



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

**DESARROLLO E IMPLEMENTACIÓN DE INDICADOR DE SEGUIMIENTO DE  
OBRAS DE GENERACIÓN Y TRANSMISIÓN EN UN MERCADO ELÉCTRICO**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL ELECTRICISTA

EUGENIO ERNESTO QUINTANA PAINEMAL

PROFESOR GUÍA:

RODRIGO PALMA BEHNKE

MIEMBROS DE LA COMISIÓN:

ARIEL VALDENEGRO ESPINOZA

CARLOS BENAVIDES FARIAS

SANTIAGO DE CHILE

Mayo de 2012

## **DESARROLLO E IMPLEMENTACIÓN DE INDICADOR DE SEGUIMIENTO DE OBRAS DE GENERACIÓN Y TRANSMISIÓN EN UN MERCADO ELÉCTRICO**

El objetivo general de este trabajo es sistematizar, a partir de la información pública disponible, los grados de avance de proyectos de generación y transmisión en un sistema eléctrico, de manera de poder establecer un juicio sobre la fecha de puesta en servicio de dichas obras. Esto con el fin de que exista una herramienta que permita clasificar a los proyectos en base a modelos matemáticos.

Dada la naturaleza de la información disponible se utiliza la metodología de KKD para obtener conocimiento a través de la información disponible, lo que permite estimar la probabilidad de que una determinada variable nominal pertenezca a cada una de las categorías que puede tomar la variable en cuestión. Para determinar esta probabilidad primero se seleccionan los atributos que permitan explicar la variable en estudio, para luego determinar los coeficientes que acompañan a cada una de las categorías de los atributos seleccionados.

Para la construcción de los modelos se utiliza la información pública disponible, por lo que en caso de disponer de mayor información, los resultados del modelo pueden cambiar y mejorar. De los resultados obtenidos se concluye que el acierto de los modelos obtenidos aplicados a los datos históricos supera el 70% en ambos casos. Además de entregar una predicción de la clase en estudio, se debe entregar la probabilidad que tiene la variable en estudio de pertenecer a cada categoría, ya que en los casos en que el modelo no entregó una predicción correcta fue debido a que la probabilidad que tenía esa categoría no era muy alta o era muy parecida a otra. Pese a que modelos de redes neuronales o árboles de decisión no entregan probabilidad, se utilizaron para comparar los resultados obtenidos, entregando estos dos últimos modelos mejores resultados en el caso que se estudia la materialización de los proyectos (cerca de un 10% de más aciertos).

Finalmente se desarrolla un conjunto de herramientas gráficas que permiten visualizar los resultados obtenidos, además de observar los estadísticos más comunes.

*Dedicada a todos mis seres queridos,  
en especial a Eugenio Agustín.*

# Agradecimientos

Primero que todo agradecer a mi familia, en especial a mis padres Eugenio y Mireya, por todo el esfuerzo que hicieron desde que yo era muy niño para que siempre hubiera un ambiente adecuado para el estudio y para que nunca faltara un nuevo libro que leer en casa. Además de ello, aprecio la importancia que le daban a estudiar, al cultivo del conocimiento en general y a una formación integral. También agradecer a mi abuela Blanca que siempre se preocupó de mí, brindándome todo su amor y ayuda, lo que siempre agradeceré de todo corazón.

También agradecer a Pía y a su madre Marlene, por el apoyo que me prestaron, especialmente cuando se acercaba la entrega de este trabajo. Gracias a mi hijo Eugenio Agustín, que con su sola presencia me motivaba a seguir adelante, incluso en los momentos en los que ya no daba más y sólo quería bajar los brazos.

Además agradecer a mis compañeros de carrera, en particular a los muchachos con los que compartí tardes de estudio, partidos de baby fútbol en la multicancha, Electrocopas junto a la Tacorne, asados y carretes. Aunque no lo crean, siempre se agradece un momento ameno de relajación y ocio en nuestra ajetreada y demandante carrera. Son tantas las personas con las que compartí, que no alcanza esta hoja para listarlos a todos, pero sólo por nombrar algunos: Monje, Salinas, Moller, Samir, Erick, Espinoza, Abelino, Serena, Nico, Calderón, Claudio, Fehlandt, Pato, Guerrero, Felipe, Rodrigo, y otros muchos más.

Con respecto a la realización de este trabajo, agradecer a Carlos Benavides por sus consejos y orientaciones, que me fueron de utilidad. Además agradecer a Francisco Orduña por la orientación que me dio en la utilización de las herramientas de Tecnologías de la Información, que fue un valioso aporte para la automatización de las pruebas que tuve que realizar.

