



**UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL**

**REDISEÑO DE PROCESOS DE FISCALIZACIÓN PROACTIVA EN LA
DIRECCIÓN DEL TRABAJO DE CHILE**

**PROYECTO DE GRADO PARA OPTAR AL GRADO DE MAGÍSTER EN
INGENIERÍA DE NEGOCIOS CON TECNOLOGÍAS DE INFORMACIÓN
MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL**

ALEJANDRO ALLENDE ACED

PROFESOR GUÍA:

EZEQUIEL MUÑOZ KRSULOVIC

PROFESOR CO-GUÍA:

LUIS ABURTO LAFOURCADE

MIEMBROS DE LA COMISIÓN:

EDUARDO CONTRERAS VILLABLANCA

CLAUDIO COLLAO BAHAMONDES

SANTIAGO DE CHILE

2018

RESUMEN EJECUTIVO

La Dirección del Trabajo es la organización que se encarga de velar por el cumplimiento de la legislación laboral a través de sus productos y servicios estratégicos, donde la fiscalización es una de sus principales herramientas de detección de infraccionalidad.

El proyecto de tesis se enfoca en el rediseño de los procesos de fiscalización proactiva en base a la Metodología de Ingeniería de Negocios, donde se ha presentado una baja tasa de detección de infraccionalidad (Eficacia) en los programas de fiscalización, la cual en los últimos 6 años ha tenido un promedio del 27% y nunca superando el 35%. Este resultado se considera bajo debido a estudios que estiman una tasa de infraccionalidad natural (teórica) del 37% (Kanbur, Ronconi, & Wedenoja, 2013), por lo cual, la fiscalización que está enfocada en encontrar infracción está bajo la tasa natural o base. Debido a la restricción de recursos de la organización, la cual se ha mantenido lo últimos 5 años con un número de fiscalizadores promedio de 450 a lo largo del país, y una demanda creciente cercana al 5% desde el 2010, la organización se ha visto en la necesidad de optimizar el manejo de sus recursos y enfocar su acción inspectiva a aquellas empresas con mayor probabilidad de infracción.

Para esto, el proyecto posee como objetivo aumentar la eficacia de la fiscalización proactiva al doble hasta el 2018, a través del rediseño del proceso de la fiscalización y el desarrollo de modelos predictivos con la información de la organización.

La metodología general del proyecto de rediseño es la Ingeniería de Negocios, mientras que para la creación de los modelos predictivos se utiliza la metodología CRISP-DM.

Los modelos predictivos utilizados en el proyecto fueron Árboles de decisión (CART, Conditional Inference Trees y Random Forest), Regresión Logística y Redes Neuronales, donde el que presentó los mejores resultados de predicción fue el Random Forest, con un *Accuracy* del 87,41% y un *Fscore* del 84,43%. Con tal modelo se realizó una prueba piloto a 560 empresas a lo largo de todo Chile, donde se obtuvo una asertividad del **42,24%**, donde en regiones se llegó a tener por sobre el 50% de acierto.

Se presenta una arquitectura basada en el levantamiento de una máquina virtual la cual alberga el desarrollo generado para los pasos de la metodología CRISP-DM, donde se muestran los casos de uso, diagrama de arquitectura y diagrama de despliegue.

Por último, se realiza una evaluación económica al proyecto, donde se plantea un indicador de costo efectivo por fiscalización (CEF), el cual se define que si el proyecto presenta un menor CEF es conveniente su realización. El escenario sin proyecto presentó un CEF de CLP\$133.650, mientras que el con Proyecto mostró un CEF de CLP\$104.810 a una tasa social de descuento del 6%.

*A mi familia, Lorena y Catalina,
gracias por todo el apoyo y amor incondicional
en los buenos y malos momentos,
por sobre todo en los malos,
los cuales han sido muchos.*

AGRADECIMIENTOS

En primer lugar, agradecer a mi madre Lorena y mi hermana Catalina quienes han sido el pilar fundamental en mi vida, me han acompañado y apoyado durante todo el camino, tanto universitario como de mi vida. De ustedes he aprendido todo lo que soy, las amo infinitamente.

A mis tíos, Paula y Roberto quienes me han acompañado en los peores momentos, son un gran apoyo en mi vida, siempre se los agradeceré.

Agradecer a mi polola Camila, quien me ha amado, aguantado y acompañado durante todo el transcurso de mi tesis. Gracias por tu amor y compañía, han sido los 2 mejores años, te amo.

También agradecer a las grandes amistades que han nacido en el transcurso de la universidad desde el día uno hasta los últimos, agradecer a Camilo Villar, Bastián Muñoz, Chih-yu Chuang, Diego García, Manuel Hernández, Pablo Moscoso quienes me han acompañado durante todo el proceso tanto de la universidad como en la vida, todos los Catanes, pokers, subidas al *ride*, y tantos panoramas. Gracias, también, a los cabros del *age* a quienes dejé botados por mi tesis, Feña, Ruso, Armando, Kevin, Sapu y Felipiño, tantos buenos momentos fuera y dentro de las aulas, perdón por lo ingrato.

Muchas Gracias a todo el MBE, al Profesor Ezequiel Muñoz, con quien tuve el agrado de conocer y trabajar en conjunto, muchas gracias por sus grandes consejos tanto profesionales como de vida, lo admiro mucho. Al Profesor Luis Aburto, quien me ayudó a direccionar y abordar el proyecto con todos sus conocimientos y grandes consejos. Muchas Gracias también a Luciano Villarroel, quien me ayudó a tomar decisiones en los momentos más críticos, gracias por todo. Obviamente, sin olvidar a Ana María y Laura, quienes me han apoyado desde el primer momento que llegué a preguntar de qué se trataba este viaje llamado MBE, de corazón, muchas gracias.

Agradecer a las grandes personas de la Dirección del Trabajo, quienes me han apoyado y ayudado mucho en esta ardua tarea, a Natalia y Cristina quienes me enseñaron mucho sobre la organización y me apoyaron durante todo este trabajo. A Claudio quien me ayudó desde un inicio a direccionar todo el trabajo que de a poco ha ido creciendo, muchas gracias por la tutoría y la gran amistad. También agradecer a Alejandro Baeza y Andrés Arenas, quienes en el corto tiempo que he llegado a conocer, han sido grandes amigos y hemos pasado muy buenos momentos, gracias por todos los consejos y buenos ratos.

Por último, agradecer a todos los actores no humanos que me ayudaron a pasar por este proceso, la música, las pesas, las series, entre otros hobbies, fueron necesarios para lidiar con todo lo que se vivió.

TABLA DE CONTENIDO

II RESUMEN EJECUTIVO

III DEDICATORIA

IV AGRADECIMIENTOS

CAPÍTULO 1: INTRODUCCIÓN Y CONTEXTO.....	1
1.1 DESCRIPCIÓN GENERAL DE LA EMPRESA.....	2
1.1.1 <i>Visión</i>	2
1.1.2 <i>Misión</i>	2
1.1.3 <i>Productos y Servicios Principales</i>	2
1.1.4 <i>Clientes</i>	2
1.1.5 <i>Organigrama</i>	3
1.2 PROBLEMA U OPORTUNIDAD IDENTIFICADA.....	3
1.3 OBJETIVOS Y RESULTADOS ESPERADOS DEL PROYECTO.....	6
1.3.1 <i>Objetivo General</i>	6
1.3.2 <i>Objetivos Específicos</i>	6
1.3.3 <i>Resultados Esperados</i>	7
1.4 ALCANCE.....	7
1.4.1 <i>Dentro del Alcance</i>	7
1.4.2 <i>Fuera del Alcance</i>	7
1.5 RIESGOS POTENCIALES.....	8
CAPÍTULO 2: MARCO TEÓRICO.....	10
2.1 METODOLOGÍA DE INGENIERÍA DE NEGOCIOS.....	10
2.2 METODOLOGÍA CRISP-DM.....	12
2.3 MINERÍA DE DATOS.....	13
2.3.1 <i>Modelos de Minería de Datos</i>	14
CAPÍTULO 3: PLANTEAMIENTO ESTRATÉGICO Y MODELO DE NEGOCIOS	20
3.1 POSICIONAMIENTO ESTRATÉGICO.....	20
3.2 BALANCED SCORECARD.....	20
3.2.1 <i>Perspectiva Financiera</i>	21
3.2.2 <i>Perspectiva Usuarios</i>	21
3.2.3 <i>Perspectiva Procesos Internos</i>	21

3.2.4	<i>Perspectiva Aprendizaje y Crecimiento</i>	21
3.3	MODELO DE NEGOCIOS	21
3.3.1	<i>Propuesta de Valor</i>	22
3.3.2	<i>Interfaz del Cliente</i>	22
3.3.3	<i>Infraestructura de Gestión</i>	22
3.3.4	<i>Flujos de Ingresos y Estructura de Costos</i>	23
	CAPÍTULO 4: ANÁLISIS SITUACIÓN ACTUAL	24
4.1	ARQUITECTURA DE PROCESOS	24
4.2	MODELAMIENTO DETALLADO DE PROCESOS	25
4.2.1	<i>Modelamiento IDEF0</i>	25
4.3	DIAGNÓSTICO DE LA SITUACIÓN ACTUAL.....	28
4.4	CUANTIFICACIÓN DEL PROBLEMA U OPORTUNIDAD.....	29
	CAPÍTULO 5: PROPUESTA DE DISEÑO DE PROCESOS	31
5.1	DIRECCIONES DE CAMBIO Y ALCANCE	31
5.1.1	<i>Estructura de la Empresa y Mercado</i>	31
5.1.2	<i>Anticipación</i>	32
5.1.3	<i>Coordinación</i>	32
5.1.4	<i>Prácticas de trabajo</i>	33
5.1.5	<i>Integración de Procesos Conexos</i>	34
5.1.6	<i>Mantenimiento Consolidada del Estado</i>	34
5.2	DISEÑO DETALLADO DE PROCESOS.....	36
5.2.1	<i>Diseño en IDEF0</i>	36
5.3	LÓGICA DE NEGOCIOS	39
5.4	PRUEBA DE LA LÓGICA DE NEGOCIOS.....	40
5.4.1	<i>Entendimiento del Negocio</i>	40
5.4.2	<i>Entendimiento de los Datos</i>	41
5.4.3	<i>Preparación de los Datos</i>	50
5.4.4	<i>Modelamiento</i>	53
5.4.5	<i>Evaluación</i>	63
5.4.6	<i>Despliegue</i>	64
	CAPÍTULO 6: PROPUESTA DE APOYO TECNOLÓGICO	65
6.1	ESPECIFICACIÓN DE REQUERIMIENTOS.....	65
6.1.1	<i>Requerimientos Funcionales</i>	65

6.1.2	<i>Requerimientos No Funcionales</i>	65
6.2	ARQUITECTURA TECNOLÓGICA.....	66
6.3	DISEÑO DEL SISTEMA VIRTUAL.....	66
6.3.1	<i>Casos de Uso</i>	66
6.3.2	<i>Diagrama de Arquitectura del Sistema</i>	69
6.3.3	<i>Diagrama de Despliegue</i>	70
	CAPÍTULO 7: GESTIÓN DEL CAMBIO	72
7.1	CONTEXTO DE LA EMPRESA.....	72
7.2	ANÁLISIS DE LOS PRINCIPIOS DE DISEÑO	73
7.2.1	<i>Liderazgo y Gestión del proyecto de cambio</i>	73
7.2.2	<i>Estrategia y sentido del proceso de cambio</i>	74
7.2.3	<i>Cambio y Conservación</i>	74
7.2.4	<i>Organización y Estructura del proyecto de cambio</i>	75
7.2.5	<i>Gestión Emocional</i>	76
7.2.6	<i>Comunicaciones</i>	76
7.2.7	<i>Desarrollo de las Habilidades</i>	76
7.2.8	<i>Gestión del Poder</i>	77
7.2.9	<i>Monitoreo y Evaluación del proceso</i>	77
7.2.10	<i>Inicio, hitos, ritos y cierre</i>	77
7.3	CARACTERIZACIÓN DEL CAMBIO.....	77
7.4	FACTORES CRÍTICOS DE ÉXITO	78
7.5	PLAN DE GESTIÓN DEL CAMBIO.....	79
	CAPÍTULO 8: EVALUACIÓN DEL PROYECTO	80
8.1	PLAN PILOTO	80
8.1.1	<i>Definición del Plan Piloto</i>	80
8.1.2	<i>Resultados Obtenidos</i>	81
8.2	DEFINICIÓN DE VARIABLES Y CRITERIOS DE EVALUACIÓN	87
8.2.1	<i>Tipo de Evaluación</i>	87
8.2.2	<i>Horizonte de Evaluación</i>	88
8.2.3	<i>Tasa de Descuento</i>	88
8.2.4	<i>Escenarios de Evaluación</i>	88
8.2.5	<i>Costos</i>	89
8.2.6	<i>Tasas Contingentes</i>	89

8.2.7	<i>Inversión</i>	90
8.2.8	<i>Consideración Social de Beneficios</i>	91
8.3	FLUJO DE CAJA.....	91
8.4	ANÁLISIS DE SENSIBILIDAD.....	93
	CAPÍTULO 9: CONCLUSIONES	94
	CAPÍTULO 10: BIBLIOGRAFÍA	98
	CAPÍTULO 11: ANEXOS	100
11.1	DIAGRAMAS ENTIDAD RELACIÓN	100

ÍNDICE DE FIGURAS

Figura 1:	Organigrama Dirección del Trabajo.	3
Figura 2:	Gráfico Infraccionalidad detectada por año.	4
Figura 3:	Gráfico Infraccionalidad detectada Reactiva por año.	5
Figura 4:	Gráfico Infraccionalidad detectada Proactiva por año.	6
Figura 5:	Modelo Ontológico.....	10
Figura 6:	Metodología Ingeniería de Negocios.....	11
Figura 7:	Fases del Modelo de referencia CRISP-DM.....	12
Figura 8:	Tareas de la Minería de Datos.....	14
Figura 9:	Ejemplo Árbol de Decisión.	16
Figura 10:	Esquema de una red neuronal.....	18
Figura 11:	Cálculo del output de una neurona de capa oculta o salida.	18
Figura 12:	Arquitectura de Procesos, Dirección del Trabajo.	24
Figura 13:	Macro 1 Fiscalización de Empresas, Dirección del Trabajo.	25
Figura 14:	Administración y Análisis de Fiscalizaciones, Dirección del Trabajo.	26
Figura 15:	Análisis de Información del Mercado Laboral, Dirección del Trabajo.	27
Figura 16:	Análisis del Comportamiento Laboral, Dirección del Trabajo.	27
Figura 17:	Planificación Estratégica de Programas de Fiscalización, Dirección del Trabajo.	28
Figura 18:	Análisis de Información del Mercado Laboral, Rediseño.	36
Figura 19:	Análisis del Comportamiento Laboral, Rediseño.....	37
Figura 20:	Desarrollo y Actualización Modelos Predictivos, Rediseño.	38
Figura 21:	Planificación Estratégica de Programas de Fiscalización, Rediseño.....	38
Figura 22:	Lógica de Negocios Propuesta.	39

Figura 23: Histogramas de la variable Número de Trabajadores, Dirección del Trabajo. .	45
Figura 24: Histogramas de la variable Número de Trabajadores, SII.	45
Figura 25: Gráfico de frecuencia e Infraccionalidad según Actividad Económica.	46
Figura 26: Gráfico de frecuencia e Infraccionalidad según Tipo de Solicitante.	47
Figura 27: Gráfico de frecuencia e Infraccionalidad según Región del País.	48
Figura 28: Indicadores de Predicción versus Cortes de selección, CART.	55
Figura 29: Indicadores de Predicción versus Cortes de selección, Conditional Inference Trees.	56
Figura 30: Indicadores de Predicción versus Cortes de selección, Random Forest.	57
Figura 31: Indicadores de Predicción versus Cortes de selección, Regresión Logística. .	58
Figura 32: Indicadores de Predicción versus Cortes de selección, Red Neuronal.	59
Figura 33: Árbol de decisión, CART con profundidad máxima de 7 nodos.	61
Figura 34: Importancia de Variables explicativas, Random Forest.	62
Figura 35: Casos de Uso, Arquitectura para el Rediseño.	67
Figura 36: Caso de Uso “Obtener Datos”, Arquitectura Propuesta.	67
Figura 37: Caso de Uso “Limpiar y Transformar Datos”, Arquitectura Propuesta.	68
Figura 38: Caso de Uso “Procesar Modelos Analíticos”, Arquitectura Propuesta.	68
Figura 39: Caso de Uso “Generar Data Set para Fiscalización Proactiva”, Arquitectura Propuesta.	69
Figura 40: Caso de Uso “Pre-procesar según Modelo”, Arquitectura Propuesta.	69
Figura 41: Diagrama de Arquitectura del Sistema, Arquitectura Propuesta.	70
Figura 42: Diagrama de Despliegue, Arquitectura para el Rediseño.	71
Figura 43: Gráfico de Infraccionalidad del Piloto según Región.	82
Figura 44: Comparación Infraccionalidad Histórica vs. Piloto.	83
Figura 45: Gráfico de Infraccionalidad del mes de noviembre según años.	84
Figura 46: Gráfico de Infraccionalidad Proactiva según mes del año.	85
Figura 47: Gráfico de Infraccionalidad del Piloto según Rama de Actividad Económica. .	86
Figura 48: Diagrama E-R de Informes de Fiscalización	100
Figura 49: Diagrama E-R Ingreso de Fiscalización.	101
Figura 50: Diagrama E-R Multas.	102
Figura 51: Coeficientes Regresión Logística.	103

ÍNDICE DE TABLAS

Tabla 1: Riesgos Potenciales del Proyecto de Tesis.	8
--	---

Tabla 2: Valorización Fiscalizaciones Efectivas y No Efectivas.....	29
Tabla 3: Variable Estructura de la Empresa y Mercado	31
Tabla 4: Variable Anticipación.....	32
Tabla 5: Variable Coordinación.....	32
Tabla 6: Variable Prácticas de trabajo	33
Tabla 7: Variable Integración de Procesos Conexos.....	34
Tabla 8: Variable Mantenimiento Consolidada del Estado	34
Tabla 9: Descripción de Datos de la Dirección del Trabajo.	43
Tabla 10: Estadísticas de variables numéricas, datos Dirección del Trabajo.	44
Tabla 11: Porcentaje de Valores Nulos en la base de datos.	48
Tabla 12: Indicadores para corte seleccionado, CART.	55
Tabla 13: Indicadores para corte seleccionado, CIT.	56
Tabla 14: Indicadores para corte seleccionado, Ranfom Forest.....	57
Tabla 15: Indicadores para corte seleccionado, Logit.	59
Tabla 16: Indicadores para corte seleccionado, Red Neuronal.	60
Tabla 17: Cuadro resumen de indicadores según modelo.	60
Tabla 18: Requerimientos para la arquitectura tecnológica del rediseño.	65
Tabla 19: Equipo de Liderazgo del Proyecto de tesis.	73
Tabla 20: Distribución de fiscalizaciones, plan piloto 2017.....	80
Tabla 21: Código de Actividades Económicas Fiscalizadas	85
Tabla 22: Estimación de Fiscalizaciones.	89
Tabla 23: Tasas contingentes para la evaluación del proyecto.	90
Tabla 24: Montos de Inversión.....	90
Tabla 25: Flujo de caja, escenario sin proyecto.	91
Tabla 26: Indicadores, escenario sin proyecto.	91
Tabla 27: Flujo de Caja, escenario con proyecto.	92
Tabla 28: Indicadores, escenario con proyecto.....	92
Tabla 29: Simulación de casos para tasas contingentes.....	93

CAPÍTULO 1: INTRODUCCIÓN Y CONTEXTO

El trabajo es el principal instrumento de un país para crecer y generar una economía estable para sus ciudadanos, además es fuente de auto-realización y permite cultivar una serie de habilidades importantes como el trabajo en equipo, además de permitir mejorar la calidad de vida. Si bien el trabajo posee estas características positivas, existen factores que hacen que se torne en una fuente de problemas y estrés, además de la existencia de irregularidades que hacen que los trabajadores sean vulnerados en sus derechos, generando en esto un problema no solo a nivel de ciudadanos sino un problema a nivel país. Es aquí donde el estado, a través de la Dirección del Trabajo, se encarga de evitar que este tipo de conflictos ocurra por medio de fiscalizaciones y también se encarga de establecer soluciones a conflictos que sucedieron en el ambiente laboral y atender denuncias y reclamos de trabajadores que pudieron ser vulnerados en sus derechos.

La Dirección del Trabajo (DT) es la herramienta directa al trabajo de campo y se encarga de fiscalizar y atender a los usuarios dentro del mix de productos y servicios que esta ofrece. Está en directa vigilancia de la Presidencia de la República a través del Ministerio del Trabajo y Prevención Social y posee personalidad jurídica y patrimonio propio, siendo una entidad autónoma con presupuesto propio.

Es dentro del marco estratégico, táctico y operativo de la DT donde se encuentran los primeros problemas que afectan directamente a los usuarios de la entidad (entiéndase por usuarios tanto a trabajadores, empleadores u otros actores laborales) afectando al cumplimiento de su misión organizacional, donde se observan altos tiempos de espera en atención presencial, bajas tasas de detección de infraccionalidad de empresas fiscalizadas, entre otros.

Cabe destacar que actualmente el Gobierno de Chile presenta un Proyecto General de Modernización del Estado el cual se llama “Programa de Mejora de la Gestión Pública y de los Servicios al Ciudadano” el cual está en manos del Ministerio de Hacienda, y es la DT una de las entidades a mejorar dentro del programa señalado, y está direccionado a la mejora de los Sistemas de Información que posee. De este programa nace el Proyecto particular “Modernización de los Procesos de Fiscalización y Gestión de los Servicios a los Usuarios de la Dirección del Trabajo” a partir de diversas consultorías realizadas a la entidad, con los dos ejes principales detallados en el nombre del proyecto.

Es aquí donde el presente documento está inmerso y se enfoca específicamente en un subtema del eje Fiscalización, el cual es la Fiscalización Proactiva o Programada. Esta fiscalización es toda aquella que no es producto de una denuncia, sino que es netamente preventiva y busca, de manera estratégica, encontrar empresas infractoras con el fin de anticiparse a posibles infracciones. Es importante, además, destacar que el otro subtema, Fiscalización Reactiva o bajo denuncia, la cual está basado en la reacción ante una denuncia realizada

por un usuario, no está exento de problemas, pero estos están fuera del alcance del proyecto realizado. Sin embargo, existe la conjetura que una Fiscalización Proactiva efectiva debiese disminuir la cantidad de Fiscalizaciones Reactivas.

1.1 Descripción General de la Empresa

La Dirección del Trabajo es un servicio público el cual posee personalidad jurídica y patrimonio propio. Está bajo la vigilancia del presidente de la República a través del Ministerio del Trabajo y Prevención Social.

1.1.1 Visión

La Visión de la Dirección del Trabajo entregada en su sitio web es: “Ser un Servicio Público moderno, respetado y confiable que garantice y promueva relaciones laborales justas, disminuyendo las desigualdades y así contribuir a la justicia social.” (Dirección del Trabajo, 2017).

1.1.2 Misión

La Misión de la organización es: “Velar por el cumplimiento de la legislación laboral, previsional y de seguridad y salud en el trabajo y promover el ejercicio pleno de la libertad sindical, fomentando el diálogo social, fortaleciendo sistemas de prevención y solución alternativa de conflictos, que permita relaciones más justas y equitativas entre trabajadores y empleadores, contribuyendo a un sistema democrático de relaciones laborales.” (Dirección del Trabajo, 2017).

1.1.3 Productos y Servicios Principales

Los principales Servicios/Productos que ofrece la Dirección del Trabajo son:

- Fiscalización
- Sistemas de Prevención y Solución alternativa de Conflictos
- Promoción de Libertad Sindical y Asistencia Técnica
- Atención de Usuarios
- Dictámenes

1.1.4 Clientes

Los clientes de la DT pueden ser: Trabajadores, empleadores, dirigentes sindicales, autoridades de gobierno, miembros de la judicatura, organizaciones de trabajadores, representantes políticos, organizaciones de empleadores, organismos internacionales, universidades y ONG's.

1.1.5 Organigrama

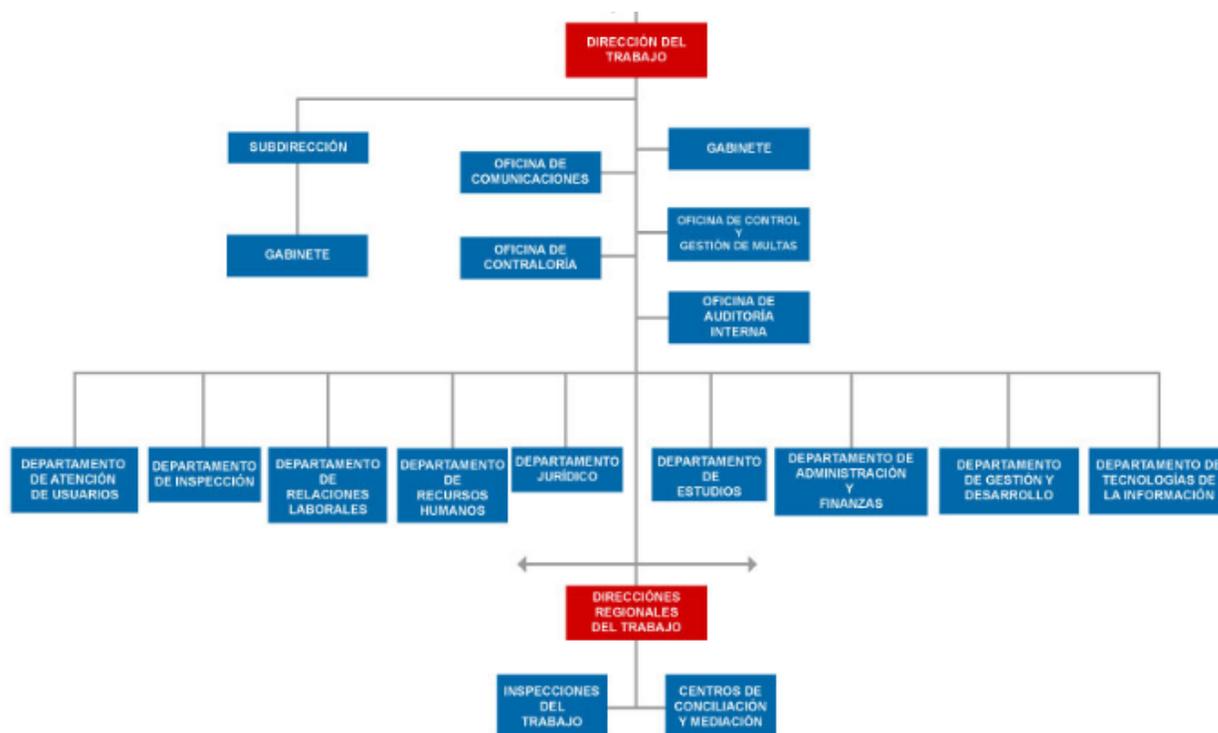


Figura 1: Organigrama Dirección del Trabajo.

Fuente: www.dt.gob.cl

1.2 Problema u Oportunidad Identificada

La DT, a modo de resumen, posee dos grandes ejes, los cuales son Fiscalización y Atención de Usuarios. Son estos procesos los cuales la organización realiza para entregar valor a sus usuarios. El Proyecto de Tesis está enfocado en el eje de Fiscalización, más específicamente, en la **Fiscalización Proactiva**.

La fiscalización se divide en dos tipos:

- Fiscalización Reactiva: La cual responde a la denuncia realizada por un usuario y la DT procede a fiscalizar a la empresa en su/s establecimiento/s.
- Fiscalización Proactiva: Esta fiscalización es netamente preventiva, y busca de manera estratégica, las empresas con mayor riesgo de ser infractoras, para luego fiscalizarla en su/s establecimiento/s.

Se observan diversos problemas tanto para la fiscalización reactiva como proactiva y se analizarán de forma más profunda. Con el objetivo de contextualizar la cantidad de fiscalizaciones que realiza la DT, se observan los datos históricos obtenidos del sistema informático de la DT (DT PLUS y SIRELA). En la Figura 2 se observa la cantidad de fiscalizaciones y su respectiva detección

de infracción a lo largo de los años, con un promedio del 40%, la cual se definirá como Eficacia, cuya ecuación es la siguiente:

$$Eficacia = \frac{\# \text{ de fiscalizaciones de empresas infractoras}}{\text{Total de fiscalizaciones realizadas}}$$

por lo cual se puede observar que el año con mayor eficacia fue el 2011, del mismo modo, el año 2007 se presenta como el que tuvo la menor.

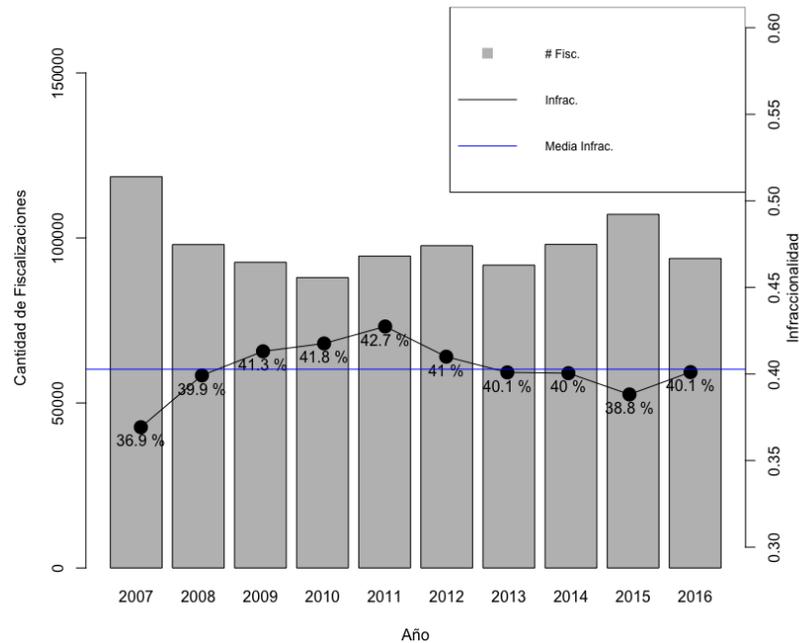


Figura 2: Gráfico Infraccionalidad detectada por año.

Fuente: Elaboración Propia.

La Figura 3 muestra el mismo gráfico de infraccionalidad, pero para las fiscalizaciones reactivas, donde el comportamiento es más estable a lo largo de los años, con una eficacia media del 46%. Este efecto se deduce a que la infraccionalidad se mantiene relativamente constante a lo largo de los años, con una cantidad de denuncias que se mantiene en el tiempo. Dado que la fiscalización reactiva nace de una denuncia, un 40% de detección de infraccionalidad por estas puede explicarse porque el denunciante no tenía una causa válida ni argumento legal, o si la tenía, el fiscalizador no logra encontrar evidencia de esto. La principal causa de esto puede ser el tiempo de demora que existe entre que se realiza la denuncia hasta que el fiscalizador va a terreno, el cual puede ser lo suficientemente extenso para que la empresa corrija la falta. Sin embargo, existe un malgasto de recursos en este aspecto, debido a que hay un fiscalizador que concurre a la empresa y no encuentra infracción, incurriendo en un costo de oportunidad de ir a otra posible empresa infractora.

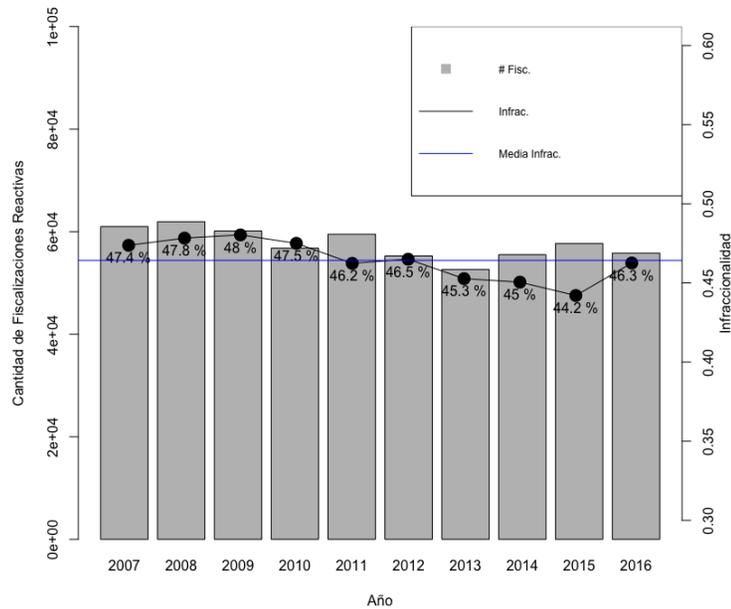


Figura 3: Gráfico Infraccionalidad detectada Reactiva por año.

Fuente: Elaboración Propia.

Por otro lado, la Figura 4 muestra la fiscalización proactiva, la cual posee una eficacia media del 27%, esto quiere decir que 1 de cada 4 empresas (aproximadamente) fiscalizadas presentan algún tipo de infracción. Para la definición que le entrega la DT a este tipo de fiscalización, este indicador es demasiado bajo, debido a que debe ser una fiscalización planificada, enfocada a los sectores donde hay mayor riesgo, por ende, una mayor probabilidad de encontrar una empresa infractora.

El comportamiento de este tipo de fiscalización es más aleatorio en términos de infraccionalidad, además de presentar una cantidad bastante menor de fiscalizaciones que la fiscalización reactiva, donde la fiscalización proactiva posee un 25% del total de fiscalizaciones.

Según esto, la fiscalización proactiva está por debajo de la línea base que calcula Kanbur, Ronconi & Wedenoja (2013) del 37%, lo cual en teoría describe en primera instancia que en promedio no se logra fiscalizar a las empresas infractoras dada esta diferencia entre el 27% de fiscalización proactiva a los 37% de infracción "natural" en Chile. Esto es un problema para la DT debido a que no se está cumpliendo su objetivo estratégico declarado del aumentar el cumplimiento de la ley laboral y denota un problema de eficacia en el modelo predictivo que actualmente se está utilizando y eficiencia de recursos en enviar fiscalizadores a empresas no infractoras.

