TESIS
UNIVERSIDAD CIENTÍFICA DEL SUR

UNIVERSIDAD CIENTÍFICA DEL SUR

FACULTAD DE CIENCIAS EMPRESARIALES

CARRERA PROFESIONAL DE INGENIERÍA EMPRESARIAL Y DE SISTEMAS

"MACHINE LEARNING Y EL RIESGO DE ACCIDENTE CEREBROVASCULAR EN PACIENTES DEL CONSULTORIO DEL DR. PINTO"

Tesis para optar el título profesional de:

Ingeniero Empresarial y de Sistemas

Presentado por:

NATHALIA ELENA CONTRERAS SIU (0000-0002-7109-9189)

Asesor:

CARLOS SALVADOR PERALTA DELGADO (0000-0003-2301-503X)

Lima – Perú

2021

Tesis publicada con autorización del autor
No olvide citar esta tesis

TESIS
UNIVERSIDAD CIENTÍFICA DEL SUR

UNIVERSIDAD CIENTÍFICA DEL SUR

ACTA DE SUSTENTACIÓN DE TESIS

Lima, 04 de febrero de 2021

Los integrantes del Jurado de tesis:

Presidente	Dra. Sandra Jeannet Meza Balvin
Miembro 1	Mg. Jean Merry Salas Rojas
Miembro 2	Mg. Wilver Auccahuasi Aiquipa

Se reúnen para evaluar la tesis titulada:
"MACHINE LEARNING Y EL RIESGO DE ACCIDENTE CEREBROVASCULAR EN PACIENTES DEL CONSULTORIO DEL DR. PINTO"

Presentado por el(la) bachiller.
NATHALIA ELENA CONTRERAS SIU

Para optar el Título Profesional de
INGENIERO DE SISTEMAS EMPRESARIALES

Asesorado(a) por:
CARLOS SALVADOR PERALTA DELGADO

Luego de haber evaluado el informe final de tesis y evaluado el desempeño de(l) (los) estudiantes de la carrera de **Ingeniería de Sistemas Empresariales** en la sustentación, concluyen de manera unánime (X) por mayoría simple ( ) calificar a:

Tesista:	**NATHALIA ELENA CONTRERAS SIU**		
Nota (en letras):	Quince		
Aprobado (x)	Aprobado - Muy buena ( )	Aprobado - Sobresaliente ( )	Desaprobado ( )

Los miembros del jurado firman en señal de conformidad.

SANDRA JEANNET MEZA BALVIN
Presidente(a) del Jurado

CARLOS SALVADOR PERALTA DELGADO
Asesor(a)

JEAN MERRY SALAS ROJAS
Miembro 1

WILVER AUCCAHUASI AIQUIPA
Miembro 2

T.//(51) 60 6738
informes@cientificaud.pe
cientificaud.pe

Tesis publicada con autorización del autor
No olvide citar esta tesis

TESIS
UNIVERSIDAD CIENTÍFICA DEL SUR

Dedicatoria

A mi madre, mi padre y mi abuela, quienes me ha apoyado en todo momento, han sido mis guías de vida y me han enseñado a dar siempre mi mayor esfuerzo, que el trabajo duro tiene su recompensa y que siempre hay una solución aunque parezca lo contrario.

Tesis publicada con autorización del autor
No olvide citar esta tesis

TESIS
UNIVERSIDAD CIENTÍFICA DEL SUR

Agradecimientos

Al Ing. Carlos Peralta Delgado, por su asesoría, tiempo y conocimientos durante esta investigación.

Al Dr. William Pinto, Dr. Juan Trujillo, y a la Lic. Mirian Huamán por sus conocimientos y su apoyo en el presente trabajo.

Tesis publicada con autorización del autor
No olvide citar esta tesis

TESIS
UNIVERSIDAD CIENTÍFICA DEL SUR

ÍNDICE GENERAL

Dedicatoria ii

Agradecimientos iii

ÍNDICE GENERAL iv

ÍNDICE DE FIGURAS vi

ÍNDICE DE TABLAS viii

ÍNDICE DE ANEXOS ix

RESUMEN x

ABSTRACT xi

INTRODUCCIÓN 1

CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA 3

1.1. Descripción de la realidad problemática 3

1.2. Formulación del problema 8

1.2.1. Problema General 8

1.2.2. Problemas Específicos 8

1.3. Justificación de la Investigación 9

1.4. Limitaciones de la Investigación 10

1.5. Viabilidad de la Investigación 10

CAPÍTULO II: MARCO TEÓRICO 11

2.1. Antecedentes de la Investigación 11

2.1.1. Antecedentes Nacionales 11

2.1.2. Antecedentes Internacionales 13

2.2. Bases Teóricas 17

2.2.1. Machine Learning 17

2.2.2. Riesgo de Accidente Cerebrovascular 52

Tesis publicada con autorización del autor
No olvide citar esta tesis

2.3. Objetivos de la Investigación 77
2.3.1. Objetivo General 77
2.3.2. Objetivos Específicos 77
2.4. Formulación de Hipótesis 78
2.4.1. Hipótesis General 78
2.4.2. Hipótesis Específicas 78
CAPÍTULO III: DISEÑO METODOLÓGICO 79
3.1. Diseño de la Investigación 79
3.2. Tipo 79
3.3. Enfoque 80
3.4. Población 80
3.5. Muestra 80
3.6. Operacionalización de Variables 82
3.7. Técnicas para la Recolección de Datos 82
3.7.1. Descripción de los Instrumentos 82
3.7.2. Validez y Confiabilidad de los Instrumentos 84
3.8. Técnica para el Procesamiento y Análisis de Datos 84
3.9. Aspectos Éticos 85
CAPÍTULO IV: RESULTADOS 86
CAPÍTULO V: DISCUSIÓN, CONCLUSIONES Y RECOMENDACIONES 111
REFERENCIAS BIBLIOGRÁFICAS 117
GLOSARIO DE TÉRMINOS Y ABREVIATURAS 123
ANEXOS 125

Tesis publicada con autorización del autor
No olvide citar esta tesis

TESIS
UNIVERSIDAD CIENTÍFICA DEL SUR

ÍNDICE DE FIGURAS

Figura 1. Funcionamiento de un Algoritmo Tradicional..... 19

Figura 2. Funcionamiento de un Algoritmo de Machine Learning 20

Figura 3. Clasificación de Algoritmos de Machine Learning 22

Figura 4. Algoritmo de Agrupamiento (Clustering)..... 26

Figura 5. Aplicación Web desde Multidispositivos 31

Figura 6. Funcionalidad de una Aplicación Web 34

Figura 7. Proceso para Desarrollo de Aplicación Web 36

Figura 8. Matriz de Confusión 39

Figura 9. Fórmula para Métrica de Exactitud..... 40

Figura 10. Fórmula para Métrica de Sensibilidad 41

Figura 11. Fórmula para Métrica de Especificidad..... 42

Figura 12. Metodología Fundamental para Ciencia de Datos..... 46

Figura 13. Categorías de la Presión Arterial..... 56

Figura 14. Accidente Cerebrovascular Isquémico..... 68

Figura 15. Accidente Cerebrovascular Hemorrágico..... 69

Figura 16. Técnica para la Muestra 80

Figura 17. Fórmula para Muestra de Población Finita..... 81

Figura 18. Estructura del Dataset Recopilado..... 93

Figura 19. Histograma por cada Variable Numérica 94

Figura 20. Gráfico de Barras por cada Variable Categórica 96

Figura 21. Carga de Datos 97

Figura 22. División de los Datos 97

Figura 23. Importación de Librerías necesarias 98

Tesis publicada con autorización del autor
No olvide citar esta tesis

Figura 24. Conversión de Variables categóricas a numéricas 98

Figura 25. Parámetros para el Algoritmo 98

Figura 26. Conversión de variables y Creación del Pipeline 99

Figura 27. Entrenamiento del Modelo de Random Forest 99

Figura 28. Matriz de Confusión usando Test Dataset 100

Figura 29. Métricas de Evaluación usando Test Dataset 100

Figura 30. Implementación de la Solución 101

Figura 31. Matriz de Confusión usando Data de Validación 103

Figura 32. Métricas de Desempeño para Validación 103

Figura 33. Histograma por Variable Numérica de la Muestra 104

Figura 34. Gráfico de Barras por Variable Categórica de la Muestra 106

Figura 35. Resultado de la Prueba de Chi-cuadrado 108

Tesis publicada con autorización del autor
No olvide citar esta tesis

TESIS
UNIVERSIDAD CIENTÍFICA DEL SUR

ÍNDICE DE TABLAS

Tabla 1. *Examen de Glucosa en Sangre en Ayunas* 57

Tabla 2. *Análisis de Colesterol Total en Sangre* 59

Tabla 3. *Análisis de Colesterol HDL en Sangre* 59

Tabla 4. *Análisis de Colesterol LDL en Sangre* 60

Tabla 5. *Niveles del Índice de Masa Corporal (IMC)* 61

Tabla 6. *Detalle de los datos* 92

Tabla 7. *Estadística Descriptiva de Variables Numéricas* 93

Tabla 8. *Estadística Descriptiva de Variables Categóricas* 95

Tabla 9. *Comparación: Real vs. Predicción* 102

Tabla 10. *Resumen Estadístico: Variables Numéricas de la Muestra* 104

Tabla 11. *Resumen Estadístico: Variables Categóricas de la Muestra* 106

Tabla 12. *Random Forest: Feature Importance* 110

Tabla 13. *Comparación de métricas entre modelos* 111

Tesis publicada con autorización del autor
No olvide citar esta tesis

TESIS
UNIVERSIDAD CIENTÍFICA DEL SUR

ÍNDICE DE ANEXOS

Anexo 1. Matriz de Consistencia 126

Anexo 2. Matriz de Operacionalización..... 127

Anexo 3. Instrumento de Evaluación 128

Anexo 4. Ficha de Validación del Instrumento por Jueces Expertos..... 129

Anexo 5. Autorización de la Institución donde se realizó la Investigación 135

Anexo 6. Ficha de Requerimientos Técnicos..... 136

Anexo 7.Tabla de Datos Recopilados para el proyecto..... 138

Anexo 8. Interfaces de la Aplicación Web..... 143

Anexo 9. Tabla de Resultados del Machine Learning..... 145

Anexo 10. Diagrama de Capas de la Solución Propuesta..... 148

Tesis publicada con autorización del autor
No olvide citar esta tesis

TESIS
UNIVERSIDAD CIENTÍFICA DEL SUR

RESUMEN

La presente investigación se basó en torno a la problemática en el consultorio médico del Dr. Pinto sobre la incidencia de accidentes cerebrovasculares. Por ello, este estudio se enfoca en el uso de machine learning para desarrollar e implementar una aplicación web que considere las principales condiciones de riesgo del paciente, y prediga esta enfermedad, sirviendo como referencia para el doctor.

Esta investigación tiene un diseño experimental y su objetivo es determinar el nivel de influencia para predecir el riesgo de accidente cerebrovascular en pacientes del consultorio en cuestión. Por ende, la solución que se desarrolló fue una aplicación web basada en un modelo de Random Forest que procesa la data, y predice accidentes cerebrovasculares. Todo esto mediante una primera interfaz en la cual se ingresan los principales riesgos del paciente tales como la edad, glucosa, hipertensión, entre otros. Mientras que en la segunda interfaz, el médico consulta los resultados procesados. Para el desarrollo de esta solución se aplicó la metodología de Data Science, considerando también los detalles de los principales requerimientos funcionales.

Los resultados indican que sí existe una relación entre machine learning y los accidentes cerebrovasculares, ya que tanto los diagnósticos predichos como los reales son iguales. Por ello, se concluye que la solución es de utilidad para el consultorio del Dr. Pinto.

Palabras claves: Machine Learning, Data Science, Accidente Cerebrovascular

Tesis publicada con autorización del autor
No olvide citar esta tesis

TESIS
UNIVERSIDAD CIENTÍFICA DEL SUR

ABSTRACT

The present investigation was based on the problem in Dr. Pinto's medical office about Stroke incidence. Therefore, this study focuses on the use of machine learning to develop and implement a web application that considers the main risk factors of the patient, and predicts this disease, serving as a reference for the doctor.

This investigation has an experimental design and its objective is to determine the level of influence in order to predict the risk of stroke in patients at the medical office in question. Hence the developed solution was a web application based on a Random Forest model that processes data, and predicts strokes. All of this through a first interface in which the main risk factors of the patient such as age, glucose, hypertension, to name but a few are registered. In the second interface, the doctor searches the patient's results. Regarding the development of this solution, Data Science methodology was applied, and the main functional requirements were considered.

The results indicate that machine learning and stroke are related, since both the predicted and the actual diagnoses are the same. Consequently, it is concluded that the proposed solution is useful for Dr. Pinto's medical office.

Keywords: Machine Learning, Data Science, Stroke

Tesis publicada con autorización del autor
No olvide citar esta tesis

TESIS
UNIVERSIDAD CIENTÍFICA DEL SUR

INTRODUCCIÓN

El riesgo de un accidente cerebrovascular es latente cuando los elementos de riesgo en un paciente no están controlados o monitoreados. Esto podría causar la interrupción del flujo sanguíneo, y por ende del oxígeno, al cerebro, produciendo un ACV. Por tanto, el énfasis en temas de prevención es fundamental para que profesionales de la salud apliquen los tratamientos correspondientes, y para que las tasas de incidencia y mortalidad disminuyan.

Por ello que el problema central de investigación es: ¿Cómo Machine Learning influye para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto?

Esta investigación muestra el potencial de una tecnología como Machine Learning en temas de salud, específicamente accidentes cerebrovasculares. Su adaptabilidad permite su implementación en una aplicación web sencilla y accesible desde cualquier dispositivo por el médico tratante.

La importancia de este tipo de soluciones en el sector salud permite que el médico tenga una referencia durante las consultas, con el fin de brindar tratamiento preventivo de la enfermedad, mediante el control de los elementos de riesgo detectados. De esta manera se mejora la calidad de vida del paciente, y se evitan secuelas y gastos adicionales que esta enfermedad trae consigo.

Tesis publicada con autorización del autor
No olvide citar esta tesis

El presente trabajo de investigación está estructurado de la siguiente manera:

- **Capítulo I:** En este capítulo se describe la situación problemática relacionada a machine learning y a los accidentes cerebrovasculares, desde el ámbito mundial hasta el centro médico de estudio. A partir de ello, se formula el problema, estableciendo las justificaciones, las limitaciones y la viabilidad.
- **Capitulo II:** En este capítulo se describen el marco teórico, que incluye los antecedentes internacionales y nacionales, las bases teóricas, los objetivos, y las hipótesis de la investigación.
- **Capitulo III:** En este capítulo se describen los detalles del diseño metodológico, la población, y la muestra. Así como también la Operacionalización de variables, técnicas de recolección y procesamiento de datos.
- **Capitulo IV:** En este capítulo se muestran los resultados de la investigación, los cuales consisten en un análisis del sector, aplicación de la metodología seleccionada, y la fase de comprobación.
- **Capítulo V:** En este capítulo se especifican las discusiones de los resultados, las conclusiones que se obtuvieron a partir de los datos de la implementación de la solución en el consultorio médico, y las recomendaciones basadas en la misma.

Asimismo, esta investigación cuenta con las siguientes partes:

- **Referencias bibliográficas:** Incluyen trabajos citados y referencia utilizadas.
- **Anexos:** Contiene la matriz de consistencia y Operacionalización, el instrumento de evaluación y su validación por jueces expertos, carta de autorización del centro médico, documentos técnicos, y tablas de resultados.

La Autora

Tesis publicada con autorización del autor
No olvide citar esta tesis

TESIS
UNIVERSIDAD CIENTÍFICA DEL SUR

CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA

1.1. Descripción de la realidad problemática

Machine Learning hace referencia a aprender de los datos, a través del desarrollo de un modelo, el cual luego se puede usar para hacer predicciones (Theodoridis, 2015).

Se denomina Accidente Cerebrovascular (ACV) al momento en que la circulación sanguínea hacia una zona del cerebro se ve interrumpida, produciendo que las células cerebrales mueran por la falta de oxígeno e irrigación (National Institute of Neurological Disorders and Stroke [NINDS], 2019).

A nivel mundial, la Organización Mundial de la Salud (OMS, 2016) señala que esta enfermedad es la segunda causa de decesos y la tercera causa de discapacidad. Asimismo, lo califica como una crisis que va en aumento y que no recibe la atención requerida. Se considera que la falta de medidas preventivas y la poca asociación de elementos de riesgo con un accidente cerebrovascular, tienen como consecuencia el aumento de los mismos principalmente en países de bajos-medianos ingresos.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Basado en lo anterior, a nivel continental, Estados Unidos, siendo un país desarrollado, debido a sus altos ingresos, y alta capacidad tecnológica, ha conseguido tener un mayor seguimiento de las incidencias de accidentes cerebrovasculares gracias a la implementación de la tecnología dentro y fuera de los centros de salud. Walker (2019) resalta que el uso de Machine Learning e inteligencia artificial podrían ayudar a reducir accidentes cerebrovasculares debido al análisis de datos de pacientes para identificar tendencias, y predecir. A pesar que Machine Learning recién está incursionando en el ámbito de la Salud, se considera que tiene potencial para hacer frente a situaciones en las cuales el tiempo es crucial.

A nivel nacional, Perú no es ajeno a la problemática de los accidentes cerebrovasculares, particularmente en la etapa de vida del adulto mayor, aunque cabe resaltar que las enfermedades no distinguen edades. El (Ministerio de Salud del Perú [MINSA], 2018) califica el accidente cerebrovascular como segundo causante de fallecimientos en el Perú, considerándolo entre las enfermedades de importancia nacional. Además, la institución señala que "(…) durante el año 2015 fallecieron 9499 personas a causa de la enfermedad cerebrovascular, produciendo una tasa específica de mortalidad de 30 muertes por 100 mil habitantes. En relación a la tasa del año 2003, se observa un aumento del 45%" (MINSA, 2018, p. 105).

Tesis publicada con autorización del autor
No olvide citar esta tesis

Con respecto al uso de Machine Learning en el Perú en el área de salud, aún no hay registro de entidades que hayan implementado dicha tecnología. Sin embargo, (Ministerio de Salud del Perú [MINSA], 2018) bajo Resolución Ministerial, publicó un proyecto relacionado al uso de tecnología en el ámbito de la Salud. Entre las mencionadas están Machine Learning, y Big Data, que buscan ser usadas para mejorar las decisiones médicas.

A nivel regional, entre las 3 regiones del Perú, la región Costa es la que reporta mayor mortalidad por accidentes cerebrovasculares, seguido por la región Sierra y la región Selva respectivamente. El MINSA (2018, p. 109) indica que la tasa de mortalidad en la Costa por esta enfermedad, entre los años 2003 al 2015, incrementó en 78%.

A nivel local, el distrito de Miraflores, según estadísticas del (Repositorio Único Nacional de Información en Salud [REUNIS], 2018), la tasa de mortalidad quinquenal, por accidente cerebrovascular, del periodo 2011 a 2015 fue de 105,78. Debido a que solo existen estadísticas hasta 2015, basándose en ellas, se pueden hacer proyecciones para el siguiente quinquenio. Por tanto, se proyecta que para el periodo 2016 a 2020, la tasa de mortalidad podría alcanzar 109,68.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Es necesario mencionar que los datos obtenidos del REUNIS del MINSA, son respecto a la tasa de mortalidad por accidente cerebrovascular. Se estima que estos representan solo la problemática en salud del sector público. Sin embargo, no hay datos sobre incidencias de la enfermedad, ya que no necesariamente una persona, que sufre un accidente cerebrovascular, muere. Existen casos en los que los pacientes, que han sido tratados con prontitud, llegan a recuperarse completamente en el transcurso de las semanas. Mientras que en otros casos, los pacientes necesitan sesiones de rehabilitación física y del lenguaje. Otros, sin embargo, adquieren una discapacidad permanente. Es por esto que los datos de la tasa de mortalidad únicamente ofrecen una visión parcial de las consecuencias de esta enfermedad.

La principal causa del riesgo de un accidente cerebrovascular en pacientes del Consultorio del Dr. Pinto, se debe a la falta de control de los elementos de riesgo, los cuales no suelen asociarse directamente a un accidente cerebrovascular, por lo que ciertos factores de riesgo como la hipertensión o diabetes se tratan como enfermedades individuales.

La consecuencia que esto siga sucediendo es la incidencia de un accidente cerebrovascular en sí, lo cual puede llevar a la muerte o a una discapacidad. Esto último afecta tanto la calidad de vida de la persona, como la salud emocional de la misma y de su entorno familiar.

Tesis publicada con autorización del autor
No olvide citar esta tesis

La solución que se plantea en esta investigación, con respecto a esta problemática es el uso de Machine Learning con el fin de encontrar patrones a partir de datos, y así poder aumentar el seguimiento sobre los aspectos médicos, y predecir el riesgo de tener un accidente cerebrovascular. Asimismo, que esto sirva como una herramienta complementaria en la decisión médica, para fomentar la prevención.

Es por ello que el problema central de investigación es: ¿Cómo Machine Learning influye para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto?

Tesis publicada con autorización del autor
No olvide citar esta tesis

1.2. Formulación del problema

1.2.1. Problema General

- ¿Cómo Machine Learning influye para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto?

1.2.2. Problemas Específicos

- ¿Cuáles son los requerimientos funcionales de Machine Learning para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto?
- ¿Cuál es el grado de exactitud de Machine Learning para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto?
- ¿Cuál es el grado de sensibilidad de Machine Learning para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto?
- ¿Cuál es el grado de exactitud de Machine Learning para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto?

Tesis publicada con autorización del autor
No olvide citar esta tesis

1.3. Justificación de la Investigación

Justificación Social

Esta investigación beneficiará a los pacientes, ya que al haber un mayor control sobre su propensión al accidente cerebrovascular, se reduciría este mediante el tratamiento preventivo que indique el médico. Evitando así, que dicha enfermedad ocurra, y por ende, previniendo las consecuencias que esto conlleva.

Justificación Práctica

La presente investigación permitirá controlar los elementos de riesgo de un accidente cerebrovascular, y predecir el riesgo del mismo, mediante el desarrollo de un machine learning y de una aplicación web, que estarán conectados. De esta manera, las predicciones complementarían las decisiones médicas para el tratamiento preventivo del paciente.

Justificación Económica

Cuando una persona sufre un accidente cerebrovascular y sobrevive, adquiere algún tipo de discapacidad, por lo cual tendrá que ir a rehabilitación y necesitar cuidados extras en su hogar. Todo esto resulta costoso y llega a exceder el presupuesto de la persona. Por tanto, el desarrollo de un machine learning con una aplicación web, permitirá controlar los elementos de riesgo de un accidente cerebrovascular, y predecir el riesgo del mismo, complementando las decisiones médicas para tratamiento preventivo, evitando las secuelas, y minimizando los gastos de salud de una persona.

Tesis publicada con autorización del autor
No olvide citar esta tesis

1.4. Limitaciones de la Investigación

En la presente investigación, la primera limitación es geográfica, porque se realiza en el consultorio médico del Dr. Pinto, ubicado en el distrito de Miraflores en Lima, Perú.

La segunda limitación que se tuvo fue en temas de movilidad, debido a la pandemia. Esto impactó el tiempo de la recolección de datos, el desarrollo, y la comprobación de la solución.

El estudio se limita al desarrollo de un machine learning y de una aplicación web (para facilitar la interacción del usuario con el programa). Esto tiene como propósito controlar a pacientes con propensión a un accidente cerebrovascular, predecirlo, y que sirva como guía para prevenirlo mediante un tratamiento del médico.

1.5. Viabilidad de la Investigación

El acceso a la información y la recopilación de la misma tuvo el total apoyo y autorización del consultorio médico, el cual permitió la revisión de las historias médicas.

Referente a la viabilidad económica de la presente, no existen grandes gastos, por lo que todo será cubierto por el propio investigador.

Tesis publicada con autorización del autor
No olvide citar esta tesis

CAPÍTULO II: MARCO TEÓRICO

2.1. Antecedentes de la Investigación

2.1.1. Antecedentes Nacionales

Ordóñez Barrios, y Vizcarra Infantes (2018), realizaron una investigación titulada "Modelo Predictivo para el diagnóstico de la Diabetes Mellitus Tipo 2 soportado por SAP Predictive Analytics", en la Facultad de Ingeniería de la Universidad Peruana de Ciencias Aplicadas (UPC), en Lima, Perú.

La investigación tuvo como objetivo la elaboración de un modelo predictivo para diagnosticar diabetes mellitus tipo 2.

Esta no realiza el desarrollo del modelo predictivo, sino que usa modelos disponibles dentro de SAP Predictive Analytics. Se realizaron 600 encuestas para recopilar datos para el modelo, y se dividieron en datos para aprendizaje y datos para pruebas. Después de dicho proceso, se determina que el algoritmo de Red Neuronal tuvo mayor precisión para el caso de estudio.

Se concluyó que el modelo diseñado puede ser implementado en centros médicos en el proceso de evaluación del paciente. Además, la viabilidad fue validada por un experto en medicina y telemedicina. También, se brindó una propuesta de implementación del proyecto.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Posadas Ruiz (2018), realizó la investigación titulada “Factores de riesgo asociados a accidente cerebrovascular en el Hospital Edgardo Rebagliati Martins, 2016”, en la Facultad de Medicina Humana de la Universidad Ricardo Palma (URP), en Lima, Perú.

La finalidad de dicho trabajo fue la identificación de factores de riesgo de un Accidente Cerebrovascular de tipo isquémico trombótico.

La investigación fue no experimental, de tipo analítico, de enfoque cuantitativo. La muestra estuvo dividida en: muestra de número de casos y muestra control, en cada uno se asignaron 76 pacientes hospitalizados en el área de Neurología del Hospital.

Como resultados se obtuvo que: la edad promedio de los casos fue 72.21, y que de 76 pacientes diagnosticados con la enfermedad, 39 eran hombres y 37 mujeres. Además, se analizó la ocurrencia de factores de riesgo como diabetes, hipertensión, tabaquismo, entre otros.

Se concluyó que las personas de 60 años a más tienen mayor probabilidad de tener un accidente cerebrovascular. Además, se demostró que la hipertensión y la dislipidemia tienen mayor relación significativa con la enfermedad. Mientras que la diabetes, y el tabaquismo también presentan una asociación significativa.

Tesis publicada con autorización del autor
No olvide citar esta tesis

2.1.2. Antecedentes Internacionales

Heo et al. (2019), realizaron una investigación titulada “Machine Learning–Based Model for Prediction of Outcomes in Acute Stroke”, en el Departamento de Neurología y el Departamento de Laboratorio Médico de Yonsei University College of Medicine, en Seúl, Corea del Sur.

La investigación se enfocó en la aplicabilidad de técnicas de Machine Learning para predecir accidentes cerebrovasculares de tipo isquémico a largo plazo.

El estudio fue retrospectivo con una cohorte prospectiva de individuos con diagnóstico confirmado de ACV isquémico. A los resultados favorables a los 3 meses, se les asignó una escala de 0, 1, o 2.

Los investigadores desarrollaron tres modelos de Machine Learning, y compararon sus predicciones: Red Neuronal, Random Forest, y Regresión Logística. Finalmente, para evaluar la precisión de dichos modelos, los resultados se compararon con la prueba de puntaje ASTRAL, por sus siglas en inglés.

Se consideró un total de 2604 pacientes, de los cuales 2043 tuvieron un resultado favorable. El resultado del modelo de Red Neuronal dio un valor significativamente mayor (0.888) que la prueba de puntaje ASTRAL (0.839), mientras que Random Forest y Regresión Logística dieron valores de 0.857 y 0.849 respectivamente.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Se concluyó que los algoritmos de Machine Learning, especialmente Red Neuronal, puede mejorar la predicción de resultados a largo plazo en pacientes con accidente cerebrovascular isquémico.

Jeena, y Kumar (2018), realizaron la investigación titulada “Machine Intelligence in Stroke Prediction”, en Kerala, India.

El objetivo de la investigación fue diseñar un sistema de predicción de accidentes cerebrovasculares, mediante el análisis de varios parámetros que serán usados como factores de riesgo.

Este estudio utiliza los algoritmos de clasificación llamados SVM (Máquina de Vectores de Soporte) y Red Neuronal. Se atribuye el uso de una base de datos de un tercero. Después del procesamiento, 350 casos fueron seleccionados, habiendo una consistencia entre el número de pacientes por sexo y por grupo de edad (desde los 35 a los 90 años). Se consideró que la fibrilación auricular y la presión sanguínea eran los mayores factores de riesgo.

La investigación concluyó que Machine Learning puede ser proyectado como una herramienta importante en la toma de decisiones médicas. Las predicciones hechas por el algoritmo de Red Neuronal tuvieron una precisión del 98.1%, mientras que el algoritmo SVM alcanzó el 91%.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Adam, Yousif, y Bashir (2016), realizaron una investigación titulada "Classification of Ischemic Stroke using Machine Learning Algorithms".

El objetivo de dicha investigación fue revisar y analizar los casos clasificados como accidente cerebrovascular isquémico, con el fin de clasificar sus tipos mediante el uso de Machine Learning.

Se desarrolló un modelo para clasificar los accidentes cerebrovasculares isquémicos usando el algoritmo de Árbol de Decisión y KNN. El modelo se basó en un conjunto de datos de 400 casos sobre esta enfermedad. Los datos fueron recolectados en diferentes hospitales y centros médicos en Sudán.

Se concluyó que el algoritmo de Árbol de Decisión tuvo un mejor desempeño que KNN, por lo que sus resultados pueden ser usados por especialistas médicos para clasificar y diagnosticar accidentes cerebrovasculares isquémicos en pacientes. Además, se descubrió que síntomas como la irritabilidad, convulsiones, y resultados de tomografías pueden determinar directamente el tipo de ACV isquémico. Finalmente, los investigadores encontraron que el ACV isquémico tipo Trombosis tiene mayor incidencia entre los casos estudiados en Sudán.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Asadi, Dowling, Yan, y Mitchell (2014) realizaron una investigación titulada "Machine Learning for Outcome Prediction of Acute Ischemic Stroke Post Intra-Arterial Therapy", en Melbourne Brain Centre, Departamento de Medicina, Royal Melbourne Hospital, Universidad de Melbourne, Victoria, Australia.

El principal objetivo fue el diseño y la comparación de diferentes métodos de Machine Learning para predecir el resultado de una intervención endovascular en un accidente cerebrovascular isquémico.

Se indica que fue un estudio retrospectivo utilizando una base de datos recopilada con 107 casos de accidente cerebrovascular isquémico tratado mediante intervención endovascular. Para el procedimiento se utilizaron estadísticas, y los algoritmos de Red Neuronal y SVM, para diseñar un machine learning capaz de clasificar el resultado en bueno o malo, basado en ciertos factores.

Como resultado se mostró que algoritmo de Red Neuronal obtuvo una congruencia del 80% entre el objetivo y el resultado. Sin embargo, después del proceso de optimización, el algoritmo SVM mostró una mejora en el desempeño, con un error cuadrático promedio de 2.064.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Se concluyó que a pesar de haber trabajado con una base de datos pequeña, existió una precisión de resultados cercana al 70%. De dicha forma, se demostró el potencial para incorporar más datos para mejorar aún más la precisión. Finalmente, se propuso un sistema robusto basado en Machine Learning con potencial para optimizar la selección de procesos endovasculares versus un tratamiento médico tradicional, en casos de accidentes cerebrovasculares isquémicos.

2.2. Bases Teóricas

2.2.1. Machine Learning

2.2.1.1. Concepto de Machine Learning

Machine Learning hace referencia a aprender de los datos, a través del desarrollo de un modelo, el cual luego se puede usar para hacer predicciones (Theodoridis, 2015).

Según Hurwitz y Kirsch (2018, p. 4) señalan que Machine Learning es una forma de Inteligencia Artificial que hace posible a un sistema aprender a partir de los datos, en lugar de la programación explicita. […] Machine Learning utiliza una serie de algoritmos que aprenden de los datos para mejorar, describirlos, y predecir resultados.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Ambas definiciones coinciden en que Machine Learning utiliza datos para su funcionamiento, y finalidad. Cuando se habla de Machine Learning, el enfoque es el aprendizaje, y para ello son necesarios los datos, de los cuales se extraen patrones con el fin de predecir.

Haciendo una comparativa, los humanos aprenden mediante la experiencia, mientras que las máquinas o sistemas aprenden de los datos. En este caso, los datos y la experiencia son equivalentes, ya que de ambos se puede extraer conocimiento. La experiencia humana son datos, que pueden ser representados y recopilados de tal manera que sean entendibles por una máquina.

El concepto de Machine Learning no es nuevo. Se estima que surgió a los fines de los cincuentas, pero no tuvo una gran presencia. Su impacto en la actualidad se debe al avance en la capacidad tecnológica, y la cantidad de datos generados diariamente y que son accesibles. Ambos factores han hecho que, en los últimos años, la popularidad de Machine Learning vaya aumentando, y que sea implementado en diversas industrias, como: negocios, salud, agricultura, clima, entre otras.

Tesis publicada con autorización del autor
No olvide citar esta tesis

2.2.1.2. Algoritmo de Machine Learning en la Aplicación Web

2.2.1.2.1. Algoritmos de Machine Learning

La (Real Academia Española [RAE], 2014) define un algoritmo como: "Conjunto ordenado y finito de operaciones que permite hallar la solución de un problema."

Generalmente se conoce que un algoritmo refiere a un grupo de instrucciones que tienen parámetros definidos, y están ordenadas lógicamente, con el fin de dar una solución a un determinado problema.

Géron (2017, pp. 4 - 5) señala que un algoritmo tradicional es desarrollado usando reglas predefinidas que suelen convertirse en una extensa y compleja lista, la cual es difícil de mantener debido a la constante adición de nuevas condiciones.



Figura 1. Funcionamiento de un Algoritmo Tradicional.

Fuente: Elaboración Propia

Tesis publicada con autorización del autor
No olvide citar esta tesis

Tradicionalmente, los algoritmos son creados dependiendo del contexto y del problema en cuestión. Siempre reciben una entrada, realizan un proceso, y proporcionan un resultado. Sin embargo, cuando el contexto es cambiante, y las condiciones que se presentan aumentan, este procedimiento puede volverse tedioso, ya que se necesitaría considerar todas las nuevas condiciones y agregarlas al algoritmo para no perder detalles del contexto o problema. En este sentido, la solución está en el uso de algoritmos de machine learning.

Brownlee (2016, p. 12) indica que los algoritmos de machine learning son técnicas que tienen como objetivo asignar una variable de salida (predicción) basándose en una o más variables de entrada. Cabe resaltar este tipo de algoritmos no dependen de reglas definidas, sino que establecen una lógica para crear supuestos.

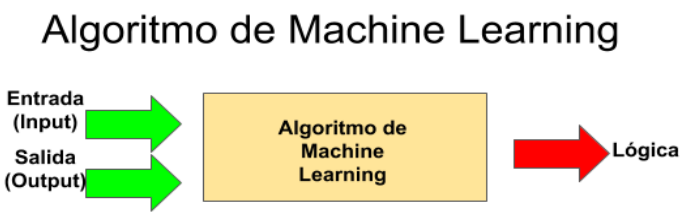


Figura 2. Funcionamiento de un Algoritmo de Machine Learning

Fuente: Elaboración Propia

Tesis publicada con autorización del autor
No olvide citar esta tesis

Un algoritmo de machine learning trabaja usando una entrada y una salida, dando como resultado una lógica. Una entrada es una característica o grupo de las mismas, presente en los datos que se tienen. Mientras que una salida es un resultado o etiqueta ya conocida en los datos. Ambos aspectos serán procesados, y el resultado es la lógica, o también llamada "patrón". La lógica será el modelo de machine learning, que se usará con nuevos datos.

