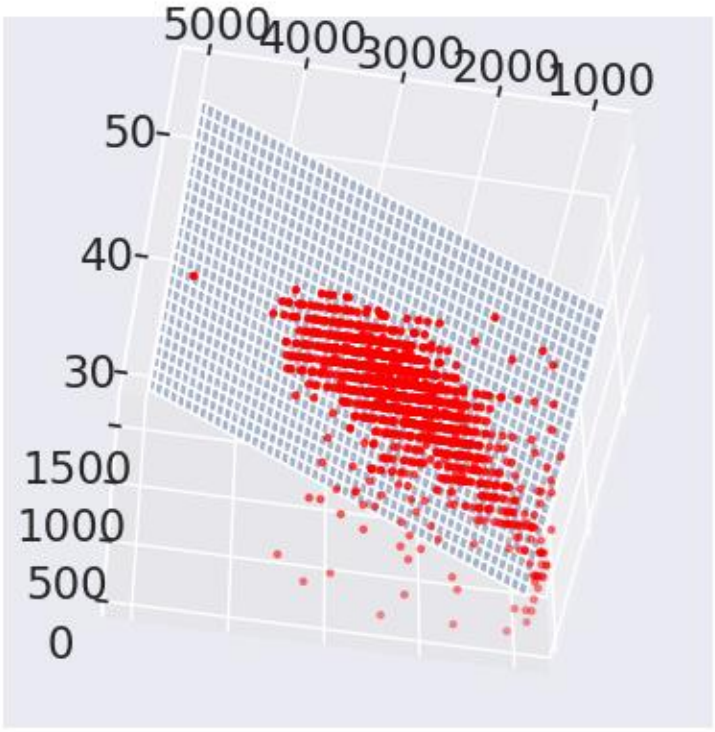
La figura 28 muestra el modelo de regresión lineal 3D codificado por scikit-learn.

6.4 Conclusiones del capítulo

A pesar de contar con una buena cantidad de fuentes de datos, durante la codificación del modelo se pudo demostrar que los datos presentaban pocas correlaciones entre sí. Lo ideal era buscar con un conjunto de variables permitiera predecir cuáles barrios, comunas o aseguradoras son las que podría presentarse los casos de desnutrición infantil. Sin embargo, con la matriz gaussiana fue solo con los datos de talla y peso al nacer permitía identificar el estado de nutrición del infante. Lo que demostró que es necesario garantizar la calidad de inserción de los datos por parte de la Secretaría de Salud Municipal de Medellín. Además, ampliar el rango de los estados de los infantes, independiente si están desnutridos, con sobrepeso o en un estado normal. Ello, podría permitir una mejor predicción o diagnóstico de la problemática.

Figura 28

Modelado 3D del modelo de regresión lineal



Nota. La figura muestra el modelado en 3D de la regresión lineal generada con scikit-learn. Fuente: Elaboración propia

Capítulo 7

Evaluación del proceso de analítica de datos

En este capítulo se presenta la evaluación del proceso. Para ello, se hace de dos maneras: la primera, evaluando a través de técnicas tradicionales de pruebas de software y la segunda, con métricas definidas para evaluar los resultados de la regresión lineal.

7.1 Pruebas del modelo de regresión lineal

Para realizar las pruebas de la respuesta del modelo de regresión lineal se seleccionaron una serie de datos originales posteriores al proceso de ETL, se seleccionarán tres datos por cada tipo de peso como se puede ver en la tabla 10.

Tabla 10

Selección de datos de prueba

Id_Tipo_peso (Resultado Esperado)	peso_nac (X1)	talla_nac (X2)
1	900	33
1	1005	35
1	1070	42
2	2800	49
2	3280	51
2	3820	52
3	4000	51
3	4050	54
3	5000	55

Nota. Datos tomados directamente de VSC resultante del proceso ETL. Fuente: Elaboración propia

Código para validar los resultados y resumen de resultados de prueba en la tabla 11:

```
peso_talla_std = sc_x.transform(np.array([X1,X2]).reshape(1, -1))
peso_std = slr.predict(peso_talla_std)
print("El niño se encuentra en un estado
de",  sc_y.inverse_transform(peso_std))
```

Tabla 11

Resultados de la prueba del modelo de regresión lineal

peso_nac (X1)	talla_nac (X2)	Resultado Obtenido	Resultado Esperado
900	33	0.54545523	1
1005	35	0.62188592	1
1070	42	0.70712082	1
2800	49	1.79560775	2
3280	51	2.0979961	2
3820	52	2.42995641	2
4000	51	2.53183482	3
4050	54	2.58170632	3
5000	55	3.16071367	3

Nota. Comparación de los resultados reales Vs los esperados. Fuente: Elaboración propia

Cabe resaltar que el modelo de regresión lineal cumple con una recta **Y = mX + b**. Por ende, los valores son reales, si aplicamos la regla de truncamiento para convertir los datos reales a cifras enteras se evidencia que se están obteniendo los valores esperados.