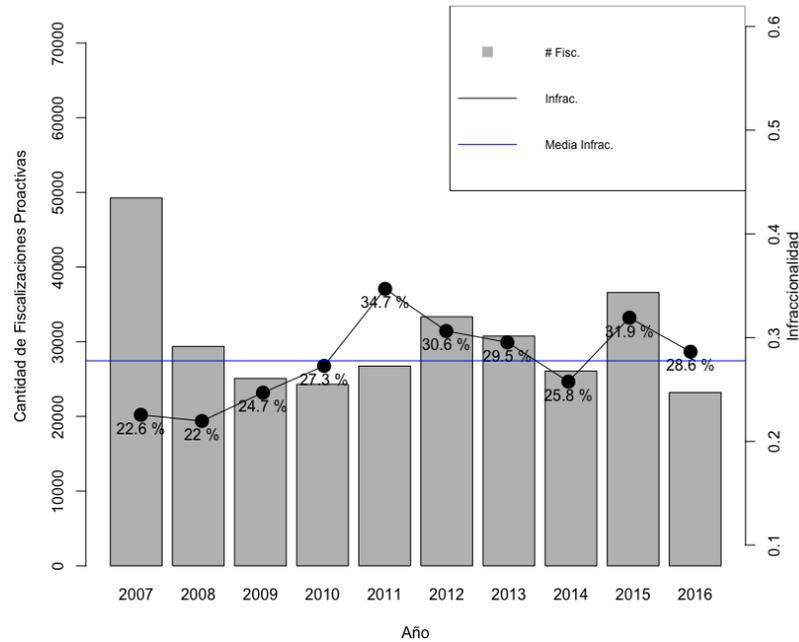


Figura 4: Gráfico Infraccionalidad detectada Proactiva por año.

Fuente: Elaboración Propia.

1.3 Objetivos y Resultados Esperados del Proyecto

1.3.1 *Objetivo General*

El proyecto a desarrollar, pretende aumentar la eficacia de la fiscalización proactiva, generando una cadena causal, la cual se inicia con el aumento de la percepción de riesgo, lo que disminuiría la tasa de Infraccionalidad base o “natural”. El Objetivo General del Proyecto se define como:

“Rediseñar Procesos de Fiscalización Programada para doblar la Eficacia al año 2018”

1.3.2 *Objetivos Específicos*

Los objetivos específicos del proyecto para poder lograr el objetivo General son los siguientes:

- Levantamiento y Análisis de los procesos involucrados en fiscalización proactiva al 2017.
- Obtención, Preparación y Análisis de los datos involucrados en el proceso general de fiscalización al 2017.
- Evaluación de Modelos de Minería de Datos e Inteligencia de Negocios que se ajuste a los datos obtenidos al 2017.
- Propuesta de Rediseño de Procesos de fiscalización Proactiva.

- Especificación UML del rediseño propuesto.

1.3.3 Resultados Esperados

Los resultados esperados a partir del proyecto son los siguientes:

- Mejor entendimiento del comportamiento de las empresas y su infraccionalidad a través del análisis de información de fiscalizaciones.
- Encontrar posibles patrones de infraccionalidad ocultos en los datos.
- Generar programas de Fiscalización apoyados en el análisis de información y modelamiento predictivo.
- Sentar bases para mejora continua del análisis de datos.

1.4 Alcance

El enfoque del proyecto está en la fiscalización proactiva de la DT, dirigida al rediseño de su proceso y en la adopción de herramientas de Minería de Datos para la predicción, conjunto a los datos y softwares necesarios para la implementación de estas herramientas.

1.4.1 Dentro del Alcance

Los elementos a considerar en el proyecto son los siguientes:

- Rediseño de Proceso de Fiscalización Proactiva, En la parte de la planificación y gestión de empresas a Fiscalizar.
- Evaluación de modelos de predicción – Se cubre la elección de cuales modelos son mejores para la predicción.
- Uso de Softwares – Se utilizan softwares que soporten los modelos.
- Análisis de Datos – Se cubre el tipo de datos necesarios para la aplicación.
- Prototipo de prueba – Se realiza un prototipo de la aplicación integrada para la predicción de empresas a fiscalizar.

1.4.2 Fuera del Alcance

- Implementación de la aplicación – No se realiza la implementación de la aplicación desarrollada, debido a que esto será hecho por la DT en el contexto del proyecto BID.
- Rediseño de fiscalización Reactiva – Debido a que se le quiere dar un mayor enfoque ala fiscalización proactiva en la DT. Además, se espera que la mejora de la fiscalización proactiva, disminuya la cantidad de fiscalizaciones reactivas.
- Optimización de la asignación de los fiscalizadores a cada fiscalización.

1.5 Riesgos Potenciales

Los posibles riesgos que posee el proyecto son presentados en la Tabla 1, conjunto a su Probabilidad, Tipo, Impacto y estrategia de mitigación.

Tabla 1: Riesgos Potenciales del Proyecto de Tesis.

Nº	Tipo	Riesgo	Probabilidad	Impacto	Estrategia de Mitigación
1	Político	Cambio de Objetivos de la Administración	100%	4	Evidenciar avances robustos para la mantención del proyecto
2	Económico	El presupuesto para el componente de fiscalización no es suficiente para la propuesta de rediseño	10%	7	Mostrar de manera temprana y robusta las necesidades monetarias del rediseño.
3	Técnico	Datos no entregados, o entregados en tiempos disconformes	40%	6	Iniciar el proceso de extracción de fuentes de datos de manera temprana
4		Falta de Requerimiento de StakeHolder importante	20%	6	Recolectar Todos los requerimientos de aquellos actores claves dentro del proceso
5		Falta de Acceso a datos de otros servicios importantes	60%	6	Conversaciones con los otros Servicios para alianzas estratégicas del uso de datos
6		Corrupción del Alcance del proyecto	50%	5	Selección de Alcance del Proyecto tempranamente
7	Organizacional	Resistencia al cambio	60%	6	Diseñar un plan de gestión del cambio, el cual integre a los trabajadores a las nuevas

					tecnologías a implementar.
8		Disponibilidad de los funcionarios durante el desarrollo del Proyecto	40%	3	Generar Instancias Eficientes para la muestra de avances y el feedback por ambas partes
9		Falta de Competencias y Conocimiento para el uso de la plataforma	20%	5	Manual de Uso de la Plataforma con los Casos de uso y su detalle de funcionamiento
10		No adopción (No Uso) de la herramienta	30%	7	Plan de Gestión de KPIs ligados al uso de la herramienta desarrollada

CAPÍTULO 2: MARCO TEÓRICO

2.1 Metodología de Ingeniería de Negocios

Dentro de la literatura se pueden encontrar diversas definiciones de la Ingeniería de Negocios, donde todas comparten puntos en común basados en la integración de Modelo de Negocios, Estructura Organizacional, Procesos de Negocios, Sistemas y Tecnologías de Información en la búsqueda del diseño de servicios. Dentro del trabajo de Barros (2017) convergen todos estos conceptos y añade una mirada sobre la Arquitectura Empresarial (Enterprise Architecture) en una Ontología para el diseño de Negocios, presentada en la Figura 5.

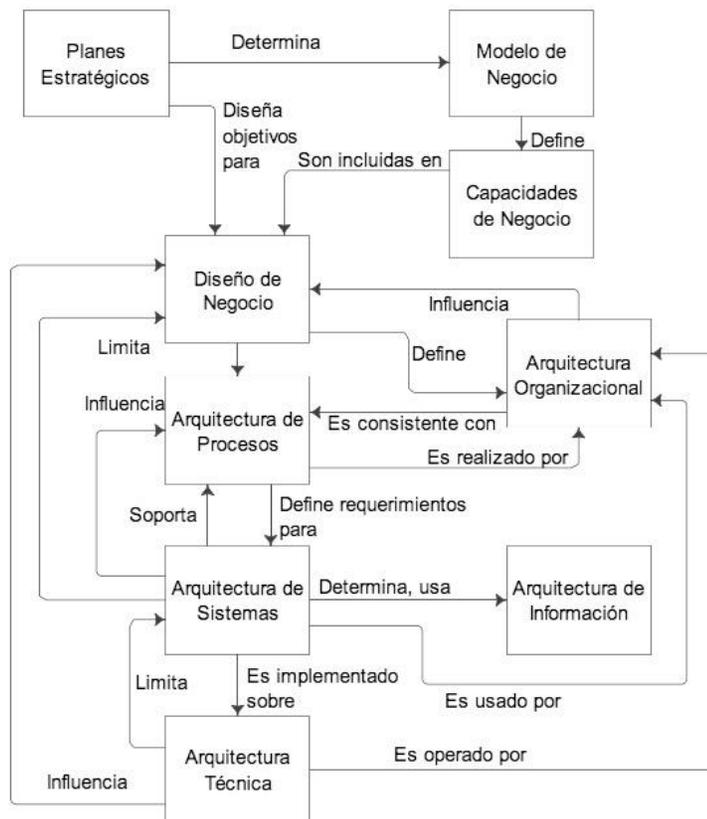


Figura 5: Modelo Ontológico.

Fuente: Barros, O. (2017).

Esta ontología incorpora el alineamiento de la planificación estratégica de la empresa con el Modelo de negocios como los principales pilares para el diseño del negocio, donde luego se encuentran las arquitecturas necesarias para el sustento del diseño y que permiten el funcionamiento de las actividades. Las arquitecturas son las siguientes:

- **Arquitectura de Procesos:** Arquitectura que establece los procesos necesarios para el diseño del negocio, la relación entre estos y la lógica del negocio.
- **Arquitectura Organizacional:** Arquitectura que se relaciona directamente con la de Procesos de acuerdo a la estructura de los roles organizacionales, el trabajo a realizar y por quien.
- **Arquitectura de Sistemas:** Arquitectura que define el sistema de información que posee la organización y el soporte a los procesos que entrega.
- **Arquitectura de Información:** Arquitectura de la Información que posee la organización en base a sus datos operacionales.
- **Arquitectura Técnica:** Arquitectura relacionada con todo el soporte físico (hardware) de los sistemas de las arquitecturas previamente mencionadas.

Este modelo ontológico demuestra un esfuerzo en la idea de hacer operativas las ideas y focos establecidos en la estrategia de la organización en base a un alineamiento de esta estrategia con las arquitecturas encargadas del funcionamiento operacional. Para cumplir esto, aparece el concepto de Metodología de Ingeniería de Negocios que comenta Barros (2016) en el mismo trabajo, la cual detalla las etapas a seguir para el diseño del negocio. La metodología queda detallada en la Figura 6.

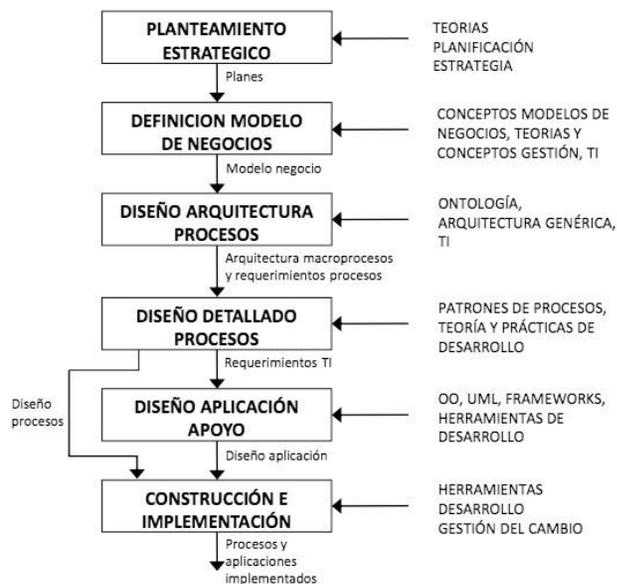


Figura 6: Metodología Ingeniería de Negocios.

Fuente: Barros, O. (2017).

2.2 Metodología CRISP-DM

La metodología CRISP-DM es una herramienta que presenta un proceso de tareas enfocadas a la aplicación de técnicas de minería de datos a empresas de diferentes industrias (Chapman, y otros, 2000). Esta metodología posee 4 niveles de abstracción donde se encuentran las fases, tareas genéricas, tareas especializadas e instancias de proceso. Cada uno de estos niveles contiene al siguiente, es decir, cada fase tiene sus tareas genéricas, las que a su vez poseen tareas especializadas y estas poseen instancias del proceso.

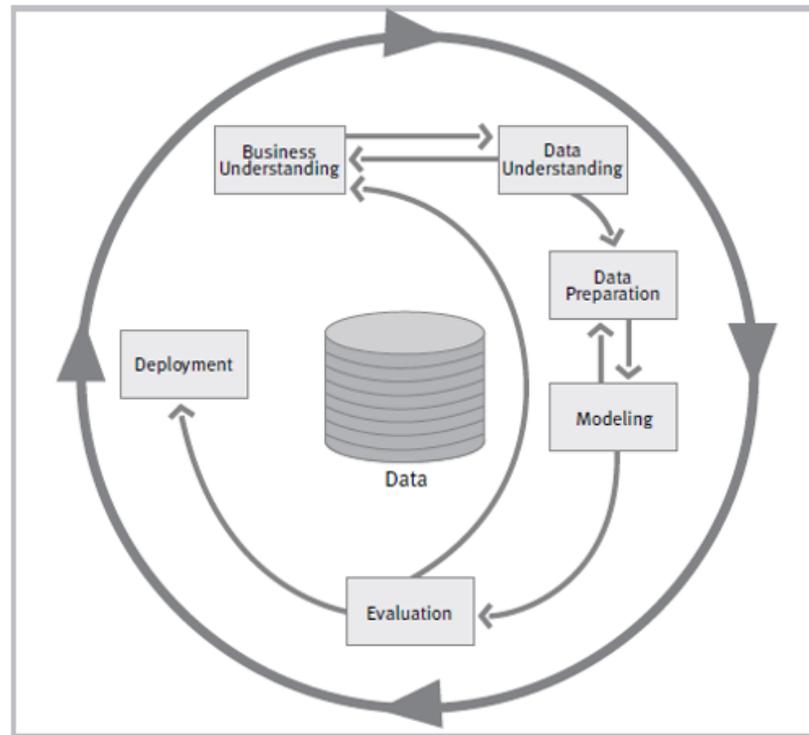


Figura 7: Fases del Modelo de referencia CRISP-DM.

Fuente: Chapman, y otros, 2000

Se pueden observar, en el modelo de referencia de la Figura 7, las fases de la metodología, donde cada una de estas se explican a continuación:

- Entendimiento del Negocio: Fase inicial la cual se enfoca en el entendimiento del negocio de la organización conjunto al entendimiento de los objetivos y requerimientos del proyecto base, con la finalidad de convertir el proyecto base en un proyecto de Minería de Datos con sus respectivos objetivos y plan de acción.
- Entendimiento de los Datos: Fase que inicia con la primera recolección de datos, donde se realizan actividades que permiten la familiarización y

entendimiento de estos. Además, se identifican problemas de calidad de los datos y posibles hipótesis en base a las primeras observaciones.

- **Preparación de los Datos:** Fase que involucra todas las tareas asociadas a la construcción de la base de datos final para el modelamiento a partir de la recolección inicial. Esta fase es iterativa e involucra tareas como: Selección de variables, limpieza de los datos, transformación de datos, entre otros.
- **Modelamiento:** Fase donde se prueban diferentes técnicas de modelamiento en la base de datos final, donde, dependiendo de la técnica, es posible tener que preparar los datos de forma diferente a la realizada en la fase previa, por lo tanto, volver a iterar entre la fase de preparación de datos y modelamiento es común. Además, se calibran los parámetros de las diferentes técnicas de modelamiento con el fin de obtener resultados óptimos.
- **Evaluación:** Fase donde se evalúan los modelos construidos en la fase previa, donde se asegura que el modelo cumpla con los objetivos de negocio definidos en la fase inicial, además de la revisión de los pasos para la creación y ejecución de este. Por otro lado, al final de esta fase se obtienen las decisiones de negocio de la utilización de los resultados de los modelos.
- **Despliegue:** Fase donde, en base al conocimiento obtenido por los modelos, se ejecutan acciones directas en los procesos de la organización, con el fin de concretar el objetivo de negocio del proyecto base. Es usual que esta tarea la tome la organización y no el analista, pero es tarea del analista el traspasar el conocimiento obtenido en los datos a la organización.

2.3 Minería de Datos

Minería de Datos (Data Mining) es un término utilizado entre industrias que concentra técnicas y disciplinas que permiten la obtención de conocimiento y patrones escondidos dentro de bases de datos, útiles para la toma de decisiones dentro de las organizaciones (Han, Kamber, & Pei, 2012). La minería comienza con la extracción y tratamiento de bases de datos de cualquier fuente y tamaño para su posterior procesamiento en base a métodos y modelos computacionales que hacen uso de técnicas estadísticas, inteligencia artificial, aprendizaje de máquina, entre otras, donde posteriormente, luego de reiteradas iteraciones y evaluaciones, se obtienen patrones ocultos en los datos que ayudan al conocimiento del negocio.

La minería de datos se utiliza para una vasta serie de tareas como se puede observar en la Figura 8, dentro de las cuales, los más usados dentro de las organizaciones son: Clasificación y Clusterización.

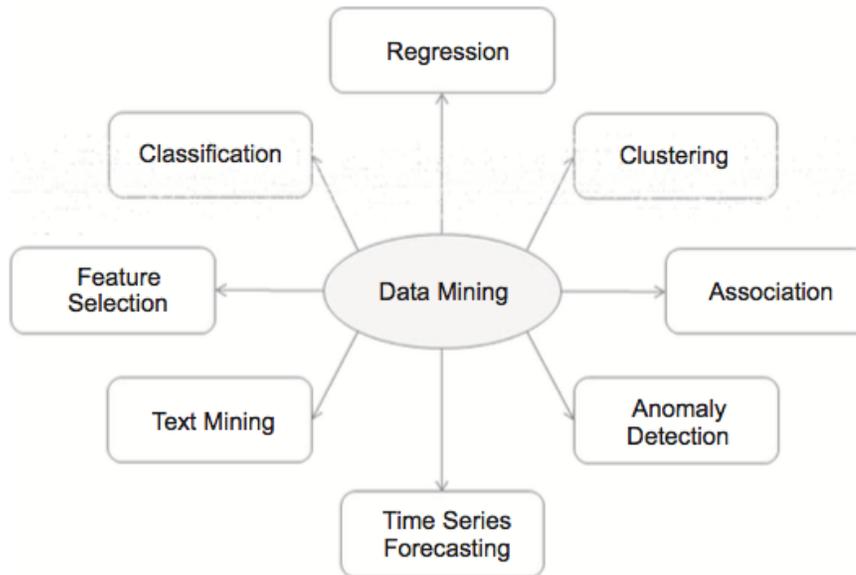


Figura 8: Tareas de la Minería de Datos.

Fuente: Han, Kamber, & Pei. (2012).

La **Clasificación** es el proceso de encontrar una función o patrón que permita distinguir (predecir) clases entre los datos. Este proceso se realiza a partir de una base de datos de entrenamiento, la cual posee etiquetadas las clases respectivas a predecir. Con este entrenamiento, se obtiene un modelo parametrizado, para posteriormente predecir una base de datos de prueba que no posea etiquetadas las clases. Modelos típicos de clasificación son: Árboles de decisión, redes neuronales, vectores de soporte, entre otros.

La **Clusterización**, a diferencia de la clasificación, es un proceso que no entrena con una base de datos etiquetada, sino que crea clases (clusters) en base al criterio de la máxima similitud dentro de los clusters y mínima similitud entre clusters. Con esto, el cada cluster puede ser visto como una clase, donde se pueden etiquetar o nombrar en función de las características similares de las observaciones, con lo cual se puede conocer de mejor manera los datos y su entendimiento en términos del negocio. Modelos típicos de clusterización son: K-means, Nearest-Neighbor Algorithm, DBSCAN, entre otros.

2.3.1 Modelos de Minería de Datos

Los modelos de Minería de Datos basados en Aprendizaje de Máquina se pueden separar en dos grandes corrientes: Modelos Supervisados y Modelos no Supervisados.

Los modelos Supervisados se basan en encontrar relaciones en set de datos categorizados con el fin último de categorizar set de datos no categorizados con las variables de entrada con las cuales fueron entrenados los modelos. Existe una variable objetivo la cual posee la categorización, donde esta puede ser

binaria o tener una serie definida de clases, la cual se intenta predecir en base a un set de variables explicativas. Este modelamiento comúnmente se realiza con una separación de la base de datos inicial, donde se separa en un set de entrenamiento (*training set*) y un set de prueba (*testing set*), ambos con la variable objetivo conocida. Inicialmente se modela con los datos de entrenamiento, donde se obtiene el modelo ya entrenado con su respectivo error, luego, con el set de prueba, se predicen los valores de la variable objetivo con el uso de las variables explicativas, para finalmente calcular el error de prueba con los valores originales de esta.

Los modelos no supervisados también ayudan a encontrar patrones y relaciones en los datos, pero a diferencia del aprendizaje supervisado, el set de datos no posee una categorización, por ende, el modelamiento se basa en la relación entre las observaciones, donde se trabaja con distancias entre ellas.

Dentro del trabajo se utilizaron los siguientes modelos:

2.3.1.1 Regresión Logística

La regresión logística es una versión compleja de una regresión lineal, donde una de las principales diferencias es que la variable objetivo es binaria, por lo cual la estimación lineal puede entregar resultados fuera de rango. Para esto, la regresión logística utiliza la función Logit, la cual es una derivación matemática del cálculo del logaritmo natural de los puntos de ventaja de probabilidad o más conocido en inglés como *Odds*. Esta función se modela de manera lineal respecto a las variables explicativas donde se obtiene la siguiente ecuación:

$$\text{Logit} = \text{Log} \left(\frac{p}{1-p} \right) = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n$$

con $p = \text{probabilidad de ocurrencia del evento}$

De la ecuación se desprende que la probabilidad p es la función sigmoideal cuyo argumento es la regresión lineal de las variables explicativas, la cual se encuentra acotada entre 0 y 1 lo que permite obtener resultados basados en la variable objetivo binaria.

$$p = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_n X_n}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_n X_n}}$$

El modelo, para el aprendizaje, debe calcular los valores de cada coeficiente para la futura clasificación de los datos. Para el cálculo de los coeficientes se estiman por máxima verosimilitud, donde los valores de la probabilidad y variables explicativas son conocidas. Luego, con cada coeficiente estimado, se obtiene el modelo parametrizado, donde se utilizan los datos de prueba para obtener el resultado de la predicción, donde se debe decidir de un punto de corte para la clasificación de las clases (si es binario el corte común es 0,5) dado que el modelo retorna valores entre 0 y 1.

El modelo de regresión logística es altamente usado en diversos estudios e industrias debido a su facilidad de uso e interpretación (Vijay, 2015), sin embargo, los coeficientes no se pueden interpretar directamente como se hace con una regresión lineal, debido a la componente no lineal del modelo y debe considerarse que cada coeficiente obtenido en la parametrización del modelo está relacionado las *Odds*. Un coeficiente, es el logaritmo de las *Odds* de la variable relacionada a este. Además, con los coeficientes obtenidos en el modelamiento, se pueden realizar test de significancia a las variables, por lo cual es un buen método para la selección de variables relevantes al análisis.

2.3.1.2 Árboles de Decisión

Los arboles de decisión son modelos predictivos para clasificación altamente utilizados en la minería de datos (Vijay, 2015) debido a su facilidad de creación desde el punto de vista de analista y su facilidad de entendimiento desde el punto de vista del negocio. Según lo descrito por Han, Kamber, & Pei (2012) estos modelos, son una estructura de nodos y ramas que explican reglas realizadas por el proceso de aprendizaje para la separación de las clases en base a los atributos o variables explicativas entregadas en el set de datos de entrenamiento.

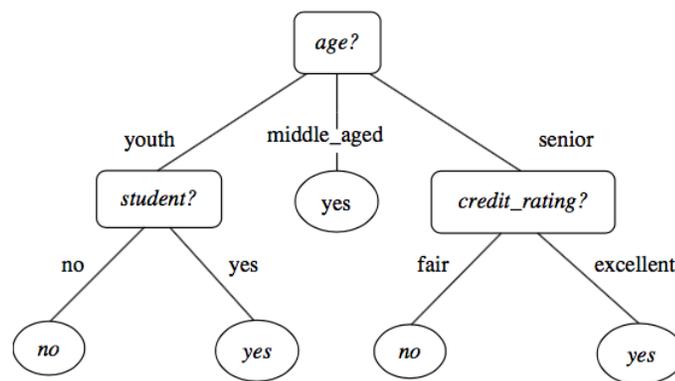


Figura 9: Ejemplo Árbol de Decisión.

Fuente Han, Kamber, & Pei (2012).

Los nodos son test realizados a algún atributo, donde este test separa los datos en base a una condición. Luego, vienen las ramas, las cuales representan el resultado de la condición entregada por el nodo previo. Por último, están los nodos hojas, los cuales poseen la clase o etiqueta predicha en base a las condiciones realizadas previamente. En la Figura 9 se observa un ejemplo de árbol de decisión y su estructura gráfica.

El método para la creación y aprendizaje de un árbol inicializa con 3 parámetros: el dataset inicial de entrenamiento, la lista de atributos a analizar y el método de selección de atributos (Índice de Gini, Ganancia de Información, entre otros). Se inicia con un nodo, si el nodo posee todos los datos de una misma clase, el nodo

se transforma en hoja, si no, se utiliza el criterio de separación el cual indica la mejor forma de separar las ramas en base al atributo del nodo, generando cortes en la variable (numérica o categórica). Dependiendo del tipo de partición, el cual a su vez depende del tipo de variable, el nodo se etiqueta en base al criterio y se generan las respectivas ramas con su resultado, donde para cada uno de estos, se vuelven a probar los atributos y se repite el algoritmo, el cual se detiene cuando ocurren 3 tipos de condiciones:

- Los datos dentro de un nodo son todos de la misma clase.
- No existen más atributos en los cuales se pueda separar los datos, por lo cual se utiliza voto mayoritario para etiquetar el nodo con la clase más representante.
- No existen datos para una rama, por lo cual se crea una hoja en el nodo del cual nace la rama y se etiqueta con la clase mayoritaria.

El resultado del algoritmo genera una serie de reglas que permiten clasificar los datos de acuerdo a la variable supervisada deseada, donde esto permite, de una manera gráfica, entender el comportamiento que presentan los datos.

Los tipos de árboles a utilizar en el proyecto son los modelos CART (Árboles de Clasificación y Regresión por sus siglas en inglés), Conditional Inference Trees (Árboles de Inferencia Condicional en su traducción) y Random Forest (Bosques Aleatorios en su traducción).

Los modelos **CART** y **Conditional Inference Trees** se diferencian por su método de selección de atributos, donde el primero utiliza medidas de información como el índice de Gini, mientras que el segundo utiliza test estadísticos como el uso del p-valor para la separación de atributos. Esto permite una comparación entre métodos debido a que modelos como el CART tienden a buscar variables que permiten una mayor cantidad de separaciones en términos de la mejor, mientras que el modelo de inferencia condicional busca una medida en términos de la significancia estadística de estas separaciones (Hothron, Hornik, & Zeileis, 2016).

Por otro lado, los **Bosques Aleatorios** son un modelo que realiza un número fijo de árboles de decisión, donde en cada uno de estos realiza una selección (aleatoria o no) de las variables explicativas del problema, donde finalmente, en base a métodos como el voto mayoritario para clasificación o promedios para la regresión obtiene un resultado final para el entrenamiento del modelo. Este tipo de modelo permite realizar un análisis de importancia de variables, debido a que en cada árbol creado y entrado permite rastrear el resultado dependiendo de las variables que utilizó, donde finalmente compara estos resultados en función del resto de los árboles.

2.3.1.3 Redes Neuronales

Las redes neuronales son un modelo de aprendizaje de máquina supervisado, cuyo nombre proviene de su similitud a la función del cerebro humano a través de neuronas conectadas. Estas redes consisten en 3 elementos: Capas,

Neuronas y Pesos donde a unidad principal de la red son las neuronas, las cuales son unidades tipo input/output que se interconectan entre ellas a través de pesos (Han, Kamber, & Pei, 2012).

Existen 3 capas en una red neuronal, la capa de entrada, capa oculta y capa de salida. La **capa de entrada** es el conjunto de neuronas que toman las observaciones del set de datos de entrenamiento, siendo estas neuronas el símil de las variables explicativas, donde son alimentadas directamente a cada neurona.

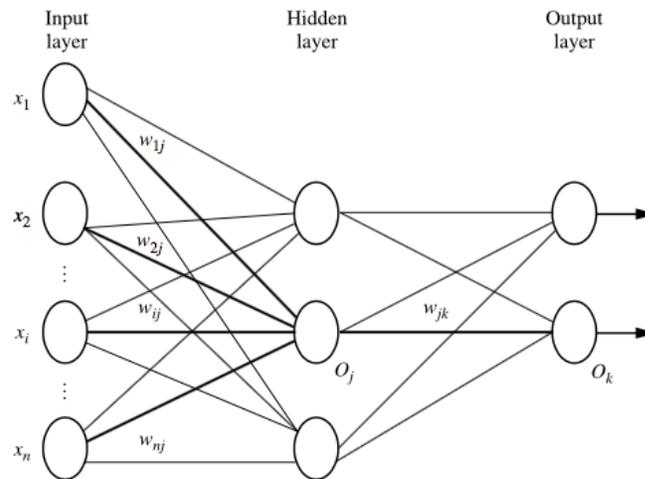


Figura 10: Esquema de una red neuronal.

Fuente Han, Kamber, & Pei (2012).

Posteriormente, se encuentra la capa oculta la cual posee un número arbitrario de capas de neuronas (seleccionado por el analista, donde usualmente es una), donde las salidas de las neuronas de estas capas son calculadas a través de la aplicación de una función de activación a la suma ponderada de los outputs de las neuronas previas.

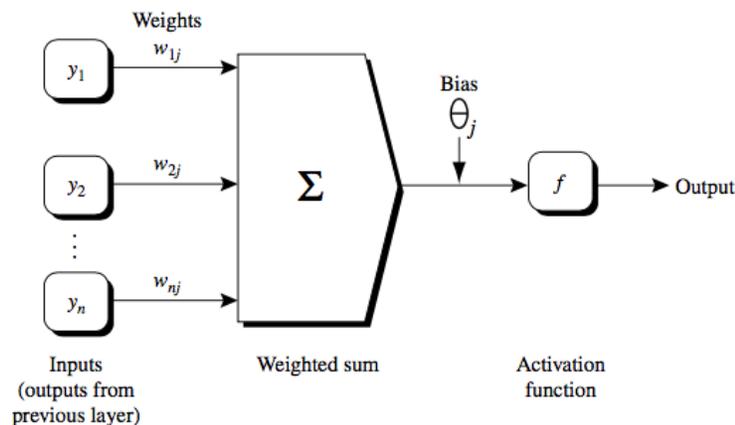


Figura 11: Cálculo del output de una neurona de capa oculta o salida.

Fuente Han, Kamber, & Pei (2012).

La Figura 11 muestra el flujo que siguen las neuronas tanto en la capa oculta como en la capa de salida, donde las funciones de activación más utilizadas son la función logística y la tangente hiperbólica. También se observa una neurona "Bias", donde cada capa posee esta neurona extra, la cual se encarga de que la función de activación no anule o active todas las entradas del set de datos, donde actúa de manera similar a la constante n de una función lineal del tipo:

$$f(x) = mx + n$$

donde no siempre los resultados pasan por el origen, por ende, un valor para esta constante debe ser distinto de 0 (para ese tipo de casos).

El mecanismo de aprendizaje de una red neuronal es a través del algoritmo de propagación hacia atrás o "Backpropagation", donde la red aprende en base a iteraciones de las observaciones del set de datos de entrenamiento conjunto a un set inicial de pesos y neuronas "bias". El algoritmo, en términos simples, inicializa los pesos y las neuronas "bias" de manera aleatoria, ingresa las observaciones del set de datos de entrenamiento a la capa de entrada y calcula el output para cada neurona de las capas posteriores, donde finalmente se obtiene el resultado final en la capa de salida el cual es comparado con el valor conocido de la clase o etiqueta, para calcular el error de predicción. El objetivo del algoritmo es minimizar este error a través de la modificación de los pesos de manera inversa o "hacia atrás" iterativamente, donde se modifican primero los pesos de la capa de salida, para seguir con los pesos de las capas ocultas donde se vuelve a calcular la predicción en base a los cambios y se repite nuevamente este paso. Eventualmente, los pesos convergen (no siempre hay convergencia) y el algoritmo se detiene y se obtiene el modelo final de predicción de la red neuronal.

Las redes neuronales poseen una alta capacidad predictiva debido a su aprendizaje iterativo, pero requiere altos tiempos de procesamiento debido a esto. Por otro lado, las redes neuronales son vistas como cajas negras, debido a su complejidad matemática, por ende, no son buenos modelos para explicar el porqué de la predicción dada.

CAPÍTULO 3: PLANTEAMIENTO ESTRATÉGICO Y MODELO DE NEGOCIOS

3.1 Posicionamiento Estratégico

Para el análisis del posicionamiento estratégico de la entidad se recurre al modelo delta de Hax para empresas sin fines de lucro (Hax, 2010). Según este modelo, la Dirección del Trabajo se encuentra cercano al vértice de Look-in Sistémico, en la dimensión Canal de Entrega, donde la DT es la única organización fiscalizadora de la legislación laboral, netamente por un tema legislativo, dado que la ley no permite tener más entes fiscalizadores en el ámbito.

Dado este posicionamiento impuesto (no es obtenido por competencia, sino por decretos de ley), debido a lo comentado previamente, la DT posee una serie de problemáticas en términos de eficiencia, similar a las ineficiencias atribuidas a los monopolios de economías perfectas, donde al ser los únicos proveedores del servicio entregado, la capacidad es limitada. De acuerdo a esto, el posicionamiento que la organización busca a futuro va en dirección de Mejor Producto, ubicado específicamente en la Eficiencia Administrativa, donde en base a los recursos limitados de la entidad conforme a su presupuesto, realice su función de manera eficiente, efectiva y eficaz.

