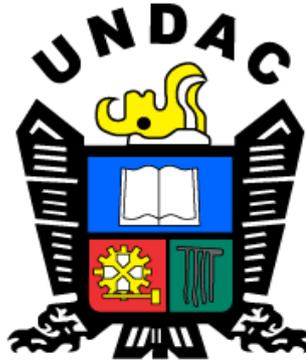


UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN
FACULTAD DE INGENIERÍA
ESCUELA DE FORMACIÓN PROFESIONAL DE SISTEMAS Y
COMPUTACIÓN



T E S I S

**Modelo de Minería de Datos para la Predicción de Casos de Anemia
en Madres Gestantes de la Provincia de Pasco - 2023**

Para optar el título profesional de:

Ingeniero de Sistemas y Computación

Autor:

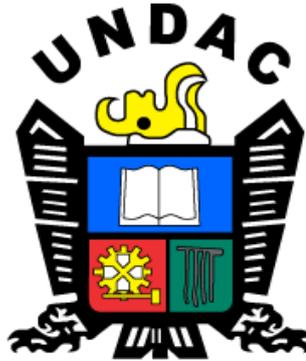
Bach. Bryan Omar YAURI SANTA CRUZ

Asesor:

Mg. Oscar Cleворio CAMPOS SALVATIERRA

Cerro de Pasco – Perú - 2024

UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN
FACULTAD DE INGENIERÍA
ESCUELA DE FORMACIÓN PROFESIONAL DE SISTEMAS Y
COMPUTACIÓN



T E S I S

**Modelo de Minería de Datos para la Predicción de Casos de Anemia
en Madres Gestantes de la Provincia de Pasco - 2023**

Sustentada y aprobada ante los miembros del jurado:

Mg. Lisbeth Gisela NEGRETE CARHUARICRA
PRESIDENTE

Mg. Pit Frank ALANIA RICALDI
MIEMBRO

Mg. Jose Luis SOSA SANCHEZ
MIEMBRO



Universidad Nacional Daniel Alcides
Carrión Facultad de Ingeniería
Unidad de Investigación

INFORME DE ORIGINALIDAD N° 194-2023-UNDAC/UIFI

La Unidad de Investigación de la Facultad de Ingeniería de la Universidad Nacional Daniel Alcides Carrión en mérito al artículo 23° del Reglamento General de Grados Académicos y Títulos Profesionales aprobado en Consejo Universitario del 21 de abril del 2022, La Tesis ha sido evaluado por el software antiplagio Turnitin Similarity, que a continuación se detalla:

Tesis:

**Modelo de Minería de Datos para la predicción de casos de anemia en
madres gestantes de la Provincia de Pasco - 2023**

Apellidos y nombres de los tesisistas:

Bach. YAURI SANTA CRUZ, Bryan Omar

Apellidos y nombres del Asesor:

Mg. CAMPOS SALVATIERRA, Oscar Cleворio

Escuela de Formación Profesional
Ingeniería Sistemas y Computación

Índice de Similitud

29%

APROBADO

Se informa el Reporte de evaluación del software similitud para los fines pertinentes:

Cerro de Pasco, 12 de enero del 2024


UNDA UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN
FACULTAD DE INGENIERÍA
UNIDAD DE INVESTIGACIÓN
Luis Villar Requís Carbajal
DOCTOR EN CIENCIAS - DIRECTOR

DEDICATORIA.

Gracias Dios por permitirme llegar a este momento tan especial en mi vida. Gracias por los triunfos y momentos difíciles que me hacen valorar cada día más, y gracias a mi madre por estar conmigo durante todo mi camino estudiantil y de vida. Gracias a mi padre por sus consejos y saber guiarme hasta el final de mi carrera.

AGRADECIMIENTO

Sin el apoyo constante de mis padres no habría llegado a este punto de mi vida. A lo largo de mi vida, siempre estuvieron ahí para animarme, empujarme y creer en mí incluso cuando no era yo mismo. Ahora, al final de mi viaje de aprendizaje, sólo puedo expresarles mi profunda gratitud. Sin tu ayuda no habría podido superar los obstáculos que se me presentaron. Gracias a su amor y apoyo llegué aquí. No tengo suficientes palabras para agradecerte.

Muchas gracias, mamá y papá. Los quiero.

RESUMEN

El trabajo de investigación que realice se titula: “Modelo de Minería de Datos para la Predicción de Casos de Anemia en Madres Gestantes de la Provincia de Pasco - 2023” El objetivo principal fue desarrollar un modelo de extracción de datos para predecir la prevalencia de anemia entre mujeres embarazadas en el condado de Pasco en 2023 utilizando un diseño de estudio preexperimental con una variable y pocos controles. Esta información se refiere a la anemia entre las mujeres embarazadas en el condado de Pasco en 2022, pero el aumento se observó entre las mujeres embarazadas, enfocándose en su anemia y una muestra de 2,779 casos de anemia. Este proyecto de minería de datos probó tres algoritmos principales: red neuronal de perceptrón multicapa, Naive Bayes y árbol de decisión, y la técnica del árbol de decisión logró la mejor precisión: análisis de perceptrón multicapa 95 %, Naive Bayes 83 %, árbol de decisión. La tecnología alcanza el 99%.

Palabras Clave. Minería de datos, Predicción de casos de anemia en gestantes.

ABSTRACT

The research work I carried out is titled: "Data Mining Model for the prediction of anemia cases in pregnant mothers in the Province of Pasco - 2023" The main objective was to develop a data extraction model to predict the prevalence of anemia among pregnant women in Pasco County in 2023 using a pre-experimental study design with one variable and few controls. This information refers to anemia among pregnant women in Pasco County in 2022, but the increase was observed among pregnant women, focusing on their anemia and a sample of 2,779 anemia cases. This data mining project tested three main algorithms: multi-layer perceptron neural network, Naive Bayes and decision tree, and the decision tree technique achieved the best accuracy: multi-layer perceptron analysis 95%, Naive Bayes 83%, decision decision tree. Technology reaches 99%.

Keywords. Data mining, Prediction of anemia cases in pregnant women.

INTRODUCCIÓN

Para encontrar las herramientas necesarias es necesario definir el problema a abordar y, para ello, actualmente existen diversos marcos que facilitan el procesamiento de datos. Las redes neuronales tienen plataformas como Weka y otros softwares, se adaptan a cualquier problema y cuentan con bibliotecas para ayudar a encontrar la mejor solución posible y representarlas gráficamente.

El propósito de este estudio es desarrollar un modelo predictivo para determinar el porcentaje de casos de anemia en madres embarazadas.

Consta de los siguientes capítulos y las partes que a continuación se mencionan:

El capítulo 1, contiene las siguientes partes: identificación y definición del problema, limitaciones del estudio, planteamiento del problema, planteamiento de objetivos, justificación del estudio y limitaciones del estudio.

El capítulo 2, contiene las siguientes partes: teoría - base científica, definición del concepto, identificación de supuestos y variables y finalmente la definición operativa.

El capítulo 3, incluye tipos de investigación, métodos de investigación y contenido de la investigación. Diseño, poblaciones primarias y muestras, métodos y herramientas de recolección de datos, métodos de procesamiento y análisis de datos, procesamiento de datos estadísticos, selección de instrumentos de investigación, validación y confiabilidad, y principios éticos.

El capítulo 4, Resultados y discusión, que comprende las siguientes partes: descripción del trabajo, análisis e interpretación de resultados, prueba de hipótesis y tratamiento de resultados.

Al final se presentan las conclusiones, recomendaciones, referencias y anexos.

El autor.

ÍNDICE

DEDICATORIA.

AGRADECIMIENTO

RESUMEN

ABSTRACT

INTRODUCCIÓN

ÍNDICE

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

CAPITULO I

PROBLEMA DE INVESTIGACIÓN

1.1.	IDENTIFICACIÓN Y DETERMINACIÓN DEL PROBLEMA	1
1.2.	DELIMITACIÓN DE LA INVESTIGACIÓN.....	3
1.2.1.	Delimitación Espacial.....	3
1.2.2.	Delimitación temporal	4
1.2.3.	Delimitación conceptual.....	4
1.3.	FORMULACIÓN DEL PROBLEMA	4
1.3.1.	Problema general	4
1.3.2.	Problemas específicos.....	4
1.4.	FORMULACIÓN DE OBJETIVOS.....	4
1.4.1.	Objetivo general.....	4
1.4.2.	Objetivos específicos.....	4
1.5.	JUSTIFICACIÓN DE LA INVESTIGACIÓN.....	5
1.6.	LIMITACIONES DE LA INVESTIGACIÓN	5

CAPITULO II

MARCO TEÓRICO

2.1.	ANTECEDENTES DE ESTUDIO	6
2.1.1.	A nivel Internacional	6

2.1.2.	A nivel Nacional	8
2.1.3.	A nivel Local	10
2.2.	BASES TEÓRICAS – CIENTÍFICAS	11
2.2.1.	Minería de datos.....	11
2.2.2.	CRISP-DM (Proceso Estándar de la Industria Cruzada para la Minería de Datos	15
2.2.3.	Redes Neuronales	17
2.2.4.	Árbol de decisión J48	19
2.2.5.	Predicción de casos de anemia en gestantes	20
2.3.	DEFINICIÓN DE TÉRMINOS BÁSICOS.....	22
2.4.	FORMULACIÓN DE HIPÓTESIS	23
2.4.1.	Hipótesis General	23
2.4.2.	Hipótesis Específicas.....	24
2.5.	IDENTIFICACIÓN DE VARIABLES.....	24
2.6.	DEFINICIÓN OPERACIONAL DE VARIABLES E INDICADORES.....	24

CAPITULO III

METODOLOGÍA Y TECNICAS DE INVESTIGACIÓN

3.1.	TIPO DE INVESTIGACIÓN.....	25
3.2.	NIVEL DE INVESTIGACIÓN	25
3.3.	MÉTODOS DE INVESTIGACIÓN	25
3.4.	DISEÑO DE INVESTIGACIÓN.....	26
3.5.	POBLACIÓN Y MUESTRA	26
3.5.1.	Población	26
3.5.2.	Muestra.....	27
3.6.	TÉCNICAS E INSTRUMENTOS DE RECOLECCIÓN DE DATOS	27
3.7.	SELECCIÓN, VALIDACIÓN Y CONFIABILIDAD DE LOS INSTRUMENTOS DE INVESTIGACIÓN.	27
3.8.	TÉCNICAS DE PROCESAMIENTO Y ANÁLISIS DE DATOS.	27

3.9. TRATAMIENTO ESTADÍSTICO.....	27
3.10. ORIENTACIÓN ÉTICA FILOSÓFICA Y EPISTÉMICA.....	28

CAPITULO IV

RESULTADOS Y DISCUSIÓN

4.1. DESCRIPCIÓN DEL TRABAJO DE CAMPO	29
4.2. PRESENTACIÓN, ANÁLISIS E INTERPRETACIÓN DE RESULTADOS	30
4.3. PRUEBA DE HIPÓTESIS.....	45
4.4. DISCUSIÓN DE RESULTADOS.....	47

CONCLUSIONES

RECOMENDACIONES

REFERENCIAS BIBLIOGRÁFICAS

ANEXOS

ÍNDICE DE TABLAS

Tabla 1 Definición Operacional de Variables.....	24
Tabla 2 Diccionario de datos.....	32
Tabla 3 Resultado de las técnicas de estudio	45

ÍNDICE DE FIGURAS

Figura 1 Proporción de anemia (OMS) en gestantes que acuden a los establecimientos de salud según Diresa 2022,	2
Figura 2 Número de evaluadas y casos de anemia en gestantes que acudieron a los EESS, por grupos etarios 2022.....	2
Figura 3 Pasos de minería de datos para descubrir el conocimiento.....	12
Figura 4 CRISP-DM, sigue siendo la mejor metodología para proyectos de análisis, minería de datos o ciencia de datos.....	15
Figura 5 Metodología CRSP-DM.....	15
Figura 6 Partes del sistema de una red neuronal.....	17
Figura 7 Estructura de la red neuronal monocapa.....	18
Figura 8 Estructura de la red neuronal multicapa.....	18
Figura 9 Sobre ajuste.....	19
Figura 10 Representación gráfica de árbol de decisión.....	19
Figura 11 Valores normales de concentración de hemoglobina y niveles de anemia.	22
Figura 12 Acuerdo de Gobernabilidad.....	30
Figura 13 Modelo de negocio del departamento de Ginecología y Obstetricia	31
Figura 14.Datos de madres Gestantes de la Provincia de Pasco.....	32
Figura 15.Datos de madres Gestantes de la Provincia de Pasco en el SPSS.....	33
Figura 16.Análisis descriptivo de madres Gestantes de la Provincia de Pasco según Provincia.....	33
Figura 17.Gráfico de barras de madres Gestantes de la Provincia de Pasco según Provincia.....	34
Figura 18.Red MLP.....	35
Figura 19.Reporte de Métricas.....	36
Figura 20.Reporte de Naive Bayes	37
Figura 21.Reporte de Árbol de decisión	38
Figura 22.Árbol de decisión	44

