

TABLA DE CONTENIDO

1.	Introducción	1
1.1	Exposición del Tema	1
1.2	Justificación.....	1
1.3	Objetivos	2
1.3.1	Objetivos Generales	2
1.3.2	Objetivos Específicos.....	2
1.4	Metodología.....	3
1.5	Alcances.....	4
1.6	Resultados Esperados	5
1.7	Planificación	5
2.	Administración Tributaria en Chile	7
2.1	Servicio de Impuestos Internos (SII)	7
2.2	Impuestos.....	8
2.2.1	Impuesto a la Renta	10
2.2.2	IVA.....	10
2.3	Brecha Tributaria.....	11
2.3.1	Evasión	12
2.4	Segmentación de Contribuyentes	12
2.4.1	Micro y Pequeña Empresa	14
3.	Credit Scoring	14
3.1	Metodología KDD	15
3.2	Definiciones Preliminares	17
3.2.1	Casos “Buenos” y “Malos”	17
3.2.2	Ventana de Muestra y de Comportamiento	17
3.3	Desarrollo del Modelo de <i>Credit Scoring</i>	18
3.3.1	Elección de la Muestra	18
3.3.2	Selección de los Datos	19
3.3.3	Preprocesamiento	20
3.3.4	Transformación de Datos	21
3.3.5	Creación de Modelos (Técnicas de Data Mining)	22
3.3.6	Comparación de Resultados	26
3.3.7	Consideraciones.....	27
4.	Descripción de los Datos	27

4.1	Datos Iniciales	28
4.2	Definición de Default	29
4.3	Elección de la Muestra	31
4.4	Definición de las Variables Independientes.....	32
5.	Tratamiento de Datos	33
5.1	Tratamiento de Datos Nulos.....	33
5.2	Tratamiento de Outliers.....	34
5.3	Selección de Variables.....	35
5.4	Transformación de Variables.....	36
6.	Aplicación de Modelos	36
6.1	Técnicas.....	37
6.1.1	Árboles de Decisión.....	38
6.1.2	Redes Neuronales	39
6.1.3	Regresión Logística	39
6.2	Prueba Ex-Ante	39
6.2.1	Árboles de Decisión.....	39
6.2.2	Redes Neuronales	39
6.2.3	Regresión Logística	40
6.3	Prueba General.....	41
6.3.1	Árboles de Decisión.....	41
6.3.2	Redes Neuronales	41
6.3.3	Regresión Logística	42
6.4	Iteraciones.....	42
6.4.1	Árboles de Decisión.....	43
6.4.2	Redes Neuronales	44
6.4.3	Regresión Logística	46
6.5	Modelo Final.....	48
6.5.1	Árboles de Decisión.....	48
6.5.2	Redes Neuronales	49
6.5.3	Regresión Logística	51
7.	Resultados	52
7.1	Análisis Validación y Robustez.....	53
7.1.1	Árboles de Decisión.....	54
7.1.2	Regresión Logística	55

7.2	Interpretación	57
7.3.1	Árboles de Decisión	58
7.3.2	Regresión Logística	60
7.3	Comportamiento por Zona Geográfica	63
8.	Conclusiones	65
8.1	Conclusiones Específicas.....	66
8.2	Propuestas de Mejora	68
8.3	Recomendaciones de Uso	69
9.	Bibliografía	71
10.	Anexos.....	73
	Anexo 1: Formulario 29	73
	Anexo 2: Interfaz de IBM SPSS Modeler	75
	Anexo 3: Campos Seleccionados.....	76
	Anexo 4: Variables Independiente.....	80
	Anexo 5: Estadísticos Descriptivos Previo a la Limpieza de Datos	82
	Anexos 6: Resumen de Tratamiento de Datos Nulos.....	84
	Anexos 7: Resumen de Tratamiento de Outliers.....	86
	Anexo 8: Importancia de Variables.....	88
	Anexo 9: Resumen de Matriz de Correlación	89
	Anexo 10: Varianza Total Explicada por Método ACP	90
	Anexo 11: Categorización de Variables.	92
	Anexo 12: Detalle de Transformación de Variables.	94
	Anexo 13: Matriz de Costos para la Técnica de Árboles de Decisión.	96
	Anexo 14: Iteraciones para la Técnica de Árboles de Decisión.....	97
	Anexo 15: Iteraciones para la Técnica de Redes Neuronales.....	98
	Anexo 16: Iteraciones para la Técnica de Regresión Logística.....	99
	Anexo 17: Iteración de Cantidad de Neuronas para el Modelo Final de Regresión Logística.....	102
	Anexo 18: Reglas del Modelo Final de Árboles de Decisión.	103
	Anexo 19: Output del Modelo Final de Regresión Logística.....	104
	Anexo 20: Análisis para la Obtención del KS de los Modelos Finales.....	105
	Anexo 21: Mapa de Predicción de Default Por Región.	106
	Anexo 22: Mapa de Predicción de Default en la RM.	107