3.2 Balanced Scorecard

La Dirección del Trabajo no posee la herramienta de gestión BSC dentro de sus prácticas, sí posee 4 objetivos estratégicos, pero no existe una herramienta que permita gestionar avances ni metas. Los objetivos estratégicos de la organización son:

- Aumentar el cumplimiento de la legislación laboral, mejorando la cobertura y oportunidad de la fiscalización laboral, a través de la fiscalización reactiva y proactiva.
- Promocionar la organización sindical y la libertad sindical, a través de la defensa de la libertad sindical y del incremento de la capacitación, difusión y atención preferencial a los actores sindicales.
- Aumentar y mejorar el acceso, la oportunidad y la entrega de productos y servicios de la institución hacia los ciudadanos, a través de las inspecciones del trabajo y sitio web.
- Incrementar la cobertura y la calidad de los sistemas de prevención y solución alternativa de conflictos y las instancias de diálogo social, a través de la conciliación y mediación.

Para el cumplimiento de los objetivos, la organización está inmersa en el proyecto de Modernización BID, el cual toma acción para todos los objetivos declarados por la DT. A modo de ejercicio se realiza un Balanced ScoreCard de la organización en base a los objetivos del proyecto BID.

3.2.1 *Perspectiva Financiera*

- **Gestión Eficiente de recursos**
 - Disminuir tiempos de ocio de mediadores y conciliadores en un 50 % hasta el 2020.

3.2.2 *Perspectiva Usuarios*

- **Aumentar el cumplimiento de la legislación laboral**
 - Aumentar el asertividad de fiscalización proactiva en un 80 % hasta el año 2020.
- **Incrementar Cobertura y calidad de los sistemas de prevención y solución alternativa**
 - Mejorar Indicadores de Satisfacción de Encuesta anual en un 10 % de forma anual.

3.2.3 *Perspectiva Procesos Internos*

- **Aumentar y mejorar el acceso, la oportunidad y la entrega de productos**
 - Aumentar la tasa de tramites web en un 50 % hasta el 2020.
 - Disminuir tasa de abandono de los llamados en un 80 % hasta el 2020.
 - Disminuir la tasa de fuga de usuarios presenciales en las 6 Oficinas con mayor diferencia al promedio, en un 20 % en promedio de las 6 oficinas hasta el 2017.
- **Mejorar asignación de fiscalizadores**
 - Disminuir los tiempos medios entre la puesta de la denuncia hasta la asignación del fiscalizador en un 50 % hasta el 2020.
- **Mejorar rastreo de requerimientos de los usuarios**
 - Disminuir número de consultas cuyas respuestas difieren en resultado dentro de un mismo requerimiento (respuestas heterogéneas) en un 100 % hasta el 2020.

3.2.4 *Perspectiva Aprendizaje y Crecimiento*

- **Aumentar instancias de retroalimentacion entre Departamentos**
 - Mejorar Instancias de retroalimentación en la DT

3.3 Modelo de Negocios

Para analizar el modelo de negocios de la DT, se utilizó el modelo presentado por Osterwalder, el cuál es el Canvas, dado por su completitud y facilitar el alineamiento con la estrategia de la organización.

3.3.1 Propuesta de Valor

La Dirección del Trabajo tiene como propuesta de valor la entrega de servicios de fiscalización, mediación, conciliación, supervisión a todos sus usuarios dentro del país, además de generar y habilitar conocimiento y estudios expertos respecto a la legislación laboral. Estos servicios se entregan con trato igualitario, eficiente en sus intervenciones, técnica especializada en materia de trabajo, moderna y alcanzable en sus medios.

3.3.2 Interfaz del Cliente

Los **clientes** de la DT son todos los actores laborales del mercado en Chile, pueden ser tanto empleados como empleadores, además de organizaciones y organismos de trabajadores. Estos usuarios son atendidos con trato igualitario.

Los canales por los cuales los usuarios acceden a los servicios son:

- Presencial, el cual representa un 55% del total de solicitudes
- Telefónica, con un 42 % del total.
- Vía web, el cual solo tiene un 3 % del total.

De esta manera, la **relación con los usuarios** es directa y especializada, de cara a solucionar de manera integral (en todas sus aristas) los requerimientos de estos, por lo que se cuenta con profesionales en temáticas laborales.

3.3.3 Infraestructura de Gestión

Para poder entregar de manera efectiva y eficiente los servicios que la DT posee, los **procesos claves** son:

- Atención Especializada de Usuarios.
- Fiscalización Proactiva.
- Fiscalización Reactiva.
- Mediación.
- Conciliación.

Estos procesos son, los que, en conjunto, los que generan valor para el usuario debido a que son aquellos procesos que requieren de mayor especialización y tienen un impacto directo en el cumplimiento de la legislación laboral. Por un lado, la atención al usuario debe ser oportuna y rápida, y posteriormente, ya sea fiscalización, mediación o conciliación, debe tener una resolución eficaz. Estos procesos se explicarán mejor posteriormente.

Al ser la DT una entidad pública, tiene como principal **recurso** a su capital humano, son las personas quienes son el principal activo dentro de la organización dado que estos son los que crean el conocimiento que entrega la DT dentro de sus dictámenes y están encargados de los aspectos técnicos que los requerimientos presentan. También son claves los recursos tecnológicos, que permiten entregar un servicio más satisfactorio para los usuarios, y la legislación

laboral, ya que es el insumo principal para la creación de toda la información y conocimiento que entrega la DT, y es la base en cual se generan las multas y resoluciones de las demandas. Por último, su único y principal **socio clave** es el Ministerio del Trabajo y Previsión Social el cual está en constante contacto a la DT en materias de inspección, debido a que es dependiente directo de este ministerio y es quien entrega el presupuesto para la realización de las actividades de la DT.

3.3.4 Flujos de Ingresos y Estructura de Costos

Los **costos** de la DT se desglosan en 3 componentes:

- Gastos de Personal, con 45.635 MM\$ en el año 2015, lo que representa un 78 % del presupuesto.
- Bienes y Servicios de consumo, con 10.633 MM\$; 18 % del presupuesto.
- Otros, con 2.260 MM\$ con el último 4 % del presupuesto.

Estos costos se mantienen porcentualmente durante el tiempo, y como era de esperar, el mayor exponente es el gasto de personal, el cual es el activo principal de la entidad. Por otro lado, el **ingreso** de la DT depende del Presupuesto anual, el cual es entregado por el Ministerio del Trabajo y Previsión Social. Las multas no son de ingreso directo a la DT, por lo que no se consideran dentro de los ingresos de la organización.

Como bien se explicó previamente, el posicionamiento estratégico de la empresa es el Canal de entrega (Lock-in Sistémico), esto presenta a la DT como el único medio el cual el Gobierno entrega estos servicios, se observa claramente en el modelo de negocios que, independiente de esta unicidad, la organización está buscando el uso óptimo de recursos y entregar un servicio de calidad con miras a la satisfacción del usuario. Esto es de suma importancia, dado que, en el sector privado un monopolio incentiva a la ineficiencia, ocurre lo mismo en el sector público, dado que al ser el único que imparte un servicio, los usuarios deben “acomodarse” al nivel de atención que este entregue. Es aquí donde se ve la oportunidad de mejorar las entidades públicas desde un punto de vista del rediseño de procesos y la modernización con tecnologías de la información. Precisamente el Ministerio de Hacienda conjunto al Banco Interamericano del Desarrollo (BID), actualmente, son pioneros del proyecto de Modernización del Estado, el cual busca promover la mejora de los procesos de las entidades públicas con el único fin de aumentar la satisfacción del usuario.

CAPÍTULO 4: ANÁLISIS SITUACIÓN ACTUAL

4.1 Arquitectura de Procesos

La arquitectura de procesos de la DT (Figura 12) consta de tres cadenas de valor, las cuales son; Consultoría Laboral, Fiscalización de empresas y solución Alternativa de Conflictos. Se destaca que existe un proceso común entre todas las cadenas, el cual es la Atención de Requerimientos, la cual posee los canales presenciales, web y telefónico, cuya recepción de requerimientos es derivada a una de las 3 cadenas de valor o el requerimiento, dadas sus características iniciales, es respondido directamente por los asistentes laborales.

La cadena de valor “Consultoría Laboral” son todos aquellos requerimientos de consulta, los cuales son atendidos por asistentes laborales dentro de las oficinas, Call center o web, siendo uno de los requerimientos con más alta frecuencia en las oficinas del país. Por otro lado, los requerimientos de “solución alternativa de conflictos” presenta una evaluación más completa, ya sea para requerimientos de conciliación o mediación, siendo necesaria una posible revisión por parte de abogados del servicio. La cadena de “Fiscalización de Empresas” consta del proceso de fiscalización el cual es explicado en la apertura de Modelamiento de los patrones de procesos siguientes.

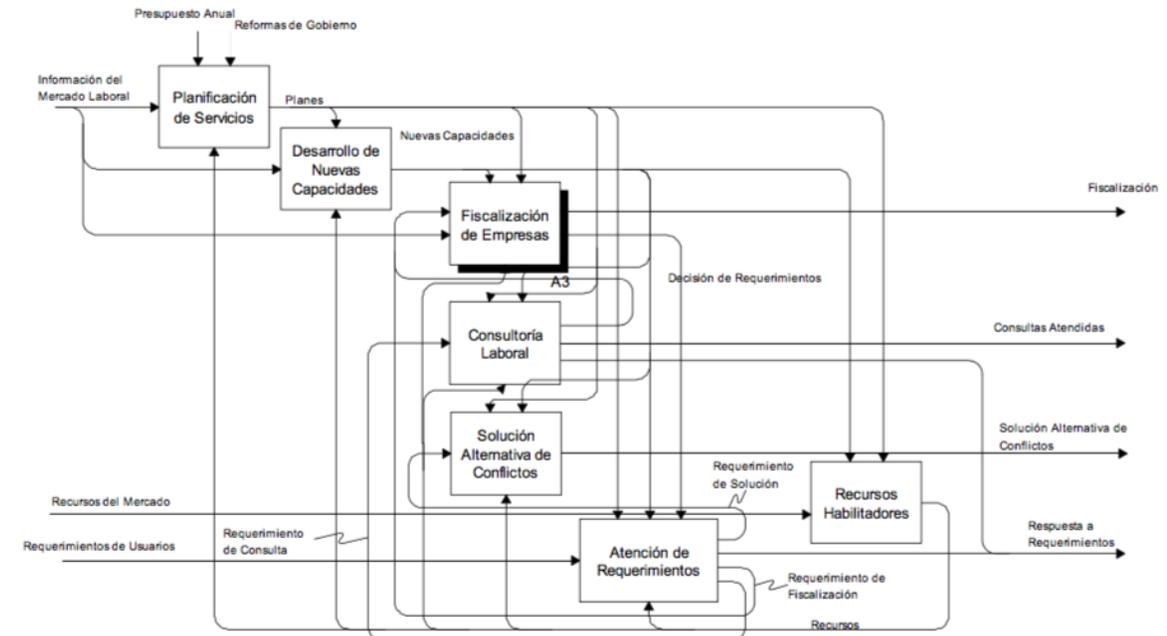


Figura 12: Arquitectura de Procesos, Dirección del Trabajo.

Fuente Elaboración Propia.

4.2 Modelamiento Detallado de Procesos

4.2.1 Modelamiento IDEF0

Al abrir el macroproceso Fiscalización de Empresas, se observan los procesos que conforman la macro, siguiendo los patrones de procesos entregados por Barros (2017). Se observa en la Figura 13 la Administración y Análisis de las Fiscalizaciones, donde se realiza la planificación de las demandas a satisfacer como a las fiscalizaciones proactivas y se toman las decisiones de cuales requerimientos a satisfacer de manera reactiva. También se tiene la “Gestión de Fiscalizaciones” donde se programa operativamente los planes anuales de fiscalización en base al análisis entregado por el primer proceso y el respectivo control de las fiscalizaciones a realizar y realizadas. Luego se encuentra la Entrega del Servicio de Fiscalización como tal, el cuál es el trabajo de campo, donde el fiscalizador concurre a la empresa/establecimiento y la evalúa de acuerdo a los formularios existentes en la DT.

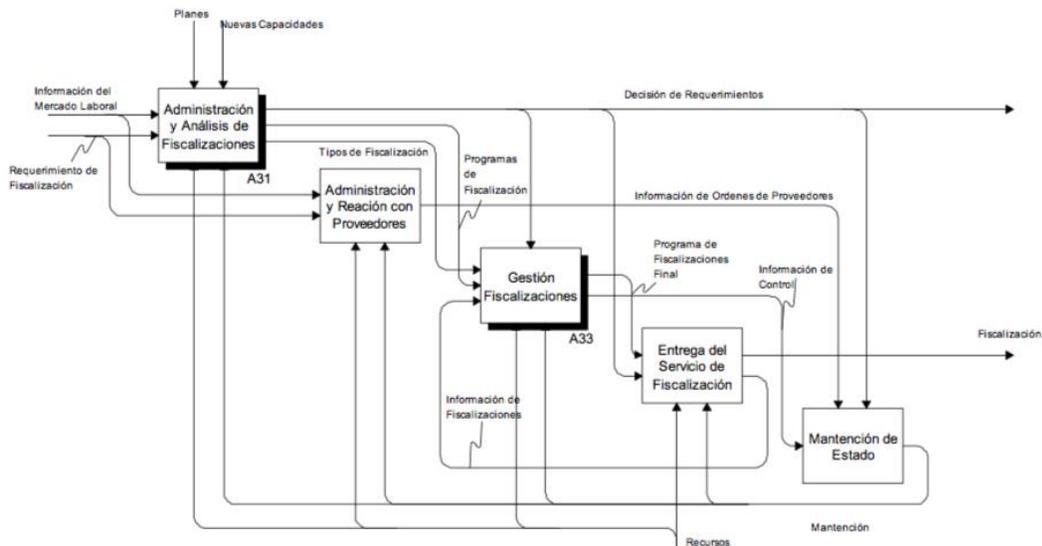


Figura 13: Macro 1 Fiscalización de Empresas, Dirección del Trabajo.

Fuente Elaboración Propia.

Debido al proyecto de tesis, el enfoque del rediseño va direccionado al proceso “Administración y Análisis de Fiscalizaciones”, donde se ubica la aplicación en primera instancia del modelo predictivo a proporcionar por el proyecto y su respectiva aplicación futura en la planificación de programas, en base a la gestión necesaria a la capacidad de la organización y los objetivos estratégicos. Con esto presente, se abre el primer proceso, Administración y Análisis de Fiscalizaciones donde se observa, en la Figura 14, el proceso de “Análisis de Información del Mercado Laboral”. En este proceso se realizan todos los estudios respecto a toda la información que posee la DT respecto al mercado laboral, tanto como información propia de fiscalizaciones realizadas, como bases de datos de

empresas del SII (Servicio de Impuestos Internos) e información gubernamental. Luego se ubica el proceso Decisión de Requerimientos a satisfacer, el cual entrega las reglas de negocio para que un requerimiento termine finalmente en una fiscalización, donde se analizan las características de esta entrada y es evaluado por los respectivos Fiscalizadores.

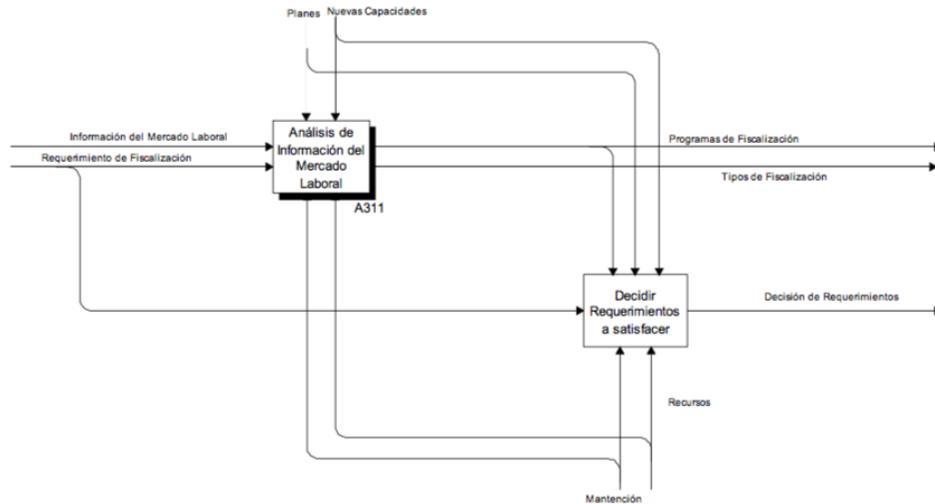


Figura 14: Administración y Análisis de Fiscalizaciones, Dirección del Trabajo.

Fuente Elaboración Propia.

Continuando con la apertura, se abre el proceso de Análisis de Información del mercado laboral en la Figura 15, en la cual se ubica la introducción de nuevas fiscalizaciones como la fiscalización electrónica empezó a formar parte del set de fiscalizaciones el año 2016. Luego se tiene el “análisis del comportamiento laboral” el cual son revisiones de fiscalizaciones pasadas y su efectividad por Actividad Económica, materias fiscalizadas, etc. donde se entregan lineamientos futuros a ser fiscalizados, ya sea por focos gubernamentales, compromisos inter-institucionales o contingencias. Por último, está la “planificación estratégica de programas de fiscalización”, el cual contempla las características de empresas a fiscalizar de manera proactiva/programada a partir de los lineamientos previamente obtenidos, es importante destacar que se fiscaliza por lineamiento, donde la decisión de a qué empresa fiscalizar queda en manos del fiscalizador y no de aquellos que planifican.

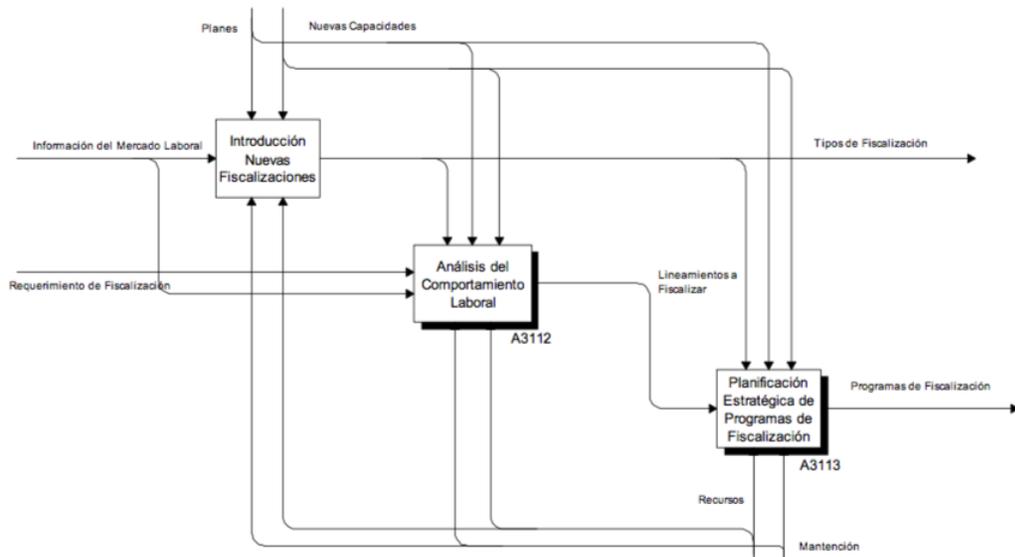


Figura 15: Análisis de Información del Mercado Laboral, Dirección del Trabajo.

Fuente Elaboración Propia.

En la apertura del “Análisis del comportamiento laboral”, se puede observar, en la Figura 16, la preparación de los datos históricos de las fiscalizaciones, los cuales entregan la información de la eficacia de estas y toda la información recopilada en el trabajo de campo realizado. Luego se pasa a la revisión de estos datos, donde se generan informes descriptivos de las fiscalizaciones realizadas de manera agregada y se generan conclusiones de los posibles lineamientos a fiscalizar, donde se generan lineamientos de comercio, construcción, migrantes, entre otros.

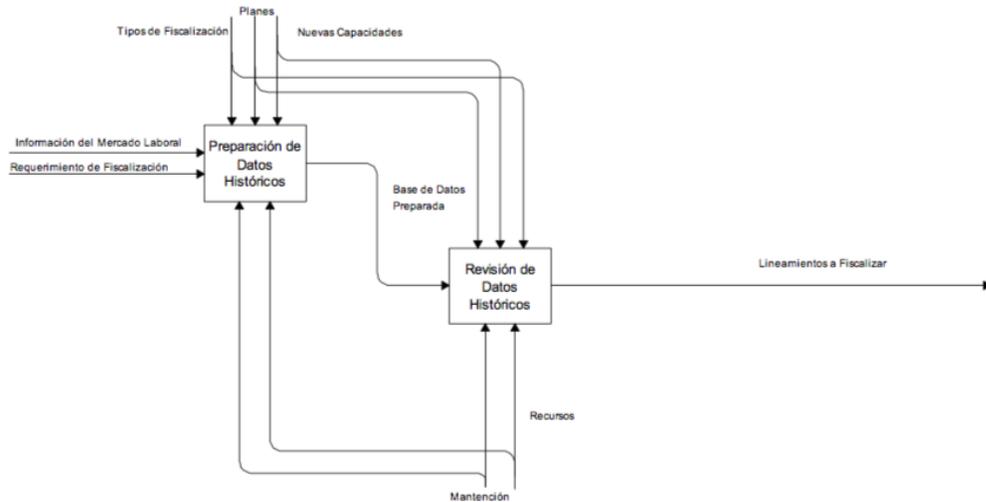


Figura 16: Análisis del Comportamiento Laboral, Dirección del Trabajo.

Fuente Elaboración Propia.

Por otro lado, se proporciona la apertura de “Planificación Estratégica de Programas de Fiscalización”, dado que este proceso es uno de los cuales se beneficiará del modelo predictivo a proporcionar por el proyecto. Se puede observar en la Figura 17 los procesos: “Recopilación de Antecedentes de Lineamientos” y “Preparación de Circulares de Programas”. El primero contempla toda aquella información por la cual realizar el programa, ya sean cifras notables, compromisos con otras instituciones y todo lo necesario para introducir y explicar la necesidad el programa de fiscalización. Por último, se encuentra la Preparación de la circular del programa, la cual es el documento que contiene los antecedentes recopilados e información de control y fechas límite. Este documento es entregado a todas las direcciones regionales.

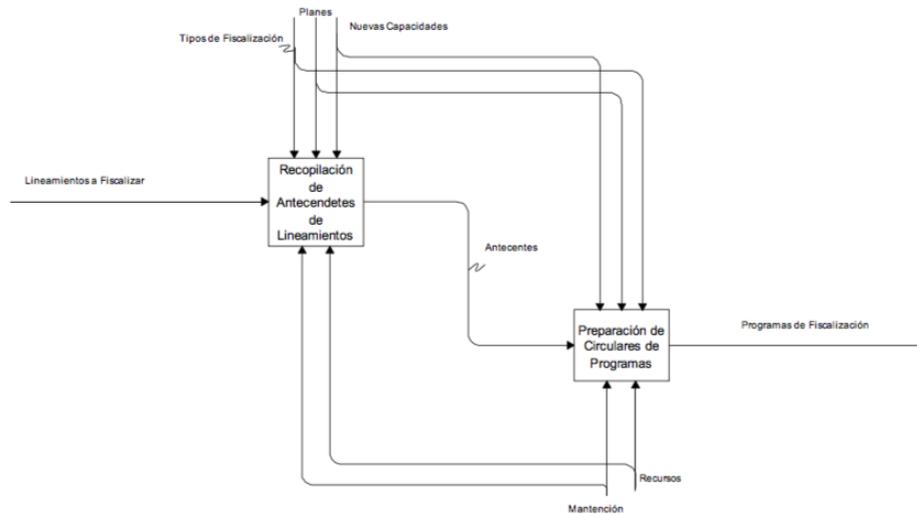


Figura 17: Planificación Estratégica de Programas de Fiscalización, Dirección del Trabajo.

Fuente Elaboración Propia.

4.3 Diagnóstico de la Situación Actual

El principal problema que presenta la DT es debido a la existencia de una cantidad, a lo largo del tiempo de fiscalizaciones reactivas (bajo denuncia), que supera 3 veces (en promedio) a las fiscalizaciones programadas, lo cual da muestras de un modelo basado en la reacción y no en la prevención, por ende, existe infraccionalidad dentro de las empresas chilenas y la DT posee información de estas, dado el ingreso de las denuncias de sus usuarios. Acá se presenta una oportunidad de tratamiento de esta información para poder clasificar y predecir posibles empresas infractoras dadas sus características observables y de comportamiento, con el fin de obtener un programa basado en un análisis de minería de datos.

También es importante destacar que la cantidad de empresas fiscalizadas anualmente, que bordean las 100.000 en promedio, son solo un 5% del total de empresas en Chile, lo que también entrega un problema de eficiencia de recursos dada la capacidad limitada de la organización para fiscalizar a las empresas. Con esto, es importante que el programa esté enfocado a la utilización eficiente de fiscalizadores y en consecuencia un foco a las empresas más probables de ser infractoras, donde el modelo de data mining que se desea desarrollar es la herramienta encargada de esto último.

4.4 Cuantificación del Problema u Oportunidad

Debido a que el problema es de eficiencia de recursos, se detalla la siguiente cuantificación según datos de fiscalización del 2014.

Se poseen costos de fiscalización realizada los cuales son estimados según el costo por hora de los fiscalizadores obtenido en la página oficial de la dirección del trabajo, la cual da un valor aproximado de **\$12.256** por hora (45 horas semanales). Con una duración promedio de 4 horas por fiscalización y un costo de transporte aproximado a **\$2.500** se tiene un costo por fiscalización de **\$51.524**.

Se considera una fiscalización Efectiva, a aquella fiscalización, independiente si es proactiva o reactiva, que resultó encontrar infracción, dado a que el costo en el cual se incurrió para realizar la investigación fue efectivo en su objetivo final. A partir de esto se toman los datos del 2014, donde se realizaron de manera completa 101.442 fiscalizaciones, donde los resultados fueron los siguientes:

Tabla 2: Valorización Fiscalizaciones Efectivas y No Efectivas.

	Total	Porcentaje	Valor en CLP
Efectivas	37.113	36,5%	1.912.210.212
No Efectivas	64.379	63,5%	3.317.063.596
Total	101.442	100%	5.229.273.808

En la Tabla 2 se puede observar que el 36,5% de las fiscalizaciones en el 2014 resultaron ser efectivas, por lo tanto, el costo por realizarlos también es efectivo, sin embargo, cerca de los 2 tercios resultaron no ser efectivas, donde se tiene un costo no efectivo de **\$3.317.063.596** anual, debido a que no se encontró infracción. Este costo no es totalmente no efectivo debido a que se obtiene información que es importante, como es saber que efectivamente no es una empresa infractora, pero ese recurso puede ser destinado de mejor manera, lo

cual lleva consigo, el disminuir la cantidad de denuncias dado que se fiscaliza a las empresas “correctas”.

CAPÍTULO 5: PROPUESTA DE DISEÑO DE PROCESOS

5.1 Direcciones de Cambio y Alcance

El actuar de la Dirección del Trabajo, en base a los últimos 10 años, se ha basado en las fiscalizaciones reactivas, donde estas son más del 70% de las fiscalizaciones totales al año, destacando un modelo en base a la reacción de lo que ocurre en el universo laboral de Chile. Si bien el espíritu de esto pudo haber sido útil años atrás, la cantidad de empresas en Chile ha crecido a un 6% anual según el SII y la tarea de visitar y fiscalizar a estas, por consiguiente, ha resultado ser más difícil. Con la limitante en términos de fiscalizadores y recursos para ir a fiscalizar, se hace necesario el focalizar la acción inspectiva, con el fin de cubrir de mejor manera posible las necesidades y requerimientos de los usuarios. Este foco se puede ver reflejado en la fiscalización proactiva de la dirección del trabajo, dado su carácter preventivo y sorpresivo, donde con la ayuda de los avances de los tratamientos y procesamiento de la información, se puedan realizar programas de fiscalización preventiva, en base a las empresas más riesgosas que ha presentado la historia de la información de la organización.

En resumen, la dirección que se desea tomar en la organización con el proyecto es de fortalecer la Fiscalización Proactiva, en base a la eficiencia de los recursos actuales, anticipándose a las posibles denuncias a futuro conforme a programas de fiscalización efectivos basados en el análisis de la información.

Para esto, se realiza un análisis de las variables de cambio según la metodología de Ingeniería de negocios de Barros (2017).

5.1.1 Estructura de la Empresa y Mercado

Tabla 3: Variable Estructura de la Empresa y Mercado

a.	Estructura de la Empresa y Mercado	Actual	Propuesto
a.1	Servicio Integral al cliente	No	Mantener situación actual.
a.2	Lock-in Sistémico	Si, Canal de Entrega	Se desea pasar a un enfoque de eficiencia administrativa.
a.3	Integración con proveedores	No	Mantener situación actual.

a.4	Estructura Interna: centralizada o descentralizada	Descentralizada	Se centraliza la decisión de cuales empresas ir a fiscalizar.
a.5	Toma de decisiones: centralizada o descentralizada	Parcialmente descentralizada	Mantener situación actual.

5.1.2 Anticipación

Tabla 4: Variable Anticipación

b.	Anticipación	Actual	Propuesto
b.1	Planificación de Fiscalizaciones Proactivas	Si, en base a la experiencia.	Planificación en base a la información y a la experiencia.
b.2	Modelos Predictivos de Empresas	No.	Incorporar modelos de inteligencia que apoyen a la planificación.

5.1.3 Coordinación

Tabla 5: Variable Coordinación

c.	Coordinación	Actual	Propuesto
c.1	Reglas	Reglas definidas en base al código del Trabajo.	Incorporar posibles nuevas reglas en base al análisis de la información.
c.2	Jerarquía	Jerarquía administrativa.	Mantener situación actual.
c.3	Colaboración	Si, formalmente en reuniones.	Mantener situación actual.

c.4	Partición	Por tipo de requerimiento.	Mantener situación actual.
-----	-----------	----------------------------	----------------------------

5.1.4 Prácticas de trabajo

Tabla 6: Variable Prácticas de trabajo

d.	Prácticas de trabajo	Actual	Propuesto
d.1	Lógica de Negocio: Automatizada o semi-automatizada	Escasa automatización de planificación. Lógica en base a la experiencia.	Se propone una nueva lógica, semi-automatizada en base a metodología de minería de datos
d.2	Lógica de apoyo a actividades tácitas	No.	Automatización de los indicadores de los modelos predictivos a realizar, para la toma de decisiones.
d.3	Procedimientos de comunicación e integración	Existe un proceso de comunicación de programas.	Agregar mejoras al sistema tecnológico que acompaña a la comunicación de los programas.
d.4	Lógica y procedimientos de medición de desempeño y control	Si, al final de los programas.	Incorporar mediciones en tiempo real en la medida que se vaya agregando datos de fiscalizaciones al sistema.

5.1.5 Integración de Procesos Conexos

Tabla 7: Variable Integración de Procesos Conexos

e.	Integración de Procesos Conexos	Actual	Propuesto
e.1	Proceso aislado	No.	Incorporar un nuevo proceso, que tiene impacto en procesos relacionados.
e.2	Todos o la mayor parte de los procesos de un macroproceso	No.	Generar una retroalimentación entre los procesos de administración, gestión y ejecución dentro de la macro 1.
e.3	Dos o más macros que interactúan	Si, las 2 cadenas interactúan en función de la fiscalización reactiva	Mantener situación actual.

5.1.6 Mantención Consolidada del Estado

Tabla 8: Variable Mantención Consolidada del Estado

f.	Mantención Consolidada del Estado	Actual	Propuesto
f.1	Datos Propios	Si. Sistema DT-PLUS y SIRELA	Se propone tener datos específicos para la evaluación de los modelos predictivos.
f.2	Integración con datos de otros sistemas de la empresa	No, los sistemas operan por	Se propone integrar los datos de forma

		separado sin integración automatizada.	automatizada para el análisis
f.3	Integración con datos de sistemas de otras empresas	No.	Se propone agregar datos de otros servicios como el SII.

5.2 Diseño Detallado de Procesos

El proyecto de rediseño no afecta a la arquitectura de procesos de la Dirección del Trabajo, por lo que solo conlleva a rediseño de procesos dentro de la cadena de valor “Fiscalización de empresas”, donde los procesos a rediseñar se ubican dentro del proceso “Administración y Análisis de las Fiscalizaciones”. En la apertura del proceso “Análisis de Información del Mercado Laboral”, se rediseñan los procesos “Análisis del Comportamiento Laboral” y “Planificación Estratégica de Programas de Fiscalización”.

5.2.1 Diseño en IDEF0

La apertura de la cadena de valor de fiscalización tampoco cambia de forma estructural, de la misma manera, al abrir el proceso de “Administración y Análisis de las Fiscalizaciones”, este no cambia dado el alcance del proyecto. El proceso que se desea rediseñar es el de Análisis de Información del Mercado Laboral, al cual, dentro de su estructura “AS IS”, se desea rediseñar el proceso “Análisis del Comportamiento Laboral” donde se incorpora el desarrollo de modelos predictivos y rediseñar la “Planificación Estratégica de Programas de Fiscalización”, donde se incorpora la aplicación del modelo predictivo desarrollado en el proceso anterior, donde se utiliza para la selección de empresas en base al “scoring” generado. La estructura se observa en la Figura 18.