CAPITULO I

PROBLEMA DE INVESTIGACIÓN

1.1. Identificación y determinación del problema

En cuanto a la anemia en niños menores de 3 años, alcanzó el 28,2 por ciento en el primer semestre de 2022, 0,5 puntos porcentuales menos que en el primer semestre de 2021. “En el primer semestre de 2022, el número de niños sometidos a pruebas de hemoglobina es un 10,2% menos que en el periodo correspondiente de 2021. Cabe decir que en el primer semestre de 2022 se registraron más de 90.000 niños menores de 3 años. En el primer semestre de 2021. También se puede observar que 2 departamentos regionales de salud presentaron la proporción de anemia clasificada como un problema importante de salud pública por la Organización Mundial de la Salud, es decir. Con más del 40% de anemia, la Diresa Huancavelica fue la más alta a nivel nacional y fue del 43,9%. 0.5 puntos porcentuales superior al primer semestre de 2021”. También es importante mencionar que solo Diresa Piura y Centro Diris Lima tuvieron proporciones inferiores al 20%, la OMS los clasificó como un problema de salud pública mediano.

Figura 1

Proporción de anemia (OMS) en gestantes que acuden a los establecimientos de salud según Diresa 2022,

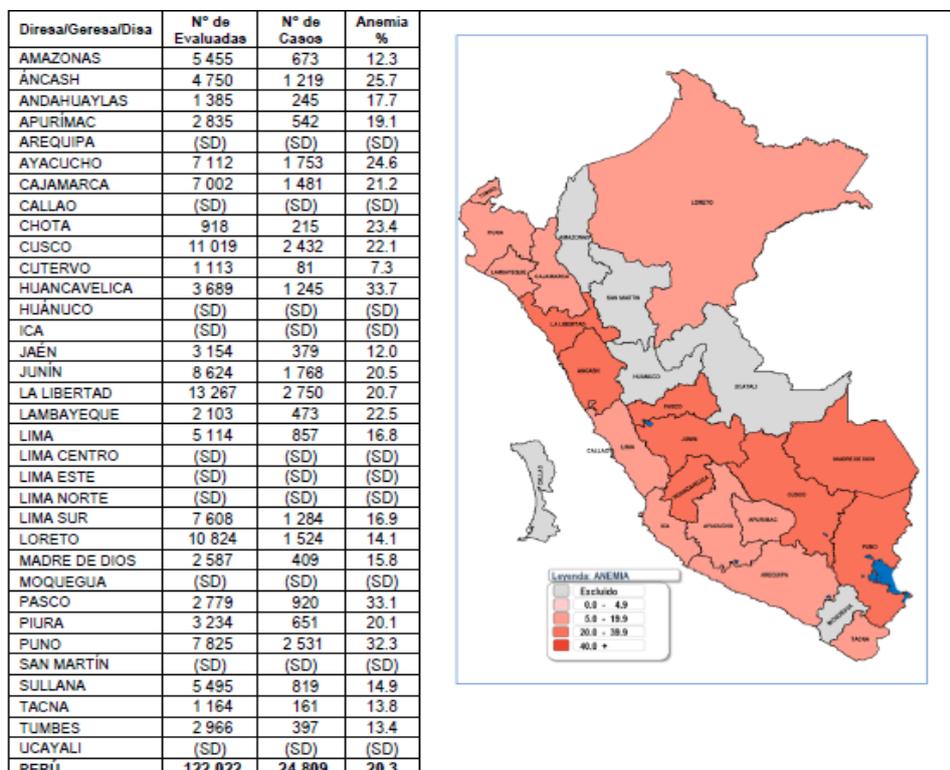
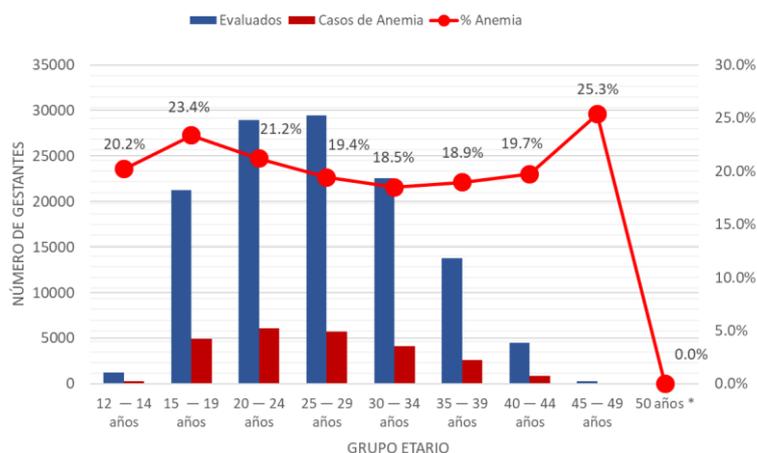


Figura 2

Número de evaluadas y casos de anemia en gestantes que acudieron a los EESS, por grupos etarios 2022.



La anemia durante el embarazo puede convertirse en un problema de salud tanto para la madre como para el feto, y las consecuencias a largo plazo pueden ser graves si no se toman las precauciones necesarias. Según la

Organización Mundial de la Salud, la deficiencia de hierro es responsable de una de cada cinco muertes perinatales y una de cada diez muertes maternas en los países en desarrollo. Por otro lado, en este estado de embarazo, debido al aumento del volumen plasmático, se presenta una evidente anemia, lo que hace que la sangre se expanda, reduciendo así el número de glóbulos rojos; también lo hará la demanda de hierro. Si este no es el caso. Si se cumple, causará anemia. Los datos actuales muestran que las mujeres embarazadas tienen un alto riesgo de sufrir anemia durante el parto, lo que afectará también al feto. Cabe señalar que en el Perú existen reportes de que la deficiencia de hierro afecta no solo el peso al nacer del niño y el estado inmunológico de la madre, sino que también aumenta el riesgo de muerte durante el embarazo y el parto. Por ello, para mejorar aún más la calidad de vida de las mujeres embarazadas y los recién nacidos, el Ministerio de Salud llama a fortalecer las estrategias de tratamiento y prevención de la anemia durante el embarazo. El propósito de este proyecto es proporcionar minería de datos. Un modelo cuyas predicciones se pueden utilizar para tomar decisiones estratégicas coherentes.

Esto permitirá la detección temprana y la predicción de la anemia en mujeres embarazadas comparando los datos médicos de mujeres embarazadas anémicas y no anémicas y centrándose en las mujeres embarazadas que tienen más probabilidades de desarrollar esta enfermedad. Utilizando esta información, el Ministerio de Salud (2019) puede implementar estrategias o campañas de prevención para ayudar a reducir o eliminar la proporción de mujeres embarazadas con la enfermedad.

1.2. Delimitación de la investigación.

1.2.1. Delimitación Espacial

Se realizó trabajo de investigación con los datos de los casos de anemia en madres gestantes de la Provincia de Pasco.

1.2.2. Delimitación temporal

La causa de investigación para la recolección de datos, el estudio de la interpretación de la información realizados en el año 2022.

1.2.3. Delimitación conceptual

Se buscará temas conceptuales del modelo de Minería de datos.

1.3. Formulación del problema

1.3.1. Problema general

¿Se podrá implementar un modelo de minería de datos para la predicción de casos de anemia en madres gestantes de la Provincia de Pasco - 2023?

1.3.2. Problemas específicos

¿Se podrá implementar un modelo de minería de datos aplicando la metodología CRISP-DM para la predicción de casos de anemia en madres gestantes de la Provincia de Pasco - 2023?

¿Se podrá validar el modelo de minería de datos a través de la validación cruzada para la predicción de casos de anemia en madres gestantes de la Provincia de Pasco - 2023?

1.4. Formulación de objetivos

1.4.1. Objetivo general

Implementar un modelo de minería de datos para la predicción de casos de anemia en madres gestantes de la Provincia de Pasco - 2023.

1.4.2. Objetivos específicos

Proponer un modelo de minería de datos para la predicción de casos de anemia en madres gestantes de la Provincia de Pasco aplicando CRISP-DM.

Validar el modelo de minería de datos aplicando validación cruzada para la predicción de casos de anemia en madres gestantes de la Provincia de Pasco - 2023.

1.5. Justificación de la investigación

Uno de los objetivos del MINSA “es brindar una mejor atención a sus pacientes, especialmente a las mujeres embarazadas. Se sabe que la gestión de los recursos de atención al paciente se basa en el promedio de la atención médica brindada y la información de apoyo a la decisión. Las predicciones de este trabajo permiten iniciativas para mejorar la atención de las mujeres embarazadas en los establecimientos de salud”.

Este trabajo es posible gracias a la información de registro y control materno acumulada en la base de “datos del sistema WawaRed, por lo que este estudio se justifica y su implementación es de gran importancia”.

1.6. Limitaciones de la investigación

Recursos financieros limitados para desarrollar, informar y validar afirmaciones de investigación.

CAPITULO II

MARCO TEÓRICO

2.1. Antecedentes de estudio

2.1.1. A nivel Internacional

Según Porto (2015) en su trabajo de investigación titulado “Diseño de un modelo para asistencia médica y predicción de preeclampsia en mujeres embarazadas” En la atención a las mujeres embarazadas se introdujo un modelo basado en un sistema de información que les permite gestionar mejor el embarazo e informar mejor a los médicos sobre la posible preeclampsia en las pacientes. Este proyecto utiliza técnicas de extracción de datos (p. ej., árboles de decisión, algoritmo C5.0, análisis de componentes principales de PCA, escalamiento de dimensiones no métricas de NMDS) para desarrollar un modelo de predicción de la preeclampsia en el segundo trimestre del embarazo. La propuesta de este modelo surgió del interés en desarrollar predictores de preeclampsia que permitieran a los obstetras revisar y monitorear fácilmente los resultados de las pacientes para respaldar la toma de decisiones. Una vez identificadas las variables predictoras más adecuadas, se selecciona el mejor método para realizar este proceso. El modelo se probó utilizando 100 conjuntos de datos simulados, y el 75% de las muestras correspondieron a los datos de entrenamiento y el 25% a los datos de validación. Los resultados obtenidos con

los métodos utilizados (C5.0, PCA, NMDS) se compararon con otros métodos de predicción encontrados en la literatura, mejorando la tasa de éxito.

Según Solano (2013) en su trabajo de investigación titulado “Aplicación de Minería de Datos para la Identificación de Patrones de comportamiento en las organizaciones enfocado en Prácticas de Impresión: Caso de Estudio” La investigación de minería de datos se realiza utilizando la metodología CRISP-DM, que incluye 6 pasos: conocimiento del negocio, comprensión de datos, preparación de datos, modelado, evaluación y transmisión final. Cada paso consta de un conjunto de actividades que permiten realizar el análisis de datos para lograr los objetivos del proyecto. En las primeras etapas del desarrollo del proyecto, la participación activa del cliente de la investigación es muy importante. Como el primero era el responsable de brindar información para comprender y descubrir el negocio e informar los análisis posteriores, se realizaron varias reuniones hasta cumplir los objetivos del primer paso.

