



UNIVERSIDAD CÉSAR VALLEJO

FACULTAD DE INGENIERIA

ESCUELA ACADEMICO PROFESIONAL DE INGENIERIA DE SISTEMAS

PROYECTO DE TESIS

“Modelo de minería de datos usando machine learning con reconocimiento de patrones de síntomas y enfermedades respiratorias en las historias clínicas para mejorar el diagnóstico de pacientes en la ciudad de Trujillo 2016.”

Autora:

Sandra Zoraida Medrano Parado

Asesor:

Ing. Ricardo Mendoza Rivera

Línea de Investigación:

Sistemas de Información Estratégicos y de Toma de Decisiones

Trujillo – Perú

2016

PAGINA DEL JURADO

Señores miembros del Jurado: En cumplimiento del Reglamento de Grados y Títulos de la Universidad César Vallejo presento ante ustedes la Tesis titulada "Modelo de minería de datos usando machine learning con reconocimiento de patrones de síntomas y enfermedades respiratorias en las historias clínicas para mejorar el diagnóstico de pacientes en la ciudad de Trujillo 2016", la misma que someto a vuestra consideración y espero que cumpla con los requisitos de aprobación para obtener el título Profesional de Ingeniería de Sistemas.

La Autora.

DEDICATORIA

Dedico esta tesis a mi familia que siempre me apoyaron y brindaron su consejo.

A mi asesor que contribuyo en mi investigación y me guio por la metodología correcta.

A mis demás docentes que me brindaron su apoyo y comprensión para lograr finalizar satisfactoriamente esta investigación.

DECLARACION DE AUTENTICIDAD

Yo Sandra Zoraida Medrano Parado con DNI N° 73273500, a efecto de cumplir con las disposiciones vigentes consideradas en el Reglamento de Grados y Títulos de la Universidad César Vallejo, Facultad de Ingeniería, Escuela de Ingeniería de Sistemas, declaro bajo juramento que toda la documentación que acompaño es veraz y auténtica. Así mismo, declaro también bajo juramento que todos los datos e información que se presenta en la presente tesis son auténticos y veraces. En tal sentido asumo la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de información aportada por lo cual me someto a lo dispuesto en las normas académicas de la Universidad César Vallejo.

Trujillo, 19 de Diciembre de 2016

Sandra Zoraida Medrano Parado

PRESENTACIÓN

Esta investigación aborda el tema de machine learning como apoyo a los diagnósticos médicos mediante la predicción de enfermedades a partir de síntomas y signos. El problema radica en el desacertado diagnóstico que el médico proporciona al paciente generando gastos en medicamentos y análisis innecesarios, además del desprestigio del doctor.

Por ello se vio por conveniente usar la minería de datos para mejorar los diagnósticos, utilizando la metodología CRISP con el algoritmo de árboles de decisión con el lenguaje de programación python.

Esta herramienta dio resultados positivos en las predicciones cumpliendo con el objetivo de mejorar los diagnósticos médicos, abriendo puertas para ampliar la investigación y no solo centrarse en una especialidad como fue en esta oportunidad las enfermedades respiratorias sino en las demás especialidades.

En conclusión, la minería de datos aportó en gran medida a la mejora de los diagnósticos médicos demostrando la relevancia que tiene y sus aportes que puede brindar a la sociedad.

INDICE DE CONTENIDO

PAGINA DEL JURADO	ii
DEDICATORIA	iii
DECLARACION DE AUTENTICIDAD.....	iv
PRESENTACIÓN	v
INDICE DE CONTENIDO.....	vi
RESUMEN.....	x
ABSTRACT	xi
I. INTRODUCCIÓN.....	12
1.1 Realidad Problemática.....	12
1.2 Trabajos previos	15
1.3 Teorías relacionadas al tema	16
1.3.1 Error Médico	16
1.3.2 Minería de datos:	17
1.3.2.1 Concepto.....	17
1.3.2.2 Patrón Estructural.....	17
1.3.2.3 Machine Learning.....	17
1.3.2.4 Atributos	17
1.3.2.5 Aplicaciones de Machine Learning	18
1.3.2.5.1 Web Mining	18
1.3.2.5.2 Imágenes.....	18
1.3.2.5.3 Marketing y Ventas	18
1.3.2.5.4 Otras Aplicaciones.....	18
1.3.2.6 Algoritmos orientados a predicciones	19
1.3.2.6.1 Algoritmos de clasificación	19
1.3.2.6.2 Agrupación de algoritmos	19
1.3.2.6.3 Algoritmo de regresión.....	19
1.3.2.7 Data Science	19
1.3.2.7.1 Etapas.....	20
1.3.2.8 Modelo CRISP	21

1.4	Formulación del problema.....	22
1.5	Justificación del estudio.....	23
1.6	Hipótesis	23
1.7	Objetivos	23
	Objetivo General:.....	23
	Objetivo Específico:.....	23
II.	METODO	24
2.1	Diseño de investigación.....	24
2.2	Variables y operacionalización de variables	24
2.3	Población y muestra.....	25
2.4	Técnicas e instrumentos de recolección de datos, validez y confiabilidad	26
2.5	Métodos de análisis de datos	27
2.6	Aspectos éticos.....	28
III.	Resultados.....	28
	MODELO CRISP	28
	PASO 1: COMPRENSIÓN DEL NEGOCIO	29
	PASO 2: COMPRENSIÓN DE LA DATA.....	29
	Descripción de los Datos.....	42
	Exploración de los Datos.....	42
	Verificación de la Calidad de los Datos.....	46
	PASO 3: PREPARACIÓN DE LOS DATOS.....	46
	Preparación de la colección de datos	46
	Selección de datos	49
	El servicio de limpieza y procesamiento de datos.....	52
	PASO 4: MODELADO	54
	Resultados:.....	61
	Prueba de Hipótesis.....	62
	▪ Prueba de Hipótesis para el indicador de Tiempo de realizar diagnósticos	62
	▪ Prueba de Hipótesis para el indicador de Diagnósticos Acertados.....	64
IV.	DISCUSIÓN.....	65
V.	CONCLUSIÓN	66

VI. RECOMENDACIÓN.....	66
VII. PROPUESTA.....	66
IV. REFERENCIAS.....	67
ANEXOS.....	69

INDICE DE ILUSTRACIONES

Ilustración 1. Descripción General del proceso de la ciencia de los datos	21
Ilustración 2. Fases del modelo de Referencia CRISP-DM.....	22
Ilustración 3: Matriz de Enfermedades y Síntomas.....	39
Ilustración 4:Modelo de Clases	40
Ilustración 5:Modelo Lógico	41
Ilustración 6: Inventario de datos de la tabla Enfermedad.....	47
Ilustración 7: Inventario de datos de la tabla Enfermedad Síntoma.....	47
Ilustración 8: Inventario de datos de la tabla Enfermedad Signo	48
Ilustración 9: Inventario de datos de la tabla Tipo síntoma	48
Ilustración 10: Inventario de datos de la tabla Antecedente Enfermedad.....	48
Ilustración 11: Variables de Estudio de la tabla Enfermedad.....	49
Ilustración 12: Variables de Estudio de la tabla Tipo Síntoma.....	49
Ilustración 13:Variables de Estudio de la tabla Síntoma.....	50
Ilustración 14: Variables de Estudio de la tabla Signo	50
Ilustración 15: Variables de Estudio de la tabla Enfermedad Antecedente	51
Ilustración 16: Variables de Estudio de la tabla Ubicación Geográfica.....	51
Ilustración 17: Clean de data.....	52
Ilustración 18:Preparar la data	54
Ilustración 19: Árbol Síntoma Enfermedad 1	57
Ilustración 20: Árbol Síntoma Enfermedad.....	57
Ilustración 21: Árbol Síntoma Enfermedad 3	57
Ilustración 22: Árbol Tipo Síntoma Enfermedad 3.....	58
Ilustración 23: Árbol Tipo Síntoma Enfermedad 2.....	58
Ilustración 24: Árbol Tipo Síntoma Enfermedad 1.....	58
Ilustración 25: Árbol Signo Enfermedad 3.....	59
Ilustración 26: Árbol Signo Enfermedad 2.....	59
Ilustración 27: Árbol Signo Enfermedad 1	59
Ilustración 28: Árbol Antecedente Enfermedad 1.....	60
Ilustración 29: Árbol Antecedente Enfermedad 2.....	60
Ilustración 30: Testeo con Python.....	61

INDICE DE TABLAS

Tabla 1: Descripción de la Base de Datos	42
Tabla 2: Descripción Tablas	42
Tabla 3: Cantidad de Registros por tablas	42
Tabla 4: Cantidad de Registros por Enfermedad	43
Tabla 5: Cantidad de Registros por Tipo Síntoma	44
Tabla 6: Cantidad de Registros por Signo.....	45
Tabla 7: Cantidad de Registros por antecedentes.....	45
Tabla 8: Objetivos definidos en la comprensión del negocio	46
Tabla 9: Registros de la tabla Ubicación Geográfica.....	51
Tabla 10: Tipo de Indicadores.....	62
Tabla 11: Tiempo de atención de paciente	63
Tabla 12: Tabulación de diagnósticos acertados	64
Tabla 13: Comparación del Indicador Ta y Td.....	65
Tabla 14: Promedio de tiempo después de aplicar la solución.....	79

RESUMEN

Esta investigación se desarrolló en la ciudad de Trujillo desde abril hasta diciembre del presente, centrándose en el problema de diagnósticos médicos y el uso de machine learning para contribuir en la mejora. La metodología que se usó para desarrollar la investigación fue CRISP. El tipo de investigación de acuerdo al fin es aplicada, de acuerdo a la técnica de contrastación es Pre Experimental, la población es la ciudad de Trujillo, los instrumentos usados en la investigación son el cuestionario y el análisis documental.

En conclusión, se logró la mejora de los diagnósticos médicos a partir de un modelo de machine learning usando el algoritmo de árbol de decisión.

Palabras Clave: machine learning, diagnóstico médico, árbol de decisión.

ABSTRACT

This research was developed in the city of Trujillo from April to December of the present, focusing on the problem of medical diagnoses and the use of the learning machine to contribute to the improvement. The methodology used to develop the research. The type of research agreed in order to apply, according to the technique of contrasts is Pre Experimental, the population is the city of Trujillo, the instruments used in the investigation of the questionnaire and documentary analysis.

In conclusion the improvement of the medical diagnoses was obtained from a model of machine learning using the algorithm of decision tree.

Key words: machine learning, medical diagnosis, decision tree.

I. INTRODUCCIÓN

1.1 Realidad Problemática

Los errores médicos son objetos de estudio a nivel internacional desde algunas décadas atrás, en el año 1999 el Instituto de Medicina de los Estados Unidos publicó un informe titulado “*To Err is human: Building a Safer Health System*” tocando temas que fueron escasamente discutidos en ese tiempo: la seguridad del paciente, donde detallaba que 98 000 personas murieron por errores médicos evitables estudio que se realizó en el año 1984 siendo la sexta causa de muerte en el mencionado país. A partir de este informe se promovieron políticas, congresos en diversas partes del mundo ya que fue un tema relevante y controversial. En el 2013 se realiza una nueva estimación de pacientes afectados según 4 estudios publicados anteriormente por el departamento de Salud y Servicios Humanos de la Oficina del Inspector General y otras entidades, usando la herramienta Global Tigger Tool para mejorar la estimación ya que en el primer estudio no se usó porque aún no existía tal herramienta concluyendo un aproximado de más de 400 mil personas fallecidas por errores médicos prevenibles aumentando la preocupación por este tema, ya que un paciente se atiende para que mejore no para su muerte siendo la tercer causa de muerte en Estados Unidos. (A New, Evidence-based Estimate of Patient Harms Associated with Hospital Care, 2013)

Este problema aqueja a nivel global y lo que se busca desde hace muchos años es tener un cultura que permita cambiar los estilos de trabajos, promover políticas y estrategias para reducir el impacto de los errores, por ello se crearon entidades en todo el mundo que velan por la seguridad del paciente como en Estados Unidos: National Patient Safety Foundation, en México Programa Nacional de la Cruzada por la Calidad de los Servicios de Salud y en otros países más, además de la Organización Mundial de la Salud formara la “*Alianza Mundial para la Seguridad del Paciente*”. A pesar de ello se ha ido mejorando lentamente y el progreso seguirá

así si no se toman las medidas para mejorar ello. (La cultura de seguridad del paciente: del pasado al futuro en cuatro tiempos, 2010)

El Dr. Elmer Huerta, Director del Preventorio del Instituto de Cáncer del MedStar Washington Hospital Center en Washington, DC en su columna Cuida tu Salud del El Comercio menciona que: *“Siempre ha ocurrido errores médicos provocando el miedo por caer en manos de una entidad médica. Además, público un estudio de la National Patienty Safety Foundation donde 42 % de los pacientes habían sido afectados ya sea la persona o algún familiar o amigo”*.

A nivel nacional, Perú participo de una investigación pionera denominada IBEAS (“Estudio Iberoamericano de Eventos Adversos”) donde además participaron Argentina, México, Costa Rica y Colombia en colaboración de la Organización Panamericana y Mundial de la Salud, siendo objeto de estudio 58 centros médicos concluyendo que cada 10 de 100 pacientes habían sufrido algún daño producido por cuidados sanitarios (Organizacion Mundial de la Salud, 2010) ya sea por diagnósticos, medicación, entre otros causas generando pérdidas económicas, pues se realizan gastos en consultas, medicinas, exámenes equivocados sin mencionar la desconfianza del paciente no solo por el médico sino de toda la institución.

También se ha investigado las denuncias judiciales por responsabilidad médica , esta investigación analizo de 1457 informes de los cuales 880 exámenes se omitieron ya que no se determinaba la especialidad y se referían a personal no médico (Estudio de las denuncias penales por responsabilidad profesional médica en el Instituto de Medicina Legal de Lima, Perú, 2013) que abarco del 2005 al 2010 mostrando que no se contaba con los recursos necesarios para realizar las investigaciones, y evidencias adecuadas, a pesar de ello se denota un crecimiento en las denuncias dando una alerta al sector salud para que genere estrategias que permita mitigar el problema.

Dando más detalle de este problema, las infracciones más frecuentes según un análisis del 1991 al 2000 es la relación médico paciente con un 35,7 % y de la siguiente década fue en los derechos del paciente con un 55 % (Denuncias éticas

de médicos colegiados en Perú, 1991-2010: frecuencia y naturaleza, 2014) dando como resultado de la mala relación que el paciente tiene con el doctor.

Otro problema en salud en el Perú según una revista de la Universidad Nacional Mayor de San Marcos es la metodología de enseñanza que se imparte a los futuros médicos, pues, los alumnos de medicina son orientados para entregar un buen diagnóstico y para ello deben aprender un gran número de enfermedades, cuanto más conocen, mejor aún; pero la memoria humana es perecedera, frágil dificultando que el diagnóstico sea más acertado además recurren a la intuición y subjetivismo contribuyendo a un mal diagnóstico. (El Registro Médico Orientado por Problemas, 1998 pág. 73)

Los errores médicos son un grave problema en la sociedad, una alternativa que contribuirían a la reducción de errores médicos es con la utilización de minería de datos para ayudar al personal médico, sugiriendo en base a síntomas enfermedades, y realizando otros análisis para la mejora del diagnóstico.

En la actualidad en toda empresa manejan un sinfín de datos. Al terminar el día tienen miles de facturas, prestamos, servicios que son registrados y almacenados. Todos esos datos pueden contribuir a brindar conocimiento a la empresa por ejemplo se puede conocer cuál es la enfermedad más frecuente por regiones del país o por edades. Para ello necesitan de una herramienta que le permita manejar todos los datos y convertir en conocimiento utilizando patrones, búsquedas heurísticas entre otros.

Con el procesamiento de los datos de las historias clínicas, se podrá obtener mediante un proceso de búsqueda patrones sobre síntomas y las enfermedades.

La situación del sector salud en el Perú con respecto al uso de minería de datos son escasas, pocos estudios abarcan sobre este tema. Si bien existen iniciativas tecnológicas, como reglamentar las historias clínicas electrónicas, o crear sistemas de información que ayuden a los procesos de las entidades de salud, aun no hay estudios de minería de datos (eSalud en Perú: implementación de políticas para el fortalecimiento de sistemas de información en salud, 2014 pág. 438).

Si bien los estudios en minería de datos en el sector salud es escaso, se demostrará la gran ayuda que puede brindar como poder crear a partir de los datos

almacenados históricamente modelos que muestren un patrón de comportamiento de los pacientes y sirva como base para reorientar los recursos.

1.2 Trabajos previos

El estudio titulado *“Técnicas de minería de datos aplicadas al diagnóstico de entidades clínicas”* que tiene como objetivo de analizar dos técnicas: clasificación y de agrupación dando como resultado dos modelos matemáticos enfocado en pacientes que sufren de hipertensión arterial contribuye a la creación de conocimiento a través del análisis de los diagnósticos médicos. (Técnicas de minería de datos aplicadas al diagnóstico de entidades clínicas, 2012 págs. 174-183)

Este antecedente también contribuye en brindar su algoritmo usado para el análisis de datos y poder así delimitar el algoritmo que se realizara y la metodología que se va a usar.

En el trabajo de investigación de (CANDELA Cáceres, 2015) titulado *“Proceso de Descubrimiento de Conocimiento para Predecir el Abandono de Tratamiento en una Entidad de Salud Pública”* realizado en la ciudad de Lima, cuyo objetivo fue el de automatizar un proceso de descubrimiento de conocimiento para una institución de salud pública que permita determinar el comportamiento de los pacientes con respecto a la continuidad en sus tratamientos. Realizo pruebas con cuatro algoritmos dando como resultado: *“Al algoritmo SVM un porcentaje de acierto de 96.4%, siendo el de mayor precisión, al algoritmo de modelos combinados un 95.9%, al algoritmo de árbol de decisión un 83.5%, y al algoritmo de redes neuronales un 53.9%.”* Concluyendo que, gracias al algoritmo SVM, se pudo determinar los factores más influyentes como son la edad, la autoestima, los medicamentos suministrados, entre otros y, gracias al algoritmo de árbol de decisión, las reglas asociadas a las categorías de tiempo de duración de la hospitalización. Esta investigación contribuye al uso de minería de datos, el beneficio de emplearlo y permite conocer el funcionamiento de los algoritmos para a partir de ello realizar el modelo propuesto.

