

Tabla de contenido

1.	Introducción.....	1
1.1.	Antecedentes Generales.....	1
1.1.1.	Servicio de Impuestos Internos.....	1
1.1.2.	Clasificación de Impuestos.....	2
1.1.3.	Importancia del IVA.....	2
1.1.4.	Clasificación de las empresas según SII.....	4
1.1.5.	Giros de impuestos.....	5
1.2.	Definición y Justificación del proyecto.....	5
1.3.	Objetivos.....	7
1.3.1.	Objetivos Generales.....	7
1.3.2.	Objetivos Específicos.....	7
1.4.	Alcances.....	8
1.5.	Resumen metodología.....	9
2.	Marco Conceptual.....	10
2.1.	Modelo de Gestión del Cumplimiento Tributario.....	10
2.1.1.	Definición del modelo.....	10
2.1.2.	Probabilidad y consecuencias de riesgo tributario.....	11
2.2.	Flujos de los giros seleccionados.....	11
2.2.1.	Multa F29.....	12
2.2.2.	Impuestos F29.....	12
2.2.3.	Rectificatoria F29.....	13
2.2.4.	Auditoría IVA.....	13
2.3.	Tratamientos de Cobranza.....	16
2.4.	Entendimiento del Problema y Generación de Variables.....	17
2.4.1.	Estudio del problema dentro del Servicio.....	17
2.4.2.	Credit Scoring.....	17
2.4.3.	Estudio de otros Servicios.....	18
2.4.4.	Grupos de variables.....	19
3.	Metodología.....	19
3.1.	Knowledge Discovery in Databases.....	19
3.2.	Tipos de Modelos.....	20
3.2.1.	Árboles de Decisión.....	21
3.2.2.	Regresión Logística.....	24
3.3.	Validación de Modelos.....	25
3.3.1.	Matriz de Confusión.....	25

3.3.2.	Accuracy	26
3.3.3.	Métricas sobre los valores predichos	26
3.3.4.	Métricas sobre los valores reales	27
3.3.5.	Curva ROC y AUC	27
3.3.6.	Curva de Ganancia de Información	28
4.	Desarrollo del trabajo en la Base de Datos.....	29
4.1.	Selección y pre procesamiento de datos	30
4.1.1.	Descripción de la base de datos y Definición de No Pago.....	30
4.1.2.	Selección de variables	32
4.1.3.	Pre procesamiento	33
4.2.	Transformación.....	34
4.3.	Análisis de variables	37
4.3.1.	Análisis Univariados	37
4.3.2.	Análisis de Correlación	41
4.4.	Detalle del Problema Final	41
5.	Resultados Modelos.....	43
5.1.	Elección modelo de tipo Árbol de Decisión	43
5.2.	Resultados Árbol CHAID.....	45
5.3.	Elección modelo de tipo Regresión Logística	47
5.4.	Resultados Regresión Logística.....	48
5.5.	Elección modelo de tipo Random Forest.....	49
5.6.	Resultados Random Forest	51
5.7.	Comparación modelos finales.....	53
6.	Resultados específicos.....	55
6.1.	Explicación de Modelo CHAID	55
6.1.1.	Nodo 1. Giros por Auditoría IVA	55
6.1.2.	Nodo 2. Giros por Impuesto F29	57
6.1.3.	Nodo 3. Giros por Multa F29.....	59
6.1.4.	Respecto a los supuestos del capítulo 4.4	61
6.2.	Caracterización de giros.....	62
6.3.	Resultados según segmentos de interés	63
6.3.1.	Segmentos SII	63
6.3.2.	Segmentos TGR	64
6.3.3.	Comparación Random Forest y CHAID.....	66
7.	Propuestas.....	67
7.1.	Propuestas para el Servicio de Impuestos Internos.....	67

7.1.1.	Propuestas sobre variables	67
7.1.2.	Propuestas sobre las predicciones de no pago	67
7.1.3.	Propuestas sobre otros estudios.....	69
7.2.	Propuestas para la TGR	69
7.3.	Análisis económico.....	70
8.	Conclusiones.....	72
8.1.	Conclusiones generales respecto al modelo generado	73
8.2.	Conclusiones específicas	73
8.2.1.	Respecto a la identificación de variables y patrones.....	73
8.2.2.	Respecto a los segmentos de contribuyentes	74
8.2.3.	Respecto a la predicción de nuevos giros	74
8.2.4.	Respecto a las propuestas para el SII y TGR.....	75
8.2.5.	Trabajos futuros	75
9.	Bibliografía.....	76
10.	Anexos.....	79
10.1.	Formulario 29	79
10.2.	Ejemplo de Giro	81
10.3.	Listado de Variables.....	82
10.4.	Categorías de variables.....	84
10.5.	Capítulo 4. Otros análisis Univariados.....	85
10.6.	Regresión logística	87
10.7.	Variación OOB.....	88
10.8.	Resultados específicos.....	89
10.9.	Código modelo Random Forest.....	91
10.10.	Modelo exploratorio Segmento Mediano y Gran Deudor	92
10.11.	Particiones modelo CHAID	93