Según Hernández (2015) en su trabajo de investigación titulado “Modelo de Minería de datos para identificación de patrones en el aprovechamiento académico” se hizo un análisis de la aplicación de técnicas de minería de datos para identificar patrones de comportamiento con el fin de predecir el fracaso escolar y el abandono. Se llevaron a cabo experimentos en la escuela secundaria privada, durante los cuales se identificaron las variables que afectan el éxito académico, Al comparar factores importantes para la toma de decisiones y acciones apropiadas, presentamos el modelo con el mejor desempeño. La implementación utiliza el método CRISP-DM, que divide de forma iterativa el proceso de exploración de datos en seis pasos interactivos. Analice el comportamiento de los estudiantes utilizando redes neuronales, árboles de decisión y modelos de agrupamiento de K-medias. La probabilidad del modelo se calcula en base a los datos experimentales que muestran que el modelo predictivo dio resultados positivos. Tomar decisiones basadas en el

conocimiento empresarial utilizando herramientas de minería de datos facilita enormemente la planificación administrativa, educativa y psicoeducativa para prevenir retrasos en los estudiantes y brindar un apoyo continuo.

2.1.2. A nivel Nacional

Según Carpio (2016) en su trabajo de investigación titulado “Modelo de predicción de la morosidad en el otorgamiento de crédito financiero aplicando metodología CRISP-DM” Para este escenario, se desarrolló un modelo para predecir la morosidad de los préstamos utilizando técnicas de minería de datos, utilizando datos históricos recopilados en la CRAC Los Andes e identificando variables clave utilizadas para desarrollar este estudio. Se utilizaron técnicas avanzadas como validación cruzada, matrices de error y curvas ROC para calibrar los modelos de calificación crediticia. Se evaluaron cuatro algoritmos de calificación crediticia para predecir los incumplimientos crediticios, siendo Random Forests el de mejor desempeño con un nivel de confianza general “del 82%; el uso de una metodología inteligente en relación con CRISP-DM fue importante. Para lograr el desarrollo del sistema web de predicción del riesgo crediticio SISMO”.

Según Condori (2019) en su trabajo de investigación titulado “Modelo de Minería de datos para la predicción de casos de anemia en gestantes de la Provincia de Ilo” En la actualidad la anemia es una enfermedad que afecta al 24.8% de la población mundial, siendo los más afectados los niños en edad preescolar y las madres gestantes, esta realidad se da en la mayoría de los países del mundo, dada su relevancia, existen muchas investigaciones abordadas desde diferentes perspectivas, entre ellas, desde el enfoque de la ciencia de la computación a través de su línea de investigación denominada minería de datos que consiste en investigaciones de predicción y clasificación utilizando los diferentes algoritmos. El propósito de esta investigación es desarrollar un modelo de minería de datos predictivo mediante la aplicación de

técnicas de aprendizaje automático, que han tenido un impacto positivo en la investigación en diversos campos como la medicina. En los análisis del contexto empresarial, se propusieron las tres técnicas anteriores para predecir los casos futuros de madres embarazadas anémicas, Se utilizan perceptrón multicapa, Naive Bayer y algoritmo de árbol de decisión J48 basado en entrenamiento de datos históricos. Entre los 422 datos sobre mujeres embarazadas anémicas en la provincia de Ilo, el algoritmo con mayor precisión fue Naive Bayes con un 89 %, el árbol de decisión J48 con un 79 % y el perceptrón multicapa con un 89 %. 62%. El desarrollo de este proyecto se basa en el método CRISP-DM, donde se desarrolla cada paso hasta llegar al resultado final..

Según Vela (2022) en su trabajo de investigación titulado “Implementación de un modelo de minería de datos para predecir la deserción de los clientes en una empresa de telecomunicaciones” se desarrolló un modelo predictivo haciendo uso de técnicas de minería de datos para analizar el comportamiento del cliente, con la finalidad de lograr identificar y clasificar a los clientes con mayor riesgo a desertar en una empresa de telecomunicaciones y así, apoyar a la empresa en la toma de decisiones certísimas y la creación de estrategias de retención. Para lograr el objetivo principal, se analizan Con base en las métricas de evaluación proporcionadas por el algoritmo XGBoost, en la literatura se han propuesto algoritmos básicos de minería de datos para determinar qué algoritmo es más adecuado para la situación actual y logra los mejores resultados. Logramos un 83% de precisión en la identificación de prospectos en riesgo de fracaso. Durante el desarrollo del módulo de predicción basado en el algoritmo seleccionado, se utilizó la metodología CRISP-DM para su creación, evaluación e implementación. La implementación del modelo se realizó utilizando la interfaz web nativa de Flask Framework, que permite a los usuarios finales generar informes globales personalizados basados en los lenguajes de programación JavaScript y Python. Finalmente, el grado de

aceptación del modelo estuvo determinado por dos indicadores. Se logró una eficiencia y precisión del 83%, una tasa de evaluación y una confiabilidad del 80%. Eficacia de la interfaz final en términos de usabilidad y rendimiento para pruebas de caja blanca y caja negra.

Según Mendoza (2023) en su trabajo de investigación titulado “Método alternativo para la detección de la anemia a través de sus factores asociados en mujeres en edad reproductiva: una aplicación de redes neuronales artificiales” tuvo como objetivo principal aplicar las redes neuronales artificiales como método alternativo para la detección de la anemia en mujeres en edad reproductiva a través de los factores asociados. Presentamos un método de predicción utilizando un enfoque cuantitativo y un diseño transversal no experimental. El método utilizado es un cuestionario y el instrumento es un cuestionario. Se utilizaron datos de la Encuesta de Población y Salud Familiar (ENDES) de 2020. En 2020 fueron 9.976.306 mujeres peruanas y la muestra estuvo conformada por 9.713 mujeres entre 15 y 49 años. Como método estadístico se utilizó una red neuronal artificial llamada Perceptrón Multicapa. La precisión del diagnóstico correcto fue del 81%. Finalmente, se concluyó que el método de redes neuronales artificiales puede ser utilizado como un método alternativo para la detección de anemia por factores relacionados en mujeres en edad fértil.

2.1.3. A nivel Local

Según Alania (2018) en su trabajo de investigación titulado “Aplicación de técnicas de minería de datos para predecir la deserción estudiantil de la facultad de ingeniería de la Universidad Nacional Daniel Alcides Carrión” La minería de datos se utiliza para explorar los datos disponibles en cualquier campo y encontrar la información que contienen. Métodos de clasificación como árboles de decisión, reglas, red bayesiana, etc. se puede aplicar a datos académicos universitarios para predecir el comportamiento de los estudiantes,

resultados de exámenes, abandono de estudiantes, etc. Esta predicción ayuda a las autoridades a identificar la deserción estudiantil y determinar los ratios y otras previsiones. actividades El algoritmo de árbol de decisión C4.5 (J48) se aplica a los datos de calificaciones del último semestre de los estudiantes para predecir si abandonarán la escuela. La puntuación del árbol de decisión predijo el número de estudiantes con probabilidades de abandonar sus carreras profesionales. El resultado se puede utilizar para mejorar la toma de decisiones para las acciones de las autoridades. Después de la evaluación con datos sin procesar, se ingresa un conjunto de datos de prueba en el sistema para analizar los resultados. Un análisis comparativo de los resultados muestra que la previsión ayudó a determinar con mayor precisión la mejora del resultado. Para analizar la precisión del algoritmo, se compara con el algoritmo Random Tree Tree y se descubre que es igualmente efectivo en términos tanto de la precisión de los resultados de aprendizaje del estudiante como del tiempo necesario para crear el árbol.

2.2. Bases teóricas – científicas

2.2.1. Minería de datos.

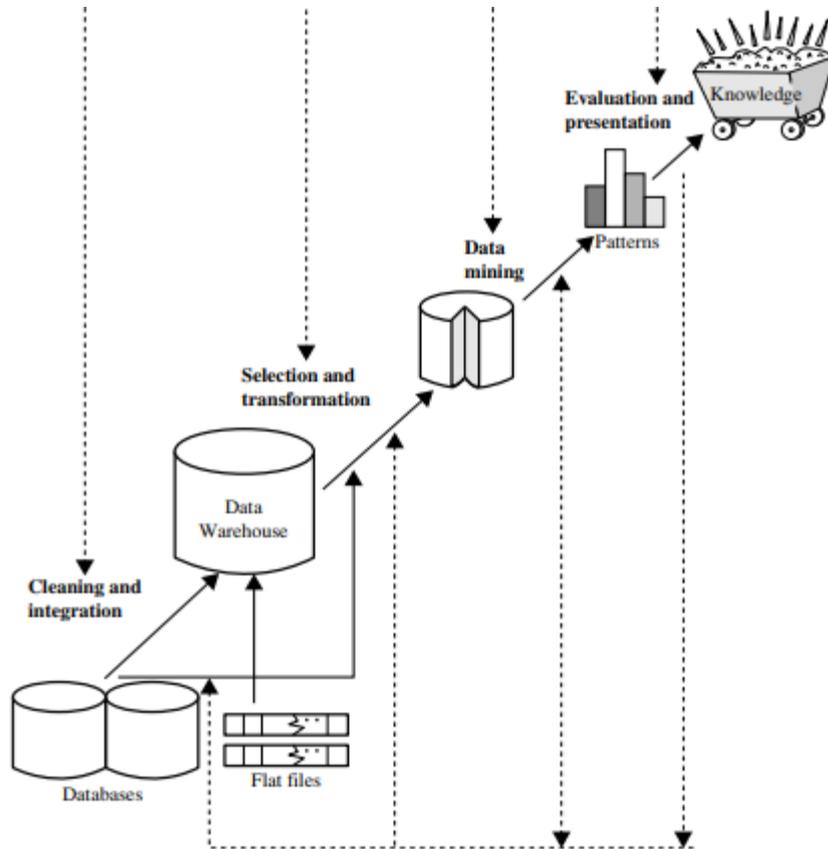
Algunos autores autorizados llaman al siglo XXI la era de la información porque el trabajo diario de las personas genera grandes cantidades de información almacenada en bases de datos de diversos sistemas de información, y esta cantidad de información valiosa puede ayudar a las instituciones a tomar decisiones precisas. El tema de este estudio es predecir casos de anemia en mujeres embarazadas en instituciones médicas.

En este argumento la tecnología de minería de datos tiene un rol muy significativo, según (Leskovec, Rajaraman, & Ullman, 2014) la definición más comúnmente aceptada de "minería de datos" es el descubrimiento de "modelos" para datos". Sin embargo, para (Han, Kamber, & Pei, 2012) minería de datos es

“descubrimiento de conocimiento a partir de datos”, además plantea una serie de pasos de minería de datos, que continuación se presenta.

Figura 3

Pasos de minería de datos para descubrir el conocimiento.



1. Limpieza de datos: Una fase que consiste en remover datos inconsistentes y ruidosos.
2. Integración de data: Donde se puede combinar diferentes datos
3. Selección de data: Una fase donde los datos relevantes son recuperadas de las bases de datos para el análisis
4. Transformación de data: Una fase donde la data es transformada y consolidada en un formato apropiado para la operación con técnicas de minería de datos
5. Minería de datos: La fase esencial de aplicación de técnicas para extracción de patrones interesantes

6. Modelo de evaluación: La fase donde se identifica a patrones de representación de conocimientos basada en características
7. Presentación de conocimiento La fase donde se identifica a patrones de representación de conocimientos y visualización

Encontrar patrones de comportamiento a partir de los big data recopilados en las empresas es posible mediante la minería de datos, que puede utilizarse para fines adicionales, como tomar buenas decisiones. Luego puede definirse como el proceso de descubrir relaciones de patrones nuevas y significativas en grandes cantidades de datos.

Sin embargo, este estudio fue desarrollado con base la metodología CRISP-DM es una metodología probada utilizada por muchos investigadores.

Técnicas de Minería de Datos:

- **Técnicas descriptivas**

Las técnicas descriptivas tienen el objetivo de descubrir y explicar la relación de variables, “estas mostrarán nuevas relaciones entre las variables o excepciones de acuerdo a la empresa en que se utilice este proceso” (Virsedá Benito & Román Carrillo).

- **Descripción de clases**

Hay tres de estas características: Descripción de datos, que recopila características comunes, e Identificación de datos, que compara las características comunes de una clase de objetos con otra.

- **Análisis de asociación**

Es el descubrimiento de reglas de asociación de datos, el proceso de buscar correlaciones significativas en un conjunto de datos.

- **Análisis de clusters.**

“En síntesis, análisis de asociación, consiste en la agrupación de objetos maximizando la similitud dentro de una clase y minimizando la similitud entre clases”.