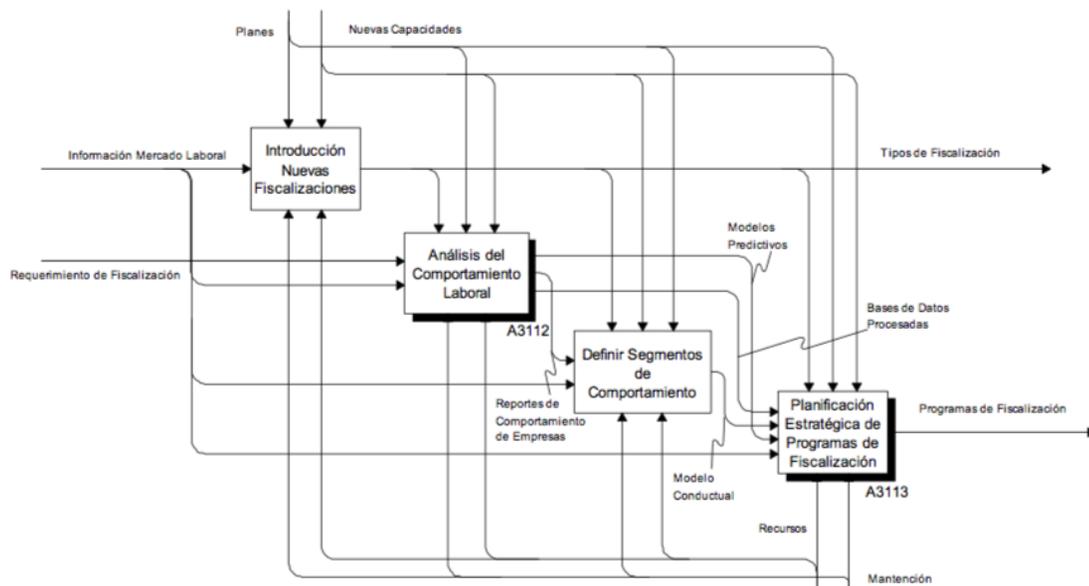


Figura 18: Análisis de Información del Mercado Laboral, Rediseño.

Fuente Elaboración Propia.

Luego, en la apertura de Análisis del Comportamiento Laboral, la estructura se rediseña, cambiando el proceso de Revisión de Datos Históricos a dos nuevos

procesos los cuales son “Desarrollar Modelos de Predicción” y “Desarrollar Modelo de Cambio Conductual”, donde ambos están enfocados en la fiscalización de las empresas infractoras de manera proactiva y en el tipo de trato que esta deba tener según el segmento de comportamiento que se encuentre. Ambos procesos serían nuevos en la Dirección, por lo cual el proceso del “Desarrollo y Actualización de modelos predictivos” es totalmente nuevo y de diseño elaborado por el autor del proyecto. La estructura del proceso rediseñado se observa en la Figura 19.

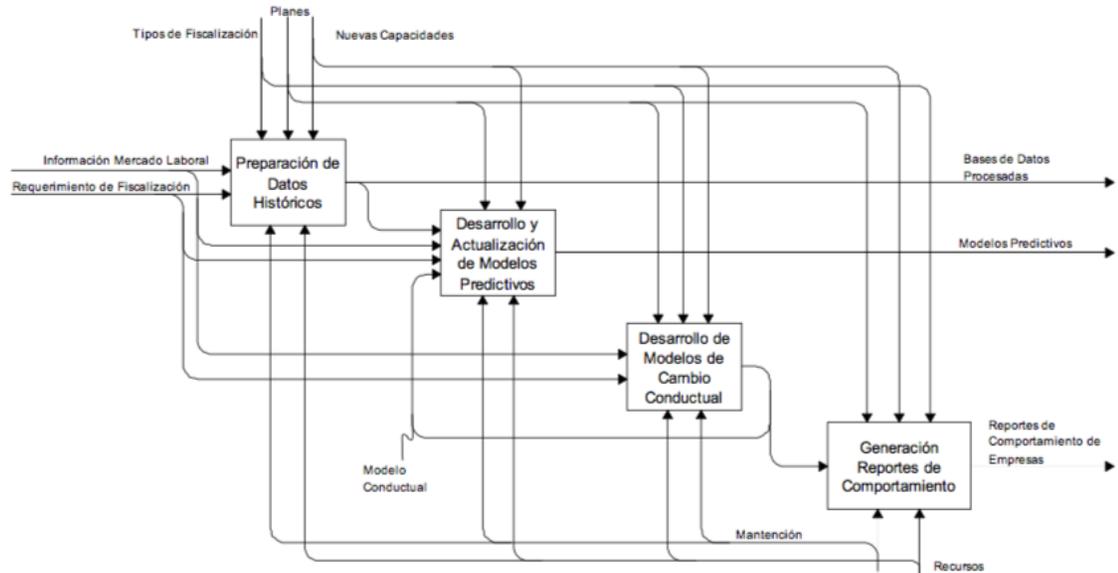


Figura 19: Análisis del Comportamiento Laboral, Rediseño.

Fuente Elaboración Propia.

Por último, en la apertura de Desarrollo de Modelos de Predicción se observa en la Figura 20 que presenta los 3 procesos genéricos planteados en la estructura de procesos, los cuales son “Evaluar desempeño de pronósticos”, “Actualizar y Evaluar Parámetros del Modelo” y “Ajustar y evaluar otros Modelos”. Se presenta un diagrama BPMN para mejor entendimiento del proceso, donde se detallan las decisiones a tomar, las tareas clave dentro del proceso y el tipo de datos a utilizar para la evaluación del modelo, además se incorpora la variable tiempo, donde este proceso será ejecutado una vez al mes para analizar su desempeño.

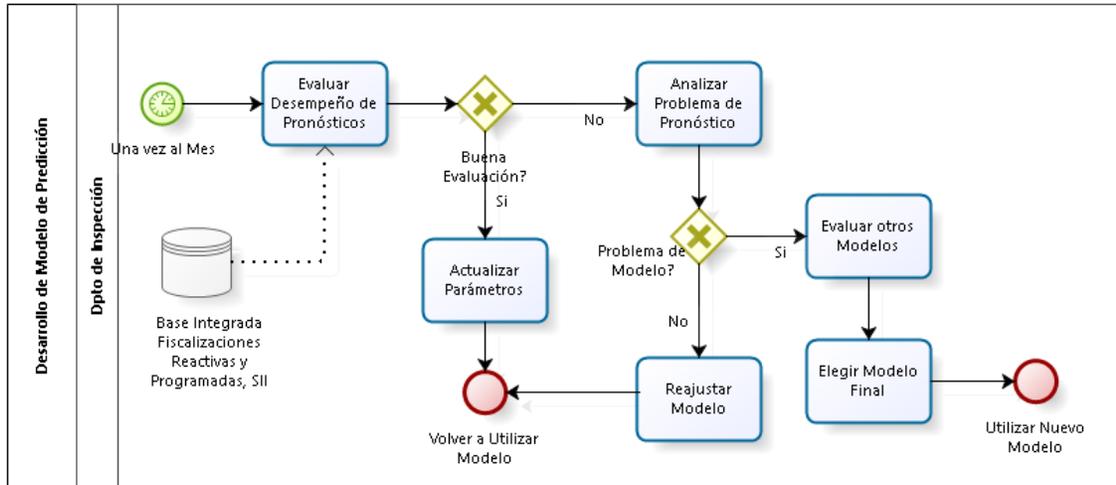


Figura 20: Desarrollo y Actualización Modelos Predictivos, Rediseño.

Fuente Elaboración Propia.

Respecto a la apertura del proceso “Planificación Estratégica de Programas de Fiscalización”, se observa en la Figura 21, la incorporación de dos nuevos procesos, los cuales se refieren a la aplicación de los modelos predictivos desarrollados y la preparación de la base de datos para la entrega a las direcciones regionales. Debido a que ahora se seleccionan las empresas de manera previa, se incorpora este nuevo proceso para ingresar los datos correspondientes a la ubicación de las empresas a fiscalizar.

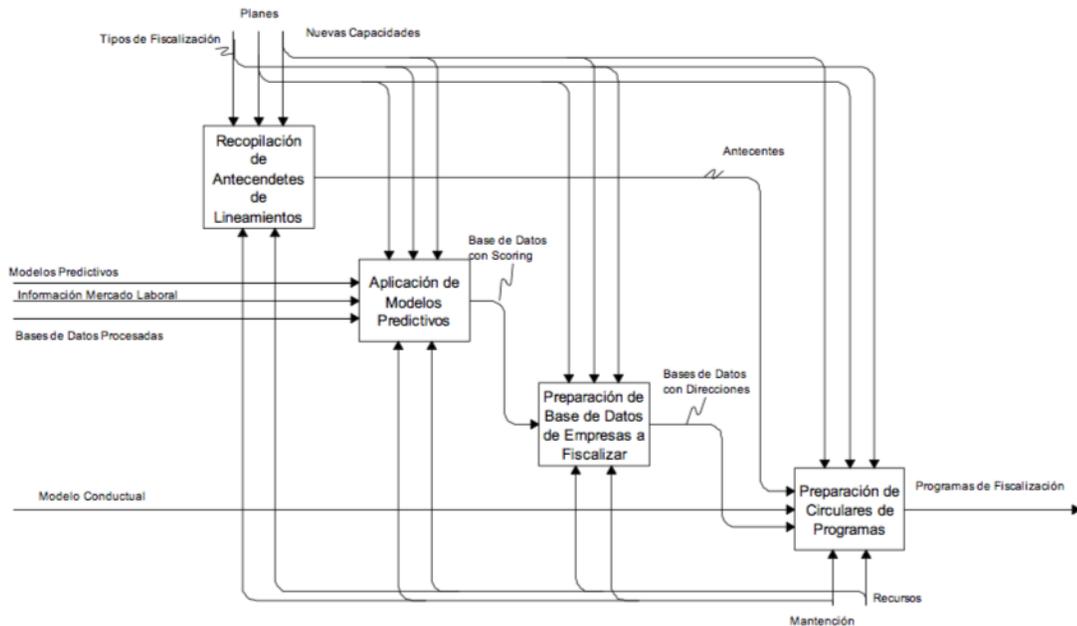


Figura 21: Planificación Estratégica de Programas de Fiscalización, Rediseño.

Fuente Elaboración Propia.

5.3 Lógica de Negocios

El motor principal que mueve el rediseño propuesto es la incorporación de Inteligencia de Negocios a los procesos de Fiscalización proactiva, por lo cual se requiere de una retroalimentación constante para poder minimizar el error que los modelos puedan entregar, dado que un supuesto general, y no alejado de la realidad, es que el comportamiento de las empresas varía con respecto al tiempo, por ende, un modelo fijo no sería flexible a este tipo de cambios ni sería un modelo basado en la realidad. Para esto, la nueva lógica de negocios está basada en una lógica iterativa de mejora continua como la propuesta por Deming la cual consiste en un ciclo llamado PDCA (Plan, Do, Check, Act) por sus siglas en inglés.

En la Figura 22, las etapas instanciadas para el problema son: Modelar, Planificar, Fiscalizar y Evaluar, las cuales son alimentadas por las fuentes de datos internas y externas. Es importante señalar que esta lógica no sigue un proceso único sino un conjunto de procesos señalados en la arquitectura y estructura de procesos, siendo únicamente etapas abstractas que señalan una acción a realizar.

El **Modelar** consiste en la creación y/o modificación de los modelos de inteligencia utilizados para la creación de programas, siendo alimentado por las bases de datos tanto internas como externas. Esta etapa constituye todos los procesos enfocados al análisis y preparación de los datos, desarrollo y modificación de modelos y creación de listados de empresas con su probabilidad de infracción.

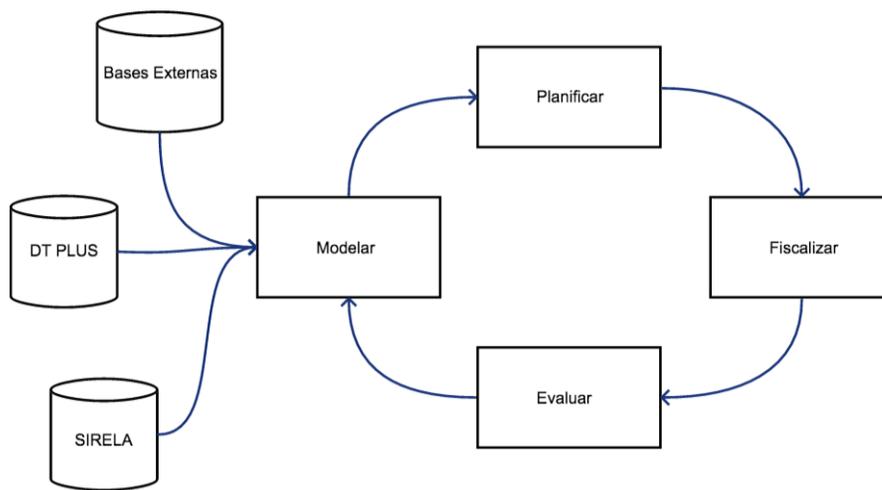


Figura 22: Lógica de Negocios Propuesta.

Fuente Elaboración Propia.

El **Planificar** es toda acción enfocada a la creación del programa y sus respectivas definiciones, como en qué regiones se realizará, cuáles serán las

materias a fiscalizar en el programa, objetivos de este, entre otras cosas. Esta etapa termina cuando cada fiscalización posee un fiscalizador encargado de realizar la inspección ya sea en terreno o de otro tipo.

Fiscalizar es toda acción inspectiva realizada por los fiscalizadores, donde estos, en base al programa, realizan las respectivas fiscalizaciones las cuales fueron planificadas previamente. Esta etapa termina cuando todas las fiscalizaciones del programa poseen un tipo de término, por ende, cada una de estas es etiquetada en términos de infraccionalidad, donde se obtiene el programa completo realizado.

El **Evaluar** es la etapa donde, en base a los resultados obtenidos por la etapa previa, se analiza la capacidad predictiva que tuvo el modelo y se evalúan diferentes indicadores para esto. Esta evaluación entrega indicios si el modelo predijo de manera aceptable o hay que realizar modificaciones para mejorar el error de predicción.

De esta manera la lógica se repite y se generan instancias de aprendizaje y retroalimentación las cuales permiten el tomar acciones de mejora dentro de las etapas y como ciclo general. Se espera que en la medida que haya iteraciones realizadas, los indicadores se estabilicen, dado que las iteraciones iniciales pueden generar cambios abruptos en los indicadores a evaluar.

5.4 Prueba de la Lógica de Negocios

Para la prueba de la lógica de negocios, se realizaron 2 grandes iteraciones a los datos entregados por la Dirección del Trabajo, donde inicialmente se trabajó con los datos desagregados por fiscalización realizada y una segunda iteración con los datos agregados por RUT de empresa y su respectiva infraccionalidad. Para esto, se utiliza la metodología CRISP-DM (*Cross Industry Standard Process for Data Mining*), la cual posee fases probadas en diferentes industrias para la aplicación de minería de datos a las empresas y organizaciones (Chapman, y otros, 2000). En los siguientes apartados se detallan las fases de la metodología y su aplicación a la organización.

5.4.1 Entendimiento del Negocio

La fase de entendimiento del negocio es en base a la experiencia que posee el autor en la organización, donde desarrolló sus actividades dentro de la Dirección del Trabajo por 12 meses, con lo cual se obtuvo conocimiento de las prácticas y el negocio que permitieron realizar el proyecto de tesis.

5.4.1.1 Determinar Objetivos de Negocios

El objetivo principal que se destaca es el objetivo del proyecto de tesis, el cuál es el aumentar la eficacia de las fiscalizaciones proactivas al doble, rediseñando sus procesos. Este objetivo tiene como directriz principal el cambio de un modelo de

fiscalización en base a la reacción a uno preventivo, con el fin de anticiparse a las infracciones.

5.4.1.2 Evaluación de Situación Actual

La evaluación de la situación actual se desarrolló en el presente capítulo, en apartados anteriores.

5.4.1.3 Determinación de objetivos de la Minería de Datos

Los principales objetivos de la Minería de Datos son:

- Análisis de los Datos de la organización.
- Entendimiento del Comportamiento a través del análisis.
- Obtención de Modelos que posean un ajuste adecuado en términos de predicción.

5.4.2 Entendimiento de los Datos

5.4.2.1 Recopilación de Datos Iniciales

Para la obtención de los datos se realizó una serie de reuniones con fiscalizadores y profesionales TI, quienes detallaron los sistemas que posee la organización y sus respectivas tablas de datos. Los sistemas de la Dirección del Trabajo son:

- DT-PLUS: Sistema donde se realiza el ingreso y egreso de las denuncias/fiscalizaciones, se suben los informes de fiscalización, se obtienen los resultados de las multas a partir de estos.
- SIRELA: Sistema donde se encuentra la información de las relaciones laborales, tales como organizaciones sindicales, negociaciones colectivas, huelgas, entre otros.

Se decide tomar como set de datos, el período 2007-2016, donde se obtiene la historia correspondiente a las fiscalizaciones realizadas.

De acuerdo al Diagrama Entidad Relación que se otorgó por la organización (ver anexo 1), se recopilamos datos de acuerdo al proceso de fiscalización que se separan en términos de datos como: **Ingreso de Fiscalización** la cual es la información que se obtiene por el solicitante cuando se ingresa al sistema, **Informe de Fiscalización** información obtenida durante la fiscalización y es ingresada al sistema por el inspector. Para cada uno de estos se recopilamos los siguientes datos:

- Ingreso de Fiscalización
 - Oficina de origen de la fiscalización.
 - Número de la comisión
 - Tipo de Solicitante de la fiscalización
 - Urgencia de la fiscalización

- Afectados
- Fecha del registro
- Informe de Fiscalización
 - Tipo de Término de la fiscalización
 - Egreso de la fiscalización
 - RUT de la empresa
 - Comuna del domicilio fiscalizado
 - Comuna del domicilio de la casa matriz
 - Número de Trabajadores
 - Tipo de Empresa
 - Código de Actividad Económica
 - Rama de Actividad Económica
 - Detección de Infracción
 - Materias Fiscalizadas e Infraccionadas

Además, a partir de los datos del sistema SIRELA, se incorpora una variable la cual es el número de sindicatos que tiene la empresa.

Por último, se obtiene una base de datos del Servicio de Impuestos Internos del listado de contribuyentes entre los años 2006-2016, donde se une, en base al RUT de la empresa los siguientes campos:

- Año
- Razón Social
- Tramo de venta
- Número de trabajadores
- Dirección
- Región
- Rubro
- Subrubro
- Actividad económica principal
- Fecha de inicio de actividad
- Fecha de término de giro
- Tipo de término de giro
- Tipo de contribuyente
- Subtipo de contribuyente
- Información del capital propio tributario, para cada contribuyente (f22c645 y f22c646).

Con esta recopilación, se obtiene una base final de 65 variables y 1.408.087 registros los cuales por fila poseen una fiscalización realizada.

5.4.2.2 Descripción los Datos

Para conocer de mejor manera los datos (solo de la DT) planteados previamente se presenta la Tabla 9, donde los tipos de datos están basados en la herramienta estadística R. Estos tipos de datos varían entre numéricos (Integer o Double),

textos (Character) los cuales no poseen una estructura definida y categóricos (Factor) los cuales poseen categorías dentro de sus registros, donde dentro de la tabla el número de categorías estará detallado entre paréntesis.

Tabla 9: Descripción de Datos de la Dirección del Trabajo.

Nombre de Variable	Descripción	Tipo Dato en R
CodOficina	Código de la oficina donde se registra la fiscalización	Integer
Agno	Año del registro de la fiscalización	Integer
NroComision	Identificador de la fiscalización	Integer
TotalAfectados	Número total de afectados por la fiscalización	Integer
Urgencia	Si la fiscalización es urgente o no	Factor (2)
SolEsAfectado	Si el solicitante es afectado o no	Factor (2)
FechaRegistro	Fecha y hora de registro de la fiscalización	Character
CodTipoSol	Código del tipo de solicitante	Integer
solicitante	Tipo de Solicitante	Factor (14)
CodUnidadOrigen	Código de la Unidad de Origen de la fiscalización	Integer
unidadOrigen	Unidad de Origen de la fiscalización	Factor (3)
CodTipoTermino	Código del tipo de término de la fiscalización	Integer
tipoTermino	Tipo de Término de la fiscalización	Factor (8)
EgresoConMulta	Si la fiscalización terminó con multa o no	Factor (2)
RutEmpresa	RUT de la Empresa	Integer
EmpDFCodComuna	Código de Comuna de Domicilio Fiscalizado	Integer
EmpDMCodComuna	Código de Comuna de Domicilio Casa Matriz	Integer
EmpTrabHombres	Número de Trabajadores de la empresa	Integer
CodCae	Código de Actividad Económica	Integer
CodTipoEmpresa	Código de Tipo de Empresa	Integer
tieneMulta	Si la empresa tiene multa o no	Factor (2)
ccae	Código Interno de Actividad Económica	Integer
gcae	Glosa Interna de Actividad Económica	Factor (710)
crae	Código de rama de Actividad Económica	Integer
grae	Glosa Interna de rama de Actividad Económica	Factor (19)
nolnfra	Cantidad de materias en que no se detecta infracción	Integer
Infra	Cantidad de materias en que se detecta infracción	Integer

derechoFund	Cantidad de materias relacionadas a derechos fundamentales	Integer
novalnfra	Cantidad de materias en que no se genera Infraccionalidad	Integer
num_materias	Cantidad de materias revisadas en la fiscalización	Integer
num_sind	Cantidad de sindicatos de la empresa	Integer
grupocodtipoMaterias	Código materias fiscalizadas, separadas por coma.	Character
grupoglosatipoMaterias	Glosa materias fiscalizadas, separadas por coma.	Character
grupocodigolnfra	Código materias infraccionadas, separadas por coma.	Character
grupoglosalnfra	Glosa materias infraccionadas, separadas por coma.	Character

5.4.2.3 Exploración de los Datos

Para la exploración de los datos, se realiza un resumen de las variables en R, donde se excluyen las variables asociadas a códigos, debido a que no presentan mayor información, además se separan en aquellas variables que son numéricas y categóricas, donde las numéricas se analiza en base a estadísticas básicas y las categóricas en gráficos de frecuencia.

Tabla 10: Estadísticas de variables numéricas, datos Dirección del Trabajo.

Variable	Min	Max	Promedio	Mediana	Desviación
TotalAfectados	1	667	173.7	100	205.47
noInfra	0	26	2.96	3	2.1
Infra	0	17	0.64	0	1.17
derechoFund	0	1	0.000022	0	0.0014
novalnfra	0	11	0.005	0	0.1
num_materias	1	29	3.614	3	2.289
num_sind	0	303	5.577	0	20.94
EmpTrabHombres	1	3264	1282	1077	1029.393

Dada la alta desviación de la variable EmpTrabHombres, se realiza un histograma para observar la distribución de la variable, acompañándola con el logaritmo de la variable a modo de comparación.

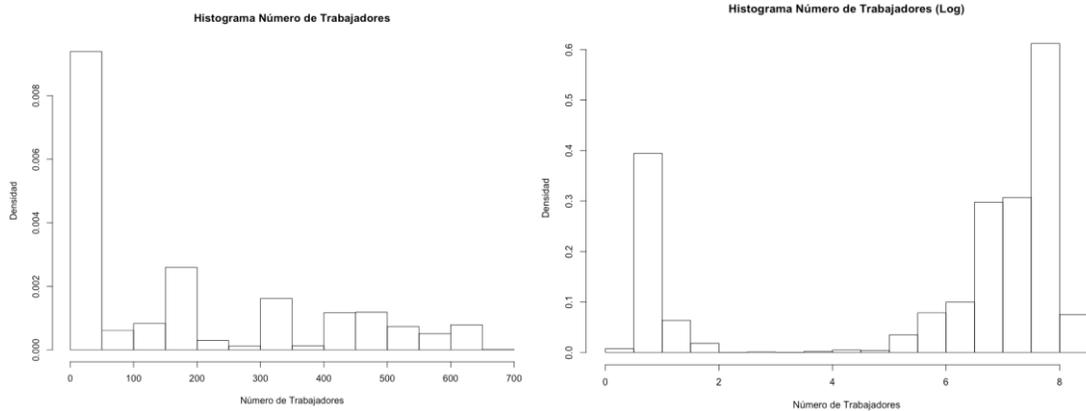


Figura 23: Histogramas de la variable Número de Trabajadores, Dirección del Trabajo.
Fuente Elaboración Propia.

Se observa que ambos gráficos no poseen una estructura saludable para el análisis, dado que según discusiones con los fiscalizadores, al momento de ingresar el dato, en ocasiones ingresan el número de personas entrevistadas, o el número de personas en el domicilio fiscalizado el cual puede diferir del número total de trabajadores de la empresa. Dado esto, se decide analizar el número de trabajadores del servicio de impuestos internos, donde se realizan los mismos gráficos.

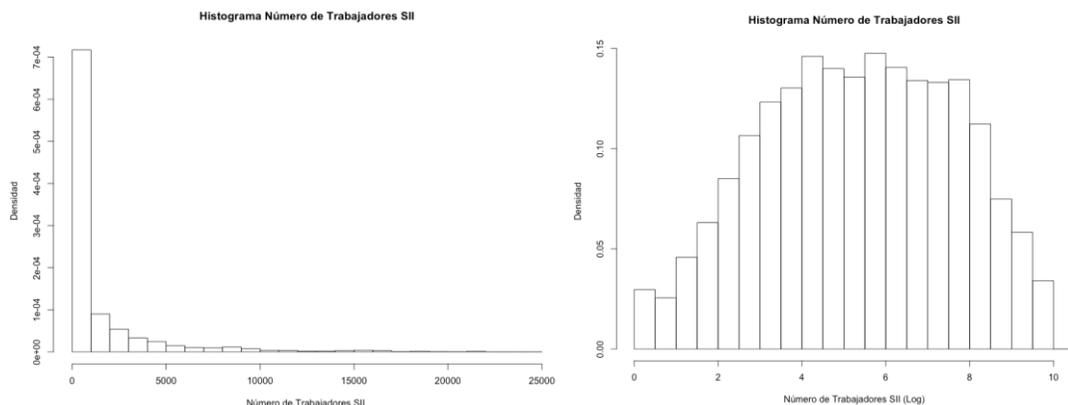


Figura 24: Histogramas de la variable Número de Trabajadores, SII.
Fuente Elaboración Propia.

Se puede apreciar que el dato entregado por el SII presenta una distribución cercana a una normal, lo cual la hace más efectiva al momento de aplicar modelos estadísticos.

Para las variables categóricas, se describen a partir de gráficos de frecuencias e infraccionalidad asociada a cada una de las categorías, donde se utilizan las variables Rama de Actividad Económica, Región y Tipo de solicitante .

En la Figura 25 se observa la frecuencia de la variable Rama de actividad económica y su respectiva infraccionalidad por categoría, donde se destacan las

actividades: Suministro de electricidad, gas y agua, Explotación de minas y canteras, Hogares privados con servicio doméstico, Industrias manufactureras y Hoteles y Restaurantes como las actividades con la mayor tasa de infraccionalidad detectada. Por otro lado, se tienen actividades cuya cantidad de fiscalizaciones es muy elevada pero su infraccionalidad detectada se encuentra cercana o baja al promedio tales como: Actividades Inmobiliarias, empresariales y alquiler, Comercio y Transporte, almacenamiento y comunicaciones. Por último, se ubican las actividades con menor infraccionalidad, donde se encuentran: Administración pública y defensa, Organizaciones y órganos extraterritoriales y Pesca. Se destaca que las organizaciones públicas no se rigen bajo el código laboral como todo el resto, sino que bajo Estatuto Administrativo, el cual es fiscalizado por la Contraloría General de la República.

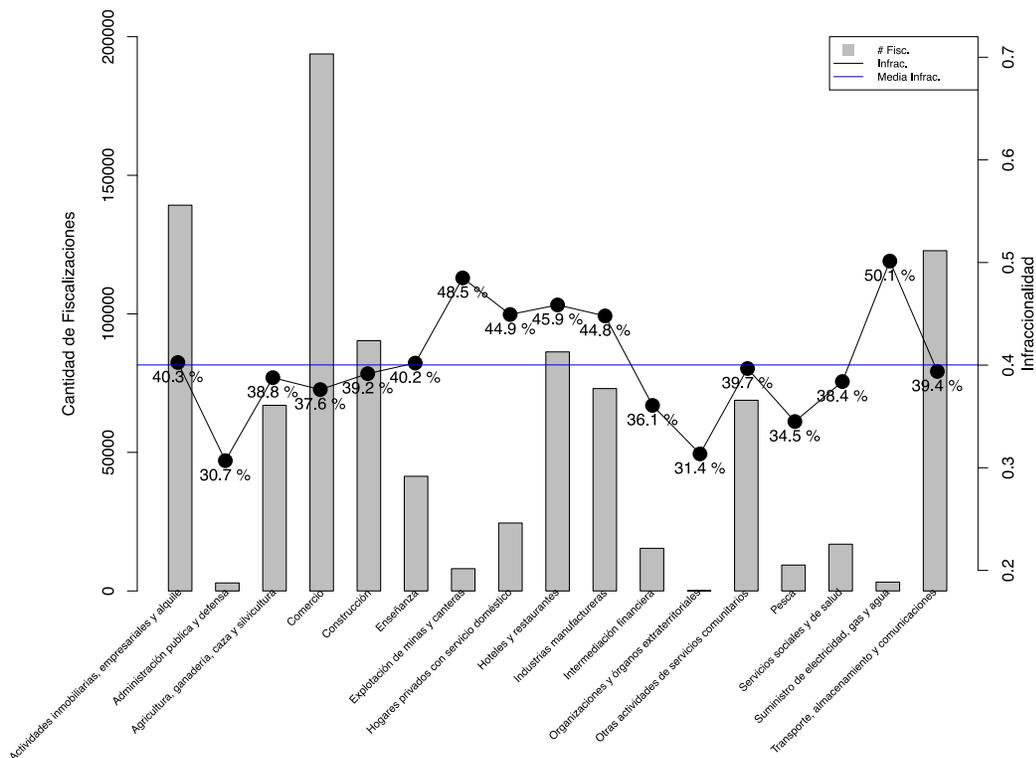


Figura 25: Gráfico de frecuencia e Infraccionalidad según Actividad Económica.

Fuente Elaboración Propia.

En la Figura 26 se muestra la frecuencia según el Tipo de solicitante y la respectiva Infraccionalidad. Se destaca, en primer lugar, que aquellas fiscalizaciones realizadas por Programa poseen un 26,8% de detección de infraccionalidad, donde se confirma la problemática abordada por el proyecto de tesis. Las fiscalizaciones reactivas son aquellas cuyo tipo de solicitante puede ser: Trabajadores, Terceros, No se Identifica y Organizaciones Sindicales, donde se presentan las tasas más altas de detección.

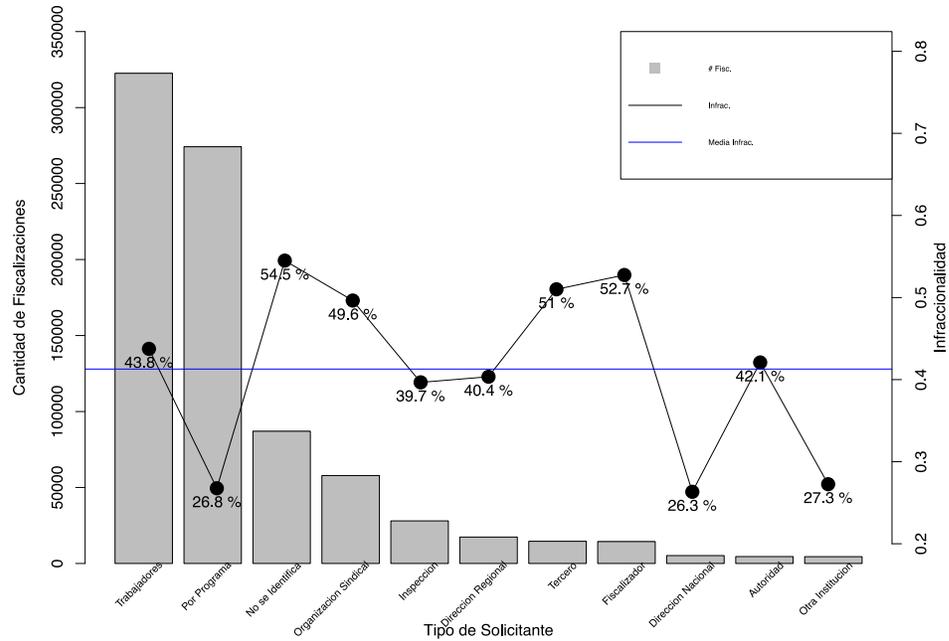


Figura 26: Gráfico de frecuencia e Infracionalidad según Tipo de Solicitante.

Fuente Elaboración Propia.

En la Figura 27 se observa la frecuencia de fiscalizaciones según región y su respectiva infraccionalidad. Se destaca que la cantidad de fiscalizaciones en la región metropolitana es la más alta, mientras que su tasa de detección se encuentra bajo el promedio. Las regiones 7, 9 y 10 poseen de las detecciones más altas llegando a infraccionalidad sobre el 50%, mientras que la región 12 posee la infraccionalidad más baja, con un 21%.

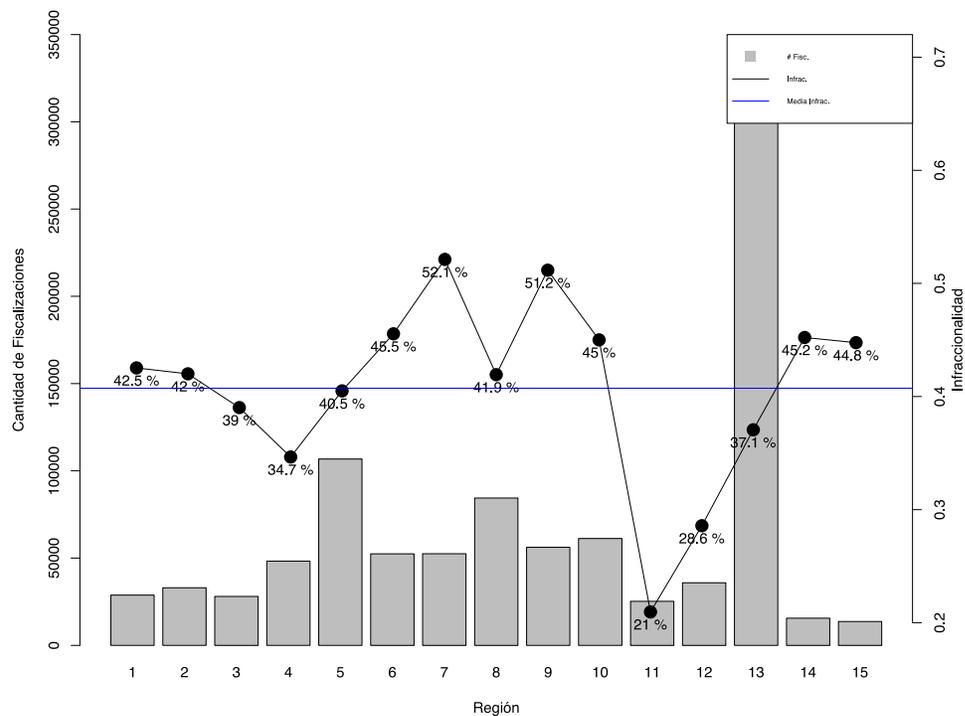


Figura 27: Gráfico de frecuencia e Infracionalidad según Región del País.