Según Hurwitz y Kirsch (2018, p. 4), a medida que los algoritmos de machine learning reciben mayor cantidad de datos de entrenamiento, es posible que la precisión sea mayor.

Un algoritmo de machine learning se basa en el procesamiento de los datos que recibe. Por ello, del grupo de datos con los que se trabaja, se dividirán en dos grupos: datos de entrenamiento (training data), y datos de prueba (testing data). Cuanto mayor sean los datos de entrenamiento, la precisión aumentará debido a que existirá un mayor número de registros o casos de los cuales se pueden "aprender" o hacer inferencias.

Es importante resaltar, que la esencia de los algoritmos de machine learning está en el uso de datos para su funcionamiento. No obstante, se debe tener en cuenta la calidad de los datos, es decir que sean válidos, lógicos, y representativos.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Existe una gran variedad de algoritmos de machine learning. Generalmente se clasifican por el enfoque, que serían los siguientes: Aprendizaje Supervisado, Aprendizaje No Supervisado, y Aprendizaje Reforzado. En el primero, el objetivo es predecir un resultado (valor numérico o categórico) basándose en datos que ya tienen una etiqueta o clasificación. Mientras que en el segundo, el objetivo es agrupar datos basándose en semejanzas. Por último, el tercero se caracteriza por la retroalimentación, es decir que se aprende mediante "prueba y error".

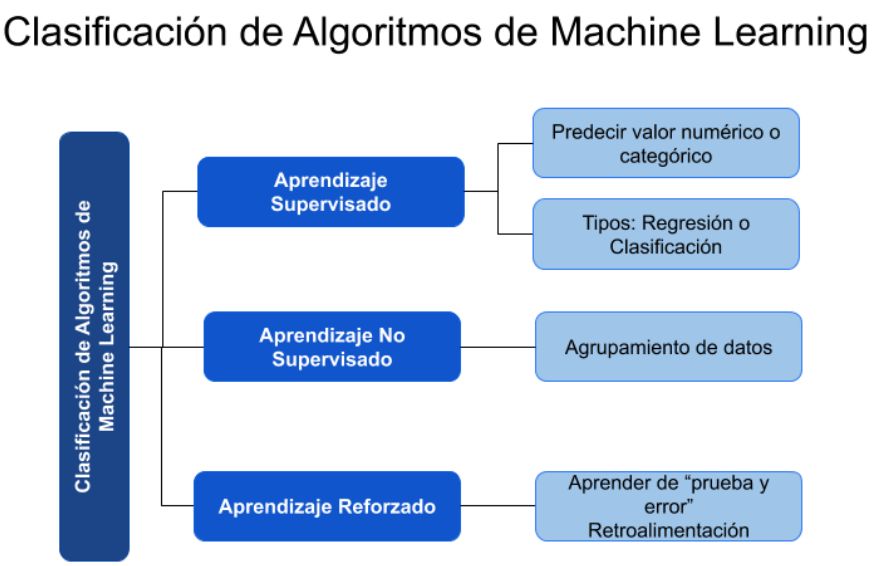


Figura 3. Clasificación de Algoritmos de Machine Learning

Fuente: Elaboración Propia

Tesis publicada con autorización del autor
No olvide citar esta tesis

A continuación, se explicaran los tipos de algoritmo más usados en machine learning. Entre ellos se tienen los siguientes:

2.2.1.2.1.1. Algoritmos de Clasificación

Theodoridis (2015, p. 60) señala que la clasificación predice una clase a la cual un objeto o patrón pertenece. Se asume que un patrón pertenece a una única clase conocida.

Por lo anterior, los algoritmos de clasificación tienen como objetivo la predicción de una clase o etiqueta, la cual será predefinida por las características en el conjunto de datos que se utilice. Por lo general, las clases son los resultados conocidos, y pueden ser varios.

Según Müller y Guido (2016, p. 25), la clasificación se separa en: clasificación binaria, y clasificación multiclase. La primera clasifica solo entre dos clases. Mientras que la segunda clasifica entre más de dos clases.

Dependiendo de los datos, la cantidad de clases con las que se trabaje, y tipo de las mismas, se seleccionará el algoritmo de clasificación más adecuado.

Tesis publicada con autorización del autor
No olvide citar esta tesis

2.2.1.2.1.2. Algoritmos de Regresión

Géron (2017, p. 8) indica que la regresión consiste en predecir un valor numérico en base a un grupo de características, a las cuales se les refiere como 'predictores'.

Asimismo, Müller y Guido (2016, p. 26) señalan que el objetivo de la regresión es predecir un número continuo, o también conocido como número flotante en programación, o número real en matemáticas.

Por los dos conceptos mencionados, los algoritmos de regresión buscan predecir un resultado de tipo numérico basándose en las relaciones que existan entre los predictores, los cuales son obtenidos del conjunto de datos seleccionado.

Los algoritmos de regresión pueden predecir precios, cantidades, entre otros valores numéricos dependiendo del problema planteado. Müller y Guido (2016, p. 26) indican que un ejemplo de este tipo de algoritmo es cuando se quiere predecir el ingreso anual de una persona basándose en su educación, su edad, y su lugar de residencia.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Del ejemplo anterior, el resultado que se busca es numérico ya que el ingreso anual hace referencia a una cantidad. Dicho ingreso anual será obtenido por los predictores, los cuales serían los tres siguientes: la educación, la edad, y el lugar de residencia. Los datos con los que se trabaja deben tener valores de ingreso anual conocidos con los mismos predictores, para que se pueda procesar, y obtener la predicción.

2.2.1.2.1.3. Algoritmos de Agrupamiento

Hurwitz y Kirsch (2018, p. 29) explican que estos algoritmos pertenecen al tipo de aprendizaje no supervisado, y que consisten en agrupar objetos con características similares.

Los algoritmos de agrupamiento, o también conocido en inglés como "Clustering", ayudan a dividir un conjunto de datos en grupos o clusters, los cuales comparten características entre sí. De esta manera, cuando este se aplique a futuros registros de datos, estos deberían ser asignados o clasificados en alguno de los grupos, dependiendo de sus características.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Géron (2017, p. 10) ejemplifica el uso del algoritmo de agrupamiento, mediante un caso en el que se tiene datos de los visitantes de un blog. Entonces, se plantea el uso de este tipo de algoritmo con el fin de detectar grupos similares de visitantes. Se enfatiza en que en ningún momento se le indica al algoritmo a que grupo pertenecen los visitantes, sino que este encuentra las conexiones por sí mismo. Por último, se menciona que existe el algoritmo de agrupamiento jerárquico, que subdivide cada grupo en grupos más pequeños.

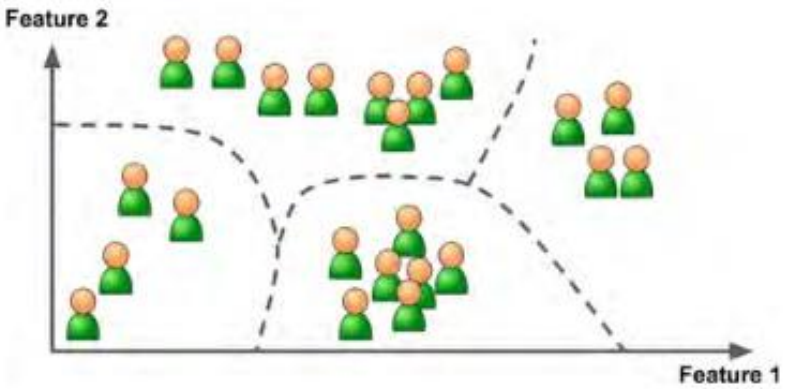


Figura 4. Algoritmo de Agrupamiento (Clustering)

Fuente: Géron (2017, p. 11)

Lo mencionado por el autor anterior, muestra que los algoritmos de agrupamiento encuentran similitudes entre los datos y los agrupan de acuerdo a las mismas, de tal manera que los grupos son diferentes entre sí. Estos algoritmos no necesitan conocer previamente la asignación de los grupos, sino que ellos lo determinan. En caso se quiera profundizar más sobre los grupos, se puede utilizar el agrupamiento jerárquico, y así obtener grupos más pequeños y selectos. Esto puede ayudar a enfocar soluciones o propuestas a grupos específicos, o detectar anomalías.

Tesis publicada con autorización del autor
No olvide citar esta tesis

2.2.1.2.1.4. Redes Neuronales y Deep Learning

Hurwitz y Kirsch (2018, p. 17) indican que una red neuronal está compuesta por tres o más capas: una capa de entrada, una o más capas ocultas, y una capa de salida. Los datos ingresan en la capa de entrada, y son procesados en la capa oculta y de salida.

Las redes neuronales son algoritmos que trabajan mediante capas, las cuales buscan abstraer características de los datos. Como mínimo una red neuronal tiene tres capas que interactúan entre sí. Generalmente estos se utilizan en problemas complejos de clasificación, regresión, o agrupamiento.

Según Hurwitz y Kirsch (2018, p. 17), Deep Learning es un método propio de machine learning que enlaza redes neuronales para aprender iterativamente de los datos. Es útil para aprender patrones de datos no estructurados.

Deep Learning es una forma de machine learning que utiliza esencialmente las redes neuronales, y trabaja por niveles o capas usando datos no estructurados, es decir aprendizaje no supervisado.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Hurwitz y Kirsch (2018, p. 17) señalan que Deep Learning y las redes neuronales complejas simulan el funcionamiento del cerebro humano, de tal forma que las máquinas puedan ser entrenadas para abstraer y solucionar problemas vagamente definidos.

La idea de Deep learning y las redes neuronales nacen por querer imitar o replicar la compleja actividad cerebral humana. De esta forma, las maquinas podrían simular el pensamiento humano y realizar actividades como lo hacen las personas.

Géron (2017, p. 253) señala que las redes neuronales son el centro del Deep Learning. Se caracterizan por ser versátiles, poderosas, y escalables, haciéndolas ideales para problemas de Machine Learning extensos y de alta complejidad, tales como clasificación de billones de imágenes (Google Imágenes), servicios de reconocimiento de voz (Siri de Apple), recomendaciones de videos (Youtube), entre otros.

Por lo anterior, se puede ver que las redes neuronales y Deep Learning están siendo utilizadas en plataformas que los usuarios consumen diariamente. Las empresas que aplican estas tecnologías buscan brindar una mejor experiencia a sus usuarios, mediante la optimización y personalización de su contenido, especialmente cuando se manejan cantidades enormes de datos.

Tesis publicada con autorización del autor
No olvide citar esta tesis

2.2.1.2.1.5. Random Forest

Shalev-Shwartz & Ben-David (2014, p.255) señalan que Random forest es un clasificador que consiste en una colección de Árboles de Decisión, donde cada árbol se construye aplicando un algoritmo en un conjunto de entrenamiento y un vector adicional. La predicción se obtiene por mayoría de votos sobre la predicción individual de cada uno de los árboles.

Brink, Richards, & Fetherolf (2017, p. 229) definen a Random forest como un método de aprendizaje por conjuntos, que se puede adaptar a varios árboles de decisiones de clasificación o regresión con subconjuntos de datos de entrenamiento, y hacer predicciones basadas en el modelo combinado. Sin embargo, Ozdemir (2016, p. 312) indica que la principal debilidad de los Árboles de decisión es las diferentes particiones de los datos de entrenamiento puede causar que los árboles sean muy diferentes.

Por lo mencionado anteriormente, Random forest consiste en un grupo de árboles de decisiones individuales que operan juntos para conseguir el mejor resultado predictivo, es decir que los resultados de cada árbol se suman al final. Cada árbol predice una clase, y aquella con más votos se convierte en la predicción final. Se considera también que este algoritmo limita el sobreentrenamiento, reduciendo la varianza en cada subconjunto de datos.

Tesis publicada con autorización del autor
No olvide citar esta tesis

2.2.1.2.2. Aplicación Web

La creación del Internet ha sido un punto clave para la evolución tecnológica. Aunque en un comienzo, este se usaba para hacer búsquedas y publicar páginas web; lo que lo convertía en un repositorio de información. Sin embargo, fue evolucionando, dando lugar a nuevas tecnologías y servicios por internet, tales como los servicios de correo electrónico, redes sociales, plataformas educativas, plataformas de compras y pagos, entre otros.

Es así como en la actualidad, muchas empresas tienen una página web, en conjunto con plataformas o aplicaciones web, las cuales albergan los principales servicios de la empresa. Los usuarios usan internet en su día a día. En consecuencia, las empresas trasladan sus servicios a la web, con el fin de facilitar la comunicación y optimizar sus procesos.

2.2.1.2.2.1. Definición de una Aplicación Web

Una aplicación web es una solución tecnológica que se conecta a un servidor web para obtener información (respuesta). Asimismo, este tipo de aplicación es adaptable, y accesible a partir de cualquier dispositivo con conexión a internet y con un navegador instalado.

Cabe destacar que se puede acceder a una aplicación web desde internet (público), o desde intranet (red interna o local). Esto depende de la usabilidad y propósito de la aplicación.

Tesis publicada con autorización del autor
No olvide citar esta tesis

En los últimos años, las aplicaciones web han escalado rápidamente debido a la adaptabilidad del diseño, que permite acceder desde todo tipo de dispositivo, y la necesidad de un navegador y conexión a internet para su visualización e interacción. De esta forma, facilitan varias actividades de la vida cotidiana, y es por ello que se están convirtiendo en las herramientas favoritas de muchos usuarios, yendo desde lo más simple como motores de búsqueda o servicios de correo electrónico, hasta algunas más complejas como comercio electrónico.

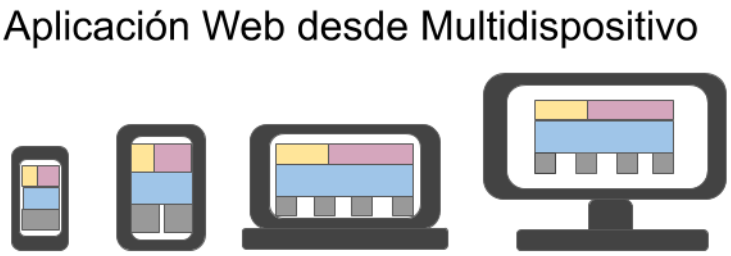


Figura 5. Aplicación Web desde Multidispositivos

Fuente: Elaboración Propia

Es preciso diferenciar una página web y una aplicación web. La primera es un conjunto de información en forma de documento, que no interactúa con un servidor web, es decir que solo muestra la información. Mientras que la segunda sí se conecta al servidor web, y es dinámica, ya que existe interacción del usuario con la aplicación.

Se suele decir que una aplicación web también es una página web; sin embargo, por lo explicado anteriormente, la diferencia radica en la complejidad y la funcionalidad de ambas.

Tesis publicada con autorización del autor
No olvide citar esta tesis

2.2.1.2.2.2. Funcionalidad de una Aplicación Web

Básicamente, una aplicación web funciona con tres elementos principales, que interactúan entre sí en el proceso de solicitudes y respuestas. Estos elementos son los siguientes:

- **Capa de Presentación:**

  Cardador Cabello (2015) señala que el objetivo de esta capa es mostrar el sistema web y la información contenida, transferir información del usuario, y enviarle las respuestas del servidor.

  Esta capa también es conocida como navegador o cliente. Es lo que el usuario visualiza cuando ingresa a la aplicación web, y con lo que va a interactuar. En otras palabras, es la estructura visual, o también llamada interfaz de usuario.

  Generalmente, para la creación de la interfaz de usuario, se emplean tecnologías como: HTML (para la estructura del contenido de la página), CSS (para el diseño y estilo de la página), y JavaScript (para el dinamismo e interactividad de la página).

  Las tecnologías anteriormente mencionadas son procesadas e interpretadas por el navegador, y luego este muestra al cliente o usuario, la interfaz. Es esta última, la encargada de realizar las solicitudes al servidor y mostrar las respuestas enviadas desde el mismo.

Tesis publicada con autorización del autor
No olvide citar esta tesis

- **Capa de Negocio:**

  Cardador Cabello (2015) indica que esta capa almacena programas que el servidor web usará. Además, recibe los datos de la capa de presentación, y los procesa para enviar respuestas al usuario.

  Por tanto, es válido decir que el servidor web es un software, al cual se le asigna un puerto o un canal, por donde se establece el contacto con el cliente, se aguarda para procesar sus solicitudes, y se envía lo debido, es decir una respuesta.

  Para crear aplicaciones web con conexión a un servidor, existen varios lenguajes de programación, tales como: PHP, Python, Java, entre otros. Actualmente, también existen entornos de ejecución, y frameworks, como por ejemplo: NodeJS, o AngularJS, respectivamente. Estos se basan en JavaScript, y su uso depende de las necesidades de la aplicación web.

- **Capa de Datos:**

  Cardador Cabello (2015) señala que en esta capa se almacenan los datos, y de mantener la accesibilidad a estos, mediante gestores de bases de datos.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Los datos en una aplicación son almacenados en una base de datos, de la cual se pueden extraer, modificar, o eliminar. Dependiendo de la funcionalidad de la aplicación, la base de datos puede almacenar contenido multimedia, texto, usuarios, entre otros.

Cuando se refiere a base de datos, existen dos tipos principales: la base de datos relacional (SQL), y la base de datos no relacional (NoSQL). La primera estructura los datos de forma consistente en tablas que guardan relación entre sí. Por otro lado, en la segunda, los datos se almacenan como documentos, y no existe relación entre ellos.

Las bases de datos relacionales más utilizadas son: MySQL, Microsoft SQL Server, Oracle, entre otras más. Mientras, que las bases de datos no relacionales más usadas son: MongoDB, Cloudant, y otras más.

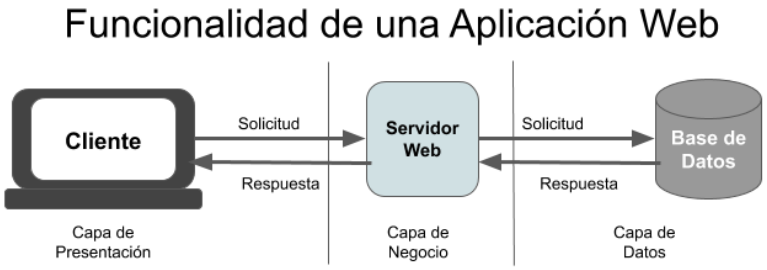


Figura 6. Funcionalidad de una Aplicación Web

Fuente: Elaboración Propia basada en Cardador Cabello (2015) y Kasireddy (2015)

Tesis publicada con autorización del autor
No olvide citar esta tesis

2.2.1.2.2.3. Proceso para Desarrollo de una Aplicación Web

El desarrollo de una aplicación web comprende los siguientes procesos elementales:

- **Definición de Requerimientos de la Aplicación Web**

  Principalmente en esta etapa se definen los objetivos y el alcance del desarrollo de la aplicación web. Además, se detallan los requerimientos funcionales y no funcionales. Los primeros se relacionan directamente con la funcionalidad de la aplicación, es decir qué va a hacer y cuáles son sus componentes. Mientras que los segundos se relacionan con la usabilidad de la aplicación, es decir, aspectos como entorno de desarrollo, seguridad, tipo de hardware, tiempo de respuesta, interfaz de usuario, entre otros.

- **Diseño de la Aplicación Web**

  El diseño va a determinar la relación entre los componentes y el comportamiento de la aplicación. Además se define la interfaz o las interfaces de usuario que se requieren. Por tanto, esto quiere decir que se define un flujo de trabajo, que contribuye al orden de las actividades en la etapa de desarrollo.

- **Implementación y Pruebas de la Aplicación Web**

  La implementación empieza desde la codificación de la aplicación hasta el primer prototipo funcional. Una vez que se obtiene el prototipo, este es sometido principalmente a pruebas funcionales, con el fin de validar el cumplimento de los requerimientos funcionales inicialmente definidos.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Como parte de la etapa de pruebas, el prototipo es sometido a pruebas de esfuerzo con la finalidad de simular la usabilidad, la carga de información, y otros aspectos. Siempre en esta etapa se procura llevar la aplicación al máximo para conocer sus límites. Los errores que se identifiquen, deben ser corregidos, y las pruebas deben ser repetidas para validar y obtener el prototipo final.



Figura 7. Proceso para Desarrollo de Aplicación Web

Fuente: Elaboración Propia

2.2.1.3. Datos para Modelo de Machine Learning

2.2.1.3.1. Datos

En la actualidad, los datos han ido incrementando su valor, cantidad, y aplicabilidad, debido a las nuevas innovaciones tecnológicas que promueven la generación de datos, su obtención, su análisis, y su utilización.

Cielen, Meysman y Ali (2016, p. 4) resaltan que existen diversos tipos de datos, tales como: Datos Estructurados, Datos No Estructurados, Audio, Video, Imágenes, Datos generados por Máquinas, entre otros. Independientemente del sector, las empresas recurren a los datos para mejorar o promover su desarrollo.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Actualmente, los datos son generados por varios canales, por lo que su estructura es variante. Por ejemplo, en la plataforma de una red social, un usuario registrado puede publicar texto, geolocalización, contenido multimedia, y más. Esto conduce a la existencia de distintos datos dentro de un mismo sitio, y los cuales la empresa proveedora se encarga de gestionar.

Cielen, Meysman y Ali (2016, pp. 2 - 3) ilustran el uso de los datos en una organización gubernamental. Señalan que este tipo de instituciones, no solo analizan sus datos, sino que también los comparten públicamente. Además, los científicos de datos trabajan en diferentes proyectos, como por ejemplo: la detección de fraude u otra actividad criminal, o la optimización financiera de un proyecto. Las fuentes de datos que estas entidades utilizan son amplias, desde Google Maps hasta mensajes de texto.

Las organizaciones gubernamentales no son las únicas que utilizan datos como parte de su actividad, sino que también lo hacen las organizaciones comerciales, o sin fines de lucro. Las razones para el uso de datos, depende del contexto de la empresa, sus objetivos, y su alcance.

Tesis publicada con autorización del autor
No olvide citar esta tesis

2.2.1.3.2. Modelo de Machine Learning

Hurwitz y Kirsch (2018, p. 4) señalan que un modelo de machine learning es el resultado obtenido del entrenamiento de un algoritmo de machine learning.

En párrafos anteriores se mencionó que el resultado de un algoritmo de machine learning, después de procesar los datos de entrenamiento, es una lógica, generada a partir de una entrada y una salida. La lógica obtenida es el llamado modelo de machine learning.

Cuando se tiene el modelo de machine learning, y se procesan nuevos datos con este, el resultado es llamado 'predicción', el cual está basado en los datos de entrenamiento usados originalmente en el algoritmo que creó el modelo. Es por ello que es necesario mantener la calidad y consistencia de los datos, para reducir errores en los resultados o predicciones del modelo de machine learning.

De un modelo de machine learning, se puede evaluar su desempeño, es decir que tan bien o mal está funcionando y haciendo predicciones. De esta manera, en caso de existir errores, se pueden implementar cambios y mejoras. Para la evaluación de desempeño, existen varias técnicas, las cuales dependerán del tipo de modelo a evaluar. Por ejemplo, existen métricas para evaluar modelos de clasificación, de regresión, o de agrupamiento.

Tesis publicada con autorización del autor
No olvide citar esta tesis

2.2.1.3.2.1. Métricas de Evaluación para Modelos de Clasificación

La forma más ideal de evaluar los modelos de clasificación es mediante la matriz de confusión. Müller y Guido (2016, p. 280) señalan que la matriz de confusión tiene una dimensión de dos por dos, donde las clases reales son las filas, y las clases predichas son las columnas. Se cuenta la frecuencia con la que un objeto de la fila, se clasifica como tal clase en la columna.

La matriz de confusión ayuda a medir el desempeño del modelo de clasificación mediante el conteo de las veces que una clase conocida es clasificada como tal, o como otra clase. Es decir que, al existir un cruce entre las filas (clases conocidas) y columnas (clases predichas), se puede observar cuantas veces el modelo predijo exactamente la clase conocida o cuantas veces la clasificó de otra manera.

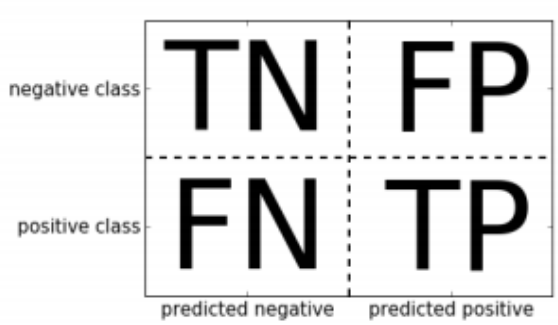


Figura 8. Matriz de Confusión

Fuente: Müller y Guido (2016, p. 281)

La figura anterior muestra el orden de las filas y columnas de la matriz de confusión. Al momento que ambas se cruzan, las cantidades representan las métricas básicas de desempeño del modelo.

Tesis publicada con autorización del autor
No olvide citar esta tesis

En la figura, estas métricas están representadas por sus siglas en inglés, y consisten en las siguientes:

- **TN:** En español, 'verdaderos negativos' (VN). Refiere al número total de predicciones donde la clase predicha fue negativa o falsa, y la clase real también era negativa o falsa.
- **FN:** En español, 'falsos negativos' (FN). Refiere al número total de predicciones inexactas donde la clase predicha fue negativa o falsa, pero la clase real era verdadera o positiva.
- **FP:** En español, 'falsos positivos' (FP). Refiere al número total de predicciones inexactas donde la clase predicha fue verdadera o positiva, pero la clase real era negativa o falsa.
- **TP:** En español, 'verdaderos positivos (VP). Refiere al número total de predicciones donde la clase predicha fue verdadera o positiva, y la clase real también era verdadera o positiva.

De la matriz de confusión, se obtienen otras métricas, tales como las siguientes:

2.2.1.3.2.1.1. Exactitud

Müller y Guido (2016, p. 282) definen la exactitud como el número de predicciones correctas (TP y TN) divididas entre el número total de registros.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Figura 9. Fórmula para Métrica de Exactitud

Fuente: Müller y Guido (2016, p. 282)

Tesis publicada con autorización del autor
No olvide citar esta tesis

En la figura anterior, se muestra la fórmula para el cálculo de la exactitud de un modelo. Se observa que se suman los verdaderos negativos y los verdaderos positivos (TN y TP), y se dividen entre la suma de todas las variables de la matriz de confusión (TP, TN, FP, y FN).

La exactitud es una métrica de desempeño que permite conocer el porcentaje de la cantidad de predicciones correctas del total de clases reales. En otras palabras, ayuda a detectar la frecuencia con la que el modelo de machine learning acierta las clases reales.

2.2.1.3.2.1.2. Sensibilidad

Müller y Guido (2016, p. 283) indican que la métrica de sensibilidad calcula cuantas clases positivas son detectadas como tal en las predicciones positivas.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}}$$

Figura 10. Fórmula para Métrica de Sensibilidad

Fuente: Müller y Guido (2016, p. 283)

En la figura anterior, se muestra la fórmula para el cálculo de la sensibilidad de un modelo. Se observa que el valor de los verdaderos positivos (TP) es dividido entre la suma de los verdaderos positivos y falsos negativos (TP y FN).

Tesis publicada con autorización del autor
No olvide citar esta tesis

La sensibilidad es la métrica de desempeño que permite conocer el porcentaje de la cantidad de predicciones positivas correctas del total de clases reales positivas. En otros términos, esta métrica permite calcular la frecuencia con la que el modelo de machine learning predice una clase positiva, y esta coincide con la clase positiva real.

2.2.1.3.2.1.3. Especificidad

Géron (2017, p. 91) define la especificidad como la proporción de muestras negativas que han sido correctamente clasificadas como negativas.

$$ESPECIFICIDAD = \frac{TN}{TN+FP}$$

Figura 11. Fórmula para Métrica de Especificidad

Fuente: Elaboración Propia

De la figura anterior, se observa que para el cálculo de la especificidad, el valor de los verdaderos negativos (TN) es dividido entre la suma de los valores de los verdaderos negativos y falsos positivos (TN y FP).

La especificidad es una métrica de desempeño que permite identificar el porcentaje de la cantidad de predicciones negativas clasificadas correctamente del total de clases reales negativas. Esto quiere decir que se calcula la frecuencia con la que el modelo de machine learning predice una clase negativa, y esta coincide con la clase negativa real.

Tesis publicada con autorización del autor
No olvide citar esta tesis

2.2.1.3.3. Data Science

En los últimos años, la cantidad de datos, que un usuario genera, ha incrementado sustancialmente, debido a su actividad en internet, donde realiza búsquedas, usa y publica en redes sociales, genera contenido multimedia, etcétera. En conjunto con esto, el acceso a dispositivos electrónicos como teléfonos inteligentes y tablets, permite a los usuarios mantener su actividad en internet constante durante el día.

En dicho contexto, los datos generados son gestionados para obtener valor de ellos. Es así como aparecen diversos términos relacionados con datos. Data Science y Machine Learning son dos términos muy distinguidos en esta tendencia. Ambos están relacionados, pero poseen diferentes denotaciones, y no deben usarse intercambiablemente.

2.2.1.3.3.1. Definición de Data Science

Cielen, Meysman y Ali (2016, p. 2) indican que Data Science comprende métodos para el análisis de grandes cantidades de datos, con el fin de obtener el máximo conocimiento de los mismos. Además, consideran que Data Science es una extensión del campo de la estadística, y que añade métodos computacionales a la misma.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Entonces, la ciencia de datos es un área que engloba y usa distintas áreas de estudio con el propósito de sustraer nuevos conocimientos de los datos, es decir información que una persona no puede detectar a simple vista debido a la complejidad o el volumen de datos con el que se trabaja.

Por otra parte, Grus (2015, p. 141) señala que Data Science consiste en convertir problemas de negocio en problemas de datos, y recolectar datos, comprenderlos, limpiarlos, y dales formato.

Por tanto, Data Science o Ciencia de Datos (en español) se encarga de trasladar problemas de la realidad a problemas computacionales para darles solución a partir del uso de datos. En este sentido, Data Science es el proceso de utilizar datos para comprender el entorno y las problemáticas que existen.

En Data Science, los datos que se recopilan son sometidos a un tratamiento en el que se limpian, corrigen, organizan, y forman, con el fin de estudiarlos y obtener información sobre tendencias o percepciones (insights) del problema o de la situación. De esta manera, se encuentran soluciones que pueden ser: idear estrategias y mejorar la toma de decisiones.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Hace 40 o 30 años, no existían las soluciones tecnológicas de hoy en día. Por ejemplo, almacenar datos no era posible, y si lo era, los costos eran demasiado elevados; sin embargo, ahora, grandes volúmenes de datos se pueden almacenar a precios accesibles. Lo mismo sucedía con la adquisición de softwares; no obstante, actualmente existen opciones gratuitas de código abierto.

Por consiguiente, el gran interés por Data Science en la actualidad, es por la existencia y accesibilidad a los datos, cuya tendencia de crecimiento va en aumento, así como también por la variedad de opciones tecnológicas que han surgido.

2.2.1.3.3.2. Metodología de Data Science

El término "metodología" refiere a un grupo de actividades o tareas ordenadas secuencialmente, con el fin de lograr determinados objetivos. A partir de esto, se puede decir que la metodología de Data Science busca proporcionar una estructura y orden para obtener resultados.

John B. Rollins es el autor de la metodología de Data Science. Él desarrolló esta metodología basándose en el proceso CRISP-DM y su experiencia profesional como científico de datos.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Rollins (2015, p. 2) menciona que esta metodología: "proporciona al científico de datos un marco sobre cómo proceder con los métodos, procesos y argumentos que se utilizarán para obtener respuestas o resultados."

Por ende, la metodología de Rollins busca poner a disposición una guía sobre el uso táctico de los métodos y actividades para conseguir resultados, dejando de lado las especificaciones de una determinada solución tecnológica.

Rollins (2015, p. 2) refiere que su metodología, en las primeras cinco etapas, es parecida a la metodología de minería de datos; sin embargo, enfatiza que su metodología acentúa la usabilidad de los grandes volúmenes de datos, el análisis de textos para modelos predictivos, y la automatización de ciertos procesos.

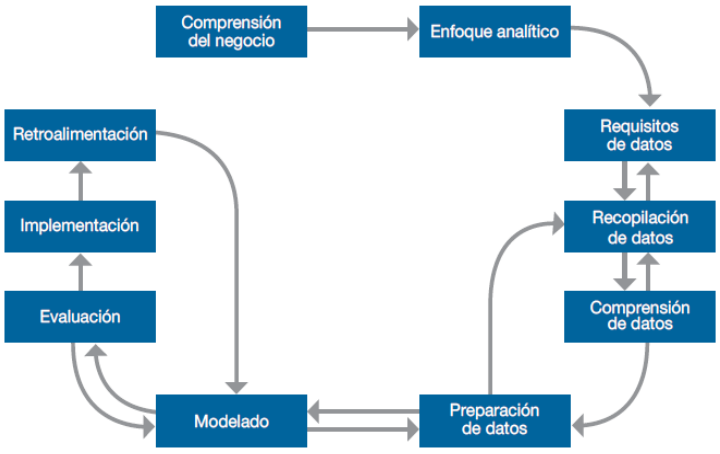


Figura 12. Metodología Fundamental para Ciencia de Datos

Fuente: Rollins (2015, p. 3)

Tesis publicada con autorización del autor
No olvide citar esta tesis

De la figura, se observa el flujo de la metodología. Muchas de las etapas tienen un flujo cíclico, lo cual representa la constante reciprocidad en aspectos de comunicación y relación existentes. Esto quiere decir que se puede añadir, o modificar aspectos, con el fin de mejorar el planteamiento de la solución.

La metodología de Data Science se caracteriza por ser iterativa y se compone de 10 etapas. Rollins (2015, pp. 3 - 5) señala que dichas etapas son las siguientes:

- **Etapa 1: Comprensión del negocio**

  En esta etapa, los dueños del negocio interesados en la solución analítica explican su situación problemática, dando lugar así, a la definición del problema, objetivos del proyecto, y características de la solución que se busca. Los dueños participan activamente durante todas las etapas, ya que proporcionan su perspectiva e información empresarial del negocio. Además, evalúan y validan los avances del proyecto.

- **Etapa 2: Enfoque analítico**

  A partir de la problemática empresarial planteada en la etapa 1, se determina el enfoque analítico, el cual representa la solución del problema usando herramientas estadísticas y de machine learning, y así determinar el más conveniente.

Tesis publicada con autorización del autor
No olvide citar esta tesis

- **Etapa 3: Requisitos de datos**

El enfoque analítico elegido en la etapa 2, decreta el contenido de los datos, el formato, y representación de los mismos.

- **Etapa 4: Recopilación de datos**

Se recopilan todos los datos disponibles, independientemente de su tipo, siempre y cuando sean de importancia para la problemática. En esta etapa, se evalúan decisiones de inversión para conseguir información no accesible o que no está al alcance inmediato. No obstante, se recomienda postergar dichas decisiones hasta conocer más sobre los datos que se tienen y el modelo con el que se trabajará.

- **Etapa 5: Comprensión de datos**

A los datos obtenidos en la etapa anterior, se les aplica estadísticas descripticas y técnicas de visualización para conocer su contenido, evaluar la calidad, etc.

- **Etapa 6: Preparación de datos**

El objetivo es generar una primera versión del conjunto de datos con el que se trabajará. Esta etapa comprende tareas como: la limpieza de datos (corrige valores inválidos, repetidos, faltantes, o en distinto formato), la combinación de datos de diversos orígenes, y la transformación de datos en información provechosa. Con la ayuda de herramientas que trabajan en paralelo, y la automatización de ciertos pasos, el tiempo de la preparación de datos se ve optimizado.

Tesis publicada con autorización del autor
No olvide citar esta tesis

- **Etapa 7: Modelado**

El objetivo es el desarrollo de modelos predictivos o descriptivos, de acuerdo al enfoque analítico elegido en la etapa 2. El modelado es iterativo debido a los insights que van obteniendo las empresas, lo que conlleva a repetir la preparación de datos y las especificaciones del modelo. Generalmente, se evalúan varios algoritmos hasta hallar el más adecuado para los datos que se tienen en ese momento.

