

# Tabla de Contenido

<b>Índice de Tablas</b>	<b>vii</b>
<b>Índice de Figuras</b>	<b>viii</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Antecedentes Generales . . . . .	1
1.1.1. WIC . . . . .	1
1.1.2. Hospital Clínico de la Universidad de Chile . . . . .	2
1.2. Justificación . . . . .	3
1.2.1. Unidades de Cuidados Intensivos . . . . .	3
1.2.2. Adultos Mayores Hospitalizados en Unidades de Cuidados Intensivos . . . . .	4
1.2.3. Herramientas de Apoyo a la Decisión . . . . .	7
1.2.4. Propuesta de solución . . . . .	8
1.3. Hipótesis de Investigación . . . . .	8
1.4. Objetivos . . . . .	8
1.4.1. Objetivo General . . . . .	8
1.4.2. Objetivos Específicos . . . . .	8
1.5. Metodología . . . . .	9
1.6. Resultados Esperados . . . . .	10
1.7. Alcances . . . . .	10
1.8. Estructura del Informe . . . . .	11
<b>2. Marco Teórico</b>	<b>13</b>
2.1. Data Mining . . . . .	13
2.1.1. KDD y Data Mining . . . . .	13
2.1.2. Machine Learning . . . . .	15
2.1.3. Aprendizaje Supervisado y No Supervisado . . . . .	15
2.2. Missing Data . . . . .	15
2.2.1. Tipos de Valores Faltantes . . . . .	15
2.2.2. Manejo de Datos Faltantes . . . . .	16
2.3. Selección de Atributos . . . . .	17
2.3.1. F-Score . . . . .	18
2.3.2. Correlación de Pearson . . . . .	19
2.4. Métodos de Clasificación . . . . .	20
2.4.1. Regresión Logística . . . . .	20
2.4.2. Naïve Bayes . . . . .	21
2.4.3. Support Vector Machines(SVM) . . . . .	21

2.4.4. Redes Neuronales Artificiales . . . . .	22
2.5. Validación de Modelos . . . . .	23
2.5.1. Entrenamiento y Prueba del Modelo . . . . .	24
2.5.2. Tipos de Error y Medidas de Evaluación . . . . .	26
2.5.3. Curva Receiver Operating Characteristics (ROC) . . . . .	27
<b>3. Modelos de Predicción de Mortalidad</b>	<b>29</b>
<b>4. Diseño del Modelo</b>	<b>32</b>
4.1. Selección y Extracción de Datos . . . . .	32
4.1.1. Ficha Clínica Electrónica . . . . .	32
4.1.2. Farmacia . . . . .	33
4.1.3. Laboratorio . . . . .	34
4.1.4. Base de Datos Preliminar . . . . .	34
4.1.5. Diseño del cohorte . . . . .	35
4.1.6. Estadística Descriptiva . . . . .	37
4.2. Selección y Transformación de Variables . . . . .	41
4.2.1. Transformación Variables Diagnóstico Principal . . . . .	41
4.2.2. Transformación Variables Diagnósticos Secundarios . . . . .	43
4.2.3. Transformación Variables Procedimientos . . . . .	45
4.2.4. Transformación Variables Farmacia . . . . .	45
4.2.5. Transformación Variables Exámenes de Laboratorio . . . . .	46
4.3. Manejo de Datos Faltantes . . . . .	47
<b>5. Modelamiento</b>	<b>51</b>
5.1. Selección de Variables . . . . .	51
5.2. Modelo 24 horas . . . . .	54
5.2.1. Selección de Variables Modelo 24 horas . . . . .	54
5.2.2. Resultados Entrenamiento Modelo 24 horas . . . . .	57
5.2.3. Interpretación Modelo 24 horas . . . . .	59
5.3. Modelo 48 horas . . . . .	63
5.3.1. Selección de Variables Modelo 48 horas . . . . .	63
5.3.2. Resultados Entrenamiento Modelo 48 horas . . . . .	67
5.3.3. Interpretación Modelo 48 horas . . . . .	69
5.4. Modelo 72 horas . . . . .	72
5.4.1. Selección de Variables Modelo 72 horas . . . . .	72
5.4.2. Resultados Entrenamiento Modelo 72 horas . . . . .	75
5.4.3. Interpretación Modelo 72 horas . . . . .	77
5.5. Discusión . . . . .	81
5.5.1. Análisis y Comparación de Resultados . . . . .	81
5.5.2. Opinión Experta . . . . .	83
<b>6. Conclusiones y Trabajo Futuro</b>	<b>85</b>
6.1. Conclusiones . . . . .	85
6.2. Trabajo Futuro . . . . .	87
<b>7. Bibliografía</b>	<b>89</b>

<b>A. Resultados Análisis Puntaje F</b>	<b>96</b>
A.1. Variables y puntaje F para corte p-valor=0.3 para modelo 24 horas . . . . .	96
A.2. Variables y puntaje F para corte p-valor=0.3 para modelo 48 horas . . . . .	99
A.3. Variables y puntaje F para corte p-valor=0.25 para modelo 72 horas . . . . .	102
<b>B. Coeficientes Logit</b>	<b>106</b>
B.1. Variables y coeficiente $\beta$ para modelo 24 horas . . . . .	106
B.2. Variables y coeficiente $\beta$ para modelo 48 horas . . . . .	109
B.3. Variables y coeficiente $\beta$ para modelo 72 horas . . . . .	112