Técnicas predictivas

Las técnicas predictivas según (Espino Timón, 2017) “consiste en la extracción de información existente en los datos y su utilización para predecir tendencias y patrones de comportamiento, pudiendo aplicarse sobre cualquier evento desconocido, ya sea en el pasado, presente o futuro”, a continuación, se presentan algunas técnicas predictivas más destacadas.

- **Clasificación y predicción.**

“Estos son métodos de clasificación de categorías basados en datos variables. También puede predecir nuevos objetos en determinadas categorías de modelos. Son adecuados tanto para datos cuantitativos como categóricos”.

- **Árboles de decisión.**

Existen diferentes definiciones de árboles de decisión, según (Unidad de Análisis e Información Financiera, 2014) “Un árbol de decisión es un modelo de clasificación que divide un conjunto de análisis y persigue grupos de resultados de la más alta pureza”

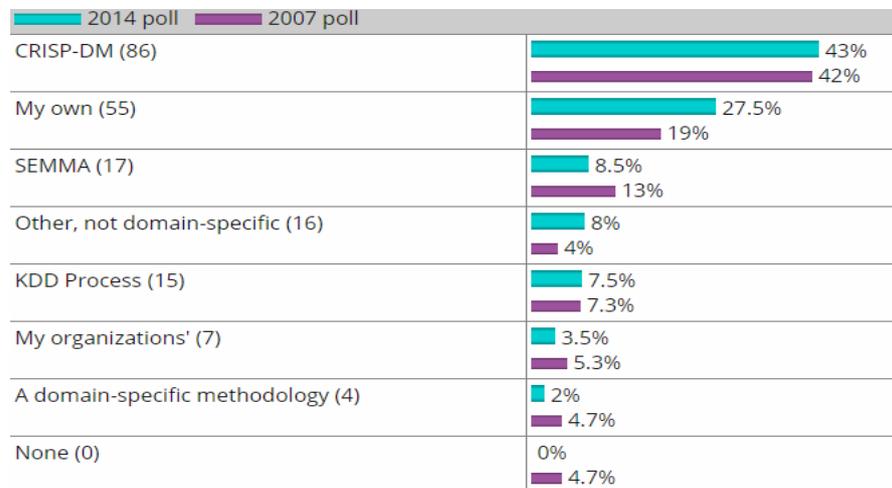
- **Redes neuronales.**

Según (Espino Timón, 2017) “Las redes neuronales se utilizan cuando no se conoce la naturaleza exacta de la relación entre los valores de entrada y de salida”, sin embargo, en los últimos tiempos “el término red neuronal ha evolucionado para abarcar una gran clase de modelos y métodos de aprendizaje” (Friedman, Tibshirani, & Hastie, 2017).

2.2.2. CRISP-DM (Proceso Estándar de la Industria Cruzada para la Minería de Datos)

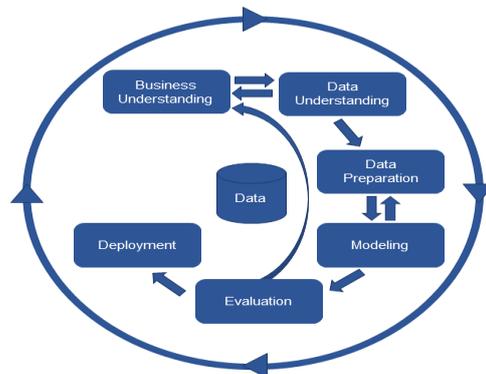
Según (Piatetsky, s.f.), “en la encuesta realizada por KDnuggets, la metodología más usada para desarrollo de proyectos de minería de datos y ciencia de datos es CRISP-DM”, como se puede apreciar en la Figura 4.

Figura 4
CRISP-DM, sigue siendo la mejor metodología para proyectos de análisis, minería de datos o ciencia de datos



El método más popular para desarrollar proyectos de minería de datos consta de 6 pasos, cada uno de los cuales conduce al correcto desarrollo del proyecto en este argumento, como se muestra en la Figura 5.

Figura 5
Metodología CRSP-DM



1. Comprensión de negocio.

Desarrollando un proyecto de minería de datos, es imprescindible tener un buen conocimiento del modelo de negocio de la industria en la que se desarrolla el proyecto de minería de textos.

Principalmente "comprender los objetivos y requisitos comerciales desde una perspectiva empresarial" (Wirth y Hipp), también planificar los objetivos del proyecto según el contexto.

2. Comprensión de data.

Esta es la etapa donde comienza la recopilación de datos, analizando e identificando datos cualitativos e identificando problemas potenciales.

3. Preparación de data

Es el asunto de estructurar conjuntos de datos, pero los pasos principales incluyen realizar "selección de datos, limpieza de datos, reestructuración de datos, integración de datos" (Brown, s.f.) y normalización de datos, donde se pueden aplicar algoritmos de minería de datos.

4. Modelado.

Esta es la fase donde se utilizan técnicas de minería de datos para comprender el procedimiento del modelo, además se planifican experimentos, pero principalmente se construye el modelo y finalmente se evalúa el modelo seleccionado para el proyecto.

5. Evaluación.

En esta etapa se evalúan los resultados obtenidos al aplicar el modelo de investigación de datos, se evalúa el proceso y finalmente se dan nuevos pasos.

6. Desarrollo.

Es el proceso de integrar un proyecto al negocio utilizando una metodología de integración, pero los resultados primero deben ser comunicados y evaluados a los involucrados en el negocio.

2.2.3. Redes Neuronales

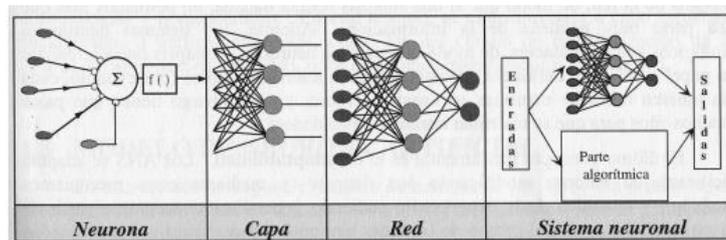
Las definiciones varían, biológicamente es "una unidad de procesamiento simple que recibe e integra señales desde y hacia otras neuronas" (Basogain Olabe).

Básicamente, una neurona recibe un conjunto de entradas que representan las salidas de otras neuronas. Cada entrada se multiplica por un peso equivalente correspondiente a su "peso" o grado de conexión sináptica. Se suman todas las entradas ponderadas y se determina el nivel de excitación o activación de la neurona. El vector de activación básico de una neurona artificial se expresa mediante la siguiente fórmula.

Una red neuronal es un sistema compuesto por neuronas, capas y redes como se muestra en la figura.

Figura 6

Partes del sistema de una red neuronal.

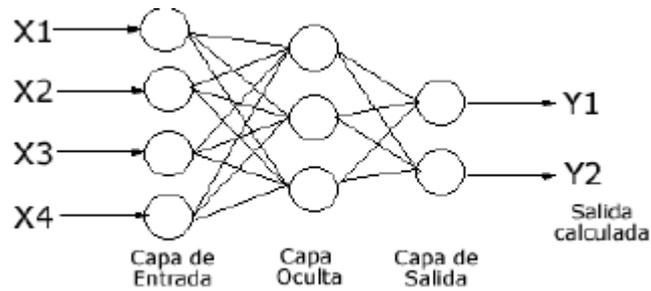


a. Redes neuronales artificiales monocapa.

Es el tipo de red neuronal conformada de una sola capa, según (Jorge Matich, 2001) "en las redes monocapa, se establecen conexiones entre las neuronas que pertenecen a la única capa que constituye la red. Las redes mono capas se utilizan generalmente en tareas relacionadas con lo que se conoce como auto asociación (regenerar información de entrada que se presenta a la red de forma incompleta o distorsionada)".

Figura 7

Estructura de la red neuronal monocapa.

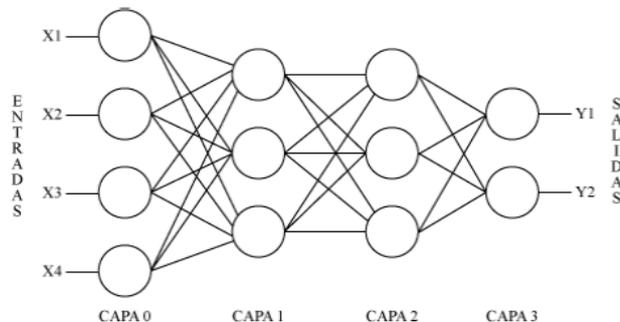


b. Redes neuronales artificiales multicapa

"Las redes neuronales multicapa constan de varias capas y están conectadas entre sí. Según (Basogain Olabe) la salida de una capa es la entrada a la siguiente capa. Se ha demostrado que las redes multicapa funcionan mejor que las redes de una sola capa en calidad y orientación de ruta"

Figura 8

Estructura de la red neuronal multicapa.

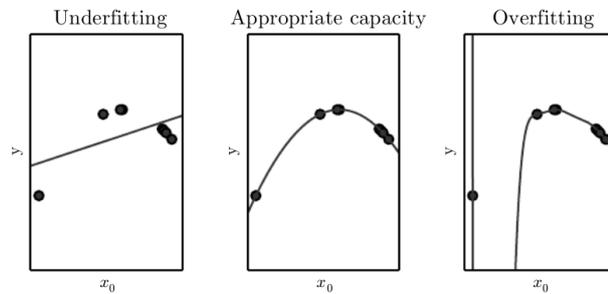


Las redes neuronales se vuelven más complejas según la cantidad de capas ocultas y la cantidad de neuronas que contienen. Esta complejidad significa que, si bien el algoritmo es confiable, el rendimiento puede variar cuando se utilizan hiperparámetros como el procesamiento de datos.

En la aplicación de redes neuronales, existe un problema común llamado bajo rendimiento, es la adaptación del aprendizaje, que generalmente tiene problemas de sobreajuste y desajuste, un factor es la pequeña cantidad de datos en la muestra de entrenamiento, y se acabó. - Cuando se corrige mediante

algoritmos de optimización, el error cuadrático medio también se utiliza para obtener un valor grande y se puede minimizar retrocediendo en el tiempo.

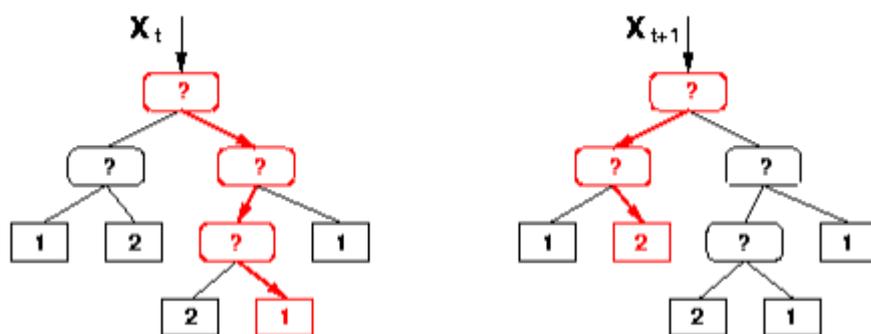
Figura 9
Sobre ajuste.



2.2.4. Árbol de decisión J48

Un algoritmo de árbol de decisiones que puede analizar decisiones basándose en el uso de resultados y probabilidades asociadas, “es un conjunto de condiciones o reglas organizadas en una estructura jerárquica, de tal manera que la decisión final se puede determinar siguiendo las condiciones que se cumplen desde la raíz hasta alguna de sus hojas”. (Vizcaino Garzon, 2008)

Figura 10
Representación gráfica de árbol de decisión.



Un árbol de decisión consta de nodos de decisión, representados por cuadrados, que indican que se debe tomar una decisión en esta etapa del proceso. Los nodos que representan la probabilidad de un evento aleatorio en

ese punto del proceso se muestran como círculos. Ramas que muestran diferentes caminos, ya sean deterministas o aleatorios.

“Otro algoritmo de árbol de decisión J48 utiliza el método Gini para crear puntos divididos. Para hallar el valor de Gini se aplica de la siguiente fórmula”:

$$Gini(D) = 1 - \sum_{i=1}^m p_i^2$$

Donde, “ p_i es la probabilidad de que una tupla en D pertenezca a la clase C_i ”.