- **Etapa 8: Evaluación**

Se evalúa la calidad del modelo, y que este sea la solución idónea y completa a la problemática empresarial planteada inicialmente. Para ello, se procede a calcular métricas de diagnóstico, en conjunto con gráficas y cuadros. En el caso de modelos predictivos, estos se evalúan con datos de pruebas, y así aplicar posibles ajustes; después se usan datos de validación para la evaluación final.

Una prueba adicional que se aplica a los modelos es la de significancia, la cual sustenta la implementación del modelo.

- **Etapa 9: Implementación**

Cuando los dueños del negocio aprueban el modelo desarrollado, este se implementa en producción o en un ambiente de pruebas. La complejidad de la implementación radica en el uso que se le dará al modelo, y las áreas empresariales que lo usarán.

Tesis publicada con autorización del autor
No olvide citar esta tesis

- **Etapa 10: Retroalimentación**

  El resultado del modelo implementado será la retroalimentación del mismo, ya que se conocerá su rendimiento y su impacto en la realidad de la empresa. La retroalimentación es analizada para mejorar el modelo, su precisión, y sus beneficios. Se pueden automatizar procesos tales como la evaluación de modelo, recopilación de retroalimentación, ajustes del modelo, y reimplementación.

Todas las etapas descritas anteriormente representan una estructura que busca tomar en cuenta varios factores que contribuyen a la resolución de un problema mediante el uso de los datos.

Se observa que la metodología empieza con la definición y delimitación de un problema o dificultad empresarial, y el entendimiento de los objetivos de los dueños del negocio. Luego, se elige el enfoque analítico (descriptivo, predictivo, análisis estadístico, etc.), el cual depende del problema planteado. A continuación, se definen los requisitos de los datos, es decir tipos de datos, sus características, el formato, las fuentes, entre otros factores para la primera recopilación. Seguidamente, se recopilan los datos de las fuentes indicadas que pueden ser base de datos internas o públicas, datos geográficos, entre otros.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Luego, para la comprensión de datos se usa estadística descriptiva, como el cálculo de la media, desviación estándar, etc., y se complementa con gráficos. De esta manera, se conoce si los datos recolectados representan la problemática definida. Posteriormente, se inicia la preparación de datos, a fin de eliminar elementos innecesarios o duplicados, completar valores, entre otras tareas. Por lo general, es la etapa que abarca más tiempo, por ello la automatización resulta favorable.

Después, en el modelado se consideran las características del problema y el enfoque analítico definido. Luego, la evaluación del modelo verifica su desempeño, si necesita ajustes, y si cumple como solución a la problemática. Acto seguido, la implementación del modelo se basa en la presentación del mismo a los dueños del negocio para su validación, y posible puesta en marcha. Finalmente, la retroalimentación favorece la mejora del modelo, mediante el conocimiento de su impacto y desempeño.

Cabe hacer hincapié en la importancia de las etapas 3 y 4, que corresponden a Requisitos de datos y Recopilación de datos respectivamente, puesto que dichos procesos influyen en la calidad y valor de los datos, lo cual tiene relación directa con el modelo. Para obtener el mejor rendimiento de este último, se necesita que los datos sean consistentes, relevantes, y de calidad.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Otra etapa a tener en cuenta es la número 6, que corresponde a Preparación de datos. Al ser una etapa que consume mucho tiempo, en algunos casos hasta un 70%, es necesario trabajar cuidadosamente, para evitar empezar de nuevo.

2.2.2. Riesgo de Accidente Cerebrovascular

2.2.2.1. Concepto de Accidente Cerebrovascular (ACV)

Un Accidente cerebrovascular es conocido con diferentes nombres, entre los cuales figuran los siguientes: Ataque Cerebral, Enfermedad Cerebrovascular, Ictus, Apoplejía, entre otros.

OMS (2016) clasifica esta enfermedad como la segunda causa de fallecimientos y la tercera causa de discapacidad, a nivel global. Esta situación se puede agravar cuando los recursos en el sector salud de un país no son los adecuados ni suficientes en cuestiones de emergencias, atenciones médicas, y prevención.

Se denomina Accidente Cerebrovascular (ACV) cuando la circulación a una zona del cerebro se detiene, produciendo que las células cerebrales mueran por la falta de oxígeno e irrigación sanguínea (National Institute of Neurological Disorders and Stroke [NINDS], 2019).

Tesis publicada con autorización del autor
No olvide citar esta tesis

Cuando disminuye o se corta la sangre al cerebro, puede ser por una obstrucción o por una hemorragia. En cualquiera de los casos, cuando el cerebro no recibe oxigeno ni irrigación sanguínea, este empieza a morir, y como consecuencias principales, la persona se enfrenta a la muerte o a una discapacidad.

2.2.2.2. Definición de Riesgo de Accidente Cerebrovascular

La (Real Academia Española [RAE], 2014) define la palabra riesgo como: "Contingencia o proximidad de un daño."

Por tanto, se entiende como riesgo de accidente cerebrovascular, a la posibilidad o tendencia de una persona a sufrir esta enfermedad. Esto se mide a partir de los factores de riesgo que se identifican en cada paciente. Por ende, a mayor cantidad de factores de riesgo, mayor es la posibilidad que ocurra un accidente cerebrovascular.

Mencionado lo anterior, es esencial el monitoreo de los factores de riesgo a fin de prevenir la incidencia de un accidente cerebrovascular y sus secuelas.

2.2.2.3. Factores de Riesgo

Este refiere a una particularidad propia de una enfermedad, la cual está presente o no en un individuo, indicando así que tan propenso puede estar a tener o no la enfermedad.

Tesis publicada con autorización del autor
No olvide citar esta tesis

En cuestión de un accidente cerebrovascular, estos factores son diversos, y se clasifican en dos grupos: Modificables y No Modificables. En los siguientes párrafos se explican los principales factores para cada uno de estos grupos.

2.2.2.3.1. Factores de Riesgo Modificables

Estos factores son características o comportamientos presentes en una persona debido a su estilo de vida. Cuando se identifican, el objetivo es monitorearlos, tratarlos (de ser necesario), y revisarlos para evitar valores o rangos elevados que aumenten la predisposición a un accidente cerebrovascular.

Entre los más importantes y frecuentes, se encuentran los siguientes:

2.2.2.3.1.1. Hipertensión

La hipertensión, o también conocida como Presión o Tensión Arterial Alta, es uno de los factores a tener en cuenta cuando se hace referencia a los accidentes cerebrovasculares.

La hipertensión es un trastorno que afecta directamente a los vasos sanguíneos. Cuando el corazón bombea sangre, los vasos se encargan de transportarla dentro del cuerpo. En dicho proceso, la sangre bombeada ejerce fuerza en las paredes de los vasos. Esta fuerza tiene el nombre de presión o tensión arterial. Cuando es alta, quiere decir que el corazón está haciendo mayor esfuerzo para bombear la sangre. (Organización Mundial de la Salud [OMS], 2013)

Tesis publicada con autorización del autor
No olvide citar esta tesis

Por lo anterior, la hipertensión requiere un control minucioso, ya que afecta tanto a los vasos sanguíneos (arterias) como al corazón (por el esfuerzo), pero también se ven comprometidas otras partes del cuerpo en las que el flujo sanguíneo es vital, como por ejemplo: el cerebro.

Es por ello que según el National Institute of Disorders and Stroke [NINDS] (2017), advierte que el riesgo de un accidente cerebrovascular en una persona hipertensa aumenta en cuatro a seis veces. Así como también, se estima que entre un 40% y 90% de personas que han tenido un accidente cerebrovascular, tienen hipertensión.

Para determinar la presión arterial, se necesitan los valores de la sístole y la diástole, las cuales se obtienen mediante el uso de un tensiómetro. Se considera que la presión arterial normal tiene que estar debajo de los valores: 120 mm Hg para sístole y 80mm Hg para diástole.

La Asociación Estadounidense del Corazón (AHA, por sus siglas en inglés), en su hoja informativa titulada "*¿Qué es la presión arterial alta?*", publica las categorías de presión arterial con fines preventivos.

Tesis publicada con autorización del autor
No olvide citar esta tesis

CATEGORÍA DE LA PRESIÓN ARTERIAL	SYSTÓLICA mm Hg (número de arriba)		DIASTÓLICA mm Hg (número de abajo)
NORMAL	MENOS DE 120	y	MENOS DE 80
ELEVADA	120 – 129	y	MENOS DE 80
PRESIÓN ARTERIAL ALTA (HIPERTENSIÓN) NIVEL 1	130 – 139	o	80 – 89
PRESIÓN ARTERIAL ALTA (HIPERTENSIÓN) NIVEL 2	140 O MÁS ALTA	o	90 O MÁS ALTA
CRISIS DE HIPERTENSIÓN (consulte a su médico de inmediato)	MÁS ALTA DE 180	y/o	MÁS ALTA DE 120

Figura 13. Categorías de la Presión Arterial

Fuente: American Heart Association [AHA] (2017)

2.2.2.3.1.2. Diabetes

La Diabetes es una enfermedad resultante de la incapacidad tanto del páncreas para producir insulina (hormona), como del cuerpo para emplearla. La insulina es la encargada de normalizar el nivel de glucosa en sangre. Esta última es esencial para el organismo, ya que es la energía que requieren las células, principalmente las cerebrales.

Según NINDS (2017), resalta que “las personas con diabetes tienen tres veces el riesgo de un accidente cerebrovascular de las personas sin diabetes.”

Tesis publicada con autorización del autor
No olvide citar esta tesis

Lo anterior demuestra que la diabetes, al igual que la hipertensión, son factores de riesgo directos tanto en los accidentes cerebrovasculares como en las condiciones cardiacas. Es esencial que la diabetes sea detectada y tratada para evitar complicaciones en la salud.

Se puede medir la glucosa mediante diferentes exámenes médicos, siendo el más común el examen de glucosa en sangre en ayunas. En dicha prueba, el valor de la glucosa óptimo debe ser menor a 100 mg/dl. Generalmente, se consideran los siguientes niveles de glucosa en el examen de sangre:

Tabla 1. *Examen de Glucosa en Sangre en Ayunas*

Categoría	Rango
Normal	Menor a 100 mg/dl
Prediabetes	100 mg/dl – 125 mg/dl
Diabetes	126 mg/dl a más

Fuente: Elaboración Propia basada en Mayo Clinic (2019)

Cuando el examen de glucosa en ayunas indica *Diabetes*, el médico otorga la orden para que el paciente se realice pruebas adicionales en laboratorio, como examen de aliento, examen de orina, entre otros.

Tesis publicada con autorización del autor
No olvide citar esta tesis

2.2.2.3.1.3. Colesterol

NINDS (2019) describe al colesterol como una sustancia grasosa que es transportada hacia las células mediante el colesterol LDL. Sin embargo, advierte que el exceso de LDL se acumula en los vasos sanguíneos, limitando su función.

El colesterol es beneficioso para el cuerpo ya que no solo ayuda en la formación celular (membranas), sino que también fomenta la producción de vitaminas como la A, y la D. No obstante, todo en exceso es dañino para la salud, por lo que tener un nivel alto de colesterol aumenta el tendencia a sufrir una enfermedad del corazón o un accidente cerebrovascular.

NINDS (2017) indica que el hígado se encarga de producir el colesterol necesario que el cuerpo requiere para sus funciones, y que este proceso natural no es un riesgo para la salud. Eventualmente, lo que pone en peligro a una persona, es la ingesta de alimentos altos en colesterol, tales como la carne o productos lácteos.

Como se mencionó en líneas anteriores, todo en exceso es dañino. El consumo desmesurado de productos altos en colesterol contribuye al aumento de problemas circulatorios, por lo que mantener una dieta balanceada es lo recomendable.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Tabla 2. *Análisis de Colesterol Total en Sangre*

Categoría	Niveles
Deseable	Menor a 200 mg/dl
Limite Superior	200 mg/dl – 239 mg/dl
Alto	240 mg/dl a más

Fuente: Elaboración Propia basada en Mayo Clinic (2019)

Existen dos tipos de colesterol: el HDL y el LDL (por sus siglas en inglés). El primero es conocido como "colesterol bueno", y el segundo es conocido como "colesterol malo". La suma de ambos resulta en el colesterol total.

NINDS (2017) señala que el LDL recorre todo el cuerpo, y su exceso se amontona en las arterias, haciendo que se endurezcan y se conviertan en placa arterial. Esta última fomenta el desarrollo de coágulos. Por otro lado, el HDL es provechoso para el organismo, ya que su exceso no se queda en las arterias sino que regresa al hígado, el cual lo elimina mediante los riñones.

Tabla 3. *Análisis de Colesterol HDL en Sangre*

Categoría	Niveles
Deficiente	Hombre: menor a 40 mg/dl Mujer: menor a 50 mg/dl
Mejor	40 mg/dl – 59 mg/dl
El mejor posible	60 mg/dl a más

Fuente: Elaboración Propia basada en Mayo Clinic (2019)

Tesis publicada con autorización del autor
No olvide citar esta tesis

El colesterol HDL (colesterol bueno) es beneficioso para el funcionamiento adecuado de los órganos del cuerpo. Además, se encarga de regular y reducir el colesterol LDL (colesterol malo), evitando así posibles bloqueos en las arterias.

Tabla 4. *Análisis de Colesterol LDL en Sangre*

Categoría	Niveles
Óptimo	Menor a 100 mg/dl
Casi óptimo	100 mg/dl – 129 mg/dl
Límite superior	130 mg/dl – 159 mg/dl
Alto	160 mg/dl – 189 mg/dl
Muy Alto	190 mg/dl a más

Fuente: Elaboración Propia basada en Mayo Clinic (2019)

El colesterol LDL en niveles elevados afecta directamente al sistema circulatorio debido a la acumulación de restos de grasa en las arterias. Esto produce un cierre gradual de dichos vasos sanguíneos, y evita que la sangre fluya como es debido. Consecuentemente, se generan afecciones cardiacas y cerebrales.

2.2.2.3.1.4. Índice de Masa Corporal

U.S. Department of Health and Human Services y U.S. Department of Agriculture (2015, p. 89) definen que el índice de masa corporal es un indicador de la grasa del cuerpo, y que se calcula considerando el peso y la talla de una persona. Asimismo, mencionan que este indicador tiene cuatro categorías.

Tesis publicada con autorización del autor
No olvide citar esta tesis

El índice de masa corporal permite conocer si una persona tiene o no, un peso óptimo respecto a su estatura. De esta manera, el valor de este indicador permite reconocer a los individuos con sobrepeso u obesos, es decir aquellos con un peso mayor al normal.

Tabla 5. *Niveles del Índice de Masa Corporal (IMC)*

IMC	Niveles
Bajo peso	Menor a 18.5
Normal	18.5 – 24.9
Sobrepeso	25.0 – 29.9
Obeso	30.0 a más

Fuente: Elaboración Propia basada en U.S. Department of Health and Human Services y U.S. Department of Agriculture (2015, p. 89)

Se atribuye el aumento del índice de masa corporal (IMC) al consumo de alimentos altos en grasas y a la inactividad física. El sobrepeso u obesidad refieren al excedente de grasa en el cuerpo, lo que resulta insano. Por tanto, se atribuye que a mayor IMC, existe mayor riesgo de ciertas enfermedades, tales como: diabetes, fallas cardiacas, entre otros.

2.2.2.3.1.5. Actividad Física

Hacer ejercicio es clave para la prevención de varias enfermedades. Sin embargo, la era tecnología ha traído consigo el sedentarismo, lo cual impide que el cuerpo no puede eliminar los excesos de sustancias como el colesterol o el azúcar, teniendo como consecuencia niveles altos de los mismos y otras enfermedades como cardiopatías y accidentes cerebrovasculares.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Cuando se habla de actividad física se refiere a una serie de ejercicios o deportes que la persona disfrute, y que pueda realizar de acuerdo a su condición médica y su edad.

U.S. Department of Health and Human Services y U.S. Department of Agriculture (2015, p. 74) recomiendan semanalmente realizar mínimo 150 minutos de actividad física de intensidad moderada, o 75 minutos de intensidad alta. Por otro lado, la Asociación Estadounidense del Corazón [AHA] (2016) señala que en caso la persona tenga sobrepeso o hipertensión, se recomienda 40 minutos de ejercicios de intensidad moderada con una frecuencia de 3 – 4 días por semana.

Las recomendaciones anteriores respecto a los tiempos, debe ser evaluada también por un médico en base a las condiciones del paciente. Siempre es aconsejable comenzar poco a poco, especialmente si se ha pasado por un tiempo prolongado de inactividad física.

2.2.2.3.1.6. Fumar

Jeena y Kumar (2018, p. 32) señalan que el humo del cigarro contiene químicos que al ser inhalados afectan el colesterol, la composición de la sangre, y la producción de plaquetas.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Fumar es un hábito nocivo, muchas veces vinculado con el ámbito social, y que se le relaciona ante todo con algún tipo de cáncer; sin embargo existen otras enfermedades a las que un fumador es propenso. La acción de fumar lesiona todos los órganos y las células del cuerpo, causando que la integridad del individuo se vea comprometida gravemente.

Existen dos tipos de fumadores: un fumador activo, que es aquella persona que fuma cigarrillos directamente, y un fumador pasivo, que es aquella persona que inhala o está expuesto al humo del cigarrillo. Independientemente del tipo de fumador, ambos se ven afectados por las consecuencias sobre su salud.

NINDS (2017) indica que fumar cigarros es uno de los factores que más influye en un accidente cerebrovascular. De esta manera, se ha reportado que en la población de adultos jóvenes, fumar es un factor directo de un accidente cerebrovascular. Mientras que en la población de adultos mayores se ve afectada por factores como diabetes, hipertensión, etc.

Tesis publicada con autorización del autor
No olvide citar esta tesis

De acuerdo con lo anteriormente referido, los adultos jóvenes son los que están sufriendo mayor incidencia de accidente cerebrovascular debido al cigarrillo. Sin embargo, se debe tener en cuenta que un adulto mayor que ha fumado por muchos años, y que ha dejado de hacerlo, aunque efectivamente reduce su riesgo de un accidente cerebrovascular, es muy difícil que dicho riesgo disminuya tanto para compararse con riesgo de una persona que jamás ha fumado.

Lo anterior se debe a que los efectos dañinos del cigarrillo, por años y por su cantidad, perjudican significativamente los órganos del cuerpo, dificultando que cuando se deje de fumar, las células puedan regenerarse por completo. Eso sí, existe una mejora notable en la salud, que aumentará a medida que pasen los años sin fumar. El NINDS (2017) indica que la propensión a un ictus se reduce después de abandonar el cigarrillo, especialmente luego de 2 a 4 años, pero que pueden pasar décadas para que el riesgo del exfumador equivalga a la de alguien que en la vida ha fumado.

2.2.2.3.2. Factores de Riesgo No Modificables

Estos factores son particularidades de una persona que no se pueden alterar o cambiar, pero que hay que tener en cuenta. Entre los principales se encuentran los siguientes:

Tesis publicada con autorización del autor
No olvide citar esta tesis

2.2.2.3.2.1. Edad

La edad es un factor que se debe tener en cuenta en todo momento, especialmente cuando se refiere a accidentes cerebrovasculares. NINDS (2019) señala que entre los 55 y 85 años, el riesgo de un accidente cerebrovascular se duplica cada 10 años.

Ninguna persona está exenta de sufrir un accidente cerebrovascular, pero se estima que a partir de cierta edad, el riesgo aumenta progresivamente. Esto se debe a que el cuerpo de un adulto mayor no funciona de la misma manera que el de un joven, ya que existen cambios en el metabolismo, enfermedades, influencia del estilo de vida, entre otros factores, que marcan una diferencia entre las ambas poblaciones. Cabe recordar, que en el Perú, los adultos mayores (65 años a más) son los más propensos a sufrir accidentes cerebrovasculares, siendo así calificada como la segunda causa de mortalidad en dicha población.

2.2.2.3.2.2. Sexo

El sexo de un individuo también influye en el aumento del riesgo de un accidente cerebrovascular. NINDS (2017) refiere que los hombres tienen más riesgo de padecer un accidente cerebrovascular, pero más mujeres mueren debido a lo mismo. Esto se debe a que los hombres sufren un accidente cerebrovascular cuando son más jóvenes que las mujeres, por lo que el porcentaje de supervivencia para ellos es mayor.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Las mujeres jóvenes no suelen sufrir de accidentes cerebrovasculares o relacionados, debido al estrógeno. Este último es una hormona que aumenta durante el ciclo menstrual, y protege de forma natural a las mujeres. Además, ayuda a regularizar el organismo. Por ejemplo, el estrógeno es parte de la síntesis de las grasas, haciendo que el HDL (colesterol bueno) aumente.

El estrógeno preserva la salud de las mujeres, haciendo que su riesgo de cardiopatías y accidentes cerebrovasculares sea mínimo. Esto es hasta que empieza la menopausia, y el estrógeno deja de producirse.

Por lo anterior, las mujeres suelen sufrir accidentes cerebrovasculares en etapas más avanzadas de su vida, y es probable que mueran debido a que a mayor edad, menor es el porcentaje de supervivencia por accidente cerebrovascular.

2.2.2.3.2.3. Antecedentes Familiares

NINDS (2017) señala que en la familia puede haber una predisposición genética a factores de riesgo relacionados a un accidente cerebrovascular, como la diabetes o hipertensión.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Enfermedades como la diabetes o hipertensión son consideradas hereditarias. Por lo que tener dichos antecedentes en la familia, es de especial cuidado, ya que junto con el estilo de vida que se adquiere, dichos factores podrían o no potenciarse.

Se debe tener en consideración los factores de riesgo y las condiciones que produjeron un accidente cerebrovascular en la familia. De este modo se conocerá la predisposición que existe, y se podrá hacer un control de los mismos.

2.2.2.4. Tipos de ACV

2.2.2.4.1. Accidente Cerebrovascular Isquémico

Se considera que el accidente cerebrovascular isquémico sucede con mayor frecuencia y prácticamente conforma el 80% de los casos.

Según NINDS (2017), describe que este tipo de accidente cerebrovascular: "ocurre cuando una arteria que suministra sangre al cerebro queda bloqueada, reduciendo repentinamente, o interrumpiendo el flujo de sangre y, con el tiempo, ocasionando un infarto en el cerebro".

Este tipo de accidente cerebrovascular se caracteriza por el bloqueo de una arteria del cerebro, lo cual interrumpe el flujo sanguíneo a dicho órgano. El bloqueo de la arteria es comúnmente causado por un coágulo, a lo que le puede acompañar la acumulación de grasa en este conducto.

Tesis publicada con autorización del autor
No olvide citar esta tesis

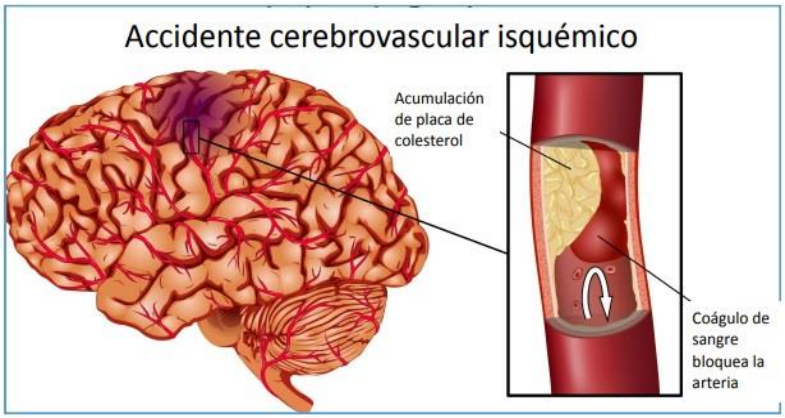


Figura 14. Accidente Cerebrovascular Isquémico

Fuente: UNC Comprehensive Stroke Center (2016, p. 3)

Dependiendo del lugar donde se forme el coágulo, el accidente cerebrovascular isquémico recibe una clasificación. NINDS (2017) indica que hay dos tipos:

- **Accidente cerebrovascular Embólico**: cuando la formación del coágulo (émbolo) es en una zona del cuerpo lejos del cerebro, y se desplaza hasta llegar a las arterias de dicho órgano.
- **Accidente cerebrovascular Trombótico:** cuando la formación del coágulo (trombo) es en una arteria del cerebro. Este se mantiene fijo, mientras aumenta de tamaño hasta bloquear la irrigación sanguínea al cerebro.

2.2.2.4.2. Accidente Cerebrovascular Hemorrágico

Posadas Ruiz (2018, p. 28) indica que un accidente cerebrovascular de tipo hemorrágico sucede por la rotura de una arteria en el cerebro.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Cuando se rompe una arteria, la sangre procedente de la misma entra en contacto directo con el cerebro, o ingresa a espacios del mismo. En cualquiera de los casos, la hemorragia tiene que ser detectada, y detenida lo antes posible para evitar daños neurológicos.

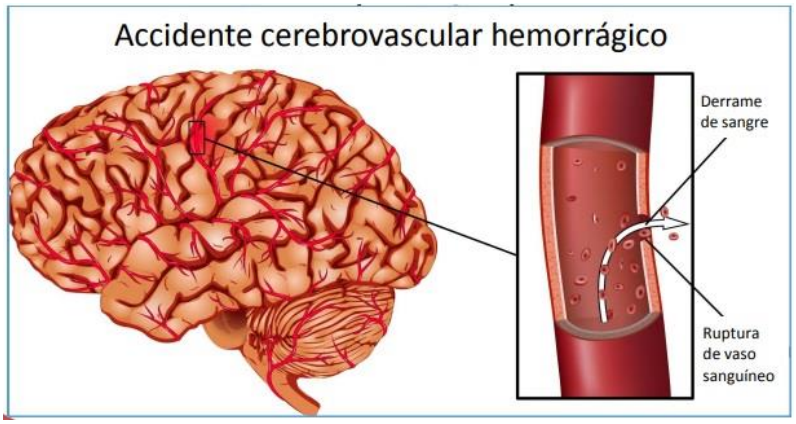


Figura 15. Accidente Cerebrovascular Hemorrágico

Fuente: UNC Comprehensive Stroke Center (2016, p. 3)

Existen dos tipos de accidente cerebrovascular hemorrágico dependiendo del lugar donde sucede la hemorragia. Según Posadas Ruiz (2018, p. 28 - 30), estas son las siguientes:

- **Hemorragia Intracerebral**: Sucede cuando una arteria del cerebro se rompe, derramando sangre en el mismo. Esto afecta a las neuronas y al tejido cerebral, ya que no obtienen la suficiente irrigación sanguínea.
- **Hemorragia Subaracnoidea:** Sucede cuando se rompe una arteria del cerebro, derramando sangre en el espacio entre el cerebro y el cráneo.

Tesis publicada con autorización del autor
No olvide citar esta tesis

De los dos tipos mencionados anteriormente, NINDS (2017) señala que la hemorragia subaracnoidea puede producir un daño en todo el cerebro, por lo que es el más letal de todos los tipos de accidente cerebrovascular.

Aunque los accidentes cerebrovasculares hemorrágicos suceden con menor frecuencia que los de tipo isquémico, son los más mortíferos debido a que el cerebro entra en contacto directo con la sangre, dañando el órgano y produciendo discapacidades. Por ello para evitar este tipo de accidente cerebrovascular, se debe controlar la presión arterial, y descartar presencia de aneurismas.

2.2.2.4.3. Accidente Isquémico Transitorio

Jeena y Kumar (2018, p. 32) señalan que a un accidente isquémico transitorio, también se le conoce como “mini accidente cerebrovascular”, y que es producido por el corte temporal de la irrigación sanguínea al cerebro. Generalmente, los síntomas que se presentan desaparecen dentro de las 24 horas.

Un accidente isquémico transitorio también es conocido como ataque isquémico transitorio o TIA (por sus siglas en inglés). Tiene la misma sintomatología que un accidente cerebrovascular tipo isquémico, por lo que también comparte su causa, siendo esta el bloqueo del flujo sanguíneo por un coágulo.

Tesis publicada con autorización del autor
No olvide citar esta tesis

NINDS (2017) señala que un accidente isquémico transitorio dura unos minutos, y los síntomas desaparecen en 1 hora aproximadamente. Sin embargo, no hay forma de determinar si los síntomas pueden ser incapacitantes o mortales. Por tanto, se debe tratar como un accidente cerebrovascular normal.

En un accidente isquémico transitorio, los síntomas que lo acompañan no son duraderos, pero se necesita una evaluación para conocer los daños que puede haber causado. Cabe resaltar que no se le debe restar importancia, ni esperar a que los síntomas desaparezcan, sino que se debe acudir lo más pronto a un centro médico.

Un accidente isquémico transitorio es una advertencia del riesgo que tiene una persona de sufrir un accidente cerebrovascular de mayor intensidad en cualquier momento. La evaluación médica es esencial para reducir dicho riesgo.

2.2.2.5. Diagnóstico

Existen varias pruebas para diagnosticar un accidente cerebrovascular. El Centro Integral de Accidente Cerebrovascular de UNC, también conocido en inglés como UNC Comprehensive Stroke Center (2016, p. 8), agrupa las pruebas de acuerdo a su fin, entre ellas se tienen:

- **Pruebas para determinar el tipo de accidente cerebrovascular:** Tomografía computarizada o Resonancia magnética.

Tesis publicada con autorización del autor
No olvide citar esta tesis

- **Pruebas para obtener más información del tipo y las razones:** Ecocardiograma, monitoreo cardiaco, angiografía cerebral, entre otros
- **Pruebas para detectar otras enfermedades:** Electroencefalograma, análisis de sangre, entre otros

Las pruebas de diagnóstico ayudan a conocer varios aspectos del accidente cerebrovascular. Pruebas como la tomografía o la resonancia magnética, otorga información sobre algún tipo de bloqueo o sangrado arterial. Pruebas a nivel cardiaco nos informan sobre el estado de las arterias y posibles coágulos o acumulaciones de grasa. El médico tratante será el encargado de decidir las pruebas basándose en el estado del paciente y su evaluación inicial.

En la evaluación inicial, que da indicios de un accidente cerebrovascular, se valoran aspectos como: los movimientos, la fuerza de las extremidades, el habla, la visión, entre otros.

2.2.2.6. Síntomas

Los síntomas de un accidente cerebrovascular se pueden presentar en cualquier momento. Es esencial la identificación de los mismos, con el fin de acudir a un centro médico cuanto antes.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Posadas Ruiz (2018, pp. 30 - 31) señala que los síntomas se presentan repentinamente y que varían dependiendo del lugar y la magnitud del accidente cerebrovascular. Se deben considerar los siguientes síntomas:

- **Dolor de cabeza:** Fuerte, inesperado, y que no cesa.
- **Alteración del tacto:** Se presentan problemas para distinguir temperaturas, o sentir dolor.
- **Problemas para comunicarse y comprender:** Signos de confusión, pérdida de memoria, no puede hablar, leer o escribir.
- **Problemas de visión:** Visión doble, reducción de la visión, o ceguera.
- **Falta de Control:** Problemas para controlar esfínteres, mantener el equilibrio, coordinar movimientos, o caminar.
- **Debilidad Muscular o Entumecimiento:** Especialmente en la cara, o una de las extremidades superiores o inferiores. Puede haber sensación de hormigueo.
- **Otros Síntomas:** Cambios emocionales o de personalidad, mareos, sueño repentino.

Los síntomas mencionados anteriormente pueden variar en duración e incluso pueden aparecer y desaparecer constantemente. Es importante reconocer los síntomas y actuar rápidamente llamando a emergencias o acudiendo al médico en seguida, ya que si no se recibe la atención medica requerida, las consecuencias podrían ser fatales.

Tesis publicada con autorización del autor
No olvide citar esta tesis

2.2.2.7. Consecuencias

Las consecuencias de sufrir un accidente cerebrovascular son: la discapacidad o la muerte. Ambos dependerán de varios factores como la zona en que sucede el accidente cerebrovascular, el tiempo en que se recibe atención médica, condición médica de la persona, entre otros.

American Heart Association [AHA] y American Stroke Association [ASA] (2015) señalan las siguientes consecuencias generales:

- Inmovilidad de alguno de los lados del cuerpo.
- Cansancio extremo.
- Problemas al hablar o ingerir alimentos.
- Problemas cognitivos (memoria, resolución de problemas).
- Cambios de humor, de comportamiento, o personalidad.
- Disminución de la visión periférica y percepción visual.

Las complicaciones mencionadas anteriormente, son las que más comúnmente se presentan tras un accidente cerebrovascular. La presencia de las mismas dependerá de la zona cerebral comprometida. Es recomendable observar cuidadosamente la situación emocional del paciente, ya que no solo puede presentar cambios de comportamiento o de humor, sino que también puede presentar cuadros de depresión, frustración, o negación, los cuales son esenciales tratar para encaminar su recuperación.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Además, en muchos casos, el paciente necesitará una persona que le ayude en sus actividades, ya que generalmente quedan limitados físicamente. No obstante, la rehabilitación es esencial para que dichos pacientes puedan recuperarse y recobrar progresivamente su independencia.

2.2.2.8. Prevención

La prevención es esencial para evitar un accidente cerebrovascular y todas sus complicaciones. De tal manera, se conserva la integridad física y mental de la persona. En caso la persona ya ha sufrido un accidente cerebrovascular, la prevención es importante para evitar otro a futuro.

El Centro Integral de Accidente Cerebrovascular de UNC, también conocido en inglés como UNC Comprehensive Stroke Center (2016, pp. 13 - 14), señala los siguientes aspectos a tener en cuenta para la prevención de un accidente cerebrovascular:

- **Controlar hipertensión, diabetes, y colesterol:** Implica realizar ejercicio, una dieta sana, y tomar los medicamentos recetados.
- **Alimentación sana:** Reducir el consumo de alimentos altos en grasas saturadas, grasas trans, sodio, y azúcar. Limitar el consumo de carne y mantequilla. Aumentar la ingesta de frutas, vegetales, pescado, carne magra, y cereales.
- **Actividad física regular:** Realizar mínimo 30 minutos diarios. Esto reduce el colesterol, azúcar, y presión sanguínea.

Tesis publicada con autorización del autor
No olvide citar esta tesis

- **Tabaco:** Si es fumador activo, dejar de hacerlo. Esto implica buscar ayuda médica, y cambiar estilo de vida. Además, evitar exposición al humo.
- **Alcohol y abuso de sustancias:** Evitar el consumo desmedido de alcohol, y drogas sociales.

De las recomendaciones anteriores, se puede observar que todas se basan en el control de factores de riesgo modificables. Esto se debe a que el estilo de vida de una persona influye considerablemente en su salud.

Lo fundamental es una dieta saludable, acompañada de una rutina de ejercicios, debido a que ambos ayudan al control de la hipertensión, la diabetes, y el colesterol. Asimismo, el médico puede recetar medicamentos complementarios para el control de los mismos. Igualmente, es crucial abandonar los cigarrillos, el alcohol, y otras drogas.

Tesis publicada con autorización del autor
No olvide citar esta tesis

2.3. Objetivos de la Investigación

2.3.1. Objetivo General

- Determinar el nivel de influencia de Machine Learning para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto.

2.3.2. Objetivos Específicos

- Determinar los requerimientos funcionales de Machine Learning para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto.
- Determinar el grado de exactitud de Machine Learning para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto.
- Determinar el grado de sensibilidad de Machine Learning para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto.
- Determinar el grado de especificidad de Machine Learning para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto.

Tesis publicada con autorización del autor
No olvide citar esta tesis

2.4. Formulación de Hipótesis

2.4.1. Hipótesis General

- Machine Learning influye significativamente para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto.