7.2 Otros criterios para evaluar el modelo de regresión lineal

Cuando se usa el modelo de regresión lineal hay otros criterios que permite rectificar la validez de la técnica. Entre ellos se encuentran R^2, MSE y valores residuales.

- **Coeficiente de determinación R^2:** Ayuda para mostrar lo bueno que se ajusta el modelo con los datos que hay, así como su variabilidad real. Es una de las más utilizadas. Cuanto el resultado se acerca a 1, mejor (Bhalla, 2023).
- **Mean Square Error (MSE):** Error cuadrático medio, se utiliza para calcular la función de perdida, los resultados se comparan a las predicciones. Si el valor llega a ser muy grande, tiende a ser impreciso (Scikit-Learn Developers, 2023).
- **Valores Residuales:** Representan el error o la discrepancia entre los datos reales y las predicciones del modelo. Interpretar los resultados de los valores residuales puede proporcionar información importante sobre el rendimiento y la precisión del modelo de regresión lineal. Entre los valores más cercanos a cero indican que el modelo ha predicho correctamente los valores observados. Esto sugiere que el modelo está ajustando bien los datos y no hay un sesgo sistemático en las predicciones (Qualtrics, 2023).

Se inserta código para evaluar el coeficiente de determinación R^2 y el error cuadrático medio.

```
import sklearn.metrics as metrics
mse = metrics.mean_squared_error(y_test,y_pred)
r2 = metrics.r2_score(y_test, y_pred)

print("r2 ", r2.round(4))
print("mse: ", mse.round(4))
```

Y para hallar los residuales del modelo:

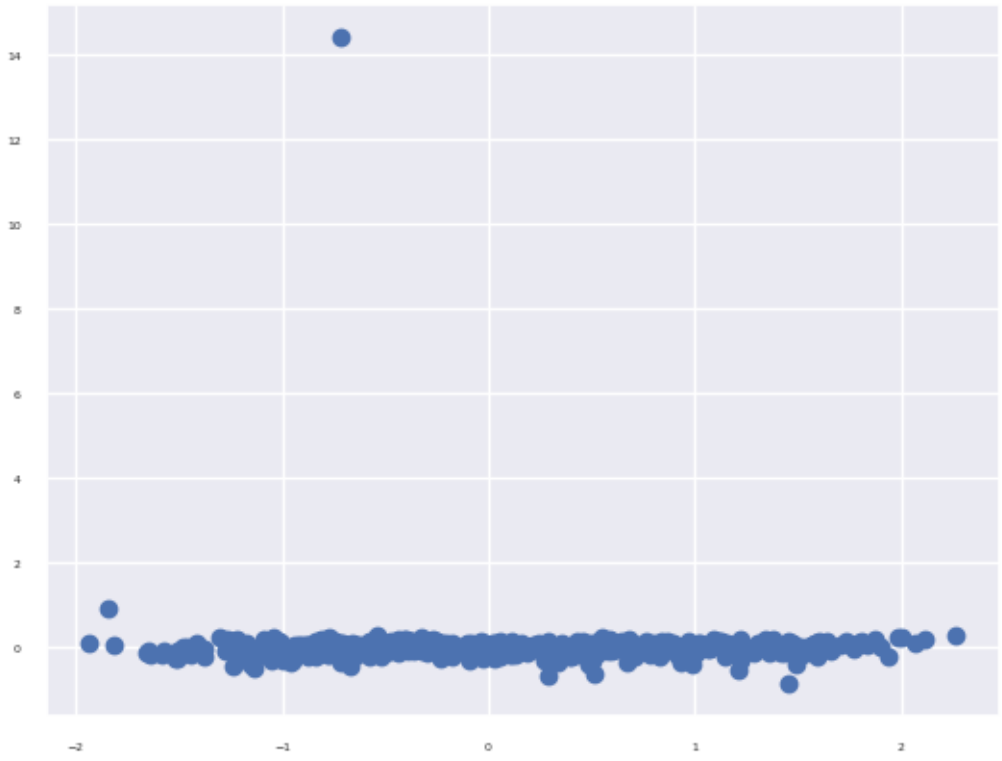
```
residuals = np.subtract(y_test, y_pred.reshape(-1))
plt.scatter(y_pred, residuals)
plt.show()
```