En este estudio de (SALAS GUILLEN, y otros, 2014) titulado “*Uso De Inferencia Basada En Ontologías Para Dar Soporte Al Diagnostico Veterinario*” tenía como objetivo: “*Aplicar inferencia basada en ontologías para dar soporte al proceso de diagnóstico encapsulando el conocimiento del especialista en una base de información que de diagnóstico consultable*” realizado en la ciudad de Lima. Mediante el uso de la propuesta por la metodología CommonKADS, siendo un modelo de comportamiento que permitirá extraer información filtrada en base a los síntomas observados. Dando como resultado la construcción de una interfaz gráfica para mostrar cómo se realiza el diagnóstico haciendo uso de inferencias basadas en Ontologías. Obteniendo un 70% de casos diagnosticados exitosamente en base a la información anotada por el veterinario.

Este estudio contribuye una metodología y una ontología permitiendo a este proyecto analizar dicha ontología para poder elaborar el modelo propuesto.

1.3 Teorías relacionadas al tema

1.3.1 Error Médico

Un error médico se puede dar por exceso de servicio cuando el tratamiento no tiene probabilidades que el paciente mejore; por insuficiencia de servicios cuando la atención no es en el tiempo adecuado; y por diagnóstico inapropiado ya que el doctor intuye el posible tratamiento.

El diagnóstico es la actividad inicial del médico es el punto de partida siendo una de las actividades principales, se recogen los síntomas para elegir el tratamiento adecuado, además tiene como elementos de entrada: las historias clínicas, el examen físico, exámenes de laboratorio. En función a ello el doctor infiere el diagnóstico más adecuado.

Un error en el diagnóstico es directamente responsable el personal médico ya sea cognitivos debido a una mala recolección de datos o desinformación, también por erro del sistema ignorancia o mala práctica de la lex artis. (Errores médicos, 2009)

Existen tipos de errores en el diagnóstico que se mencionó anteriormente, terapéuticos, preventivos entre otros

Los errores terapéuticos por ejemplo mala indicación de dosificación del medicamento, también en realizar algún procedimiento o tratamiento de manera inadecuada.

Los errores preventivos se dan al realizar seguimiento del tratamiento inadecuado. (Errores médicos, 2009)

El acto médico es la conducta bajo la que se rige el medico esta actitud es seria, comprometida con la relación médico paciente y con su relación social, es decir la influencia del entorno que rodean al médico y al paciente. (2011).

1.3.2 Minería de datos:

1.3.2.1 Concepto

Minería de datos: nos permite descubrir patrones de datos y convertir datos en conocimiento que representan una ventaja competitiva mayormente económica. Teniendo un aprendizaje automático o semiautomático (H. Witten, y otros, 2011 pág. 4)

1.3.2.2 Patrón Estructural

Patrón Estructural: ayuda a identificar la estructura de los datos, no necesariamente son fórmulas, pero sirve para clasificar los datos. (H. Witten, y otros, 2011 págs. 5-6)

1.3.2.3 Machine Learning

Machine Learning: el aprendizaje es adquirir conocimiento a través de experiencias, enseñanzas e investigaciones, y machine learning busca adquirir este conocimiento a través de patrones de manera automática, explicando datos y poder hacer predicciones de ello. (H. Witten, y otros, 2011 pág. 7)

1.3.2.4 Atributos

Atributos: los datos en estudio son los atributos partiendo la investigación por ello y su clasificación que pueden ser nominal y numéricos (H. Witten, y otros, 2011 págs. 3-8)

1.3.2.5 Aplicaciones de Machine Learning

1.3.2.5.1 Web Mining

Los datos que existen en internet a crecido exponencialmente y por ello las empresas no dudan en usar este medio para obtener información de sus clientes, Google para medir la posición de la página web lo hace según el prestigio, mientras más páginas se enlazan con su sitio web tiene mayor prestigio. También analizan las consultas a los motores de búsqueda. (H. Witten, y otros, 2011 págs. 21-22) .

1.3.2.5.2 Imágenes

El procesamiento de imágenes también es una herramienta de gran impacto usada por ejemplo para la detección de derrames de petróleo por satélite, reconocimiento facial gracias a la minería de datos, los pixeles de entrada se convierten en información. (H. Witten, y otros, 2011 pág. 23)

1.3.2.5.3 Marketing y Ventas

Los clientes son una pieza importante para toda empresa por ello que la confiabilidad que una empresa puede representar es significativa y se usan herramientas para mantener el cliente determinando qué nuevos servicios son adecuados, entregando incentivos en quienes se obtenga mayores beneficios.

En marketing realizando un análisis demográfico se puede determinar que producto de puede vender más, también otro ejemplo es el envío de ofertas personalizadas por el tipo de cliente. (H. Witten, y otros, 2011 págs. 26-27)

1.3.2.5.4 Otras Aplicaciones

El uso de machine Learning es muy amplia algunos ejemplos en biología se usa para ayudar a identificar genes dentro de cada genoma, en la televisión para conocer que programas son más

vistos, en lo informático para detectar la intrusión mediante un análisis de operaciones inusuales. (H. Witten, y otros, 2011 págs. 27-28)

1.3.2.6 Algoritmos orientados a predicciones

1.3.2.6.1 Algoritmos de clasificación

Usualmente utilizadas para las predicciones hay muchas herramientas para la clasificación: regresión logística, redes neuronales, arboles de decisión, para más complejos se usan técnicas multimodales.

1.3.2.6.2 Agrupación de algoritmos

Utilizan un aprendizaje no supervisado, cuando el resultado no se conoce. La herramienta más conocida es el algoritmo k-medias que agrupa datos similares entre si utilizando una métrica de distancia.

1.3.2.6.3 Algoritmo de regresión

Son usadas para predecir variables con resultado numéricas, algunas técnicas usadas son la regresión lineal, árboles de decisión, redes neuronales que imita el cerebro humano. (WEE Hyong, y otros, 2015 pág. Capitulo 1)

1.3.2.7 Data Science

Data Science: la ciencia del dato tiene como propósito brindar técnicas y un proceso de recolección de datos que recae en un esfuerzo de equipo. El análisis de datos comprende, análisis descriptivo usando gráficos, histogramas; análisis de diagnósticos utilizado en la inteligencia de Negocio para explicar las causas de los hechos utilizando técnicas de agrupamiento, clasificación y árboles de decisión; análisis predictivo usado para toma de decisiones de acuerdo a las predicciones; análisis prescriptivo combina el análisis predictivo y sirve para la optimización utilizando las técnicas

programación lineal y no lineal además de árboles de decisión. (WEE Hyong, y otros, 2015)

1.3.2.7.1 Etapas

1. Definir el problema de negocio:

Este paso guía el proyecto, si no está definido correctamente se terminará realizando una solución a un problema equivocado. Esta etapa es muy importante definirlo minuciosamente, precisa y refleje lo que se quiere solucionar desde el punto de vista del negocio fijando objetivos para ello.

2. Adquirir y preparar los datos:

Es un proceso es muy laborioso pues uno desea analizar algunas variables y no existen datos suficientes y es la fuente principal para investigar, se realiza los ETL extraer, transformar y cargar los datos en bruto desde fuentes diversas como bases de datos, archivos de registros web entre otros.

3. Desarrollar el modelo:

Se determina el algoritmo según el modelo planteado a ejecutar por ejemplo un árbol de decisión para el modelo predicciones entre otros según los objetivos del negocio.

4. Implementar el modelo:

Aplicarlo en la realidad, en el negocio ya sea mediante una aplicación web o móvil.

5. Modelo de monitor rendimiento:

Seguimiento al modelo implementado, el modelo planteado va perdiendo la precisión al pasar el tiempo por diversas razones, se recomienda reconstruir cada año de acuerdo al entorno si es muy dinámico cada semana o diario.

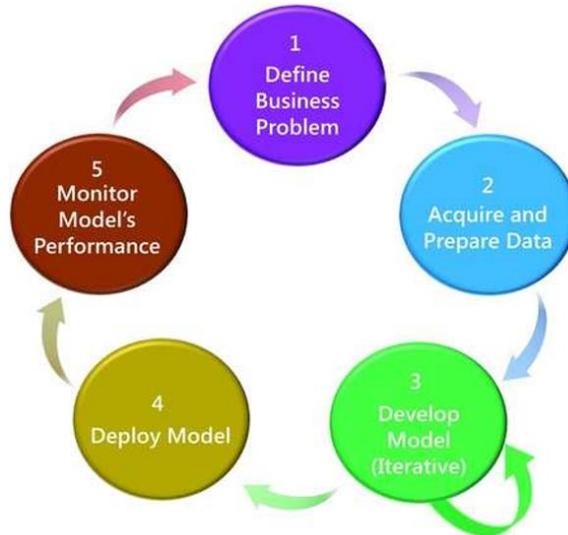


Ilustración 1. Descripción General del proceso de la ciencia de los datos

1.3.2.8 Modelo CRISP

1. Comprensión del negocio

Es la primera fase que se centra en comprender al negocio fijando objetivos de proyecto además se debe definir el problema siempre en función al negocio en un plan preliminar para alcanzar los objetivos trazados.

2. Comprensión de los datos

Esta segunda fase se centra en los datos identificando problemas de calidad de datos para formar una hipótesis adecuada de acuerdo a la información analizada.

3. Preparación de datos

Esta fase tiene una duración mayor y de gran importancia, se prepara los datos a partir de registros, atributos entre otros se transforma y luego se limpian para ser usadas por el modelo.

4. Modelado

En esta fase se selecciona las técnicas de modelado de acuerdo a los requerimientos sobre la forma de los datos. Se puede

desarrollar de uno a más técnicas. También se genera la prueba del diseño para probar la calidad y validez del modelo escogido.

5. Evaluación

Después de construir el modelo es importante evaluarlo con los objetivos del proyecto planteados en función del negocio, ya que deben estar alineados, además de comparar con los objetivos se puede encontrar información de suma importancia para el negocio para futuras predicciones.

6. Desarrollo

El construir el modelo no es el fin del proyecto, aun se debe presentar al cliente del modo que pueda usarlo ya sea con una aplicación web, móvil entre otros , ordenando los datos , mostrando de manera entendible para el usuario. (CHAPMAN, y otros, 2000)

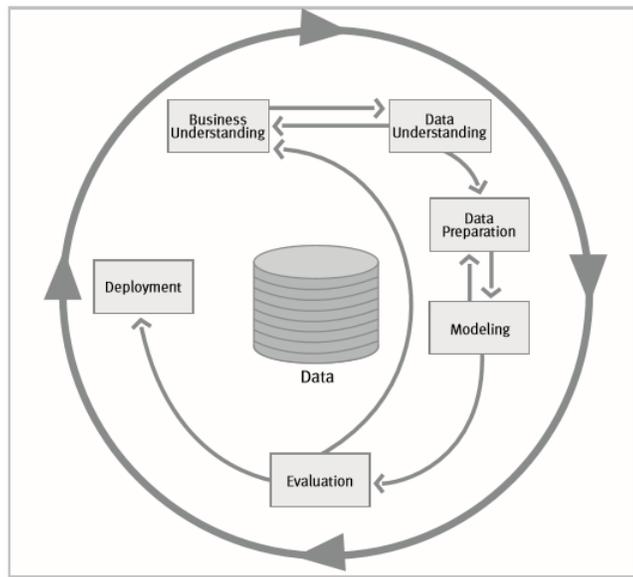


Ilustración 2. Fases del modelo de Referencia CRISP-DM

1.4 Formulación del problema

¿De qué manera el desarrollo de un algoritmo de minería de datos con reconocimiento de patrones de síntomas y enfermedades respiratorias influirá en el diagnóstico de pacientes?

1.5 Justificación del estudio

La investigación propuesta busca mediante la aplicación de la teoría y los conceptos básicos sobre minería de datos y los algoritmos de minería de datos ayudar al sector salud con el análisis de las historias clínicas.

En cuanto a plataforma tecnológica el desarrollo del algoritmo será compatible con todos los sistemas y para construirlo se usarán herramientas de software privativo y libre.

Esta investigación se enfoca en todas las entidades del sector salud ya sea privada o pública.

1.6 Hipótesis

H₁: Un modelo de minería de datos para el reconocimiento de patrones de síntomas y enfermedades mejora significativamente el diagnóstico de pacientes en la ciudad de Trujillo.

H₀: Un modelo de minería de datos para el reconocimiento de patrones de síntomas y enfermedades no mejora significativamente el diagnóstico de pacientes en la ciudad de Trujillo

1.7 Objetivos

Objetivo General:

- Mejorar el diagnóstico de pacientes en la ciudad de Trujillo con el desarrollo de un modelo de minería de datos con reconocimiento de patrones de síntomas y enfermedades respiratorias.

Objetivo Específico:

- Identificar los escenarios y características de las enfermedades y síntomas
- Reducir el tiempo de diagnosticar enfermedades en un paciente.
- Acertar los diagnósticos con el modelo de minería de datos.

II. METODO

2.1 Diseño de investigación

Pre Experimental

Diseño pre-test post –test con su solo grupo, cuyo esquema es el siguiente:

G: O₁ X O₂

Donde:

O₁: pre-test

x: Aplicación del modelo de minería de Datos

O₂: post – test

2.2 Variables y operacionalización de variables

Variable	Conceptualización	Operacionalización	Dimensiones	Indicadores	Escala
Modelo de minería de datos	<p>Minería de datos: nos permite descubrir patrones de datos y convertir datos en conocimiento que representan una ventaja competitiva mayormente económica. Teniendo un aprendizaje automático o semiautomático (H. Witten, y otros, 2011 pág. 4)</p>	<p>instrumento Valorada: Tiempo de realizar el diagnóstico con el modelo es menor al tiempo de aplicar sin el modelo</p>	<p>Tiempo de realizar diagnóstico</p>	$\frac{\sum ta}{n} - \frac{\sum td}{n2}$ <p>Donde: ta: tiempo de atención antes n : número pacientes atendidos anteriormente td: tiempo de atención después n2: número de pacientes atendidos después.</p>	<p>Cuantitativa de razón.</p>

Diagnóstico de Pacientes	Actividad inicial del médico es el punto de partida de la relación paciente médico, tiene como elementos de entrada: las historias clínicas, el examen físico, exámenes de laboratorio. (Errores médicos, 2009)	Valorada: 70 - 100 % aceptable, 50 – 69% parcialmente aceptable. 1 – 49 % inaceptable	Diagnósticos acertados	Diagnósticos acertados / diagnóstico total	Cuantitativa Porcentual.
---------------------------------	--	--	------------------------	--	--------------------------

2.3 Población y muestra

De acuerdo la Encuesta Nacional de Satisfacción de Usuarios en Salud (ENSUSALUD) efectuado en los establecimientos del Ministerio de Salud (MINSA) y ESSALUD se realizó la muestra siguiente:

Población: 1,859,640 Habitantes en la ciudad de Trujillo

Se realiza el cálculo de la muestra con la siguiente fórmula

$$n_0 = \frac{Nz^2 * p * q}{(N - 1) * E^2 + z^2 * p * q} * def * tnr$$

Donde:

n_0 : muestra

Z: nivel de confianza – 1.96 (95%)

p: proporción de éxito – 38%

q: proporción fracaso – 62%

E: error de Estimación – margen aceptado en la estimación de p

N: Número de habitantes

def: Efecto de diseño al usar muestreo con más de una etapa de selección
– 1.2 en base a resultados de la ENDES 2000.

tnr: Coeficiente que ajuste al tamaño muestral debido a la no respuesta.
Reemplazando los datos da como resultado 790

Muestra: Usuarios de Consulta Externa en la región de la Libertad 790.

Unidad de Análisis: Usuarios en Consulta Externa.

2.4 Técnicas e instrumentos de recolección de datos, validez y confiabilidad

TECNICAS	INSTRUMENTOS
<p>A. Análisis documental: <i>“Comprende el procesamiento analítico- sintético que, a su vez, incluye la descripción bibliográfica y general de la fuente, la clasificación, indización, anotación, extracción, traducción y la confección de reseñas”.</i> (Análisis documental y de información: dos componentes de un mismo proceso., 2004)</p>	<ul style="list-style-type: none"> - Fichas Resumen - Fichas Bibliográficas - Fichas textuales
<p>B. Encuesta: se recoge información usando procedimientos estandarizados de manera que a cada individuo se le hacen las mismas preguntas en más o menos la misma manera. La intención de la encuesta no es describir los individuos particulares quienes, por azar, son parte de la muestra, sino obtener un perfil compuesto de la población. (BEHAR Rivero, 2008)</p>	<ul style="list-style-type: none"> - Cuestionario

Validación y Confiabilidad del Instrumento:

La elaboración de la encuesta ha sido realizada por la entidad privada ENSUSALUD y no es necesaria realizar la validación y confiabilidad.

Anexo 1 Formato de Encuesta del ENSUSALUD

2.5 Métodos de análisis de datos

Análisis ligados a las hipótesis

- **Paramétrica**

- **Prueba Z**

Es una prueba estadística de hipótesis que con más frecuencia. La gráfica de su función es curva normal de frecuencias tiene una forma de campana y en el centro se sitúan tres medidas de tendencia central. Esta distribución es de gran relevancia ya que permite modelar numerosos fenómenos naturales, sociales y psicológicos.

$$Z = \frac{X - \bar{X}}{\sigma}$$

Donde:

Z = valor estadístico de la curva normal de frecuencias

X = cualquier valor de una muestra estadística

\bar{X} = promedio o media aritmética obtenido de la muestra estadística, valor representativo.

σ = desviación estándar

Hipótesis Nula

$$H_0 : \mu_B - \mu_A = 0$$

El indicador del modelo actual es mejor que el indicador del modelo propuesto

Hipótesis Alternativa

$$H_1 : \mu_B - \mu_A > 0$$

El indicador del modelo propuesto es mejor que el indicador del modelo actual

2.6 Aspectos éticos

Respeto a la privacidad, proteger la identidad de los individuos que participan en el estudio.