Finalmente agradecer al Profesor Rodrigo Palma por invitarme a participar de un proyecto tan interesante como fue la implementación de un sistema de monitoreo de mercado energético, que abrió mis horizontes de conocimientos y me permitió acercarme a una dimensión desconocida como lo es la integración de distintas ramas del saber con un único objetivo en común.

# Índice de contenidos

RESUMEN DE LA MEMORIA.....	i
Agradecimientos .....	iii
Índice de contenidos.....	iv
Índice de figuras.....	vii
Índice de tablas.....	ix
Glosario.....	x
Capítulo 1 – Introducción.....	1
1.1. Motivaciones .....	1
1.2. Objetivos .....	3
1.2.1. Objetivo general .....	3
1.2.2. Objetivos específicos .....	3
1.3. Alcances .....	4
1.4. Estructura del documento.....	4
Capítulo 2 –Antecedentes .....	6
2.1. Mercado eléctrico chileno .....	6
2.2. Demanda futura de energía .....	6
2.3. Institucionalidad.....	7
2.3.1. Ministerio de Energía.....	7
2.3.2. Ministerio del Medio Ambiente .....	9
2.4. Estado del arte de los sistemas de monitoreo.....	10
2.4.1. Literatura científica .....	10
2.4.2. Índice de credibilidad de las inversiones .....	10
2.4.3. Casos de otros países.....	11
2.5. Sistema de monitoreo de los mercados energéticos en Chile .....	12
2.6. Etapas de un proyecto eléctrico.....	13
2.6.1. Pre-factibilidad.....	13

2.6.2. Factibilidad.....	13
2.6.3. Inversión y construcción .....	14
2.6.4. Operación .....	14
2.7. Permisos aplicables a los proyectos .....	14
2.7.1. Permisos que presentan dificultades e instituciones responsables .....	16
2.8. Fuentes de información .....	20
2.8.1. Prensa especializada.....	20
2.8.2. Sistema de Evaluación de Impacto Ambiental.....	20
2.8.3. Informe de Precio de nudo de corto plazo .....	21
2.8.4. Centros Económicos de Despacho de Carga (CDEC) .....	26
2.9. Implicancias de los atrasos en el sistema .....	26
2.9.1. Programación de Largo Plazo (PLP).....	26
2.9.2. Estudio de Transmisión Troncal (ETT) .....	27
2.9.3. Precio Básico de la Energía.....	28
Capítulo 3 – Metodología propuesta .....	29
3.1. Descripción general de la metodología .....	29
3.2. Detalle de etapas .....	31
3.2.1. Extracción de la información .....	31
3.2.2. Almacenamiento y transformación .....	34
3.2.3. Obtención de modelos.....	35
3.2.4. Formas de visualización .....	38
3.3. Aspectos de la implementación.....	41
3.3.1. Softwares utilizados .....	41
3.3.2. Esquema del procesamiento.....	42
Capítulo 4 – Resultados .....	44
4.1. Aplicación de la metodología.....	44
4.1.1. Atraso de las obras .....	44
4.1.2. Materialización de la construcción.....	51
4.2. Visualización.....	56
4.2.1. Resultados .....	56
4.2.2. Árboles de decisión.....	59

4.2.3. Diagramas de cajas.....	60
4.2.4. Histogramas .....	63
Capítulo 5 – Análisis y discusiones .....	65
5.1. Análisis de resultados.....	65
5.1.1. Importancia de la selección de atributos .....	65
5.1.2. Atraso de las obras .....	65
5.1.3. Materialización de la construcción.....	66
5.2. Discusiones .....	68
5.2.1. Base de datos.....	68
5.2.2. Metodología .....	69
5.2.3. Modelos.....	69
Capítulo 6 – Conclusiones .....	71
Capítulo 7 – Bibliografía.....	73
Capítulo 8 – Anexos.....	76
8.1. Anexo 1: Inteligencia de Negocios .....	76
8.1.1. Bases de datos multidimensionales.....	76
8.1.2. Justificación de los softwares utilizados .....	77
8.2. Anexo 2: Knowledge Discovery in Databases (KDD) .....	77
8.3. Anexo 3: Atributos cualitativos de los proyectos eléctricos .....	78
8.4. Diagramas de cajas.....	79