Fuente Elaboración Propia.

5.4.2.4 Verificación de calidad de los Datos

Para verificar si los datos no poseen muchos registros perdidos, se realiza una tabla de las variables con el porcentaje de datos perdidos.

Tabla 11: Porcentaje de Valores Nulos en la base de datos.

Variable	% de NA's
CodOficina	0
Agno	0
NroComision	0
TotalAfectados	13,27
Urgencia	6,1
SolEsAfectado	41,57
FechaRegistro	0
CodTipoSol	0
solicitante	0
CodUnidadOrigen	0

unidadOrigen	0
CodTipoTermino	0
tipoTermino	0
EgresoConMulta	0
RutEmpresa	0,98
EmpDFCodComuna	0
EmpDMCodComuna	49,87
EmpTrabHombres	4,49
CodCae	0
CodTipoEmpresa	26,30
tieneMulta	0
ccae	0,07
gcae	0,07
crae	0,07
grae	0,07
noInfra	0
Infra	0
derechoFund	0
novalnfra	0
num_materias	0
num_sind	0
grupocodtipoMaterias	0
grupoglosatipoMaterias	0
grupocodigolnfra	0
grupoglosalnfra	0
actividadeconomicaSII	35,96
agnoSII	35,96
bloqueSII	96,55
calleSII	35,96
comunaSII	35,96

deptoSII	84,73
dvSII	35,96
f22c645SII	51,78
f22c646SII	92,21
fechainicioSII	36,96
fechaterminoSII	95,88
ntrabajadoresSII	39,88
numeroSII	36,71
razonsocialSII	35,96
regionSII	35,96
rubroSII	35,96
rutSII	35,96
subrubroSII	35,96
subtipocontribuyenteSII	35,96
tipocontribuyenteSII	35,96
tipotermينوSII	95,88
tramoventasSII	35,96
villapoblacionSII	86,24

Se puede observar que los registros obtenidos por el SII presentan un alto número de valores nulos. Por otro lado, de los datos obtenidos en la organización, las variables SolEsAfectado, TotalAfectados y EmpDMCodComuna presentan el mayor porcentaje de valores perdidos en la base.

5.4.3 Preparación de los Datos

5.4.3.1 Selección de Atributos

Dada los resultados de la fase previa, Entendimiento de los Datos, se puede observar que una cantidad considerable de variables del SII no son fidedignas debido a la alta cantidad de valores nulos. De esta manera, y de acuerdo a los objetivos del proyecto y necesidades de negocio se dejan para el análisis las variables: **ntrabajadoresSII** y **tramoventasSII**, dejando fuera del análisis el resto de las variables por su baja cantidad de datos.

Por otro lado, dentro de los datos propios de la DT se realizan las siguientes selecciones de datos:

- CodOficina

- Solicitante y CodTipoSol
- CodUnidadOrigen y unidadOrigen
- CodTipoTermino y TipoTermino
- Fecharegistro
- RUT Empresa
- Ccae y gcae
- Crae y grae
- noInfra
- Infra
- Num_materias
- Num_sind
- grupoCodigoMaterias
- grupoglosatipoMaterias
- grupocodigInfra
- grupoglosaInfra

Además, la variable solicitante posee 2 categorías que no están alineadas al análisis predictivo debido a que son inicializadas con objetivos que no buscan infracción, si no otro tipo. Estas categorías son: **Institución Previsional** y **Empleador**, por ende, los registros que posean esta categoría en la variable son removidos del análisis.

Con esta selección, se queda con una base de datos de 21 variables y 830.366 observaciones.

5.4.3.2 Limpieza de Datos

Dada la selección de atributos, la nueva base posee solo los valores nulos de las variables del SII, por lo que se decide sacar del análisis aquellas filas que no posean tales valores, dado que, al ser una base externa, las técnicas de imputación de datos en base a similaridad de casos no es una técnica fiable.

5.4.3.3 Construcción, Integración y Formateo de Datos

Para el análisis predictivo de la primera iteración, se requiere de una variable dependiente, la cual sea el objeto de predicción del modelo. Para esto se crea una variable la cual se llama **Infractor**. Esta variable se construye a partir de la variable **Infra**, donde la nueva variable toma el valor 1 si **Infra** es mayor o igual a 1 y 0 si es igual a 0. Con esto la variable **Infractor** es 1 si dentro de la fiscalización hubo una o más materias en las que se encontró infractor y 0 si no hubo ninguna, siendo una variable binaria, en resumen, de si fue infractor o no.

Se crea una variable binaria que detalle si existe o no sindicato en la empresa, debido a que la variable original **num_sind** no posee datos fidedignos, según conversaciones con fiscalizadores, debido a que no se actualiza cuando estos sindicatos dejan de existir, por lo cual se decide generar una variable que posea la información de que si existe o no. Para ello, se realiza la misma operación que

la variable Infractor, donde la nueva variable **Exsind** toma el valor 1 si **num_sind** es mayor o igual a 1, y 0 en caso contrario.

También se crea la variable **Mes**, a partir de la fecha de registro, la cual se transforma al formato "Date" de la herramienta R, donde se puede extraer, en base a la función *month(date)* el mes directamente, sin realizar cálculos intermedios.

Además, se crean variables de comportamiento **Recency**, **NumFisc** y **NumInf**.

Recency: La variable recency se crea a partir de la variable **Fecharegistro**, donde recency es la diferencia entre la fecha de la observación de un RUT determinado menos la fecha de la última fiscalización del mismo RUT. Esta variable esta medida en días y si es la primera fecha toma el valor 0.

NumFisc: La variable NumFisc es la suma acumulada hasta la fecha de la cantidad de fiscalizaciones realizadas al RUT correspondiente. Si no hay fiscalizaciones pasadas, toma el valor 0.

NumInf: La variable NumInf es la suma acumulada de las fiscalizaciones donde se encontró infracción al RUT correspondiente. Si no hay infracción en fiscalizaciones pasadas, toma el valor 0.

Para la **segunda iteración**, la base debió ser transformada completamente debido a que la recopilación inicial de datos fue realizada por fiscalización, por ende, cada registro representa una fiscalización finalizada, dado esto, se agrupan las fiscalizaciones por la variable **RutEmpresa**, donde se crea, para cada año, la cantidad de fiscalizaciones e infracciones del respectivo RUT. Con esta lógica se crean las siguientes variables tipo **Agno_01** y **inf_06**, las cuales, para este caso, la variable Agno_01 representa la cantidad de fiscalizaciones que tuvo la empresa el año 2006 y la variable inf_06 es la cantidad de veces que fue descubierto infractor en alguna de las fiscalizaciones realizadas el mismo año, por consiguiente, se tiene para cada año el respectivo par de variables.

También se crean las siguientes variables de resumen:

Fiscmean: Promedio anual de fiscalizaciones.

Infmean: Promedio anual de fiscalizaciones infraccionadas.

Fiscsd: Desviación estándar anual de fiscalizaciones.

Infstd: Desviación estándar anual de fiscalizaciones infraccionadas.

Qfisc: Suma total de fiscalizaciones.

Qinf: Suma total de fiscalizaciones infraccionadas.

Recency: Años desde la última fiscalización (Medidos desde un año a fijar).

Infrac: Infraccionalidad de la empresa a lo largo de la historia, se calcula como *qinf* sobre *qfisc*.

En términos de la variable objetivo, debido a que la base ya no es por fiscalización si no por empresa, se debe crear una nueva variable que permita supervisar el análisis. Para esto, se crea una variable llamada **Infractor**, la cual toma el valor 1 si en el año N a evaluar la variable **inf_ON**, sea mayor o igual a 0, lo que indica que la empresa, en tal año, fue infractora al menos una vez, y toma el valor 0 si no.

5.4.4 Modelamiento

5.4.4.1 Selección de Técnicas de Modelamiento

Tanto para la primera iteración, la cual es el estudio predictivo de la base desagregada por fiscalizaciones, como la segunda iteración, la cual se basa en la predicción basada en la agregación de las fiscalizaciones por RUT de empresa fiscalizada, se realizan pruebas con los modelos (revisar capítulo 2): Regresión Logística, Árboles de Decisión (Random Forest, Conditional Inference Trees, CART) y Redes Neuronales (Multilayer Perceptron). Es importante señalar que para algunos de estos modelos se debe tener datos completos (no contener datos nulos), por lo cual la base a entrenar debe estar limpia de valores nulos. Por otro lado, dada la herramienta a utilizar para el modelamiento, ciertos algoritmos permiten el ingreso de las variables categóricas en su forma común, donde cada observación posee el valor de una categoría, por otro lado, algoritmos como las redes neuronales requieren solo datos numéricos, por lo cual se requiere de una transformación de las variables categóricas a utilizar a sus respectivas variables *dummy* o binarias por cada categoría.

Es necesario destacar que, dado que es una predicción, las variables a utilizar como explicativas, deben ser conocidas de manera ex ante a la fiscalización, donde variables como el Total de Afectados, no entra a este grupo, debido a que es una variable que se obtiene posterior a la fiscalización en terreno.

Por consiguiente, las variables explicativas a utilizar en la primera iteración serán: **Exsind, Region, grae, ntrabajadoresSII, numfisc, numinf, recency y tramodeventas**. La variable a predecir será: **Infractor**, que toma valores 0 y 1, previamente explicados.

Para la segunda iteración, las variables explicativas serán: Para todos los años las respectivas variables **Agno_N, inf_N, Region, grae, fiscmean, infmean, fiscsd, infsd, qfisc, qinf, recency** y por último **Infrac**. Para la variable objetivo a predecir se utiliza **Infractor**, explicada en apartados anteriores.

5.4.4.2 Generar Diseño de Prueba

Para la prueba de los modelos de la iteración uno, se separa la base de datos en un set de prueba y un set de testeo. El set de entrenamiento posee una muestra del 75% de los datos y el set de testeo el 25% restante.

Para la segunda iteración, el diseño se basa en la separación de la base de datos en el set de prueba donde se considera la historia de las empresas entre el 2006 y el 2014 y el set de testeo toma la historia del año 2015 y 2016.

Dado que los modelos a probar son supervisados el rendimiento de estos se mide con el análisis de la matriz de confusión y sus respectivos indicadores tales como: *Recall*, *Precision*, *Specificity*, *Accuracy*, *Fscore*, entre otros.

El método de selección de los modelos se basa principalmente en el balance que existe entre los indicadores *Precision* y *Recall*, donde el indicador *Fscore* es el promedio armónico entre ambos. Con esto presente, un modelo aceptable para el análisis es aquel que tenga el mejor *Fscore*.

5.4.4.3 Construcción de Modelos

Para la construcción de los modelos señalados se utiliza el software R con una serie de paquetes que facilitan el trabajo de las fases del CRISP-DM. Los modelos para la **segunda iteración** con la base agregada por empresas se señalan a continuación.

5.4.4.3.1 Classification and Regression Trees

Para la construcción de árboles de decisión tipo CART, se utiliza paquete “rpart” el cual posee el método “class” para la clasificación y el método “anova” para la regresión. Además, posee parámetros de control de crecimiento del árbol tales como el parámetro “cp” (*complexity parameter*) el cual es una medida tipo corte que evita que el árbol siga creciendo si es que no se mejora el error, “minsplit” el cual fija el número mínimo de observaciones que debe tener un nodo para intentar una separación, “xval” el cual es el número de validaciones a realizar, entre otros. Permite tanto variables categóricas en su forma común, variables numéricas y valores nulos entre los datos.

Para la iteración 2, los parámetros fijados del árbol son: Método “anova”, Complexity parameter 0.001 y 10 validaciones cruzadas. Se utiliza el método “anova” debido a que se desea encontrar el mejor corte de balance entre *precision* y *recall*. Con este fin, se modela con los parámetros previos y se evalúan los indicadores de matriz de confusión para diversos cortes. La Figura 28 muestra los diferentes indicadores señalados previamente, donde se puede ver que las curvas Recall y Precision se cruzan en el corte 0.125.

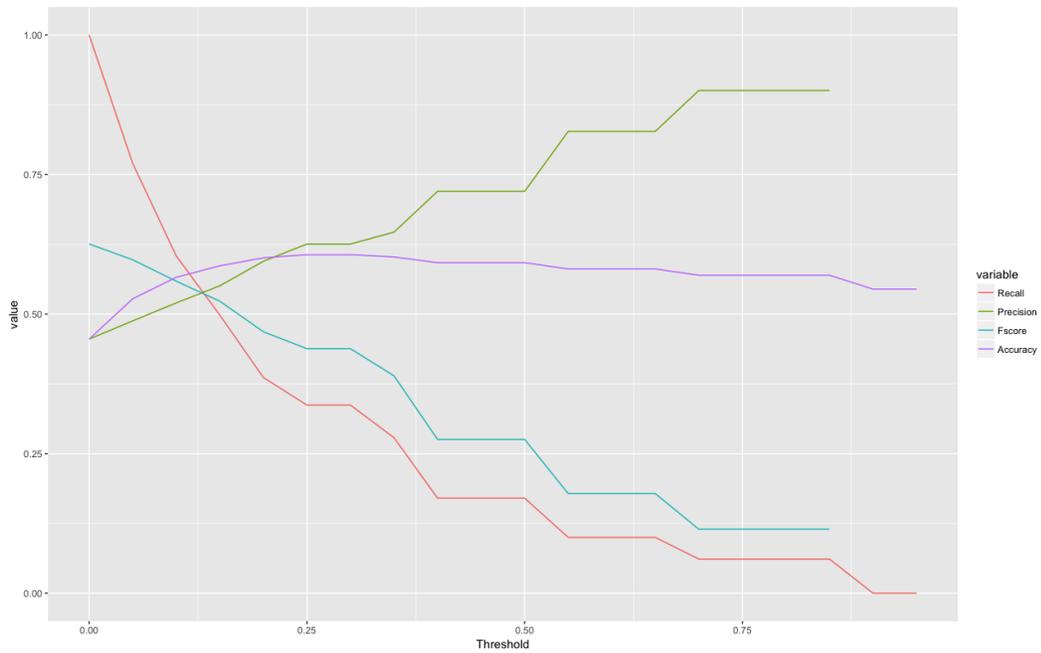


Figura 28: Indicadores de Predicción versus Cortes de selección, CART.

Fuente Elaboración Propia.

Con este corte, los valores para los indicadores son los siguientes:

Tabla 12: Indicadores para corte seleccionado, CART.

Indicador	Valor
Accuracy	60.63%
Fscore	43.80%
Recall	33.70%
Precision	62.54%
Specificity	83.14%

5.4.4.3.2 Conditional Inference Trees

Para los Árboles de Inferencia Condicional se utiliza el paquete “party” el cual permite realizar un árbol regresivo. Este paquete permite parametrizar el árbol en medidas como el tipo de test a realizar para la selección de atributos donde permite Test estadísticos, test Montecarlo, entre otros, donde permite fijar el valor del estadístico p-valor para la implementación de una separación. Al igual que para “rpart” permite fijar el valor de observaciones en un nodo intentar separaciones, permite variables categóricas en su forma común y valores nulos entre los datos.

Los parámetros definidos para la Iteración 2 son: test de tipo estadístico donde se utiliza el p-valor como criterio, donde se especifica que el test debe tener un p valor menor al 5% (mincriterion = 0.95 en la fórmula). Nuevamente, se entregan los resultados en forma de probabilidad, donde se busca, al igual que el modelo

CART, el corte óptimo para el modelo. Se observa en la Figura 29 el gráfico de los indicadores respecto a una serie de cortes, donde es similar al corte obtenido con el modelo CART, con un corte 0.25 para el cruce entre precision y recall, por ende, tanto los criterios de información como criterios estadísticos se comportan de similar medida con el set de datos.

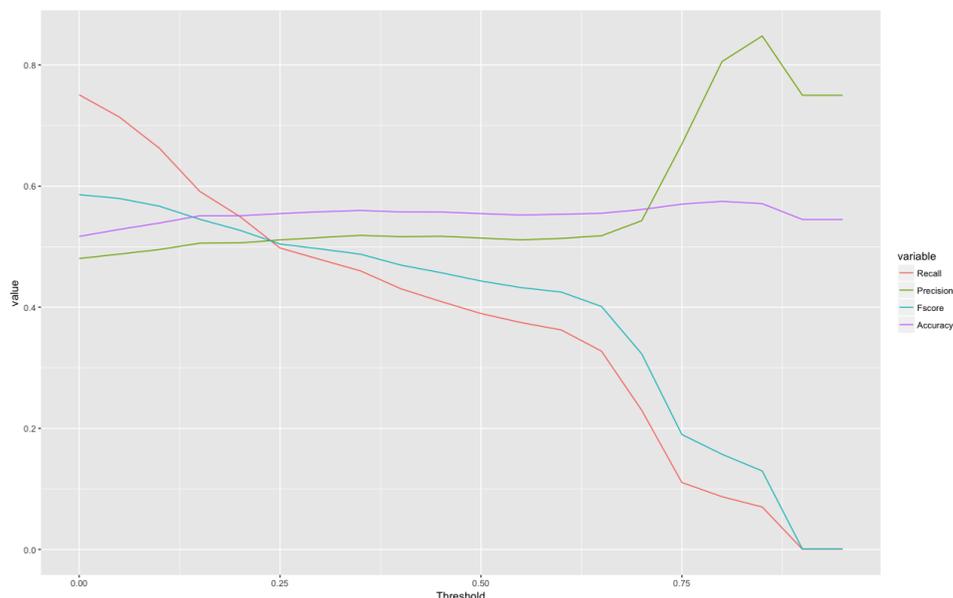


Figura 29: Indicadores de Predicción versus Cortes de selección, Conditional Inference Trees.

Fuente Elaboración Propia.

Con el corte seleccionado se obtienen los siguientes resultados:

Tabla 13: Indicadores para corte seleccionado, CIT.

Indicador	Valor
Accuracy	55.47%
Fscore	50.43%
Recall	49.78%
Precision	51.11%
Specificity	60.22%

5.4.4.3.3 Random Forest

Para el tipo de modelo de Árboles, Bosques Aleatorios se utiliza el paquete “randomForest”, donde se busca mejorar los resultados obtenidos por los árboles de decisión. Este paquete permite parametrizar la cantidad de árboles a crear, la cantidad de variables a ir seleccionando aleatoriamente entre los diversos árboles, permite conservar los resultados de las variables más importantes en términos de resultados del modelo, y los parámetros clásicos de los árboles mencionados previamente.

Los parámetros seleccionados para el modelo son 100 árboles a crear, donde puede tomar las 20 variables o menos para la selección aleatoria con reemplazo. También se conservan los valores de la importancia de cada variable en los respectivos árboles.

Con esta parametrización se entrena el modelo y al predecir en el set de prueba, se obtienen valores de probabilidad de infracción, por lo cual se realiza el mismo gráfico obtenido en los árboles previos. Se puede observar en la Figura 30 que el *Recall* con el *Precision* se juntan en un corte cercano al 0.12, pero el *Fscore* y *Accuracy* presentan un máximo, el cual se ubica en el corte cercano al 0.25, por lo cual se define como corte de selección.

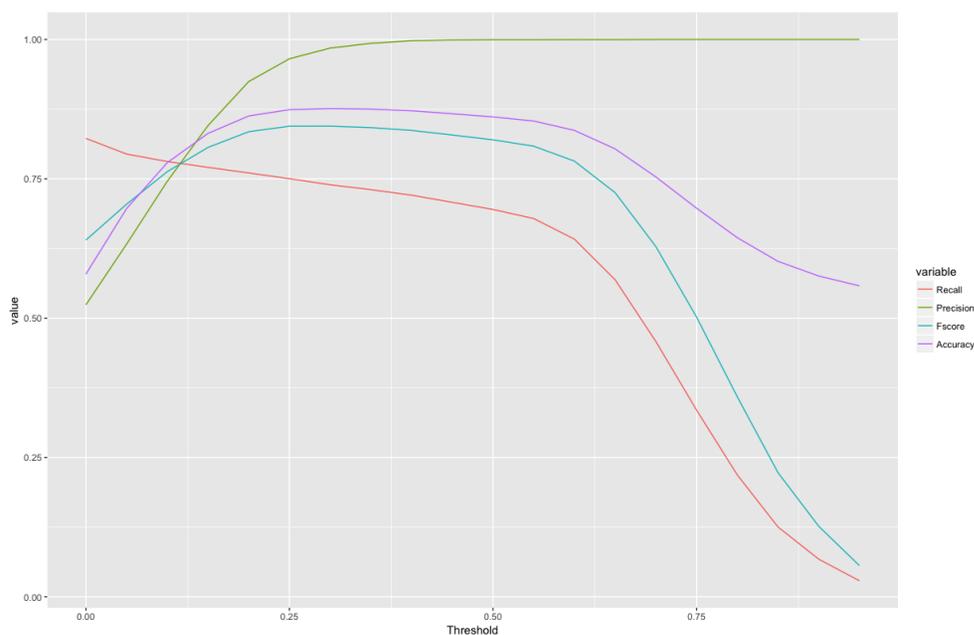


Figura 30: Indicadores de Predicción versus Cortes de selección, Random Forest.

Fuente Elaboración Propia.

Los resultados del modelo son los siguientes:

Tabla 14: Indicadores para corte seleccionado, Ranfom Forest.

Indicador	Valor
Accuracy	87.41%
Fscore	84.43%
Recall	75.03%
Precision	96.53%
Specificity	97.75%

5.4.4.3.4 Regresión Logística

El modelo de regresión logística se realiza a través del paquete básico de R, donde la función para este tipo de modelos es “glm” por sus siglas en inglés

Generalized Linear Models. Los parámetros disponibles para su modificación son la cantidad máxima de iteraciones que el modelo puede realizar para su convergencia, además del índice de tolerancia de esta (epsilon). Además, se puede cambiar el tipo de función a utilizar, donde la función Logit es una de estas (existen otras como probit o cauchy) y la diferencia entre las distintas funciones es la distribución de sus errores.

Finalmente, los parámetros utilizados para la regresión logística son la utilización de la función logit, una cantidad máxima de iteraciones de 1000 y un epsilon de 10^{-14} . Con estos parámetros se predice el set de prueba, donde se entregan resultados de probabilidad de infraccionalidad como los modelos previos detallados, donde se realiza el gráfico de corte para la selección del óptimo. En la Figura 31 se puede observar que el corte se ubica cercano al 0.05, pero existe un máximo de Accuracy cercano al 0.125.

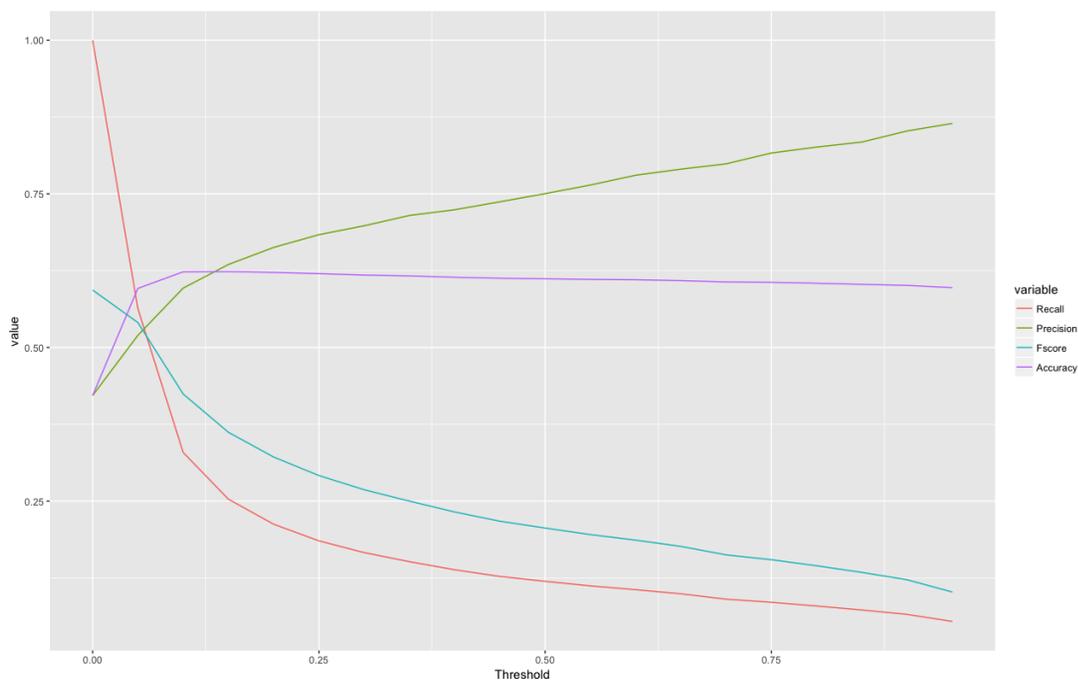


Figura 31: Indicadores de Predicción versus Cortes de selección, Regresión Logística.

Fuente Elaboración Propia.

Los resultados del modelo son los siguientes:

Tabla 15: Indicadores para corte seleccionado, Logit.

Indicador	Valor
Accuracy	62.31%
Fscore	42.43%
Recall	32.93%
Precision	59.66%
Specificity	83.75%

5.4.4.3.5 Red Neuronal (Multilayer Perceptron)

Para el modelo de redes neuronales se utilizó el paquete “mlp”, el cual permite parametrizar los valores de cantidad de neuronas por capa, cantidad de capas, ratio de aprendizaje (*Learning Rate*), cantidad de iteraciones para el aprendizaje, diferentes funciones de activación, entre otros.

Debido a los altos tiempos de procesamiento, una parametrización base para la red neuronal de 20 neuronas con una única capa, una cantidad de iteraciones de 10000. Para el corte óptimo se realiza el mismo gráfico de los modelos anteriores, el cual se puede observar en la Figura 32, este corte se encuentra cercano al 0.8, donde se aplica la selección de infractores.

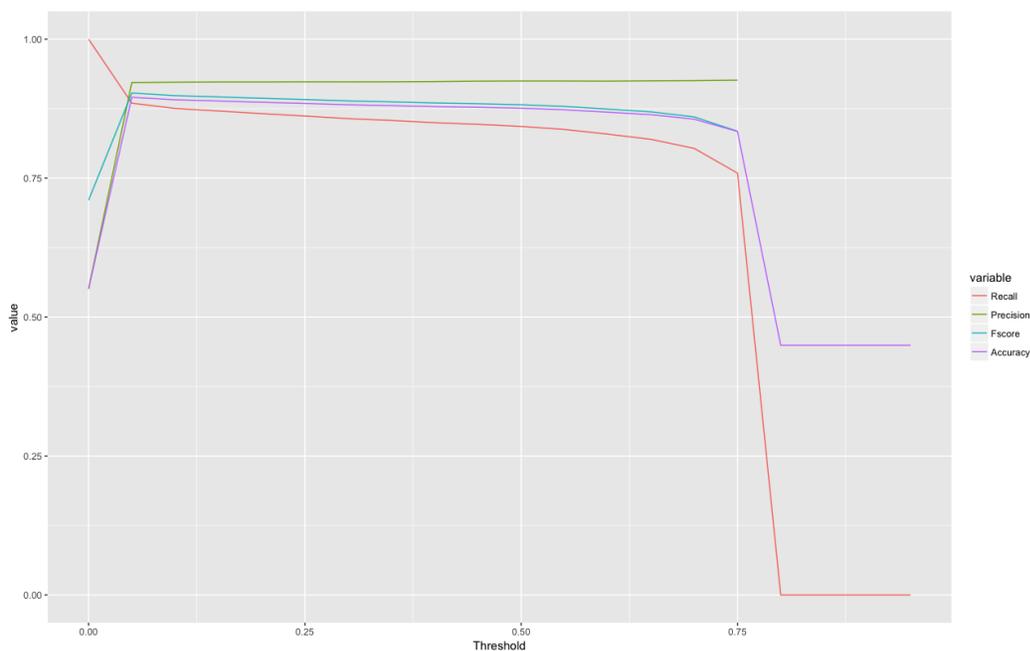


Figura 32: Indicadores de Predicción versus Cortes de selección, Red Neuronal.

Fuente Elaboración Propia.

Los resultados del modelo son los siguientes:

Tabla 16: Indicadores para corte seleccionado, Red Neuronal.

Indicador	Valor
Accuracy	83.40%
Fscore	83.43%
Recall	75.90%
Precision	92.62%
Specificity	92.58%

5.4.4.4 Evaluación de Modelos

En la Tabla 17 se observa el resumen de los resultados en la base de testeo de los modelos, donde los mejores resultados fueron obtenidos por Random Forest y por la Red Neuronal (MultiLayer Perceptron), donde se obtuvieron resultados sobre el 80% de *Accuracy*. Si bien son los mejores resultados, ambos modelos son poco interpretables respecto a las reglas tomadas para generar la predicción.

Tabla 17: Cuadro resumen de indicadores según modelo.

Modelo/KPI	Accuracy	Fscore	Recall	Precision	Specificity
Rpart	60.63%	43.80%	33.70%	62.54%	83.14%
CIT	55.47%	50.43%	49.78%	51.11%	60.22%
RandomForest	87.41%	84.43%	75.03%	96.53%	97.75%
Logit	62.31%	42.43%	32.93%	59.66%	83.75%
MLP	83.40%	83.43%	75.90%	92.62%	92.58%

Se presenta una evaluación de los resultados de acuerdo a los distintos modelos utilizados.

Para los árboles de decisión, tanto **CART** como **CIT** entregan resultados interpretables de acuerdo a los cortes realizados de acuerdo al algoritmo respectivo. Debido a la complejidad y el número de cortes realizados de los árboles, no se puede analizar gráficamente el árbol respectivo, pero se aplica un parámetro de profundidad para poder visualizar como quedaría un árbol con una menor cantidad de cortes. En la Figura 33 se observa un árbol con una profundidad máxima de 7 nodos, donde los principales cortes realizados son basados en la infraccionalidad promedio histórica y si ha sido fiscalizadas en los 2 años previos al año objetivo. Es importante destacar que las características propias de la empresa no se dan a luz a esta profundidad del árbol, donde ni la Región ni la Actividad Económica aparecen como cortes.

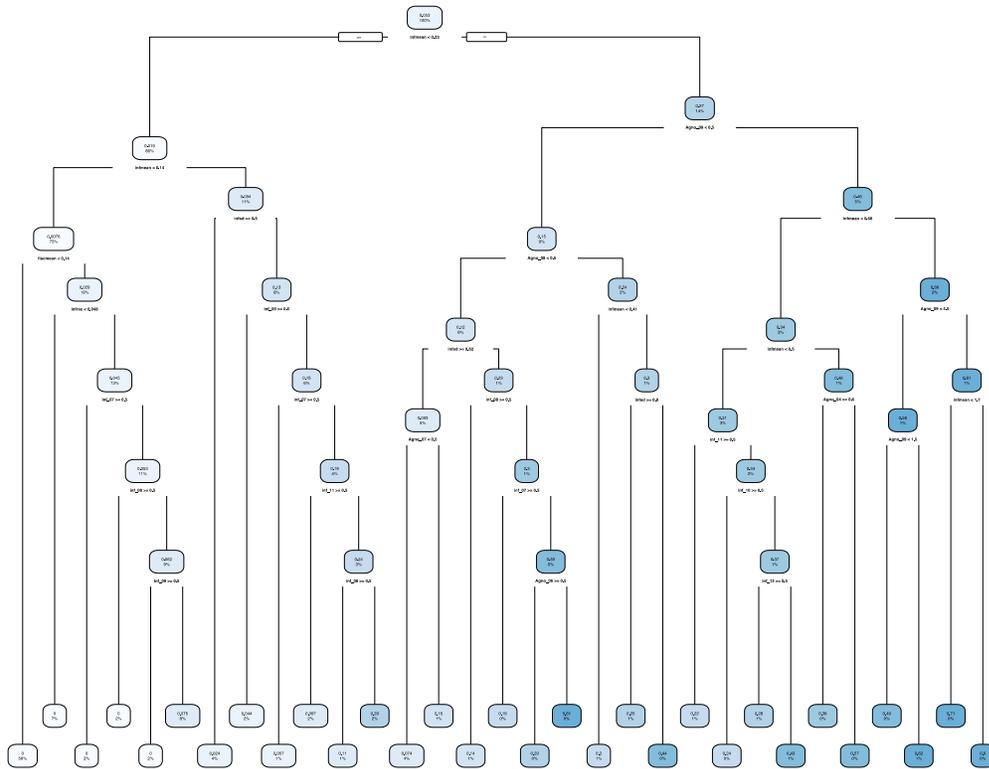


Figura 33: Árbol de decisión, CART con profundidad máxima de 7 nodos.

Fuente Elaboración Propia.

Para el **Random Forest** no es posible realizar una representación gráfica debido a que es un conjunto de árboles, pero es posible analizar la importancia de las variables en términos de información (Índice de Gini) y de *accuracy* de las variables explicativas, donde en la Figura 34 se puede ver el efecto de eliminar la variable respecto a los índices indicados en el eje horizontal de ambos gráficos. Las variables más importantes en términos de información son la infraccionalidad histórica de las empresas (previas a los años 2015 y 2016 en este caso), la actividad económica de la empresa y su Región, seguido posteriormente de otros indicadores como el promedio de fiscalizaciones. Por otro lado, en términos del *Accuracy* del modelo, las variables más importantes son la infraccionalidad histórica y la cantidad de fiscalizaciones realizadas en años anteriores. Esto destaca la importancia del comportamiento en el análisis realizado, donde se infiere que aquellas empresas que han sido infractoras en el tiempo, lo seguirán siendo en el futuro. Este modelo entrega un *Accuracy* y *Fscore* sobre el 80%, entregando el mejor resultado de los modelos.