2.4.2. Hipótesis Específicas

- Los requerimientos funcionales de Machine Learning influyen significativamente para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto.
- El grado de exactitud de Machine Learning influye significativamente para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto.
- El grado de sensibilidad de Machine Learning influye significativamente para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto.
- El grado de especificidad de Machine Learning influye significativamente para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto.

Tesis publicada con autorización del autor
No olvide citar esta tesis

CAPÍTULO III: DISEÑO METODOLÓGICO

3.1. Diseño de la Investigación

El diseño de la presente investigación es experimental puro, ya que mediante un experimento, se busca conocer el efecto que tiene machine learning sobre el riesgo de accidente cerebrovascular. Además, la aplicación se diseña y desarrolla de manera completa, y se implementa en el consultorio médico, el cual es un ambiente controlado.

Para el desarrollo del Machine Learning se utilizará Python con librería PySpark en Jupyter Notebooks. Mientras que para desarrollar la Aplicación Web, se considerarán los siguientes componentes: En la Capa de Presentación se utilizará HTML, CSS y JavaScript; en la Capa de Negocio se utilizará JavaScript en plataforma NodeRed; y en la Capa de Datos se utilizará Cloudant.

El método es Hipotético - Deductivo debido a que se observa la realidad problemática, se formula una hipótesis y se hacen deducciones sobre los efectos o consecuencias de la misma, y finalmente se ejecuta un experimento para contrastar y comprobar lo planteado con la realidad.

3.2. Tipo

Esta investigación es de tipo correlacional, ya que emplea técnicas estadísticas para determinar el nivel de significancia entre las variables estudiadas.

Tesis publicada con autorización del autor
No olvide citar esta tesis

3.3. Enfoque

La investigación tiene un enfoque mixto porque combina la variable cuantitativa (aspecto numérico) y la variable cualitativa (opiniones y preguntas). Esto implica recolectar y analizar de datos en ambos aspectos.

3.4. Población

La Unidad de Análisis son aquellos pacientes con riesgo de accidente cerebrovascular.

Durante la entrevista, el Dr. William Pinto Zeballos, mencionó que existe un total de 520 pacientes atendidos por él en un año en su consultorio particular.

3.5. Muestra

La muestra se basa en una población de 520 pacientes. Se aplicará la siguiente técnica para determinar la correlación entre las variables, tal como se indica en la siguiente figura:

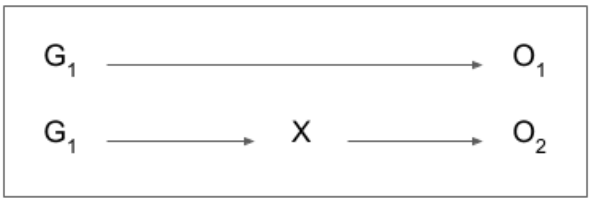


Figura 16. Técnica para la Muestra

Fuente: Elaboración Propia

Tesis publicada con autorización del autor
No olvide citar esta tesis

Mediante la fórmula probabilística, se obtiene lo siguiente:

$$n = \frac{N \cdot Z^2 \cdot (p.q)}{N \cdot E^2 + Z^2 \cdot (p.q)}$$

Figura 17. Fórmula para Muestra de Población Finita

Fuente: Peralta (2017)

De la fórmula:

N = Tamaño de la población; n = Tamaño de la Muestra

Z = Constante de Nivel de Confianza; E = Error Absoluto;

P = Porcentaje de la población; Q = (1-P);

Aplicando:

N = 520; **Z =** 1,96 al 95% de confianza; **E =** 5%; **P =** 0,5; **Q =** 0,5

Los valores para P y Q se deben a que no se conoce la composición de la muestra, por lo que se asume una estructura estándar.

Entonces: El resultado de la muestra es: 221 pacientes.

Esto quiere decir que se recopila 221 registros médicos de pacientes ya atendidos, que corresponden al último resultado obtenido en su historial. Esto significa que la relación es: 1 paciente es igual a 1 registro, el cual contiene la siguiente información para la investigación: edad, sexo, antecedentes familiares, fumador, hipertensión, glucosa, colesterol, índice de masa corporal, ejercicio, e incidencia de ACV.

Tesis publicada con autorización del autor
No olvide citar esta tesis

- **Criterio de Inclusión:**
  - Personas mayores de 18 años.
  - Personas atendidas por el Dr. Pinto.
  - 221 pacientes cuyo historial médico se considera para la validación del Machine Learning.
- **Criterio de Exclusión:**
  - Personas menores de 18 años
  - Personas atendidas por algún médico diferente al Dr. Pinto.

Con la muestra se realizó la validación de la solución de Machine Learning, y con la población restante, que equivale a 299 pacientes, se utilizó para que el algoritmo aprenda a realizar la predicción.

3.6. Operacionalización de Variables

Para mayor detalle sobre la Operacionalización de las variables de la presente investigación, referir al Anexo 2.

3.7. Técnicas para la Recolección de Datos

3.7.1. Descripción de los Instrumentos

- **Métricas de Desempeño:** Se utilizan para conocer el desempeño del modelo de machine learning desarrollado.

Tesis publicada con autorización del autor
No olvide citar esta tesis

a) **Indicador de Sensibilidad:**

La fórmula para este indicador es: **TP/TN+FN**

La sensibilidad es la métrica de desempeño que permite conocer el porcentaje de la cantidad de predicciones positivas correctas del total de clases reales positivas. En otros términos, esta métrica permite calcular la frecuencia con la que el modelo de machine learning predice una clase positiva, y esta coincide con la clase positiva real.

b) **Indicador de Especificidad:**

La fórmula para este indicador es: **TN/TN+FP**

La especificidad es una métrica de desempeño que permite identificar el porcentaje de la cantidad de predicciones negativas clasificadas correctamente del total de clases reales negativas. Esto quiere decir que se calcula la frecuencia con la que el modelo de machine learning predice una clase negativa, y esta coincide con la clase negativa real.

- **Ficha Técnica de Entrevista:** Se aplica a los pacientes para registrar y procesar sus principales factores de riesgo en el programa (experimento). Las variables en esta ficha se sustentan en las bases teóricas y en el juicio de expertos.

Tesis publicada con autorización del autor
No olvide citar esta tesis

- **Revisión de Historias Médicas:** Se revisan las historias médicas de pacientes ya atendidos, y se registran sus factores de riesgo, sin el uso del experimento.

3.7.2. Validez y Confiabilidad de los Instrumentos

- Con respecto a las métricas de desempeño, el uso, validez y confiabilidad de estas, están avaladas por varios autores, quienes las recomiendan para conocer el desempeño de un modelo de machine learning desarrollado.
- Para la ficha técnica de entrevista, su validación se da por juicio de expertos en Medicina. (Ver Anexo 4).

3.8. Técnica para el Procesamiento y Análisis de Datos

- Para este proceso se usó Python conjuntamente con librerías para análisis y visualización de datos, tales como: Pandas, Matplotlib, y Seaborn.
- El dataset estuvo compuesto por variables que representan los principales factores modificables y no modificables relacionados a un ACV. Asimismo se consideró una variable para la clasificación binaria de la enfermedad. El total de registros equivale a 520, de los cuales se considera 221 para la muestra y validación final, y 299 para el desarrollo de la solución. Del último grupo, el 70% de 299 se utilizó como training set y el 30% restante como testing set del desarrollo. Para más detalle, ver el punto 4.2 sobre Metodología del Proyecto.

Tesis publicada con autorización del autor
No olvide citar esta tesis

- Para la comprobación estadística de la Hipótesis, se usó la prueba de Chi-Cuadrado, puesto que se buscó comparar dos grupos de datos categóricos de tipo binario, y determinar el nivel de asociación entre los mismos. Hernández, Fernández, & Baptista (2014) la definen como: "prueba estadística para evaluar acerca de la relación entre dos variables categóricas."

3.9. Aspectos Éticos

El conjunto de datos recopilado de las historias médicas fueron proporcionadas con autorización y consentimiento del Consultorio Médico, no reúne datos personales de ningún paciente, y son de uso exclusivo para la elaboración del machine learning. Esto también aplica a los datos recopilados durante las pruebas reales de la aplicación de machine learning. En todo momento, se mantiene la confidencialidad de los datos.

Tesis publicada con autorización del autor
No olvide citar esta tesis

TESIS
UNIVERSIDAD CIENTÍFICA DEL SUR

CAPÍTULO IV: RESULTADOS

4.1. Diagnóstico del Sector

En el sector salud, basándose en el Análisis de las Causas de Mortalidad en Perú 1986-2015, del Ministerio de Salud del Perú [MINSA] (2018), se observa que la institución considera que los accidentes cerebrovasculares son de importancia nacional, ya que son la segunda causa de muerte en el país. Además, resalta que a nivel nacional, la tasa de mortalidad al 2015 aumentó en 45% en comparación al 2003. Mientras que a nivel regional, la Costa registró un aumento del 78% en la tasa de mortalidad desde el 2003 hasta el 2015.

En vista de la problemática en el Sector Salud sobre Accidentes Cerebrovasculares, se observa que tanto la incidencia como la mortalidad de la enfermedad, han incrementado en la última década. Asimismo, se estima una tendencia ascendente para los siguientes años.

Con respecto al uso de tecnologías emergentes como Machine Learning, el sector salud no ha reportado entidades que hayan implementado dichas técnicas. No obstante, el Ministerio de Salud del Perú [MINSA] (2018) bajo Resolución Ministerial, publicó un proyecto en que estima el uso de Machine Learning, Big Data, entre otras, para complementar las decisiones médicas.

Tesis publicada con autorización del autor
No olvide citar esta tesis

4.1.1. Análisis FODA del Sector:

A continuación se presenta un Análisis FODA sobre Machine Learning aplicado al Sector Salud en Perú:

a) Fortalezas:

1. Generación de médicos que consideran a la tecnología como un complemento eficaz durante las atenciones ambulatorias y emergencias.
2. Digitalización de historias médicas.

b) Debilidades:

1. Baja capacidad tecnológica en los establecimientos públicos de salud.
2. Reducido almacenamiento de datos para diseño de indicadores (KPIs) de Atención, Seguimiento de Incidencias, y Prevención.
3. Reducido interés en el uso de tecnología basada en datos, tales como Machine Learning, u otros, como complementos en las atenciones médicas y en temas de prevención.

c) Oportunidades:

1. Investigaciones internacionales que demuestran los beneficios y la eficacia del uso de Machine Learning en diversas cuestiones de salud.
2. Variedad de especialidades médicas, en las cuales Machine Learning puede utilizarse con fines de prevención y orientación médico-paciente.
3. Aumento del poder adquisitivo de los establecimientos de salud públicos para mejorar la capacidad tecnológica en los centros médicos.

Tesis publicada con autorización del autor
No olvide citar esta tesis

d) Amenazas:

1. Aumento de la incidencia y la tasa de mortalidad en enfermedades tratables, y prevenibles con un mejor seguimiento.
2. Baja concientización de la población en temas de prevención de enfermedades.
3. Dificultad en el monitoreo de pacientes con alto riesgo de una enfermedad.

Considerando lo anterior, se procede a cruzar los aspectos internos y externos, obteniendo lo siguiente:

a) Estrategia Fortalezas-Oportunidades (F-O):

- Implementación de tecnología basada en machine learning, orientada a contrarrestar enfermedades con mayor tasa de mortalidad como accidentes cerebrovasculares.
- Inversión en soluciones de almacenamiento y análisis de datos en Cloud para diseño de indicadores y control evolutivo mensual.
- Considerar que machine learning es una herramienta cognitiva solvente, que mediante el cálculo de aspectos como exactitud, sensibilidad, y especificidad, permite evaluar su desempeño y respuesta hacia diagnósticos de enfermedades.

Tesis publicada con autorización del autor
No olvide citar esta tesis

b) Estrategia Debilidades-Oportunidades (D-O):

- Considerar el grado de exactitud de machine learning para identificar accidentes cerebro-vasculares basándose en datos sobre la predisposición a la enfermedad; convirtiéndose en un complemento en cuestiones de diagnóstico médico.
- Considerar requerimientos funcionales necesarios para fomentar la implementación en centros médicos de aplicaciones webs o móviles, que incorporen servicios de machine learning.

c) Estrategia Fortalezas-Amenazas (F-A):

- Incentivos para promover el uso de la tecnología en distintas áreas de los establecimientos de salud.
- Monitoreo de pacientes y sus tratamientos, complementado con herramientas tecnológicas con grado de exactitud, sensibilidad, y especificidad significativos.

d) Estrategia Debilidades-Amenazas (D-A):

- Considerar iniciativas tecnológicas para el seguimiento del estado del paciente, incluyendo como requerimiento funcional el almacenamiento de datos y digitalización de historias médicas.
- Incorporar medidas preventivas para enfermedades con mayor mortalidad e incidencia.
- Establecer convenios con entidades para fomentar el uso de herramientas cognitivas para el sector salud.

Tesis publicada con autorización del autor
No olvide citar esta tesis

4.2. Metodología de desarrollo del Proyecto

Para el desarrollo de la presente investigación, se usó la metodología de Data Science elaborada por John B. Rollins, ya que refleja un marco de trabajo iterativo que consta de 10 etapas, cuyo objetivo es la comprensión del negocio, la implementación de una solución, y la retroalimentación para ajustes y mejoras a futuro (Rollins, 2015).

Esta metodología consiste en las siguientes etapas:

1. **Etapa 1: Comprensión del Negocio**

   Se trabajó en el consultorio del Dr. William Pinto, quien tiene 25 años de experiencia, y resalta que en el año 2019 ha observado una mayor incidencia de accidentes cerebrovasculares.

   Se establecen los siguientes objetivos del proyecto:

   - **Objetivo del Consultorio Médico:**

     Determinar el nivel de influencia de Machine Learning para predecir el riesgo de ACV en pacientes del Consultorio del Dr. Pinto.

   - **Objetivo de la Solución:**
     - Crear un modelo de Machine Learning basándose en los factores de riesgo de la enfermedad.
     - Crear una aplicación web que esté conectado al modelo de Machine Learning, para la mejor interacción con el usuario.

Tesis publicada con autorización del autor
No olvide citar esta tesis

2. **Etapa 2: Enfoque Analítico**

Se usa un enfoque predictivo, ya que se busca clasificar y conocer la probabilidad de incidencia de un accidente cerebrovascular, es decir el riesgo que el paciente tiene, basándose en una combinación de factores de riesgo.

Se usará un algoritmo de Machine Learning conocido como Bosque Aleatorio, o Random Forest (en Inglés), ya que su aleatoriedad reduce la varianza y mejora la precisión de los resultados. Además, no se ve afectado por valores atípicos, y el sobre-entrenamiento disminuye debido a las ejecuciones paralelas.

3. **Etapa 3: Requisitos de Datos**

Los datos necesarios, para la creación de la solución, se componen de la siguiente manera:

- **Factores de Riesgo Modificables:** Fumador (últimos 5 años), Hipertensión, Glucosa, Colesterol, Índice de Masa Corporal (IMC), y Ejercicio (minutos/semana).
- **Factores de Riesgo No Modificables:** Edad, Sexo, y Antecedentes Familiares.
- **ACV:** El paciente ha tenido un accidente cerebrovascular o no.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Tabla 6. *Detalle de los datos*

Datos	Tipo Dato	Descripción
Edad	Número	Edad del paciente
Sexo	Texto	Sexo del paciente
Antecedentes Familiares	Texto	Paciente tiene o no antecedentes familiares
Fumador (últimos. 5 años)	Texto	Paciente es o no fumador
Hipertensión	Texto	Paciente tiene o no antecedentes hipertensión
Glucosa	Número	Nivel de glucosa del paciente
Colesterol	Número	Nivel de colesterol del paciente
Índice de Masa Corporal	Número	IMC del paciente
Ejercicio (minutos/semana)	Número	Cantidad de minutos de ejercicio por semana
ACV	Texto	Paciente ha tenido o no ACV

Fuente: Elaboración Propia

4. Etapa 4: Recopilación de Datos

Se recopilan los datos requeridos utilizando una Ficha Técnica de Entrevista (ver Anexo 3) y las historias médicas de los pacientes. Estas últimas son accesibles y proveídas por el Consultorio Médico. No existe necesidad de invertir en base de datos de terceros.

Para mayor detalle sobre los datos recopilados en la presente investigación, referir al Anexo 7.

5. Etapa 5: Comprensión de Datos

En esta etapa, para comprender y conocer las variables, se aplica estadística descriptiva, junto con gráficos.

Tesis publicada con autorización del autor
No olvide citar esta tesis

a. Estructura del Dataset:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 520 entries, 0 to 519
Data columns (total 10 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   EDAD           520 non-null    int64
 1   SEXO           520 non-null    object
 2   HIPERTENSION   520 non-null    object
 3   GLUCOSA        520 non-null    int64
 4   COLESTEROL     520 non-null    int64
 5   IMC            520 non-null    int64
 6   HISTFAM        520 non-null    object
 7   FUMAR          520 non-null    object
 8   EJERCMINXSEM   520 non-null    int64
 9   ACV            520 non-null    object
dtypes: int64(5), object(5)
memory usage: 40.8+ KB
```

Figura 18. Estructura del Dataset Recopilado

Fuente: Elaboración Propia

La figura anterior muestra la estructura y los detalles del dataset. Este último consiste de 520 registros, 0 valores nulos, y 10 variables (5 variables numéricas y 5 variables categóricas).

b. Variables Numéricas: Estadística Descriptiva y Gráficos

Aplicando Estadística descriptiva, se obtiene el siguiente resumen:

Tabla 7. *Estadística Descriptiva de Variables Numéricas*

	Edad	Glucosa	Colesterol	IMC	Ejercicio (min/semana)
Conteo	520	520	520	520	520
Media	49.13	96.10	194.88	26.88	110.52
Std	12.97	19.01	26.89	3.94	68.05
Min	28	70	150	20	0
25%	38	81	172	23	51
50%	49	92	195	27	108
75%	60	108	218	30	164
Max	72	160	245	34	249

Fuente: Elaboración Propia

Tesis publicada con autorización del autor
No olvide citar esta tesis

A partir de la Tabla 7, sobre las variables numéricas se puede conocer los valores mínimos y máximos, la cantidad total de valores, la media, desviación estándar, entre otros.

Para visualizar este tipo de variables, se utiliza el histograma, el cual es un gráfico que muestra la distribución de las frecuencias. A continuación, se muestra el resultado para cada variable:

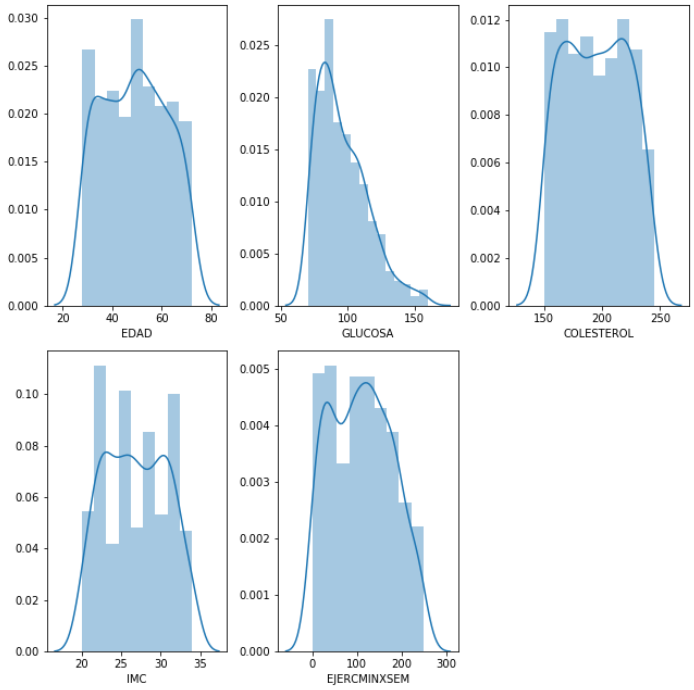


Figura 19. Histograma por cada Variable Numérica

Fuente: Elaboración Propia

Tesis publicada con autorización del autor
No olvide citar esta tesis

c. Variables Categóricas: Estadística Descriptiva y Gráficos

Aplicando la estadística descriptiva, se obtiene el siguiente resumen:

Tabla 8. *Estadística Descriptiva de Variables Categóricas*

	Sexo	Hipertensión	Historia Familiar	Fumar	ACV
Conteo	520	520	520	520	520
Único	2	2	2	2	2
Superior	M	S	N	N	S
Frecuencia	260	312	330	498	312

Fuente: Elaboración Propia

A partir de la Tabla 8Tabla **8**, sobre las variables categóricas, se puede conocer la cantidad de registros, los valores únicos, el primer valor que aparece, y la frecuencia del mismo.

Para visualizar las variables, se utiliza el gráfico de barras, el cual muestra las frecuencias según las categorías. A continuación, se muestra el gráfico para cada variable:

Tesis publicada con autorización del autor
No olvide citar esta tesis

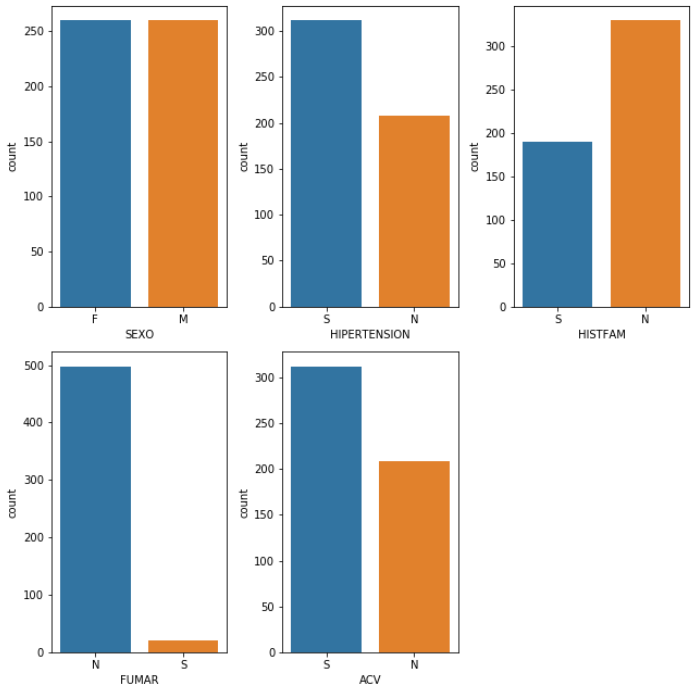


Figura 20. Gráfico de Barras por cada Variable Categórica

Fuente: Elaboración Propia

6. Etapa 6: Preparación de Datos

Se descarta que existan valores vacíos, nulos o repetidos, ya que estos se evitaron durante la recopilación de datos. Por tanto, se obtiene un total de 520 registros válidos, que forman parte de la primera versión del Dataset. (Ver Anexo 7Anexo 7).

7. Etapa 7: Modelado

Se desarrolla un modelo predictivo, ya que el enfoque analítico que se estableció fue el Predictivo usando el algoritmo de Machine Learning: Random Forest.

Tesis publicada con autorización del autor
No olvide citar esta tesis

a. Distribución de Datos:

Para este modelo predictivo, se usa el dataset recopilado en la etapa anterior, y se dividió de la siguiente manera:

- Del total de registros (520), se separaron 221 registros, los cuales representan la muestra de la investigación, y se usaron para la validación final.
- Los registros restantes, en este caso 299, se utilizan para el desarrollo del Machine Learning. De esta manera, los datos se dividieron en 70-30, es decir que el 70% de los datos son considerados para training set, y el 30% para testing set.

b. Desarrollo del Modelo:

A continuación, se detallan los pasos para la creación del Modelo Predictivo:

1. Se carga y divide el dataset.

```
from pyspark.sql import SparkSession
spark = SparkSession.builder.getOrCreate()

df_data_1 = spark.read\
  .format('org.apache.spark.sql.execution.datasources.csv.CSVFileFormat')\
  .option('header', 'true')\
  .option('inferSchema','True')\
  .load(cos.url('CompleteDataACV_v3-2_ML.csv'))
```

Figura 21. Carga de Datos

Fuente: Elaboración Propia

```
split_data = df_data_1.randomSplit([0.7, 0.30], 24)
train_data = split_data[0]
test_data = split_data[1]

print("Training Set: " + str(train_data.count()))
print("Testing Set : " + str(test_data.count()))
```

```
Training Set: 204
Testing Set : 95
```

Figura 22. División de los Datos

Fuente: Elaboración Propia

Tesis publicada con autorización del autor
No olvide citar esta tesis

2. Se importaron las librerías de pyspark necesarias.

```
from pyspark.ml.feature import StringIndexer, IndexToString, VectorAssembler
from pyspark.ml.classification import RandomForestClassifier
from pyspark.ml.evaluation import MulticlassClassificationEvaluator
from pyspark.ml import Pipeline, Model
```

Figura 23. Importación de Librerías necesarias

Fuente: Elaboración Propia

3. Se convirtieron las variables categóricas en índices numéricos.

```
stringIndexer_sexo = StringIndexer(inputCol="SEXO", outputCol="SEXO_IX")
stringIndexer_hip = StringIndexer(inputCol="HIPERTENSION", outputCol="HIPERTENSION_IX",stringOrderType="alphabetAsc")
stringIndexer_histfam = StringIndexer(inputCol="HISTFAM", outputCol="HISTFAM_IX")
stringIndexer_fumar = StringIndexer(inputCol="FUMAR", outputCol="FUMAR_IX")
stringIndexer_label = StringIndexer(inputCol="ACV", outputCol="label", stringOrderType="alphabetAsc").fit(df_data_1)
```

Figura 24. Conversión de Variables categóricas a numéricas

Fuente: Elaboración Propia

4. Se combinaron todas las variables (features) en un vector de una columna. Después, se definió el label y los features como parámetros en el algoritmo seleccionado.

```
vectorAssembler_features = VectorAssembler(inputCols=["EDAD","SEXO_IX","HIPERTENSION_IX","GLUCOSA",
                                                      "COLESTEROL","IMC","HISTFAM_IX",
                                                      "FUMAR_IX","EJERCMINXSEM"], outputCol="features")
```

```
rf = RandomForestClassifier(labelCol="label", featuresCol="features")
```

Figura 25. Parámetros para el Algoritmo

Fuente: Elaboración Propia

5. Se convirtieron las variables a su forma original. Además, se creó un pipeline para especificar las etapas de las conversiones en el modelo predictivo.

Tesis publicada con autorización del autor
No olvide citar esta tesis

```
labelConverter = IndexToString(inputCol="prediction", outputCol="predictedLabel", labels=stringIndexer_label.labels)

transform_df_pipeline = Pipeline(stages=[stringIndexer_sexo, stringIndexer_hip,
                                         stringIndexer_histfam, stringIndexer_fumar, stringIndexer_label, vectorAssembler_features])
transformed_df = transform_df_pipeline.fit(df_data_1).transform(df_data_1)

pipeline_rf = Pipeline(stages=[stringIndexer_sexo, stringIndexer_hip, stringIndexer_histfam,
                               stringIndexer_fumar, stringIndexer_label, vectorAssembler_features, rf, labelConverter])
```

Figura 26. Conversión de variables y Creación del Pipeline

Fuente: Elaboración Propia

6. Finalmente, se entrenó el modelo de Random Forest usando el training set y el pipeline.

```
stroke_rf = pipeline_rf.fit(train_data)
```

Figura 27. Entrenamiento del Modelo de Random Forest

Fuente: Elaboración Propia

c. Desarrollo de la Aplicación Web:

Se procede al desarrollo de la aplicación web para conectarlo con el componente de machine learning creado, y así facilitar su uso en un ambiente de pruebas.

La aplicación web consiste de 2 interfaces: en la primera se ingresan los datos del paciente, y en la segunda el doctor consulta los datos ingresados y el resultado. (Ver Anexo 8). Para mayor detalle, referir al Anexo 10 sobre el Diagrama de Capas de la solución.

Tesis publicada con autorización del autor
No olvide citar esta tesis

8. Etapa 8: Evaluación

El modelo creado se evalúa usando el testing set definido. Esto con el fin de validar los resultados y comprobar que el modelo funciona correctamente. Al término de la evaluación, se utiliza la matriz de confusión para conocer el desempeño del modelo, y calcular otras métricas de evaluación.

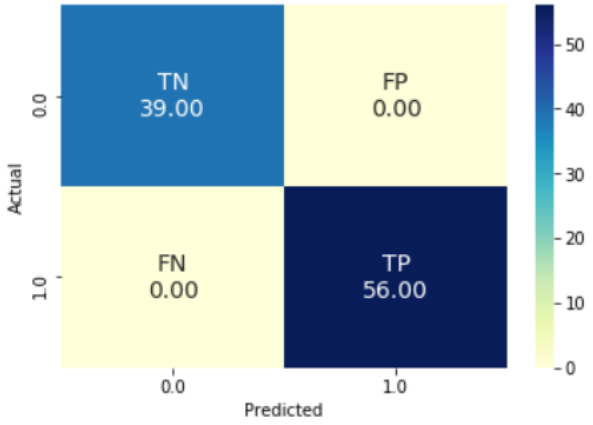


Figura 28. Matriz de Confusión usando Test Dataset

Fuente: Elaboración Propia

Basándose en los resultados de la matriz de confusión (Figura 28), se determina lo siguiente sobre el modelo desarrollado:

```
TN = conf_matrix[0][0]
FP = conf_matrix[0][1]
FN = conf_matrix[1][0]
TP = conf_matrix[1][1]
```

```
exct = (TP + TN)/(TP+TN+FP+FN)
senst = TP/(TP+FN)
especif = TN/(TN+FP)
print("Exactitud: ",exct)
print("Sensibilidad: ",senst)
print("Especificidad: ",especif)
```

```
Exactitud:  1.0
Sensibilidad:  1.0
Especificidad:  1.0
```

Figura 29. Métricas de Evaluación usando Test Dataset

Fuente: Elaboración Propia

Tesis publicada con autorización del autor
No olvide citar esta tesis

Según la Figura 29, se observa lo siguiente respecto a los resultados de las métricas de evaluación sobre el testing set:

- La Exactitud es 100%. Esto significa que el total de ACV ha sido predicho correctamente por el modelo.
- La Sensibilidad es 100%. Esto significa que del total de ACV que son "SI", todos han sido predichos por el modelo como tal.
- La Especificidad es 100%, lo que significa que del total de ACV que son "NO", todos han sido predichos como tal.

Considerando lo anterior, por el momento, no existe necesidad de realizar ajustes al modelo.

9. Etapa 9: Implementación

Luego de desarrollar la aplicación web, conectarlo con el machine learning, y obtener la aprobación del Doctor, se procede con la implementación dentro del consultorio médico para su uso durante las consultas médicas que se realicen. Por tanto, el flujo del proceso de atención será el siguiente:

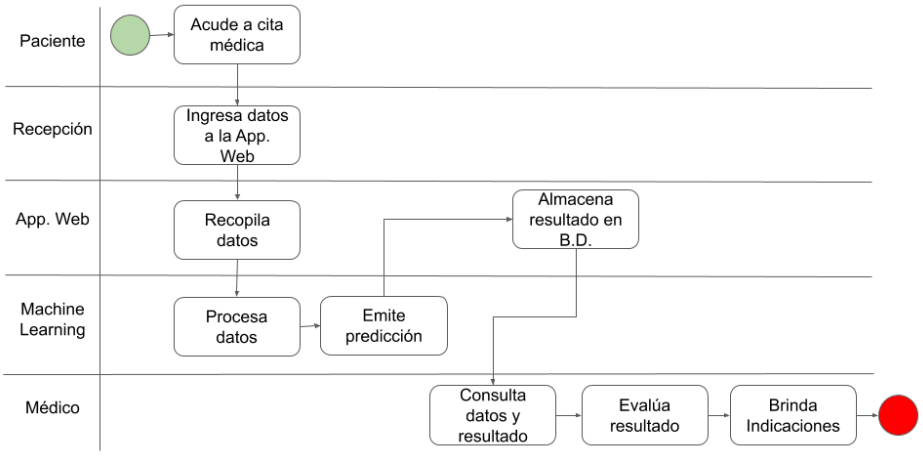


Figura 30. Implementación de la Solución

Fuente: Elaboración Propia

Tesis publicada con autorización del autor
No olvide citar esta tesis

10. Etapa 10: Retroalimentación

Después de utilizar la aplicación web con el modelo de machine learning en un grupo de 221 pacientes (muestra), los datos recopilados servirán para realizar ajustes y mejorar el modelo. Además, se conocerá su impacto en la realidad del Consultorio Médico.

4.3. Comprobación

4.3.1. Test – Validación

La validación se realizó usando los 221 registros correspondientes a la muestra. Estos datos fueron ingresados en la aplicación web, y procesados por el modelo de machine learning; obteniendo así, una tabla con 19 columnas (ver Anexo 9), de la cual se extraen las variables más relevantes para la comparación entre el resultado predicho y el diagnóstico del médico.

Tabla 9. C*omparación: Real vs. Predicción*

	Etiqueta Real	Predicción	Etiqueta Predicha	Probabilidad
1	N	0.0	N	[0.958, 0.042]
2	S	1.0	S	[0.0, 1.0]
3	S	1.0	S	[0.012, 0988]
4	N	0.0	N	[0.964,0.036]
5	S	1.0	S	[0.014,0.986]
6	N	0.0	N	[0.864,0.136]
7	N	0.0	N	[0.976,0.024]
8	S	1.0	S	[0.052,0.948]
9	N	0.0	N	[0.926,0.074]
10	N	0.0	N	[0.976,0.024]

Fuente: Elaboración Propia

Tesis publicada con autorización del autor
No olvide citar esta tesis

En la Tabla 9, se muestra un extracto con los 10 primeros registros de la comparación entre la columna Etiqueta Real (determinado por el médico) y la Etiqueta Predicha, realizada por el machine learning. Además, se muestran detalles como la codificación de la predicción, y la probabilidad de la misma. Para mayor detalle referir al Anexo 9.

El resultado de la matriz de confusión (con la cual se ve el trabajo del algoritmo) para la fase de la validación, se muestra a continuación:

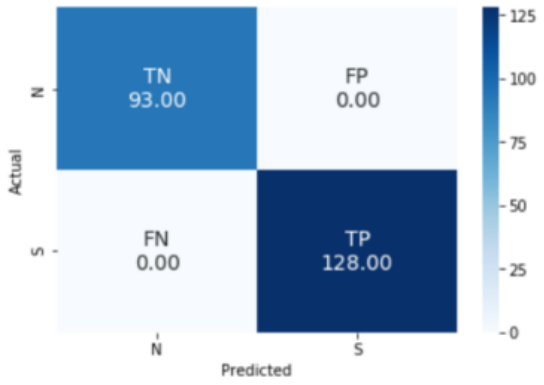


Figura 31. Matriz de Confusión usando Data de Validación

Fuente: Elaboración Propia

De la Figura 31, se calcularon las métricas de desempeño para esta etapa:

```
TN = conf_matrix['N']['N']
FP = conf_matrix['N']['S']
FN = conf_matrix['S']['N']
TP = conf_matrix['S']['S']
```

```
exct = (TP + TN)/(TP+TN+FP+FN)
senst = TP/(TP+FN)
especif = TN/(TN+FP)
print("Exactitud: ",exct)
print("Sensibilidad: ",senst)
print("Especificidad: ",especif)
```

