

# TABLA DE CONTENIDO

<b>INTRODUCCIÓN.....</b>	<b>1</b>
1.1 Antecedentes generales.....	1
1.1.1 Características de la organización .....	1
1.1.2 Marco institucional .....	4
1.2 Descripción del proyecto y justificación .....	5
1.2.1 Información del área de la organización.....	5
1.2.2 Identificar el problema u oportunidad y su relevancia, con sus efectos y posibles causas. .....	8
1.2.3 Identificar hipótesis y posibles alternativas de solución para resolver el problema u oportunidad .....	10
1.2.4 Propuesta de valor de las posibles soluciones o impacto del cambio propuesto .....	13
1.3 Objetivos .....	14
1.3.1 Objetivo General.....	14
1.3.2 Objetivos Específicos .....	14
1.4 Marco conceptual.....	15
1.4.1 Bases de Datos .....	15
1.4.2 Python.....	15
1.4.3 Algoritmos de aprendizaje de máquinas .....	15
1.4.4 Algoritmos Generadores de datos.....	16
1.4.5 Evaluación de modelos generadores .....	20
1.4.6 Interpretabilidad versus funcionalidad de los modelos.....	24
1.5. Metodología.....	25
1.6 Alcances.....	27
<b>ESTADO DEL ARTE .....</b>	<b>28</b>
2.1 Problemas asociados a tener escasez de datos .....	28
2.2 Utilización de herramientas de aprendizaje de máquinas en el área de la salud.....	30
2.3 Generación de datos en la literatura .....	31
2.4 Arquitectura de los algoritmos.....	32
2.1.1 Conditional Tabular GAN (CTGAN):.....	32
2.1.2 Tabular VAE (TVAE):.....	33
2.1.3 Gaussian Copula (GC) .....	34
<b>GENERACIÓN Y MANIPULACIÓN DE DATOS REALES .....</b>	<b>35</b>
3.1 Proceso de Búsqueda de Pacientes .....	36
3.2 Realización de los exámenes .....	37
3.3 Procesamiento de los datos.....	39
3.4 Generalización del proceso.....	41
<b>GENERACIÓN DE DATOS SINTÉTICOS.....</b>	<b>43</b>
4.1 Características de las bases de datos .....	43
4.2 Implementación de los algoritmos generativos.....	44
4.3 Indicadores de la generación de datos .....	50
<b>EVALUACIÓN DE ALGORITMOS GENERADORES.....</b>	<b>56</b>
5.1 Índice de similitud.....	56
5.2 Evaluación utilizando un modelo de clasificación .....	59
5.2.1 Datos sintéticos .....	60
5.2.2 Datos mixtos .....	63

5.3 Análisis de sensibilidad .....	66
<b>EVALUACIÓN DEL IMPACTO ECONÓMICO .....</b>	<b>69</b>
6.1 Costos de generar exámenes .....	69
6.2 Mercado de investigación .....	72
6.3 Beneficios de utilizar datos sintéticos .....	73
<b>DATOS SINTÉTICOS COMO UN PRODUCTO O SERVICIO .....</b>	<b>76</b>
7.1 Aplicaciones que utilizan datos sintéticos .....	76
7.1.1 M-Sense .....	76
7.1.2 Medkit-learn .....	76
7.1.3 DECAF .....	76
7.2 Restricciones en el área de la salud .....	77
7.3 Modelando datos sintéticos como un producto o servicio .....	78
7.4 Requerimientos funcionales y no funcionales .....	79
<b>CONCLUSIONES Y RECOMENDACIONES .....</b>	<b>83</b>
8.1 Conclusiones .....	83
8.2 Recomendaciones .....	84
<b>BIBLIOGRAFÍA.....</b>	<b>88</b>
<b>ANEXOS .....</b>	<b>97</b>
A. Resultados de la generación de datos .....	97
A.1. Test de navegación .....	97
A.2. Cámara de movimiento ocular .....	98
A.3. Electroencefalograma .....	99
A.4. Base de datos ya procesados y unificados .....	100
A.5. 150 Datos para base ya procesada y unificada .....	102
B. Mercado de investigación relacionada al área de la salud .....	103