Resultados:

- R^2 = 0.9595
- MSE = 0.0314
- Valores Residuales = Ver figura 29

Figura 29

Cálculo de los valores residuales



Nota. Valores residuales representados en forma gráfica. Fuente: Elaboración propia

Los resultados de R^2, MSE y valores residuales nos permite concluir que el modelo de regresión lineal es adecuado y confiable, ya que el valor del coeficiente de determinación es cercano a 1 (0.9595), es cercano a cero según el error cuadrático medio y finalmente gráficamente los valores oscilan entre el eje cero lo que permite afirmar que los valores están ajustado a los datos.

7.3 Conclusiones del capítulo

El alcance de la evaluación del proceso de analítica de datos permitió determinar que la técnica de regresión lineal aplicada tiene un buen índice de predicción. Fue evaluada mediante dos estrategias: la primera, a través de los datos reales iniciales de entrada originales y a través de funciones de métricas disponibles dentro de la librería scikit-learn.

Capítulo 8

Conclusiones y trabajos futuros

En este capítulo se registran las principales contribuciones obtenidas del trabajo de grado, igualmente, se enuncian los trabajos futuros que se derivan de este.

8.1 Conclusiones

- El modelo de regresión lineal fue acorde a los datos de entrada y salida, dado una talla y peso al nacimiento me permitirá predecir si el infante presentará o no riesgo de desnutrición, como se pudo evidenciar en la evaluación desarrollada en el capítulo seis.
- Se evidenció la necesidad de asegurar la calidad de los datos desde el inicio del registro de los datos sobre desnutrición infantil, por ende, se recomienda que la secretaría de Salud Municipal de Medellín tenga un ingeniero de datos con el fin de garantizar el formato de los datos desde el inicio y no se pierdan valores durante el proceso de ETL y modelado de la analítica.
- La cantidad de valores representativos dentro del conjunto de datos no siempre van a garantizar un grupo de datos adecuado y representativo. En este caso, el conjunto de datos pasó de 23 datos a cuatro. Y al momento de codificar la técnica de regresión lineal multivariable quedaron tres. Sería más interesante si dado un conjunto de datos, hubiere predicho en qué barrios, comunas o aseguradora podría presentarse un niño o niña con desnutrición infantil.
- Es importante ampliar el rango de datos acerca de los estados nutricionales, ya que los datos solo se basaron en desnutrición aguda. La analítica de datos mejoraría si adicionamos por cada consulta pediátrica, independiente del diagnóstico, si es leve, severo o en estado normal.

8.2 Trabajos futuros

- A través del conjunto de datos producto de este trabajo de grado, aplicar otras técnicas o estrategias de analítica de datos para generar otros modelos y evaluar su factibilidad.
- Estandarizar el reporte y formato de datos al portal de datos abiertos para evitar la pérdida de datos durante el proceso de ETL y aplicación de analítica de datos.
- Explorar otras herramientas o frameworks desde el proceso de ETL o aplicación de analítica y ciencia de datos. Inclusive, la oportunidad de generar pipelines automáticos.
- Un sistema de recolección y reporte de datos continuo a través de un pipeline, en el cual se recolecte en las distintas Instituciones Prestadoras de Salud (IPS) y Empresas Prestadoras de Salud (EPS) donde se realicen consultas médicas pediátricas y permita recolectar el estado nutricional del niño y generar un conjunto más amplio de metadatos para la ciencia de datos.

Referencias

Alonso Amo, F. -M.-S. (2005). Introducción a la Ingeniería del Software: modelos de desarrollo de programas. En F. -M.-S. Alonso Amo, *Introducción a la Ingeniería del Software: modelos de desarrollo de programas* (pág. 77). Delta Publicaciones.

Amat Rodrigo, J. (Febrero de 2023). *Machine learning con Python y Scikit-learn.* https://www.cienciadedatos.net/documentos/py06_machine_learning_python_scikitlearn

Ayala Gaytán, A. D.-H. (2 de Enero de 2015). *Infraestructura, ingreso y desnutrición infantil en México.* Infraestructura, ingreso y desnutrición infantil en México: https://www.scielosp.org/article/ssm/content/raw/?resource_ssm_path=/media/assets/spm/v57n1/v57n1a5.pdf

Bhalla, D. (12 de Junio de 2023). *Difference between Adjusted R-squared and R-squared*.