```
Exactitud:  1.0
Sensibilidad:  1.0
Especificidad:  1.0
```

Figura 32. Métricas de Desempeño para Validación

Fuente: Elaboración Propia

Tesis publicada con autorización del autor
No olvide citar esta tesis

4.3.2. Estadística Descriptiva

a. **Datos de la Muestra:** De la muestra de 221 pacientes, se obtienen los resultados estadísticos mostrados a continuación:

- **Variables Numéricas:**

Tabla 10. *Resumen Estadístico: Variables Numéricas de la Muestra*

	Edad	Glucosa	Colesterol	IMC	Ejercicio (min/semana)
Conteo	221	221	221	221	221
Media	49.71	96.19	193.89	27.04	106.19
Desv. Std	13.06	19.20	27.41	4.00	70.25
Min	28	70	150	20	0
25%	38	80	170	24	46
50%	49	94	194	27	102
75%	60	108	218	30	161
Max	72	153	245	34	249

Fuente: Elaboración Propia

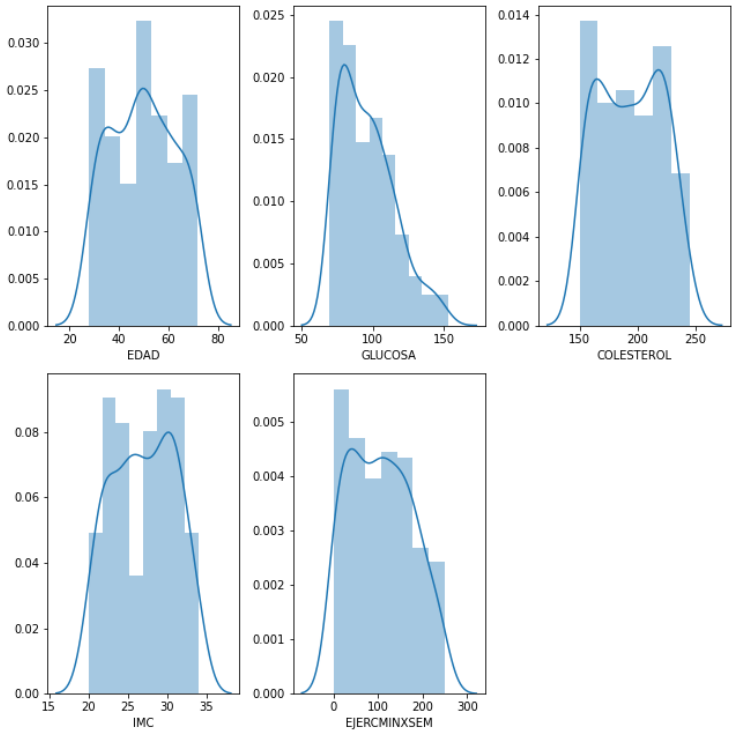


Figura 33. Histograma por Variable Numérica de la Muestra

Fuente: Elaboración Propia

Tesis publicada con autorización del autor
No olvide citar esta tesis

Basándose en la Tabla 10 y la Figura 33, se determina lo siguiente sobre cada variable:

- **Edad:** El rango de edades en los pacientes está entre 28 a 72 años. El 75% de ellos tiene una edad menor o igual a 60, lo cual está en relación al promedio general de edades que es 49.7 (Media), con una desviación estándar de 13.06 años.
- **Glucosa:** En general, los niveles de glucosa oscilan entre 70 y 153 mg/dl. El 50% de los pacientes tiene un nivel de glucosa igual o menor a 94 mg/dl, lo cual está en relación a la media que es 96.19 mg/dl, con una desviación de 19.20 unidades.
- **Colesterol:** Comprende un rango entre 150 y 245 mg/dl. El 50% de los pacientes tiene un nivel igual o menor a 194 mg/dl, lo cual es el nivel deseable. Esto está relacionado con la media que es 193.89 mg/dl, con una desviación de 27.40 unidades.
- **Índice de Masa Muscular (IMC):** El rango de IMC está entre 20 y 34, con una desviación estándar de 4. Solo el 25% de los pacientes tiene un IMC igual o menor a 24 (Normal). La media es 27.04, lo que indica sobrepeso.
- **Ejercicio (Minutos por semana):** En general, los pacientes hacen ejercicio entre 0 y 249 minutos por semana. El 50% de ellos realiza un tiempo igual o menor a 102 minutos, lo cual se relaciona con la media que es 106.19, con una desviación estándar de 70.25 minutos por semana.

Tesis publicada con autorización del autor
No olvide citar esta tesis

- **Variables Categóricas:**

Tabla 11. *Resumen Estadístico: Variables Categóricas de la Muestra*

	Sexo	Hipertensión	Historia Familiar	Fumar	ACV
Conteo	221	221	221	221	221
Único	2	2	2	2	2
Superior	F	S	N	N	S
Frecuencia	116	128	145	214	128

Fuente: Elaboración Propia

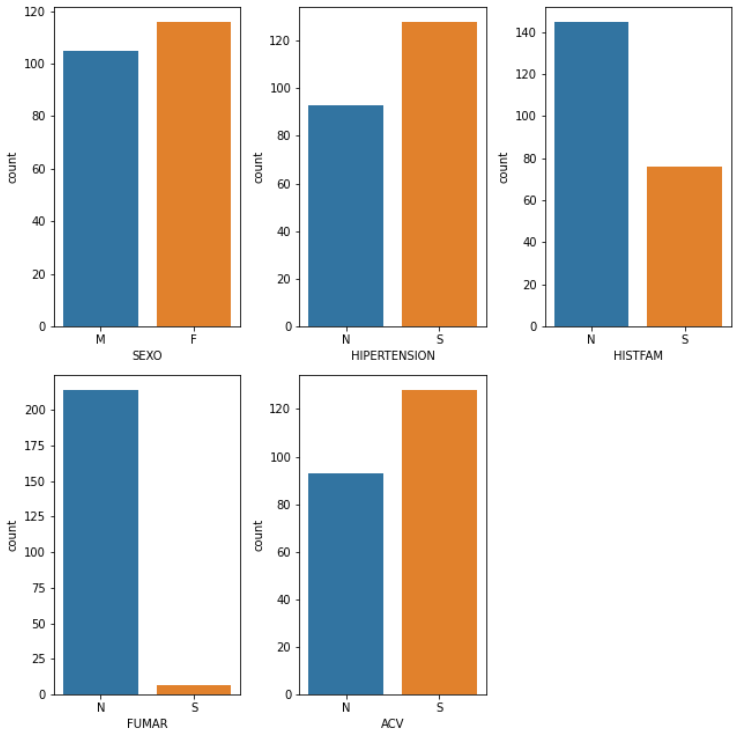


Figura 34. Gráfico de Barras por Variable Categórica de la Muestra

Fuente: Elaboración Propia

Basándose en la Tabla 11 y la Figura 34, se determina lo siguiente:

- **Sexo:** Consiste en dos valores: Femenino (F) y Masculino (M). En la muestra hay 116 mujeres y 105 hombres.

Tesis publicada con autorización del autor
No olvide citar esta tesis

- **Hipertensión:** Tiene dos valores: Sí (S) y No (N). 128 pacientes tienen hipertensión, mientras que 93 no tienen.
- **Historia Familiar:** Existen 145 pacientes que no tienen historia o antecedente familiar, mientras que 76 pacientes si tienen.
- **Fumar:** Tiene dos valores: Sí (S) y No (N). De los 221 pacientes, 214 no fuman, en tanto 7 si lo hacen.
- **Accidente Cerebrovascular (ACV):** Del total de pacientes en la muestra, 128 han tenido un ACV, y 93 no han tenido.

b. Resultados por Indicador:

- **Indicador 1: Requerimientos Funcionales**

  En el Anexo 6, correspondiente a la Ficha de Requerimientos Técnicos, se detallaron los requerimientos funcionales de la solución. Los cuales se cumplieron durante el desarrollo de la metodología de este proyecto.

- **Indicador 2: Porcentaje Total de ACV clasificados correctamente**

  Considerando la Figura 32, la exactitud es del 100%.

- **Indicador 3: Porcentaje de ACV clasificados como "SI"**

  Considerando la Figura 32, el porcentaje de ACV clasificados como "SI" es 100% (Sensibilidad).

- **Indicador 4: Porcentaje de ACV clasificados como "NO"**

  Considerando la Figura 32, el porcentaje de ACV clasificados como "NO" es 100% (Especificidad).

Tesis publicada con autorización del autor
No olvide citar esta tesis

4.3.3. Estadística Inferencial

Para la prueba de Hipótesis, se aplica el test de Chi-Cuadrado, puesto que se busca comparar dos grupos de datos categóricos de tipo binario, y determinar el nivel de asociación entre los mismos.

Se consideran las etiquetas reales (criterio médico) y las etiquetas predichas (resultado de Machine Learning).

a. Declaratoria de Hipótesis:

Hipótesis Nula (H_0): Etiquetas reales $\neq$ Etiquetas Predichas

Hipótesis Alternativa (H_A): Etiquetas reales $=$ Etiquetas Predichas

b. Prueba de Hipótesis: Chi-cuadrado

Los resultados son los siguientes:

```
chi2, p, ddof, expected = chi2_contingency(freq_tab)
txt = "Test Statistic: {}\np-value: {}\nDegrees of Freedom: {}\n"
print( txt.format( chi2, p, ddof ) )
print( expected )
```

```
Test Statistic: 216.91613622218395
p-value: 4.2567100507437095e-49
Degrees of Freedom: 1