El índice de Gini tiene en cuenta la distribución binaria de cada atributo. Podrás calcular la cantidad ponderada de aditivos en cada apartado. Si una partición binaria del atributo A divide los datos D en D_1 y D_2 , entonces D es el índice de Gini:

$$Gini_A(D) = \frac{|D_1|}{|D|} Gini(D)_1 + \frac{|D_2|}{|D|} Gini(D)_2 + \frac{|D_n|}{|D|} Gini(D)_n$$

“Para un atributo de valor discreto, el subconjunto que produce el índice de Gini mínimo seleccionado se elige como atributo de distribución. Para atributos valorados continuamente, la estrategia es seleccionar cada par de valores adyacentes como un posible punto de separación con el índice de Gini más pequeño como ese punto”.

2.2.5. Predicción de casos de anemia en gestantes

Anemia:

“La anemia es una enfermedad de la sangre líquida que afecta la salud y calidad de vida de las personas, existen muchos tipos de anemia, como: anemia ferropénica, anemia perniciosa, anemia aplásica y anemia hemolítica. Esta enfermedad puede afectar a personas de diferentes edades”.

Causas de anemia:

“Los glóbulos rojos contienen hemoglobina, una proteína que transporta oxígeno por el cuerpo del ser humano, cuando mejor producción de estos glóbulos rojos, o se destruye demasiados, puede generar malestar al organismo del ser humano y por ende causar cansancio otros síntomas” (US Department of Health and Human Services, 2011).

A continuación, se presenta algunas posibles signos y síntomas de la anemia:

- Cansancio o debilidad
- Piel pálida o amarilla
- Desaliento o mareo
- Sed en exceso
- Sudor
- Pulso débil y rápido
- Respiración rápida
- Problemas en el corazón

La anemia por deficiencia de hierro durante el embarazo también se asocia con bajo peso al nacer y depresión posparto. Algunos estudios también muestran que los bebés tienen un mayor riesgo de morir antes o poco después del nacimiento. El informe de Perú sigue estándares técnicos para el tratamiento y prevención de la anemia en niños, adolescentes, mujeres embarazadas y puérperas:

Figura 11

Valores normales de concentración de hemoglobina y niveles de anemia.

Población	Con Anemia Según niveles de Hemoglobina (g/dL)			Sin anemia según niveles de Hemoglobina
	Severa	Moderada	Leve	
Niños				
Niños Prematuros				
1ª semana de vida		≤ 13.0		>13.0
2ª a 4ta semana de vida		≤ 10.0		>10.0
5ª a 8va semana de vida		≤ 8.0		>8.0
Niños Nacidos a Término				
Menor de 2 meses		< 13.5		13.5-18.5
Niños de 2 a 6 meses cumplidos		< 9.5		9.5-13.5
	Severa	Moderada	Leve	
Niños de 6 meses a 5 años cumplidos	< 7.0	7.0 - 9.9	10.0 - 10.9	≥ 11.0
Niños de 5 a 11 años de edad	< 8.0	8.0 - 10.9	11.0 - 11.4	≥ 11.5
Adolescentes				
Adolescentes Varones y Mujeres de 12 - 14 años de edad	< 8.0	8.0 - 10.9	11.0 - 11.9	≥ 12.0
Varones de 15 años a más	< 8.0	8.0 - 10.9	11.0 - 12.9	≥ 13.0
Mujeres NO Gestantes de 15 años a más	< 8.0	8.0 - 10.9	11.0 - 11.9	≥ 12.0
Mujeres Gestantes y Puérperas				
Mujer Gestante de 15 años a más (*)	< 7.0	7.0 - 9.9	10.0 - 10.9	≥ 11.0
Mujer Puérpera	< 8.0	8.0 - 10.9	11.0 - 11.9	≥ 12.0

2.3. Definición de términos básicos.

- **Árbol de decisión**

Un árbol de decisiones es un mapa de los posibles resultados de una serie de decisiones. Permite a un individuo u organización comparar posibles acciones en función de sus costos, probabilidades y beneficios.

- **Redes neuronales**

En aprendizaje de máquina, la propagación hacia atrás de errores o retro programación es un algoritmo de aprendizaje supervisado utilizado para entrenar redes neuronales prealimentadas.

- **Red neuronal convolucional**

Una red neuronal convolucional es un tipo de red neuronal artificial en la que las neuronas artificiales responden a campos receptivos de la misma manera que las neuronas de la corteza visual primaria del cerebro biológico.

- **Conjunto de algoritmos**

El aprendizaje profundo es un conjunto de algoritmos de aprendizaje automático que intentan modelar abstracciones de datos de alto nivel utilizando arquitecturas informáticas que admiten múltiples

transformaciones iterativas no lineales de datos expresados en forma matricial o tensorial.

- **Data Mining**

El Data Mining o minería de datos Es un conjunto de tecnologías que hacen que el análisis de datos sea mucho más accesible. Con la ayuda de la minería de datos, gracias al uso de diversos algoritmos matemáticos, es posible extraer datos para identificar ciertos patrones de comportamiento.

- **Machine Learning**

El Machine Learning es una rama de la inteligencia artificial que utiliza algoritmos para permitir que las computadoras reconozcan patrones en datos masivos y hagan predicciones (análisis predictivo).

- **Perceptrón Multicapa**

Un perceptrón multicapa es una red neuronal artificial que consta de múltiples capas, por lo que tiene la capacidad de resolver problemas que no son separables linealmente, que es la principal limitación de un perceptrón. Un perceptrón multicapa puede conectarse total o localmente.

- **Red Neuronal Recurrente**

Red Neuronal Recurrente: La estructura de una red neuronal artificial es relativamente simple y está relacionada principalmente con la multiplicación de matrices.

2.4. Formulación de Hipótesis

2.4.1. Hipótesis General

La Implementación de un modelo de minería de datos predecirá los casos de anemia en madres gestantes de la Provincia de Pasco - 2023.

2.4.2. Hipótesis Específicas

Aplicando la metodología CRISP-DM se pudo implementar un modelo de minería de datos para la predicción de casos de anemia en madres gestantes de la Provincia de Pasco-2023.

Aplicando la validación cruzada se podrá validar el modelo de minería de datos para la predicción de casos de anemia en madres gestantes de la Provincia de Pasco-2023.

2.5. Identificación de Variables

2.5.1. Variables independientes

Minería de Datos.

2.5.2. Variables dependientes

Predicción de casos de anemia en gestantes.

2.6. Definición Operacional de variables e indicadores

Tabla 1

Definición Operacional de Variables

VARIABLE	DEFINICIÓN CONCEPTUAL	DIMENSIONES
INDEPENDIENTE Minería de Datos	Es el proceso de hallar anomalías, patrones y correlaciones en grandes conjuntos de datos para predecir resultados. Empleando una amplia variedad de técnicas, puede utilizar esta información para incrementar sus ingresos, recortar costos, mejorar sus relaciones con clientes, reducir riesgos y más.	- Aplicando la metodología CRISP-DM. - Validación cruzada.
DEPENDIENTE Predicción de casos de anemia en gestantes	Es la falta de hierro en la sangre, que es necesario para producir hemoglobina (la parte de la sangre que distribuye oxígeno desde los pulmones hacia los tejidos del cuerpo).	- Reportes de madres gestantes

Nota. La tabla muestra la definición de variables

CAPITULO III

METODOLOGÍA Y TECNICAS DE INVESTIGACIÓN

3.1. Tipo de Investigación

El tipo de la investigación del presente estudio es aplicada. Según Ortega (2017), es el estudio científico que busca resolver un problema o planteamiento específico, que se caracteriza por buscar la aplicación o utilización de los conocimientos que se adquieren.

3.2. Nivel de investigación

Este trabajo de investigación “adopta un nivel cuantitativo porque se basa en el estudio y análisis de la realidad a través de diversos procedimientos basados en mediciones”.

3.3. Métodos de investigación

La investigación que realizo es analítica e inductiva y soy responsable de descubrir las causas raíces de los incidentes a través del descubrimiento de relaciones causa-efecto. En este sentido, la investigación explicativa también puede ocuparse de determinar las causas (investigación post facto), como de los efectos (investigación experimental), mediante la prueba de hipótesis. Sus resultados y conclusiones constituyen el nivel más profundo de conocimientos” según (Hernández Sampieri & Mendoza Torres, 2018).

3.4. Diseño de investigación

El proyecto está basado en una investigación pre-experimental, en este tipo de investigación se analiza una sola variable y prácticamente no existe ningún tipo de control. No existe manipulación de la variable independiente ni se utiliza grupo control (Ávila Baray, 2006). En una investigación pre-experimental el tipo de diseño consiste en administrar un tratamiento a un determinado grupo y después aplicar una medición en una o más variables orientado a identificar algunos patrones que permitan pronosticar un aproximado de nacimientos para el MINSA de la provincia de Pasco. Para el diseño de contrastación de la hipótesis, se usará lo siguiente.

G X O

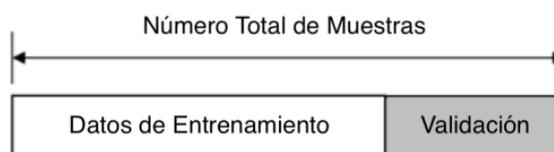
Donde:

O: Medición de los sujetos del grupo (Entrenamiento, validación)

G: Grupo de sujetos (Grupo1, Grupo2)

X: Aplicación del Estímulo (Modelo de minería de datos)

Aplicando las técnicas de machine learning se pretende realizar un diseño que entrene y valide datos del algoritmo propuesto.



3.5. Población y muestra

3.5.1. Población

Según Hernández Sampieri (2018) menciona que: “Es un grupo de personas de la industria que nos apoyan en la obtención de muestras y resultados”.

Los datos incluyen casos 3174 mujeres con anemia durante el período 2022 Provincia de Pasco, pero aumentaron el número de mujeres embarazadas y se centraron en sus casos de anemia.

3.5.2. Muestra

Para Arias, Villasis y Miranda (2016) comento que “la muestra es una parte de la población que es característico de un todo y se opta por adquirir información acerca de las variables”.

Se tomo una muestra por conveniencia de 2779 casos de anemia en gestantes de las provincias de Pasco (Daniel Alcides Carrión, Oxapampa y Pasco), recolectando información de todos los centros de salud con casos de anemia en gestantes.

3.6. Técnicas e instrumentos de recolección de datos

“Datos recopilados del 2022 en Wawared, bajo previo coordinación y autorización de dato para la presente investigación”.

3.7. Selección, validación y confiabilidad de los instrumentos de investigación.

Base de datos con casos reales de madres gestantes con o sin anemia que se encuentran en la página Wawared de la provincia de Pasco.

3.8. Técnicas de procesamiento y análisis de datos.

Los datos lo encontramos en el Microsoft Excel la cual para el estudio se paso al Software SPSS 26 para luego analizar descriptivamente los datos.

Usare el Software Weka 3.8.6. para realizar el modelo de los algoritmos.

3.9. Tratamiento Estadístico.

Se aplicaron las fases de la metodología CRISP–DM:

- Comprensión del negocio
- Comprensión de data
- Preparación de data

- Modelado
 - Evaluación

3.10. Orientación ética filosófica y epistémica.

En la presente investigación se destacó:

Mantendremos el formulario de registro confidencial y garantizaremos el anonimato, por lo tanto, su perfil será ignorado. Información transparente sobre los resultados., es decir no se añade información adicional o distorsionada sobre el estudio.

CAPITULO IV

RESULTADOS Y DISCUSIÓN

4.1. Descripción del trabajo de campo

El trabajo de investigación se basa en el análisis de los datos recopilados de casos reales de madres gestantes con anemia y/o sin anemia por Wawared de la provincia de Pasco (Daniel Alcides Carrión, Oxapampa y Pasco) para ello se vio el Acuerdo de Gobernabilidad para el Desarrollo Integral de la Región Pasco 2023-2026.

Estando presentes o acercándonos a los planes de gobierno presentados en el proceso electoral regional en curso, nos comprometemos a promover el seguimiento conjunto del Estado y la sociedad civil sobre el avance de la implementación de este acuerdo de gobernanza, para lo cual utilizaremos la matriz de seguimiento que se acordó. La mesa redonda del Paseo sobre la lucha contra la pobreza a partir de una propuesta al efecto. La siguiente imagen explica lo siguiente:

Figura 12

Acuerdo de Gobernabilidad.