Blum, B. I. (1996). *Beyond Programming: To a New Era of Design.* Oxford University Press.

Carreño, R. R. (Septiembre de 2017). *Desnutrición infantil y factores de riesgo en niños menores de 5 años.* Desnutrición infantil y factores de riesgo en niños menores de 5 años: http://repositorio.unesum.edu.ec/bitstream/53000/916/1/UNESUM-ECU-EMFER-2017-08.pdf

Coronado Escobar, Z. (15 de Septiembre de 2014). *Factores Asociados A La Desnutrición En Niños Menores De 5 Años.* Retrieved 1 de Noviembre de 2022, from http://biblio3.url.edu.gt/Tesario/2014/09/15/Coronado-Zully.pdf

Escobar, Z. Y. (15 de Septiembre de 2014). *Crai Landívar - Red de Bibliotecas.* Crai Landívar - Red de Bibliotecas: http://biblio3.url.edu.gt/Tesario/2014/09/15/Coronado-Zully.pdf

Fan, W., & Bifet, A. (2013). Mining big data: current status, and forecast to the future. *ACM SIGKDD explorations newsletter, 14*(2), 1-5. https://doi.org/https://doi.org/10.1145/2481244.2481246

Frankenfield, J. (27 de Junio de 2022). *Data Analytics: What It Is, How It's Used, and 4 Basic Techniques*. Data Analytics: What It Is, How It's Used, and 4 Basic Techniques: https://www.investopedia.com/terms/d/data-analytics.asp

Google. (Junio de 2023). *Colaboratory, preguntas frecuentes*. https://research.google.com/colaboratory/intl/es/faq.html

Instituto Colombiano de Bienestar Familiar. (Noviembre de 2017). *ENSIN: Encuesta Nacional de Situación Nutricional 2015.* Retrieved 7 de Noviembre de 2022, from https://www.icbf.gov.co/bienestar/nutricion/encuesta-nacional-situacion-nutricional

Khare, S. K. (2017). Investigation of Nutritional Status of Children based on Machine Learning Techniques using Indian Demographic and Health Survey Data. *the 7th International Conference on Advances in Computing & Communications*.

M, G. C. (2014). Desnutrición infantil en menores de cinco años en Perú: tendencias y factores determinantes. *Revista Panamericana de Salud Publica, 35*(2), 104-112.

Martínez, A. F. (Diciembre de 2006). *Modelo de análisis del impacto social y económico de la desnutrición infantil en América Latina.* Modelo de análisis del impacto social y económico de la desnutrición infantil en América Latina: https://repositorio.cepal.org/bitstream/handle/11362/5491/S0600972_es.pdf?sequence=1&isAllowed=y

Medellín Cómo Vamos. (27 de Septiembre de 2021). *¿Cómo va la primera infancia en Medellín?* ¿Cómo va la primera infancia en Medellín?: https://www.medellincomovamos.org/system/files/2022-10/docuprivados/Informe%20Primera%20Infancia%202021%20MCV%20pdf.pdf

Minsalud. (22 de Febrero de 2016). *Ministerio de Salud y Protección social.* Ministerio de Salud y Protección social: https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VS/PP/SNA/abc-desnutricion-aguda.pdf

MinTic. (23 de Abril de 2023). *Gobierno Digital*. https://gobiernodigital.mintic.gov.co/portal/Iniciativas/Datos-abiertos/

Moine, J. (Abril de 2013). *Metodologías para el descubrimiento de conocimiento en bases de datos: un estudio comparativo.* https://core.ac.uk/download/pdf/16703288.pdf

Moine, J. M., Gordillo, S., & Haedo, A. S. (2011). Análisis comparativo de metodologías para la gestión de proyectos de minería de datos. *XVII Congreso Argentino de Ciencias de la Computación (CACIC 2011).* http://sedici.unlp.edu.ar/handle/10915/18749

Mountain Goat Software. (12 de Junio de 2023). *User Stories*. https://www.mountaingoatsoftware.com/agile/user-stories

Oficina de Perspectivas y Políticas Mundiales de UNICEF. (Octubre de 2019). *Crecer bien en un mundo en transformación. Niños, alimentos y nutrición.* Retrieved 4 de Diciembre de 2022, from https://www.unicef.org/media/61091/file/Estado-mundial-infancia-2019-resumen-ejecutivo.pdf

Organization, W. H. (Mayo de 2007). *World Health Organization, World Food Programme, United Nations System Standing Committee on Nutrition & United Nations Children's Fund (UNICEF).* World Health Organization, World Food Programme, United Nations System Standing Committee on Nutrition & United Nations Children's Fund (UNICEF): https://apps.who.int/iris/bitstream/handle/10665/44295/9789280641479_eng.pdf?sequence=1&isAllowed=y

Perez Lopez, C., & Santin Gonzalez , D. (2007). *Minería de datos. Técnicas y herramientas.* Ediciones Paraninfo, S.A.