III. Resultados

MODELO CRISP

DESCRIPCIÓN

VISION GENERAL

Según lo investigado existen denuncias judiciales por responsabilidad médica, desde el 2005 al 2010 mostrando que no se contaba con los recursos necesarios para realizar las investigaciones, y evidencias adecuadas, a pesar de ello se denota un crecimiento en las denuncias dando una alerta al sector salud para que genere estrategias que permita mitigar el problema.

Otro problema en salud en el Perú según una revista de la Universidad Nacional Mayor de San Marcos es la metodología de enseñanza que se imparte a los futuros médicos, pues, los alumnos de medicina son orientados para entregar un buen diagnóstico y para ello deben aprender un gran número de enfermedades, cuanto más conocen, mejor aún; pero la memoria humana es perecedera, frágil dificultando que el diagnóstico sea más acertado además recurren a la intuición y subjetivismo contribuyendo a un mal diagnóstico.

VISION ESTRATEGICA

Se desea implantar mejoras en el sector salud para brindar una mejor atención a los pacientes, para contribuir al desarrollo sostenible del país.

PASO 1: COMPRENSIÓN DEL NEGOCIO

DETERMINACIÓN DE OBJETIVOS DE NEGOCIO

Dado los problemas mencionados anteriormente se justifican con una encuesta desarrollada el 2015 por ENSUSALUD desean mejorar la satisfacción de sus usuarios no solo con el tiempo de atención sino en diagnósticos más certeros.

Objetivo del negocio:

- Mejorar el diagnóstico de pacientes en la ciudad de Trujillo con el desarrollo de un modelo de minería de datos con reconocimiento de patrones de síntomas y enfermedades respiratorias.

Objetivo de la minería de Datos:

- Crear un modelo predictivo de clasificación de las enfermedades a partir de los síntomas y signos que corresponden a cada enfermedad.

PASO 2: COMPRENSIÓN DE LA DATA

Recolección de Datos

Identificación de signos y síntomas:

- Fiebre
 - Fiebre continua
 - Fiebre febrícula
 - Fiebre febril
 - Fiebre remitente
 - Fiebre intermitente
 - Fiebre recurrente
 - Fiebre Hética
 - Fiebre ondulante
 - Fiebre invertida
 - Fiebre bicotidiana
 - Fiebre nosocomial
 - Fiebre de origen desconocido
- Astenia
- Disnea
 - Ortopnea
 - Platipnea
 - Disnea espontanea

- Disnea de rápida progresión
 - Disnea de lenta progresión
- Ronquidos
 - Ronquido de cubito dorsal
- Afonía
 - Afonía total
 - Afonía parcial
- Estridor
 - Estridor inspiratorio Nasofaringe
 - Estridor inspiratorio Orofaringe
 - Estridor inspiratorio Laringe
 - Estridor inspiratorio compresión externa
 - Estridor inspiratorio en vías respiratorias bajas
 - Estridor Mixto
- Tos
 - Tos seca
 - Tos productiva
- Expectoración
 - Esputo mucoso
 - Esputo espumoso y rosado
 - Esputo negro, con partículas de hollín
 - Esputo purulento
 - Esputo herrumbroso
 - Esputo similar a la jalea
 - Esputo con rasgos de sangre
- Hemoptisis
 - Hemoptisis idiopática
- Dolor torácico
- Hiperventilación
- Hipoventilación
- Palpitación pulmonar
 - Vibraciones vocales aumentadas
 - Vibraciones vocales disminuidas
- Percusión pulmonar
 - Sonoridad disminuida
 - Sonoridad aumentada
- Auscultación pulmonar
 - Murmullo vesicular
 - Soplos pulmonares
 - Estertores secos

- Estertores húmedos
 - Crepitantes
- Alteraciones ventilatorias
 - Alteraciones ventilatorias obstructivas
 - Alteraciones ventiladores restrictivas
- Insuficiencia respiratoria
- Enfisema pulmonar
 - Tipo PP
 - Tipo BB
- Neumotórax
 - Neumotórax espontaneo
 - Neumotórax traumático
- Fibrosis pulmonar
- Edema pulmonar
- Derrame pleural
 - Exudado pleural
 - Trasudado pleural
 - Quilotórax
 - Exudado plural hemorrágico
- Absceso pulmonar
 - Absceso pulmonar aspiración
 - Abscesos específicos
- Tumores mediastínicos
 - Mediastino central
 - Mediastino superior anterior
 - Mediastino inferior anterior
 - Mediastino posterior
- Metástasis pulmonares
- Insomnio
 - Primario
 - Secundario
- Somnolencia
 - Primaria
 - Periódica
 - Crónica
- Prurito generalizado
- Rubicundez
 - Rubicundez generalizada
 - Rubicundez facial
- Cefalea

- Migraña
- Cefalea tensional
- Cefalea en racimos
- Miscelánea de cefalea
- Cefalea aguda generalizada
- Cefalea aguda localizada
- Cefalea aguda recurrente
- Cefalea crónica progresiva
- Cefalea crónica no progresiva
- Algas faciales
- Epistaxis
 - Epistaxis anterior
 - Epistaxis posterior
- Caída de dientes
- Proptosis
 - Proptosis unilateral agudo
 - Proptosis unilateral crónico
- Triada de horner
- Cifosis
- Dolor lumbar
- Dolor óseo
- Algas vertebrales
- Artritis
 - Monoartritis
 - Poliartritis
- Debilidad muscular aguda
- Dolor de hombro
- Dolor cervicobraquial
- Palpitaciones
- Hipertensión arterial
 - Hipertensión arterial sistólica
 - Hipertensión arterial sistólica y diastólica
- Hipotensión arterial
 - Hipotensión crónica
 - Hipotensión aguda
- Ingurgitación yugular
- Ruidos cardiacos
 - Primer ruido
 - Segundo ruido

- Tercer ruido
 - Cuarto ruido
- Soplos cardiacos
 - Soplos diastólicos
 - Soplos sistólicos
 - Soplos continuos
 - Soplos sistólicos – diastólicos
 - Soplo carotideo
- Taquicardia
 - Taquicardia sinusal
 - Taquicardia regular
- Pericarditis
- Cor pulmonale
- Cianosis
 - Cianosis central
 - Cianosis periférica
- Insuficiencia cardiaca
 - Insuficiencia cardiaca izquierda
 - Insuficiencia cardiaca derecha
- Shock
 - Shock hipovolémico
 - Shock distributivo
 - Shock obstructivo
 - Shock cardiogénico
- Miocarditis
- Infarto de miocardio
- Tromboembolismo pulmonar
- Halitosis
- Disfagia
- Dolor al masticar
- Vómito
- Meteorismo
- Mala absorción
 - Primaria
 - Secundaria
- Constipación
 - Estreñimiento agudo
 - Estreñimiento progresivo y reciente
 - Estreñimiento crónico

- Hepatomegalia
 - Hígado palpable sin hepatomegalia
 - Hepatomegalia verdadera
- Esplenomegalia
- Ascitis
 - Ascitis libre
 - Ascitis tabicada
- Líquido ascítico
 - Trasudado
 - Exudado
 - Quiloso
- Ictericia
 - Ictericia prehepática
 - Ictericia intrahepática
 - Ictericia posthepática
 - Formas mixtas de ictericia
- Dolor abdominal
- Íleo paralítico
- Hematuria
 - hematuria intensa
 - hematuria moderada
 - hematuria microscópica
- uremia
- acidosis respiratoria
- alcalosis respiratoria
 - alcalosis respiratoria aguda
- síndrome de secreción inapropiada de HAD (SIADH)
- insuficiencia suprarrenal
 - Is primaria
 - Is secundaria
- Anemia hemolítica
 - AH congénitas
 - AH adquiridas
- Dolor de garganta
- Congestión nasal
- Dolor muscular
- Estornudos

Características de los signos y síntomas

- Temperatura
- Ambiente
- Duración
- Importancia
- Frecuencia
- Peso
- Presión arterial
- Herencia
- antecedentes
- edad

Enfermedad:

- Infecciones respiratorias superiores
 - Rinofaringitis aguda
 - Resfrio por bacteria
 - Rinitis alérgica
 - Rinitis estacional
 - Rinitis crónica
 - Sinusitis aguda
 - Sinusitis Subaguda
 - Sinusitis recurrente
 - Sinusitis crónica
 - Faringoamigdalitis aguda
 - Faringoamigdalitis crónica
 - Laringitis subglótica – CRUP
 - Laringotraqueitis
 - Laringotraqueobronquitis
 - Epiglotitis
 - Tosferina
 - Difteria
 - Nasofaringitis
 - Adenoiditis

- Infecciones respiratorias inferiores
 - Asma
 - Bronquiolitis
 - Bronquitis aguda
 - Neumonía adquirida por la comunidad
 - Neumonía nosocomial
 - Neumonía por bacteria
 - Neumonía atípica Mycoplasma
 - Tuberculosis pulmonar
 - Tuberculosis ganglionar
 - Tuberculosis meníngea
- Otras enfermedades
 - SAOS
 - Influenza

Matriz de Enfermedades y síntomas

Enfermedades		rino-faringitis aguda	resfriado por bacteria	rinitis alérgica (tipo de rinitis crónica)	rinitis estacional	rinitis crónica	sinusitis aguda
Síntomas							
tos	seca	si	si	no	no	no	si
	productiva	no	no	no	no	no	no
	aguda	si	si	no	no	no	si
	subaguda	no	no	no	no	no	no
	crónica	no	no	no	no	no	no
	espasmódica	no	no	no	no	no	no
	perruna						
dolor	garganta u odinofagia	si	si	si	no	si	si
	torácico	si	si	no	no	no	no
	lumbar	no	no	no	no	no	no
	ósea	no	no	no	no	no	no
	hombro	no	no	no	no	no	no
	cervicobranquial	no	no	no	no	no	no
	al masticar	no	no	no	no	no	no
	abdominal	si	si	no	no	no	no
	oído u otalgia	no	si	no	no	no	no
	muscular	si	si	no	no	no	no
asimetría torácica	no	no	no	no	no	no	
náuseas	si	si	no	no	no	no	
mareo	no	no	no	no	no	no	
vómitos	si	si	no	no	no	no	
diarrea	si	si	no	no	no	no	
fiebre	febril	si	si	no	no	no	si
	febrícula	no	no	no	no	no	no
	continua	no	si	no	no	no	no
	remitente	no	no	no	no	no	no
	intermitente	no	si	no	no	no	no
	recurrente	no	no	no	no	no	no
	hética	no	no	no	no	no	no

Enfermedades		rinoфарингитис aguda	resfrio por bacteria	rinitis alergica (tipo de rinitis cronica)	rinitis estacional	rinitis cronica	sinusiti aguda
Síntomas							
	ondulante	no	no	no	no	no	no
	invertida	no	no	no	no	no	no
	bicotidiana	no	no	no	no	no	no
	nosocomial	no	no	no	no	no	no
	FOD	no	no	no	no	no	no
estornudo		si	si	si	no	si	no
picazon o prurito	nariz	no	no	si	no	si	no
	boca	no	no	si	no	si	no
	ojos	no	no	si	no	si	no
	garganta	no	no	si	no	si	no
	piel	no	no	si	no	si	no
adinamia		no	no	no	no	no	si
cansancio		no	no	no	no	no	si
cefalea	migraña	no	no	no	no	no	no
	cefalea tensional	p	p	no	no	no	no
	cefalea en racimos	p	p	no	no	no	no
	cefalea aguda localizada	p	p	no	no	no	si
	cefalea aguda generalizada	p	p	no	no	no	no
	cefalea aguda recurrente	p	p	no	no	no	no
	cefalea cronica progresiva	p	p	no	no	no	no
	cefalea cronica no progresiva	p	p	no	no	no	no
ronquidos	cubito dorsal	no	no	no	no	no	no
	normal	no	no	no	no	no	no
apnea		no	no	si	no	no	no
pesadillas		no	no	no	no	no	no
insomnio	primario	no	no	si	no		no
	secundario	no	no	si	no	no	no
	agitado	no	no	si	no	no	no
sueño		no	no	no	no	no	no

Enfermedades		rino faringitis aguda	resfio por bacteria	rinitis alergica (tipo de rinitis cronica)	rinitis estacional	rinitis cronica	sinusiti aguda
Síntomas							
insuficiencia suprarrenal	IS primaria	no	no	no	no	no	no
	IS secundaria	no	no	no	no	no	no
anemia hemolítica	AH congénitas	no	no	no	no	no	no
	AH adquiridas	no	no	no	no	no	no
rinorrea	recurrente	si	si	no	no	nose	nose
	hialina	no	no	si	no	no	no
	clara	no	no	no	no	no	no
	serosa	no	no	no	no	no	no
	purulenta	no	si	no	no	no	si
obstruccion nasal		si	no	no	no	no	si
impetigo nasal							
adenitis cervical							
epifora - lagrimeo		si	si	si	no	si	no
mialgia		si	si	no	no	no	si
hiposmia		no	no	si	no	no	no
incontinencia urinaria		no	no	no	no	no	no
hereditario		no	no	no	no	no	no
hiperemia conjuntival		no	no	si	no	no	no
ojeras		no	no	si	no	no	no
quemosis		no	no	si	no	no	no
conjuntivitis		no	no	no	no	no	no
eritema		no	no	no	no	no	no
exantema		no	no	no	no	no	no
petequias	paladar	no	no	no	no	no	no
ulceras	orofaringe	no	no	no	no	no	no
otitis media		no	no	no	no	no	no
hipoxia		no	no	no	no	no	no

Ilustración 3: Matriz de Enfermedades y Síntomas

Solo es muestra parte de la matriz elaborada de enfermedades y síntomas

Modelado de Clases:

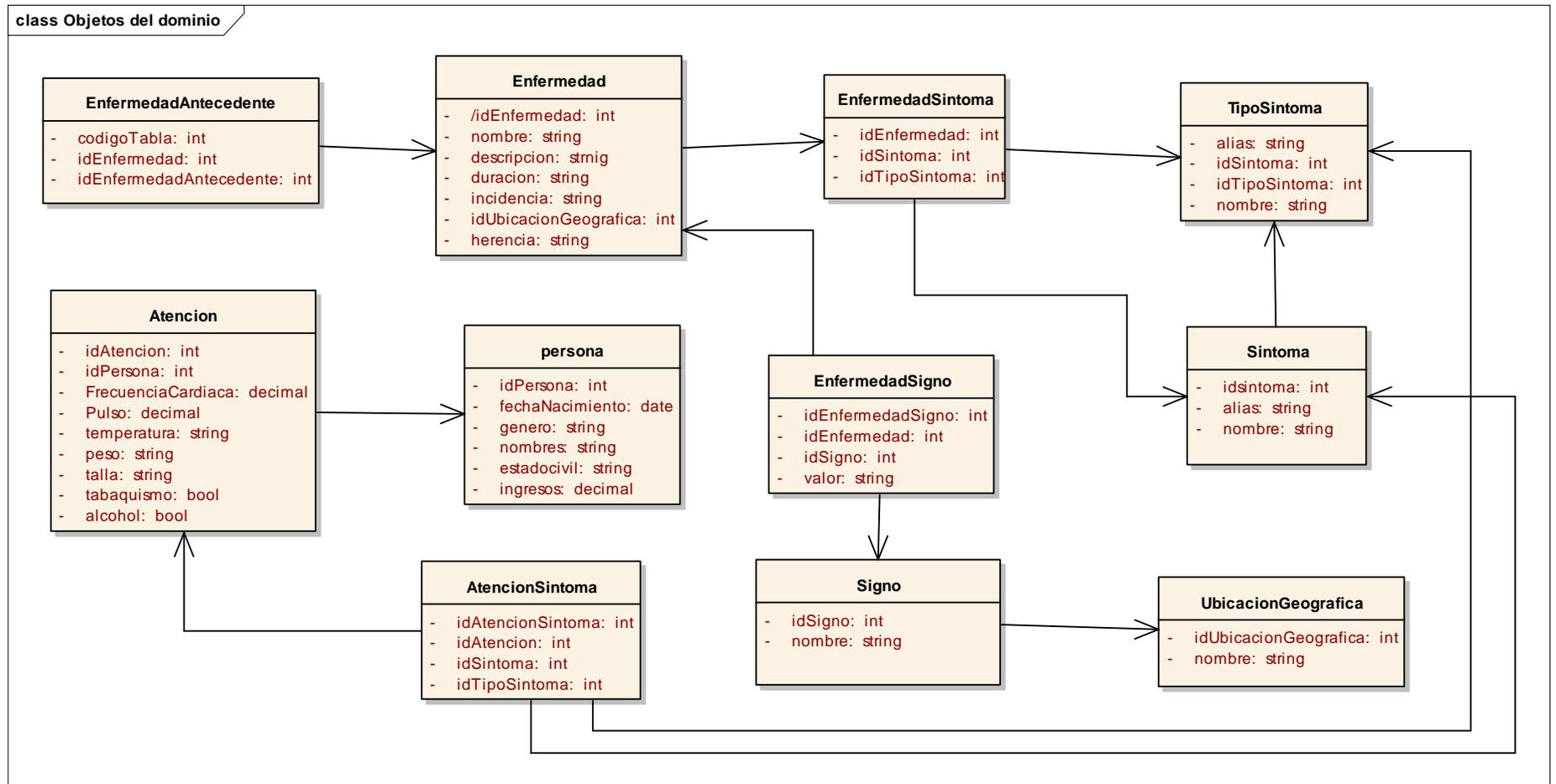


Ilustración 4: Modelo de Clases

Modelo Lógico

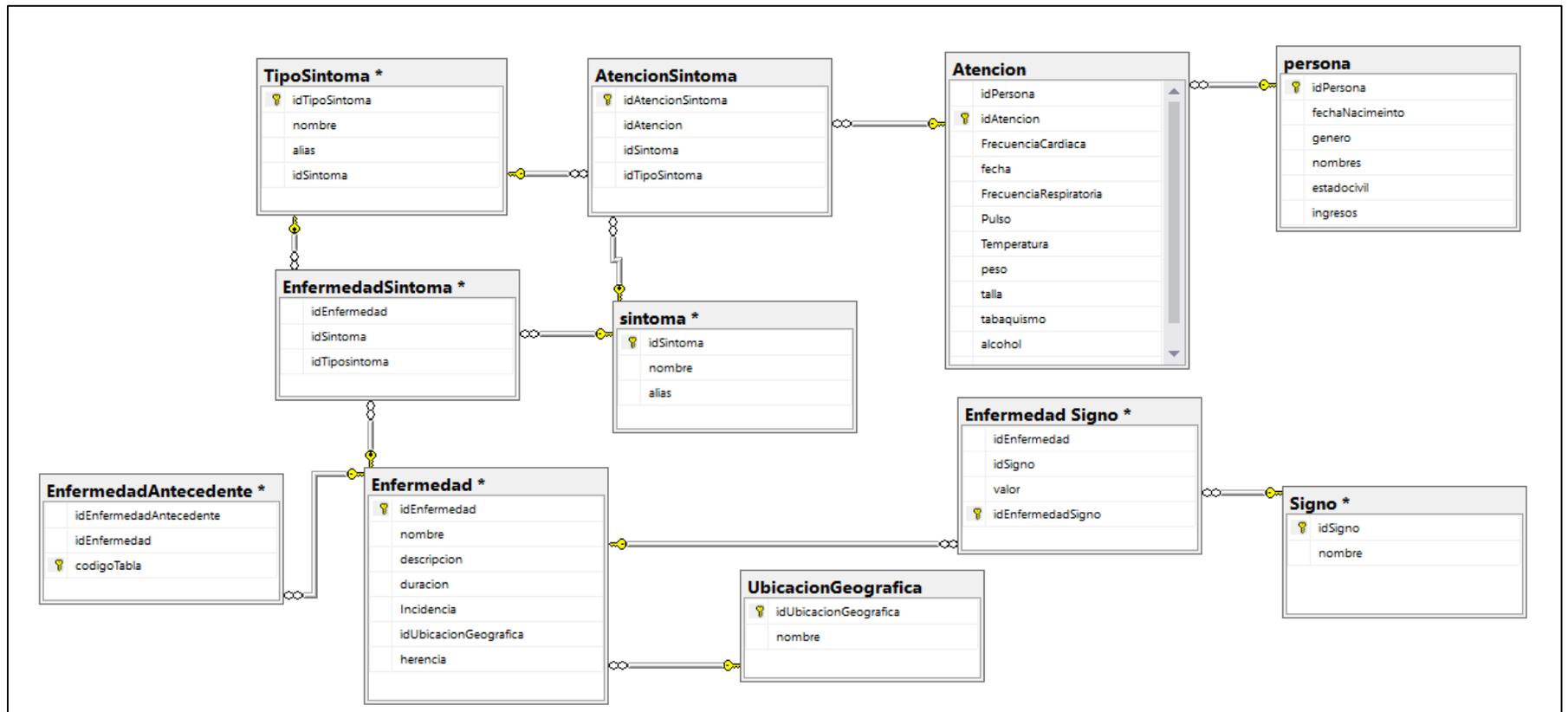


Ilustración 5: Modelo Lógico

Descripción de los Datos

Base de Datos

Base de Datos	Tamaño
Clinicabd	100MB

Tabla 1: Descripción de la Base de Datos

Descripción de las tablas a usar

TABLA	DESCRIPCION
Enfermedad	Enfermedades Respiratorias
Signo	Signos de la enfermedad
Sintoma	Síntomas de la enfermedad
Enfermedad Antecedente	Antecedente de enfermedades
Tipo Síntoma	Tipo de Síntoma
Persona	Paciente de un centro médico
Atencion	Atenciones generadas por la persona

Tabla 2: Descripción Tablas

Cantidad de Registros de los datos muestras asignados

TABLA	Nro Registros
Enfermedad	35
Signo	7
Sintoma	127
Enfermedad Antecedente	20
Tipo Síntoma	174
Persona	18484
Atencion	3000

Tabla 3: Cantidad de Registros por tablas

Exploración de los Datos

- Enfermedades y el número de sus síntomas

Enfermedad	Nro de Sintoma
Adenoiditis	5
bronquiolitis	9
bronquitis aguda	4
Difteria	1
epiglotitis	6
faringoamidgalitis aguda bacteria	11
faringoamidgalitis aguda virus	7
faringoamidgalitis cronica	3

Influenza	4
laringitis	4
laringotraqueitis	5
laringotraqueobronquitis	5
nasofaringitis	9
neumonía adquirida en la comunidad	8
Neumonía atípica Mycoplasma	14
neumonía bacteria	13
neumonía bacteria meningococo	14
neumonía bacteria neumococo	15
neumonía virus Haemophilus influenzae tipo b	13
resfrío por bacteria	7
rinitis alérgica (tipo de rinitis crónica)	8
rinitis crónica	3
rinofaringitis aguda	8
saos	2
sinusitis aguda	6
sinusitis crónica	6
sinusitis recurrente	6
sinusitis subaguda	6
Tosferina	7
Tuberculosis pulmonar	4
tuberculosis ganglionar	1

Tabla 4: Cantidad de Registros por Enfermedad

- Enfermedades y el número de sus Tipos de síntomas

Enfermedad	Nro Tipo Síntoma
Adenoiditis	6
bronquiolitis	16
bronquitis aguda	10
Difteria	2
epiglotitis	8
faringoamigdalitis aguda bacteria	8
faringoamigdalitis aguda virus	7
faringoamigdalitis crónica	4
Influenza	7
laringitis	7
laringotraqueitis	7

laringotraqueobronquitis	6
nasofaringitis	4
neumonía adquirida en la comunidad	16
Neumonía atípica Mycoplasma	12
neumonía bacteria	12
neumonía bacteria meningococo	13
neumonía bacteria neumococo	12
neumonía virus Haemophilus influenzae tipo b	13
resfrío por bacteria	12
rinitis alérgica (tipo de rinitis crónica)	10
rinitis crónica	6
rinofaringitis aguda	9
saos ()	4
sinusitis aguda	7
sinusitis crónica	6
sinusitis recurrente	5
sinusitis subaguda	5
tosferina	4
Tuberculosis pulmonar	6
tuberculosis ganglionar	1

Tabla 5: Cantidad de Registros por Tipo Síntoma

- Enfermedades y el número de signos

Enfermedad	Nro Signo
bronquiolitis	3
Difteria	2
epiglotitis	1
faringoamigdalitis aguda bacteria	2
faringoamigdalitis aguda virus	2
faringoamigdalitis crónica	2
Influenza	1
laringitis	1
laringotraqueitis	1
laringotraqueobronquitis	2
nasofaringitis	2
neumonía adquirida en la comunidad	3
Neumonía atípica Mycoplasma	1
neumonía bacteria	2
neumonía bacteria meningococo	1

neumonia bacteria neumococo	2
neumonia virus Haemophilus influenzae tipo b	2
resfrio por bacteria	1
saos ()	1
sinusiti aguda	2
sinusitis cronica	2
sinusitis recurrente	2
sinusitis subaguda	2
tosferina	1
Tuberculosis pulmonar	1
tuberculosis ganglionar	1

Tabla 6: Cantidad de Registros por Signo

- Enfermedades y sus antecedentes

Enfermedad	Nro Antecedente
asma	1
bronquiolitis	2
epiglotitis	1
faringoamidgalitis aguda bacteria	3
Influenza	1
laringitis	1
laringotraqueitis	1
laringotraqueobronquitis	1
neumonia adquirida en la comunidad	4
Neumonia atipica Mycoplasma	2
neumonia bacteria	1
neumonia bacteria meningococo	2
neumonia bacteria neumococo	4
neumonia virus Haemophilus influenzae tipo b	1
rinitis alergica (tipo de rinitis cronica)	1
Tuberculosis pulmonar	1

Tabla 7: Cantidad de Registros por antecedentes

Verificación de la Calidad de los Datos

- Integridad de Entidad:
Se puede apreciar que las tablas suministradas mantienen una clave primaria lo cual asegura unicidad de las operaciones.
- Integridad de Dominio:
Se aplicaron las siguientes órdenes:

```
Select distinct idTipoSintoma from EnfermedadSintoma  
Select distinct idSintoma from EnfermedadSintoma  
Select distinct idSigno from [EnfermedadSigno]
```

- Integridad Referencial
Las tablas se encuentran referenciada

PASO 3: PREPARACIÓN DE LOS DATOS

Para la preparación de los datos se hará lo siguiente:

- Preparar la colección de datos a usar
- Selección de los Datos
- El servicio de limpieza y procesamiento de datos
- Selección de características

Preparación de la colección de datos

Conjunto de Datos:

- Valores observados hoy en las variables que se espera predecir con el modelo.

Estas variables a predecir están en función a cada objetivo definido en la etapa de comprensión del negocio, por ello se observa el siguiente cuadro:

OBJETIVOS DEL NEGOCIO	OBJETIVOS MINERIA DE DATOS	VARIABLE A PREDECIR
Mejorar el diagnóstico de pacientes	Modelo de Clasificación de enfermedades	Enfermedades a partir de síntomas y signos

Tabla 8: Objetivos definidos en la comprensión del negocio

El inventario de datos de trabajo a usar será el siguiente:

Enfermedad

Feature	Count	Unique Value Count	Missing Value Count	Min	Max	Mean	Mean Deviation	1st Quartile	Median	3rd Quartile	Mode
	38	38	0	1	39	20.184211	9.815789	10.25	20.5	29.75	(1,2,3,4,5,6,7,15,16,17,18,19,4,25,26,27,2,3,34,35,36,3
nombre	38	38	0								
duracion	21	19	17								
Incidencia	25	16	13								
herencia	0	1	38								
UbicacionGeografica	13	2	25								

Ilustración 6: Inventario de datos de la tabla Enfermedad

Enfermedad Síntoma

Feature	Count	Unique Value Count	Missing Value Count	Min	Max	Mean	Mean Deviation	1st Quartile	Median	3rd Quartile	Mode	Range
	366	29	0	1	35	19.065574	8.407059	11.25	22	27	25	34
enfermedad	366	29	0									
idSintoma	366	51	0	3	125	70.557377	44.131506	15.75	96	114.75	6	122
sintoma	366	51	0									
alias	44	6	322									

Ilustración 7: Inventario de datos de la tabla Enfermedad Síntoma

Enfermedad Signo

Feature	Count	Unique Value Count	Missing Value Count	Min	Max	Mean	Mean Deviation	1st Quartile	Median	3rd Quartile	Mode
idEnfermedadSigno	60	60	0	1	60	30.5	15	15.75	30.5	45.25	{1,2,3,4,5,6,7,14,15,16,17,18,24,25,26,27,33,34,35,36,42,43,44,45,50,51,52,53,59,60}
idEnfermedad	60	26	0	2	34	17.466667	8.166667	9	17.5	25	{6,7,8,9,20}
Enfermedad	60	26	0								
idSigno	60	6	0	1	6	2.083333	0.869444	1	2	2	2

Ilustración 8: Inventario de datos de la tabla Enfermedad Signo

Tipo síntoma

Feature	Count	Unique Value Count	Missing Value Count	Min	Max	Mean	Mean Deviation	1st Quartile	Median	3rd Quartile	Mode	Range
idEnfermedad	490	31	0	1	35	17.77551	8.10329	10	20	25	{20,22}	34
enfermedad	490	31	0									
idTipoSintoma	490	73	0	1	173	49.110204	39.385789	8	34	73	3	172
Tipo_sintoma	490	73	0									

Ilustración 9: Inventario de datos de la tabla Tipo síntoma

Enfermedad Antecedente

Feature	Count	Unique Value Count	Missing Value Count	Min	Max	Mean	Mean Deviation	1st Quartile	Median	3rd Quartile	Mode	Range
idEnfermedad	27	16	0	3	29	20.555556	5.08642	16.5	22	25	{22,25}	26
enfermedad	27	16	0									
idAntecedente	27	14	0	6	39	26.740741	10.30727	17	34	36.5	34	33
Antecedente	27	14	0									

Ilustración 10: Inventario de datos de la tabla Antecedente Enfermedad

Selección de datos

De acuerdo a la data preparada procederemos a la selección respectiva.

