



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

### 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**

### **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.

### **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.

**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
<b>CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA</b> .....	<b>3</b>
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
<b>CAPÍTULO II: MARCO TEÓRICO</b> .....	<b>11</b>
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

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
<b>CAPÍTULO III: DISEÑO METODOLÓGICO .....</b>		<b>79</b>
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
<b>CAPÍTULO IV: RESULTADOS .....</b>		<b>86</b>
<b>CAPÍTULO V: DISCUSIÓN, CONCLUSIONES Y RECOMENDACIONES .....</b>		<b>111</b>
<b>REFERENCIAS BIBLIOGRÁFICAS .....</b>		<b>117</b>
<b>GLOSARIO DE TÉRMINOS Y ABREVIATURAS.....</b>		<b>123</b>
<b>ANEXOS .....</b>		<b>125</b>

## ÍNDICE DE FIGURAS

<i>Figura 1.</i> Funcionamiento de un Algoritmo Tradicional.....	19
<i>Figura 2.</i> Funcionamiento de un Algoritmo de Machine Learning .....	20
<i>Figura 3.</i> Clasificación de Algoritmos de Machine Learning .....	22
<i>Figura 4.</i> Algoritmo de Agrupamiento (Clustering).....	26
<i>Figura 5.</i> Aplicación Web desde Multidispositivos .....	31
<i>Figura 6.</i> Funcionalidad de una Aplicación Web .....	34
<i>Figura 7.</i> Proceso para Desarrollo de Aplicación Web .....	36
<i>Figura 8.</i> Matriz de Confusión .....	39
<i>Figura 9.</i> Fórmula para Métrica de Exactitud.....	40
<i>Figura 10.</i> Fórmula para Métrica de Sensibilidad .....	41
<i>Figura 11.</i> Fórmula para Métrica de Especificidad.....	42
<i>Figura 12.</i> Metodología Fundamental para Ciencia de Datos.....	46
<i>Figura 13.</i> Categorías de la Presión Arterial.....	56
<i>Figura 14.</i> Accidente Cerebrovascular Isquémico .....	68
<i>Figura 15.</i> Accidente Cerebrovascular Hemorrágico.....	69
<i>Figura 16.</i> Técnica para la Muestra .....	80
<i>Figura 17.</i> Fórmula para Muestra de Población Finita .....	81
<i>Figura 18.</i> Estructura del Dataset Recopilado .....	93
<i>Figura 19.</i> Histograma por cada Variable Numérica .....	94
<i>Figura 20.</i> Gráfico de Barras por cada Variable Categórica .....	96
<i>Figura 21.</i> Carga de Datos .....	97
<i>Figura 22.</i> División de los Datos .....	97
<i>Figura 23.</i> Importación de Librerías necesarias .....	98

<i>Figura 24.</i> Conversión de Variables categóricas a numéricas .....	98
<i>Figura 25.</i> Parámetros para el Algoritmo .....	98
<i>Figura 26.</i> Conversión de variables y Creación del Pipeline.....	99
<i>Figura 27.</i> Entrenamiento del Modelo de Random Forest .....	99
<i>Figura 28.</i> Matriz de Confusión usando Test Dataset .....	100
<i>Figura 29.</i> Métricas de Evaluación usando Test Dataset .....	100
<i>Figura 30.</i> Implementación de la Solución .....	101
<i>Figura 31.</i> Matriz de Confusión usando Data de Validación .....	103
<i>Figura 32.</i> Métricas de Desempeño para Validación .....	103
<i>Figura 33.</i> Histograma por Variable Numérica de la Muestra .....	104
<i>Figura 34.</i> Gráfico de Barras por Variable Categórica de la Muestra.....	106
<i>Figura 35.</i> Resultado de la Prueba de Chi-cuadrado .....	108



## ÍNDICE DE TABLAS

Tabla 1. <i>Examen de Glucosa en Sangre en Ayunas</i> .....	57
Tabla 2. <i>Análisis de Colesterol Total en Sangre</i> .....	59
Tabla 3. <i>Análisis de Colesterol HDL en Sangre</i> .....	59
Tabla 4. <i>Análisis de Colesterol LDL en Sangre</i> .....	60
Tabla 5. <i>Niveles del Índice de Masa Corporal (IMC)</i> .....	61
Tabla 6. <i>Detalle de los datos</i> .....	92
Tabla 7. <i>Estadística Descriptiva de Variables Numéricas</i> .....	93
Tabla 8. <i>Estadística Descriptiva de Variables Categóricas</i> .....	95
Tabla 9. <i>Comparación: Real vs. Predicción</i> .....	102
Tabla 10. <i>Resumen Estadístico: Variables Numéricas de la Muestra</i> .....	104
Tabla 11. <i>Resumen Estadístico: Variables Categóricas de la Muestra</i> .....	106
Tabla 12. <i>Random Forest: Feature Importance</i> .....	110
Tabla 13. <i>Comparación de métricas entre modelos</i> .....	111

**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

## 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

## 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