Principales políticas	Principales resultados	Principales indicadores
	R. Se mejora el acceso a establecimientos de salud por parte de las personas adultas mayores.	<ul style="list-style-type: none"> • Establecimientos de salud abastecidos con medicamentos. • Adultos mayores atendidos por especialistas de geriátrica y gerontológica.
Garantizar la continuidad del crecimiento y una vida saludable sin anemia y desnutrición crónica, desde la concepción.	R. Se mejora la nutrición y el estado de la salud en mujeres en edad fértil, gestantes y niños y niñas.	<ul style="list-style-type: none"> • Madres gestantes con anemia.
	R. Se incrementa la lactancia materna exclusiva.	<ul style="list-style-type: none"> • Recién nacidos que reciben lactancia materna en la primera hora y exclusiva hasta los seis meses.
	R. Se reduce la desnutrición crónica infantil en menores de 5 años de edad.	<ul style="list-style-type: none"> • Niños y niñas menores de 5 años con desnutrición crónica (OMS³). • Niños y niñas menores de 36 meses con control de CRED⁴ completo.

4.2. Presentación, análisis e interpretación de resultados

4.2.1. Comprensión de Negocio.

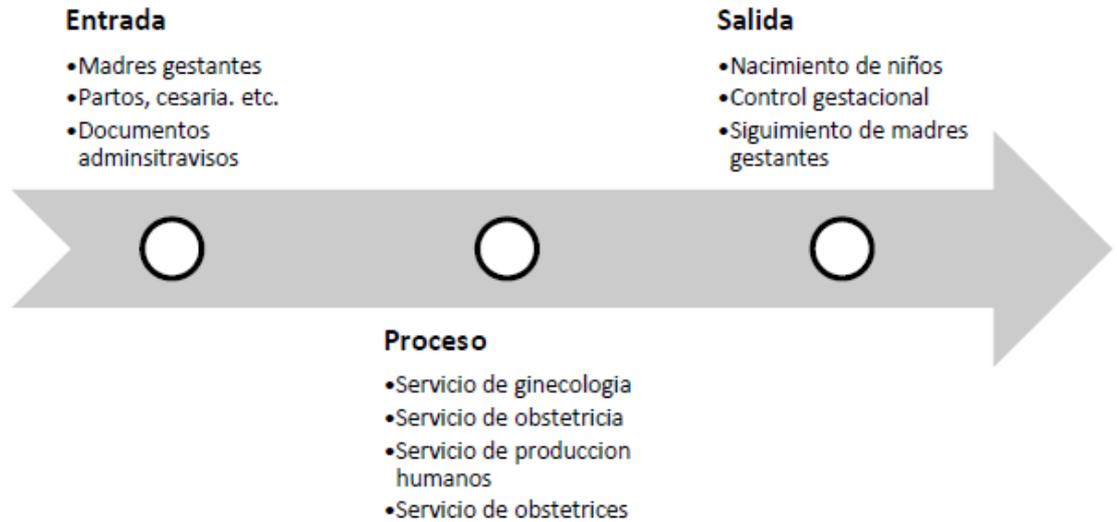
“De acuerdo al organigrama del Departamento de Ginecología y Obstetricia la máxima autoridad es el director de programas de la sucursal, seguido del técnico administrativo y otras subdivisiones de ser necesario, el departamento tiene como objetivo brindar cuatro tipos de servicios, servicios ginecológicos, obstetricia, servicio de reproducción humana y obstetricia, la dirección estratégica está integrada por el director de la rama de ginecología y obstetricia, técnicos administrativos de apoyo a la actividad o servicio principal, obstetras y ginecólogos de acuerdo al organigrama del departamento, así lo establece el consejo. el técnico tiene un papel de apoyo muy importante en la toma de decisiones estratégicas del sector, porque guarda toda la información en el sistema Wawared, y el objetivo de este proyecto es utilizar la información de madres embarazadas con y sin anemia para predecir casos de anemia en mujeres embarazadas en el condado de Pasco”.

Estas y otras funciones especiales tienen por objeto: Recepción, almacenamiento, clasificación y transmisión oportuna de los documentos entrantes y preparados por el departamento para decisión de la sede. Realiza y

supervisa servicios de secretaría muy complejos de acuerdo con los objetivos operativos de la institución.

Figura 13

Modelo de negocio del departamento de Ginecología y Obstetricia



4.2.2. Comprensión de data

Recolección de data y descripción

La 1ª fase de la metodología de CRISP-DM describe la adaptación de los objetivos del proyecto al trabajo de investigación, por lo que la recolección de datos se recopiló a través del sistema Wawared, abarcando los años 2018-2022, la información mencionada está en línea con los objetivos del proyecto. El proyecto y los datos obtenidos se utilizan para analizarlo e interpretar el modelo diseñado.

Recolectando información con los indicadores expuestos:

Tabla 2.

Diccionario de datos

Variable	Descripción
EESS	Establecimiento de Salud
Edad	Edad de la madre gestante
Edad_Gestacional	Meses de gestación
Peso	Peso de la madre gestante
Talla	Talla de la madre gestante
Tipo de embarazo	Tipo de embarazo de la madre gestante
Distrito	Lugar de atención
Hemoglobina	Hemoglobina de la madre gestante
Dx_Anemia	Clasificador de anemia

4.2.3. Preparación de data

Para preparar los datos, todos los registros se describieron estadísticamente determinando la media y la desviación estándar gráficos de datos también se presentan en un formato prescrito para garantizar la coherencia entre las secciones. En esta etapa, el factor clave es la muestra en la categoría minoritaria y ésta se identifica durante la observación del material.

Figura 14.

Datos de madres Gestantes de la Provincia de Pasco

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
EESS	RENIPRES	Fecha	Edad	Edad_Gest	Peso	Talla	PPG	Tipo_Emb	UBIGEO	Provincia	Distrito	Altitud_D	Hemoglol	Hbc	Dx_Anemia
2	C.S. ULUACHI	1009	1/03/2022	35	38	67	149	54.5	190101 PASCO	CHAUPIMAR	4338	14.5	10.37	Anemia Leve	
3	C.S. ULUACHI	1009	1/03/2022	31	14	51.2	142	51.5	190101 PASCO	CHAUPIMAR	4338	15.9	11.77	Normal	
4	C.S. ULUACHI	1009	1/04/2022	35	38	66.8	147	56.5	190101 PASCO	CHAUPIMAR	4338	13.5	9.37	Anemia Moderada	
5	C.S. ULUACHI	1009	1/04/2022	37	34	50.4	145	46.5	190101 PASCO	CHAUPIMAR	4338	15.8	11.67	Normal	
6	C.S. ULUACHI	1009	1/04/2022	31	35	69.2	154	66.5	190101 PASCO	CHAUPIMAR	4338	11.5	7.37	Anemia Moderada	
7	C.S. ULUACHI	1009	1/05/2022	31	34	56.7	160	54.5	190101 PASCO	CHAUPIMAR	4338	15.1	10.97	Anemia Leve	
8	C.S. ULUACHI	1009	1/05/2022	19	35	52	142	43.5	190101 PASCO	CHAUPIMAR	4338	15.1	10.97	Anemia Leve	
9	C.S. ULUACHI	1009	1/05/2022	23	39	65	152	47.5	190101 PASCO	CHAUPIMAR	4338	15.1	10.97	Anemia Leve	
10	C.S. ULUACHI	1009	1/06/2022	32	29	66.1	156	56.5	190101 PASCO	CHAUPIMAR	4338	13.1	10.97	Anemia Leve	
11	C.S. ULUACHI	1009	1/06/2022	19	39	71.2	154	53.5	190101 PASCO	CHAUPIMAR	4338	13.2	9.07	Anemia Moderada	
12	C.S. ULUACHI	1009	1/07/2022	26	30	69.3	148	64.5	190101 PASCO	CHAUPIMAR	4338	15.5	11.37	Normal	
13	C.S. ULUACHI	1009	1/07/2022	21	38	59.2	162	59.5	190101 PASCO	CHAUPIMAR	4338	14.8	10.67	Anemia Leve	
14	C.S. ULUACHI	1009	1/11/2022	23	32	53.4	140	42.5	190101 PASCO	CHAUPIMAR	4338	14.8	10.67	Anemia Leve	
15	C.S. ULUACHI	1009	1/14/2022	30	7	80	152	80.5	190101 PASCO	CHAUPIMAR	4338	16.5	12.37	Normal	
16	C.S. ULUACHI	1009	1/14/2022	29	12	75.3	158	60.5	190109 PASCO	SIMON BOLI	4200	17.3	13.52	Normal	
17	C.S. ULUACHI	1009	1/16/2022	18	10	59.2	158	60.5	190101 PASCO	CHAUPIMAR	4338	17.4	13.27	Normal	
18	C.S. ULUACHI	1009	1/17/2022	29	36	68.5	154	66.5	190101 PASCO	CHAUPIMAR	4338	15.5	11.37	Normal	
19	C.S. ULUACHI	1009	1/18/2022	18	37	64	154	62.5	190109 PASCO	SIMON BOLI	4200	18.6	14.42	Normal	
20	C.S. ULUACHI	1009	1/21/2022	21	10	53.4	158	54.5	190101 PASCO	CHAUPIMAR	4338	15.5	11.37	Normal	
21	C.S. ULUACHI	1009	1/22/2022	22	33	66	159	58.5	190101 PASCO	CHAUPIMAR	4338	16.1	11.97	Normal	
22	C.S. ULUACHI	1009	1/27/2022	40	32	66.8	150	57.5	190101 PASCO	CHAUPIMAR	4338	15.1	10.97	Anemia Leve	
23	C.S. ULUACHI	1009	1/28/2022	23	38	72.5	151	63.5	190101 PASCO	CHAUPIMAR	4338	13.2	9.07	Anemia Moderada	

Parte de la preparación pasa por la distribución prevista de la etiqueta prediciendo, una forma de hacerlo es trazar un gráfico cuando la clase esté equilibrada, porque el desequilibrio no permite realizar la predicción deseada.

Por ello prepararemos los datos pasando al Software SPSS v26.

Figura 15.

Datos de madres Gestantes de la Provincia de Pasco en el SPSS

Edad	Edad_gestacional	Peso	Talla	PPG	Tipo_Embarazo	Ubigeo	Provincia	Distrito	Altitud	Hemoglobina	Hbc	Dx_Anemia
35	38	67,0	149	54,0	S	190101	PASCO	CHAUPIMARCA	4338	14,5	10,37	Anemia Leve
31	14	51,2	142	51,0	S	190101	PASCO	CHAUPIMARCA	4338	15,9	11,77	Normal
35	38	66,6	147	56,0	S	190101	PASCO	CHAUPIMARCA	4338	13,5	9,37	Anemia Moderada
37	34	50,4	145	46,0	S	190101	PASCO	CHAUPIMARCA	4338	15,8	11,67	Normal
31	35	69,2	154	66,0	S	190101	PASCO	CHAUPIMARCA	4338	11,5	7,37	Anemia Moderada
31	34	58,7	160	54,0	S	190101	PASCO	CHAUPIMARCA	4338	15,1	10,97	Anemia Leve
19	35	52,0	142	43,5	S	190101	PASCO	CHAUPIMARCA	4338	15,1	10,97	Anemia Leve
23	39	65,0	152	47,0	S	190101	PASCO	CHAUPIMARCA	4338	15,1	10,97	Anemia Leve
32	29	66,1	156	56,0	S	190101	PASCO	CHAUPIMARCA	4338	15,1	10,97	Anemia Leve
19	39	71,2	154	53,0	S	190101	PASCO	CHAUPIMARCA	4338	13,2	9,07	Anemia Moderada
26	30	69,3	148	64,0	S	190101	PASCO	CHAUPIMARCA	4338	15,5	11,37	Normal
21	38	59,2	162	59,0	S	190101	PASCO	CHAUPIMARCA	4338	14,8	10,67	Anemia Leve
23	32	53,4	140	42,0	S	190101	PASCO	CHAUPIMARCA	4338	14,8	10,67	Anemia Leve
30	7	80,0	152	80,0	S	190101	PASCO	CHAUPIMARCA	4338	16,5	12,37	Normal
29	12	75,3	158	60,0	S	190109	PASCO	SIMON BOLIVAR	4200	17,3	13,52	Normal
18	10	59,2	158	60,0	S	190101	PASCO	CHAUPIMARCA	4338	17,4	13,27	Normal
29	36	68,5	154	66,0	S	190101	PASCO	CHAUPIMARCA	4338	15,5	11,37	Normal
18	37	64,0	154	62,0	S	190109	PASCO	SIMON BOLIVAR	4200	18,6	14,82	Normal
21	10	53,4	158	54,0	S	190101	PASCO	CHAUPIMARCA	4338	15,5	11,37	Normal
22	33	66,0	159	58,0	S	190101	PASCO	CHAUPIMARCA	4338	16,1	11,97	Normal
40	32	66,8	150	57,0	S	190101	PASCO	CHAUPIMARCA	4338	15,1	10,97	Anemia Leve
23	38	72,5	151	63,0	S	190101	PASCO	CHAUPIMARCA	4338	13,2	9,07	Anemia Moderada

Figura 16.