Variables incluidas

Enfermedades

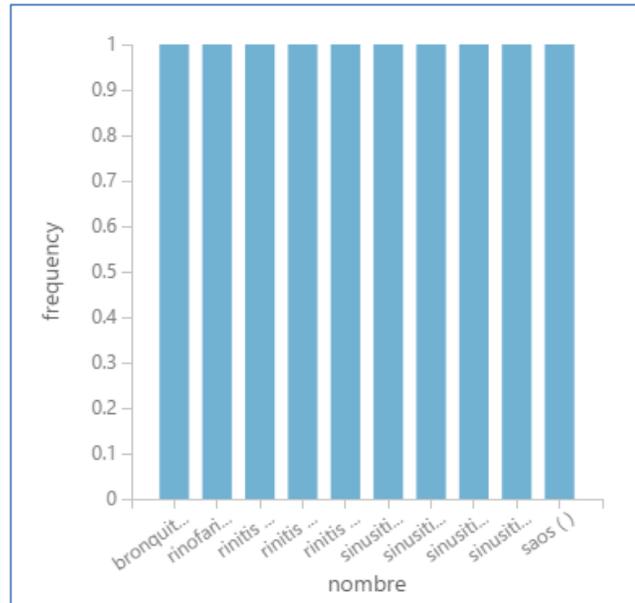


Ilustración 11: Variables de Estudio de la tabla Enfermedad

Tipo síntoma

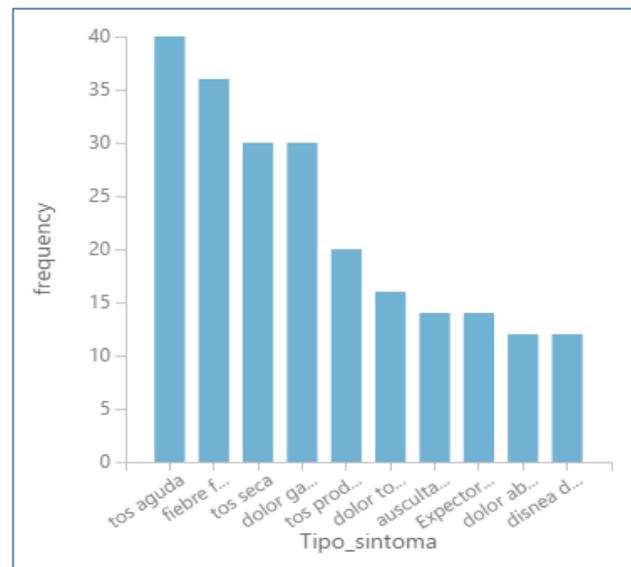


Ilustración 12: Variables de Estudio de la tabla Tipo Síntoma

Síntoma

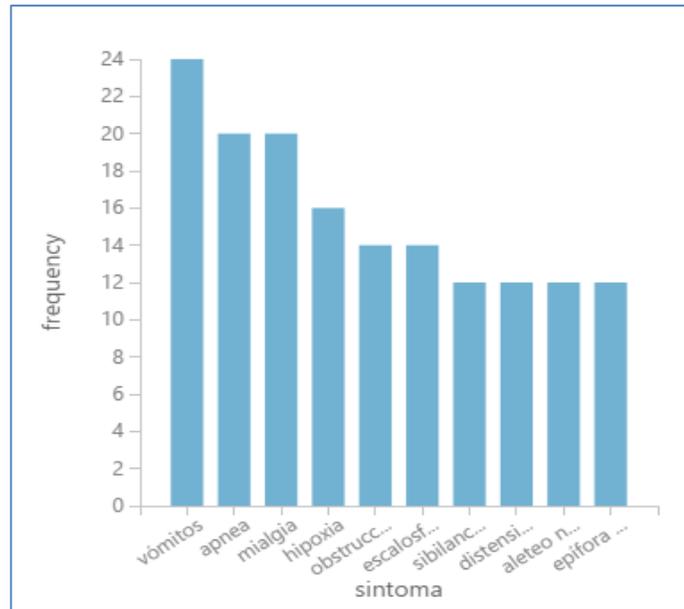


Ilustración 13: Variables de Estudio de la tabla Síntoma

Signo

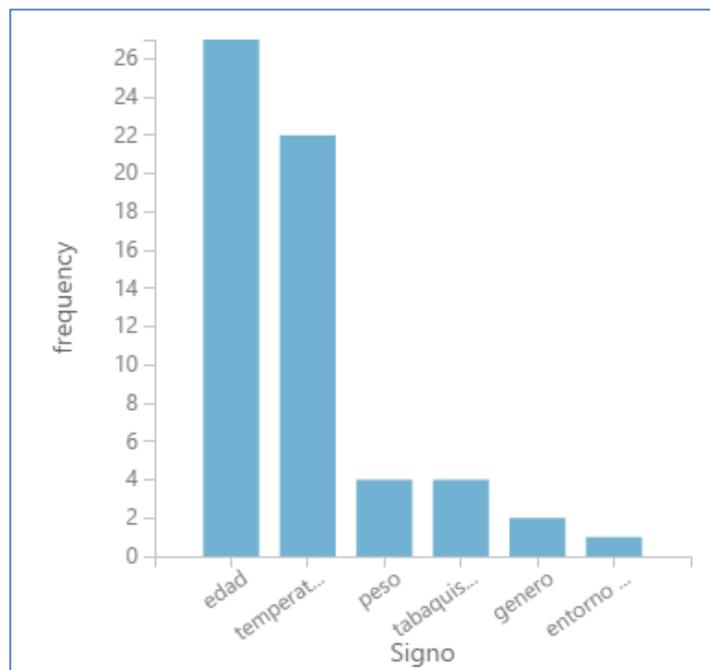


Ilustración 14: Variables de Estudio de la tabla Signo

Enfermedad Antecedente

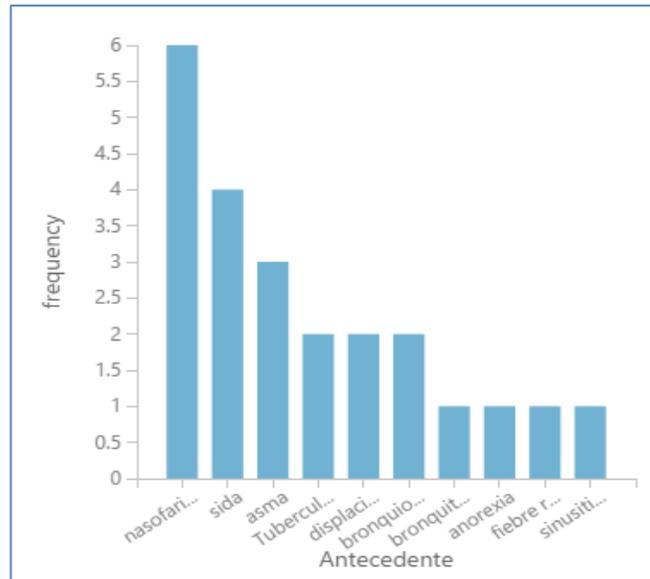


Ilustración 15: Variables de Estudio de la tabla Enfermedad Antecedente

Variables excluidas

La ubicación geográfica

Tabla 9: Registros de la tabla Ubicación Geográfica

nombre	nro
ninguno	21
zonas frías	13
zonas lluviosa	1

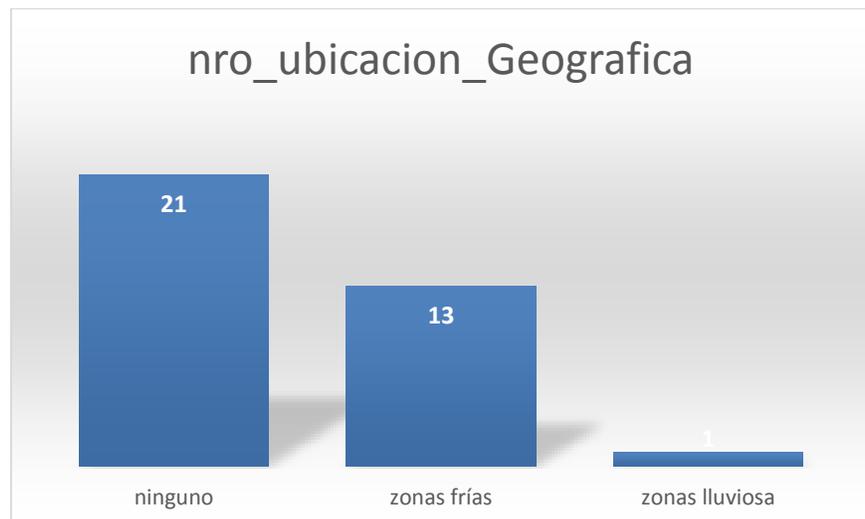


Ilustración 16: Variables de Estudio de la tabla Ubicación Geográfica

Se observa que el 60% es ninguno y zona fría es el 37.15% por lo que el estudio se realizara omitiendo esta variable.

El servicio de limpieza y procesamiento de datos

Estos son los registros a depurar:

Enfermedades

Al validar la data el campo descripción en los datos null serán cambiados a “ “. Al igual que en duración, incidencia será cambiado a no, ubicación geográfica será cambiado a 0 y herencia cambiado a 0.

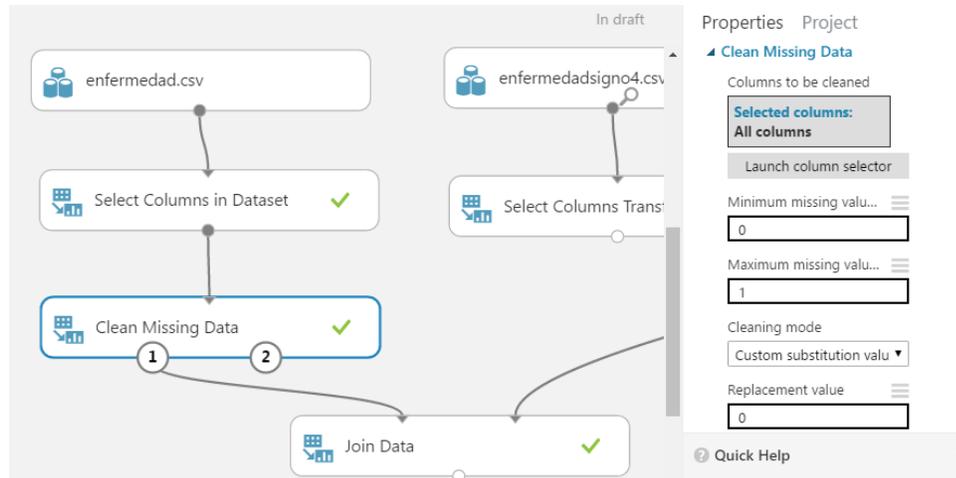


Ilustración 17: Clean de data

Procesamiento de Datos

Obteniendo la edad de la persona

Aplicamos la función datediff que calcula la diferencia en una frecuencia de tiempo determinada, en este caso en años respecto a la fecha de hoy

```
select p.*, edad = DATEDIFF(YY,p.fechaNacimiento,GETDATE())  
from persona p
```

Selección de características

La primera vista será correspondiente a los signos y enfermedades:

```
SELECT DISTINCT e.nombre, dbo._getSignos(e.nombre) AS signos  
FROM          dbo.[Enfermedad Signo] AS es INNER JOIN  
              dbo.Enfermedad AS e ON es.idEnfermedad =  
e.idEnfermedad
```

La siguiente vista será correspondiente a los síntomas y las enfermedades:

```
SELECT DISTINCT e.nombre, dbo._getSintomas(e.nombre) AS
Sintomas
FROM      dbo.EnfermedadSintoma AS es INNER JOIN
          dbo.Enfermedad AS e ON es.idEnfermedad =
e.idEnfermedad
```

La siguiente vista será correspondiente a los tipos de síntomas y las enfermedades:

```
SELECT DISTINCT e.nombre, dbo._getTipoSintomas(e.nombre) AS
Tipo_Sintomas
FROM  dbo.EnfermedadSintoma AS es INNER JOIN
      dbo.Enfermedad AS e ON es.idEnfermedad = e.idEnfermedad
```

La vista Enfermedad

```
create view enfermedad as
select e.nombre,e.duracion,e.Incidencia,e.herencia ,u.nombre
from Enfermedad e left join UbicacionGeografica u on
e.idUbicacionGeografica = u.idUbicacionGeografica
where e.idUbicacionGeografica is null or e.idUbicacionGeografica =
u.idUbicacionGeografica
```

La vista Antecedente

```
SELECT DISTINCT e.nombre, dbo._getAntecedentes(e.nombre) AS
Antecedentes
FROM  dbo.EnfermedadAntecedente AS es INNER JOIN
      dbo.Enfermedad AS e ON es.idEnfermedad = e.idEnfermedad
```

La vista atenciones:

```
SELECT a.idAtencion, a.idPersona, p.nombres, DATEDIFF(YY,
p.fechaNacimeinto, GETDATE()) AS edad,
a.peso, a.Temperatura, a.tabaquismo, a.alcohol, p.genero
FROM  dbo.Atencion AS a INNER JOIN
      dbo.persona AS p ON a.idPersona = p.idPersona
```

PASO 4: MODELADO

Seleccionar la técnica de modelado

Para el modelo de clasificación se realizará con el algoritmo de árboles de decisión porque permite valores continuos y discretos. Adecuado para la predicción

Construir el modelo

Se usará los servicios de Azure Machine Learning.

Para la construcción del modelado se aplica el anterior paso que es preparar data y queda de la siguiente manera:



Ilustración 18: Preparar la data

Construcción de Algoritmo

Se va a usar Ipython para el análisis de datos el algoritmo usado es el siguiente:

Para construir el árbol de decisiones se realiza lo siguiente:

```
import pandas as pd
from math import log
def calcShannonEnt(dataSet):
    numEntries = len(dataSet)
    labelCounts = {}
    for featVec in dataSet:
        currentLabel = featVec[-1]
        if currentLabel not in labelCounts.keys():
            labelCounts[currentLabel] = 0
            labelCounts[currentLabel] += 1
    shannonEnt = 0.0
    for key in labelCounts:
        prob = float(labelCounts[key])/numEntries
        shannonEnt -= prob * log(prob,2)
    return shannonEnt
```

```

def splitDataSet(dataSet, axis, value):
    retDataSet = []
    for featVec in dataSet:
        if featVec[axis] == value:
            reducedFeatVec = featVec[:axis]
            reducedFeatVec.extend(featVec[axis+1:])
            retDataSet.append(reducedFeatVec)
    return retDataSet

def chooseBestFeatureToSplit(dataSet):
    numFeatures = len(dataSet[0]) - 1
    baseEntropy = calcShannonEnt(dataSet)
    bestInfoGain = 0.0; bestFeature = -1
    for i in range(numFeatures):
        featList = [example[i] for example in dataSet]
        uniqueVals = set(featList)
        newEntropy = 0.0
        for value in uniqueVals:
            subDataSet = splitDataSet(dataSet, i, value)
            prob = len(subDataSet)/float(len(dataSet))
            newEntropy += prob * calcShannonEnt(subDataSet)
        infoGain = baseEntropy - newEntropy
        if (infoGain > bestInfoGain):
            bestInfoGain = infoGain
            bestFeature = i
    return bestFeature

def majorityCnt(classList):
    classCount={}
    for vote in classList:
        if vote not in classCount.keys(): classCount[vote] = 0
        classCount[vote] += 1
    sortedClassCount = sorted(classCount.iteritems(),key=operator.itemgetter(1), reverse=True)
    return sortedClassCount[0][0]

```

```

def createTree(dataSet, labels):
    classList = [example[-1] for example in dataSet]
    if classList.count(classList[0]) == len(classList):
        return classList[0]
    if len(dataSet[0]) == 1:
        return majorityCnt(classList)
    bestFeat = chooseBestFeatureToSplit(dataSet)
    bestFeatLabel = labels[bestFeat]
    myTree = {bestFeatLabel: {}}
    del(labels[bestFeat])
    featValues = [example[bestFeat] for example in dataSet]
    uniqueVals = set(featValues)
    for value in uniqueVals:
        subLabels = labels[:]
        myTree[bestFeatLabel][value] = createTree(splitDataSet\
            (dataSet, bestFeat, value), subLabels)
    return myTree

```

Evaluar el modelo

Se realizó un árbol de decisión para síntoma_enfermedad, tipoSíntoma_enfermedad, signo_enfermedad, antecedente_enfermedad y se muestra a continuación:

Árbol Síntoma Enfermedad:

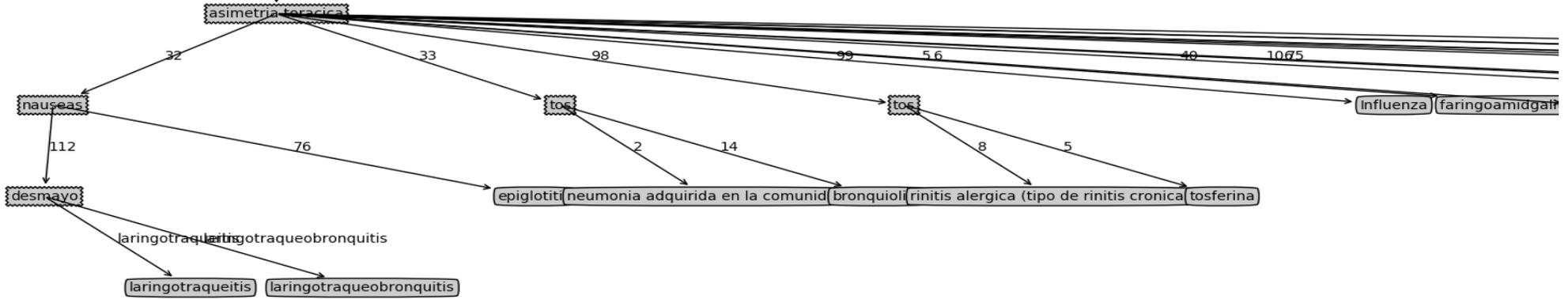


Ilustración 19: Árbol Síntoma Enfermedad 1

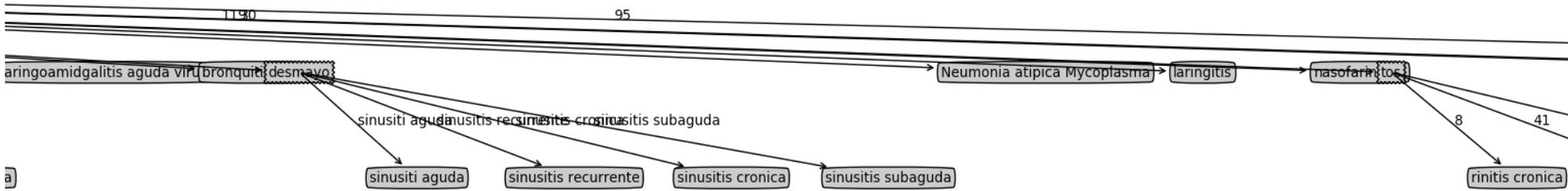


Ilustración 20: Árbol Síntoma Enfermedad

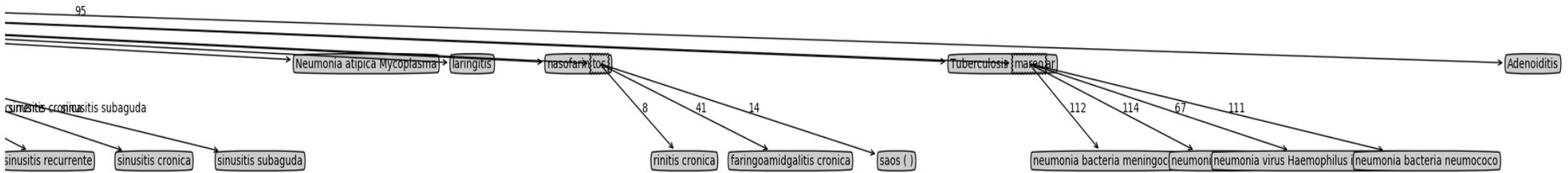


Ilustración 21: Árbol Síntoma Enfermedad 3

Árbol Tipo Síntoma Enfermedad:

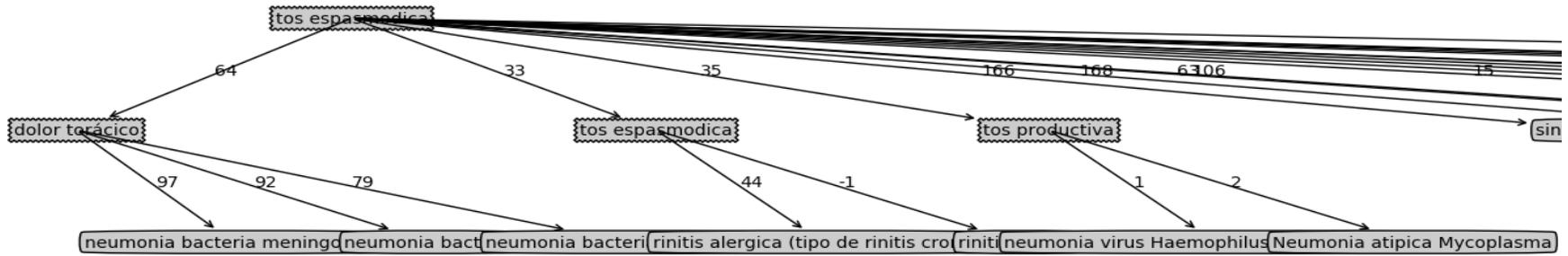


Ilustración 24: Árbol Tipo Síntoma Enfermedad 1

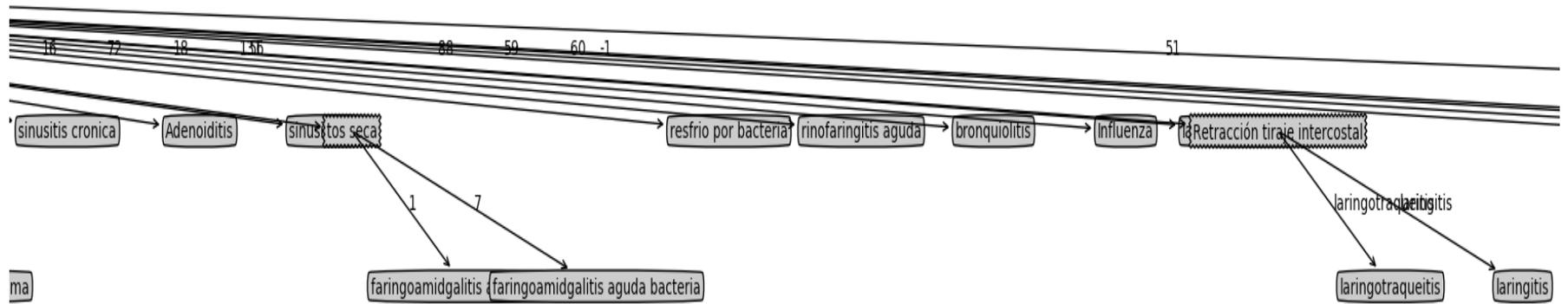


Ilustración 23: Árbol Tipo Síntoma Enfermedad 2

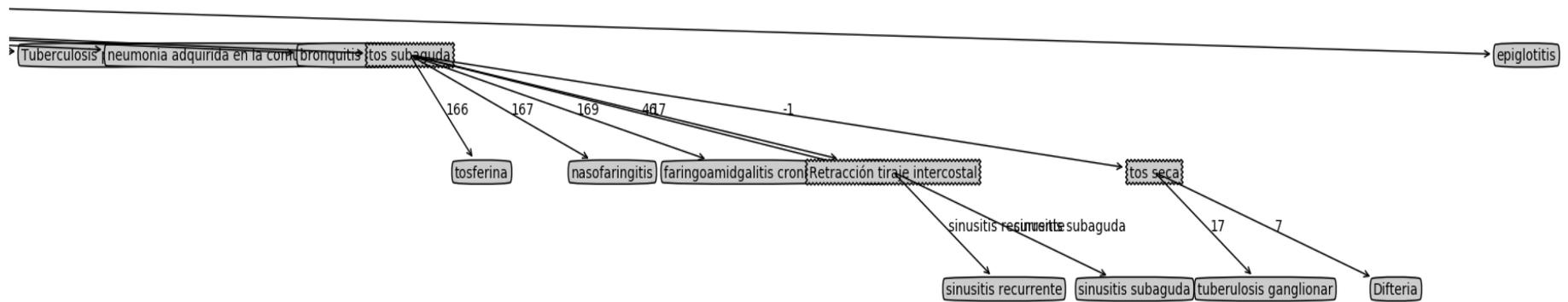


Ilustración 22: Árbol Tipo Síntoma Enfermedad 3

Árbol Signo Enfermedad:

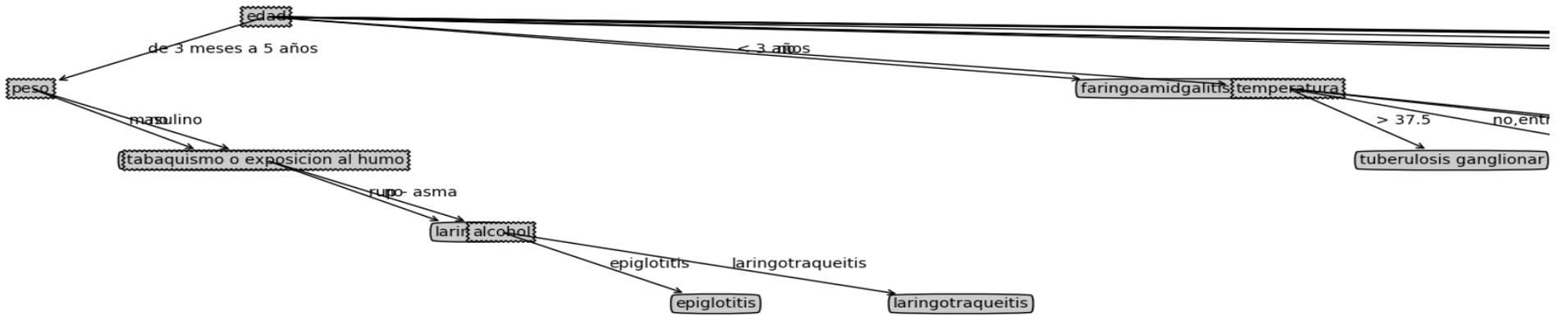


Ilustración 27: Árbol Signo Enfermedad 1

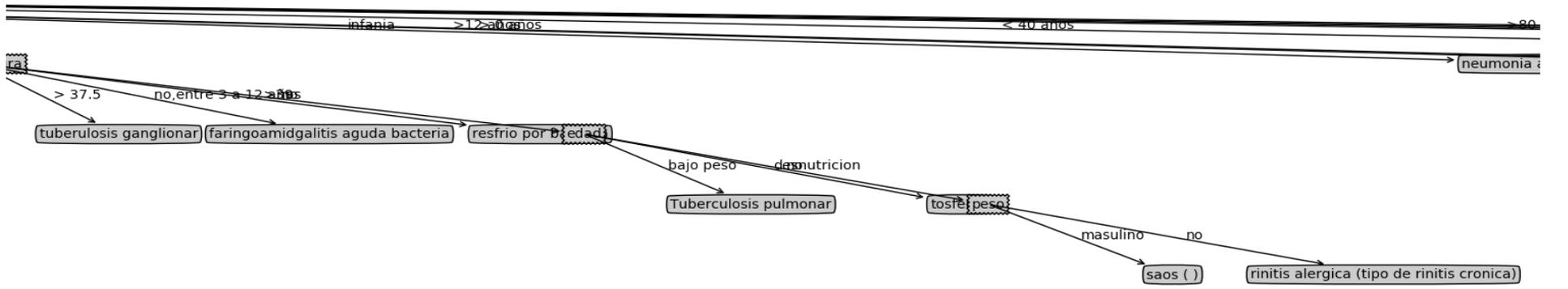


Ilustración 26: Árbol Signo Enfermedad 2

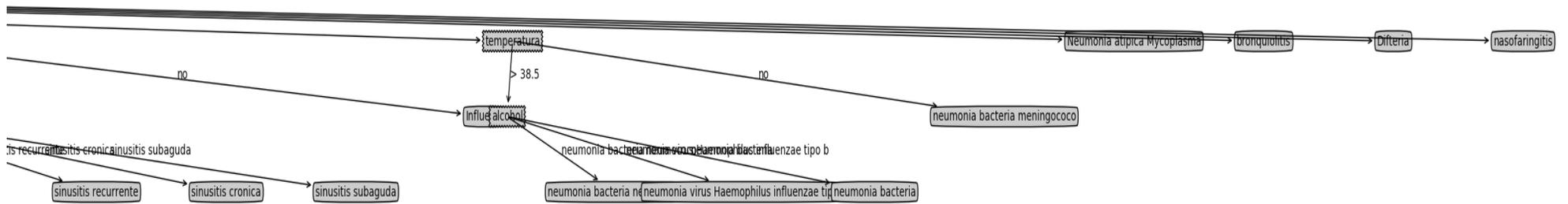


Ilustración 25: Árbol Signo Enfermedad 3

Árbol Antecedente Enfermedad:



Ilustración 28: Árbol Antecedente Enfermedad 1

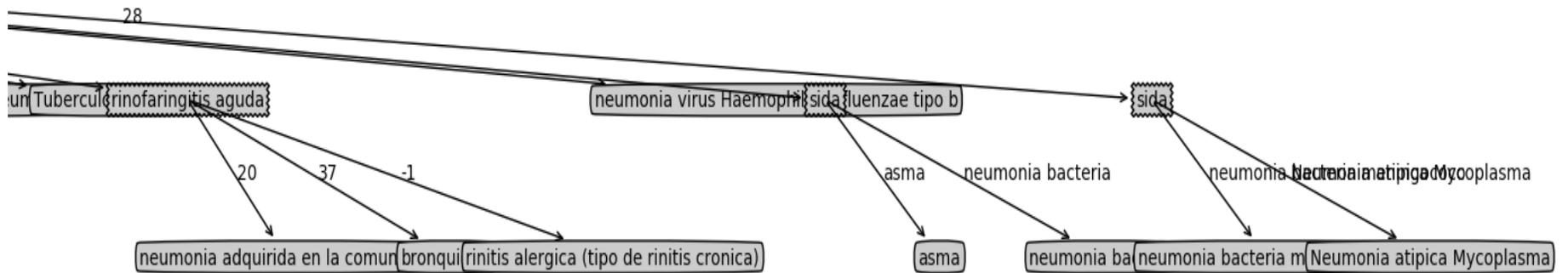


Ilustración 29: Árbol Antecedente Enfermedad 2

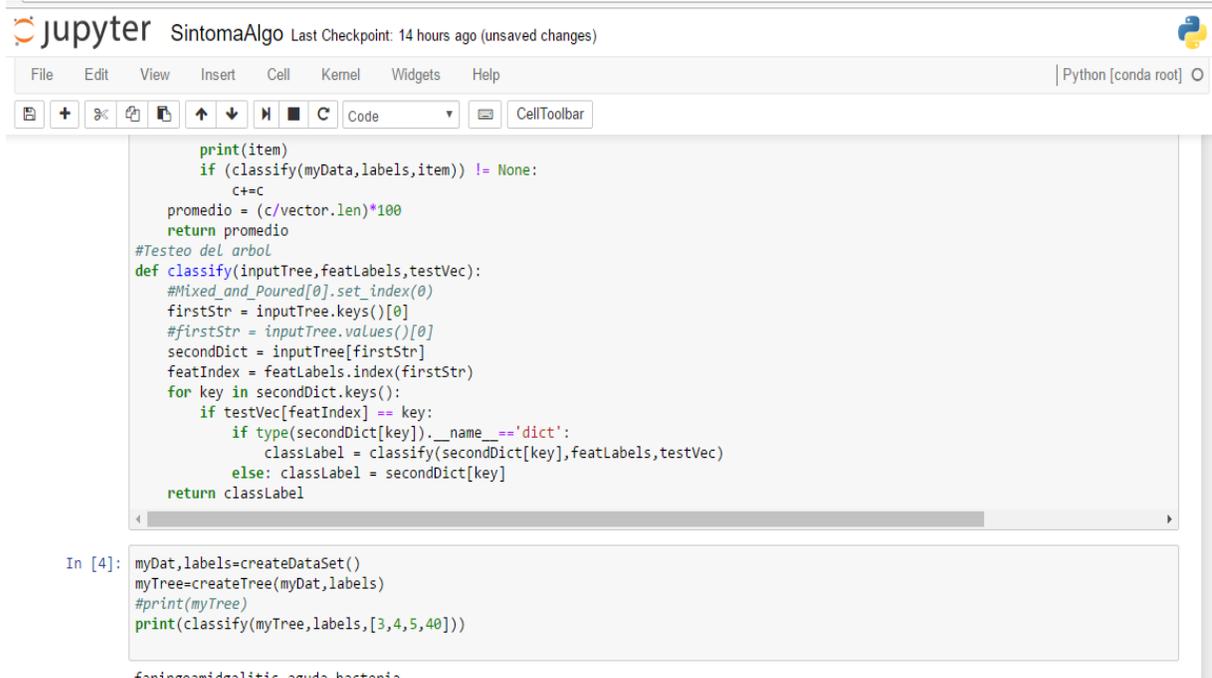
Evaluación del Modelo

Se ejecuta el siguiente algoritmo para realizar el testeo del árbol construido este algoritmo se adapta para los 4 árboles de decisión:

```
def classify(inputTree, featLabels, testVec):
    #Mixed_and_Poured[0].set_index(0)
    firstStr = inputTree.keys()[0]
    #firstStr = inputTree.values()[0]
    secondDict = inputTree[firstStr]
    featIndex = featLabels.index(firstStr)
    for key in secondDict.keys():
        if testVec[featIndex] == key:
            if type(secondDict[key]).__name__=='dict':
                classLabel = classify(secondDict[key], featLabels, testVec)
            else: classLabel = secondDict[key]
    return classLabel
```

En la siguiente ventana se está enviando el vector siguiente:[3,4,5,40] solo algunos síntomas de la faringoamidgalitis aguda por bacteria y te devuelve la enfermedad que coincide con ello.

Resultados:



```
Jupyter SintomaAlgo Last Checkpoint: 14 hours ago (unsaved changes) Python [conda root]

File Edit View Insert Cell Kernel Widgets Help

print(item)
if (classify(myData,labels,item)) != None:
    C+=C
promedio = (C/vector.len)*100
return promedio
#Testeo del arbol
def classify(inputTree, featLabels, testVec):
    #Mixed_and_Poured[0].set_index(0)
    firstStr = inputTree.keys()[0]
    #firstStr = inputTree.values()[0]
    secondDict = inputTree[firstStr]
    featIndex = featLabels.index(firstStr)
    for key in secondDict.keys():
        if testVec[featIndex] == key:
            if type(secondDict[key]).__name__=='dict':
                classLabel = classify(secondDict[key], featLabels, testVec)
            else: classLabel = secondDict[key]
    return classLabel

In [4]: myDat,labels=createDataSet()
myTree=createTree(myDat,labels)
#print(myTree)
print(classify(myTree,labels,[3,4,5,40]))

faringoamidgalitis aguda bacteria
```

Ilustración 30: Testeo con Python

Prueba de Hipótesis

La Contrastación de Hipótesis se realizado de acuerdo al Método Propuesto Pre Test - Pos Test, para poder aceptar o rechazar la hipótesis. Así mismo, para la realización de este diseño se identificaron indicadores cuantitativos, los cuales se describen a continuación:

INDICADOR	TIPO
Tiempo de realizar diagnóstico al paciente	Cuantitativo
Diagnósticos acertados	Cuantitativo

Tabla 10: Tipo de Indicadores

- **Prueba de Hipótesis para el indicador de Tiempo de realizar diagnósticos**
 - **Definición de Variables**
 - Ta = Tiempo antes de aplicar el modelo de minería de datos
 - Td = Tiempo después de aplicar el modelo de minería de datos
 - **Hipótesis Estadística**
 - Hipotesis Ho=** el tiempo de atención antes de aplicar el modelo de minería de datos es menor o igual que el tiempo de atención después de aplicar el modelo de minería de datos(Minutos)
 - $H_o = T_a - T_d \leq 0$**
 - Hipotesis Ha =** el tiempo de atención antes de aplicar el modelo de minería de datos es mayor que el tiempo de atención después de aplicar el modelo de minería de datos(Minutos)
 - $H_a = T_a - T_d > 0$**
 - **Nivel de Significancia**
 - Se define un nivel de confiabilidad de 95%.
 - Usando el nivel de significancia($\alpha=0.05$) del 5 % será el margen de error.
 - Por lo tanto el nivel de confianza($1 - \alpha = 0.95$) será del 95%.

- **Estadígrafo de contraste**

Tipo de atención	Antes	Después	Antes	Después
	\bar{T}_a	\bar{T}_d	$(\bar{T}_a)^2$	$(\bar{T}_d)^2$
Minsa	13	11	169	121
Essalud	17	12	289	144
Clinica	21	15	441	225
Sumatoria	51	38	899	390
Promedio	17	11.667	289	136.12
Varianza				

Tabla 11: Tiempo de atención de paciente

Valores obtenidos para antes de aplicar el patrón de la Encuesta realizada por ENSUSALUD 2014 .

Valores obtenidos para después de aplicar el patrón consulta a dos Doctores que dicto los tiempos promedios (ver Anexo 2).

Varianza:

$$\sigma_a^2 = \frac{\sum_{i=1}^n T_a - \bar{T}_a^2}{n_a} = \frac{12194 - 289}{790} = 15,07$$

$$\sigma_d^2 = \frac{\sum_{i=1}^n T_d - \bar{T}_d^2}{n_d} = \frac{9217 - 136.12}{790} = 11.495$$

Calculo de Z:

$$Z_c = \frac{\bar{T}_a - \bar{T}_d}{\sqrt{\left(\frac{\sigma_a^2}{n_a} + \frac{\sigma_d^2}{n_d}\right)}} = \frac{17 - 11.667}{0.205301} = 25.977$$

- **Región Crítica:**

Para $\alpha = 0.05$, en la tabla N° ,(Anexo N°) encontramos $Z_\alpha = 1.645$.

Entonces la región crítica de la prueba es $Z_c < 1.645 >$.

- **Conclusión**

Puesto que $Z_c = 25.997$ calculado, es mayor que $Z_\alpha = 1.645$ y estando este valor dentro de la región de rechazo $> 1.645 <=$, entonces se rechaza la H_0 y por consiguiente se acepta H_a . Se concluye entonces que el tiempo de atención al paciente después de

aplicar el modelo es menor al tiempo de atención antes de aplicar el modelo con un nivel de error del 5% y un nivel de confianza del 95%.

▪ **Prueba de Hipótesis para el indicador de Diagnósticos Acertados**

○ **Definición de Variables**

R_p = número de aciertos

○ **Hipótesis Estadística**

Hipótesis H_0 = el número de aciertos es mayor al 70% de lo estimado

$$H_0 = R_p > 70\%$$

Hipótesis H_a = el número de aciertos es menor o igual al 70% de lo estimado

$$H_a = R_p \leq 70\%$$

○ **Nivel de significancia**

Se define el margen de error en ($\alpha=0.05$) 5% por lo tanto con un nivel de confianza de ($1-\alpha = 0.95$) 95%

○ **Estadígrafo de contraste**

$n = 3000$ y coloco la formula

Para calcular el rendimiento de construyo un algoritmo y se sacara un promedio de ello con cada árbol de decisión

Para Enfermedad_Sintoma = 1000

Para Enfermedad_TipoSintoma = 1000

Para Enfermedad_Signo = 190

Para Enfermedad_Antecedente = 10

Enfermedad Síntoma		Enfermedad Tipo Síntoma		Enfermedad Signo		Enfermedad Antecedente	
Aciertos	Promedio %	Acierto	Promedio %	Acierto	Promedio %	Acierto	Promedio %
854	85.4%	781	78.10%	156	82.11%	8	80.00%
Total	100%	2200					

Tabla 12: Tabulación de diagnósticos acertados

Calculo Da:

$$Da = \frac{\sum_{i=1}^n \text{aciertos}}{N} * 100 = \left(\frac{1797}{22} \right) = 81.68$$

○ **Conclusión**

Puesto que $Da = 81.68$ es mayor que 70% por lo tanto se acepta la H_0 y se rechaza la H_a

IV. DISCUSIÓN

Tiempo de realizar diagnósticos

Como se puede observar en la tabla N° 13 el tiempo promedio de atención antes de aplicar el modelo es de 17 minutos y el tiempo promedio de atención después de aplicar el modelo es de 11.667 minutos lo que representa una reducción de 5.333 minutos que se representaría de la siguiente manera:

Ta		Td		Decremento	
	%		%		%
17	100.00	11.667	68.63	5.333	31.371

Tabla 13: Comparación del Indicador Ta y Td

De acuerdo al cuadro, la reducción representa el 31.371%

Diagnósticos Acertados

Sobre este indicador se observa que el número de aciertos por cada árbol sobrepasa el 70% dando como porcentaje final promedio de 81.68%, por lo tanto, el modelo aplicado es aceptable de acuerdo al cuadro de Operacionalización de variable.

V. CONCLUSIÓN

- El primer indicador alcanzo un 81.77% en promedio de diagnósticos acertados demostrando así parte de la efectividad del proyecto.
- El tiempo de atención también aporta a la efectividad con una reducción de 31.371% en comparación al tiempo promedio de atención anterior a la aplicación de la solución.
- De acuerdo a lo descrito en las anteriores viñetas se concluye que se logró el cumplimiento del objetivo principal que consiste en mejorar los diagnósticos a partir de síntomas y signos.
- Se demuestra los aportes de la minería de datos a la sociedad son de relevancia, claro ejemplo es este proyecto de investigación.