# Índice de Tablas

3.1. Sistemas Scoring Clásicos . . . . .	30
3.2. Modelos predictivos machine learning . . . . .	31
4.1. Descripción de atributos set de datos ficha clínica electrónica . . . . .	33
4.2. Descripción variables exámenes de laboratorio . . . . .	34
4.3. Resumen variables extracción inicial . . . . .	35
4.4. Clasificación de comorbilidades de Elixhauser CIE-10 . . . . .	43
4.5. Clasificación de Medicamentos . . . . .	46
4.6. Variables derivadas resultados laboratorio . . . . .	47
4.7. Portcentaje de complitud variables de laboratorio por configuración de predicción	47
4.8. Variables de rango laboratorio . . . . .	48
5.1. Resumen variables transformadas . . . . .	52
5.2. Puntaje F modelo 24 horas . . . . .	55
5.3. Resumen resultados promedio AUC de validación cruzada para cuatro casificado- res, según cada umbral de corte para base de datos 24 horas . . . . .	56
5.4. Resumen resultados entrenamiento de modelo de 24 horas para cuatro clasifica- dores . . . . .	57
5.5. Puntaje F modelo 48 horas . . . . .	65
5.6. Resumen resultados promedio AUC de validación cruzada para cuatro casificado- res, según cada umbral de corte para base de datos 48 horas . . . . .	66
5.7. Resumen resultados entrenamiento de modelo de 48 horas para cuatro clasifica- dores . . . . .	67
5.8. Puntaje F modelo 72 horas . . . . .	73
5.9. Resumen resultados promedio AUC de validación cruzada para cuatro casificado- res, según cada umbral de corte para base de datos 72 horas . . . . .	74
5.10. Resumen resultados entrenamiento de modelo de 72 horas para cuatro clasifica- dores . . . . .	75
A.1. Puntaje F variables relevantes, modelo 24 horas . . . . .	96
A.2. Puntaje F variables relevantes, modelo 48 horas . . . . .	99
A.3. Puntaje F variables relevantes, modelo 72 horas . . . . .	102
B.1. Variables y coeficiente $\beta$ para modelo 24 horas . . . . .	106
B.2. Variables y coeficiente $\beta$ para modelo 48 horas . . . . .	109
B.3. Variables y coeficiente $\beta$ para modelo 72 horas . . . . .	112

# Índice de Figuras

2.1. Proceso KDD . . . . .	14
2.2. <i>Scatterplot de data set de dos clases, con hiperplanos de máxima distancia encontrados mediante algoritmo SVM con núcleo linear. Las instancias que coinciden en los hiperplanos de máxima distancia (puntos rojos) se llaman vectores de soporte</i> . . . . .	22
2.3. Estructura MLP . . . . .	23
2.4. <i>Esquema de división del conjunto de datos para validación cruzada de 4 folds</i> . . . . .	25
2.5. <i>Validación cruzada "Leave one out"(LOOCV)</i> . . . . .	25
2.6. <i>Estructura de una matriz de confusión para problema de clasificación binario</i> . . . . .	26
2.7. <i>Espacio ROC</i> . . . . .	28
4.1. Cantidad de observaciones disponibles por configuración de predicción . . . . .	37
4.2. Histogramas variables demográficas y descriptores generales . . . . .	38
4.3. Histograma variables laboratorio . . . . .	39
4.4. Histograma variables laboratorio . . . . .	40
4.5. Jerarquía CIE 10 representada por cuatro niveles . . . . .	41
4.6. Distribución Diagnóstico-Paciente Nivel 4 Jerarquía CIE-10 . . . . .	42
4.7. Distribución Diagnóstico-Paciente Nivel 3 Jerarquía CIE-10 . . . . .	42
5.1. Mapa de calor covariables modelo 24 horas . . . . .	54
5.2. Puntaje F modelo 24 horas . . . . .	55
5.3. Gráfico promedio AUC de validación cruzada para cuatro clasificadores, según cada umbral de corte para base de datos 24 horas . . . . .	56
5.4. Comparación curva ROC promedio para cuatro clasificadores en modelo 24 horas . . . . .	58
5.5. Curva ROC para cuatro clasificadores en modelo 24 horas. La curva azul opaca corresponde al promedio de cada curva generada por las iteraciones de la validación cruzada, las cuales se muestran en azul claro . . . . .	59
5.6. Coeficientes Logit modelo 24 horas . . . . .	60
5.7. Puntaje F modelo 48 horas . . . . .	64
5.8. Gráfico promedio AUC de validación cruzada para cuatro clasificadores, según cada umbral de corte para base de datos 48 horas . . . . .	66
5.9. Curva ROC para cuatro clasificadores en modelo 48 horas. La curva azul opaca corresponde al promedio de cada curva generada por las iteraciones de la validación cruzada, las cuales se muestran en azul claro . . . . .	68
5.10. Comparación curva ROC promedio para cuatro clasificadores en modelo 48 horas . . . . .	69
5.11. Coeficientes Logit para 20 variables más relevantes, modelo 48 horas . . . . .	70
5.12. Puntaje F modelo 72 horas . . . . .	73

5.13. Gráfico promedio AUC de validación cruzada para cuatro clasificadores, según cada umbral de corte para base de datos 72 horas . . . . .	74
5.14. Curva ROC para cuatro clasificadores en modelo 72 horas. La curva azul opaca corresponde al promedio de cada curva generada por las iteraciones de la validación cruzada, las cuales se muestran en azul claro . . . . .	76
5.15. Comparación curva ROC promedio para cuatro clasificadores en modelo 72 horas	77
5.16. Coeficientes Logit modelo 72 horas . . . . .	78
5.17. Comparación AUC . . . . .	81