[[39.13574661 53.86425339]
 [53.86425339 74.13574661]]
```

Figura 35. Resultado de la Prueba de Chi-cuadrado

Fuente: Elaboración Propia

Del resultado de la prueba de Chi-cuadrado (Figura 35), se observa:

- Chi-cuadrado = 216.92 con grado de libertad = 1
- El valor de p es menor que el nivel de significancia (0.05), ya que se aproxima a 0.

Tesis publicada con autorización del autor
No olvide citar esta tesis

TESIS
UNIVERSIDAD CIENTÍFICA DEL SUR

Por tanto, considerando el valor de p, se rechaza la hipótesis nula en favor de la hipótesis alternativa. Esto significa que existe relación entre ambas etiquetas.

4.3.4. Random Forest: Funcionamiento e Importancia de las variables

El algoritmo Random Forest, utilizado durante el desarrollo de la solución, consiste en una serie de árboles de decisiones individuales que operan juntos para conseguir el mejor resultado predictivo. Dicha característica de individualismo significa que los árboles no están relacionados entre sí al momento de decidir. Cada árbol predice una clase, y aquella con más votos se convierte en la predicción final.

Brownlee (2016, p.128) explica que Random forest es una versión mejorada de los árboles de decisiones al utilizar sub-árboles con baja correlación para obtener predicciones. Al mismo tiempo, cada árbol predice sobre un grupo de variables seleccionadas aleatoriamente, y no sobre el total.

Por otro lado, la importancia de las variables, o Feature Importance (en inglés), proporciona detalles sobre las variables más relevantes e influyentes entre las clases.

Tesis publicada con autorización del autor
No olvide citar esta tesis

A continuación se muestra la importancia que le asigna Random Forest a los datos ingresados en el desarrollo:

Tabla 12. *Random Forest: Feature Importance*

	Variable	Puntaje
1	Hipertensión	0.6229
2	Ejercicio (min/semana)	0.1224
3	Historia Familiar	0.0984
4	Colesterol	0.0455
5	IMC	0.0365
6	Glucosa	0.0346
7	Edad	0.0331
8	Sexo	0.0047
9	Fumar	0.0018

Fuente: Elaboración Propia

Basándose en la Tabla 12, se observa que la variable Hipertensión es la más relevante (0.62) en las predicciones, seguido de las variables Ejercicio (0.12) e Historia familiar (0.1).

Tesis publicada con autorización del autor
No olvide citar esta tesis

CAPÍTULO V: DISCUSIÓN, CONCLUSIONES Y RECOMENDACIONES

5.1. Discusión

Hipótesis General:

En relación a la hipótesis general: "Machine Learning influye significativamente para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto", Asadi et al. (2014) consideraron que los modelos de machine learning pueden ser de gran utilidad no solo para prognosis y predicción de resultados, sino que también como asistentes en la toma de decisiones clínicas para casos particulares. Considerando ello, en la presente investigación se obtuvo que el desempeño de machine learning fue óptimo al cumplir los requerimientos funcionales (ver Anexo 6), y los resultados de la prueba de hipótesis evidencian que sí existe relación entre los diagnósticos de accidente cerebrovascular reales y predichos, debido a que el valor de chi-cuadrado es 216.92 y el valor de p se aproxima a 0, lo cual es menor al nivel de significancia (0.05). Por consiguiente, se cumple el objetivo general que es determinar el nivel de influencia de machine learning para predecir el riesgo de accidente cerebrovascular. Adicionalmente, la tabla 13 muestra que las métricas de desempeño varían, cuando se remueve Hipertensión, aún con esta sensibilidad el modelo es mayor al 60% de confianza.

Tabla 13. *Comparación de métricas entre modelos*

	Exactitud	Sensibilidad	Especificidad
Modelo original (Todas las variables)	100%	100%	100%
Modelo 2 (Removiendo: Hipertensión)	65.26%	65.75%	63.63%

Fuente: Elaboración Propia

Tesis publicada con autorización del autor
No olvide citar esta tesis

Hipótesis Específicas:

- Para la hipótesis específica N°1 relacionada con los requerimientos funcionales, Asadi et al. (2014) propusieron que el modelo creado se incorporase en un sistema online que sirva como guía para decisiones terapéuticas, y que permita la incorporación de datos para mejoras en el desempeño. Considerando esto, en la presente investigación se puede evidenciar mediante lo descrito en el Capítulo IV, que para los requerimientos funcionales de machine learning se seleccionó el algoritmo de Random Forest, el cual demostró un alto desempeño y optimo tiempo de respuesta desde el entrenamiento hasta la fase de predicciones. Esto integrado con la Aplicación web, cuyos requerimientos funcionales consistieron en dos interfaces tanto para registro como búsqueda de información, y conexiones tanto al módulo de machine learning como al de base de datos, permitió que la solución sea interactiva, y que cumpla con la recolección de nuevos datos, procesamiento, y almacenamiento de los mismos.

Tesis publicada con autorización del autor
No olvide citar esta tesis

- Para la hipótesis específica N°2 relacionada al grado de exactitud, en la investigación “Machine Learning for Outcome Prediction of Acute Ischemic Stroke Post Intra-Arterial Therapy”, Asadi et al. (2014) consideraron prometedor obtener una exactitud cercana a 70% a pesar de trabajar con una muestra pequeña (107), ya que se podrían incorporar más datos para mejorar las predicciones. Esto demuestra que las predicciones y la cantidad de datos están relacionados. En la presente investigación, la exactitud alcanzó 100% (ver Figura 32) considerando una muestra de 221, por lo que no se descarta la incorporación de nuevas fuentes de datos con fines de mejora.

- En referencia a las hipótesis específicas N°2, N°3, y N°4 relacionadas al grado de exactitud, sensibilidad, y especificidad, en la investigación “Machine Intelligence in Stroke Prediction”, Jeena y Kumar (2018), usando una muestra de 350, mostraron que el algoritmo de red neuronal alcanzó una precisión de 98.1%, sensibilidad del 100%, y especificad del 95.8%. Esto demuestra que las métricas de desempeño de un algoritmo pueden alcanzar valores superiores a 90%. En la presente investigación, para el algoritmo de Random Forest, las métricas de exactitud, sensibilidad, y especificidad alcanzaron el 100% (ver Figura 32).

Tesis publicada con autorización del autor
No olvide citar esta tesis

5.2. Conclusiones

- De acuerdo al Objetivo específico N°1: “Determinar los requerimientos funcionales de Machine Learning para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto.”, se logró determinar los requerimientos funcionales para la construcción del algoritmo y funciones que están detallados en el Anexo 6, mediante la entrevista con el Dr. Pinto. Asimismo, para el desarrollo (ver punto 4.2. Metodología de desarrollo del Proyecto), se usó el lenguaje Python y la librería de PySpark para la creación del machine learning, a nivel de la aplicación web, el diseño se encuentra en el Anexo 8 y el diagrama de capas en el Anexo 10, para el frontend se trabajó con HTML, CSS, y JavaScript, en el backend, Node-RED (Runtime Node.js), y Cloudant como base de datos (servicio de IBM Cloud).
- Referente al Objetivo específico N°2: “Determinar el grado de exactitud de Machine Learning para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto.”, se logró determinar que el grado de exactitud es 100% (ver Figura 32), lo que significa que el machine learning clasifica correctamente los accidentes cerebrovasculares positivos y negativos. Además, la comparativa de los resultados mediante la prueba de Chi-cuadrado, evidenció la concordancia con el criterio médico. Cabe recordar que según la teoría, la exactitud es una métrica de desempeño que permite conocer el porcentaje de la cantidad de predicciones correctas del total real. En otras palabras, ayuda a detectar la frecuencia con la que el modelo de machine learning acierta las clases reales.

Tesis publicada con autorización del autor
No olvide citar esta tesis

- En relación al Objetivo específico N°3: “Determinar el grado de sensibilidad de Machine Learning para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto.”, se logró determinar que el grado de sensibilidad es 100% (ver Figura 32), es decir que el machine learning identifica adecuadamente los accidentes cerebrovasculares positivos, lo que coincide con la clase positiva real del criterio médico. Cabe resaltar que la sensibilidad es la métrica de desempeño que permite conocer el porcentaje de la cantidad de predicciones positivas correctas del total de clases reales positivas. En otros términos, esta métrica permite calcular la frecuencia con la que el modelo de machine learning predice una clase positiva, y esta coincide con la clase positiva real.

- Respecto al Objetivo específico N°4: “Determinar el grado de especificidad de Machine Learning para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto.”, se logró determinar que el grado de especificidad es 100% (ver Figura 32). Esto significa que el machine learning predice acertadamente los accidentes cerebrovasculares negativos, y estas coinciden con lo estipulado por el médico. Cabe indicar que, la especificidad es una métrica de desempeño que permite identificar el porcentaje de la cantidad de predicciones negativas clasificadas correctamente del total de clases reales negativas. Esto quiere decir que se calcula la frecuencia con la que el modelo de machine learning predice una clase negativa, y esta coincide con la clase negativa real.

Tesis publicada con autorización del autor
No olvide citar esta tesis

5.3. Recomendaciones

- Referente a la conclusión N°1 se recomienda evaluar el impacto de la implementación de una mejora al machine learning creado, usando los datos recopilados en la etapa de retroalimentación. Adicionalmente, se recomienda evaluar otros algoritmos de clasificación como el de Redes Neuronales, el cual gestiona el procesamiento de datos por capas, para optimizar las predicciones.
- Considerando que la conclusión N° 2, relacionada a la métrica de exactitud, se recomienda aumentar la cantidad de registros del dataset, ya que esto mejoraría la etapa de entrenamiento del machine learning, obteniendo métricas más precisas sobre su impacto en la realidad, y validando un margen de mejora, ante la utilización de un mayor número de casos.
- Para la conclusión N°3, relacionada a la métrica de sensibilidad, se recomienda variar la proporción entre los casos positivos de ACV y los casos negativos, mediante técnicas de ajuste en la distribución de las muestras.
- Según la conclusión N°4, relacionada a la métrica de especificidad, se recomienda, igual que el anterior, cambiar la distribución entre los casos de ACV.
- Teniendo en cuenta todas las conclusiones, se recomienda considerar nuevas variables de acuerdo al contexto médico donde se implemente la solución, ya que estas pueden mejorar el algoritmo y la precisión de predicción según los objetivos de cada entidad de salud, modificando este modelo para obtener procesos de inteligencia artificial.

Tesis publicada con autorización del autor
No olvide citar esta tesis

TESIS
UNIVERSIDAD CIENTÍFICA DEL SUR

REFERENCIAS BIBLIOGRÁFICAS

Adam, S. Y., Yousif, A., & Bashir, M. (Setiembre de 2016). Classification of Ischemic Stroke using Machine Learning Algorithms. *International Journal of Computer Applications, 149*(10), 26-31.

American Heart Association [AHA]. (2016). *Respuestas del Corazón - ¿Cómo puede la actividad física convertirse en un estilo de vida?* Obtenido de American Heart Association [AHA]: https://www.heart.org/-/media/data-import/downloadables/1/6/c/howcanphyactivitybecomewayoflife_span-ucm_316254.pdf

American Heart Association [AHA]. (2017). *Respuestas del Corazón - ¿Qué es la presión arterial alta?* Obtenido de American Heart Association [AHA]: https://www.heart.org/-/media/data-import/downloadables/4/c/5/whatishighbloodpressure_span-ucm_316246.pdf

American Heart Association [AHA] y American Stroke Association [ASA]. (2015). *Let's Talk About Stroke - Changes Caused by Stroke.* Obtenido de American Stroke Association [ASA]: https://www.stroke.org/-/media/stroke-files/stroke-resource-center/recovery/patient-focused/spost_changescaused_2015.pdf

Asadi, H., Dowling, R., Yan, B., & Mitchell, P. (10 de Febrero de 2014). Machine Learning for Outcome Prediction of Acute Ischemic Stroke Post Intra-Arterial Therapy. *PLOS ONE, 9*(2).

Bowles, M. (2015). *Machine Learning in Python: Essential Techniques for Predictive Analysis.* Indiana, Estados Unidos: John Wiley & Sons, Inc.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Brink, H., Richards, J., & Fetherolf, M. (2017). *Real-World Machine Learning.* Nueva York, Estados Unidos: Manning Publications Co.

Brownlee, J. (2016). *Master Machine Learning Algorithms: Discover How They Work and Implement Them From Scratch.* Machine Learning Mastery.

Cardador Cabello, A. L. (2015). *Implantación de aplicaciones web en entornos internet, intranet y extranet. IFCD0210.* Málaga, España: IC Editorial.

Cielen, D., Meysman, A., & Ali, M. (2016). *Introducing Data Science: Big Data, Machine Learning, and More, Using Python Tools.* Nueva York, Estados Unidos: Manning Publications Co.

Cohen, N., & Gómez Rojas, G. (2019). *Metodología de la Investigación, ¿para qué?: La Producción de los Datos y los Diseños.* Buenos Aires, Argentina: Editorial Teseo.

Domínguez-Rios, V., & López-Santillán, M. (2016). Teoría General de Sistemas, un enfoque práctico. *TECNOCIENCIA CHIHUAHUA, 10*(3), 125-132.

Géron, A. (2017). *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems.* California, Estados Unidos: O'Reilly Media, Inc.

Grus, J. (2015). *Data Science from Scratch: First Principles with Python.* California, Estados Unidos: O'Reilly Media, Inc.

Haider, M. (2015). *Getting Started with Data Science: Making Sense of Data with Analytics.* Indiana, Estados Unidos: IBM Press.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Heo, J., Yoon, J., Park, H., Kim, Y., Nam, H., & Heo, J. (20 de Marzo de 2019). Machine Learning–Based Model for Prediction of Outcomes in Acute Stroke. *AHA/ASA Journals, 50*(5), 1263-1265.

Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, M. (2014). *Metodología de la Investigación* (6ta ed.). Ciudad de México, México: McGraw Hill Interamericana Editores.

Hurwitz, J., & Kirsch, D. (2018). *Machine Learning For Dummies, IBM Limited Edition.* Nueva Jersey: John Wiley & Sons, Inc.

Jeena, R., & Kumar, S. (2018). Machine Intelligence in Stroke Prediction. *International Journal of Bioinformatics Research and Applications, 14*(1/2), 29-48.

Kasireddy, P. (17 de Diciembre de 2015). *How the Web Works Part II: Client-Server Model & the Structure of a Web Application*. Obtenido de FreeCodeCamp: https://www.freecodecamp.org/news/how-the-web-works-part-ii-client-server-model-the-structure-of-a-web-application-735b4b6d76e3/

Mayo Clinic. (15 de Enero de 2019). *Diabetes - Diagnóstico y Tratamiento*. Obtenido de Mayo Clinic: https://www.mayoclinic.org/es-es/diseases-conditions/diabetes/diagnosis-treatment/drc-20371451

Mayo Clinic. (12 de Julio de 2019). *Prueba de colesterol*. Obtenido de Mayo Clinic: https://www.mayoclinic.org/es-es/tests-procedures/cholesterol-test/about/pac-20384601

Ministerio de Salud del Perú [MINSA]. (Setiembre de 2018). *Análisis de las Causas de Mortalidad en el Perú, 1986-2015.* Lima, Perú.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Ministerio de Salud del Perú [MINSA]. (31 de Diciembre de 2018). *Resolución Ministerial N° 1367-2018/MINSA.* Lima, Perú: Plataforma Digital Única del Estado Peruano.

Müller, A., & Guido, S. (2016). *Introduction to Machine Learning with Python: A Guide for Data Scientists.* California, Estados Unidos: O'Reilly Media, Inc.

National Institute of Neurological Disorders and Stroke [NINDS]. (03 de Enero de 2017). *Accidente cerebrovascular: Esperanza en la investigación*. Obtenido de National Institute of Neurological Disorders and Stroke: https://espanol.ninds.nih.gov/trastornos/accidente_cerebrovascular.htm

National Institute of Neurological Disorders and Stroke [NINDS]. (2019). *Brain Basics: Preventing Stroke*. Obtenido de National Institute of Neurological Disorders and Stroke: https://www.ninds.nih.gov/Disorders/Patient-Caregiver-Education/Preventing-Stroke

Ordóñez Barrios, D. A., & Vizcarra Infantes, E. R. (31 de Julio de 2018). *Modelo Predictivo para el diagnóstico de la Diabetes Mellitus Tipo 2 soportado por SAP Predictive Analytics (Tesis de Pregrado, Universidad Peruana de Ciencias Aplicadas, Lima, Perú).* Obtenido de Repositorio Académico de la Universidad Peruana de Ciencias Aplicadas (UPC): http://hdl.handle.net/10757/624417

Organización Mundial de la Salud [OMS]. (Setiembre de 2013). *Información General sobre la Hipertensión en el Mundo.* Ginebra, Suiza: Organización Mundial de la Salud.

Organización Mundial de la Salud [OMS]. (Setiembre de 2016). Stroke: a global response is needed. *Bulletin of the World Health Organization, 94*(9), 634-634A.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Ozdemir, S. (2016). *Principles of Data Science.* Birmingham, Reino Unido: Packt Publishing Ltd.

Peralta, C. (2017). *Separata Tesis.* Lima.

Posadas Ruiz, L. E. (2018). *Factores de riesgo asociados a accidente cerebrovascular en el Hospital Edgardo Rebagliati Martins, 2016 (Tesis de Pregrado, Universidad Ricardo Palma, Lima, Perú).* Obtenido de Repositorio Institucional - URP: http://repositorio.urp.edu.pe/handle/urp/1154

Real Academia Española [RAE]. (2014). *Diccionario de la lengua española* (Vigesimotercera ed.). Madrid, España.

Repositorio Único Nacional de Información en Salud [REUNIS]. (Setiembre de 2018). *Mortalidad por Causas, Perú: 1986 - 2015*. Obtenido de Ministerio de Salud del Perú: https://www.minsa.gob.pe/reunis/Mortalidad/tabla_causa.html

Rollins, J. (2015). *Metodología Fundamental para la Ciencia de Datos.* Nueva York, Estados Unidos: IBM Analytics.

Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding Machine Learning: From Theory to Algorithms.* Nueva York, Estados Unidos: Cambridge University Press.

Theodoridis, S. (2015). *Machine Learning: A Bayesian and Optimization Perspective.* Estados Unidos: Academic Press.

U.S. Department of Health and Human Services y U.S. Department of Agriculture. (2015). *2015–2020 Dietary Guidelines for Americans* (8 ed.).

Tesis publicada con autorización del autor
No olvide citar esta tesis

TESIS
UNIVERSIDAD CIENTÍFICA DEL SUR

UNC Comprehensive Stroke Center. (26 de Setiembre de 2016). *Guía sobre Accidentes Cerebrovasculares: Un resumen para el paciente y sus cuidadores.* Obtenido de Facultad de Medicina de la Universidad de Carolina del Norte (UNC): https://www.med.unc.edu/neurology/files/2018/05/overview-of-stroke-for-patients-and-caregivers-spanish.pdf

Walker, M. (30 de Abril de 2019). Can Machine Learning and AI Head Off Heart Attacks and Strokes? *The Wall Street Journal*.

Tesis publicada con autorización del autor
No olvide citar esta tesis

TESIS
UNIVERSIDAD CIENTÍFICA DEL SUR

GLOSARIO DE TÉRMINOS Y ABREVIATURAS

- **ACV:** Accidente Cerebrovascular.
- **Accidente Cerebrovascular:** Es una enfermedad mortal e incapacitante, que sucede cuando la irrigación sanguínea disminuye o se corta debido a una obstrucción o hemorragia; pudiendo afectar gravemente el cerebro y su estructura.
- **Aplicación Web:** Es una solución tecnológica que interactúa con un servidor web y que es adaptable y accesible desde cualquier dispositivo con conexión a internet.
- **Conjunto de Datos o Dataset**: Es un grupo de datos que ha sido recopilado, procesado, y unificado en un solo archivo, para desarrollar modelos de machine learning. Además, este se puede fraccionar en diferentes subconjuntos para procesos de entrenamiento, pruebas, o validación.
- **Data Science:** Es un área de estudio que traslada problemas de la realidad a problemas computacionales para darles solución a partir del uso de datos, y la aplicación de otras áreas de estudio. Logrando así, sustraer nuevos conocimientos, y comprender mejor lo que nos rodea.
- **Factor de Riesgo:** Es una característica propia de una enfermedad presente o no en un individuo, indicando así, que tan propenso está a tener o no la enfermedad.
- **Label:** Etiqueta asignada para clasificación de datos. Puede ser binario o múltiple.
- **Machine Learning:** Es el área de estudio que utiliza algoritmos y modelos para identificar y extraer patrones de un conjunto de datos, a fin de hacer predicciones.
- **Modelo de Machine Learning:** Es la lógica resultante de un algoritmo de machine learning, después que este ha procesado los datos de entrenamiento.
- **Objetivos:** Son representaciones de lo que se quiere alcanzar. Se caracterizan por ser medibles por tiempo, porcentajes, niveles, etc.

Tesis publicada con autorización del autor
No olvide citar esta tesis

- **Predicción:** Es el resultado de un modelo de machine learning que se expone a nuevos datos en la realidad, es decir datos diferentes a los del conjunto de datos inicial.
- **Prevención de Accidente Cerebrovascular:** Es un proceso que comprende la consideración de los elementos de riesgo, para evitar un ictus y sus complicaciones, y conservar la integridad física y mental de una persona.

Tesis publicada con autorización del autor
No olvide citar esta tesis

ANEXOS

Tesis publicada con autorización del autor
No olvide citar esta tesis

Anexo 1. Matriz de Consistencia

Problema general	Objetivo general	Hipótesis general	MÉTODO
¿Cómo Machine Learning influye para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto?	Determinar el nivel de influencia de Machine Learning para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto.	Machine Learning influye significativamente para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto.	**Metodología:** Diseño Experimental Puro. Método Hipotético Deductivo. **Tipo:** Correlacional. **Enfoque:** Mixto **Población:** 520 pacientes. **Muestra:** 221 pacientes. **Técnica de recolección de datos:** Métricas de Desempeño, Ficha Técnica de Entrevista, y Revisión de Historias Médicas. **Técnica de procesamiento de datos:** Python con librerías pandas, matplotlib, y seaborn.
Problema Especifico	**Objetivo Especifico**	**Hipótesis Específicos**	
¿Cuáles son los requerimientos funcionales de Machine Learning para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto?	Determinar los requerimientos funcionales de Machine Learning para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto.	Los requerimientos funcionales de Machine Learning influyen significativamente para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto.	
¿Cuál es el grado de exactitud de Machine Learning para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto?	Determinar el grado de exactitud de Machine Learning para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto.	El grado de exactitud de Machine Learning influye significativamente para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto.	
¿Cuál es el grado de sensibilidad de Machine Learning para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto?	Determinar el grado de sensibilidad de Machine Learning para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto.	El grado de sensibilidad de Machine Learning influye significativamente para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto.	
¿Cuál es el grado de especificidad de Machine Learning para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto?	Determinar el grado de especificidad de Machine Learning para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto.	El grado de especificidad de Machine Learning influye significativamente para predecir el riesgo de Accidente Cerebrovascular en pacientes del Consultorio del Dr. Pinto.	

Fuente: Elaboración Propia

Tesis publicada con autorización del autor
No olvide citar esta tesis

TESIS
UNIVERSIDAD CIENTÍFICA DEL SUR

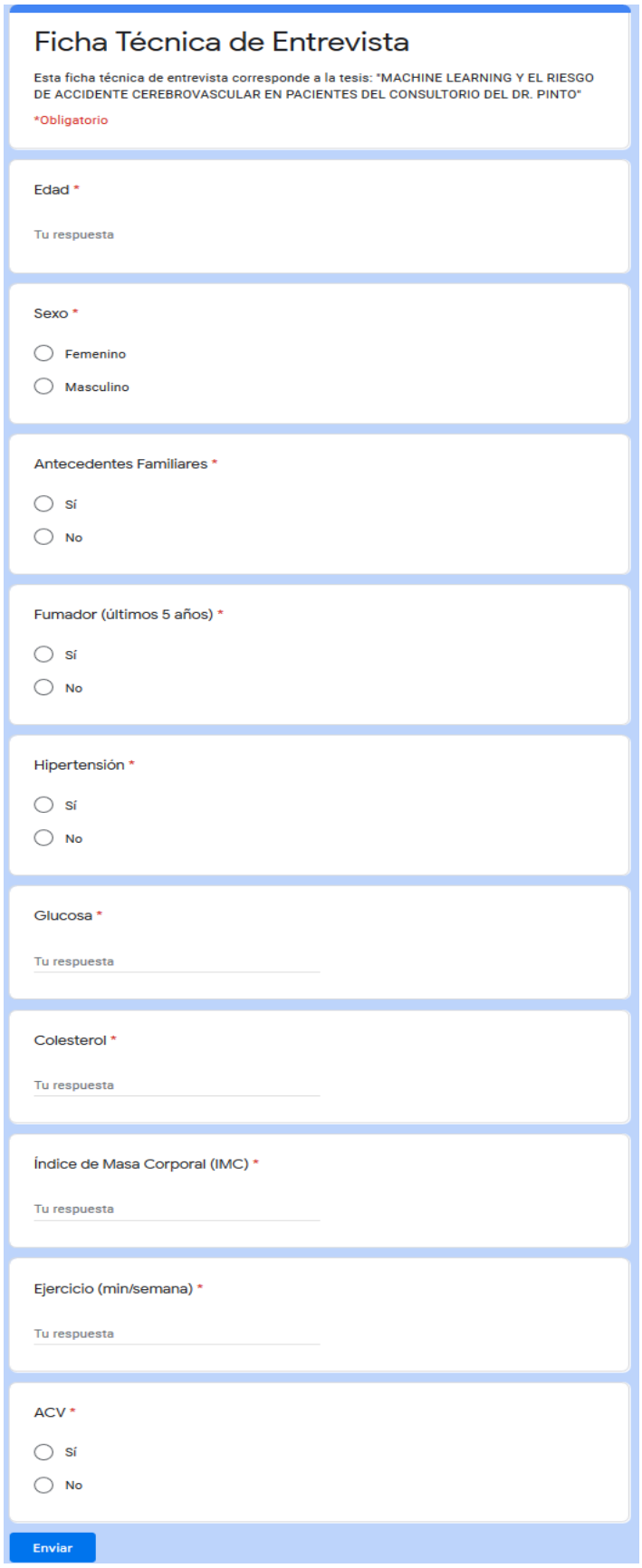
Anexo 2. Matriz de Operacionalización

VARIABLES	DEFINICIÓN	DIMENSIÓN	SUBDIMENSIÓN	INDICADOR	ESCALA	INSTRUMENTO
Variable Independiente: Machine Learning	El nombre "Machine Learning" indica aprender de los datos, a través del desarrollo de un modelo, el cual luego se puede usar para hacer predicciones. (Theodoridis, 2015)	**Dimensión 1:** Algoritmo de Machine Learning en la Aplicación Web	Funcionalidad	Requerimientos funcionales	Nominal	Ficha de Requerimientos Técnicos
		Dimensión 2: Datos para Modelo de Machine Learning	Exactitud	Porcentaje total de ACV clasificados correctamente	Razón	Ficha Técnica de Entrevista - Métricas de Desempeño
			Sensibilidad	Porcentaje de ACV clasificados como "SI"		
			Especificidad	Porcentaje de ACV clasificados como "NO"		
Variable Dependiente: Riesgo de Accidente Cerebrovascular	Un Accidente Cerebrovascular sucede cuando la circulación sanguínea a una zona del cerebro se interrumpe, produciendo que las células cerebrales mueran por la falta de oxígeno e irrigación (National Institute of Neurological Disorders and Stroke [NINDS], 2019).	**Dimensión 1:** Prevención	Control de Factores de Riesgo	Límites de Control Sobre Estándares Bases	Razón	Reporte

Fuente: Elaboración Propia

Tesis publicada con autorización del autor
No olvide citar esta tesis

Anexo 3. Instrumento de Evaluación



Ficha Técnica de Entrevista

Esta ficha técnica de entrevista corresponde a la tesis: "MACHINE LEARNING Y EL RIESGO DE ACCIDENTE CEREBROVASCULAR EN PACIENTES DEL CONSULTORIO DEL DR. PINTO"

*Obligatorio

Edad *

Tu respuesta

Sexo *

- Femenino
- Masculino

Antecedentes Familiares *

- Sí
- No

Fumador (últimos 5 años) *

- Sí
- No

Hipertensión *

- Sí
- No

Glucosa *

Tu respuesta

Colesterol *

Tu respuesta

Índice de Masa Corporal (IMC) *

Tu respuesta

Ejercicio (min/semana) *

Tu respuesta

ACV *

- Sí
- No

Enviar

Tesis publicada con autorización del autor
No olvide citar esta tesis

TESIS
UNIVERSIDAD CIENTÍFICA DEL SUR

Anexo 4. Ficha de Validación del Instrumento por Jueces Expertos

Juez Experto #1

Lima, 07 de Diciembre del 2019

Presente

Estimado Magister:

Siendo conocedor de su trayectoria académica y profesional, me he tomado la libertad de elegirlo como JUEZ EXPERTO para revisar el contenido de la ficha técnica de entrevista, que se pretende utilizar para evaluar el instrumento sobre la Tesis de: MACHINE LEARNING Y EL RIESGO DE ACCIDENTE CEREBROVASCULAR EN PACIENTES DEL CONSULTORIO DEL DR. PINTO.

A continuación le presento una lista de ítems, relacionadas a cada concepto teórico. Lo que se le solicita es marcar con una X el grado de pertenencia de cada ítem con su respectivo concepto, de acuerdo a su propia experiencia y visión profesional. No se le pide que responda las preguntas de cada área, sino que indique si cada pregunta es apropiada o congruente con el concepto o variable que se pretende medir.

Los resultados de esta evaluación, servirá para determinar los coeficientes de validez de contenido del presente cuestionario. De ante mano le agradezco su cooperación.



Firma del Tesista
Nombre: Nathalia Contreras Siu
DNI: 70246162

Tesis publicada con autorización del autor
No olvide citar esta tesis

FICHA DE EVALUACIÓN DE INSTRUMENTO

Nombre del Especialista:	William Pinto Zeballos
Sexo:	Varón (X) Mujer ( )
Edad:	(52) años
Profesión o Especialidad:	Médico
Años de Experiencia Laboral:	25