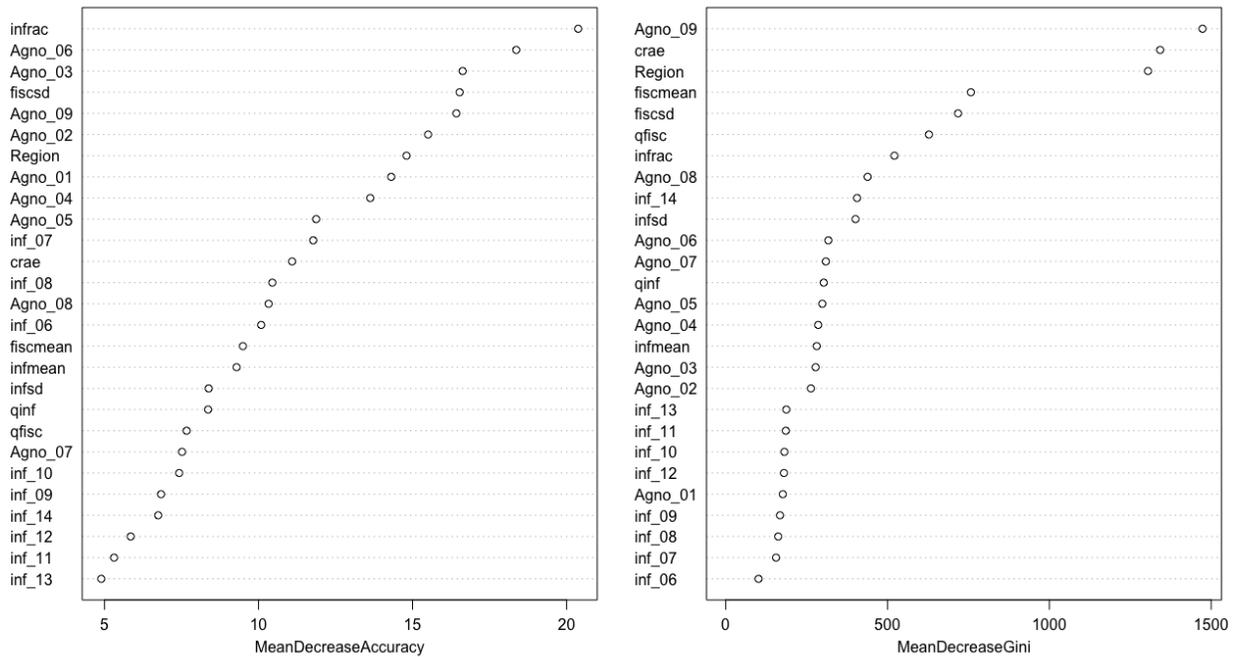


Figura 34: Importancia de Variables explicativas, Random Forest.

Fuente Elaboración Propia.

Para la **Regresión Logística** se revisan los coeficientes para cada una de las variables (ver Anexos, Figura 51) donde nuevamente las variables más importantes son de la cantidad de infracciones en años pasados, coherente con el resto de los modelos. Además, los coeficientes relacionados con la región y la actividad económica no son significativos, nuevamente confirmando análisis posteriores. Esta técnica entrega un *Accuracy* cercano al 65%, siendo el tercer mejor modelo bajo los indicadores.

Por último, las **redes neuronales** no son directamente interpretables dado el algoritmo de aprendizaje que utilizan, siendo este uno de los grandes contras que tiene al ser utilizado, sin embargo, entrega resultados altamente cercanos a la realidad. En particular, es el segundo mejor modelo desarrollado en el proyecto de tesis, pero es poco interpretable.

A modo de resumen, el Random Forest entrega el mejor resultado en base a que entrega *accuracy* más alto y es más fácil de interpretar que una red neuronal. Se decidió utilizar esta técnica para el desarrollo del piloto presentado en capítulos posteriores.

5.4.5 Evaluación

5.4.5.1 Evaluación de Resultados

Debido al tipo de modelo que se utiliza y el tipo de problema que se enfrenta, la interpretación de resultados es relativamente sencilla, donde en una predicción binaria solo existen 4 tipos de resultados, el predecir bien positivos y negativos y los tipos de error I y II, de acuerdo a la predicción.

Con los resultados del modelo se puede deducir que el comportamiento de las empresas puede ser predicho, pero en gran medida en base a su comportamiento, por lo cual empresas que nunca han sido fiscalizadas generan un problema al incorporar este comportamiento, el cual no poseen, a los modelos. Por otro lado, un descubrimiento importante es que el análisis debiese ser realizado con la información agregada y no por fiscalización, lo cual es una tarea extra al momento de realizar el modelamiento y la respectiva preparación de datos.

5.4.5.2 Revisión del Proceso

Cada fase realizada fue de vital importancia para el transcurso del proyecto, donde el entendimiento del negocio y el entendimiento de los datos fueron posibles gracias a la alta participación de profesionales de la organización a lo largo del trabajo.

En las fases de preparación de los datos y el modelamiento se utilizó la gran parte del tiempo, dado que hubo una gran cantidad de iteraciones y re-evaluaciones de las decisiones tomadas debido a resultados bajos. Se realizaron frecuentemente reuniones con profesionales en la medida que se iban obteniendo resultados tempranos con el fin de entender el porqué de estos resultados, y en caso de ser necesario, tomar medidas para una posible nueva iteración.

Respecto a las iteraciones, los resultados iniciales, al analizar los datos como fueron extraídos, a nivel de fiscalización, daban resultados no superiores al 55-60% de *Accuracy*, razón por la cual se agruparon en función de los RUT's de empresas, donde los modelos aumentaron los indicadores de validación.

Un aspecto importante a considerar en la revisión y modelamiento de análisis supervisado es la correcta selección y transformación de variables, debido a que se debe evitar utilizar variables posteriores a la fiscalización (para este caso), dado que deben ser predictores a la acción de fiscalizar y tener esa información, en el modelamiento presente, para utilizar el modelo entrenado. Esto, en diversas ocasiones, generó problemas durante el proceso evaluativo del modelo, dado que entregaban indicadores muy altos, donde ocurría que una variable explicativa posterior a la acción de fiscalizar comprometía los resultados.

En términos finales, los resultados obtenidos por los modelos CART, Conditional Inference Trees y regresión logística son similares, donde los que sobresalen y entregan los mejores resultados son los random forest y redes neuronales.

Por otro lado, el análisis constante de los datos produjo un alto conocimiento para los profesionales en términos de conocer de manera profunda los datos y sus problemas.

5.4.5.3 Determinación de siguientes pasos

El proceso realizado durante el proyecto constó de muchos cambios a través de las iteraciones de los modelos, lo cual abre la puerta a que se siga parametrizando y mejorando los resultados. Si bien existe esta oportunidad, es necesarios probar en terreno los resultados intermedios que presenten resultados decentes, con el fin de validar si se está apuntando a la dirección correcta o se debe modificar alguna decisión que se tomó o tomar decisiones nuevas en términos de negocios o de los datos. Esto con el fin de que al momento de desplegar esto a programas nacionales, el error sea el mínimo y que a futuro se siga desarrollando y modificando modelos.

Uno de los principales futuros trabajos a desarrollar en base a los modelos es la obtención de datos de caracterización de empresas, los cuales son limitados en los registros de la organización.

Además, se promueve la incorporación de datos de comportamiento de otros servicios públicos, como por ejemplo del poder judicial, donde se pueden revisar casos de las mismas empresas y sus respectivos resultados, lo cual puede entregar un insumo importante al momento de modelar y predecir si la empresa, dado que ha tenido un comportamiento pasado en la corte, pueda seguir presentando malas prácticas.

Por último, se desea incorporar variables macroeconómicas que permitan distinguir como, por ejemplo, movimientos de precios del cobre y si estos afectan a diversas actividades en términos laborales.

5.4.6 Despliegue

El despliegue e implementación de la obtención del análisis realizado se basa en la creación de los programas de fiscalización proactiva con el apoyo de modelos matemáticos, debido a que en la actualidad la tasa de asertividad es baja y la utilización eficiente de recursos es clave para una organización pública.

Este análisis se utilizará en los procesos explicados en apartados anteriores, donde se incorpora como una capacidad a los procesos de análisis del comportamiento laboral. Este proceso se realiza en la unidad de planificación de la organización, la cual genera los programas en la actualidad.

CAPÍTULO 6: PROPUESTA DE APOYO TECNOLÓGICO

6.1 Especificación de Requerimientos

El Proyecto está basado en el desarrollo de modelos de Minería de Datos, donde este proceso de modelamiento se integrará a la arquitectura tecnológica de la organización, la cual es una arquitectura de 3 capas desarrollada en .NET Framework 1.0 en su mayoría, debido a que poseen varios frameworks de trabajo, todos basados en Microsoft en distintas versiones. Se realizan los requerimientos en función de la nemotécnica FURPS por sus siglas en inglés.

6.1.1 *Requerimientos Funcionales*

Debido a que se desea incorporar el modelo predictivo a la arquitectura, los requerimientos funcionales están basados a la incorporación de un sistema virtual que posibilite el procesamiento realizado por el software utilizado para el modelamiento. Dicho esto, el sistema debe:

- Permitir el acceso y conexión a las bases de datos institucionales.
- Permitir el procesamiento de información para fines de modelamiento y reportería.
- Permitir el uso de modelos analíticos para predicción de infraccionalidad.
- Permitir la exportación de las bases de datos procesadas con el respectivo *scoring* del modelamiento.

6.1.2 *Requerimientos No Funcionales*

Según la nemotécnica FURPS, se analizan: Functionality (Funcionalidad), Usability (Usabilidad), Reliability (Fiabilidad), Performance (Desempeño) y Support (Soporte), la Tabla 18 muestra lo propuesto por el autor.

Tabla 18: Requerimientos para la arquitectura tecnológica del rediseño.

Functionality	Requerimientos Funcionales previamente señalados
Usability	Debe tener la misma facilidad de uso para el analista de Inteligencia de Negocios.
Reliability	El sistema debe tener un sistema que le permita recuperarse ante caídas en un tiempo prudente. No debe permitir pérdidas de datos.
Performance	El sistema debe correr en los computadores institucionales. La capacidad de procesamiento debe ser alta, de acuerdo a la cantidad de datos que se procesarán y los modelos analíticos que se utilizarán.

Support	Debe tener un soporte similar al existente, del cual está encargado el departamento de tecnología, el cual permita modificar el sistema dependiendo de las reglas que se apliquen a las empresas dependiendo de decisiones estratégicas.
---------	--

6.2 Arquitectura Tecnológica

Como se mencionó anteriormente, la arquitectura tecnológica de la DT está basada en una arquitectura de 3 capas desarrollada en frameworks de Microsoft, mayoritariamente .NET Framework 1.0, donde se separan las lógicas de negocio, datos y visualización. El lenguaje predominante es Visual Basic, con algunos scripts .ASP, share point, entre otros. El sistema posee funcionalidades tanto para la atención de usuarios como para la fiscalización, donde está unido la recepción de una denuncia a la asignación de un fiscalizador para su revisión y posterior trabajo investigativo o de campo. Además, la Dirección de trabajo posee la capacidad de generación de máquinas virtuales en caso de ser solicitadas.

6.3 Diseño del Sistema Virtual

Para el diseño del sistema virtual, se levanta una máquina virtual en Linux, a la cual se incorpora el software R y R Studio para el modelamiento analítico, y además, se incorpora el motor de bases de datos PostgreSQL.

6.3.1 Casos de Uso

Para el desarrollo del Proyecto, se detallan los casos de uso que para el sistema que albergue el proceso del Modelamiento. Algunos de los casos de uso son fases de la metodología CRISP-DM.

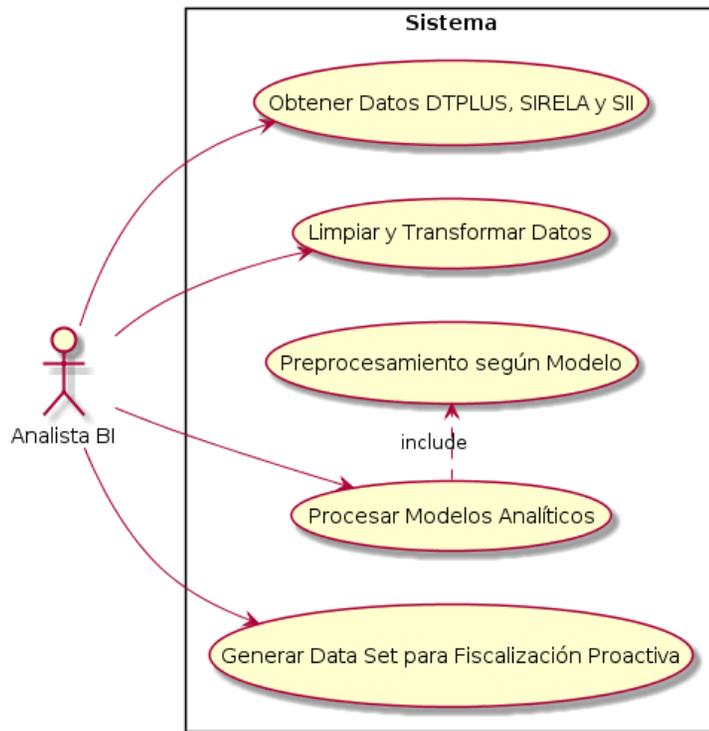


Figura 35: Casos de Uso, Arquitectura para el Rediseño.

Fuente Elaboración Propia.

La Figura 35 muestra los casos de uso para el sistema virtual, se observan: “Obtener Datos DTPLUS, SIRELA y SII”, “Limpiar y Transformar Datos”, “Preprocesamiento según Modelo”, “Procesar Modelos Analíticos” y “Generar Data Set para Fiscalización Proactiva”. El único actor presente es el Analista BI, el cual es encargado del modelamiento y la obtención de las bases de datos para el procesamiento. Para el entendimiento de los casos de Uso se realizan plantillas de descripción para cada uno de los casos en las siguientes Figuras.

Caso de Uso	Obtener Datos
Descripción	Extraer los datos de los sistemas internos y externos
Escenario de Éxito	<ol style="list-style-type: none"> 1. Abrir navegador web 2. Conectarse al Server 3. Aplicar Queries de extracción de datos 4. Importar datos a estructura Data Frame

Figura 36: Caso de Uso “Obtener Datos”, Arquitectura Propuesta.

Fuente Elaboración Propia.

Caso de Uso	Limpiar y Transformar Datos
Descripción	Pre-procesar los datos obtenidos de los sistemas internos y externos en términos de limpieza y transformación
Escenario de Éxito	<ol style="list-style-type: none"> 1. Llamar a la estructura DataFrame de datos 2. Generar un DataFrame transitorio 3. Aplicar scripts y códigos de limpieza a dataFrame transitorio 4. Generar nuevo DataFrame con limpieza

Figura 37: Caso de Uso “Limpiar y Transformar Datos”, Arquitectura Propuesta.

Fuente Elaboración Propia.

Caso de Uso	Procesar Modelos Analíticos
Descripción	Aplicar Modelos Analíticos seleccionados al Data Set limpio y pre-procesado según el tipo de Modelo
Escenario de Éxito	<ol style="list-style-type: none"> 1. Include (Pre-procesamiento según Modelo) 2. Se llama al Data Set preprocesado según modelo 3. Se separa en Entrenamiento y Testeo 4. Se utiliza el Set de Entrenamiento para el aprendizaje del modelo 5. Se aplica el modelo entrenado al Set de testeo 6. Se obtienen resultados de indicadores

Figura 38: Caso de Uso “Procesar Modelos Analíticos”, Arquitectura Propuesta.

Fuente Elaboración Propia.

Caso de Uso	Generar Data Set para Fiscalización Proactiva
Descripción	Generar Base de Datos con Scoring realizado por los Modelos Analíticos Seleccionados según Indicadores
Escenario de Éxito	<ol style="list-style-type: none"> 1. Evaluar Indicadores de Modelos 2. Seleccionar Mejores Modelos 3. Tomar Data Set para predecir 4. Aplicar Scoring 5. Generar Nuevo Archivo de Datos para la entrega a Unidad Fiscalizadora

Figura 39: Caso de Uso “Generar Data Set para Fiscalización Proactiva”, Arquitectura Propuesta.

Fuente Elaboración Propia.

Caso de Uso	Pre-procesar según Modelo
Descripción	Preparar el Set de Datos en Base al Modelo a aplicar/entrenar
Escenario de Éxito	<ol style="list-style-type: none"> 1. Se llama al Set de Datos limpio y transformado 2. Se especifica el Modelo a utilizar 3. Se revisan las necesidades de acuerdo al modelo seleccionado 4. Se realizan las depuraciones y transformaciones correspondientes al Modelo 5. Se genera nuevo Data Frame preprocesado

Figura 40: Caso de Uso “Pre-procesar según Modelo”, Arquitectura Propuesta.

Fuente Elaboración Propia.

6.3.2 Diagrama de Arquitectura del Sistema

En la Figura 41 se observa el diagrama de arquitectura del sistema virtual a desarrollar, el cual posee los actores principales para el funcionamiento del proceso de modelamiento generado. El sistema virtual levantado incluye el software R conectado a su interfaz RStudio y el motor postgresQL. Además, se incluyen los scripts generados en la iteración de la lógica de negocios. Por último, se presentan las conexiones respectivas a las bases de datos de los sistemas institucionales, donde cabe destacar que el actor “SII” es la base de datos compartida por el Servicio de Impuestos Internos a través de un convenio realizado con la organización. El diagrama que muestra las relaciones y

protocolos de comunicación entre los distintos actores se presenta en el siguiente apartado.

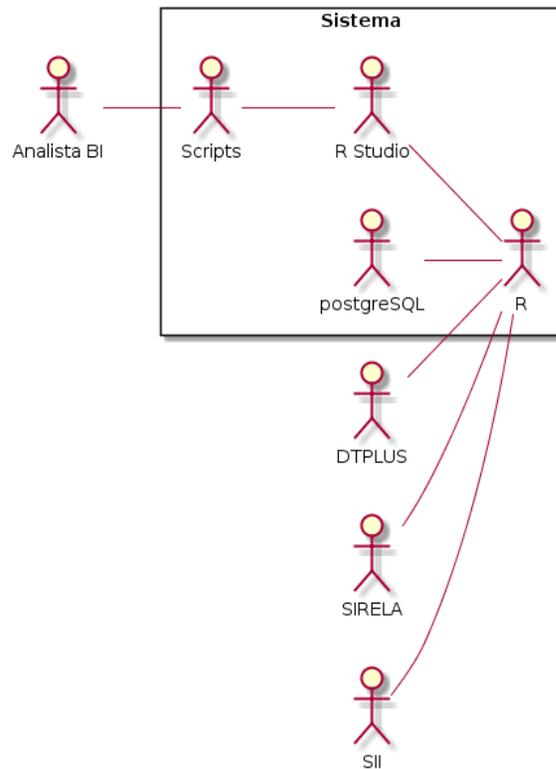


Figura 41: Diagrama de Arquitectura del Sistema, Arquitectura Propuesta.

Fuente Elaboración Propia.

6.3.3 Diagrama de Despliegue

En la Figura 42 se observa el diagrama de despliegue con el fin de identificar la estructura física de la arquitectura propuesta. El actor “Analista BI” se conecta a través de la interfaz gráfica del nodo web y del nodo consola a la aplicación virtual dentro del servidor (Nodo ServidorAplicacion), el cual se realiza a través de protocolo TCP/IP. Dado que en el proyecto se utilizó R y RStudio como los softwares principales, se levanta un servidor apache para el uso de RStudio vía web con R Server para su ejecución remota, donde se albergan los scripts desarrollados para la aplicación de los modelos, donde RStudio se conecta a la sesión de R al ser un ambiente de desarrollo integrado (IDE por sus siglas en inglés). El sistema operativo se conecta a su terminal a través del puerto 22, donde R trabaja directamente en el terminal dado que es un lenguaje interpretado y no requiere ser compilado. R se conecta a las bases de datos de la organización, las cuales se ubican en un servidor SQL (Nodo SQLSERVER), mediante el cual se tiene acceso a los registros de fiscalizaciones y de los sistemas de la DT.

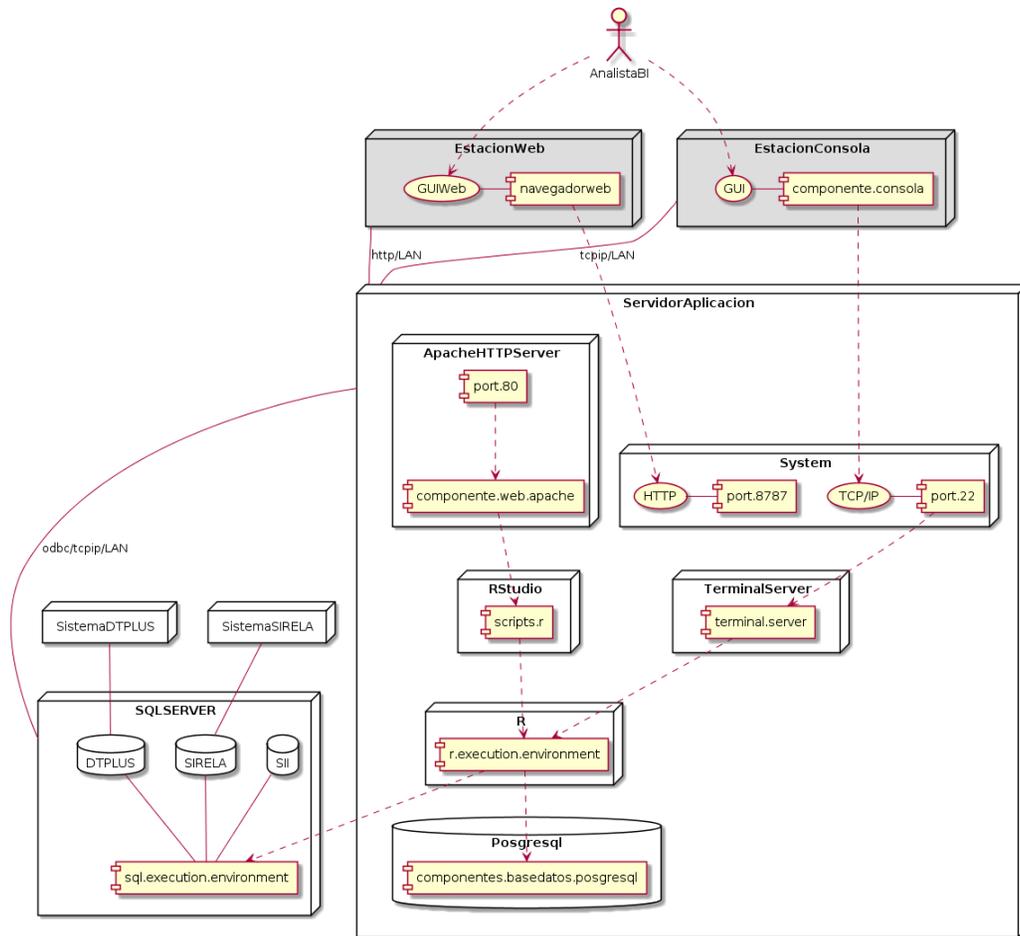


Figura 42: Diagrama de Despliegue, Arquitectura para el Rediseño.
Fuente Elaboración Propia.

CAPÍTULO 7: GESTIÓN DEL CAMBIO

7.1 Contexto de la Empresa

La Dirección del Trabajo es una organización cuya función es vital en la sociedad chilena, dado que es la única entidad que se encarga de velar por la legislación laboral en Chile. La importancia de la organización viene entregada a sus usuarios por Fiscalizadores y Asesores Laborales, quienes cumplen con la misión Institucional en el día a día.

Una de las desventajas de la organización es la baja capacidad en términos de fiscalizadores, debido al creciente número de empresas en los últimos 15 años (SII, 2017), la cobertura de las fiscalizaciones se ha visto afectada y en la actualidad el número de fiscalizaciones anuales está alrededor del 20% de la cantidad de empresas vigentes. Este concepto se considera como **Percepción de Riesgo** el cual es la Probabilidad de que un fiscalizador llegue a una empresa chilena. Con este escenario presente, se hace necesario la optimización y la efectividad de los recursos en su operación, donde el principal producto entregado por la dirección debe tener foco primario, el cuál es la fiscalización.

Como se explica anteriormente en la tesis, la fiscalización se separa en dos grandes subproductos, fiscalización Reactiva o bajo denuncia y la fiscalización proactiva o bajo programa, siendo esta ultima el principal instrumento de fiscalización en términos preventivos de infraccionalidad. Históricamente, la fiscalización proactiva mantiene una asertividad en torno al 25%, lo cual muestra que en una de cuatro fiscalizaciones, se logra encontrar infracción, siendo las 3 fiscalizaciones restantes un costo en términos de horas de fiscalizador y un costo de oportunidad de poder fiscalizar una empresa que si pueda presentar infracción. Bajo este espíritu se desarrolla el proyecto de Modernización del Estado, donde una de sus líneas de fortalecimiento es la fiscalización proactiva, incorporando técnicas de Inteligencia de Negocios para el análisis de la información institucional y la documentación del conocimiento encontrado en base a esta. Esta incorporación de técnicas, atenta contra los actuales paradigmas en los cuales se basa la fiscalización actual en la Dirección, siendo esto la motivación a realizar un plan de gestión del cambio que permita a los fiscalizadores y los trabajadores en general incorporar de manera exitosa nuevas prácticas que ayuden al cumplimiento de la misión institucional con una mayor efectividad y asertividad.

7.2 Análisis de los Principios de Diseño

Los principios analizados en este apartado están basados en el Modelo Integral de Liderazgo y Gestión del Cambio (Olguín, Crawford, & Soto, 2016). Este modelo incorpora al análisis diferentes dominios a considerar para la realización de un plan de gestión del cambio. Estos dominios se analizan de manera directa respecto al proyecto de tesis, además de los cambios que ocurren en el entorno de otros proyectos inmersos en la organización, en específico el proyecto de Modernización del Estado, previamente mencionado, donde se añaden diversos aspectos que afectan de manera directa al presente proyecto.

7.2.1 Liderazgo y Gestión del proyecto de cambio

El proyecto de tesis fue liderado por un equipo dentro de la organización conformado por el autor, Alejandro Allende Aced, quien realizó los modelos de Inteligencia de negocios y el rediseño de procesos que soportan el cambio, por otro lado, existen otros actores atinentes al proyecto de tesis y al proyecto de Modernización del estado, donde se encuentra el coordinador del proyecto de modernización, coordinador del programa de Modernización, el subdirector de la Dirección del Trabajo, el Jefe de Fiscalización y Profesionales de Fiscalización, quienes están presentes en las discusiones y reuniones de dirección y seguimiento del proyecto.

Un factor fundamental del equipo de liderazgo fue el alineamiento de los objetivos del proyecto de modernización con los objetivos del proyecto de tesis, donde la planificación y gestión se manejó en base a los tiempos del proyecto de modernización. Esto permitió la realización de reuniones de seguimiento del proyecto de tesis, además de mesas de trabajo, incorporando a fiscalizadores profesionales quienes poseen la información y el entendimiento del negocio.

El equipo de liderazgo se puede observar en la Tabla 19, donde se describe el rol de la persona en la organización y se señala su nombre.

Tabla 19: Equipo de Liderazgo del Proyecto de tesis.

Nombre	Rol
Alejandro Allende Aced	Tesista de Magíster Ingeniería de Negocios
Claudio Collao Bahamondes	Coordinador de Proyecto de Modernización, Dirección del Trabajo
Rafael Pereira Lagos	Subdirector, Dirección del Trabajo

Jorge Guzmán Kusanovic	Jefe de Dpto. de Inspección, Dirección del Trabajo
Jose Inostroza	Coordinador y Director de Programa, Ministerio de Hacienda

7.2.2 Estrategia y sentido del proceso de cambio

La estrategia del proyecto de tesis está alineada al Proyecto de Modernización en el cual está inmerso, donde se detalla que el objetivo principal es aumentar la satisfacción de los usuarios del servicio. Este objetivo principal se desagrega en diversos objetivos específicos que incorporan los beneficios tanto al usuario final como organizacionales.

La tarea de comunicar a los diversos actores del proyecto de tesis fue compleja, debido a que la construcción de estos modelos se inició en formato laboratorio con Fiscalizadores profesionales a nivel de Departamento Inspectivo, pero no a nivel de fiscalizador en Terreno, donde la posterior

7.2.3 Cambio y Conservación

Si bien existen deficiencias en la actual operación y planificación de la organización, hay una serie de prácticas institucionales que han resultado en buenos resultados a lo largo de la historia de la Dirección, donde es importante destacar en este ejercicio de cambio no se eliminarán las prácticas que han otorgado buenos resultados.

Por otro lado, debido a las características del proyecto, aumentar la tasa de asertividad de las fiscalizaciones proactivas se podría realizar a través del incremento de capacidad fiscalizadora, pero una restricción importante es el no contratar ni despedir a ningún trabajador de la organización. Con esto definido, el cumplimiento del objetivo del proyecto se aborda en base a mejorar la eficiencia y efectividad de las fiscalizaciones través de la incorporación de herramientas de Inteligencia de negocios y las tecnologías que soporten este nuevo proceso, donde parte importante de este es la incorporación de un proceso de Desarrollo de Pronósticos en base a Información, lo cual no existe en la organización pre-proyecto.

Espacio de Conservación

Los Fiscalizadores de la Dirección del Trabajo poseen un valor institucional muy arraigado a su función inspectiva, debido a la relevancia de su función en el servicio. Es la fiscalización uno de los productos estratégicos más importantes de

la organización, por ende, el fiscalizador debe ser el actor ilustre. Con esto en consideración, el aspecto técnico, conocimientos, procedimientos y experiencia en la fiscalización de terreno se conservan en base a esta identidad institucional y su experiencia en el rubro. Además, se conserva la autonomía que tienen al momento de organizar la carga de trabajo que poseen, donde son ellos quienes tienen la libertad de estructurar sus tiempos en base a sus necesidades.

Espacio de Cambio

En función del objetivo del proyecto, el gran espacio de cambio es en base a la incorporación de nuevas tecnologías que permitan hacer de las fiscalizaciones un proceso más eficiente y efectivo. Es por eso que se desea implantar como práctica en el servicio, el desarrollo de programas de fiscalización (fiscalizaciones proactivas) en base a la información dentro de los sistemas de la organización y además, incorporar bases de datos de servicios relacionados que permitan mejorar el análisis, en conjunto a esto se deben incorporar las capacidades necesarias para que este desarrollo sea exitoso en el futuro de la organización y no sea un esfuerzo puntual, acá se pueden incorporar metodologías tipo coaching, que permitan generar un aprendizaje transformacional.

Otro cambio que se busca implementar es el dejar de entregar programas en base a cuotas de fiscalizaciones (cantidad de fiscalizaciones a realizar) y modificarlo a programas en base a características definidas específicas, incorporando como variable necesaria el RUT de la empresa. Esto se desea aplicar dada la baja cartera de empresas inscritas o fiscalizadas en los sistemas del servicio, con el fin de incorporar nuevas empresas al análisis y posteriormente una posible fiscalización.

Lo anterior se basa en el apoyo tecnológico que se tendrá en función de esta nueva forma de fiscalizar, además de dejar el conocimiento implantado en los sistemas del servicio.

7.2.4 Organización y Estructura del proyecto de cambio

Los principales actores y entidades involucrados en el proyecto son: Jefe de Departamento de Fiscalización (Nivel Dirección Nacional), Jefes de Fiscalización (Nivel Dirección Regional), Unidad de Planificación (Dirección Nacional), Fiscalizadores de terreno y Jefes de Fiscalización (Nivel Inspección). Las funciones no varían de manera abrupta, salvo el fiscalizador de terreno, quien ya no posee la facultad de elegir a cuál empresa fiscalizar y la Unidad de Planificación, la cual ahora realiza su planificación en base a información.

El flujo de información y conversaciones comienza desde la unidad de Planificación, donde de forma anual, en conjunto con la jefatura, realizarán el análisis de la información y planifican los programas a realizar durante el año, para posteriormente difundir este programa a las Direcciones Regionales para

seguir el flujo a las distintas inspecciones en todas las regiones, donde el Jefe de Fiscalización recibe el programa ya en un nivel enfocado en la ubicación geográfica de la Inspección, para distribuir las fiscalizaciones entre los fiscalizadores, quienes en terreno, realizar los procedimientos correspondientes de inspección.

7.2.5 Gestión Emocional

La organización posee una historia y trayectoria que es valorada por sus trabajadores, los cuales poseen una identidad en base a esto, donde los fiscalizadores son de especial relevancia en términos estratégicos. El fiscalizador se caracteriza por su nivel de autonomía en su función, por lo cual se debe tener especial cuidado, debido a que el proyecto de forma directa, le quita la facultad del criterio de elección. Si bien esto puede resultar altamente sensible, la introducción y la explicación del porqué se realiza tal acto, se debe declarar que es para mejorar la tanto la atención a los usuarios finales como los beneficios que esto tendrá para ellos y para la organización. Es importante difundir de manera correcta estos beneficios, dado que son de vital importancia para la recepción oportuna del proyecto.

7.2.6 Comunicaciones

El proyecto ha generado de por sí innumerables instancias de comunicación y conversación, donde diversos actores han podido dar sus puntos de vista al respecto y ser considerados dentro del desarrollo de esto. En particular, gran parte de las preocupaciones que nacen de los actores en base a este proyecto han sido revisadas, salvo la expectación del posible éxito de este.

7.2.7 Desarrollo de las Habilidades

Las habilidades a incorporar y gestionar dentro del proyecto se dividen en dos niveles. En primer lugar, están las habilidades técnicas y que están basadas en el aprendizaje metodológico, nuevos procedimientos, nuevas tecnologías, etc. De lo cual el proyecto presenta una ardua tarea, debido a que la incorporación de modelos de predicción es un aprendizaje que requiere de práctica constante.