Índice de tablas

Tabla 1: Origen Giros de Impuestos	5
Tabla 2: Matriz de Confusión	25
Tabla 3: Ejemplo simplificado tabla giros	31
Tabla 4: Listado parcial de Variables	35
Tabla 5: Variables por Grupo	37
Tabla 6 : Montos y giros por concepto	38
Tabla 7: Variables correlacionadas	41
Tabla 8 : Segmentos de contribuyentes	42
Tabla 9: Desempeño de árboles SPSS	44
Tabla 10: Comparación árboles SPSS	44
Tabla 11: Matriz de Confusión CHAID	46
Tabla 12: Desempeño Regresión Logística	47
Tabla 13: Matriz de Confusión Regresión Logística	48
Tabla 14: Análisis de Sensibilidad Random Forest	50
Tabla 15: Matriz de Confusión modelo Random Forest	52
Tabla 16: Desempeño Modelos	53
Tabla 17: Comparación modelos	54
Tabla 18: Impuesto F29. Mayores probabilidades de no pago	58
Tabla 19: Impuesto F29. Menores probabilidades de no pago	59
Tabla 20: Multa F29. Mayores probabilidades de no pago	60
Tabla 21: Caracterización Giros	62
Tabla 22: Comparación Random Forest y CHAID SII	66
Tabla 23: Comparación Random Forest y CHAID TGR	66
Tabla 24: Beneficios y costos del modelo	71

Índice de Ilustraciones

Ilustración 1: Clasificación de los impuestos en Chile.....	2
Ilustración 2: Aporte por Impuesto al Ingreso Tributario Neto.....	3
Ilustración 3: Relación Cargo y Pago a través del tiempo.....	6
Ilustración 4: Monto de la deuda por tipo de giro.....	8
Ilustración 5: Matriz de riesgo Modelo Gestión del Cumplimiento Tributario.....	11
Ilustración 6: Flujo de giros del tipo Multa F29 e Impuesto F29.....	12
Ilustración 7: Flujo de giro del tipo Auditoria IVA - Notificación.....	14
Ilustración 8: Flujo de giro del tipo Auditoria IVA – Citación.....	15
Ilustración 9 : Ejemplo cuadro total a pagar de un giro.....	15
Ilustración 10: Tiempos de cobranza.....	16
Ilustración 11 : Modelo de cumplimiento de la ATO.....	18
Ilustración 12: Esquema proceso KDD.....	19
Ilustración 13: Curva ROC.....	27
Ilustración 14: Clasificación modelos según AUC.....	28
Ilustración 15: Curva Ganancia de Información.....	29
Ilustración 16: Data Inicial.....	33
Ilustración 17: No pago según tipo de giro.....	38
Ilustración 18: No pago según Cargo.....	39
Ilustración 19 : No pago según Ventas.....	40
Ilustración 20: No pago según Cargo/C91.....	40
Ilustración 21: Data Final.....	42
Ilustración 22: Importancia de las variables Modelo CHAID.....	45
Ilustración 23: Curva ROC CHAID.....	46
Ilustración 24: Curva de Ganancia CHAID.....	47
Ilustración 25: Curva ROC Regresión Logística.....	49
Ilustración 26: Curva de Ganancia Regresión Logística.....	49
Ilustración 27: Importancia de las variables Random Forest.....	51
Ilustración 28: Curva ROC Random Forest.....	52
Ilustración 29: Curva de Ganancia Random Forest.....	53
Ilustración 30 : Árbol CHAID, Primera ramificación.....	55
Ilustración 31: Nodo 1. Auditoria IVA, Segundo nivel de profundidad.....	56
Ilustración 32: Auditoria IVA, Categorías de Cargo 8,9 y 10.....	57
Ilustración 33: Distribución de las probabilidades de no pago según Cargo.....	61
Ilustración 34: Métricas de desempeño según Segmentos SII.....	64
Ilustración 35: Métricas de desempeño según Segmentos TGR.....	65
Ilustración 36: Utilidad del modelo CHAID.....	71