EVALUACIÓN DEL INSTRUMENTO

Favor de evaluar la definición de conceptos y pertinencia de cada ítem:

1. ***Dimensión:*** Datos para Modelo de Machine Learning

   Descripción: Refiere a los factores de riesgo (modificables y no modificables) a considerar en la prevención de un accidente cerebrovascular.

   Indicador: Porcentaje de Accidente Cerebrovascular (ACV)

N	Ítems	Escala	¿Es pertinente con el concepto?		¿La escala utilizada es conveniente?		Comentario
			SI	NO	SI	NO	
1	Edad	Ordinal	X		X		
2	Sexo	Nominal	X		X		
3	Antecedentes Familiares	Nominal	X		X		
4	Fumador (últimos 5 años)	Nominal	X		X		
5	Hipertensión	Nominal	X		X		
6	Glucosa	Ordinal	X		X		
7	Colesterol	Ordinal	X		X		
8	Índice de Masa Corporal (IMC)	Ordinal	X		X		
9	Ejercicio (min/semana)	Ordinal	X		X		
10	ACV	Nominal	X		X		

DR. WILLIAM PINTO ZEBALLOS
Medicina Interna - Geriatría
CMP 26657 RNE 13468

Firma _____

Nombre del Evaluador: William Pinto Zeballos

Tesis publicada con autorización del autor
No olvide citar esta tesis

Juez Experto #2

Lima, 07 de Diciembre del 2019

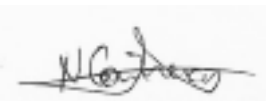
Presente

Estimado Magister:

Siendo conocedor de su trayectoria académica y profesional, me he tomado la libertad de elegirlo como JUEZ EXPERTO para revisar el contenido de la ficha técnica de entrevista, que se pretende utilizar para evaluar el instrumento sobre la Tesis de: MACHINE LEARNING Y EL RIESGO DE ACCIDENTE CEREBROVASCULAR EN PACIENTES DEL CONSULTORIO DEL DR. PINTO.

A continuación le presento una lista de ítems, relacionadas a cada concepto teórico. Lo que se le solicita es marcar con una X el grado de pertenencia de cada ítem con su respectivo concepto, de acuerdo a su propia experiencia y visión profesional. No se le pide que responda las preguntas de cada área, sino que indique si cada pregunta es apropiada o congruente con el concepto o variable que se pretende medir.

Los resultados de esta evaluación, servirá para determinar los coeficientes de validez de contenido del presente cuestionario. De ante mano le agradezco su cooperación.



Firma del Tesista
Nombre: Nathalia Contreras Siu
DNI: 70246162

Tesis publicada con autorización del autor
No olvide citar esta tesis

FICHA DE EVALUACIÓN DE INSTRUMENTO

Nombre del Especialista:	Juan Trujillo Martinez
Sexo:	Varón (X) Mujer ( )
Edad:	(50) años
Profesión o Especialidad:	Médico
Años de Experiencia Laboral:	20

EVALUACIÓN DEL INSTRUMENTO

Favor de evaluar la definición de conceptos y pertinencia de cada ítem:

1. ***Dimensión:*** Datos para Modelo de Machine Learning

   Descripción: Refiere a los factores de riesgo (modificables y no modificables) a considerar en la prevención de un accidente cerebrovascular.

   Indicador: Porcentaje de Accidente Cerebrovascular (ACV)

N	Items	Escala	¿Es pertinente con el concepto?		¿La escala utilizada es conveniente?		Comentario
			SI	NO	SI	NO	
1	Edad	Ordinal	X		X		
2	Sexo	Nominal	X		X		
3	Antecedentes Familiares	Nominal	X		X		
4	Fumador (últimos 5 años)	Nominal	X		X		
5	Hipertensión	Nominal	X		X		
6	Glucosa	Ordinal	X		X		
7	Colesterol	Ordinal	X		X		
8	Índice de Masa Corporal (IMC)	Ordinal	X		X		
9	Ejercicio (min/semana)	Ordinal	X		X		
10	ACV	Nominal	X		X		

Firma
Nombre del Evaluador: Juan Trujillo Martinez



Tesis publicada con autorización del autor
No olvide citar esta tesis

Juez Experto #3

Lima, 07 de Diciembre del 2019

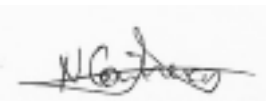
Presente

Estimado Magister:

Siendo conocedor de su trayectoria académica y profesional, me he tomado la libertad de elegirlo como JUEZ EXPERTO para revisar el contenido de la ficha técnica de entrevista, que se pretende utilizar para evaluar el instrumento sobre la Tesis de: MACHINE LEARNING Y EL RIESGO DE ACCIDENTE CEREBROVASCULAR EN PACIENTES DEL CONSULTORIO DEL DR. PINTO.

A continuación le presento una lista de ítems, relacionadas a cada concepto teórico. Lo que se le solicita es marcar con una X el grado de pertenencia de cada ítem con su respectivo concepto, de acuerdo a su propia experiencia y visión profesional. No se le pide que responda las preguntas de cada área, sino que indique si cada pregunta es apropiada o congruente con el concepto o variable que se pretende medir.

Los resultados de esta evaluación, servirá para determinar los coeficientes de validez de contenido del presente cuestionario. De ante mano le agradezco su cooperación.



Firma del Tesista
Nombre: Nathalia Contreras Siu
DNI: 70246162

Tesis publicada con autorización del autor
No olvide citar esta tesis

FICHA DE EVALUACIÓN DE INSTRUMENTO

Nombre del Especialista:	Mirian Huamán Torres
Sexo:	Varón ( ) Mujer (X)
Edad:	(46) años
Profesión o Especialidad:	Enfermera con Especialidad en Emergencias y Desastres
Años de Experiencia Laboral:	25 años

EVALUACIÓN DEL INSTRUMENTO

Favor de evaluar la definición de conceptos y pertinencia de cada ítem:

1. ***Dimensión:*** Datos para Modelo de Machine Learning

   Descripción: Refiere a los factores de riesgo (modificables y no modificables) a considerar en la prevención de un accidente cerebrovascular.

   Indicador: Porcentaje de Accidente Cerebrovascular (ACV)

N	Ítems	Escala	¿Es pertinente con el concepto?		¿La escala utilizada es conveniente?		Comentario
			SI	NO	SI	NO	
1	Edad	Ordinal	X		X		
2	Sexo	Nominal	X		X		
3	Antecedentes Familiares	Nominal	X		X		
4	Fumador (últimos 5 años)	Nominal	X		X		
5	Hipertensión	Nominal	X		X		
6	Glucosa	Ordinal	X		X		
7	Colesterol	Ordinal	X		X		
8	Índice de Masa Corporal (IMC)	Ordinal	X		X		
9	Ejercicio (min/semana)	Ordinal	X		X		
10	ACV	Nominal	X		X		

Firma

Nombre del Evaluador: Mirian Huamán Torres (CEP: 28761)

Tesis publicada con autorización del autor
No olvide citar esta tesis

TESIS
UNIVERSIDAD CIENTÍFICA DEL SUR

Anexo 5. Autorización de la Institución donde se realizó la Investigación

Lima, 02 de Noviembre del 2019

Señores Autoridades
Universidad Científica del Sur
Facultad de Ciencias Empresariales

CARTA DE AUTORIZACIÓN

Estimados Señores:

Me es grato extenderles un cordial saludo y comunicarles que la alumna **NATHALIA ELENA CONTRERAS SIU** con DNI N° **70246162**, de la carrera de **INGENIERÍA DE SISTEMAS EMPRESARIALES** en la Universidad Científica del Sur, cuenta con la debida autorización para desarrollar su tesis en mi consultorio médico para que obtenga la información necesaria que le permita desarrollar su trabajo profesional.

Además, se precisa que el uso de la información que solicite es con los fines relacionados a la sustentación de su tesis y no se mostrará o divulgará información para temas no educativos.

Atentamente,

DR. WILLIAM PINTO ZEBALLOS
Medicina Interna - Geriatría
CMP 25862 RNE 2:583

Nombre y Apellidos: William Pinto Zeballos
Cargo: Director Médico

Tesis publicada con autorización del autor
No olvide citar esta tesis

Anexo 6. Ficha de Requerimientos Técnicos

Control de Versiones					
Versión	Hecho Por	Revisado Por	Aprobado Por	Fecha	Motivo
0.1	Nathalia Contreras	Dr. Pinto	Dr. Pinto	06/03/20	Inicio
0.2	Nathalia Contreras	Dr. Pinto	Dr. Pinto	12/03/20	Actualización

FICHA DE REQUERIMIENTOS TÉCNICOS

NOMBRE DEL PROYECTO
Machine Learning y El Riesgo de Accidente Cerebrovascular en pacientes del consultorio del Dr. Pinto

Objetivos

• **Objetivo del Consultorio Médico:** Determinar el nivel de influencia de Machine Learning para prevenir el riesgo de Accidente Cerebrovascular en pacientes del Distrito de Miraflores.
• **Objetivo de la Solución:**
  o Crear un modelo de machine learning basándose en los factores de riesgo de la enfermedad
  o Crear una aplicación web que esté conectado al modelo de machine learning, para la mejor interacción con el usuario.

Requerimientos Funcionales del Sistema	
Código	**Descripción**
REQF01	Creación de un modelo machine learning para el procesamiento de datos.
REQF02	Creación de una aplicación web para interactuar con el machine learning.
REQF03	Los datos deben ser almacenados en una base de datos.
REQF04	Registro de datos y búsqueda de resultados mediante ID de historia médica.

Requerimientos No Funcionales del Sistema	
Código	**Descripción**
REQNF01	Accesibilidad desde el navegador de la PC y desde dispositivos móviles.
REQNF02	Tiempo de respuesta de la aplicación web, no mayor a 5 segundos.

Reglas del Negocio
• Priorizar la salud y bienestar de los pacientes. • Brindar el mejor tratamiento según condición médica detectada. • Diagnosticar y prevenir a tiempo complicaciones médicas.

Tesis publicada con autorización del autor
No olvide citar esta tesis

Control de Versiones					
Versión	Hecho Por	Revisado Por	Aprobado Por	Fecha	Motivo
0.1	Nathalia Contreras	Dr. Pinto	Dr. Pinto	06/03/20	Inicio
0.2	Nathalia Contreras	Dr. Pinto	Dr. Pinto	12/03/20	Actualización

Componentes de la Solución

- **Machine Learning:**
  - ***Lenguaje de Programación:*** Python
  - ***Librerías:*** PySpark (ml, sql)
  - ***Deployment:*** Watson Machine Learning
- **Aplicación Web:**
  - ***Herramienta:*** Node-RED (Runtime Node.js)
  - ***Lenguaje de Programación:*** JavaScript
  - ***Presentación:*** HTML, CSS
  - ***Base de Datos:*** Cloudant
- La solución estará alojada en la plataforma de IBM Cloud.

Criterios de Aceptación	
Concepto	**Descripción**
Técnicos	• La solución se debe desarrollar según los requerimientos estipulados. • La exactitud de la solución debe ser 80% o más.
Calidad	Funcionamiento ininterrumpido durante jornada laboral (los 7 días de la semana).
Administrativo	Aprobación del Dr. William Pinto.

Aprobación del Entregable		
Nombre	**Fecha**	**Firma**
William Pinto	16/03/2020	DR. WILLIAM PINTO ZEBALLOS Medicina Interna - Geriatría CMP 29862 RNE 15588

Tesis publicada con autorización del autor
No olvide citar esta tesis

TESIS
UNIVERSIDAD CIENTÍFICA DEL SUR

Anexo 7.Tabla de Datos Recopilados para el proyecto

N°	EDAD	SEXO	HIPERTENSION	GLUCOSA	COLESTEROL	IMC	HISTFAM	FUMAR	EJERCMINXSEM	ACV
1	65	F	S	89	191	26	S	N	55	S
2	30	M	S	84	201	31	N	N	80	S
3	30	F	S	77	200	20	N	N	86	S
4	66	F	S	124	159	27	S	N	184	S
5	30	F	N	73	164	31	N	N	182	N
6	61	M	N	81	156	24	N	N	31	N
7	29	M	S	83	184	30	S	N	187	S
8	41	F	N	79	238	21	N	N	178	N
9	49	M	N	100	153	32	N	N	244	N
10	47	M	N	97	232	27	N	N	231	N
11	44	F	N	103	230	24	N	N	77	N
12	60	F	N	90	195	30	N	N	212	N
13	65	F	N	86	215	26	N	N	114	N
14	39	M	S	80	217	23	S	N	105	S
15	72	F	S	87	195	29	S	S	93	S
16	55	M	S	121	192	26	S	N	81	S
17	42	F	S	113	174	34	S	N	104	S
18	60	M	S	85	240	23	S	N	93	S
19	30	M	S	83	217	24	N	N	139	S
20	37	F	S	106	163	24	S	N	101	S
21	56	M	S	111	205	32	S	N	33	S
22	66	M	S	119	242	23	N	N	34	S
23	42	F	N	80	227	24	N	N	131	N
24	33	F	N	80	197	23	N	N	215	N
25	41	F	S	81	184	22	S	N	67	S
26	43	F	S	74	234	28	N	N	159	S
27	32	F	S	89	153	25	N	S	123	S
28	52	F	N	83	185	30	N	N	249	N
29	67	F	N	89	239	30	N	N	102	N
30	51	F	S	131	163	28	N	N	87	S
31	44	F	S	130	191	23	N	N	162	S
32	30	F	S	83	157	32	S	N	19	S
33	47	M	S	127	175	32	S	N	143	S
34	58	F	S	108	225	32	S	N	140	S
35	72	M	N	73	172	31	N	N	10	N
36	54	M	S	103	229	34	S	N	134	S
37	56	M	S	83	157	32	N	N	16	S
38	30	F	S	102	162	28	N	N	143	S
39	40	F	S	93	205	24	S	N	4	S
40	34	M	S	111	239	27	S	N	105	S
41	51	F	N	81	237	25	N	N	70	N
42	38	F	N	105	164	27	N	N	171	N
43	50	F	S	114	163	28	S	N	121	S
44	33	F	S	121	187	25	S	N	129	S
45	57	M	N	91	176	21	N	N	142	N
46	66	F	N	85	198	32	N	N	27	N
47	54	F	N	83	200	22	N	N	162	N
48	62	F	S	76	163	28	S	N	195	S
49	59	F	N	103	186	30	N	N	228	N
50	39	F	S	152	218	23	S	N	157	S
51	39	F	N	84	232	30	N	N	241	N
52	53	F	S	77	178	28	N	N	138	S
53	39	M	N	113	220	28	N	N	26	N
54	37	F	S	85	160	34	N	N	126	S
55	54	M	N	87	238	27	N	N	125	N
56	58	M	S	75	233	25	N	N	146	S
57	33	F	N	109	210	21	N	N	138	N
58	28	M	N	87	174	21	N	N	70	N
59	33	M	S	131	204	32	S	N	107	S
60	31	F	N	74	227	29	N	N	70	N
61	55	F	N	77	207	26	N	N	145	N
62	60	F	S	95	155	34	S	N	193	S
63	37	M	N	72	239	24	N	S	58	N
64	44	M	S	128	217	24	S	N	12	S
65	49	M	S	75	212	27	N	N	247	S
66	28	M	S	108	178	24	S	N	133	S
67	63	F	S	86	220	25	N	S	181	S
68	45	F	N	81	232	23	N	N	42	N
69	39	F	N	75	184	20	N	N	3	N
70	39	F	N	94	176	20	N	N	236	N
71	54	F	S	90	178	26	S	N	2	S
72	54	F	N	90	188	28	N	N	171	N
73	34	M	N	78	228	20	N	N	123	N
74	48	F	S	122	199	25	N	N	136	S
75	29	M	S	108	176	33	S	N	179	S
76	41	M	S	83	240	25	S	N	108	S
77	50	M	S	79	176	22	N	N	40	S
78	40	F	N	114	150	29	N	N	19	N
79	54	F	S	99	195	29	S	N	126	S
80	71	M	N	74	220	31	N	N	224	N
81	61	M	S	110	150	26	S	N	106	S
82	41	M	N	91	151	28	N	N	66	N
83	71	F	N	101	173	31	N	N	116	N
84	43	M	N	80	219	24	N	N	9	N
85	41	M	S	108	153	28	S	N	85	S
86	30	M	S	84	152	27	S	N	116	S
87	28	F	S	107	204	27	N	N	107	S
88	52	F	S	117	214	24	S	N	3	S
89	61	F	S	89	161	25	N	N	38	S
90	53	F	S	80	185	25	N	N	112	S
91	42	M	S	86	152	27	S	N	87	S
92	30	F	S	94	212	29	N	N	169	S
93	58	M	S	106	199	32	N	N	163	S
94	52	M	N	84	229	22	N	N	36	N
95	32	F	S	95	225	24	S	N	158	S
96	66	F	N	89	200	32	N	N	162	N
97	31	F	N	103	191	22	N	N	203	N
98	29	F	N	77	193	29	N	N	127	N
99	52	M	S	82	205	22	S	N	17	S
100	63	M	S	72	192	27	N	N	122	S
101	31	M	S	76	204	22	N	N	134	S
102	69	F	S	85	215	31	S	N	103	S
103	60	F	S	95	224	30	S	N	178	S
104	52	F	S	108	228	25	N	N	78	S

Tesis publicada con autorización del autor
No olvide citar esta tesis

TESIS
UNIVERSIDAD CIENTÍFICA DEL SUR

105	54	M	S	136	201	23	S	N	149	S
106	34	F	S	89	207	22	N	N	130	S
107	28	M	S	91	154	31	N	N	147	S
108	60	M	N	87	161	29	N	N	239	N
109	69	M	S	82	238	27	N	N	194	S
110	29	F	S	108	166	21	N	N	91	S
111	50	M	S	157	154	25	S	N	45	S
112	61	M	N	88	208	20	N	N	128	N
113	46	M	N	89	211	29	S	N	33	N
114	55	F	N	106	214	22	N	N	212	N
115	38	M	S	81	236	32	N	N	113	S
116	67	F	S	122	191	22	N	N	120	S
117	58	F	N	79	231	21	N	N	189	N
118	43	M	N	72	224	24	N	N	45	N
119	63	F	S	70	172	34	N	N	175	S
120	38	M	N	71	168	31	N	N	33	N
121	49	M	N	93	199	30	N	N	228	N
122	52	M	N	110	169	33	N	N	21	N
123	52	M	N	110	183	34	S	N	33	N
124	46	M	S	75	228	23	N	N	203	S
125	29	M	S	128	179	23	S	N	159	S
126	41	F	S	87	171	30	S	N	138	S
127	53	F	S	97	182	25	S	N	149	S
128	48	M	S	84	214	23	S	N	83	S
129	38	F	S	92	220	24	N	N	91	S
130	51	M	S	107	189	32	N	N	124	S
131	31	M	S	85	150	23	N	S	93	S
132	44	M	S	85	218	31	S	N	31	S
133	43	M	S	157	180	34	S	N	21	S
134	64	F	S	84	171	33	N	N	198	S
135	37	M	S	87	156	29	N	N	180	S
136	31	M	N	72	175	28	N	N	233	N
137	54	F	S	114	234	30	S	N	197	S
138	59	F	S	93	208	23	S	S	5	S
139	43	M	S	88	152	22	S	N	42	S
140	50	M	S	100	204	31	S	N	133	S
141	55	M	S	131	209	30	N	N	67	S
142	68	M	S	85	191	29	S	N	15	S
143	34	F	N	83	180	28	S	N	150	N
144	33	F	N	81	206	30	N	N	125	N
145	57	F	S	113	208	32	S	N	16	S
146	70	F	S	85	234	32	N	N	15	S
147	31	F	S	72	181	26	N	N	89	S
148	68	F	S	81	200	21	N	N	82	S
149	68	F	N	110	194	30	N	N	221	N
150	41	F	N	85	194	26	N	N	5	N
151	53	M	S	118	182	29	N	N	31	S
152	58	M	N	77	200	23	N	N	41	N
153	58	F	S	70	218	32	S	S	38	S
154	63	M	S	155	158	28	S	N	85	S
155	37	M	S	73	189	31	N	N	163	S
156	34	F	N	108	157	28	N	N	151	N
157	49	M	S	84	207	26	S	N	85	S
158	60	F	N	82	169	30	N	N	118	N
159	57	F	S	94	173	26	N	N	54	S
160	47	M	N	98	226	24	S	N	163	N
161	52	M	N	73	174	21	N	N	102	N
162	65	M	N	76	225	26	N	N	139	N
163	34	F	N	87	158	27	N	N	246	N
164	39	M	N	91	185	26	N	N	170	N
165	65	M	S	92	195	26	N	N	92	S
166	32	F	N	100	168	29	N	S	224	N
167	37	M	S	86	215	28	N	N	212	S
168	63	F	S	81	157	34	S	N	164	S
169	28	F	S	160	217	25	S	N	180	S
170	30	F	S	108	162	30	S	N	6	S
171	33	M	S	146	205	28	N	N	19	S
172	44	F	N	88	191	30	N	N	55	N
173	67	F	N	85	203	20	N	N	25	N
174	55	M	N	85	222	22	N	N	217	N
175	65	M	S	93	163	32	S	N	239	S
176	41	F	S	87	176	34	S	N	96	S
177	28	F	N	79	169	30	S	N	181	N
178	58	F	N	106	162	30	S	N	124	N
179	72	F	N	82	232	25	N	N	35	N
180	50	F	N	76	188	29	N	N	5	N
181	52	M	N	120	238	26	N	N	63	N
182	56	F	N	111	236	31	N	N	74	N
183	28	F	S	91	228	24	N	N	130	S
184	44	M	S	82	206	22	S	N	142	S
185	67	M	S	117	161	22	S	N	174	S
186	61	M	N	115	207	24	N	N	121	N
187	52	F	S	139	195	22	S	S	135	S
188	65	M	S	101	184	22	N	N	82	S
189	34	F	N	116	237	32	N	N	73	N
190	60	M	N	90	163	23	N	N	116	N
191	63	M	S	84	152	26	N	N	5	S
192	64	M	N	81	179	30	N	N	74	N
193	64	M	S	90	207	23	S	N	198	S
194	41	M	S	122	169	23	S	N	28	S
195	57	M	S	113	242	31	N	N	175	S
196	68	M	S	115	196	33	S	N	30	S
197	42	M	N	116	159	25	N	N	123	N
198	66	F	S	124	176	23	S	N	181	S
199	72	F	N	81	206	32	N	N	119	N
200	39	F	S	85	173	31	S	N	132	S
201	55	M	N	72	196	21	N	N	184	N
202	70	M	S	76	190	27	N	N	195	S
203	49	M	S	94	179	21	N	N	171	S
204	69	F	S	91	223	20	N	N	38	S
205	44	M	N	136	233	26	N	N	195	N
206	56	F	S	72	174	31	S	N	150	S
207	47	F	N	124	169	20	S	N	47	N
208	51	M	N	85	182	21	N	N	18	N

Tesis publicada con autorización del autor
No olvide citar esta tesis

TESIS
UNIVERSIDAD CIENTÍFICA DEL SUR

209	60	M	S	76	156	34	N	N	167	S
210	41	M	S	73	166	34	S	N	30	S
211	69	M	N	81	220	24	N	N	64	N
212	38	F	N	110	180	32	N	N	32	N
213	29	M	S	91	154	21	N	N	203	S
214	43	M	S	70	230	26	N	N	52	S
215	39	F	S	102	190	34	N	S	43	S
216	45	M	N	80	204	26	N	N	174	N
217	32	F	N	100	207	25	S	N	76	N
218	66	F	S	98	164	22	S	N	139	S
219	50	M	S	137	174	27	S	N	193	S
220	54	M	S	78	166	29	N	N	234	S
221	65	F	S	123	154	23	N	N	185	S
222	30	F	N	88	215	26	N	N	123	N
223	41	F	S	78	222	32	N	N	19	S
224	34	M	S	126	212	23	S	N	94	S
225	61	F	S	147	213	23	S	N	21	S
226	49	F	S	107	169	32	S	N	35	S
227	35	M	S	93	165	23	S	N	181	S
228	35	M	S	126	166	33	N	N	145	S
229	43	M	S	102	232	28	S	N	20	S
230	33	M	S	91	180	26	S	S	104	S
231	30	M	S	103	182	31	N	N	182	S
232	47	M	S	89	230	28	N	N	42	S
233	67	F	S	110	200	27	N	N	28	S
234	46	F	N	79	171	25	N	N	238	N
235	52	F	N	113	219	21	N	N	187	N
236	72	F	S	92	234	29	N	N	100	S
237	33	M	N	112	210	20	S	N	196	N
238	66	M	S	75	241	30	S	N	43	S
239	56	M	N	76	201	23	N	N	197	N
240	34	M	N	104	214	22	N	N	100	N
241	41	F	N	100	173	27	N	N	31	N
242	62	M	N	88	171	23	N	N	75	N
243	64	M	S	106	197	27	S	N	84	S
244	46	M	S	123	216	34	S	N	87	S
245	29	F	S	97	179	26	S	N	189	S
246	48	F	S	73	219	24	N	N	51	S
247	28	F	S	101	188	21	N	N	226	S
248	57	M	S	96	203	27	S	S	118	S
249	70	M	S	82	214	30	S	N	140	S
250	36	F	N	78	176	25	N	N	165	N
251	44	M	S	109	190	22	N	N	125	S
252	49	F	N	74	168	24	N	N	92	N
253	47	F	S	100	189	28	N	N	169	S
254	38	F	S	144	237	23	S	S	85	S
255	36	M	S	91	157	26	S	N	63	S
256	61	M	N	102	182	32	N	N	221	N
257	32	F	N	91	176	20	S	N	216	N
258	30	F	N	92	236	30	N	N	108	N
259	50	F	S	122	241	23	N	N	101	S
260	28	M	S	86	229	23	N	N	194	S
261	50	F	N	77	235	27	N	N	23	N
262	39	M	N	89	168	23	N	N	12	N
263	38	M	S	90	167	32	S	N	145	S
264	43	F	S	89	240	26	S	N	187	S
265	67	M	N	100	186	22	N	N	88	N
266	28	F	S	101	163	24	S	N	125	S
267	64	M	N	95	192	28	S	N	208	N
268	57	M	S	110	215	22	N	N	96	S
269	59	F	S	76	241	23	S	N	103	S
270	48	F	S	70	219	23	S	N	25	S
271	69	F	S	80	231	29	S	N	120	S
272	29	M	S	97	217	32	N	N	119	S
273	71	M	S	120	198	32	N	N	81	S
274	48	M	N	76	220	29	N	N	133	N
275	52	M	N	74	221	31	N	N	171	N
276	45	F	N	75	189	26	N	N	7	N
277	64	M	S	85	238	34	S	N	50	S
278	63	F	S	100	152	31	S	S	67	S
279	49	M	S	109	214	23	S	N	152	S
280	32	F	N	80	202	27	S	N	31	N
281	59	M	S	107	240	28	S	S	132	S
282	59	M	S	94	204	22	S	N	161	S
283	53	M	S	107	172	24	S	N	162	S
284	45	M	S	124	173	28	N	N	21	S
285	71	M	S	102	182	23	N	N	246	S
286	72	M	N	91	226	23	N	N	98	N
287	31	M	N	108	178	21	N	N	44	N
288	58	M	N	93	189	31	S	N	82	N
289	70	M	N	91	227	26	N	N	150	N
290	54	M	S	70	151	31	S	N	81	S
291	46	F	S	103	190	31	S	N	195	S
292	38	M	N	120	197	29	N	N	165	N
293	57	F	S	84	222	26	N	N	9	S
294	37	F	S	156	203	25	N	N	77	S
295	39	M	S	93	199	24	N	N	35	S
296	54	M	S	114	208	27	N	N	116	S
297	60	M	S	108	156	26	S	N	31	S
298	65	M	N	124	222	30	N	N	66	N
299	51	M	S	89	213	26	N	N	64	S
300	51	M	N	102	175	30	N	N	227	N
301	60	M	S	144	183	31	S	N	69	S
302	28	F	S	99	223	29	N	N	121	S
303	55	M	N	101	209	27	N	N	199	N
304	40	M	S	102	211	27	N	N	102	S
305	72	F	N	97	154	28	N	N	72	N
306	41	F	N	84	214	23	N	N	24	N
307	30	M	S	140	226	28	S	N	122	S
308	51	F	N	96	157	29	N	N	129	N
309	46	F	N	77	166	23	N	N	7	N
310	62	M	N	76	201	20	N	N	165	N
311	70	M	S	85	154	22	S	N	53	S
312	49	F	S	91	197	24	N	N	58	S

Tesis publicada con autorización del autor
No olvide citar esta tesis

TESIS
UNIVERSIDAD CIENTÍFICA DEL SUR

313	38	M	S	117	153	26	S	S	120	S
314	54	F	S	114	231	28	S	N	64	S
315	55	M	N	71	174	30	N	N	227	N
316	33	F	S	77	195	31	S	N	107	S
317	72	F	N	102	190	29	N	N	216	N
318	56	F	S	118	220	34	S	N	60	S
319	52	M	S	143	194	27	S	N	67	S
320	72	M	S	111	165	33	S	N	122	S
321	54	M	N	101	227	21	N	N	216	N
322	49	M	N	73	224	32	N	N	174	N
323	60	F	N	87	157	26	N	N	202	N
324	42	M	N	104	239	25	N	N	121	N
325	57	M	S	94	207	31	S	N	168	S
326	46	F	S	115	235	34	S	N	131	S
327	54	M	S	91	162	31	S	S	46	S
328	70	F	N	93	187	21	N	N	51	N
329	43	M	S	95	208	27	N	N	55	S
330	44	M	S	98	179	30	S	N	151	S
331	43	M	N	101	180	24	N	N	73	N
332	68	M	N	92	176	22	S	N	207	N
333	60	F	S	120	202	29	N	N	37	S
334	59	M	N	88	193	29	N	N	17	N
335	59	F	S	134	221	31	S	N	47	S
336	33	M	S	88	185	23	N	N	136	S
337	43	F	S	79	182	25	N	N	241	S
338	37	F	S	118	202	30	N	N	52	S
339	48	F	S	82	170	24	N	N	5	S
340	58	M	S	134	210	25	S	N	126	S
341	72	F	N	77	150	29	N	S	2	N
342	34	F	N	71	224	31	S	N	2	N
343	31	M	N	111	195	27	N	N	202	N
344	48	M	S	80	154	34	S	N	72	S
345	67	M	N	87	161	31	N	N	157	N
346	37	F	S	88	245	30	S	N	69	S
347	28	M	S	73	176	22	N	N	82	S
348	42	F	N	94	161	32	N	N	67	N
349	48	F	S	78	189	27	S	N	112	S
350	48	M	N	81	217	31	N	N	249	N
351	36	M	S	107	219	25	N	S	9	S
352	47	M	N	99	231	25	N	N	139	N
353	28	F	S	123	230	33	N	N	84	S
354	29	M	S	94	176	31	N	N	159	S
355	61	M	S	75	225	32	S	N	171	S
356	62	M	S	122	200	30	N	N	156	S
357	59	M	S	124	157	31	S	N	107	S
358	31	M	S	104	229	23	S	N	86	S
359	40	F	S	147	214	30	S	N	46	S
360	29	F	N	75	192	26	N	N	171	N
361	65	F	N	74	217	29	N	N	102	N
362	62	F	S	110	242	27	N	N	173	S
363	51	M	N	84	184	30	N	N	161	N
364	40	M	S	97	212	33	S	N	100	S
365	32	M	S	83	194	29	N	N	44	S
366	49	F	S	77	161	27	N	N	24	S
367	35	F	S	79	244	26	N	N	62	S
368	28	F	S	153	201	34	N	N	24	S
369	61	M	S	101	164	22	N	N	0	S
370	29	M	S	84	221	26	N	N	159	S
371	36	F	S	108	196	34	S	S	9	S
372	69	F	S	108	185	31	S	N	112	S
373	70	F	S	84	156	30	N	N	84	S
374	57	F	S	84	210	21	N	N	11	S
375	32	M	N	77	182	20	S	N	1	N
376	51	M	N	116	167	30	N	N	220	N
377	30	F	N	82	163	29	N	N	245	N
378	69	M	S	76	171	34	S	N	145	S
379	68	M	S	108	186	25	S	N	156	S
380	60	M	N	74	233	21	N	N	47	N
381	48	F	S	74	150	33	S	N	162	S
382	56	F	S	97	219	20	N	N	10	S
383	41	F	N	83	173	20	N	N	88	N
384	34	M	N	94	196	20	N	N	233	N
385	45	F	S	102	167	24	N	N	114	S
386	29	F	S	93	171	28	N	N	90	S
387	71	M	N	93	168	26	N	N	18	N
388	63	M	S	124	217	23	S	N	13	S
389	47	M	S	105	179	27	N	N	112	S
390	28	F	S	136	151	30	S	N	93	S
391	63	F	N	80	207	27	N	N	138	N
392	50	M	S	80	151	22	N	N	198	S
393	44	M	N	107	218	30	N	N	143	N
394	38	M	N	103	216	24	N	N	248	N
395	43	M	N	85	230	32	N	N	80	N
396	40	F	N	71	162	27	N	N	4	N
397	52	F	N	76	158	25	N	N	177	N
398	55	F	S	114	179	32	N	N	18	S
399	48	F	N	73	208	32	N	N	151	N
400	65	F	S	102	159	24	N	N	205	S
401	36	F	S	111	222	27	S	N	60	S
402	49	F	S	134	171	22	S	N	189	S
403	67	F	S	100	191	23	S	N	29	S
404	71	M	N	92	217	21	N	N	32	N
405	43	M	S	75	220	22	N	N	100	S
406	34	M	S	102	211	29	N	N	99	S
407	36	F	S	96	150	21	N	N	44	S
408	63	M	N	83	164	30	N	N	52	N
409	34	F	S	111	200	25	S	N	52	S
410	66	M	N	124	164	24	N	N	166	N
411	70	F	N	77	212	24	N	N	184	N
412	54	F	N	119	152	25	N	N	53	N
413	48	M	N	93	222	31	N	N	73	N
414	33	M	S	149	207	23	S	N	199	S
415	38	M	N	76	229	25	N	N	1	N
416	28	M	S	76	244	23	S	N	96	S

Tesis publicada con autorización del autor
No olvide citar esta tesis

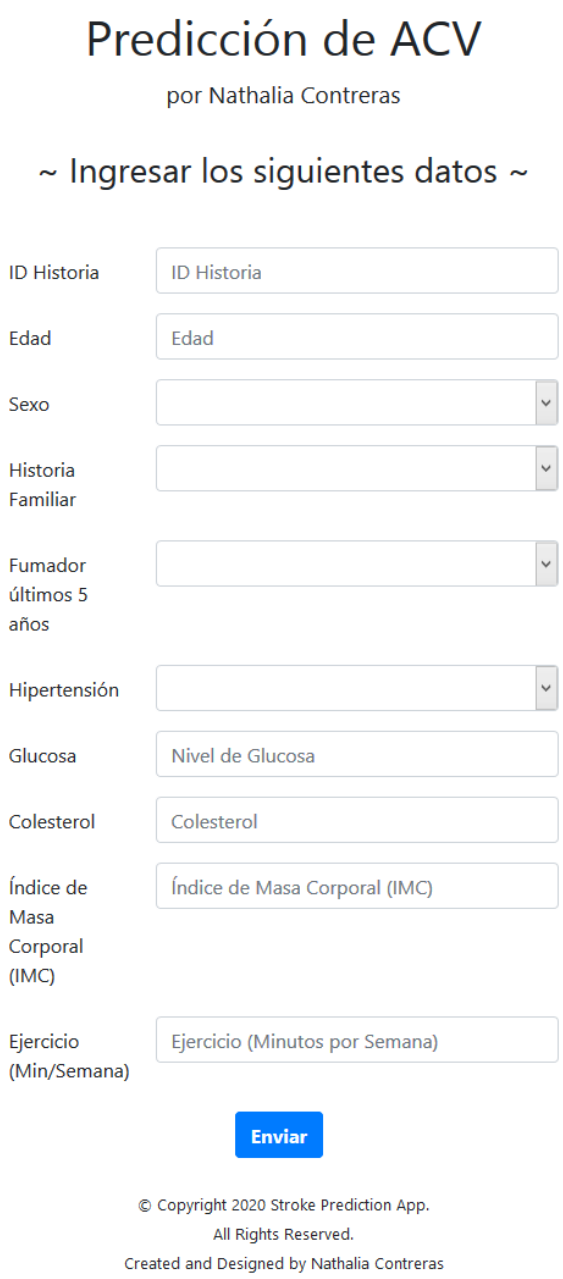
TESIS
UNIVERSIDAD CIENTÍFICA DEL SUR

417	63	F	S	95	173	25	S	N	91	S
418	57	F	S	112	150	34	N	N	171	S
419	51	F	S	117	152	26	S	N	98	S
420	47	F	S	88	153	22	S	N	2	S
421	49	F	S	82	218	27	N	N	182	S
422	59	M	S	118	165	32	S	N	3	S
423	48	F	N	82	224	31	N	N	18	N
424	70	M	S	85	158	24	S	N	143	S
425	31	M	S	86	224	20	N	N	34	S
426	67	M	N	75	154	24	N	N	3	N
427	67	F	S	96	239	26	N	N	79	S
428	64	F	N	86	216	22	N	N	94	N
429	49	M	N	70	192	30	N	N	164	N
430	72	F	S	127	164	22	S	N	41	S
431	50	F	N	103	219	31	N	N	60	N
432	50	F	S	106	193	32	S	N	149	S
433	55	F	N	81	157	22	N	N	107	N
434	59	F	N	112	235	21	N	N	67	N
435	68	M	S	84	232	25	N	N	97	S
436	57	F	S	84	202	27	S	N	143	S
437	71	F	N	87	173	20	N	N	245	N
438	39	F	S	147	242	23	S	N	58	S
439	40	F	S	111	166	30	S	N	143	S
440	71	M	S	116	219	23	S	N	140	S
441	60	F	N	99	194	28	N	N	169	N
442	31	M	S	139	152	26	S	N	101	S
443	37	F	S	114	182	31	S	N	183	S
444	71	M	N	102	196	25	N	N	126	N
445	72	M	N	102	229	28	N	N	102	N
446	62	M	S	108	160	34	N	N	141	S
447	49	F	S	114	199	30	N	N	51	S
448	35	M	S	134	214	25	S	N	28	S
449	32	M	N	83	178	28	N	N	35	N
450	59	F	S	120	223	29	S	N	51	S
451	57	M	N	87	206	26	N	N	110	N
452	63	F	N	101	205	22	N	N	46	N
453	58	F	S	71	213	28	S	N	168	S
454	49	F	N	72	150	29	N	N	36	N
455	48	M	S	107	223	31	S	N	195	S
456	71	M	S	92	190	20	N	N	246	S
457	67	F	N	113	162	21	N	N	174	N
458	46	F	N	90	177	20	N	N	221	N
459	55	M	S	72	157	26	N	N	39	S
460	34	F	N	87	192	20	N	S	34	N
461	33	M	S	80	174	33	S	N	1	S
462	39	F	N	70	171	21	N	N	2	N
463	35	F	N	94	153	22	N	N	8	N
464	57	M	N	73	177	22	N	N	172	N
465	44	F	N	99	213	32	N	N	2	N
466	53	F	N	80	208	32	N	N	193	N
467	50	M	S	80	179	29	N	N	76	S
468	32	F	N	94	192	22	N	N	193	N
469	49	F	S	79	205	34	N	N	15	S
470	32	F	S	98	180	22	N	N	111	S
471	59	M	S	77	191	28	S	N	157	S
472	50	M	N	76	170	29	N	N	40	N
473	50	F	S	116	243	28	S	N	122	S
474	33	M	N	126	221	22	N	N	231	N
475	50	F	N	103	157	28	N	N	235	N
476	51	M	S	83	153	27	S	N	34	S
477	62	F	S	104	228	27	S	N	138	S
478	63	M	N	120	215	28	N	N	40	N
479	38	M	N	110	231	31	N	N	4	N
480	48	M	S	87	165	23	S	N	193	S
481	41	M	S	89	185	30	S	N	26	S
482	68	F	S	72	174	26	N	N	97	S
483	33	M	S	75	202	25	S	N	0	S
484	28	F	S	88	185	33	S	N	8	S
485	64	F	N	89	156	23	N	N	97	N
486	45	M	S	71	233	22	N	S	190	S
487	54	M	N	111	236	26	N	N	89	N
488	71	F	S	98	221	31	S	N	131	S
489	68	F	S	78	155	30	N	N	122	S
490	45	M	S	100	191	29	S	N	122	S
491	47	M	S	74	232	31	N	N	15	S
492	53	F	S	83	210	33	N	N	145	S
493	52	F	S	71	199	25	N	N	107	S
494	49	F	S	96	228	28	N	N	196	S
495	48	F	N	95	233	22	N	N	160	N
496	40	M	S	106	168	33	S	N	142	S
497	31	F	S	73	223	32	S	N	130	S
498	36	F	S	81	201	30	S	N	41	S
499	36	F	S	105	205	29	N	N	108	S
500	71	F	N	124	230	25	S	N	217	N
501	53	F	S	73	217	25	S	N	107	S
502	41	M	S	140	227	25	S	N	92	S
503	43	M	N	92	161	24	N	N	182	N
504	67	M	S	131	161	33	N	N	51	S
505	28	F	N	74	227	31	N	N	152	N
506	64	M	N	74	193	22	N	N	183	N
507	57	F	N	75	208	32	N	N	134	N
508	35	F	N	80	191	26	N	N	238	N
509	34	F	S	94	233	25	N	N	194	S
510	40	M	S	81	212	34	S	N	24	S
511	37	M	N	110	185	27	N	N	6	N
512	51	F	N	74	194	32	N	N	36	N
513	39	F	S	82	220	27	S	N	149	S
514	66	M	S	127	242	22	S	N	30	S
515	58	F	N	74	188	22	N	N	130	N
516	56	F	S	115	172	30	N	N	183	S
517	33	F	N	77	166	31	N	N	91	N
518	65	F	N	83	164	23	N	N	2	N
519	29	M	N	106	172	25	N	N	226	N
520	56	M	N	99	175	32	N	N	228	N

Tesis publicada con autorización del autor
No olvide citar esta tesis

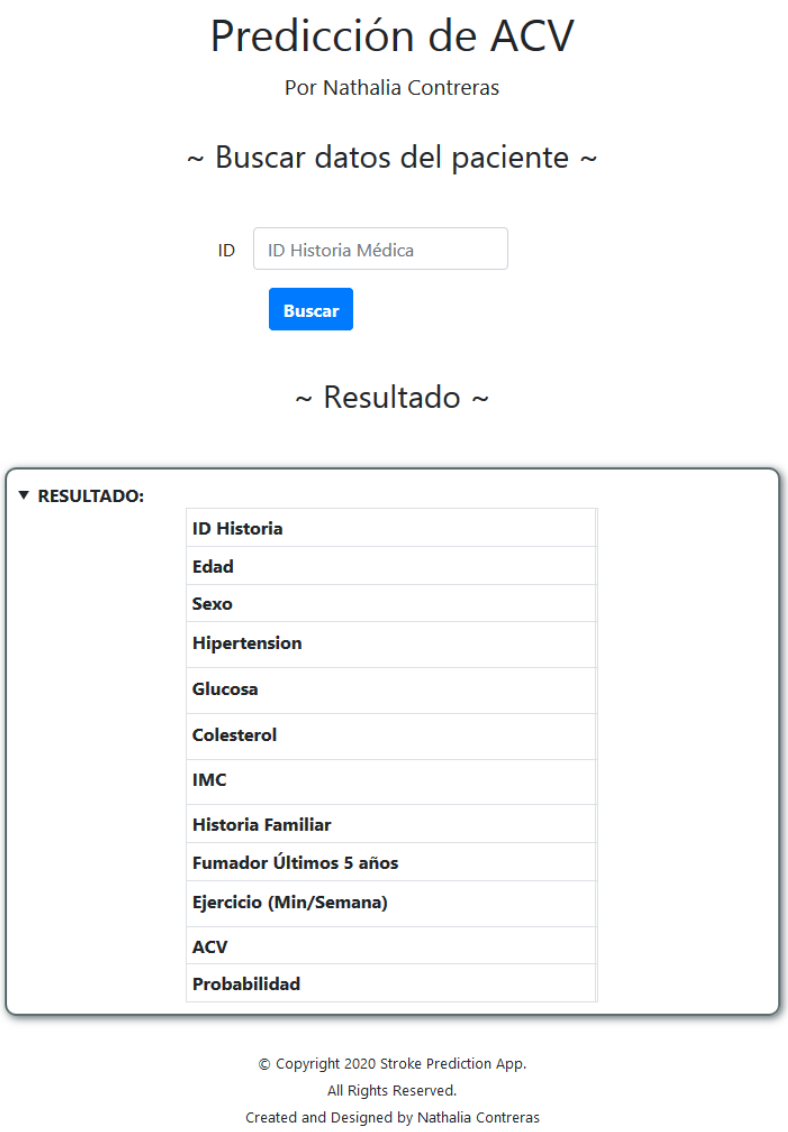
Anexo 8. Interfaces de la Aplicación Web

Interfaz #1: Ingreso de Datos



Tesis publicada con autorización del autor
No olvide citar esta tesis

Interfaz #2: Consulta de Datos y Resultados



Tesis publicada con autorización del autor
No olvide citar esta tesis

Anexo 9. Tabla de Resultados del Machine Learning

	EDAD	SEXO	HIPERTENS	GLUCOSA	COLESTEROL	IMC	HISTFAM	FUMAR	EJERCMINXSEM	SEXO_IX	HIPERTENS_IX	HISTFAM_IX	FUMAR_IX	label	features	raw Prediction	probability	prediction	predicted Label
1	51	M	N	102	175	30	N	N	227	0	0	0	0	0	[51.0, 0.0, 0.0, 102.0, 175.0, 30.0, 0.0, 0.0, 227.0]	[19.159, 0.841]	[0.958, 0.042]	0	N
2	60	M	S	144	183	31	S	N	69	0	1	1	0	0	[60.0, 0.0, 1.0, 144.0, 183.0, 31.0, 1.0, 0.0, 69.0]	[0.0, 20.0]	[0.0, 1.0]	1	S
3	28	F	S	99	223	29	N	N	121	1	1	0	0	0	[28.0, 1.0, 1.0, 99.0, 223.0, 29.0, 0.0, 0.0, 121.0]	[0.235, 19.765]	[0.012, 0.988]	1	S
4	55	M	N	101	209	27	N	N	199	0	0	0	0	0	[55.0, 0.0, 0.0, 101.0, 209.0, 27.0, 0.0, 0.0, 199.0]	[19.282, 0.718]	[0.964, 0.036]	0	N