ÍNDICE DE TABLAS

Tabla 1: Selección de tablas del DW del SII	28
Tabla 2: Categoría de las variables independientes	32
Tabla 3: Resumen de tratamiento de <i>outliers</i>	34
Tabla 4: Decisión de eliminación de variables de acuerdo a la correlación. .	35
Tabla 5: Partición de la muestra.	36
Tabla 6: Comparación de diferentes modelos de riesgo de crédito.	38
Tabla 7: Comparación de algoritmos de árboles de decisión.	38
Tabla 8: Comparación de resultados prueba ex-ante.	40
Tabla 9: Comparación de resultados modelo general.....	42
Tabla 10 : Rango de los parámetros de modificación.	43
Tabla 11 : Rangos de los parámetros de modificación.	45
Tabla 12 : Rango de modificación de los parámetros.	46
Tabla 13: Resumen de los resultados de la etapa iteración.	47
Tabla 14: Resultados de los modelos finales.....	53
Tabla 15 : Tasa de error para distintos cortes de probabilidad.	57

ÍNDICE DE ILUSTRACIONES

Figura 1: Tasa de evasión del IVA.	2
Figura 2: Modelo de gestión de servicios del SII.....	8
Figura 3: Clasificación de impuestos en Chile.	9
Figura 4: Distribución de ingresos tributarios por tipo de impuesto.....	10
Figura 5: Descomposición de la brecha tributaria.	11
Figura 6: Aporte de cada segmento empresarial a las ventas del país.....	13
Figura 7: Porcentaje de empresas según clasificación acuñada por el SII. ...	14
Figura 8: Diagrama de etapas del KDD	16
Figura 9: Evolución de casos "malos" en el tiempo.	18
Figura 10: Diagrama de modelos de data mining.	23
Figura 11: Diagrama de redes neuronales.....	26
Figura 12: Tipos de registro.	29
Figura 13: Porcentaje de atrasos para contribuyentes multi-declarantes.	30
Figura 14: Porcentaje de atrasos para contribuyentes uni-declarantes.	30
Figura 15: Porcentaje de default en cada período tributario.	31
Figura 16: Porcentaje de default en cada período tributario.	31
Figura 17: Análisis de default para la eliminación de <i>outliers</i>	34
Figura 18: Importancia relativa de variables para árboles de decisión.	44
Figura 19: Importancia relativa de las variables para redes neuronales.	45
Figura 20: Importancia relativa de las variables para árboles de decisión....	48
Figura 21: Importancia de variables en modelo final de redes neuronales... 50	
Figura 22: Curva ROC para el modelo final de redes neuronales.....	50
Figura 23: Evolución de indicadores según cantidad de variables.....	51
Figura 24: Curva ROC para el modelo final de regresión logística.	52
Figura 25: Indicadores del modelo de árboles de decisión	54
Figura 26: Análisis de robustez para el modelo de árboles de decisión.	54
Figura 27: Indicadores del modelo de regresión logística	56
Figura 28: Análisis de robustez para el modelo de regresión logística.	56
Figura 29: Robustez del KS y AUC para el modelo de regresión logística ...	57