Análisis descriptivo de madres Gestantes de la Provincia de Pasco según Provincia.

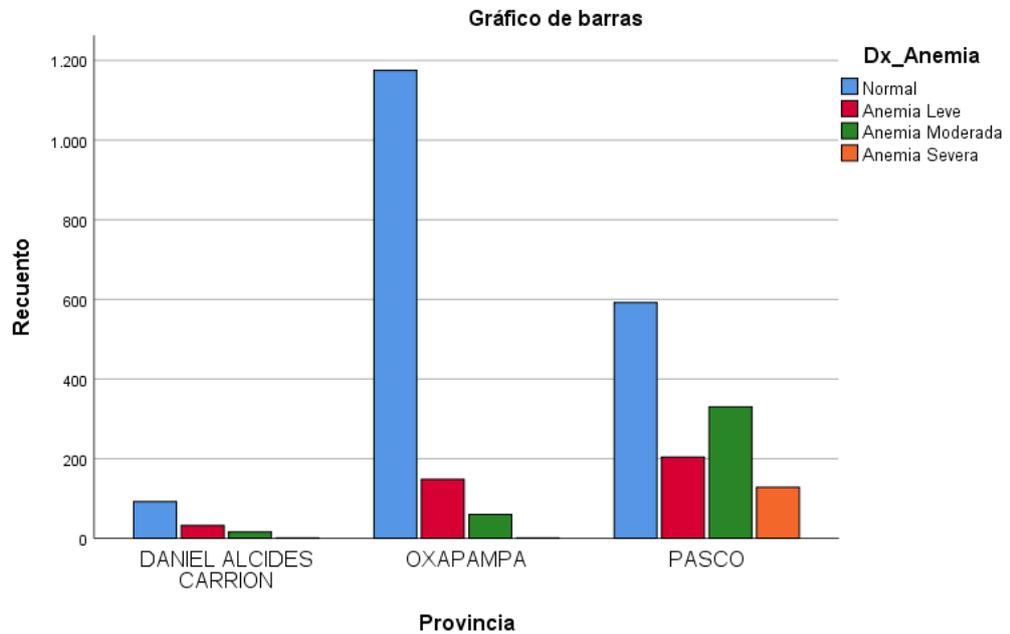
Tabla cruzada Provincia*Dx_Anemia

Recuento

Provincia	Dx_Anemia	Dx_Anemia				Total
		Normal	Anemia Leve	Anemia Moderada	Anemia Severa	
DANIEL ALCIDES CARRION		92	32	16	1	141
OXAPAMPA		1175	148	60	1	1384
PASCO		592	204	330	128	1254
Total		1859	384	406	130	2779

Figura 17.

Gráfico de barras de madres Gestantes de la Provincia de Pasco según Provincia.



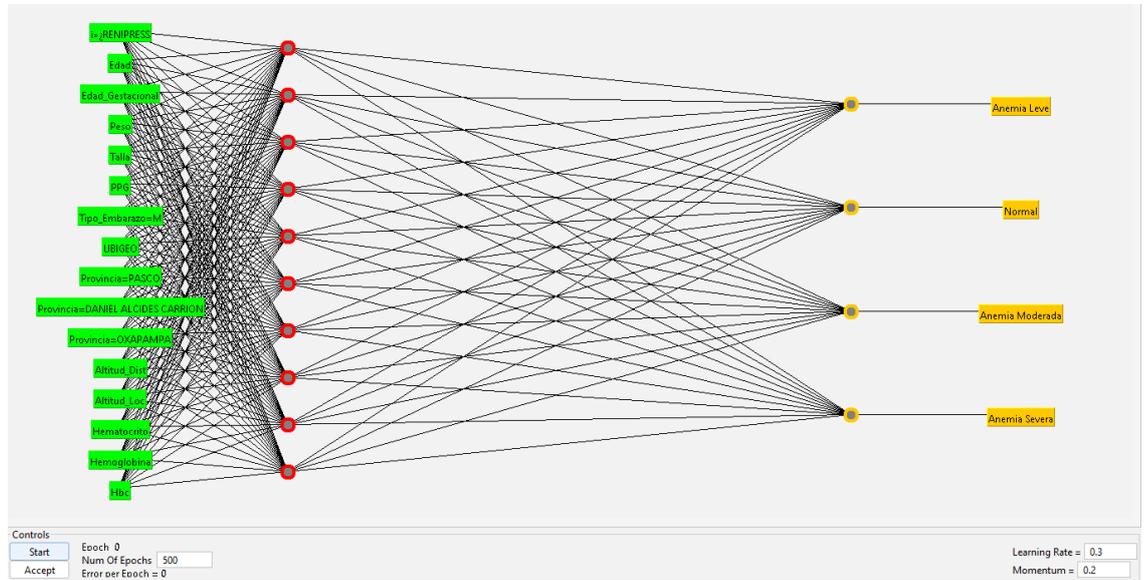
Interpretación: Las madres afectadas con anemia severa con un 98.4% en la Provincia de Pasco.

4.2.4. Modelado.

Construcción del Modelo Perceptrón Multicapa, se grafica con 3 capas.

- La primera capa corresponde a los valores de entrada que son 15 variables analizadas y procesadas en las fases de la metodología CRISP DM
- La segunda capa corresponde a la capa oculta que es parte del proceso en las redes neuronales para realizar una retroprogramación y el modelado sea optimo, esta capa consta de 4

Figura 18.
Red MLP.



La figura muestra el modelo secuencial de la capa clasificadora. Una función densa "describe las variables de la capa 1. En la primera capa, describe el valor unidades como neuronas en la primera capa. La capa de entrada consta de 15 neuronas que crean etiquetas, inicialización". Los pesos obtenidos con la derivada "kernel" están entre -0,5 y 0,5 porque tiene valores mayores y es óptima para lograr la convergencia.

4.2.5. Evaluación.

Matriz de clasificación de MLP

Primero se valida los modelos de precisión es una matriz de clasificación definida por las siguientes fórmulas:

La precisión.

$$Precisión_i = \frac{N_{ii}}{\sum_{k=1}^n N_{ki}}$$

Recall.

$$Recall_i = \frac{N_{ii}}{\sum_{k=1}^n N_{ik}}$$

F1 Score.

$$F - Score_i = \frac{2 \times Precision_i * Recall_i}{Precision_i + Recall_i}$$

El modelo obtuvo el siguiente reporte de métricas:

Figura 19.
Reporte de Métricas.

```
=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC   ROC Area  PRC Area  Class
0.831  0.023  0.853  0.831  0.842  0.817  0.979  0.905  Anemia Leve
0.983  0.054  0.973  0.983  0.978  0.934  0.995  0.997  Normal
0.919  0.007  0.956  0.919  0.937  0.927  0.997  0.978  Anemia Moderada
0.985  0.003  0.934  0.985  0.959  0.957  0.999  0.963  Anemia Severa
Weighted Avg.  0.953  0.041  0.952  0.953  0.953  0.918  0.993  0.980

=== Confusion Matrix ===
  a  b  c  d  <-- classified as
319  50  15  0  |  a = Anemia Leve
  31 1828  0  0  |  b = Normal
  24  0  373  9  |  c = Anemia Moderada
  0  0  2  128  |  d = Anemia Severa
```

Construcción del Modelo Naive Bayes

Hacer predicciones requiere calcular la probabilidad de que cada variable o categoría pertenezca a un evento de datos. Este proceso consta de los siguientes cuatro pasos: cálculo de la función de densidad de probabilidad gaussiana, precisión de la predicción y estimaciones:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Donde:

$P(A/B)$: la probabilidad de ocurrencia del evento A, dado el evento B, ya ha ocurrido.

$P(A)$ - Probabilidad de ocurrencia del evento A.

$P(B)$ - Probabilidad de ocurrencia del evento B.

$P(B/A)$ - Probabilidad de ocurrencia del evento B, dado que el evento A ya ha ocurrido.

Figura 20.

Reporte de Naive Bayes

```

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.466	0.033	0.696	0.466	0.559	0.516	0.951	0.673	Anemia Leve
	0.954	0.204	0.904	0.954	0.928	0.774	0.969	0.986	Normal
	0.648	0.038	0.745	0.648	0.693	0.647	0.915	0.719	Anemia Moderada
	0.977	0.031	0.611	0.977	0.751	0.759	0.989	0.929	Anemia Severa
Weighted Avg.	0.843	0.148	0.838	0.843	0.835	0.719	0.959	0.901	

```

=== Confusion Matrix ===

```

a	b	c	d	<-- classified as
179	138	67	0	a = Anemia Leve
66	1773	20	0	b = Normal
12	50	263	81	c = Anemia Moderada
0	0	3	127	d = Anemia Severa

Con 2779 registros equilibrados por SMOTE, la precisión de la predicción de Naive Bayes alcanzó el 83%, que es la predicción de este estudio.

Construcción del Modelo Árbol de decisión

Extracción de variables redundantes utilizando "Importancia de características, que es el grado de importancia de las variables para construir un modelo probabilístico, que es una clasificación artificial, donde el número de variables (n_feature) es el número de valores de entrada. árbol de retorno y relevancia gráfica (n_informativa) de características informativas"

Figura 21.
Reporte de Árbol de decisión

```

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      0.997    0.000    1.000     0.997    0.999     0.998    0.999     0.998    Anemia Leve
      1.000    0.001    0.999     1.000    1.000     0.999    0.999     0.999    Normal
      1.000    0.000    0.998     1.000    0.999     0.999    1.000     0.998    Anemia Moderada
      0.992    0.000    1.000     0.992    0.996     0.996    0.996     0.993    Anemia Severa
Weighted Avg.  0.999    0.001    0.999     0.999    0.999     0.999    0.999     0.999