Por otro lado, existe un nivel el cual es el desarrollo de la habilidad de trabajar en conjunto y poder llevar a cabo el proyecto, debido al tamaño de la organización y las diversas variables que afectan a su labor, se requiere de la habilidad de poder conectar a las personas en una dirección en común, trabajando en conjunto para cumplir los objetivos del proyecto.

7.2.8 Gestión del Poder

Dentro de las fuentes de poder de la organización, existe un sentimiento latente en términos de la autoridad, a que lo que se discute en cargos directivos y baja a los cargos operativos se debe de alguna forma acatar, dadas las experiencias previas de proyectos. Es por eso que se busca incentivar el proyecto en base a diferentes fuentes, como el incentivar a partir de la identidad de la dirección, dada el existente valor que existe al respecto, e incorporar la necesidad de los nuevos usuarios, comenzar a plasmar estas nuevas necesidades en nuevas capacidades para la organización, lo cual viene en conjunto a la incorporación y mejora de otras fuentes de poder como al Know-How y el conocimiento de estas nuevas capacidades, las cuáles serán vitales para la implementación del proyecto.

7.2.9 Monitoreo y Evaluación del proceso

El proceso de cambio se va a monitorear en base a dos grandes ejes, primero se monitorearán los resultados en base a las fiscalizaciones propuestas, y se evaluará su resultado, donde como segundo eje, se revisarán todos los posibles aspectos relacionados al éxito o fallo de este, pudiendo ser a partir de dos grandes fuentes, errores en el modelamiento y análisis de la información o fallas en la implementación y adaptación a los fiscalizadores.

Con este monitoreo, se iterará y evaluará en términos de efectividad del proyecto, generando espacios de mejora entre iteraciones para mejorar la probabilidad de éxito del proyecto.

7.2.10 Inicio, hitos, ritos y cierre

Un aspecto importante dentro del proyecto es comunicar las etapas del proyecto e ir mostrando los avances que este posee, es por eso que cada mes se presentan los avances en reuniones donde diversos actores comentan y discuten de las preocupaciones en base al rediseño y los modelos incorporados.

Por otro lado, existen hitos predefinidos, los cuales se basan en el inicio, piloto y cierre, siendo los principales dentro del proyecto y deben ser señalados y difundidos a los diversos actores.

7.3 Caracterización del Cambio

El gran cambio a realizar con este proyecto es la reestructuración del proceso de selección de empresas al momento de fiscalizar de manera proactiva, donde el estado sin proyecto es en base al criterio del fiscalizador y sin ningún tipo de control salvo la restricción de la actividad económica que puede tener el programa

y el número de empresas totales las cuales debe fiscalizar, conocida como cuota. Por otro lado, el estado con proyecto dispone de un programa en base a modelos predictivos, el cual posee las empresas a ir a fiscalizar y la materia relacionada a la fiscalización. Este cambio puede atentar a las emociones de los fiscalizadores y su sentido de autonomía respecto a su función, lo cual debe tratarse con cuidado, dado que el proyecto no atenta a su planificación personal ni la forma en la cual fiscaliza, la gran diferencia es otorgarle una empresa objetivo a la cual determinar si es o no infractora, donde el fiscalizador no tendrá la facultad de elegir el mismo la empresa.

Se espera que este cambio genere un aumento en términos de la asertividad de las fiscalizaciones en términos muy generales, donde luego, en términos organizacionales, se espera una serie de cambios relacionados a esta mejora. En primer lugar, se espera que con una fiscalización proactiva eficaz disminuya la cantidad de denuncias (fiscalizaciones reactivas) que se realizan en las inspecciones del trabajo. En segundo lugar, se pretende pasar a ser una organización preventiva en vez de reactiva, dado que actualmente el 75% de las fiscalizaciones reactivas, queriendo pasar totalmente a lo contrario.

En general, el cambio va enfocado al usuario final, donde se desea aumentar la satisfacción de estos en base a la tecnologización de los procesos de la organización. Estas mejoras vienen dadas por la espera disminución de tiempos de demora en las inspecciones y en términos de fiscalizaciones se espera bajar la tasa de infraccionalidad dado el aumento de la tasa de acierto de las fiscalizaciones preventivas.

Debido a que el proyecto está inmerso en un proyecto más grande, es difícil separar los aspectos más importantes respecto a la gestión del cambio, dado que los diversos componentes generan, de forma directa o indirecta, cambios en los actores dentro de la DT.

7.4 Factores Críticos de Éxito

Los factores críticos para el éxito del proyecto se detallan en la siguiente lista:

- Apoyo a nivel estratégico por parte de los cargos Directivos y jefaturas, donde se busca alinear los objetivos de la empresa conjunto a los beneficios que trae consigo la incorporación de Inteligencia al análisis de Datos.
- Relación directa con los departamentos de Inspección y Tecnologías, debido a que son ambos los involucrados en el proceso de la generación de modelos, tanto a nivel de conocimiento del negocio por parte de Fiscalización como conocimiento de los datos por parte de Tecnologías.

- Constante educación y comunicación con los fiscalizadores debido a que son el eje principal de la organización, los cuales requieren de explicación y comprensión de los beneficios esperados que toma el nuevo alineamiento de la organización, con el fin de que sean actores protagonistas en el cambio a generar.

7.5 Plan de Gestión del Cambio

Las principales acciones a tomar para la gestión del cambio son las siguientes:

- Reuniones a nivel de dirección y jefaturas con el fin de explicar los alineamientos entre los objetivos de la organización y los beneficios de la creación de programas con inteligencia de Negocios.
- Creación de programas tipo piloto para pruebas tempranas (*quick wins*) de los modelos desarrollados en base a información de la organización, con el fin de obtener resultados tempranos y validación de los modelos.
- Capacitaciones a Jefes de Fiscalización en las inspecciones nacionales conjunto a los fiscalizadores para entregar todas las herramientas y conocimiento necesario para el cambio con profesionales capacitados de la Dirección del Trabajo para responder dudas y consultas respecto a procedimientos y cambios que generarían los nuevos programas.
- Comunicación constante de resultados a todo actor en la organización, con el fin de entregar confianza de los beneficios que traería el cambio.
- Reuniones de retroalimentación entre fiscalizadores y equipos de planificación encargados de la generación de programas, para enriquecer el proceso en términos de falencias o problemas.

CAPÍTULO 8: EVALUACIÓN DEL PROYECTO

Dentro del presente apartado se detalla la evaluación económica del proyecto MBE, donde se discute el tipo de evaluación debido a la naturaleza de la organización y su rol social dentro del mercado laboral, se detalla también el plan piloto desarrollado a 560 empresas durante el proyecto y sus respectivos resultados. Por otro lado, se construyen y desarrollan los costos relacionados al proyecto y sus principales beneficios, donde posteriormente son reflejados en el flujo de caja.

8.1 Plan Piloto

Con el objetivo de obtener resultados reales del modelo predictivo, se realiza una prueba tipo piloto de este con los datos de los últimos 2 años (2015, 2016), donde se dispone de una capacidad fiscalizadora de 560 empresas en un mes.

8.1.1 Definición del Plan Piloto

La selección de las empresas para el piloto se realiza en base a la probabilidad de infracción entregada por el modelo analítico, también como en la distribución de las empresas respecto a las regiones.

Debido a la alta carga operativa de las regiones más grandes como la Metropolitana, Valparaíso y el Bío Bío, se decide distribuir equitativamente las empresas a fiscalizar a lo largo del país, donde el resultado de la distribución se muestra en la Tabla 20, siendo la Región Metropolitana la única diferente con 42 empresas a fiscalizar.

Tabla 20: Distribución de fiscalizaciones, plan piloto 2017.

REGIÓN	CANTIDAD DE EMPRESAS
TARAPACÁ	37
ANTOFAGASTA	37
ATACAMA	37
COQUIMBO	37
VALPARAÍSO	37
LIB. BDO. O'HIGGINS	37
MAULE	37
BÍO BÍO	37
ARAUCANÍA	37
LOS LAGOS	37

AYSEN DEL G. C. IBÁÑEZ	37
MAGALLANES	37
LOS RIOS	37
ARICA-PARINACOTA	37
METROPOLITANA	42
TOTAL	560

Respecto a las materias a fiscalizar, se decidió realizar un barrido general por los tres tipos de materias más característicos de la inspección laboral, los cuales toman materias **Laborales, Previsionales y de Seguridad y Salud en el Trabajo.**

Respecto al procedimiento de fiscalización, se realiza en base a visitas presenciales, donde se incurre, sin previo aviso, al domicilio del empleador, donde se revisa la documentación de acuerdo a las materias a fiscalizar.

8.1.2 Resultados Obtenidos

El plan piloto se realizó durante el mes de noviembre del año 2017, donde del total de 560 empresas a fiscalizar, fueron ubicadas y respectivamente fiscalizadas 418 empresas, donde el resto no pudo ser ubicada por diversas razones. Del porcentaje ubicado, el porcentaje de infraccionalidad fue de un **42,24%**, el cual logra superar el porcentaje histórico de las fiscalizaciones proactivas.

Con el fin de generar un análisis más específico del piloto, se separa por ubicadas y no ubicadas, donde se analiza respecto a la infraccionalidad para el primer grupo y un análisis en términos de los motivos de aquellas que no fueron ubicadas.

8.1.2.1 Ubicadas

De las 418 empresas que fueron ubicadas y fiscalizadas exitosamente, se realiza una descripción en términos de infraccionalidad respecto a las variables que caracterizaron a las empresas, las cuales son Región y Rama de la Actividad Económica. En la Figura 1 se observan las fiscalizaciones realizadas según la región, donde se puede apreciar regiones con una alta infraccionalidad como también regiones con muy baja. Las regiones notables en términos de resultados son la sexta, la cual obtuvo una muy alta infraccionalidad sobre el promedio con un 67,8% de detección, y la región 11, con la infraccionalidad más baja del piloto, con un 17,2% de detección.

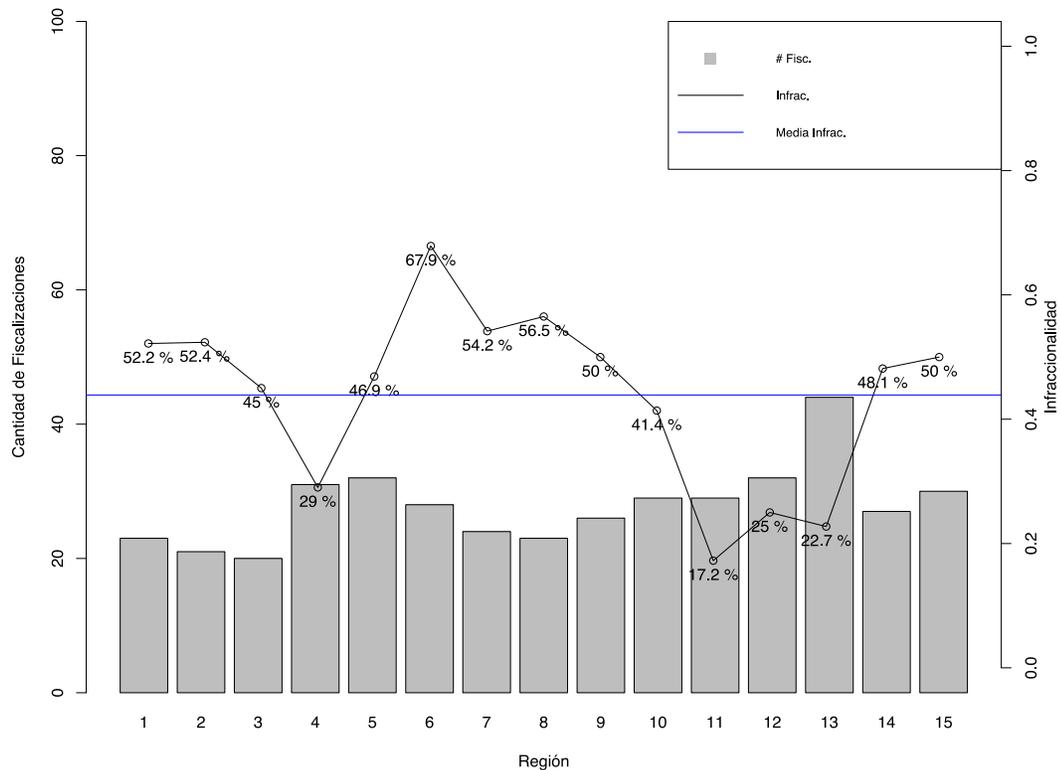


Figura 43: Gráfico de Infraccionalidad del Piloto según Región.

Fuente Elaboración Propia.

Para generar un análisis comparativo, en la Figura 44 se observa la infraccionalidad proactiva del piloto y la historia respecto a las regiones donde, en 14 regiones, el piloto logra superar el promedio histórico de la región, salvo la región Metropolitana, la cual bajó 2 puntos porcentuales de la historia. Las regiones 6 y 8 obtuvieron los dos resultados más altos, duplicando su infraccionalidad histórica, mientras que las regiones 11 y 12 presentan un comportamiento similar a esta, donde la infraccionalidad detectada supera ligeramente a su historia proactiva.

Dos de las regiones más infractoras las cuales se pueden observar en la curva de infraccionalidad histórica, dentro del piloto mantienen este comportamiento de alta detección, donde, el modelo predictivo, en su mayoría, logra captar estos patrones de comportamiento similares, tanto para las regiones que poseen una alta infraccionalidad, la 7 y 9, como para las regiones de baja infraccionalidad como la 11 y 12.

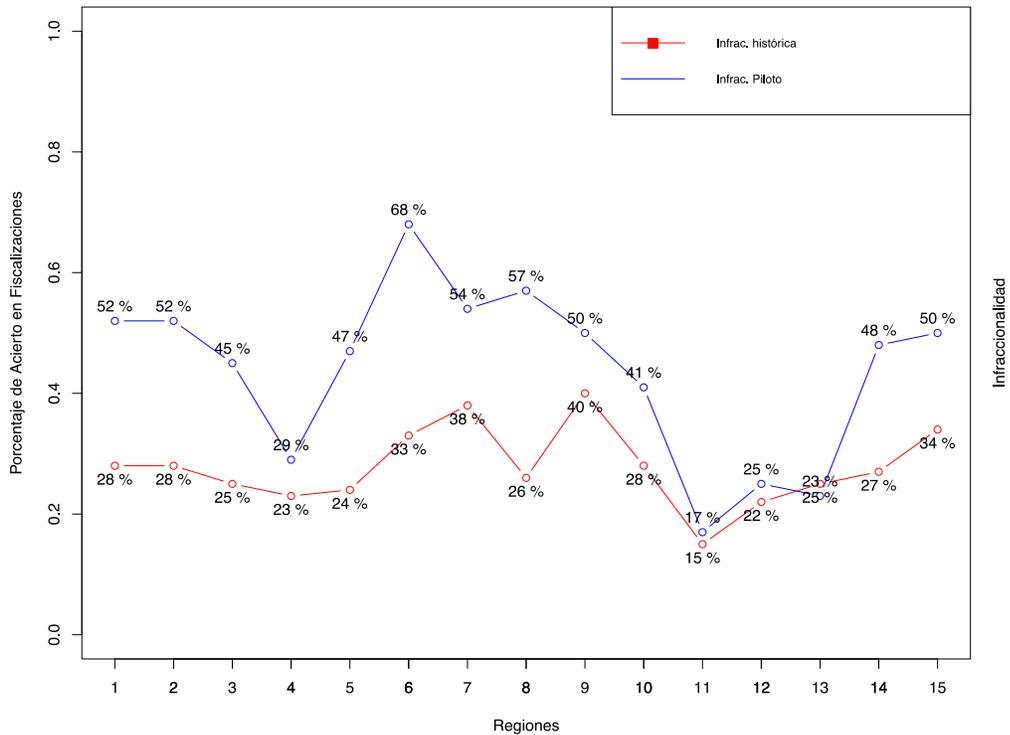


Figura 44: Comparación Infracionalidad Histórica vs. Piloto.

Fuente Elaboración Propia.

Respecto a la temporalidad durante la cual fue realizada el piloto, pueden existir componentes exógenas que hayan comprometido el resultado general del piloto, tales como rubros con una mayor actividad productiva durante cierto mes en particular, o trimestres, etc. como el rubro agrícola el cual es altamente productivo el último trimestre del año, por ende, se hace hincapié en la fiscalización del rubro en tal fecha.

El piloto fue realizado durante el mes de noviembre del año 2017, donde se realizaron las fiscalizaciones asignadas por el modelo predictivo, por lo cual uno de los cuestionamientos es si este período es más probable de tener una mayor infraccionalidad o si efectivamente el modelo genera valor agregado a la fiscalización.

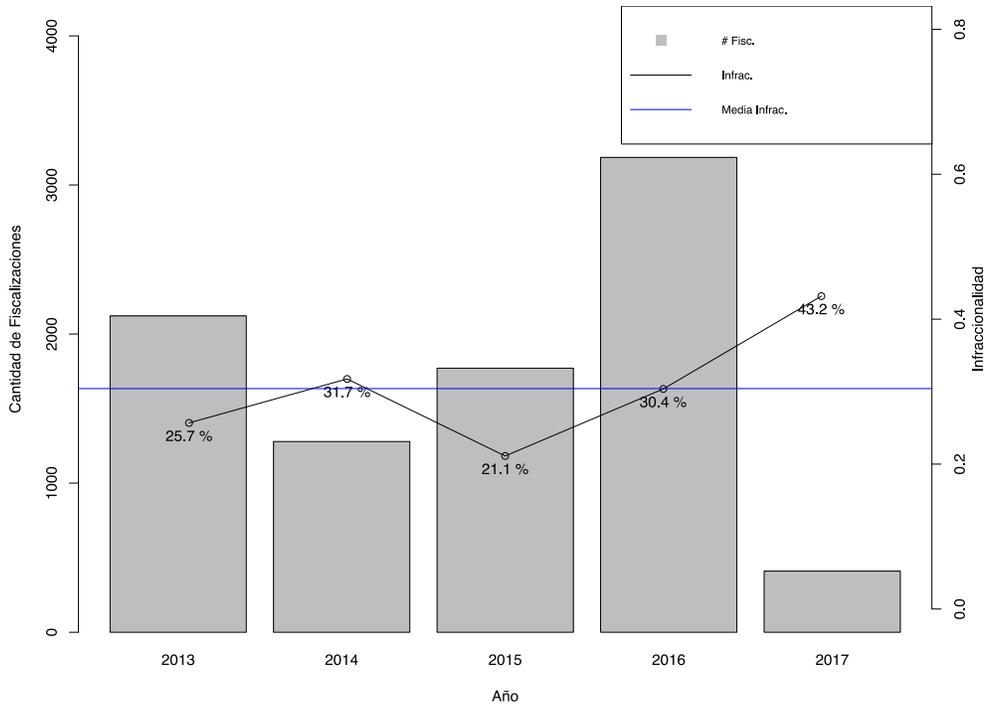


Figura 45: Gráfico de Infracionalidad del mes de noviembre según años.

Fuente Elaboración Propia.

Con el fin de analizar tal problemática, en la Figura 45 se observa un gráfico comparativo del mes de noviembre durante el periodo 2013-2016 respecto a su infraccionalidad promedio proactiva, donde en último lugar se presenta el piloto realizado por el proyecto. Se observa que entre los años 2013 y 2016 el promedio de infraccionalidad proactiva es de un 27,2%. Los *peaks* ocurren en los años 2014 y 2016, donde noviembre del 2016 se realizan más de 1000 fiscalizaciones proactivas más que el 2014, pero con una similar tasa de acierto, lo cual es debido al proyecto de Modernización BID, el cual realizó un programa de fiscalización con el fin de obtener la tasa de infraccionalidad base o “natural” de Chile. En noviembre del 2015 se obtiene la tasa más baja con un 21% y con casi 2000 fiscalizaciones. El piloto realizado en noviembre del año 2017 en base al modelo predictivo realizado por el proyecto de tesis del autor, obtiene con 560 fiscalizaciones un 43% de asertividad.

Si bien la cantidad de fiscalizaciones es baja respecto a los otros años, la tasa de infraccionalidad proactiva tiende a la baja en la medida que los meses del año pasan tal como se observa en la Figura 46, sin embargo, la cantidad de fiscalizaciones va en aumento. Esto se puede deber a la acumulación de carga de trabajo de los fiscalizadores, quienes además de fiscalizar deben cumplir con las cargas administrativas posteriores.

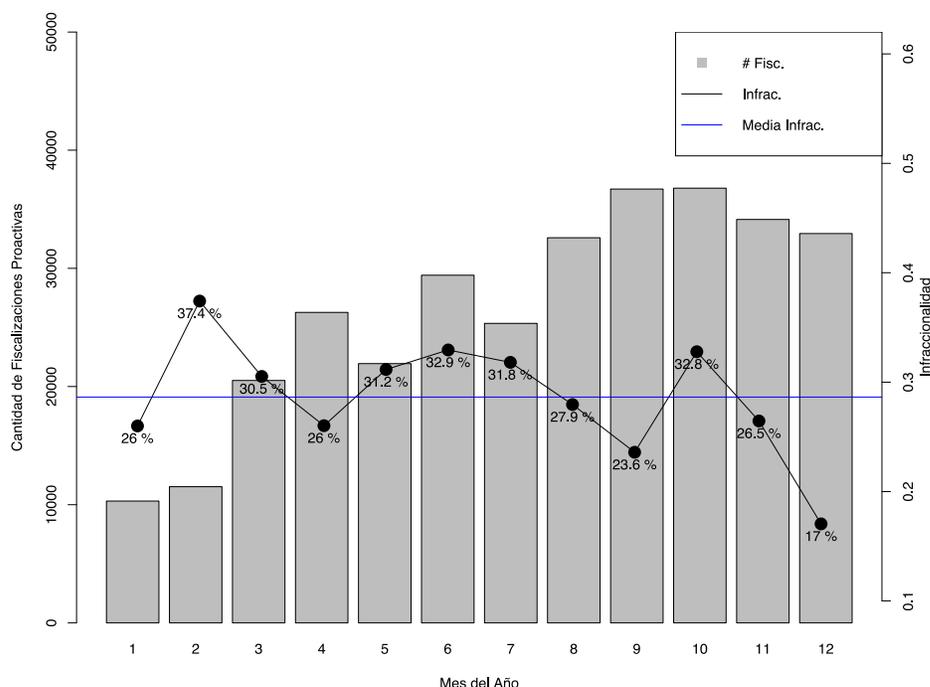


Figura 46: Gráfico de Infraccionalidad Proactiva según mes del año.

Fuente Elaboración Propia.

De acuerdo a lo planteado, se infiere que el mes de noviembre no presenta una variable que aumente la probabilidad de encontrar infracción, sino que tiende a bajar dado que la carga aumenta en los últimos meses. De acuerdo a esto, el resultado del piloto se considera un aumento de la tasa de acierto, cuyo modelo predictivo genera un valor positivo para la fiscalización.

Respecto a las actividades económicas, se puede observar en la Tabla 21 todas aquellas que fueron fiscalizadas en el piloto.

Tabla 21: Código de Actividades Económicas Fiscalizadas

Código	Glosa
0	Otras Actividades de Servicios Comunitarios
1	Agricultura, Ganadería, Caza y Silvicultura
2	Explotación de Minas y Canteras
3	Industrias Manufactureras
4	Suministro de Electricidad, Gas y Agua
5	Construcción
6	Comercio
7	Transporte, Almacenamiento y Comunicaciones
8	Actividades Inmobiliarias, Empresariales y Alquiler
9	Hoteles y Restaurantes

Se grafican las actividades respecto a su infraccionalidad detectada en la Figura 47, donde se encuentran las altamente infractoras, como “Suministro de electricidad, gas y agua”, “Agricultura”, “Comercio” y “Hoteles y Restaurantes”. De acuerdo a la historia de la organización, las empresas de suministro de electricidad, gas y agua y los Hoteles y Restaurantes presentan una alta infraccionalidad, donde el piloto demuestra que logra capturar tal patrón. Por otro lado, las actividades: “Explotación de minas y canteras”, “Construcción” e “Industrias Manufactureras” presentaron una baja asertividad, siendo que históricamente son de las actividades más infraccionadas. Esto se puede deber a la baja cantidad de empresas fiscalizadas de tales actividades, y también la alta rotación de acuerdo a las características de los rubros, donde, por ejemplo, la construcción posee faenas, las cuales pueden terminar al cabo de 2,3 años dependiendo del proyecto, por lo cual es una actividad altamente dinámica que, para encontrar y fiscalizar, se debe tener una componente temporal anticipada.

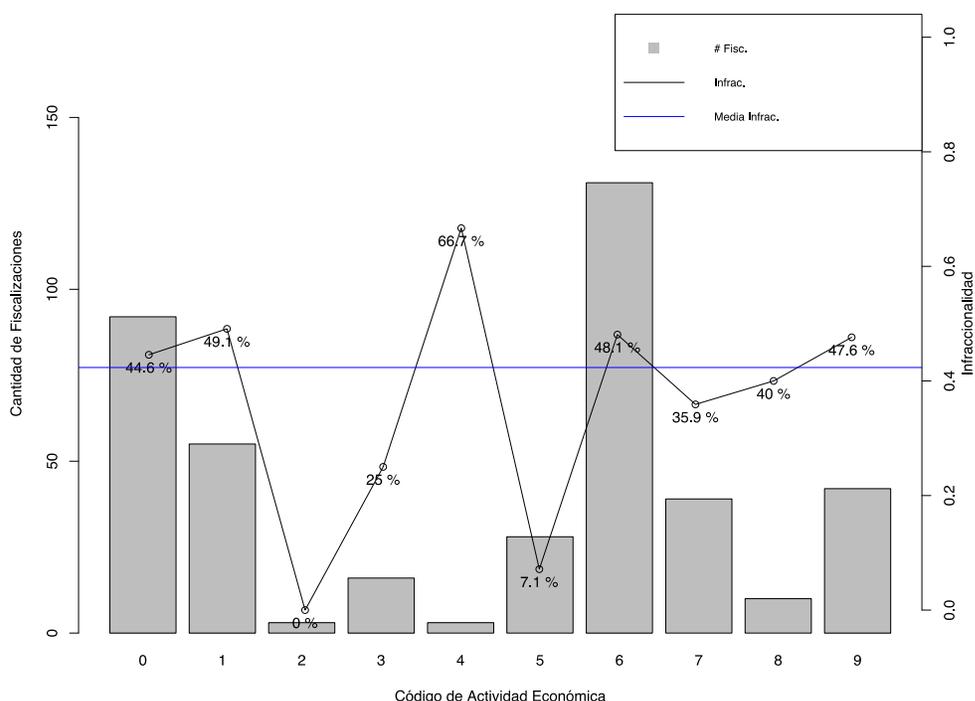


Figura 47: Gráfico de Infraccionalidad del Piloto según Rama de Actividad Económica.

Fuente Elaboración Propia.

8.1.2.2 No Ubicadas

Respecto a aquellas empresas no ubicadas por el fiscalizador, se observan y discuten diversos motivos por lo cual puede ocurrir este efecto no menor, dado que es un 25% aproximado del total de fiscalizaciones del piloto.

Este problema nace debido a las direcciones obtenidas para el set de empresas, las cuales están basadas en las direcciones que son obtenidas por las

fiscalizaciones realizadas por la DT, donde para una misma dirección, se tienen diferentes formas de escritura, abreviaciones, no enumeradas, entre otros problemas. Estas direcciones se validaron de manera manual en la web donde se generó un listado final con la dirección a fiscalizar.

De las fiscalizaciones no ubicadas, se revisan ciertos criterios que provocaron que no se encontrara la empresa, los cuales son los siguientes:

- **MYPE:** Aproximadamente el 90% de las empresas que no fueron ubicadas son clasificadas como MYPE (Micro y Pequeñas Empresas). Esto puede indicar que las direcciones proporcionadas por estas empresas podrían ser domicilios particulares de los respectivos dueños o domicilios del representante legal.
- **No ubicadas Anteriormente:** El 15% aproximado de las empresas no ubicadas, no habían sido encontradas previamente en los últimos 2 años debido a la poca estandarización de las direcciones.
- **Situaciones Irregulares en el SII:** Aproximadamente, el 27% de los empleadores no ubicados poseen situaciones tributarias irregulares, los cuales algunos ya no eran ubicados por el SII, otros con cierre de actividades, entre otras situaciones.

A partir de esta problemática, nace la iniciativa de que dada la alta cantidad de direcciones diferentes para la ubicación de un mismo lugar se incorporen estrategias en base a puntos geo-localizados, con el fin de evitar el gran número de formas diferentes de escribir una misma dirección.

8.2 Definición de Variables y Criterios de Evaluación

8.2.1 *Tipo de Evaluación*

La Dirección del Trabajo, al igual que otros servicios públicos, basa sus ingresos respecto al presupuesto anual que se le otorga, y se re-evalúa según el cumplimiento de este año tras año, por consecuencia, los ingresos son invariables respecto al servicio que otorga a sus usuarios. Con esto presente, y como ha sido comentado en el documento, el proyecto se basa en la eficiencia y optimización de los recursos actuales de la DT, donde se incorpore Inteligencia a sus procesos con el fin de lograr ese objetivo, por lo cual el tipo de evaluación se focaliza en los costos en que incurre el servicio.

Si bien los costos se basan en su mayoría en costos de personal, la efectividad de estos se ve comprometida al momento de realizar fiscalizaciones de manera no eficaz, por consiguiente, se denomina como **costo no efectivo** al costo en que se incurre en realizar una fiscalización y que esta obtenga un resultado de no infracción, por lo cual se gasta tiempo de fiscalizador y también el costo de oportunidad de fiscalizar una empresa que si pueda estar cometiendo infracción, donde esta estimación se revisa en la sección de costos.

Dada esta definición, la evaluación a realizar es de **Costo-Efectividad**, donde se evaluará el Costo Anual Equivalente (CAE) para los escenarios con y sin proyecto, para luego incorporar un nuevo indicador el cual divide el CAE sobre la cantidad de fiscalizaciones efectivas que se obtiene en cada escenario, el cual se denomina como **Costo Efectivo por fiscalización**. Luego, el criterio de selección será en base a aquel escenario que presente el menor Costo Efectivo por fiscalización.

8.2.2 ***Horizonte de Evaluación***

Debido a la aleatoriedad del mercado laboral en Chile, conjunto al avance de las tecnologías, se decide trabajar con un horizonte de evaluación de **3 años**, con el fin de no entrar en estimaciones que puedan distorsionar el valor real del proyecto.

8.2.3 ***Tasa de Descuento***

El tipo de evaluación, en términos estrictos, es una evaluación privada, dado que no se incorporan los beneficios sociales dada la complejidad de su cálculo en estas materias, pero se realiza desde una perspectiva del estado, debido a que se basa en la efectividad de sus costos en base a la implementación del proyecto. De esta manera, la tasa de descuento a utilizar es la tasa social de descuento entregada por el Ministerio de Desarrollo Social, la cual es de un **6%**.

8.2.4 ***Escenarios de Evaluación***

Los escenarios a evaluar se diferencian como Escenario sin Proyecto y Escenario con Proyecto.

El **Escenario sin Proyecto** representa a la situación actual optimizada, donde se estima la cantidad de fiscalizaciones por año y sus respectivas eficacias en base al tipo de fiscalización, donde estas tasas están dadas por los datos históricos de la Dirección del trabajo. Este escenario es el escenario base, donde se realizan las funciones inspectivas tal cual como se hacen en el presente.

El **Escenario con Proyecto** representa a la incorporación del modelo de Inteligencia de Negocios y el rediseño de los procesos que lo sustente, donde la cantidad de fiscalizaciones es la misma que el escenario base, pero las tasas de eficacia y tipo de fiscalización son modificadas dada la incorporación del modelo. Además, el escenario incorpora las horas hombre que se utilizaron para el desarrollo del modelo en la etapa inicial (Año 0) y luego se incorpora la necesidad de un nuevo rol en la DT que esté encargado de la mantención y el cuidado regular de este, debido a los diversos cambios y resultados obtenidos por parte de las fiscalizaciones realizadas.

8.2.5 Costos

La estructura de Costos del proyecto está en base a la redistribución de fiscalizaciones de los actuales trabajadores de la Dirección del Trabajo que presentan los escenarios previos.

La **estructura de costos sin Proyecto** es la situación actual optimizada, donde el costo total es calculado como la cantidad de fiscalizaciones estimadas a realizar en el año, por el costo en términos de Horas Hombre estimado.

Las estimaciones de fiscalizaciones a realizar se pueden observar en la Tabla 22, donde se obtienen a partir de las fiscalizaciones realizadas desde los años 2010 hasta el 2016, donde muestran una tendencia al alza cercana al 5 %.

Tabla 22: Estimación de Fiscalizaciones.

Año	Fiscalizaciones
Año 1	129.004
Año 2	131.031
Año 3	133.057

Para la estimación del costo por fiscalización, se obtienen aproximaciones de los fiscalizadores donde en promedio se comenta que las fiscalizaciones duran 4 horas, a lo cual se le agrega el costo por en términos de transporte. Con fuente de Gobierno Transparente, se obtuvo un promedio ponderado del valor por hora de los fiscalizadores, llegando a una suma de **CLP\$12.256**, conjunto a **CLP\$2.500** en términos de transporte. Con estos valores, se llega a la suma de **CLP\$51.524** por fiscalización.

Para la **estructura de costos con proyecto**, la cantidad de fiscalizaciones es la misma, solo se incorporan costos anuales en términos de la incorporación de un Ingeniero en Inteligencia de Negocios lo cual, en términos de los grados del servicio, posee un sueldo anual de **CLP\$20.225.630** y también un costo de mantención por parte del Departamento de Tecnologías, lo cual posee un costo anual de **CLP\$830.400**.

8.2.6 Tasas Contingentes

De acuerdo a los objetivos que presenta el proyecto, las tasas que cambiarían por la implementación de este son la Tasa de Asertividad, la cual es la división entre las fiscalizaciones que encontraron infracción sobre el total de fiscalizaciones, y la tasa de tipo de fiscalización sobre el total, la cual muestra la relación entre fiscalizaciones reactivas y proactivas.