Pressman, R. (2010). *Ingeniería del software, un enfoque práctico.* Pearson Educación.

Qualtrics. (12 de Junio de 2023). *Interpretar diagramas residuales para mejorar su regresión*. https://www.qualtrics.com/support/es/stats-iq/analyses/regression-guides/interpreting-residual-plots-improve-regression/?rid=cookie&prevsite=en&newsite=es&geo=CO&geomatch=es-la

Red de Ciudades Cómo Vamos. (23 de Abril de 2023). *La red*. https://redcomovamos.org/la-red/

Rodrigo Martínez, A. F. (Diciembre de 2006). *Modelo de análisis del impacto social y económico de la desnutrición infantil en América Latina.* Retrieved 6 de

Diciembre de 2022, from https://repositorio.cepal.org/bitstream/handle/11362/5491/S0600972_es.pdf?sequence=1&isAllowed=y

Ruiz Borja, J. E. (2018). *Comparación de herramientas ETL de código abierto.*

Sangita Khare, S. K. (2017). Investigation of Nutritional Status of Children based on Machine Learning Techniques using Indian Demographic and Health Survey Data. *7th International Conference on Advances in Computing & Communications, ICACC-2017, 22-.* Cochin, India.

Scikit-Learn Developers. (12 de Junio de 2023). *Metrics and scoring: quantifying the quality of predictions*. https://scikit-learn.org/stable/modules/model_evaluation.html

Sobrino, M., Gutiérrez, C., Cunha, A. J., Dávila, M., & Alarcón, J. (2014). Desnutrición infantil en menores de cinco años en Perú: tendencias y factores determinantes. *Revista panamericana de salud pública*(35), 104-112.

Tapia, T. A. (2021). *Percepciones y prácticas sobre cuidado familiar como factorde riesgo socio cultural asociado a la desnutrición crónica en niños entre 1 a 5 años atendidos en el centro de salud de la comunidad cebadas-parroquia cebadas- cantón guamote.* Percepciones y prácticas sobre cuidado familiar como factorde riesgo socio cultural asociado a la desnutrición crónica en niños entre 1 a 5 años atendidos en el centro de salud de la comunidad cebadas-parroquia cebadas- cantón guamote: http://repositorio.puce.edu.ec/bitstream/handle/22000/19982/TESIS%20TERESA%20YEPEZ-FINAL%20%282022%29%20%281%29.pdf?sequence=1&isAllowed=y

UNICEF, O. d. (Octubre de 2019). *ESTADO MUNDIAL DE LA INFANCIA 2019. Niños, alimento y nutrición. Crecer bien en un mundo en transformación.* ESTADO MUNDIAL DE LA INFANCIA 2019. Niños, alimento y nutrición. Crecer bien en un mundo en transformación: https://www.unicef.org/media/61091/file/Estado-mundial-infancia-2019-resumen-ejecutivo.pdf

Vallejo Ballesteros, H., Guevara Iñiguez, E., & Medina Velasco, S. (2018). Minería de Datos. *Revista Científica Mundo de la Investigación y el Conocimiento,*

2(especial), 339-349. https://doi.org/10.26820/recimundo/2.esp.2018.339-349

Vesoulis, A. N. (Agosto de 2022). *Pediatric Research.* Pediatric Research: https://www.nature.com/articles/s41390-022-02264-9

World Health Organization, World Food Programme, United Nations System Standing Committee on Nutrition & United Nations Children's Fund (UNICEF). (Mayo de 2007). *World Health Organization.* Retrieved 1 de Noviembre de 2022, from https://apps.who.int/iris/bitstream/handle/10665/44295/9789280641479_eng.pdf?sequence=1&isAllowed=y

Zachary A. Vesoulis, A. N. (16 de Agosto de 2022). *Pediatric Research.* Retrieved 2 de Noviembre de 2022, from https://www.nature.com/articles/s41390-022-02264-9