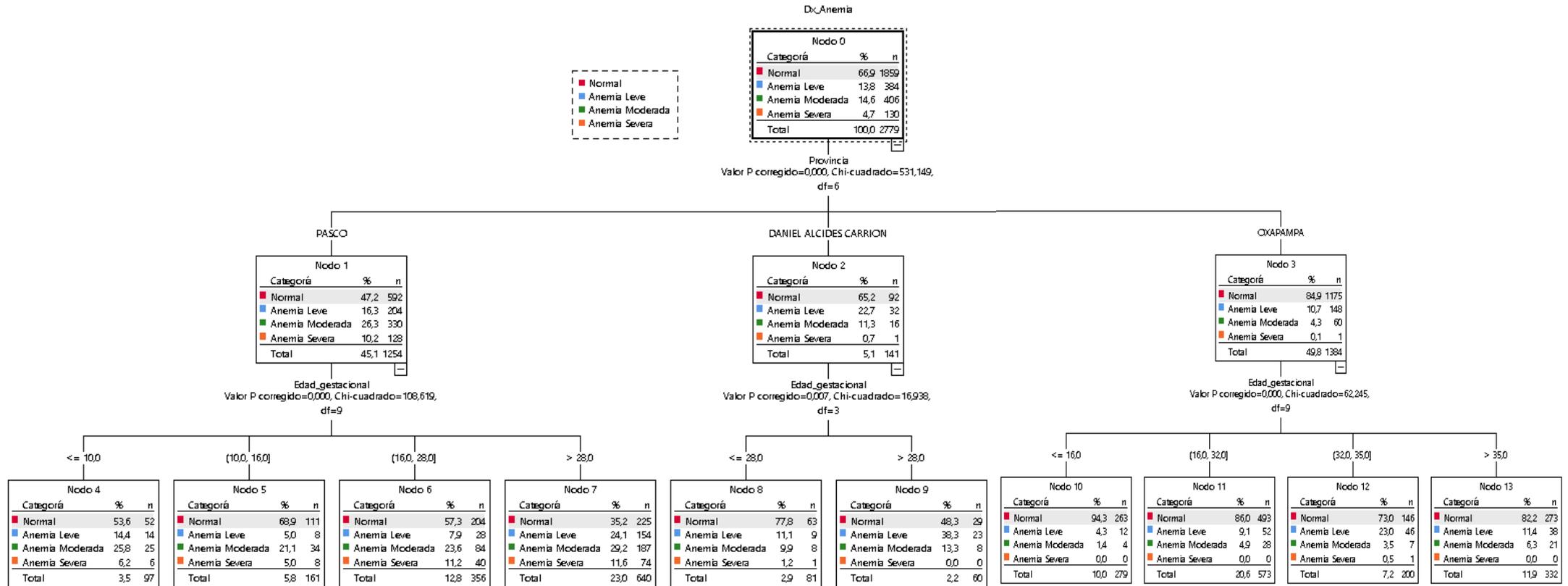
=== Confusion Matrix ===

  a  b  c  d  <-- classified as
383  1  0  0 |  a = Anemia Leve
  0 1859  0  0 |  b = Normal
  0  0 406  0 |  c = Anemia Moderada
  0  0  1 129 |  d = Anemia Severa

```

La importancia de la variable de datos experimentales en este estudio, que también se representa como la variable más influyente según estimaciones calculadas utilizando la tecnología de árbol de decisión de J48 para lograr una precisión del 98%. Los resultados son una representación general del impacto sin especificar las posibles categorías de problemas.

Figura 22.
Árbol de decisión



Evaluación de precisión de las técnicas

Este proyecto de minería de datos probó 3 algoritmos: red neuronal de “perceptrón multicapa, Naive Bayes y árbol de decisión”, y la tecnología de árbol de decisión logró la mayor precisión.

Tabla 3.

Resultado de las técnicas de estudio

Ítem	Técnica	Precisión
1	Perceptrón Multicapa	95%
2	Naive Bayes	83%
3	Árbol de decisión	99%

4.3. Prueba de Hipótesis

4.3.1. Hipótesis General

H_0 : La Implementación de un modelo de minería de datos **no predecirá** los casos de anemia en madres gestantes para de la Provincia de Pasco - 2023

H_a : La Implementación de un modelo de minería de datos **predecirá** los casos de anemia en madres gestantes para de la Provincia de Pasco - 2023.

Según el estudio que se realizó se acepta la Hipótesis alterna ya que se analizo con las tres técnicas y se puede observar que si la Implementación de un modelo de minería de datos predecirá los casos de anemia en madres gestantes para de la Provincia de Pasco – 2023.

4.3.2. Hipótesis Especificas.

H_1 : Aplicando la metodología CRISP-DM se pudo implementar un modelo de minería de datos para la predicción de casos de anemia en madres gestantes de la provincia de Pasco-2023

H0: Aplicando la metodología CRISP-DM **no se pudo implementar** un modelo de minería de datos para la predicción de casos de anemia en madres gestantes de la provincia de Pasco-2023

Ha: Aplicando la metodología CRISP-DM **se pudo implementar** un modelo de minería de datos para la predicción de casos de anemia en madres gestantes de la provincia de Pasco-2023

El primer nivel corresponde a los valores de entrada, es decir, las 15 variables analizadas y procesadas en los diferentes pasos del método CRISP DM.

La 2° capa corresponde a la capa oculta, que es la parte de la red neuronal que realiza el proceso de retro programación para el cual el modelo es óptimo.

La capa tiene 4 neuronas. Finalmente, la capa de predicción es de 1 neurona con valores 0,1,2,3. Utilizando el enfoque CRISP-DM se puede implementar un modelo de exploración de datos para predecir la incidencia de anemia en mujeres embarazadas en la provincia de Pasco-2023 y aceptar la hipótesis alternativa.

H2 Aplicando la validación cruzada se podrá validar el modelo de minería de datos para la predicción de casos de anemia en madres gestantes de la provincia de Pasco

Ho: Aplicando la validación cruzada no se podrá validar el modelo de minería de datos para la predicción de casos de anemia en madres gestantes de la provincia de Pasco

Ha: aplicando la validación cruzada se podrá validar el modelo de minería de datos para la predicción de casos de anemia en madres gestantes de la provincia de Pasco

Logrando la validación cruzada se podrá validar el modelo de minería de datos para la predicción de casos de anemia en madres gestantes de la provincia de Pasco - 2023

4.4. Discusión de resultados

El modelo sigue siendo una propuesta de la Red Provincial de Salud Pasco, porque se utilizan los datos de esta provincia, es decir, los datos del Departamento de Ginecología y Obstetricia, y su implementación y ejecución quedan a criterio de los respectivos organismos.

En este proyecto de minería de datos se probaron 3 algoritmos principales: redes neuronales de perceptrón multicapa, bahías ingenuas y árboles de decisión, incluida la técnica del árbol de decisión logró la mayor precisión: En el análisis de Perceptrón Multicapa con un 95%, Naive Bayes 83% y Árbol de decisión 99%.

Según (Condori Bellido, 2019) en los resultado obtenidos de su trabajo en los análisis del contexto empresarial, se propusieron las tres técnicas anteriores para predecir los casos futuros de madres embarazadas anémicas, para lo cual se aplicaron algoritmos de Perceptrón multicapa, Naive Bayer y Árbol de decisión J48 entrenados en datos históricos. Entre 422 registros de gestantes anémicas de la provincia de Ilo, el algoritmo con mayor precisión fue Naive Bayes, que alcanzó el 89%, seguido del árbol de decisión J48, que alcanzó el 79%, y finalmente, el perceptrón multicapa, que alcanzó el 62%.

CONCLUSIONES

Se utilizó un modelo exploratorio de datos para predecir la incidencia de anemia en mujeres embarazadas en el condado de Pasco, permitiendo la clasificación de las madres embarazadas anémicas y no anémicas e identificar características clave que requieren estrategias sólidas para la prevención de la anemia durante el embarazo.

Aplicando la metodología CRISP-DM para:

- Clasificar a las mujeres embarazadas con anemia o sin anemia utilizando tres algoritmos de predicción: perceptrón multicapa, árbol de decisión, Naive Bayes.
- Utilice la tecnología de árbol de decisión J48 para determinar con 99% de precisión el orden de las variables en mujeres embarazadas anémicas y no anémicas, donde las tres variables más influyentes son la edad gestacional, la anemia y la provincia.
- Realizando la validación cruzada del modelo, se obtuvo lo siguiente en puntuaciones: Perceptrón multicapa (95%), Árbol de decisión (99%), Naive Bayes (83%).

RECOMENDACIONES

- Centrarse en el modelo de minería de datos aplicándolo en instituciones del sector público para ayudar a prevenir y tratar oportunamente enfermedades.
- Usar los Árboles de decisión J48 predecir datos cuantitativos y categóricos para casos como este estudio.
- Utilice este trabajo como información para futuras investigaciones sobre este tema.

REFERENCIAS BIBLIOGRÁFICAS

- Alania, P. (2018). Aplicación de técnicas de minería de datos para predecir la deserción estudiantil de la facultad de ingeniería de la Universidad Nacional Daniel Alcides Carrión. *Universidad Nacional "Daniel Alcides Carrión."*
- Carpio Ticona, J. (2016). *Modelo de predicción de la morosidad en el otorgamiento de crédito financiero aplicando metodología CRISP-DM.* UNIVERSIDAD ANDINA "NÉSTOR CÁCERES VELÁSQUEZ."
- Condori Bellido, S. (2019). Modelo de Minería de datos para la predicción de casos de anemia en gestantes de la Provincia de Ilo. In *Progress in Retinal and Eye Research* (Vol. 561, Issue 3). UNIVERSIDAD NACIONAL DE MOQUEGUA.
- Hernández, J. (2015). *Modelo de Minería de datos para identificación de patrones en el aprovechamiento académico* (Issue 612). INSTITUTO TECNOLÓGICO DE LA PAZ BAJA CALIFORNIA MÉXICO.
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2018). *Metodología de la Investigación* (S. A. D. C. . McGRAW-HILL / INTERAMERICANA EDITORES (ed.); Sexta).
- Hernández Sampieri, R., & Mendoza Torres, C. (2018). *Metodología de la Investigación.*
- Mendoza Poma, K. (2023). *Método alternativo para la detección de la anemia a través de sus factores asociados en mujeres en edad reproductiva: una aplicación de redes neuronales artificiales.* Universidad Nacional Mayor de San Marcos.
- MINSA. (2019, October). *Minsa reúne a expertos para fortalecer las acciones contra la enfermedad de equinocosis | Gobierno del Perú.*
- Porto Solano, R. (2015). *Diseño de un modelo para asistencia médica y predicción de preeclampsia en mujeres embarazadas.* UNIVERSIDAD DEL NORTE - BARRANQUILLA.
- Solano Oviedo, D. (2013). *Aplicación de Minería de Datos para la Identificación de Patrones de comportamiento en las organizaciones enfocado en Prácticas de Impresión: Caso de Estudio.* PONTIFICIA UNIVERSIDAD JAVERIANA.

Vela Lopez, M. (2022). *Implementación de un modelo de minería de datos para predecir la deserción de los clientes en una empresa de telecomunicaciones.*
UNIVERSIDAD CATÓLICA SANTO TORIBIO DE MOGROVEJO.

ANEXOS

Matriz de Consistencia

Tema: “Modelo de Minería de Datos para la predicción de casos de anemia en madres gestantes de la Provincia de Pasco - 2023”

PROBLEMA GENERAL	OBJETIVO GENERAL	HIPÓTESIS GENERAL	VARIABLE INDEPENDIENTE	DIMENSIÓN	DISEÑO	POBLACIÓN Y MUESTRA
¿Se podrá implementar un modelo de minería de datos para la predicción de casos de anemia en madres gestantes para de la Provincia de Pasco - 2023?	Implementar un modelo de minería de datos para la predicción de casos de anemia en madres gestantes para de la Provincia de Pasco - 2023.	La Implementación de un modelo de minería de datos predecirá los casos de anemia en madres gestantes para de la Provincia de Pasco - 2023	Minería de Datos.	-Aplicando la metodología CRISP-DM. -Validación cruzada	Diseño: Pre - Experimental Tipo de Investigación Aplicada	POBLACIÓN La información abarca los casos de anemia en niños, adultos y gestantes, en el periodo 2022 de la provincia Pasco, sin embargo, se observó en el grupo de gestantes un crecimiento y se hizo énfasis en los casos de anemia de los mismos. MUESTRA Se toma como muestra los casos de anemia en gestantes de la provincia de Pasco, recolectando información de todos los centros de salud con casos de anemia en gestantes.
PROBLEMA ESPECÍFICO	OBJETIVO ESPECÍFICO	HIPÓTESIS ESPECÍFICA	VARIABLE DEPENDIENTE	DIMENSIÓN	MÉTODO DE INVESTIGACIÓN	TÉCNICAS - INSTRUMENTOS
¿Se podrá implementar un modelo de minería de datos aplicando la metodología CRISP-DM para la predicción de casos de anemia en madres gestantes de la Provincia de Pasco - 2023?	Proponer un modelo de minería de datos para la predicción de casos de anemia en madres gestantes de la provincia de Pasco aplicando CRISP-DM. Validar el modelo de minería de datos aplicando validación cruzada para la predicción de casos de anemia en madres gestantes	Aplicando la metodología CRISP-DM se pudo implementar un modelo de minería de datos para la predicción de casos de anemia en madres gestantes de la provincia de Pasco-2023. Aplicando la validación cruzada se podrá validar el modelo de minería de datos para la predicción de casos	Predicción de casos de anemia en gestantes.	- Reportes de madres gestantes	Método Analítica, inductiva Enfoque Cuantitativo	Técnicas: - Reportes.
¿Se podrá validar el modelo de minería de datos a través de la						

validación cruzada para la predicción de casos de anemia en madres gestantes de la Provincia de Pasco - 2023?

de la Provincia de Pasco - 2023.

de anemia en madres gestantes de la provincia de Pasco-2023.