El objetivo del proyecto plantea aumentar la tasa de asertividad al doble, y con esto, se estima que la tasa de tipo de fiscalización (Proactivas sobre el total)

aumente a un 40 % (4 de cada 10 fiscalizaciones serían proactivas). Cada una de estas tasas poseería, además una tasa de crecimiento anual, las cuales se pueden observar en la Tabla 23. Cabe destacar que el crecimiento de la asertividad reactiva no aumenta de forma anual debido a que el proyecto no contempla el mejoramiento de los procedimientos de fiscalización.

Tabla 23: Tasas contingentes para la evaluación del proyecto.

Tasa	Valor sin Proyecto	Valor con Proyecto	Crecimiento Anual (Solo con Proyecto)
Asertividad Fisc. Proactiva	25,5%	55%	5%
Asertividad Fisc. Reactiva	40,5%	40,5%	0%
Fiscalizaciones Proactivas sobre el Total	27,8%	40%	5%

8.2.7 Inversión

En términos de inversión del proyecto, los costos asociados son en base a las reuniones existentes en los 6 meses de creación de proyecto que tienen un valor respecto al tiempo utilizado en reuniones de trabajo por los actores presentados. En la Tabla 24 se muestran las horas por mes en las cuales fueron partícipes diversos actores y su valor final en los 6 meses.

Tabla 24: Montos de Inversión.

Actor	Horas Mensuales	Costo por Hora (CLP)	Costo Final (CLP)
Subdirector	1	39.187	235.120
Jefe de Fiscalización	4	35.900	861.610
Profesional TI	16	17.300	1.660.810
Fiscalizador Senior	16	17.920	1.720.330
Fiscalizador Profesional	16	10.729	1.030.070
Tesista	40	3.000	2.880.000

Por otro lado, se solicita el uso de un computador, el cual fue adquirido con un valor de **CLP\$1.169.000**. En resumen, la inversión total del proyecto es de **CLP\$9.556.950**.

8.2.8 Consideración Social de Beneficios

Si bien el beneficio social no es calculado ni evaluado en el flujo de caja, es importante destacar la importancia social que posee la acción fiscalizadora en Chile, donde la DT es el pilar de apoyo para los trabajadores que son vulnerados en sus derechos. Estos beneficios vienen otorgados por las mejoras de las condiciones de trabajo o los respectivos problemas incurridos en base a las causales de la infracción. Dentro de estos beneficios se encuentran:

- El beneficio por el pago correcto de las cotizaciones previsionales a los trabajadores afectados.
- Beneficios por la existencia de contrato en términos legales.
- Beneficio por el pago correcto de las horas extra-programáticas a los trabajadores.
- Beneficios por condiciones higiénicas laborales mínimas.

8.3 Flujo de Caja

Para el análisis del flujo de caja, se realiza un flujo por cada escenario, donde posteriormente se calcula el CAE de cada uno, para finalmente dividirlo por la cantidad de fiscalizaciones efectivas, obteniendo el Costo Efectivo por Fiscalización, o CEF.

Para el escenario sin proyecto, se puede observar en la Tabla 25 el flujo (en miles de pesos) por parte de la cantidad de fiscalizaciones y su costo unitario calculado previamente, donde se obtiene un CAE de **CLP\$6.362.290.070** y un CEF de **CLP\$133.650** (ver Tabla 26).

Tabla 25: Flujo de caja, escenario sin proyecto.

Escenario sin Proyecto	Año 0	Año 1	Año 2	Año 3
Fiscalizaciones Totales		129004	131031	133057
Fiscalizaciones Proactivas		35863	36427	36990
Fiscalizaciones Reactivas		93141	94604	96067
Flujos Finales	\$-	\$(6,643,706.00)	\$(6,748,096.50)	\$(6,852,435.50)

Tabla 26: Indicadores, escenario sin proyecto.

Tasa descuento	6%
VAN	\$(17,006,477.37)
CAE	\$(6,362,290.07)
CEF	\$(133.65)

Para el flujo de caja con proyecto, se observa en la Tabla 27 el cual entrega un CAE **CLP\$6.385.555.520**, ligeramente más alto que el CAE sin proyecto, debido a la inversión y los costos operacionales extras, y entrega un CEF de **CLP\$104.810** el cual es menor al CEF sin proyecto.

Según lo planteado previamente en el tipo de evaluación a realizar, el criterio de aprobación del proyecto era si el CEF con proyecto era menor al CEF sin proyecto, se realiza, debido al aumento de efectividad en términos de costos. Por ende, según esta evaluación **se recomienda la realización del proyecto** en base a las variables analizadas.

Tabla 27: Flujo de Caja, escenario con proyecto.

Escenario Con Proyecto	Año 0	Año 1	Año 2	Año 3
Fiscalizaciones Totales		129004	131031	133057
Fiscalizaciones				
Proactivas		51602	55033	58678
Fiscalizaciones Reactivas		77402	75998	74379
Costos Operacionales				
Mantención (1 hr TI)		\$830.40	\$830.40	\$830.40
Ingeniero BI		\$20,255.63	\$20,255.63	\$20,255.63
Inversión (6 meses)	\$9,556.95			
Subdirector (1 hr x mes)	\$235.12			
Jefe de Fiscalización (4 hrs x mes)	\$861.61			
Profesional TI (16 hrs x mes)	\$1,660.81			
Fiscalizador (16 hrs x mes)	\$1,720.33			
Fiscalizador (16 hrs x mes)	\$1,030.07			
Tesista	\$2,880.00			
Computador	\$1,169.00			
		\$(6,664,792.03)	\$(6,769,182.53)
Flujos Finales	\$(9,556.95)		\$(6,769,182.53)	\$(6,873,521.53)

Tabla 28: Indicadores, escenario con proyecto.

Tasa descuento	6%
VAN	\$(17,068,666.21)
CAE	\$(6,385,555.52)
CEF	\$(104.81)

8.4 Análisis de Sensibilidad

De acuerdo a las variables planteadas previamente, la que afecta principalmente a la efectividad del proyecto es la tasa de asertividad de las fiscalizaciones proactivas, la cual según objetivo se define que será el doble de la histórica, por lo cual se realizó un análisis *What If* en el software Excel, donde se detallan escenarios moviendo la tasa de asertividad y su crecimiento anual, otorgando el indicador CEF dentro de su matriz. La Tabla 29 muestra que, si la tasa de asertividad se mantiene igual con un crecimiento bajo el 20 %, el proyecto no se debería realizar debido a que el CEF es mayor que la situación sin proyecto. Por otro lado, si se aumenta a un 30 % de Asertividad, para crecimientos bajo el 6%, el proyecto es nuevamente no conveniente en términos del CEF. Por otro lado, asertividades sobre el 32%, sin un crecimiento necesario anual, ya muestran CEF menor al de sin proyecto, siendo este un espacio de aprobación en términos de las dos variables revisadas.

Tabla 29: Simulación de casos para tasas contingentes.

\$	(104.81)	26%	30%	35%	40%	45%	50%	55%						
0%	\$	(141.66)	\$	(135.05)	\$	(127.62)	\$	(120.96)	\$	(114.96)	\$	(109.52)	\$	(104.58)
2%	\$	(140.72)	\$	(134.07)	\$	(126.59)	\$	(119.91)	\$	(113.89)	\$	(108.45)	\$	(103.50)
4%	\$	(139.78)	\$	(133.09)	\$	(125.57)	\$	(118.86)	\$	(112.83)	\$	(107.38)	\$	(102.43)
6%	\$	(138.84)	\$	(132.11)	\$	(124.55)	\$	(117.82)	\$	(111.77)	\$	(106.32)	\$	(101.37)
8%	\$	(137.90)	\$	(131.13)	\$	(123.54)	\$	(116.78)	\$	(110.72)	\$	(105.26)	\$	(100.32)
10%	\$	(136.96)	\$	(130.15)	\$	(122.52)	\$	(115.75)	\$	(109.68)	\$	(104.21)	\$	(99.27)
12%	\$	(136.03)	\$	(129.17)	\$	(121.52)	\$	(114.72)	\$	(108.64)	\$	(103.17)	\$	(98.23)
14%	\$	(135.09)	\$	(128.20)	\$	(120.51)	\$	(113.70)	\$	(107.61)	\$	(102.14)	\$	(97.20)
16%	\$	(134.16)	\$	(127.23)	\$	(119.51)	\$	(112.68)	\$	(106.59)	\$	(101.12)	\$	(96.18)
18%	\$	(133.22)	\$	(126.26)	\$	(118.52)	\$	(111.67)	\$	(105.57)	\$	(100.10)	\$	(95.17)
20%	\$	(132.29)	\$	(125.30)	\$	(117.53)	\$	(110.66)	\$	(104.56)	\$	(99.09)	\$	(94.17)
22%	\$	(131.36)	\$	(124.33)	\$	(116.54)	\$	(109.67)	\$	(103.56)	\$	(98.09)	\$	(93.18)
24%	\$	(130.43)	\$	(123.38)	\$	(115.56)	\$	(108.67)	\$	(102.56)	\$	(97.10)	\$	(92.19)
26%	\$	(129.51)	\$	(122.42)	\$	(114.58)	\$	(107.69)	\$	(101.58)	\$	(96.12)	\$	(91.22)
28%	\$	(128.59)	\$	(121.47)	\$	(113.61)	\$	(106.71)	\$	(100.60)	\$	(95.15)	\$	(90.25)
30%	\$	(127.67)	\$	(120.52)	\$	(112.65)	\$	(105.74)	\$	(99.62)	\$	(94.18)	\$	(89.30)
32%	\$	(126.75)	\$	(119.58)	\$	(111.69)	\$	(104.77)	\$	(98.66)	\$	(93.22)	\$	(88.35)
34%	\$	(125.84)	\$	(118.64)	\$	(110.73)	\$	(103.81)	\$	(97.70)	\$	(92.27)	\$	(87.42)
36%	\$	(124.93)	\$	(117.71)	\$	(109.78)	\$	(102.86)	\$	(96.76)	\$	(91.34)	\$	(86.49)
38%	\$	(124.02)	\$	(116.78)	\$	(108.84)	\$	(101.91)	\$	(95.82)	\$	(90.41)	\$	(85.57)
40%	\$	(123.11)	\$	(115.86)	\$	(107.91)	\$	(100.98)	\$	(94.88)	\$	(89.48)	\$	(84.67)
42%	\$	(122.21)	\$	(114.94)	\$	(106.98)	\$	(100.05)	\$	(93.96)	\$	(88.57)	\$	(83.77)
44%	\$	(121.32)	\$	(114.02)	\$	(106.05)	\$	(99.12)	\$	(93.04)	\$	(87.67)	\$	(82.88)
46%	\$	(120.42)	\$	(113.11)	\$	(105.13)	\$	(98.21)	\$	(92.14)	\$	(86.77)	\$	(82.00)
48%	\$	(119.53)	\$	(112.21)	\$	(104.22)	\$	(97.30)	\$	(91.24)	\$	(85.89)	\$	(81.13)
50%	\$	(118.65)	\$	(111.31)	\$	(103.32)	\$	(96.40)	\$	(90.35)	\$	(85.01)	\$	(80.27)

Debido a que el piloto entrega resultados promedios de un 43% de infraccionalidad detectada, dado el análisis de sensibilidad se comprueba que es económicamente factible al ser mayor al 30% de asertividad.

CAPÍTULO 9: CONCLUSIONES

El servicio que proporciona la Dirección del Trabajo a sus usuarios es de vital importancia para un entorno laboral justo y ameno, donde es esta organización la cual vela por el cumplimiento de la ley laboral y el entendimiento de esta. Debido a ser el único servicio encargado de fiscalizar en el mercado laboral posee ineficiencias de operación en uno de sus ejes principales, la Fiscalización, la cual ha presentado ser una fiscalización basada en la reacción ante denuncias de sus usuarios sobre una fiscalización proactiva, siendo las fiscalizaciones reactivas 3 veces más que las proactivas. Esto posee grandes complicaciones, donde el hecho de que haya un alto número de denuncias las cuales son por la llegada de un usuario a reclamar (considerable aleatoria) provoca un desajuste de los programas planificados debido a que se debe atender un número de requerimientos nuevos no planificados (denuncias). Además, las fiscalizaciones proactivas poseen una eficacia baja del 27% aprox. donde se están destinando recursos a fiscalizar empresas las cuales no son proclives a la infracción y no existe un foco a las empresas más riesgosas en base a análisis de información pasada, sino que se le deje (mayoritariamente) a criterio del fiscalizador la elección de la empresa a visitar, lo que trae consigo una serie de incentivos perversos. De esto nace la oportunidad de rediseñar el proceso de fiscalización proactiva conjunto a la incorporación de modelos de Minería de Datos para focalizar a las empresas con una mayor probabilidad de ser infractoras y de esta manera utilizar de forma más eficiente los recursos limitados de la organización.

Se recopiló información de los procesos y sus principales actores a partir de documentos entregados por la organización, pero además se entró en detalle en reuniones y entrevistas con los actores para conocer de manera más detallada estos procesos, lo cual relevante para el entendimiento de la organización y la dinámica de estos procesos. Con esto se levanta la arquitectura y estructura de procesos de la organización y se puede conceptualizar el trabajo realizado de una manera macro. Dado esto, se pudo identificar los procesos que provocan las fallas que han causado los problemas dentro de los ejes, donde se enfocan los esfuerzos y la dirección del rediseño con el fin de generar un proceso en base a un modelo que promueva un cambio conductual en las empresas y no solo un modelo en base a esperar que los usuarios lleguen a denunciar por irregularidades. Conjunto a esto, se revisó el caso del SII el cual posee un modelo de gestión de cumplimiento tributario el cual segmenta a las empresas según su comportamiento y realiza estrategias respecto al segmento al cual correspondan, donde esto se quiere simular en la DT, con el fin de generar un tratamiento del comportamiento según la información de las empresas respecto a fiscalizaciones pasadas. En base a este modelo, se desea desarrollar el modelo de datos y data mining que acompañen el correcto enfoque a empresas prospectas a la infracción.

Además, se ha estudiado sobre la Minería de Datos y los modelos que esta ciencia trae consigo para el posible enfoque que se desarrollará en la

organización, donde se iteró con datos de prueba para obtener primeros alcances de lo que se necesitará como prueba de conceptos para la obtención de un ajuste que sea preciso bajo criterios otorgados por la literatura y finalmente lograr el objetivo de negocios el cual es mejorar eficacia de las fiscalizaciones y rediseñar el proceso a partir de este modelo.

Una de las principales conclusiones del trabajo realizado es que en la DT existe una inercia organizacional basada en la experiencia de los fiscalizadores dada la cantidad de formularios y papeles a rellenar al momento de realizar una fiscalización, por lo cual esto genera un incentivo para la elección de empresas previamente fiscalizadas u otros criterios que faciliten la negligencia de realizar el trabajo inspectivo, lo cual provoca que la eficacia de estas decaiga y no se esté encontrando a las empresas que efectivamente sean infractoras y que finalmente, llegue un usuario a reclamar este hecho. Para esto se propone lo planteado en el informe y lo explicado previamente y se pretende disminuir la cantidad de denuncias a partir de una fiscalización proactiva enfocada.

Otro tema importante, es la resistencia al cambio y las bajas expectativas por el proyecto a desarrollarse, donde se puede observar a partir de las reuniones tenidas con los actores principales que involucra el proyecto BID, existe poca disposición a ser parte de un cambio, el cual, si bien puede ser mejor para la organización, puede ser invasivo para los actores dado que afectaría sus prácticas operacionales del día a día.

Por otro lado, la DT posee un sistema tecnológico que permite un buen flujo de datos entre los operadores (fiscalizadores particularmente), basado en una arquitectura de 3 capas, la cual permite una mejor integración con el modelo a desarrollar, donde este rediseño no sería un cambio radical en el modo de visualización y tratamiento de las fiscalizaciones, permitiendo un aprendizaje más sencillo y ameno para los principales actores de este proceso.

Si bien se obtuvo una buena aproximación en términos de la creación de programas de fiscalización con una nueva capacidad predictiva en base a modelos y dado el piloto realizado, el cual obtuvo un resultado general de un 43% de asertividad, la DT posee una gran cantidad de trabajos asociados a mejorar el proceso de fiscalización completo, donde se puede incorporar automatizaciones e inteligencia a tareas que actualmente son realizados por humanos, como por ejemplo la asignación automática e inteligente de recursos o reglas automatizadas para la selección de requerimientos a fiscalizar con el fin de disminuir la cantidad de fiscalizaciones realizadas por mal ingreso de denuncia, entre otras. También existe una necesidad de construcción de un nuevo set de Indicadores de medición de procesos y resultados, dado el nuevo enfoque que desea crear la organización, donde la fiscalización focalizada sea la principal herramienta de detección, por lo que actualmente, no existen indicadores que reflejen e incentiven el cumplir tal propósito.

Lecciones Aprendidas

Dentro de las principales lecciones aprendidas dentro del proyecto, se encuentra el trabajo directo con profesionales respecto al tema a focalizar, dado que la aplicación de técnicas no es efectiva cuando no existe un entendimiento del negocio, lo cual, a través de las diferentes iteraciones realizadas a lo largo del trabajo, se pudo interpretar los resultados y tomar decisiones acordes al análisis. Esto produjo que el trabajo de preparación de datos y modelamiento tomara más tiempo, pero evitó obtener resultados incorrectos o desalineados con el objetivo del proyecto.

La metodología de Ingeniería de Negocios permite observar los procesos de manera general y permiten, en base a lógicas probadas en diferentes industrias, agrupar los procesos en diferentes macroprocesos. Esto permite identificar de manera directa los procesos involucrados en las organizaciones, dando paso al rediseño a través de los patrones identificados por Barros.

Por otro lado, el utilizar una metodología reconocida como CRISP-DM permitió, de forma estructurada, generar un proyecto de minería de datos incorporando las diversas etapas necesarias en sus distintas fases. Además, el incorporar estudios y trabajos similares en otros tópicos, permite guiar el trabajo a desarrollar y tener expectativas respecto a lo que se obtiene y los diversos problemas que este tipo de proyectos puede presentar.

Una parte importante dentro del proceso es la comunicación del proyecto, lo cual es fundamental para generar comprensión y conocimiento sobre el tema, dado que dentro de una organización basada en temas legales y con actores poco relacionado con modelos de minería de datos, la incorporación de estas técnicas puede ser muy beneficioso, pero se hace necesaria la difusión y la obtención temprana de resultados.

El trabajo desarrollado permite afirmar que en base a la información que posee el servicio, se pueden desarrollar modelos analíticos que permitan mejorar los niveles de detección de la DT y que pueden ser aún mejores si se obtuviese una mayor granularidad en términos de características de una empresa y comportamiento en otros servicios públicos. Sin embargo, es importante señalar que este esfuerzo mejorará el rendimiento de la organización durante los próximos años (5 – 10 años), pero es necesario generar un cambio importante en términos del tratamiento de los infractores, dado que las variables más importantes dentro del modelamiento son la infraccionalidad histórica, donde las empresas prefieren volver a ser infractoras, independiente si son descubiertos o no. Este cambio se alinea a cumplir de mejor manera la misión institucional, donde los esfuerzos de hallazgo de empresas infractoras vallan en conjunto a esfuerzos de cambio conductual, donde uno de los ejemplos a seguir es el SII con su plan de cumplimiento tributario.

Trabajos Futuros

El trabajo desarrollado en el proyecto se basó en rediseñar el proceso de fiscalización proactiva en base a programas creados por modelos predictivos.

Estos modelos fueron desarrollados en base a la información recaudada por la organización, la cual posee diferentes problemas. Dentro de los problemas se presentan oportunidades para generar proyectos que mejoren este rediseño en base a modelos de inteligencia de negocios. Dentro de estos trabajos de mejoramiento se encuentran:

- Probar Modelos con una caracterización más específica, dado que en la mayoría de los resultados de Región y Actividad Económica no presentaron ser importantes para las decisiones de corte, por lo cual se plantea desarrollar modelos Regionales a través de las comunas e incorporar el sub-rubro de las actividades.
- Incorporar relaciones entre industrias tipo redes sociales, con el fin de identificar posibles movimientos generados por núcleos productivos.
- Incorporar variables macroeconómicas al modelamiento, con el fin de analizar variables exógenas a las empresas.
- Generar un análisis profundo respecto a las materias infraccionadas, donde existen más de dos mil distintos tipos de infracciones a multar, donde en diferentes actividades económicas pueden tener mayor probabilidad de infracción.
- En términos más operativos se propone realizar un modelo de optimización para la programación de fiscalizaciones respecto a características de los fiscalizadores en términos de especialización y capacitación.

CAPÍTULO 10: BIBLIOGRAFÍA

- Araya, J. (2012). *Rediseño de Procesos de la Fiscalización Preventiva*, Dirección Nacional del Trabajo. Tesis de Magíster, Universidad de Chile, Santiago.
- Barros, O. (2012). *Igeniería de Negocios Diseño Integrado de Negocios, Procesos y Aplicaciones TI*. Chile: Universidad de Chile.
- Barros, O. (2017). *Business Engineering and Service Design* (Second ed., Vol. I). New York: Business Expert Press.
- Castellón, P., & Velásquez, J. (2011). Caracterización de Contribuyentes que presentan facturas falsas al SII mediante técnicas de Data Mining. *Revista Ingeniería de Sistemas, XXV*.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0 Step by step Data Mining guide*. CRISP-DM Consortium. SPSS.
- Contreras, E. (2004). *Evaluación Social de Inversiones: Enfoques Alternativos y su Aplicabilidad para Latinoamérica*. (CEPAL, Ed.) Santiago, Chile: Naciones Unidas.
- Dirección del Trabajo. (2017). *DT - Portal Institucional*. Dirección del Trabajo. Gobierno de Chile. Obtenido de Dt.gob.cl: <http://www.dt.gob.cl/1601/w3-channel.html>
- Echeverría, M. (2015). *Diseño del proceso de Monitoreo a distancia y de predicción del riesgo de crisis en pacientes con hospitalización ambulatoria domiciliaria*, Hospital Exequiel González Cortéz. Tesis de Magíster, Universidad de Chile, Santiago.
- Garrido, N. (2015). *Mejora en los procesos de planificación, control y monitoreo del servicio de urgencia del hospital clínico de la Universidad de Chile*. Tesis de Magíster, Universidad de Chile, Santiago.
- Gobierno de Chile. (2017). *Gobierno Transparente*. Obtenido de Dt.gob.cl: <http://www.dt.gob.cl/transparencia/>
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques* (Third Edition ed.). USA: Elsevier.
- Hax, A. (2010). *The Delta Model, Reinventing your Business Strategy*. New York, USA: Springer.
- Hothron, T., Hornik, K., & Zeileis, A. (Septiembre de 2016). *Conditional Inference Trees*. Obtenido de CRAN: <https://cran.r-project.org/web/packages/partykit/vignettes/ctree.pdf>
- Hsueh, S., Huang, C., & Tseng, C. (2013). Using Data Mining technology to explore labor safety strategy - A lesson from the construction Industry. *Pakistan Journal of Statistics*.

- Kanbur, R., Ronconi, L., & Wedenoja, L. (2013). Incumplimiento a la Normativa laboral en Chile. *International Labour Review*, 153, 431-444.
- Liao, C., & Perng, Y. (2008). Data Mining for occupational injuries in the Taiwan construction Industry. *Procedia Engineering*, XXIX, 3240-3244.
- MDS. (2017). *Ministerio de Desarrollo Social*. Obtenido de Sni.ministeriodesarrollosocial.gob.cl:
<http://sni.ministeriodesarrollosocial.gob.cl/download/precios-sociales-vigentes-2017/?wpdmdl=2392>
- Olguín, E., Crawford, B., & Soto, R. (2016). Change Management in Technology Projects Using a Comprehensive Model of Change Management. *Sistemas y Tecnologías de la Información, I*.
- SII. (2016). *Servicio de Impuestos Internos*. Obtenido de Sii.cl:
http://www.sii.cl/sobre_el_sii/Plan_Cumplimiento_tributario2015.pdf
- SII. (2017). *Estadísticas de empresa*. Obtenido de Sii.cl:
http://www.sii.cl/estadisticas/empresas_segun_vigencia.htm
- Torgo, L. (2011). *Data Mining with R, Learning with Case Studies*. Minnesota, USA: Vipin Kumar.
- Vijay, K. (2015). *Predictive Analytics and Data Mining, Concepts and Practice in RapidMiner*. USA: Elsevier.
- Xu, S., & Chen, L. (2008). A novel approach for Determining the Optimal number of hidden layer neurons for Forward Neural Networks and its applications in Data Mining. *5th International Conference of Information Technology and applications*.
- Zhang, G. P. (2007). Avoiding Pitfalls in Neural Network Research. *IEEE Transactions on systems, man, and cybernetics, III*.

CAPÍTULO 11: ANEXOS

11.1 Diagramas Entidad Relación

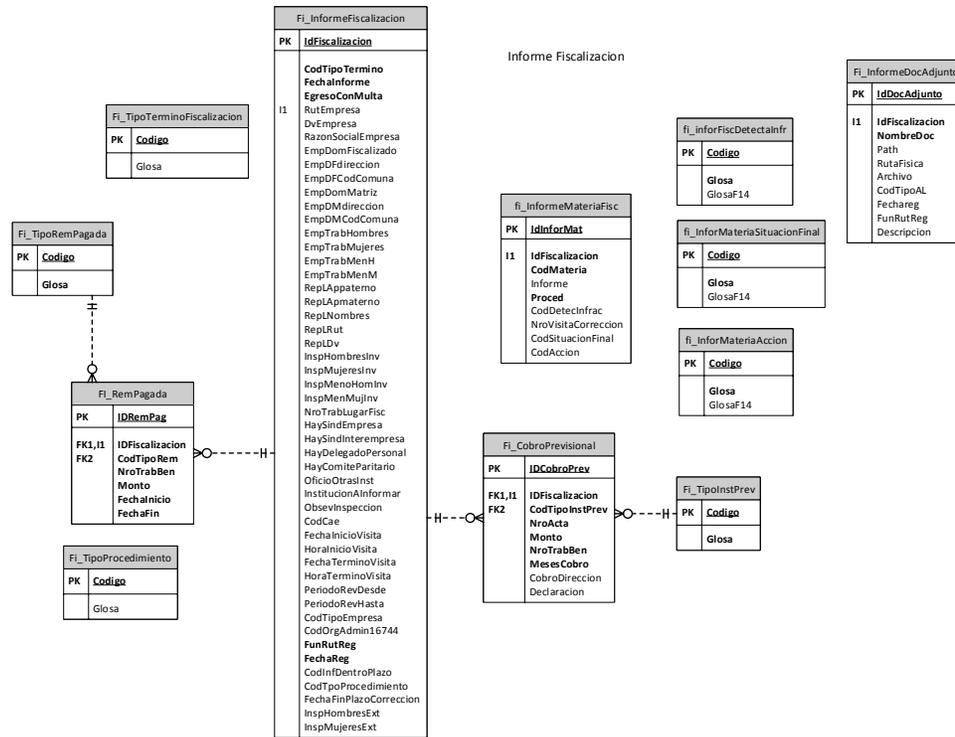


Figura 48: Diagrama E-R de Informes de Fiscalización

Ingreso de Fiscalización

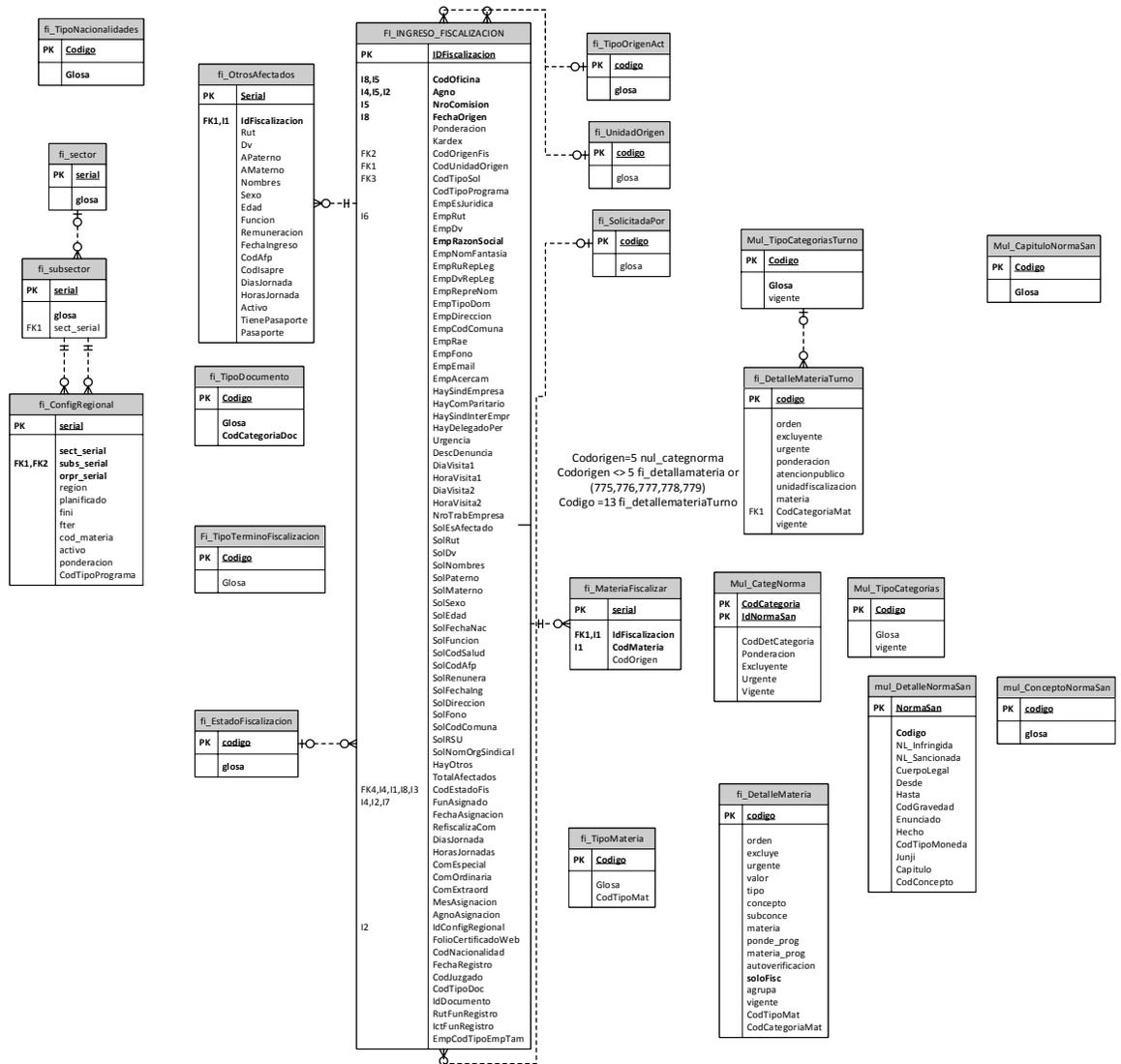


Figura 49: Diagrama E-R Ingreso de Fiscalización.

11.2 Coeficientes Modelo Logit

	Estimate	Std.	Error	z\Value
(Intercept)	-3.442339	0.091421	-37.654	2.00E-16
crae102	0.064387	0.165361	0.389	0.697001
crae103	0.279018	0.134019	2.082	0.037349
crae104	0.331069	0.061838	5.354	8.61E-08
crae105	0.068091	0.182506	0.373	0.709082
crae106	-0.200388	0.065689	-3.051	0.002284
crae107	0.228109	0.057085	3.996	6.44E-05
crae108	0.326277	0.063526	5.136	2.80E-07
crae109	0.258187	0.062449	4.134	3.56E-05
crae110	0.279562	0.106586	2.623	0.008719
crae111	-0.004855	0.061393	-0.079	0.936965
crae112	0.10702	0.194056	0.551	0.581299
crae113	0.495936	0.07339	6.758	1.40E-11
crae114	0.322772	0.09937	3.248	0.001161
crae115	-0.381286	0.068732	-5.547	2.90E-08
crae116	-0.2464	0.094407	-2.61	0.009055
crae117	-10.102419	60.407149	-0.167	0.867182
Region2	0.406545	0.092145	4.412	1.02E-05
Region3	-0.152338	0.113518	-1.342	0.179605
Region4	0.038142	0.095297	0.4	0.68898
Region5	0.107594	0.08139	1.322	0.186182
Region6	0.313963	0.086887	3.613	0.000302
Region7	0.152471	0.087748	1.738	0.082281
Region8	0.048365	0.082918	0.583	0.559701
Region9	0.196319	0.08845	2.22	0.026449
Region10	0.192771	0.086647	2.225	0.026096
Region11	0.087316	0.141877	0.615	0.538265
Region12	0.362266	0.118538	3.056	0.002242
Region13	0.032463	0.075918	0.428	0.66894
Region14	0.376779	0.108238	3.481	0.0005
Region15	0.266877	0.123223	2.166	0.030325
Agno_01	0.017794	0.028265	0.63	0.528994
inf_06	-0.04508	0.044282	-1.018	0.308671
Agno_02	0.080877	0.018623	4.343	1.41E-05
inf_07	-0.115436	0.030035	-3.843	0.000121
Agno_03	0.032422	0.01821	1.78	0.074999
inf_08	-0.052892	0.030546	-1.732	0.08335
Agno_04	0.119422	0.016856	7.085	1.39E-12
inf_09	-0.130767	0.028269	-4.626	3.73E-06
Agno_05	0.02878	0.017109	1.682	0.092534
inf_10	-0.043485	0.029947	-1.452	0.146489
Agno_06	0.03277	0.01702	1.925	0.054178
inf_11	0.025859	0.029354	0.881	0.378356
Agno_07	0.069173	0.017849	3.875	0.000106
inf_12	0.052145	0.031073	1.678	0.093321
Agno_08	0.234549	0.017957	13.061	2.00E-16
inf_13	-0.001141	0.032045	-0.036	0.971604
Agno_09	0.597534	0.018032	33.137	2.00E-16
inf_14	0.036696	0.032223	1.139	0.254787
fiscsd	-0.719191	0.050747	-14.172	2.00E-16
infsd	1.203776	0.086926	13.848	2.00E-16
infrac	-0.304497	0.041517	-7.334	2.23E-13

Figura 51: Coeficientes Regresión Logística.