VI. RECOMENDACIÓN

- Delimitar el estudio, como este proyecto se delimito a enfermedades respiratorias, las futuras investigaciones deberían fijar objetivos alcanzables.
- Si se desea aprender más sobre data mining es mejor tener un inglés básico a intermedio ya que la mayor documentación está en lengua americana.
- Si desean realizar este tipo de investigación lo ideal es tener a expertos en el tema para que los patrones sean más acertados.

VII. PROPUESTA

- Acrecentar este modelo no solo para enfermedades respiratorias y pueda predecir otras especialidades medicas

IV. REFERENCIAS

A New, Evidence-based Estimate of Patient Harms Associated with Hospital Care. **JAMES, John T.** 2013. 3, s.l. : Journal of Patient Safety, Setiembre de 2013, Vol. 9, págs. 122-128.

Análisis documental y de información: dos componentes de un mismo proceso. **DULZAIDES IGLESIAS, María Elinor y MOLINA GOMEZ, Ana María.** 2004. 2, Ciudad de La Habana : s.n., 2004, ACIMED, Vol. 12. ISSN 1024-9435.

BEHAR Rivero, Daniel Salomón. 2008. *Introducción a la Metodología de la investigación.* s.l. : Shalom, 2008. ISBN 978-959-212-783-7.

CANDELA Cáceres, Julio Christians. 2015. *Proceso de Descubrimiento de Conocimiento para Predecir el Abandono de Tratamiento en una Entidad de Salud Pública.* Lima : s.n., 2015. ISSN 2310-8894.

CHAPMAN, Pete, y otros. 2000. *Step by step data mining guide.* 2000.

Colegio Médico del Perú. 2011. *El Acto Médico.* Colegio Medico del Perú. s.l. : Fondo Editorial Comunicacional del Colegio Médico del Perú, 2011. ISBN: 978-612-45410-2-5.

Denuncias éticas de médicos colegiados en Perú, 1991-2010: frecuencia y naturaleza. **SOGI Uematzu, Cecilia, CARDENAS Díaz, Maximiliano y ZAVALA Sarrio, Salomon.** 2014. 3, Lima : s.n., Julio/Setiembre de 2014, Acta Médica Peruana, Vol. 31, págs. 165-175. ISSN 1728-5917.

El Registro Médico Orientado por Problemas. **MUÑOZ, Julio.** 1998. 1, Callao : Universidad Mayor de San Marcos, 1998, Vol. 59. 1025-5583.

Errores médicos. **ALVARADO Guevara, Ana Teresa y FLORES Sandí, Grettchen.** 2009. 1, Costa Rica : s.n., 2009, Vol. 51, págs. 16-23. ISSN 0001-6002.

eSalud en Perú: implementación de políticas para el fortalecimiento de sistemas de información en salud. **CURIOSO, Walter H.** 2014. 5-6, Washington : s.n., May/Jun de 2014, Revista Panamericana Salud Publica, Vol. 35, págs. 437-441. ISSN 1680-5348.

Estudio de las denuncias penales por responsabilidad profesional médica en el Instituto de Medicina Legal de Lima, Perú. **NAVARRO Sandoval, Cleyber.** 2013. 3, Lima : s.n., 2013, Rev. perú. med. exp. salud publica, Vol. 30, págs. 408-414. ISSN 1726-4634.

H. Witten, Ian, FRANK, Eibe y A. Hall, Mark. 2011. *Data mining : practical machine learning tools and techniques*. 3. Burlington : Elsevier, 2011. pág. 629. ISBN 978-0-12-374856-0.

La cultura de seguridad del paciente: del pasado al futuro en cuatro tiempos. **M. Aranaz, Jesús y AGRA, Yolanda. 2010.** Suplemento 1, Julio de 2010, Medicina Clínica, Vol. 135, págs. 243-291. ISSN :0025-7753.

LÓPEZ Pérez, César. 2007. *Minería de datos: técnicas y herramientas*. Madrid, España : Paraninfo, 2007. 978-84-9732-492-2.

Organizacion Mundial de la Salud. 2010. *IBEAS: red pionera en la seguridad del paciente en Latinoamérica*. Ginebra : Organización Mundial de la Salud, 2010.

SALAS GUILLEN, DIEGO ANDRES y GONZALES MACEDA, JACKLIN DEL ROCIO. 2014. *USO DE INFERENCIA BASADA EN ONTOLOGIAS PARA DAR SOPORTE AL DIAGNOSTICO VETERINARIO*. Lima : s.n., 2014. 2310-8894.

Técnicas de minería de datos aplicadas al diagnóstico de entidades clínicas. **DÁVILA Hernández, Frank y SANCHEZ Corales, Yovannys. 2012.** 2, Ciudad de la Habana : RCIM [online], 2012, Revista Cubana de Informática Médica, Vol. 4. 1684-1859.

WEE Hyong, Tok, FONTAMA, Valentine y BARGA, Roger. 2015. *Predictive Analytics with Microsoft Azure Machine Learning*. 2. New York : Apress, 2015. ISBN 978-1-4842-1200-4.

ANEXOS

**ANEXO 1: FORMATO DE ENCUESTA NACIONAL DE USUARIOS EN SALUD
DIRIGIDO A USUARIOS DE CONSULTA EXTERNA ENSUAUS 2014**



ENCUESTA NACIONAL DE SATISFACCIÓN DE USUARIOS DEL ASEGURAMIENTO UNIVERSAL EN SALUD

INFORMACIÓN CONFIDENCIAL AMPLIADA EN EL DECRETO SUPLENTO 043-2001-PCM DEL SECRETO ESTADÍSTICO



DIRIGIDA A USUARIOS DE CONSULTA EXTERNA

Hospitales, Clínicas, Centros Médicos, Establecimientos
Minsa, EsSalud, Privados y FFAA. - PNP

ENSUAUS.01

N° Cuestionario

I. LOCALIZACIÓN DEL ESTABLECIMIENTO DE SALUD

1. DEPARTAMENTO			
2. PROVINCIA			
3. DISTRITO			
4. CENTRO POBLADO			

5. COORDENADAS DEL PUNTO GPS

S									
W									
NP Punto									
Cód. Patrimonial									

II. IDENTIFICACIÓN DEL ESTABLECIMIENTO DE SALUD

1. Nombre oficial del establecimiento de salud

2. Código del Establecimiento de Salud Código EE.SS.

3. Institución

Minsa - GR..... 1	ESSALUD..... 2	FFAA. y PNP..... 3	CLÍNICAS..... 4
-------------------	----------------	--------------------	-----------------

4. Consultorio

Medicina..... 1	Cirugía..... 2	Gineco/obstetricia..... 3	Subespecialidades..... 4
-----------------	----------------	---------------------------	--------------------------

5. Turno:

Mañana..... 1	Tarde..... 2	Noche..... 3
---------------	--------------	--------------

III. ENTREVISTA Y SUPERVISIÓN

Visita Nº	EVALUADOR/A TÉCNICO/A					SUPERVISOR/A				
	Fecha	Hora		Próxima Visita		Resultado de la Visita (*)	Fecha	Hora		Resultado de la Visita (*)
		De	A	Fecha	Hora			De	A	
1	/					/				
2	/					/				
3	/					/				
4	/					/				
5	/					/				
6	/					/				

RESULTADO FINAL DE LA ENCUESTA	
FECHA	/ /
RESULTADO	

(*) CÓDIGOS DE RESULTADO				
1. COMPLETO	2. INCOMPLETO	3. RECHAZO ROTUNDO	4. RECHAZO POR FALTA DE TIEMPO	5. OTROS (Especificar)

OBSERVACIÓN:

LUGAR DE ENTREVISTA

Puerta de salida dentro del EESS..... 1
Área de farmacia..... 2
Área de laboratorio..... 3
Área de Rayos X / Ecografía..... 4
Otros..... 5
(Especificar)

POR OBSERVACIÓN DIRECTA

Sexo	Rango de edades	Fenotipo
Hombre..... 1	De 15 a 24 años..... 1	Piel clara..... 1
Mujer..... 2	De 25 a 44 años..... 2	Piel oscura..... 2
	De 45 a 64 años..... 3	Piel morena..... 3
	De 65 años a más..... 4	Piel oriental..... 4

IV. FUNCIONARIOS DE LA ENCUESTA

Cargo	DNI	Nombre y Apellidos
EVALUADOR/A TÉCNICO/A		
SUPERVISOR/A LOCAL		
COORDINADOR/A DEPARTAMENTAL		
SUPERVISOR/A NACIONAL		

SOLO ENTREVISTAR A PACIENTES QUE SALGAN DE LA CONSULTA EXTERNA Y QUE HAYAN SIDO ATENDIDOS POR UN MÉDICO

PRESENTACIÓN:

Buenos días / tardes, mi nombre es....., y trabajo para el Instituto Nacional de Estadística e Informática (INEI), órgano rector del Sistema Estadístico Nacional. En esta ocasión, por encargo de la Superintendencia Nacional de Aseguramiento en Salud, estamos realizando entrevistas a usuarios de los servicios de salud, para medir su percepción del Aseguramiento Universal de Salud. Agradecería, me permita conversar con usted, para hacerle algunas preguntas al respecto. Sus opiniones nos permitirán sugerir mejoras al sistema para en el futuro brindar una mejor atención a los afiliados. Todas sus respuestas serán tratadas en forma confidencial. ¿Me permite entrevistarla?

V. MOTIVO DE LA ATENCIÓN ACTUAL

1. ¿EL PROBLEMA DE SALUD O MOLESTIA POR LA CUAL SE ATENDIÓ HOY, ES:

- Una enfermedad o síntoma de reciente aparición (menor de 15 días)?..... 1
- Una enfermedad antigua (de 15 días a más)?..... 2

Otro?..... 3 → **FASE A 3**
(Específica)

8. EN LA FARMACIA DE ESTE ESTABLECIMIENTO, ¿LE ENTREGARON LOS MEDICAMENTOS QUE LE RECETARON:

- Todos?..... 1
- La mayoría?..... 2
- Algunos?..... 3
- NINGUNO..... 4
- NO HIZO TRAMITE..... 5

2. ¿POR ÉSTA ENFERMEDAD O DOLENCIA UD. VIENE POR:

- Paciente referido?..... 1
- Primera vez?..... 2
- Control de consulta anterior?..... 3

9. EN LA FARMACIA DE ESTE ESTABLECIMIENTO, ¿LE INDICARON QUE COMPRA ALGUN MEDICAMENTO EN OTRO ESTABLECIMIENTO?

- Sí..... 1
- No..... 2
- NO APLICA..... 3 } **FASE A 11**

3. UD. VIENE:

- Solo?..... 1 → **FASE A 4**
- Acompañado?..... 2

¿CUÁL ES LA RELACIÓN DE PARENTESCO QUE TIENE CON SU ACOMPAÑANTE?

- Padre..... 1
- Madre..... 2
- Otro familiar..... 3

Otro..... 4
(Específica)

10. EL LUGAR DONDE LE INDICARON QUE TIENE QUE COMPRAR MEDICAMENTOS, ¿ES:

- En la farmacia de un establecimiento público?..... 1
- En un establecimiento privado (Farmacia, botica)?..... 2
- Otro..... 3
(Específica)

EVALUADOR TÉCNICO:

Solo preguntar por aquellos servicios que recibió o ha sido atendido.

4. ¿CUÁNTO TIEMPO TRANSCURRIÓ, DESDE QUE INGRESÓ HASTA QUE SALIÓ DEL CONSULTORIO MÉDICO?

Añote la respuesta en el recuadro →
(En minutos)

NO SABE / NO RECUERDA..... 1

11. USTED, ¿PAGÓ DE SU BOLSILLO POR:

MOSTRAR TARJETA Nº 1

	Todo	Una parte	Nada	No corresponde
1 Consulta médica?	3	2	1	0
2 Medicamentos?	3	2	1	0
3 Análisis laboratorio?	3	2	1	0
4 Rayos X- Ecografía?	3	2	1	0
5 Procedimientos especiales (endoscopia, tomografía, electrocardiograma)?	3	2	1	0
6 Otro?..... (Específica)	3	2	1	0

5. EL TIEMPO QUE DURÓ LA ATENCIÓN MÉDICA, ¿LE PARECIÓ ADECUADO?

- Sí..... 1
- No..... 2

6. EN ESTA ATENCIÓN MÉDICA RECIBIDA, ¿LE HAN INDICADO: (Acepte una o más alternativas)

- Consulta médica?..... 1
- Medicamentos?..... 2
- Análisis laboratorio?..... 3
- Rayos X/ Ecografía?..... 4
- Procedimientos especiales (endoscopia, electrocardiograma, tomografía)?..... 5

Otro?..... 6
(Específica)

12. ANTES DE LA ATENCIÓN MÉDICA, ¿UD. HA TENIDO EL DÍA DE HOY CONTACTO CON ALGUNA PERSONA DE:

(Acepte una o más alternativas)

- Portería, seguridad?..... 1
- Recepción, admisión?..... 2
- Informes?..... 3
- Caja?..... 4
- Otro personal administrativo?..... 5

NINGUNO..... 6 → **FASE A 14**
(Específica)

7. RESPECTO A LA RECETA BRINDADA, ¿EL MÉDICO LE INDICÓ QUE LOS MEDICAMENTOS LOS ADQUIRIERA FUERA DE ESTE ESTABLECIMIENTO DE SALUD, EN FORMA:

- Total?..... 1
- Parcial?..... 2
- NO LE INDICÓ..... 3
- NO APLICA (No hubo receta)..... 4 → **FASE A 11**

13. ¿COMO CALIFICARÍA UD. LA ATENCIÓN RECIBIDA POR EL PERSONAL ADMINISTRATIVO, EN CUANTO A:

MOSTRAR TARJETA Nº 2

	MB	S	R	NI	MM	NSMR
1 Amabilidad y cordialidad?	5	4	3	2	1	0
2 Respeto?	5	4	3	2	1	0
3 Intención / disposición por atender?	5	4	3	2	1	0
4 Confianza y seguridad que le inspira?	5	4	3	2	1	0
5 Vestuario (Uniforme)?	5	4	3	2	1	0
6 Claridad de la información?	5	4	3	2	1	0

14. ANTES DE LA ATENCIÓN MÉDICA, ¿UD. HA SIDO ATENDIDO O TENIDO CONTACTO CON ALGUN(A):
(Acepte una o más alternativas)

Enfermera?..... 1
Técnico/a en enfermería?..... 2
Técnico/a de laboratorio?..... 3
Farmacéutico/a o técnico/a de farmacia?..... 4
Otro personal asistencial?..... 5
(Especifique)

NINGUNO..... 6 → **FASE A 18**

15. ¿CÓMO CALIFICARÍA UD. LA ATENCIÓN RECIBIDA POR EL PERSONAL NO MÉDICO, EN CUANTO A:

MOstrar TARJETA Nº 2

	MB	B	N	IN	MM	NS/NR
1 Amabilidad y cordialidad?	5	4	3	2	1	0
2 Respeto?	5	4	3	2	1	0
3 Intera / disposición por atender?	5	4	3	2	1	0
4 Confianza y seguridad que le inspira?	5	4	3	2	1	0
5 Vestuario (Uniforme)?	5	4	3	2	1	0
6 Claridad de la información?	5	4	3	2	1	0

16. ¿CÓMO CALIFICARÍA UD. LA ATENCIÓN RECIBIDA EL DÍA DE HOY DEL MEDICO TRATANTE, EN CUANTO A:

MOstrar TARJETA Nº 2

	MB	B	N	IN	MM	NS/NR
1 Amabilidad y cordialidad?	5	4	3	2	1	0
2 Respeto?	5	4	3	2	1	0
3 Intera / disposición por atender?	5	4	3	2	1	0
4 Confianza y seguridad que le inspira?	5	4	3	2	1	0
5 Vestuario (Uniforme)?	5	4	3	2	1	0
6 Claridad de la información?	5	4	3	2	1	0

17. ¿SABE UD. CUÁL FUE EL DIAGNÓSTICO QUE LE DIÓ EL MEDICO QUE LE ATENDIÓ?

Sí..... 1
No..... 2
No le dio ningún diagnóstico..... 3
No hubo diagnóstico (control, resultados de análisis, etc.)... 4

18. ¿EL MÉDICO LE EXPLICÓ A UD. SOBRE LA ENFERMEDAD O PROBLEMA DE SALUD QUE SUFRE?

Sí..... 1
No..... 2 → **FASE A 20**

19. ¿UD. ENTENDIÓ LO QUE EL MÉDICO LE EXPLICÓ SOBRE LA ENFERMEDAD O PROBLEMA DE SALUD QUE PADECE?

Sí..... 1
No..... 2
NO SABE / NO RESPONDE..... 3

20. ¿EL MÉDICO LE EXPLICÓ A UD. CON PALABRAS SENCILLAS SOBRE:

	SI	NO	NS/NR
1 El tratamiento que recibió?	1	2	3
2 El tipo de medicamento?	1	2	3
3 La forma de administración?	1	2	3
4 La dosis de los mismos?	1	2	3
5 Los efectos adversos del medicamento?	1	2	3

21. EN LA CONSULTA RECIBIDA, ¿LE HAN INDICADO QUE:

No requiere más atenciones?..... 1
Requiere atenciones adicionales en este mismo establecimiento?..... 2
Requiere transferencia a un establecimiento de mayor complejidad del MINSA / ESSALUD?..... 3
Requiere atención en un establecimiento particular?..... 4
NO LE DIERON NINGUNA INDICACIÓN..... 5

22. EN LA ATENCIÓN RECIBIDA, ¿LE HAN INDICADO QUE TIENE QUE HACERSE EXÁMENES DE LABORATORIO EN OTRO ESTABLECIMIENTO?

Sí..... 1
No..... 2
NO APLICA..... 3 } → **FASE A 24**

23. ¿EL LUGAR DÓNDE LE INDICARON QUE TIENE QUE HACERSE EXÁMENES DE LABORATORIO ES:

En un establecimiento público de mayor complejidad?..... 1
En un establecimiento privado (clínica, policlínica, laboratorio o consultorio particular)?..... 2
Otro?..... 3
(Especifique)

24. EN LA ATENCIÓN RECIBIDA, ¿LE HAN INDICADO QUE TIENE QUE HACERSE EXÁMENES DE IMÁGENES (RADIOGRAFIAS, ECOGRAFIAS, ETC.), EN OTRO ESTABLECIMIENTO?