5	40	M	S	102	211	27	N	N	102	0	1	0	0	0	[40.0, 0.0, 1.0, 102.0, 211.0, 27.0, 0.0, 0.0, 102.0]	[0.282, 19.718]	[0.014, 0.986]	1	S
6	72	F	N	97	154	28	N	N	72	1	0	0	0	0	[72.0, 1.0, 0.0, 97.0, 154.0, 28.0, 0.0, 0.0, 72.0]	[17.282, 2.718]	[0.864, 0.136]	0	N
7	41	F	N	84	214	23	N	N	24	1	0	0	0	0	[41.0, 1.0, 0.0, 84.0, 214.0, 23.0, 0.0, 0.0, 24.0]	[19.529, 0.471]	[0.976, 0.024]	0	N
8	30	M	S	140	226	28	S	N	122	0	1	1	0	0	[30.0, 0.0, 1.0, 140.0, 226.0, 28.0, 1.0, 0.0, 122.0]	[1.037, 18.963]	[0.052, 0.948]	1	S
9	51	F	N	96	157	29	N	N	129	1	0	0	0	0	[51.0, 1.0, 0.0, 96.0, 157.0, 29.0, 0.0, 0.0, 129.0]	[18.517, 1.483]	[0.926, 0.074]	0	N
10	46	F	N	77	166	23	N	N	7	1	0	0	0	0	[46.0, 1.0, 0.0, 77.0, 166.0, 23.0, 0.0, 0.0, 7.0]	[19.529, 0.471]	[0.976, 0.024]	0	N
11	62	M	N	76	201	20	N	N	165	0	0	0	0	0	[62.0, 0.0, 0.0, 76.0, 201.0, 20.0, 0.0, 0.0, 165.0]	[19.529, 0.471]	[0.976, 0.024]	0	N
12	70	M	S	85	154	22	S	N	53	0	1	1	0	0	[70.0, 0.0, 1.0, 85.0, 154.0, 22.0, 1.0, 0.0, 53.0]	[0.529, 19.471]	[0.026, 0.974]	1	S
13	49	F	S	91	197	24	N	N	58	1	1	0	0	0	[49.0, 1.0, 1.0, 91.0, 197.0, 24.0, 0.0, 0.0, 58.0]	[0.529, 19.471]	[0.026, 0.974]	1	S
14	38	M	S	117	153	26	S	S	120	0	1	1	1	0	[38.0, 0.0, 1.0, 117.0, 153.0, 26.0, 1.0, 1.0, 120.0]	[0.037, 19.963]	[0.002, 0.998]	1	S
15	54	F	S	114	231	28	S	N	64	1	1	1	0	0	[54.0, 1.0, 1.0, 114.0, 231.0, 28.0, 1.0, 0.0, 64.0]	[2.0, 18.0]	[0.1, 0.9]	1	S
16	55	M	N	71	174	30	N	N	227	0	0	0	0	0	[55.0, 0.0, 0.0, 71.0, 174.0, 30.0, 0.0, 0.0, 227.0]	[18.159, 1.841]	[0.908, 0.092]	0	N
17	33	F	S	77	195	31	S	N	107	1	1	1	0	0	[33.0, 1.0, 1.0, 77.0, 195.0, 31.0, 1.0, 0.0, 107.0]	[0.091, 19.909]	[0.005, 0.995]	1	S
18	72	F	N	102	190	29	N	N	216	1	0	0	0	0	[72.0, 1.0, 0.0, 102.0, 190.0, 29.0, 0.0, 0.0, 216.0]	[19.909, 0.091]	[0.995, 0.005]	0	N
19	56	F	S	118	220	34	S	N	60	1	1	1	0	0	[56.0, 1.0, 1.0, 118.0, 220.0, 34.0, 1.0, 0.0, 60.0]	[2.0, 18.0]	[0.1, 0.9]	1	S
20	52	M	S	143	194	27	S	N	67	0	1	1	0	0	[52.0, 0.0, 1.0, 143.0, 194.0, 27.0, 1.0, 0.0, 67.0]	[1.0, 19.0]	[0.05, 0.95]	1	S
21	72	M	S	111	165	33	S	N	122	0	1	1	0	0	[72.0, 0.0, 1.0, 111.0, 165.0, 33.0, 1.0, 0.0, 122.0]	[1.5, 18.5]	[0.075, 0.925]	1	S
22	54	M	N	101	227	21	N	N	216	0	0	0	0	0	[54.0, 0.0, 0.0, 101.0, 227.0, 21.0, 0.0, 0.0, 216.0]	[20.0, 0.0]	[1.0, 0.0]	0	N
23	49	M	N	73	224	32	N	N	174	0	0	0	0	0	[49.0, 0.0, 0.0, 73.0, 224.0, 32.0, 0.0, 0.0, 174.0]	[19.235, 0.765]	[0.962, 0.038]	0	N
24	60	F	N	87	157	26	N	N	202	1	0	0	0	0	[60.0, 1.0, 0.0, 87.0, 157.0, 26.0, 0.0, 0.0, 202.0]	[18.529, 1.471]	[0.926, 0.074]	0	N
25	42	M	N	104	239	25	N	N	121	0	0	0	0	0	[42.0, 0.0, 0.0, 104.0, 239.0, 25.0, 0.0, 0.0, 121.0]	[20.0, 0.0]	[1.0, 0.0]	0	N
26	57	M	S	94	207	31	S	N	168	0	1	1	0	0	[57.0, 0.0, 1.0, 94.0, 207.0, 31.0, 1.0, 0.0, 168.0]	[1.0, 19.0]	[0.05, 0.95]	1	S
27	46	F	S	115	235	34	S	N	131	1	1	1	0	0	[46.0, 1.0, 1.0, 115.0, 235.0, 34.0, 1.0, 0.0, 131.0]	[1.091, 18.909]	[0.055, 0.945]	1	S
28	54	M	S	91	162	31	S	S	46	0	1	1	1	0	[54.0, 0.0, 1.0, 91.0, 162.0, 31.0, 1.0, 1.0, 46.0]	[0.0, 20.0]	[0.0, 1.0]	1	S
29	70	F	N	93	187	21	N	N	51	1	0	0	0	0	[70.0, 1.0, 0.0, 93.0, 187.0, 21.0, 0.0, 0.0, 51.0]	[18.282, 1.718]	[0.914, 0.086]	0	N
30	43	M	S	95	208	27	N	N	55	0	1	0	0	0	[43.0, 0.0, 1.0, 95.0, 208.0, 27.0, 0.0, 0.0, 55.0]	[0.282, 19.718]	[0.014, 0.986]	1	S
31	44	M	S	98	179	30	S	N	151	0	1	1	0	0	[44.0, 0.0, 1.0, 98.0, 179.0, 30.0, 1.0, 0.0, 151.0]	[0.319, 19.681]	[0.016, 0.984]	1	S
32	43	M	N	101	180	24	N	N	73	0	0	0	0	0	[43.0, 0.0, 0.0, 101.0, 180.0, 24.0, 0.0, 0.0, 73.0]	[19.282, 0.718]	[0.964, 0.036]	0	N
33	68	M	N	92	176	22	S	N	207	0	0	1	0	0	[68.0, 0.0, 0.0, 92.0, 176.0, 22.0, 1.0, 0.0, 207.0]	[16.282, 3.718]	[0.814, 0.186]	0	N
34	60	F	S	120	202	29	N	N	37	1	1	0	0	0	[60.0, 1.0, 1.0, 120.0, 202.0, 29.0, 0.0, 0.0, 37.0]	[1.0, 19.0]	[0.05, 0.95]	1	S
35	59	M	N	88	193	29	N	N	17	0	0	0	0	0	[59.0, 0.0, 0.0, 88.0, 193.0, 29.0, 0.0, 0.0, 17.0]	[19.279, 0.721]	[0.964, 0.036]	0	N
36	59	F	S	134	221	31	S	N	47	1	1	1	0	0	[59.0, 1.0, 1.0, 134.0, 221.0, 31.0, 1.0, 0.0, 47.0]	[2.0, 18.0]	[0.1, 0.9]	1	S
37	33	M	S	88	185	23	N	N	136	0	1	0	0	0	[33.0, 0.0, 1.0, 88.0, 185.0, 23.0, 0.0, 0.0, 136.0]	[1.529, 18.471]	[0.076, 0.924]	1	S
38	43	F	S	79	182	25	N	N	241	1	1	0	0	0	[43.0, 1.0, 1.0, 79.0, 182.0, 25.0, 0.0, 0.0, 241.0]	[2.0, 18.0]	[0.1, 0.9]	1	S
39	37	F	S	118	202	30	N	N	52	1	1	0	0	0	[37.0, 1.0, 1.0, 118.0, 202.0, 30.0, 0.0, 0.0, 52.0]	[2.0, 18.0]	[0.1, 0.9]	1	S
40	48	F	S	82	170	24	N	N	5	1	1	0	0	0	[48.0, 1.0, 1.0, 82.0, 170.0, 24.0, 0.0, 0.0, 5.0]	[1.779, 18.221]	[0.089, 0.911]	1	S
41	58	M	S	134	210	25	S	N	126	0	1	1	0	0	[58.0, 0.0, 1.0, 134.0, 210.0, 25.0, 1.0, 0.0, 126.0]	[0.0, 20.0]	[0.0, 1.0]	1	S
42	72	F	N	77	150	29	N	S	2	1	0	0	1	0	[72.0, 1.0, 0.0, 77.0, 150.0, 29.0, 0.0, 1.0, 2.0]	[16.529, 3.471]	[0.826, 0.174]	0	N
43	34	F	N	71	224	31	S	N	2	1	0	1	0	0	[34.0, 1.0, 0.0, 71.0, 224.0, 31.0, 1.0, 0.0, 2.0]	[17.091, 2.909]	[0.855, 0.145]	0	N
44	31	M	N	111	195	27	N	N	202	0	0	0	0	0	[31.0, 0.0, 0.0, 111.0, 195.0, 27.0, 0.0, 0.0, 202.0]	[19.282, 0.718]	[0.964, 0.036]	0	N
45	48	M	S	80	154	34	S	N	72	0	1	1	0	0	[48.0, 0.0, 1.0, 80.0, 154.0, 34.0, 1.0, 0.0, 72.0]	[0.037, 19.963]	[0.002, 0.998]	1	S
46	67	M	N	87	161	31	N	N	157	0	0	0	0	0	[67.0, 0.0, 0.0, 87.0, 161.0, 31.0, 0.0, 0.0, 157.0]	[18.235, 1.765]	[0.912, 0.088]	0	N
47	37	F	S	88	245	30	S	N	69	1	1	1	0	0	[37.0, 1.0, 1.0, 88.0, 245.0, 30.0, 1.0, 0.0, 69.0]	[1.091, 18.909]	[0.055, 0.945]	1	S
48	28	M	S	73	176	22	N	N	82	0	1	0	0	0	[28.0, 0.0, 1.0, 73.0, 176.0, 22.0, 0.0, 0.0, 82.0]	[1.529, 18.471]	[0.076, 0.924]	1	S
49	42	F	N	94	161	32	N	N	67	1	0	0	0	0	[42.0, 1.0, 0.0, 94.0, 161.0, 32.0, 0.0, 0.0, 67.0]	[18.235, 1.765]	[0.912, 0.088]	0	N
50	48	F	S	78	189	27	S	N	112	1	1	1	0	0	[48.0, 1.0, 1.0, 78.0, 189.0, 27.0, 1.0, 0.0, 112.0]	[0.62, 19.38]	[0.031, 0.969]	1	S
51	48	M	N	81	217	31	N	N	249	0	0	0	0	0	[48.0, 0.0, 0.0, 81.0, 217.0, 31.0, 0.0, 0.0, 249.0]	[19.909, 0.091]	[0.995, 0.005]	0	N
52	36	M	S	107	219	25	N	S	9	0	1	0	1	0	[36.0, 0.0, 1.0, 107.0, 219.0, 25.0, 0.0, 1.0, 9.0]	[2.0, 18.0]	[0.1, 0.9]	1	S
53	47	M	N	99	231	25	N	N	139	0	0	0	0	0	[47.0, 0.0, 0.0, 99.0, 231.0, 25.0, 0.0, 0.0, 139.0]	[20.0, 0.0]	[1.0, 0.0]	0	N
54	28	F	S	123	230	33	N	N	84	1	1	0	0	0	[28.0, 1.0, 1.0, 123.0, 230.0, 33.0, 0.0, 0.0, 84.0]	[2.0, 18.0]	[0.1, 0.9]	1	S
55	29	M	S	94	176	31	N	N	159	0	1	0	0	0	[29.0, 0.0, 1.0, 94.0, 176.0, 31.0, 0.0, 0.0, 159.0]	[1.235, 18.765]	[0.062, 0.938]	1	S
56	61	M	S	75	225	32	S	N	171	0	1	1	0	0	[61.0, 0.0, 1.0, 75.0, 225.0, 32.0, 1.0, 0.0, 171.0]	[0.0, 20.0]	[0.0, 1.0]	1	S
57	62	M	S	122	200	30	N	N	156	0	1	0	0	0	[62.0, 0.0, 1.0, 122.0, 200.0, 30.0, 0.0, 0.0, 156.0]	[0.235, 19.765]	[0.012, 0.988]	1	S
58	59	M	S	124	157	31	S	N	107	0	1	1	0	0	[59.0, 0.0, 1.0, 124.0, 157.0, 31.0, 1.0, 0.0, 107.0]	[1.0, 19.0]	[0.05, 0.95]	1	S
59	31	M	S	104	229	23	S	N	86	0	1	1	0	0	[31.0, 0.0, 1.0, 104.0, 229.0, 23.0, 1.0, 0.0, 86.0]	[1.037, 18.963]	[0.052, 0.948]	1	S
60	40	F	S	147	214	30	S	N	46	1	1	1	0	0	[40.0, 1.0, 1.0, 147.0, 214.0, 30.0, 1.0, 0.0, 46.0]	[1.091, 18.909]	[0.055, 0.945]	1	S
61	29	F	N	75	192	26	N	N	171	1	0	0	0	0	[29.0, 1.0, 0.0, 75.0, 192.0, 26.0, 0.0, 0.0, 171.0]	[18.529, 1.471]	[0.926, 0.074]	0	N
62	65	F	N	74	217	29	N	N	102	1	0	0	0	0	[65.0, 1.0, 0.0, 74.0, 217.0, 29.0, 0.0, 0.0, 102.0]	[18.765, 1.235]	[0.938, 0.062]	0	N
63	62	F	S	110	242	27	N	N	173	1	1	0	0	0	[62.0, 1.0, 1.0, 110.0, 242.0, 27.0, 0.0, 0.0, 173.0]	[1.0, 19.0]	[0.05, 0.95]	1	S
64	51	M	N	84	184	30	N	N	161	0	0	0	0	0	[51.0, 0.0, 0.0, 84.0, 184.0, 30.0, 0.0, 0.0, 161.0]	[18.765, 1.235]	[0.938, 0.062]	0	N
65	40	M	S	97	212	33	S	N	100	0	1	1	0	0	[40.0, 0.0, 1.0, 97.0, 212.0, 33.0, 1.0, 0.0, 100.0]	[0.037, 19.963]	[0.002, 0.998]	1	S
66	32	M	S	83	194	29	N	N	44	0	1	0	0	0	[32.0, 0.0, 1.0, 83.0, 194.0, 29.0, 0.0, 0.0, 44.0]	[1.279, 18.721]	[0.064, 0.936]	1	S
67	49	F	S	77	161	27	N	N	24	1	1	0	0	0	[49.0, 1.0, 1.0, 77.0, 161.0, 27.0, 0.0, 0.0, 24.0]	[1.529, 18.471]	[0.076, 0.924]	1	S
68	35	F	S	79	244	26	N	N	62	1	1	0	0	0	[35.0, 1.0, 1.0, 79.0, 244.0, 26.0, 0.0, 0.0, 62.0]	[0.0, 20.0]	[0.0, 1.0]	1	S
69	28	F	S	153	201	34	N	N	24	1	1	0	0	0	[28.0, 1.0, 1.0, 153.0, 201.0, 34.0, 0.0, 0.0, 24.0]	[0.0, 20.0]	[0.0, 1.0]	1	S
70	61	M	S	101	164	22	N	N	0	0	1	0	0	0	[61.0, 0.0, 1.0, 101.0, 164.0, 22.0, 0.0, 0.0, 0.0]	[1.282, 18.718]	[0.064, 0.936]	1	S
71	29	M	S	84	221	26	N	N	159	0	1	0	0	0	[29.0, 0.0, 1.0, 84.0, 221.0, 26.0, 0.0, 0.0, 159.0]	[0.0, 20.0]	[0.0, 1.0]	1	S
72	36	F	S	108	196	34	S	S	9	1	1	1	1	0	[36.0, 1.0, 1.0, 108.0, 196.0, 34.0, 1.0, 1.0, 9.0]	[0.591, 19.409]	[0.03, 0.97]	1	S
73	69	F	S	108	185	31	S	N	112	1	1	1	0	0	[69.0, 1.0, 1.0, 108.0, 185.0, 31.0, 1.0, 0.0, 112.0]	[0.5, 19.5]	[0.025, 0.975]	1	S
74	70	F	S	84	156	30	N	N	84	1	1	0	0	0	[70.0, 1.0, 1.0, 84.0, 156.0, 30.0, 0.0, 0.0, 84.0]	[1.765, 18.235]	[0.088, 0.912]	1	S
75	57	F	S	84	210	21	N	N	11	1	1	0	0	0	[57.0, 1.0, 1.0, 84.0, 210.0, 21.0, 0.0, 0.0, 11.0]	[2.529, 17.471]	[0.126, 0.874]	1	S

Tesis publicada con autorización del autor
No olvide citar esta tesis

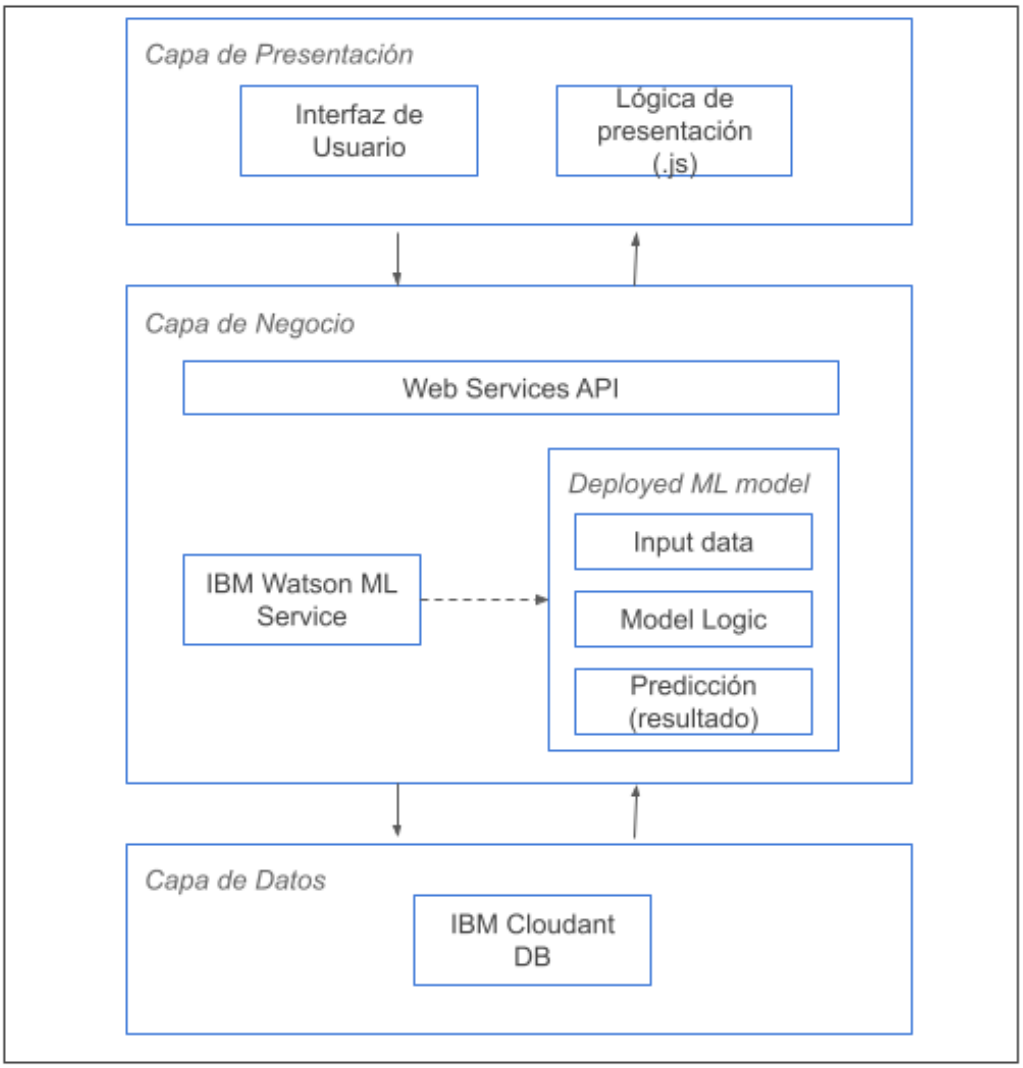
76	32	M	N	77	182	20	S	N	1	0	0	1	0	0	[32.0, 0.0, 0.0, 77.0, 182.0, 20.0, 1.0, 0.0, 1.0]	[17.529, 2.471]	[0.876, 0.124]	0	N
77	51	M	N	116	167	30	N	N	220	0	0	0	0	0	[51.0, 0.0, 0.0, 116.0, 167.0, 30.0, 0.0, 0.0, 220.0]	[19.159, 0.841]	[0.958, 0.042]	0	N
78	30	F	N	82	163	29	N	N	245	1	0	0	0	0	[30.0, 1.0, 0.0, 82.0, 163.0, 29.0, 0.0, 0.0, 245.0]	[19.909, 0.091]	[0.995, 0.005]	0	N
79	69	M	S	76	171	34	S	N	145	0	1	1	0	0	[69.0, 0.0, 1.0, 76.0, 171.0, 34.0, 1.0, 0.0, 145.0]	[0.0, 20.0]	[0.0, 1.0]	1	S
80	68	M	S	108	186	25	S	N	156	0	1	1	0	0	[68.0, 0.0, 1.0, 108.0, 186.0, 25.0, 1.0, 0.0, 156.0]	[0.282, 19.718]	[0.014, 0.986]	1	S
81	60	M	N	74	233	21	N	N	47	0	0	0	0	0	[60.0, 0.0, 0.0, 74.0, 233.0, 21.0, 0.0, 0.0, 47.0]	[20.0, 0.0]	[1.0, 0.0]	0	N
82	48	F	S	74	150	33	S	N	162	1	1	1	0	0	[48.0, 1.0, 1.0, 74.0, 150.0, 33.0, 1.0, 0.0, 162.0]	[0.091, 19.909]	[0.005, 0.995]	1	S
83	56	F	S	97	219	20	N	N	10	1	1	0	0	0	[56.0, 1.0, 1.0, 97.0, 219.0, 20.0, 0.0, 0.0, 10.0]	[2.0, 18.0]	[0.1, 0.9]	1	S
84	41	F	N	83	173	20	N	N	88	1	0	0	0	0	[41.0, 1.0, 0.0, 83.0, 173.0, 20.0, 0.0, 0.0, 88.0]	[18.779, 1.221]	[0.939, 0.061]	0	N
85	34	M	N	94	196	20	N	N	233	0	0	0	0	0	[34.0, 0.0, 0.0, 94.0, 196.0, 20.0, 0.0, 0.0, 233.0]	[20.0, 0.0]	[1.0, 0.0]	0	N
86	45	F	S	102	167	24	N	N	114	1	1	0	0	0	[45.0, 1.0, 1.0, 102.0, 167.0, 24.0, 0.0, 0.0, 114.0]	[0.532, 19.468]	[0.027, 0.973]	1	S
87	29	F	S	93	171	28	N	N	90	1	1	0	0	0	[29.0, 1.0, 1.0, 93.0, 171.0, 28.0, 0.0, 0.0, 90.0]	[1.282, 18.718]	[0.064, 0.936]	1	S
88	71	M	N	93	168	26	N	N	18	0	0	0	0	0	[71.0, 0.0, 0.0, 93.0, 168.0, 26.0, 0.0, 0.0, 18.0]	[17.532, 2.468]	[0.877, 0.123]	0	N
89	63	M	S	124	217	23	S	N	13	0	1	1	0	0	[63.0, 0.0, 1.0, 124.0, 217.0, 23.0, 1.0, 0.0, 13.0]	[0.6, 19.4]	[0.03, 0.97]	1	S
90	47	M	S	105	179	27	N	N	112	0	1	0	0	0	[47.0, 0.0, 1.0, 105.0, 179.0, 27.0, 0.0, 0.0, 112.0]	[1.282, 18.718]	[0.064, 0.936]	1	S
91	28	F	S	136	151	30	S	N	93	1	1	1	0	0	[28.0, 1.0, 1.0, 136.0, 151.0, 30.0, 1.0, 0.0, 93.0]	[0.091, 19.909]	[0.005, 0.995]	1	S
92	63	F	N	80	207	27	N	N	138	1	0	0	0	0	[63.0, 1.0, 0.0, 80.0, 207.0, 27.0, 0.0, 0.0, 138.0]	[19.529, 0.471]	[0.976, 0.024]	0	N
93	50	M	S	80	151	22	N	N	198	0	1	0	0	0	[50.0, 0.0, 1.0, 80.0, 151.0, 22.0, 0.0, 0.0, 198.0]	[1.529, 18.471]	[0.076, 0.924]	1	S
94	44	M	N	107	218	30	N	N	143	0	0	0	0	0	[44.0, 0.0, 0.0, 107.0, 218.0, 30.0, 0.0, 0.0, 143.0]	[18.517, 1.483]	[0.926, 0.074]	0	N
95	38	M	N	103	216	24	N	N	248	0	0	0	0	0	[38.0, 0.0, 0.0, 103.0, 216.0, 24.0, 0.0, 0.0, 248.0]	[20.0, 0.0]	[1.0, 0.0]	0	N
96	43	M	N	85	230	32	N	N	80	0	0	0	0	0	[43.0, 0.0, 0.0, 85.0, 230.0, 32.0, 0.0, 0.0, 80.0]	[19.235, 0.765]	[0.962, 0.038]	0	N
97	40	F	N	71	162	27	N	N	4	1	0	0	0	0	[40.0, 1.0, 0.0, 71.0, 162.0, 27.0, 0.0, 0.0, 4.0]	[19.529, 0.471]	[0.976, 0.024]	0	N
98	52	F	N	76	158	25	N	N	177	1	0	0	0	0	[52.0, 1.0, 0.0, 76.0, 158.0, 25.0, 0.0, 0.0, 177.0]	[19.529, 0.471]	[0.976, 0.024]	0	N
99	55	F	S	114	179	32	N	N	18	1	1	0	0	0	[55.0, 1.0, 1.0, 114.0, 179.0, 32.0, 0.0, 0.0, 18.0]	[2.5, 17.5]	[0.125, 0.875]	1	S
100	48	F	N	73	208	32	N	N	151	1	0	0	0	0	[48.0, 1.0, 0.0, 73.0, 208.0, 32.0, 0.0, 0.0, 151.0]	[18.235, 1.765]	[0.912, 0.088]	0	N
101	65	F	S	102	159	24	N	N	205	1	1	0	0	0	[65.0, 1.0, 1.0, 102.0, 159.0, 24.0, 0.0, 0.0, 205.0]	[0.282, 19.718]	[0.014, 0.986]	1	S
102	36	F	S	111	222	27	S	N	60	1	1	1	0	0	[36.0, 1.0, 1.0, 111.0, 222.0, 27.0, 1.0, 0.0, 60.0]	[2.091, 17.909]	[0.105, 0.895]	1	S
103	49	F	S	134	171	22	S	N	189	1	1	1	0	0	[49.0, 1.0, 1.0, 134.0, 171.0, 22.0, 1.0, 0.0, 189.0]	[0.091, 19.909]	[0.005, 0.995]	1	S
104	67	F	S	100	191	23	S	N	29	1	1	1	0	0	[67.0, 1.0, 1.0, 100.0, 191.0, 23.0, 1.0, 0.0, 29.0]	[0.282, 19.718]	[0.014, 0.986]	1	S
105	71	M	N	92	217	21	N	N	32	0	0	0	0	0	[71.0, 0.0, 0.0, 92.0, 217.0, 21.0, 0.0, 0.0, 32.0]	[18.282, 1.718]	[0.914, 0.086]	0	N
106	43	M	S	75	220	22	N	N	100	0	1	0	0	0	[43.0, 0.0, 1.0, 75.0, 220.0, 22.0, 0.0, 0.0, 100.0]	[0.0, 20.0]	[0.0, 1.0]	1	S
107	34	M	S	102	211	29	N	N	99	0	1	0	0	0	[34.0, 0.0, 1.0, 102.0, 211.0, 29.0, 0.0, 0.0, 99.0]	[0.517, 19.483]	[0.026, 0.974]	1	S
108	36	F	S	96	150	21	N	N	44	1	1	0	0	0	[36.0, 1.0, 1.0, 96.0, 150.0, 21.0, 0.0, 0.0, 44.0]	[0.282, 19.718]	[0.014, 0.986]	1	S
109	63	M	N	83	164	30	N	N	52	0	0	0	0	0	[63.0, 0.0, 0.0, 83.0, 164.0, 30.0, 0.0, 0.0, 52.0]	[19.279, 0.721]	[0.964, 0.036]	0	N
110	34	F	S	111	200	25	S	N	52	1	1	1	0	0	[34.0, 1.0, 1.0, 111.0, 200.0, 25.0, 1.0, 0.0, 52.0]	[0.373, 19.627]	[0.019, 0.981]	1	S
111	66	M	N	124	164	24	N	N	166	0	0	0	0	0	[66.0, 0.0, 0.0, 124.0, 164.0, 24.0, 0.0, 0.0, 166.0]	[18.0, 2.0]	[0.9, 0.1]	0	N
112	70	F	N	77	212	24	N	N	184	1	0	0	0	0	[70.0, 1.0, 0.0, 77.0, 212.0, 24.0, 0.0, 0.0, 184.0]	[19.529, 0.471]	[0.976, 0.024]	0	N
113	54	F	N	119	152	25	N	N	53	1	0	0	0	0	[54.0, 1.0, 0.0, 119.0, 152.0, 25.0, 0.0, 0.0, 53.0]	[17.0, 3.0]	[0.85, 0.15]	0	N
114	48	M	N	93	222	31	N	N	73	0	0	0	0	0	[48.0, 0.0, 0.0, 93.0, 222.0, 31.0, 0.0, 0.0, 73.0]	[19.235, 0.765]	[0.962, 0.038]	0	N
115	33	M	S	149	207	23	S	N	199	0	1	1	0	0	[33.0, 0.0, 1.0, 149.0, 207.0, 23.0, 1.0, 0.0, 199.0]	[3.037, 16.963]	[0.152, 0.848]	1	S
116	38	M	N	76	229	25	N	N	1	0	0	0	0	0	[38.0, 0.0, 0.0, 76.0, 229.0, 25.0, 0.0, 0.0, 1.0]	[20.0, 0.0]	[1.0, 0.0]	0	N
117	28	M	S	76	244	23	S	N	96	0	1	1	0	0	[28.0, 0.0, 1.0, 76.0, 244.0, 23.0, 1.0, 0.0, 96.0]	[0.037, 19.963]	[0.002, 0.998]	1	S
118	63	F	S	95	173	25	S	N	91	1	1	1	0	0	[63.0, 1.0, 1.0, 95.0, 173.0, 25.0, 1.0, 0.0, 91.0]	[0.282, 19.718]	[0.014, 0.986]	1	S
119	57	F	S	112	150	34	N	N	171	1	1	0	0	0	[57.0, 1.0, 1.0, 112.0, 150.0, 34.0, 0.0, 0.0, 171.0]	[1.5, 18.5]	[0.075, 0.925]	1	S
120	51	F	S	117	152	26	S	N	98	1	1	1	0	0	[51.0, 1.0, 1.0, 117.0, 152.0, 26.0, 1.0, 0.0, 98.0]	[0.091, 19.909]	[0.005, 0.995]	1	S
121	47	F	S	88	153	22	S	N	2	1	1	1	0	0	[47.0, 1.0, 1.0, 88.0, 153.0, 22.0, 1.0, 0.0, 2.0]	[0.62, 19.38]	[0.031, 0.969]	1	S
122	49	F	S	82	218	27	N	N	182	1	1	0	0	0	[49.0, 1.0, 1.0, 82.0, 218.0, 27.0, 0.0, 0.0, 182.0]	[0.529, 19.471]	[0.026, 0.974]	1	S
123	59	M	S	118	165	32	S	N	3	0	1	1	0	0	[59.0, 0.0, 1.0, 118.0, 165.0, 32.0, 1.0, 0.0, 3.0]	[1.0, 19.0]	[0.05, 0.95]	1	S
124	48	F	N	82	224	31	N	N	18	1	0	0	0	0	[48.0, 1.0, 0.0, 82.0, 224.0, 31.0, 0.0, 0.0, 18.0]	[20.0, 0.0]	[1.0, 0.0]	0	N
125	70	M	S	85	158	24	S	N	143	0	1	1	0	0	[70.0, 0.0, 1.0, 85.0, 158.0, 24.0, 1.0, 0.0, 143.0]	[0.529, 19.471]	[0.026, 0.974]	1	S
126	31	M	S	86	224	20	N	N	34	0	1	0	0	0	[31.0, 0.0, 1.0, 86.0, 224.0, 20.0, 0.0, 0.0, 34.0]	[1.0, 19.0]	[0.05, 0.95]	1	S
127	67	M	N	75	154	24	N	N	3	0	0	0	0	0	[67.0, 0.0, 0.0, 75.0, 154.0, 24.0, 0.0, 0.0, 3.0]	[17.529, 2.471]	[0.876, 0.124]	0	N
128	67	F	S	96	239	26	N	N	79	1	1	0	0	0	[67.0, 1.0, 1.0, 96.0, 239.0, 26.0, 0.0, 0.0, 79.0]	[0.0, 20.0]	[0.0, 1.0]	1	S
129	64	F	N	86	216	22	N	N	94	1	0	0	0	0	[64.0, 1.0, 0.0, 86.0, 216.0, 22.0, 0.0, 0.0, 94.0]	[19.529, 0.471]	[0.976, 0.024]	0	N
130	49	M	N	70	192	30	N	N	164	0	0	0	0	0	[49.0, 0.0, 0.0, 70.0, 192.0, 30.0, 0.0, 0.0, 164.0]	[18.765, 1.235]	[0.938, 0.062]	0	N
131	72	F	S	127	164	22	S	N	41	1	1	1	0	0	[72.0, 1.0, 1.0, 127.0, 164.0, 22.0, 1.0, 0.0, 41.0]	[0.0, 20.0]	[0.0, 1.0]	1	S
132	50	F	N	103	219	31	N	N	60	1	0	0	0	0	[50.0, 1.0, 0.0, 103.0, 219.0, 31.0, 0.0, 0.0, 60.0]	[19.235, 0.765]	[0.962, 0.038]	0	N
133	50	F	S	106	193	32	S	N	149	1	1	1	0	0	[50.0, 1.0, 1.0, 106.0, 193.0, 32.0, 1.0, 0.0, 149.0]	[0.091, 19.909]	[0.005, 0.995]	1	S
134	55	F	N	81	157	22	N	N	107	1	0	0	0	0	[55.0, 1.0, 0.0, 81.0, 157.0, 22.0, 0.0, 0.0, 107.0]	[18.529, 1.471]	[0.926, 0.074]	0	N
135	59	F	N	112	235	21	N	N	67	1	0	0	0	0	[59.0, 1.0, 0.0, 112.0, 235.0, 21.0, 0.0, 0.0, 67.0]	[20.0, 0.0]	[1.0, 0.0]	0	N
136	68	M	S	84	232	25	N	N	97	0	1	0	0	0	[68.0, 0.0, 1.0, 84.0, 232.0, 25.0, 0.0, 0.0, 97.0]	[0.0, 20.0]	[0.0, 1.0]	1	S
137	57	F	S	84	202	27	S	N	143	1	1	1	0	0	[57.0, 1.0, 1.0, 84.0, 202.0, 27.0, 1.0, 0.0, 143.0]	[0.529, 19.471]	[0.026, 0.974]	1	S
138	71	F	N	87	173	20	N	N	245	1	0	0	0	0	[71.0, 1.0, 0.0, 87.0, 173.0, 20.0, 0.0, 0.0, 245.0]	[17.25, 2.75]	[0.863, 0.138]	0	N
139	39	F	S	147	242	23	S	N	58	1	1	1	0	0	[39.0, 1.0, 1.0, 147.0, 242.0, 23.0, 1.0, 0.0, 58.0]	[2.091, 17.909]	[0.105, 0.895]	1	S
140	40	F	S	111	166	30	S	N	143	1	1	1	0	0	[40.0, 1.0, 1.0, 111.0, 166.0, 30.0, 1.0, 0.0, 143.0]	[0.373, 19.627]	[0.019, 0.981]	1	S
141	71	M	S	116	219	23	S	N	140	0	1	1	0	0	[71.0, 0.0, 1.0, 116.0, 219.0, 23.0, 1.0, 0.0, 140.0]	[1.0, 19.0]	[0.05, 0.95]	1	S
142	60	F	N	99	194	28	N	N	169	1	0	0	0	0	[60.0, 1.0, 0.0, 99.0, 194.0, 28.0, 0.0, 0.0, 169.0]	[19.282, 0.718]	[0.964, 0.036]	0	N
143	31	M	S	139	152	26	S	N	101	0	1	1	0	0	[31.0, 0.0, 1.0, 139.0, 152.0, 26.0, 1.0, 0.0, 101.0]	[0.037, 19.963]	[0.002, 0.998]	1	S
144	37	F	S	114	182	31	S	N	183	1	1	1	0	0	[37.0, 1.0, 1.0, 114.0, 182.0, 31.0, 1.0, 0.0, 183.0]	[0.591, 19.409]	[0.03, 0.97]	1	S
145	71	M	N	102	196	25	N	N	126	0	0	0	0	0	[71.0, 0.0, 0.0, 102.0, 196.0, 25.0, 0.0, 0.0, 126.0]	[19.282, 0.718]	[0.964, 0.036]	0	N
146	72	M	N	102	229	28	N	N	102	0	0	0	0	0	[72.0, 0.0, 0.0, 102.0, 229.0, 28.0, 0.0, 0.0, 102.0]	[20.0, 0.0]	[1.0, 0.0]	0	N
147	62	M	S	108	160	34	N	N	141	0	1	0	0	0	[62.0, 0.0, 1.0, 108.0, 160.0, 34.0, 0.0, 0.0, 141.0]	[1.5, 18.5]	[0.075, 0.925]	1	S
148	49	F	S	114	199	30	N	N	51	1	1	0	0	0	[49.0, 1.0, 1.0, 114.0, 199.0, 30.0, 0.0, 0.0, 51.0]	[2.282, 17.718]	[0.114, 0.886]	1	S
149	35	M	S	134	214	25	S	N	28	0	1	1	0	0	[35.0, 0.0, 1.0, 134.0, 214.0, 25.0, 1.0, 0.0, 28.0]	[0.637, 19.363]	[0.032, 0.968]	1	S
150	32	M	N	83	178	28	N	N	35	0	0	0	0	0	[32.0, 0.0, 0.0, 83.0, 178.0, 28.0, 0.0, 0.0, 35.0]	[19.529, 0.471]	[0.976, 0.024]	0	N
151	59	F	S	120	223	29	S	N	51	1	1	1	0	0	[59.0, 1.0, 1.0, 120.0, 223.0, 29.0, 1.0, 0.0, 51.0]	[2.0, 18.0]	[0.1, 0.9]	1	S
152	57	M	N	87	206	26	N	N	110	0	0	0	0	0	[57.0, 0.0, 0.0, 87.0, 206.0, 26.0, 0.0, 0.0, 110.0]	[19.529, 0.471]	[0.976, 0.024]	0	N
153	63	F	N	101	205	22	N	N	46	1	0	0	0	0	[63.0, 1.0, 0.0, 101.0, 205.0, 22.0, 0.0, 0.0, 46.0]	[19.282, 0.718]	[0.964, 0.036]	0	N
154	58	F	S	71	213	28	S	N	168	1	1	1	0	0	[58.0, 1.0, 1.0, 71.0, 213.0, 28.0, 1.0, 0.0, 168.0]	[0.529, 19.471]	[0.026, 0.974]	1	S
155	49	F	N	72	150	29	N	N	36	1	0	0	0	0	[49.0, 1.0, 0.0, 72.0, 150.0, 29.0, 0.0, 0.0, 36.0]	[18.529, 1.471]	[0.926, 0.074]	0	N
156	48	M	S	107	223	31	S	N	195	0	1	1	0	0	[48.0, 0.0, 1.0, 107.0, 223.0, 31.0, 1.0, 0.0, 195.0]	[2.037, 17.963]	[0.102, 0.898]	1	S
157	71	M	S	92	190	20	N	N	246	0	1	0	0	0	[71.0, 0.0, 1.0, 92.0, 190.0, 20.0, 0.0, 0.0, 246.0]	[1.0, 19.0]	[0.05, 0.95]	1	S
158	67	F	N	113	162	21	N	N	174	1	0	0	0	0	[67.0, 1.0, 0.0, 113.0, 162.0, 21.0, 0.0, 0.0, 174.0]	[17.282, 2.718]	[0.864, 0.136]	0	N
159	46	F	N	90	177	20	N	N	221	1	0	0	0	0	[46.0, 1.0, 0.0, 90.0, 177.0, 20.0, 0.0, 0.0, 221.0]	[19.25, 0.75]	[0.963, 0.037]	0	N
160	55	M	S	72	157	26	N	N	39	0	1	0	0	0	[55.0, 0.0, 1.0, 72.0, 157.0, 26.0, 0.0, 0.0, 39.0]	[1.529, 18.471]	[0.076, 0.924]	1	S

Tesis publicada con autorización del autor
No olvide citar esta tesis

161	34	F	N	87	192	20	N	S	34	1	0	0	1	0	[34.0, 1.0, 0.0, 87.0, 192.0, 20.0, 0.0, 1.0, 34.0]	[18.529, 1.471]	[0.926, 0.074]	0	N
162	33	M	S	80	174	33	S	N	1	0	1	1	0	0	[33.0, 0.0, 1.0, 80.0, 174.0, 33.0, 1.0, 0.0, 1.0]	[1.037, 18.963]	[0.052, 0.948]	1	S
163	39	F	N	70	171	21	N	N	2	1	0	0	0	0	[39.0, 1.0, 0.0, 70.0, 171.0, 21.0, 0.0, 0.0, 2.0]	[19.529, 0.471]	[0.976, 0.024]	0	N
164	35	F	N	94	153	22	N	N	8	1	0	0	0	0	[35.0, 1.0, 0.0, 94.0, 153.0, 22.0, 0.0, 0.0, 8.0]	[18.282, 1.718]	[0.914, 0.086]	0	N
165	57	M	N	73	177	22	N	N	172	0	0	0	0	0	[57.0, 0.0, 0.0, 73.0, 177.0, 22.0, 0.0, 0.0, 172.0]	[17.779, 2.221]	[0.889, 0.111]	0	N
166	44	F	N	99	213	32	N	N	2	1	0	0	0	0	[44.0, 1.0, 0.0, 99.0, 213.0, 32.0, 0.0, 0.0, 2.0]	[19.0, 1.0]	[0.95, 0.05]	0	N
167	53	F	N	80	208	32	N	N	193	1	0	0	0	0	[53.0, 1.0, 0.0, 80.0, 208.0, 32.0, 0.0, 0.0, 193.0]	[18.909, 1.091]	[0.945, 0.055]	0	N
168	50	M	S	80	179	29	N	N	76	0	1	0	0	0	[50.0, 0.0, 1.0, 80.0, 179.0, 29.0, 0.0, 0.0, 76.0]	[1.765, 18.235]	[0.088, 0.912]	1	S
169	32	F	N	94	192	22	N	N	193	1	0	0	0	0	[32.0, 1.0, 0.0, 94.0, 192.0, 22.0, 0.0, 0.0, 193.0]	[19.282, 0.718]	[0.964, 0.036]	0	N
170	49	F	S	79	205	34	N	N	15	1	1	0	0	0	[49.0, 1.0, 1.0, 79.0, 205.0, 34.0, 0.0, 0.0, 15.0]	[0.0, 20.0]	[0.0, 1.0]	1	S
171	32	F	S	98	180	22	N	N	111	1	1	0	0	0	[32.0, 1.0, 1.0, 98.0, 180.0, 22.0, 0.0, 0.0, 111.0]	[1.282, 18.718]	[0.064, 0.936]	1	S
172	59	M	S	77	191	28	S	N	157	0	1	1	0	0	[59.0, 0.0, 1.0, 77.0, 191.0, 28.0, 1.0, 0.0, 157.0]	[0.529, 19.471]	[0.026, 0.974]	1	S
173	50	M	N	76	170	29	N	N	40	0	0	0	0	0	[50.0, 0.0, 0.0, 76.0, 170.0, 29.0, 0.0, 0.0, 40.0]	[18.529, 1.471]	[0.926, 0.074]	0	N
174	50	F	S	116	243	28	S	N	122	1	1	1	0	0	[50.0, 1.0, 1.0, 116.0, 243.0, 28.0, 1.0, 0.0, 122.0]	[1.091, 18.909]	[0.055, 0.945]	1	S
175	33	M	N	126	221	22	N	N	231	0	0	0	0	0	[33.0, 0.0, 0.0, 126.0, 221.0, 22.0, 0.0, 0.0, 231.0]	[20.0, 0.0]	[1.0, 0.0]	0	N
176	50	F	N	103	157	28	N	N	235	1	0	0	0	0	[50.0, 1.0, 0.0, 103.0, 157.0, 28.0, 0.0, 0.0, 235.0]	[20.0, 0.0]	[1.0, 0.0]	0	N
177	51	M	S	83	153	27	S	N	34	0	1	1	0	0	[51.0, 0.0, 1.0, 83.0, 153.0, 27.0, 1.0, 0.0, 34.0]	[0.566, 19.434]	[0.028, 0.972]	1	S
178	62	F	S	104	228	27	S	N	138	1	1	1	0	0	[62.0, 1.0, 1.0, 104.0, 228.0, 27.0, 1.0, 0.0, 138.0]	[1.0, 19.0]	[0.05, 0.95]	1	S
179	63	M	N	120	215	28	N	N	40	0	0	0	0	0	[63.0, 0.0, 0.0, 120.0, 215.0, 28.0, 0.0, 0.0, 40.0]	[19.0, 1.0]	[0.95, 0.05]	0	N
180	38	M	N	110	231	31	N	N	4	0	0	0	0	0	[38.0, 0.0, 0.0, 110.0, 231.0, 31.0, 0.0, 0.0, 4.0]	[19.75, 0.25]	[0.988, 0.013]	0	N
181	48	M	S	87	165	23	S	N	193	0	1	1	0	0	[48.0, 0.0, 1.0, 87.0, 165.0, 23.0, 1.0, 0.0, 193.0]	[0.566, 19.434]	[0.028, 0.972]	1	S
182	41	M	S	89	185	30	S	N	26	0	1	1	0	0	[41.0, 0.0, 1.0, 89.0, 185.0, 30.0, 1.0, 0.0, 26.0]	[1.566, 18.434]	[0.078, 0.922]	1	S
183	68	F	S	72	174	26	N	N	97	1	1	0	0	0	[68.0, 1.0, 1.0, 72.0, 174.0, 26.0, 0.0, 0.0, 97.0]	[0.779, 19.221]	[0.039, 0.961]	1	S
184	33	M	S	75	202	25	S	N	0	0	1	1	0	0	[33.0, 0.0, 1.0, 75.0, 202.0, 25.0, 1.0, 0.0, 0.0]	[0.566, 19.434]	[0.028, 0.972]	1	S
185	28	F	S	88	185	33	S	N	8	1	1	1	0	0	[28.0, 1.0, 1.0, 88.0, 185.0, 33.0, 1.0, 0.0, 8.0]	[1.091, 18.909]	[0.055, 0.945]	1	S
186	64	F	N	89	156	23	N	N	97	1	0	0	0	0	[64.0, 1.0, 0.0, 89.0, 156.0, 23.0, 0.0, 0.0, 97.0]	[17.529, 2.471]	[0.876, 0.124]	0	N
187	45	M	S	71	233	22	N	S	190	0	1	0	1	0	[45.0, 0.0, 1.0, 71.0, 233.0, 22.0, 0.0, 1.0, 190.0]	[0.0, 20.0]	[0.0, 1.0]	1	S
188	54	M	N	111	236	26	N	N	89	0	0	0	0	0	[54.0, 0.0, 0.0, 111.0, 236.0, 26.0, 0.0, 0.0, 89.0]	[20.0, 0.0]	[1.0, 0.0]	0	N
189	71	F	S	98	221	31	S	N	131	1	1	1	0	0	[71.0, 1.0, 1.0, 98.0, 221.0, 31.0, 1.0, 0.0, 131.0]	[0.0, 20.0]	[0.0, 1.0]	1	S
190	68	F	S	78	155	30	N	N	122	1	1	0	0	0	[68.0, 1.0, 1.0, 78.0, 155.0, 30.0, 0.0, 0.0, 122.0]	[0.765, 19.235]	[0.038, 0.962]	1	S
191	45	M	S	100	191	29	S	N	122	0	1	1	0	0	[45.0, 0.0, 1.0, 100.0, 191.0, 29.0, 1.0, 0.0, 122.0]	[0.319, 19.681]	[0.016, 0.984]	1	S
192	47	M	S	74	232	31	N	N	15	0	1	0	0	0	[47.0, 0.0, 1.0, 74.0, 232.0, 31.0, 0.0, 0.0, 15.0]	[0.75, 19.25]	[0.037, 0.963]	1	S
193	53	F	S	83	210	33	N	N	145	1	1	0	0	0	[53.0, 1.0, 1.0, 83.0, 210.0, 33.0, 0.0, 0.0, 145.0]	[0.0, 20.0]	[0.0, 1.0]	1	S
194	52	F	S	71	199	25	N	N	107	1	1	0	0	0	[52.0, 1.0, 1.0, 71.0, 199.0, 25.0, 0.0, 0.0, 107.0]	[0.529, 19.471]	[0.026, 0.974]	1	S
195	49	F	S	96	228	28	N	N	196	1	1	0	0	0	[49.0, 1.0, 1.0, 96.0, 228.0, 28.0, 0.0, 0.0, 196.0]	[0.0, 20.0]	[0.0, 1.0]	1	S
196	48	F	N	95	233	22	N	N	160	1	0	0	0	0	[48.0, 1.0, 0.0, 95.0, 233.0, 22.0, 0.0, 0.0, 160.0]	[20.0, 0.0]	[1.0, 0.0]	0	N
197	40	M	S	106	168	33	S	N	142	0	1	1	0	0	[40.0, 0.0, 1.0, 106.0, 168.0, 33.0, 1.0, 0.0, 142.0]	[0.037, 19.963]	[0.002, 0.998]	1	S
198	31	F	S	73	223	32	S	N	130	1	1	1	0	0	[31.0, 1.0, 1.0, 73.0, 223.0, 32.0, 1.0, 0.0, 130.0]	[0.091, 19.909]	[0.005, 0.995]	1	S
199	36	F	S	81	201	30	S	N	41	1	1	1	0	0	[36.0, 1.0, 1.0, 81.0, 201.0, 30.0, 1.0, 0.0, 41.0]	[0.62, 19.38]	[0.031, 0.969]	1	S
200	36	F	S	105	205	29	N	N	108	1	1	0	0	0	[36.0, 1.0, 1.0, 105.0, 205.0, 29.0, 0.0, 0.0, 108.0]	[0.517, 19.483]	[0.026, 0.974]	1	S
201	71	F	N	124	230	25	S	N	217	1	0	1	0	0	[71.0, 1.0, 0.0, 124.0, 230.0, 25.0, 1.0, 0.0, 217.0]	[20.0, 0.0]	[1.0, 0.0]	0	N
202	53	F	S	73	217	25	S	N	107	1	1	1	0	0	[53.0, 1.0, 1.0, 73.0, 217.0, 25.0, 1.0, 0.0, 107.0]	[0.529, 19.471]	[0.026, 0.974]	1	S
203	41	M	S	140	227	25	S	N	92	0	1	1	0	0	[41.0, 0.0, 1.0, 140.0, 227.0, 25.0, 1.0, 0.0, 92.0]	[1.037, 18.963]	[0.052, 0.948]	1	S
204	43	M	N	92	161	24	N	N	182	0	0	0	0	0	[43.0, 0.0, 0.0, 92.0, 161.0, 24.0, 0.0, 0.0, 182.0]	[19.282, 0.718]	[0.964, 0.036]	0	N
205	67	M	S	131	161	33	N	N	51	0	1	0	0	0	[67.0, 0.0, 1.0, 131.0, 161.0, 33.0, 0.0, 0.0, 51.0]	[1.0, 19.0]	[0.05, 0.95]	1	S
206	28	F	N	74	227	31	N	N	152	1	0	0	0	0	[28.0, 1.0, 0.0, 74.0, 227.0, 31.0, 0.0, 0.0, 152.0]	[18.235, 1.765]	[0.912, 0.088]	0	N
207	64	M	N	74	193	22	N	N	183	0	0	0	0	0	[64.0, 0.0, 0.0, 74.0, 193.0, 22.0, 0.0, 0.0, 183.0]	[19.529, 0.471]	[0.976, 0.024]	0	N
208	57	F	N	75	208	32	N	N	134	1	0	0	0	0	[57.0, 1.0, 0.0, 75.0, 208.0, 32.0, 0.0, 0.0, 134.0]	[18.235, 1.765]	[0.912, 0.088]	0	N
209	35	F	N	80	191	26	N	N	238	1	0	0	0	0	[35.0, 1.0, 0.0, 80.0, 191.0, 26.0, 0.0, 0.0, 238.0]	[20.0, 0.0]	[1.0, 0.0]	0	N
210	34	F	S	94	233	25	N	N	194	1	1	0	0	0	[34.0, 1.0, 1.0, 94.0, 233.0, 25.0, 0.0, 0.0, 194.0]	[0.0, 20.0]	[0.0, 1.0]	1	S
211	40	M	S	81	212	34	S	N	24	0	1	1	0	0	[40.0, 0.0, 1.0, 81.0, 212.0, 34.0, 1.0, 0.0, 24.0]	[0.037, 19.963]	[0.002, 0.998]	1	S
212	37	M	N	110	185	27	N	N	6	0	0	0	0	0	[37.0, 0.0, 0.0, 110.0, 185.0, 27.0, 0.0, 0.0, 6.0]	[19.282, 0.718]	[0.964, 0.036]	0	N
213	51	F	N	74	194	32	N	N	36	1	0	0	0	0	[51.0, 1.0, 0.0, 74.0, 194.0, 32.0, 0.0, 0.0, 36.0]	[19.0, 1.0]	[0.95, 0.05]	0	N
214	39	F	S	82	220	27	S	N	149	1	1	1	0	0	[39.0, 1.0, 1.0, 82.0, 220.0, 27.0, 1.0, 0.0, 149.0]	[0.091, 19.909]	[0.005, 0.995]	1	S
215	66	M	S	127	242	22	S	N	30	0	1	1	0	0	[66.0, 0.0, 1.0, 127.0, 242.0, 22.0, 1.0, 0.0, 30.0]	[1.6, 18.4]	[0.08, 0.92]	1	S
216	58	F	N	74	188	22	N	N	130	1	0	0	0	0	[58.0, 1.0, 0.0, 74.0, 188.0, 22.0, 0.0, 0.0, 130.0]	[19.529, 0.471]	[0.976, 0.024]	0	N
217	56	F	S	115	172	30	N	N	183	1	1	0	0	0	[56.0, 1.0, 1.0, 115.0, 172.0, 30.0, 0.0, 0.0, 183.0]	[1.159, 18.841]	[0.058, 0.942]	1	S
218	33	F	N	77	166	31	N	N	91	1	0	0	0	0	[33.0, 1.0, 0.0, 77.0, 166.0, 31.0, 0.0, 0.0, 91.0]	[18.235, 1.765]	[0.912, 0.088]	0	N
219	65	F	N	83	164	23	N	N	2	1	0	0	0	0	[65.0, 1.0, 0.0, 83.0, 164.0, 23.0, 0.0, 0.0, 2.0]	[18.529, 1.471]	[0.926, 0.074]	0	N
220	29	M	N	106	172	25	N	N	226	0	0	0	0	0	[29.0, 0.0, 0.0, 106.0, 172.0, 25.0, 0.0, 0.0, 226.0]	[19.0, 1.0]	[0.95, 0.05]	0	N
221	56	M	N	99	175	32	N	N	228	0	0	0	0	0	[56.0, 0.0, 0.0, 99.0, 175.0, 32.0, 0.0, 0.0, 228.0]	[18.159, 1.841]	[0.908, 0.092]	0	N

Tesis publicada con autorización del autor
No olvide citar esta tesis

Anexo 10. Diagrama de Capas de la Solución Propuesta



Fuente: Elaboración Propia

Tesis publicada con autorización del autor
No olvide citar esta tesis