

# Tabla de contenido

<b>Resumen Ejecutivo</b>	<b>I</b>
<b>Agradecimientos</b>	<b>III</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Contexto	1
1.1.1. Registro Clínico Electrónico	2
1.1.2. Minería de Datos en Campos de Texto	3
1.1.3. Situación Actual y Oportunidad en la Salud Pública	4
1.1.4. Descripción del Problema y Justificación	7
1.2. Objetivos	10
1.2.1. Objetivo General	10
1.2.2. Objetivos Específicos	10
1.3. Hipótesis de Investigación	10
1.4. Metodología	11
1.5. Resultados Esperados	12
1.6. Contribución de la Memoria	12
1.7. Estructura del Contenido	13
<b>2. Marco Conceptual</b>	<b>15</b>
2.1. CRISP-DM	15
2.2. Data Mining y Machine Learning	17
2.2.1. Problema de Alta Dimensionalidad	18
2.2.2. Modelos de Clasificación	19
2.2.3. Evaluación de Algoritmos Supervisados	20
2.2.4. Itemset Mining	24
2.3. Text Mining y Natural Language Processing	26
2.4. Registro Clínico Electrónico	30
2.5. Prevención del Riesgo Cardiovascular	31

2.6.	Otros trabajos que utilizan el RCE . . . . .	32
2.6.1.	Trabajo 1: Modelamiento Predictivo usando datos de RCE . . . . .	32
2.6.2.	Trabajo 2: Identificación Automática de Factores de Riesgo Cardiovascular, usando análisis de texto del RCE . . . . .	36
<b>3.</b>	<b>Caracterización de los Datos sobre Enfermedades Cardiovasculares</b>	<b>39</b>
3.1.	Entendimiento del Negocio . . . . .	40
3.1.1.	Salud Pública en Chile . . . . .	40
3.1.2.	Las Enfermedades Cardiovasculares . . . . .	45
3.1.3.	La Aterosclerosis . . . . .	48
3.1.4.	Insuficiencia Cardiovascular . . . . .	59
3.1.5.	Programa de Salud Cardiovascular . . . . .	64
3.2.	Entendimiento de los Datos . . . . .	66
3.2.1.	Módulos Principales . . . . .	66
3.2.2.	Procesos Principales . . . . .	69
<b>4.</b>	<b>Propuesta de Investigación</b>	<b>72</b>
4.1.	Detalles Propuesta de Investigación . . . . .	72
4.1.1.	Smarter Care . . . . .	72
4.1.2.	Estrategias Preventivas para las ECV . . . . .	80
4.1.3.	Healthcare Analytics . . . . .	82
4.2.	Diseño del Experimento . . . . .	87
4.2.1.	Plataforma Healthcare Analytics . . . . .	87
4.2.2.	Resultados Esperados . . . . .	90
<b>5.</b>	<b>Experimento y Resultados</b>	<b>95</b>
5.1.	Preparación de los Datos . . . . .	95
5.1.1.	Extracción de Datos . . . . .	95
5.1.2.	Selección de Variables . . . . .	100
5.2.	Análisis de Contenido . . . . .	102
5.2.1.	Diseño Análisis de Contenido . . . . .	103
5.2.2.	Resultados Análisis de Contenido . . . . .	106
5.3.	Análisis Predictivo . . . . .	112
5.3.1.	Modelamiento Análisis Predictivo . . . . .	113
5.3.2.	Resultados Análisis Predictivo . . . . .	116
5.4.	Discusiones . . . . .	121

5.4.1. Validación de los resultados . . . . .	121
5.4.2. Uso de los resultados del estudio . . . . .	122
5.4.3. Comparación con otros estudios . . . . .	123
5.4.4. Aprendizaje a partir de los datos . . . . .	124
5.4.5. Impacto en la Gestión de la Salud Preventiva . . . . .	125
5.4.6. Aprovechar cada atención al máximo . . . . .	125
5.4.7. Investigaciones Relacionadas . . . . .	126
5.5. Trabajos Futuros . . . . .	127
5.5.1. Similitud de Pacientes . . . . .	128
<b>6. Conclusiones</b>	<b>129</b>
6.1. Resultados Esperados . . . . .	129
6.2. Comentarios Finales . . . . .	130
<b>Bibliografía</b>	<b>132</b>
<b>Apéndices</b>	<b>136</b>
A . Tablas de Framingham . . . . .	136
B . Modelo Atenciones . . . . .	139
C . Diccionario de Datos . . . . .	141
D . Procesamiento de Texto en R . . . . .	148
E . Query de Extracción . . . . .	149
F . Procesamiento en Rapid Miner - Análisis de Contenido . . . . .	150
G . Procesamiento en Rapid Miner - Análisis Predictivo . . . . .	151