Sí..... 1
No..... 2
NO APLICA..... 3 } → **FASE A 22**

25. ¿EL LUGAR DÓNDE LE INDICARON QUE TIENE QUE HACERSE EXÁMENES DE IMÁGENES ES:

En un establecimiento público de mayor complejidad?..... 1
En un establecimiento privado (clínica, policlínica, laboratorio o consultorio particular)?..... 2
Otro?..... 3
(Especifique)

VI. SATISFACCIÓN CON EL SERVICIO

26. ¿CÓMO CALIFICARÍA UD. A ESTE ESTABLECIMIENTO EN CUANTO A LA:

MOSTRAR TARIETA Nº 2

	MB	B	N	M	DM	MBNR
1 Ubicación – localización?	5	4	3	2	1	0
2 Comodidad de los ambientes?	5	4	3	2	1	0
3 Accesibilidad a los ambientes?	5	4	3	2	1	0
4 Asientos o área de espera?	5	4	3	2	1	0
5 Señalización de ambientes?	5	4	3	2	1	0
6 Limpieza – higiene?	5	4	3	2	1	0
7 Privacidad de la atención?	5	4	3	2	1	0

27. RESPECTO AL SERVICIO RECIBIDO EL DÍA DE HOY EN ESTE ESTABLECIMIENTO, ¿COMO CALIFICARÍA SU NIVEL DE SATISFACCIÓN?

MOSTRAR TARIETA Nº 3

- Muy satisfecho/a.....1
- Satisfecho/a.....2
- Ni satisfecho/a / ni insatisfecho/a.....3
- Insatisfecho/a.....4
- Muy insatisfecho/a.....5

28. CONSIDERA UD. QUE EN GENERAL EN ESTE ESTABLECIMIENTO, ¿LOS MÉDICOS CUMPLEN CON SU HORARIO DE TRABAJO:

- Siempre?.....1
- Casi siempre?.....2
- Casi nunca?.....3
- Nunca?.....4
- NO SABE / NO RESPONDE.....5

29. DE LOS SERVICIOS RECIBIDOS EN ESTE ESTABLECIMIENTO, ¿PRINCIPALMENTE DEBE MEJORAR EN:

- Farmacia?.....1
 - Olas o admisión?.....2
 - Imágenes (Rayos X- Ecografía, etc.?.....3
 - Laboratorio?.....4
 - Consultorio médico?.....5
 - Caja?.....6
 - Oficina de seguros SIS (MNSA)?.....7
 - Otro?.....8
- (Especifique)
- NO SABE / NO RESPONDE.....9

30. BASADO EN SU ATENCIÓN ACTUAL, ¿PRINCIPALMENTE QUE LE RECOMENDARÍA AL DIRECTOR DE ESTE ESTABLECIMIENTO?

MOSTRAR TARIETA Nº 4

- Que mejore el trato del personal de salud en la atención.....1
 - Que mejore el trato del personal administrativo en la atención.....2
 - Que no haya mucho trámite administrativo.....3
 - Que mejoren el área de espera.....4
 - Que existan más médicos especialistas.....5
 - Que las farmacias tengan o dispongan de medicamentos.....6
 - Otro.....7
- (Especifique)

31. SI ALGÚN FAMILIAR O AMIGO PRESENTARA UNA ENFERMEDAD O DOLENCIA, ¿LE RECOMENDARÍA ATENDERSE EN ESTE ESTABLECIMIENTO DE SALUD:

- Sin lugar a dudas?.....1
- Tal vez?.....2
- Lo pensaría dos veces?.....3
- De ninguna manera?.....4

VII. SEGUROS DE SALUD

32. ¿TIENE O ESTÁ AFILIADO A ALGÚN SEGURO DE SALUD?

- Sí.....1
- No.....2 → **PASE A 36**

33. ¿QUÉ SEGUROS DE SALUD TIENE UD.? (Acepte una o más alternativas)

- Seguro Integral de Salud (SIS).....1
 - EsSalud – Seguro Social (ex-IPSS).....2
 - Seguro de la Entidad Prestadora de Salud (EPS).....3
 - Seguro de Salud de Aseguradoras privadas.....4
 - Seguro de salud de las Clínicas.....5
 - Seguro de salud Universitario.....6
 - Sanidad de las Fuerzas armadas y policiales.....7
 - Otro.....8
- (Especifique)

SI CIRCULÓ UNA SOLA ALTERNATIVA PASE A 36

EVALUADOR TÉCNICO: Sólo preguntar por aquellos que tienen más de un seguro, de lo contrario pase a la pregunta 36.

34. SI TIENE MÁS DE UN SEGURO, ¿CUAL UTILIZA CON MAYOR FRECUENCIA?

- Seguro Integral de Salud (SIS).....1
 - EsSalud – Seguro Social (ex-IPSS).....2
 - Seguro de la Entidad Prestadora de Salud (EPS).....3
 - Seguro de Salud de Aseguradoras privadas.....4
 - Seguro de salud de las Clínicas.....5
 - Seguro de salud Universitario.....6
 - Sanidad de las Fuerzas armadas y policiales.....7
 - Otro.....8
- (Especifique)
- NUNCA LO HA UTILIZADO.....9 → **PASE A 36**

EVALUADOR TÉCNICO: Solo preguntar por el seguro que utiliza con mayor frecuencia mencionado en la pregunta 33 o 34.	
35. EN CUÁNTO A LOS SERVICIOS QUE LE BRINDA SU SEGURO DE..... ¿CÓMO CALIFICARÍA SU NIVEL DE SATISFACCIÓN?	
MOSTRAR TARJETA N° 3	
Muy satisfecho/a.....	1
Satisfecho/a.....	2
Ni satisfecho/a / ni insatisfecho/a.....	3
Insatisfecho/a.....	4
Muy insatisfecho/a.....	5
36. LA CONSULTA MÉDICA, ¿HA SIDO PAGADA O CUBIERTA POR:	
Seguro Integral de Salud (SIS)?.....	1
EsSalud – Seguro Social (ex-IPSS)?.....	2
Seguro de la Entidad Prestadora de Salud (EPS).....	3
Seguro de Salud de Aseguradoras privadas?.....	4
Seguro de salud de las Clínicas?.....	5
Seguro de salud universitario?.....	6
Salud de las Fuerzas armadas y policiales?.....	7
Usted lo pagó totalmente de su bolsillo?.....	8
Otro?.....	9
(Especifique)	

VII. ACCESIBILIDAD AL ESTABLECIMIENTO DE SALUD Y OPORTUNIDAD DE LA ATENCIÓN	
37. ¿CÓMO OBTUVO LA CITA PARA ESTA CONSULTA MÉDICA?	
Por teléfono.....	1
Por internet.....	2
Se le dieron en la consulta anterior.....	3
Tuvo que acercarse al establecimiento para obtenerla.....	4
Otro.....	5
(Especifique)	
NO REQUIERE CITA, ATIENDEN POR ORDEN DE LLEGADA.....	6 → FASE A.41
38. ¿CUÁNTAS VECES TUVO QUE INSISTIR PARA QUE LE DIERAN LA CITA?	
Añote la respuesta en el recuadro <input type="text"/>	
(N° veces)	
NO SABE / NO RECUERDA.....	1 → FASE A.41
39. ¿LA CITA PARA ESTA ATENCIÓN DE SALUD SE LA DIERON PARA EL MISMO DÍA QUE UD. SOLICITÓ?	
Sí.....	1 → FASE A.41
No.....	2
40. ¿DESPUÉS DE CUÁNTOS DÍAS QUE UD. SOLICITÓ LA CITA, SE LA DIERON?	
Añote la respuesta en el recuadro <input type="text"/>	
(N° de días)	
41. ¿CUÁNTO TIEMPO DEMORÓ EN LLEGAR DESDE SU VIVIENDA AL ESTABLECIMIENTO DE SALUD?	
Añote la respuesta en el recuadro <input type="text"/>	
(En minutos)	
42. DESDE QUE INGRESÓ AL ESTABLECIMIENTO, ¿CUÁNTO TIEMPO DEMORÓ PARA QUE LO ATENDIERAN EN LA CONSULTA?	
Añote la respuesta en el recuadro <input type="text"/>	
(En minutos)	
43. EL TIEMPO QUE ESPERÓ UD. DESDE QUE LLEGÓ AL ESTABLECIMIENTO HASTA QUE LO ATENDIERON, ¿LE PARECIÓ:	
Adecuado?.....	1
Largo?.....	2
Muy largo?.....	3

IX. RECLAMOS	
44. EN LOS ÚLTIMOS 3 MESES, ¿UD. HA TENIDO ALGÚN PROBLEMA CON SU ATENCIÓN DE SALUD EN ESTE ESTABLECIMIENTO?	
Sí.....	1
NO / NO APLICA.....	2 → FASE A.48
45. ¿QUÉ PROBLEMA TUVO? (Acepte una o más alternativas)	
Maltrato en la atención.....	1
El seguro no cubre todas las especialidades.....	2
Falta de medicamentos.....	3
Incumplimiento en la programación de citas, cirugías u otros.....	4
Demora en la atención.....	5
Médico o personal no acudió a su turno.....	6
Problemas para su atención como asegurado.....	7
Le realizaron algún cobro no advertido.....	8
No se respetó mi privacidad.....	9
Otro.....	10
(Especifique)	
46. CON RESPECTO AL PROBLEMA QUE TUVO, ¿PRESENTÓ UD. ALGÚN RECLAMO?	
Sí.....	1
No.....	2 → FASE A.48
47. ¿CREE UD. QUE LE SOLUCIONARON EL PROBLEMA QUE MOTIVÓ SU RECLAMO:	
Totalmente?.....	1
Parcialmente?.....	2
En nada?.....	3
48. DE PRESENTARSE ACTUALMENTE ALGÚN PROBLEMA EN LA ATENCIÓN, ¿DONDE PRESENTARÍA SU RECLAMO O QUEJA?	
En la dirección del establecimiento de salud.....	1
En la defensoría del asegurado.....	2
En la defensoría de la salud.....	3
En INDECOP.....	4
En la defensoría del pueblo.....	5
En SUNASA.....	6
En las oficinas del seguro.....	7
En la Sede Central del SIS / UDR SIS.....	8
Otro.....	9
(Especifique)	
NO SABE / NO RESPONDE.....	10

XII. CARACTERIZACIÓN DEL ENTREVISTADO

55. ¿QUÉ EDAD TIENE EN AÑOS CUMPLIDOS Y CUÁL ES SU FECHA DE NACIMIENTO?

Edad	Día	Mes	Año

56. SEXO

- Hombre.....1
 Mujer.....2

57. ¿EN QUÉ DEPARTAMENTO, PROVINCIA Y DISTRITO VIVE UD.?

Departamento _____

Provincia _____

Distrito _____

58. ¿CUÁL ES EL NIVEL EDUCATIVO QUE APROBÓ?

- Sin nivel.....1
 Inicial.....2
 Primaria incompleta.....3
 Primaria completa.....4
 Secundaria incompleta.....5
 Secundaria completa.....6
 Superior no universitaria incompleta.....7
 Superior no universitaria completa.....8
 Superior universitaria incompleta.....9
 Superior universitaria completa.....10
 Postgrado.....11

59. ¿CUÁL ES EL IDIOMA CON EL QUE SE COMUNICAN EN EL HOGAR?

- Español.....1
 Quechua.....2
 Otro _____ 3
 (Especifique)

60. ¿CUÁNTAS PERSONAS VIVEN PERMANENTEMENTE EN SU HOGAR?

Anote la respuesta en el recuadro

(N° personas)

61. SIN CONTAR BAÑO, COCINA, PASADIZOS NI GARAJE, ¿CUÁNTAS HABITACIONES EN TOTAL TIENE LA VIVIENDA?

Anote la respuesta en el recuadro

(N° habitaciones)

62. ¿CUÁNTAS HABITACIONES SE USAN PARA DORMIR?

Anote la respuesta en el recuadro

(N° habitaciones)

XIII. CARACTERÍSTICAS DE LA VIVIENDA Y SOCIO ECONÓMICAS

63. ACTUALMENTE, ¿EL TIPO DE VIVIENDA QUE OCUPA ES:

- Casa independiente?.....1
 Departamento en edificio?.....2
 Vivienda en quinta?.....3
 Vivienda en casa de vecindad (callejón o solar)?.....4
 Chaca o cabaña?.....5
 Vivienda improvisada?.....6
 Local no destinado para habitación humana?.....7
 Otro? _____ 8
 (Especifique)

65. EL MATERIAL PREDOMINANTE EN LOS TECHOS ES:

- ¿Concreto armado?.....1
 ¿Madera?.....2
 ¿Tejas?.....3
 ¿Planchas de calamina, fibra de cemento o similares?.....4
 ¿Caña o estera con torta de barro?.....5
 ¿Esteras?.....6
 ¿Paja, hojas de palmera, etc.?.....7
 Otro? _____ 8
 (Especifique)

64. EL MATERIAL PREDOMINANTE EN LAS PAREDES EXTERIORES ES:

- ¿Ladrillo o bloque de cemento?.....1
 ¿Piedra o sillar con cal o cemento?.....2
 ¿Adobe?.....3
 ¿Tapia?.....4
 ¿Quincha (caña con barro)?.....5
 ¿Piedra con barro?.....6
 ¿Madera?.....7
 ¿Esteras?.....8
 Otro? _____ 9
 (Especifique)

66. EL MATERIAL PREDOMINANTE EN LOS PISOS ES:

- ¿Parquet o madera pulida?.....1
 ¿Láminas asfálticas, vitílicas o similares?.....2
 ¿Losetas, terrazos o similares?.....3
 ¿Madera (entablados)?.....4
 ¿Cemento?.....5
 ¿Tierra?.....6
 Otro? _____ 7
 (Especifique)

<p>67. ¿CUÁL ES EL TIPO DE ALUMBRADO QUE TIENE SU HOGAR: (Acepte una o más alternativas)</p> <p>Electricidad?.....1 Kerosene (medero/lamparín)?.....2 Petróleo / gas (lámpara)?.....3 Vela?.....4 Generador?.....5 Otro?.....6 (Especifique) NO UTILIZA.....7</p>	<p>71. LA SEMANA PASADA, DEL _____ AL _____, EL JEFE DE HOGAR SE DESEMPEÑO EN SU OCUPACIÓN PRINCIPAL O NEGOCIO COMO:</p> <p>¿Empleado/a o patrono/a?.....1 ¿Trabajador/a independiente?.....2 ¿Empleado/a?.....3 ¿Obrero/a?.....4 ¿Trabajador/a Familiar No Remunerado/a.....5 ¿Trabajador/a del hogar?.....6 Otro?.....7 (Especifique) NO TRABAJÓ.....8</p>
<p>68. EL ABASTECIMIENTO DE AGUA EN SU HOGAR PROCEDE DE:</p> <p>¿Red pública, dentro de la vivienda?.....1 ¿Red pública, fuera de la vivienda pero dentro del edificio?.....2 ¿Red de uso público?.....3 ¿Carrión - sistema u otro similar?.....4 ¿Pozo?.....5 ¿Río, acequia, manantial o similar?.....6 Otro?.....7 (Especifique)</p>	<p>72. APROXIMADAMENTE, ¿CUÁNTO ES EL INGRESO FAMILIAR MENSUAL?</p> <p>Añote la respuesta en el recuadro → <input type="text"/> (En dólares) NO SABE / NO RECUERDA.....1</p>
<p>69. EL BAÑO O SERVICIO HIGIÉNICO QUE TIENE SU HOGAR ESTA CONECTADO A:</p> <p>¿Red pública de desagüe dentro de la vivienda?.....1 ¿Red pública de desagüe fuera de la vivienda pero dentro del edificio?.....2 ¿Letrina?.....3 ¿Pozo séptico?.....4 ¿Pozo negro o negro?.....5 ¿Río, acequia o canal?.....6 Otro?.....7 (Especifique) NO TIENE.....8</p>	<p>73. ¿CUÁL ES EL NIVEL EDUCATIVO APROBADO DEL JEFE DE HOGAR?</p> <p>Sin nivel.....1 Inicial.....2 Primaria incompleta.....3 Primaria completa.....4 Secundaria incompleta.....5 Secundaria completa.....6 Superior no universitaria incompleta.....7 Superior no universitaria completa.....8 Superior universitaria incompleta.....9 Superior universitaria completa.....10 Postgrado.....11 NO SABE / NO RECUERDA.....12</p>
<p>70. ¿SU HOGAR TIENE: (Acepte una o más alternativas)</p> <p>Equipo de sonido?.....1 Tv a color?.....2 Dvd?.....3 Licuadora?.....4 Refrigerador/congelador?.....5 Cocina a gas?.....6 Teléfono fijo?.....7 Plancha eléctrica?.....8 Lavadora?.....9 Computadora?.....10 Horno microondas?.....11 Internet?.....12 Tv Cable(Movistar TV, Claro TV, DirecTV y otro)?.....13 Celular con servicio de Internet?.....14 Celular sin servicio de Internet?.....15 NO TIENE NINGUNO.....16</p>	<p>74. ¿CUÁNTAS PERSONAS DEPENDEN ECONÓMICAMENTE DEL JEFE DE HOGAR?</p> <p>Añote la respuesta en el recuadro → <input type="text"/> (Nº Personas) (Si no hay dependientes anote "0")</p>
	<p>75. EN SU HOGAR, ¿HAY NIÑOS/AS DE 6 A 12 AÑOS?</p> <p>Sí.....1 No.....2 → CONCLUYA LA ENTREVISTA</p>
	<p>76. ACTUALMENTE, ¿ALGUNOS DE ESTOS NIÑOS/AS NO ASISTEN AL COLEGIO?</p> <p>Sí.....1 No.....2</p>

ANEXO 2: TABLA DE PROMEDIO DE TIEMPO DESPUES DE APLICAR LA SOLUCION PROPUESTA.

Doctor	Minsa	Essalud	Clinica
Ana Medrano	10	13	15
Osmar Pillaca	11	11	14
Promedio	11	12	15

Tabla 14: Promedio de tiempo después de aplicar la solución