# Índice de figuras

Figura 2.1 Etapas de un proyecto y lugar donde se ubica el ingreso al SEIA.....	14
Figura 2.2 Permisos sectoriales que tienen como pre-requisitos otros permisos. Elaboración propia a partir de la información obtenida de [18]......	16
Figura 2.3 Atrasos que han presentados los anuncios de las centrales a carbón que se están construyendo en la región del Bío Bío .....	22
Figura 2.4 Atraso en los anuncios que han presentado las últimas centrales a carbón construidas en el SING .....	23
Figura 2.5 Atraso en los anuncios que han presentado los parques eólicos construidos en el SIC23	
Figura 2.6 Atraso en los anuncios que han presentado dos grandes proyectos de Tinguiririca S.A. ....	24
Figura 2.7 Atraso en los anuncios que han presentado centrales hidroeléctricas de menos de 100 MW en el SIC.....	24
Figura 2.8 Atraso que han presentado grandes proyectos hidroeléctricos de grandes empresas en el SIC .....	25
Figura 2.9 Atrasos que han presentado turbinas de respaldo para el SIC .....	25
Figura 3.1 Esquema de la metodología propuesta.....	29
Figura 3.2 Diagrama de caja.....	40
Figura 3.3 Histograma .....	40
Figura 3.4 Esquema de implementación del procesamiento de la información propuesta.....	<b>¡Error!</b>
<b>Marcador no definido.</b>	
Figura 4.1 Resultados de aplicar el modelo de regresión logística a las obras que se encuentran en construcción en el Informe de Precio de Nudo de Octubre de 2011 .....	57
Figura 4.2 Tabla que muestra las predicciones hechas por el modelo, entregando un 1 si la mayor probabilidad es superior al 50% .....	58
Figura 4.3 Visualización de aplicar el modelo de regresión logística al caso de la materialización de los proyectos .....	59



Figura 4.4	Árbol de decisión J48 aplicado al caso que trata la materialización de los proyectos .	60
Figura 4.5	Boxplot del atraso con respecto a las distintas regiones .....	61
Figura 4.6	Boxplot del atraso con respecto a la proyección o lejanía de la fecha proyectada vs la fecha en la que se hizo el anuncio .....	61
Figura 4.7	Boxplot del atraso con respecto al combustible principal que utiliza la central .....	62
Figura 4.8	Boxplot del atraso con respecto a si aparece como en construcción o recomendado en el plan de obras de la CNE .....	62
Figura 4.9	Boxplot del atraso con respecto al tipo de empresa dueña de la central .....	63
Figura 4.10	Histograma de los atrasos .....	64
Figura 8.1	Procedimiento general del proceso de KDD.....	78
Figura 8.2	Boxplot Atraso vs Tipo de aparición en plan de obras CNE .....	79
Figura 8.3	Boxplot Atraso vs Clasificación del proyecto.....	80
Figura 8.4	Boxplot Atraso vs Existencia previa de proyecto .....	80
Figura 8.5	Boxplot Atraso vs Zona .....	81
Figura 8.6	Boxplot Atraso vs Macrozona.....	81
Figura 8.7	Boxplot Atraso vs Tipo de empresa .....	82
Figura 8.8	Boxplot Atraso vs Región .....	82
Figura 8.9	Boxplor Atraso vs Potencia.....	83
Figura 8.10	Boxplot Atraso vs atributo PMG .....	83
Figura 8.11	Boxplot Atraso vs Fuente.....	84
Figura 8.12	Boxplot Atraso vs Atributo ERNC .....	84
Figura 8.13	Boxplot Atraso vs Combustible principal .....	85
Figura 8.14	Boxplot Atrasos vs Tecnología .....	85
Figura 8.15	Boxplot Atrasos vs Anuncios.....	86
Figura 8.16	Boxplot Atraso vs Proyección.....	86

# Índice de tablas

Tabla 4.1 Atributos candidatos a describir la Clase “Atraso”, junto con los niveles que pueden tomar.....	47
Tabla 4.2 p-valor resultante de los test chi-cuadrados realizados entre la clase atraso y los atributos pre-seleccionados.....	48
Tabla 4.3 Atributos seleccionados para el modelo de regresión multinomial.....	49
Tabla 4.4 Matriz de confusión del modelo de regresión logística multinomial obtenido. ....	50
Tabla 4.5 Comparación de los resultados de los distintos modelos aplicados al caso de los atrasos de los anuncios .....	51
Tabla 4.6 Atributos para el caso de estudio sobre la materialización de los proyectos. ....	53
Tabla 4.7 p-valor resultante de aplicar los test chi-cuadrado entre la clase construcción y cada uno de los atributos.....	54
Tabla 4.8 Atributos seleccionados para explicar la clase construcción.....	54
Tabla 4.9 Matriz de confusión para el caso de la materialización de los proyectos.....	55
Tabla 4.10 Tabla que compara los distintos modelos de predicción para el caso de estudio acerca de la materialización de los proyectos eléctricos.....	56
Tabla 8.1 Regiones y Zonas .....	78
Tabla 8.2 Sistema interconectado.....	79

# Glosario

CADE:	Comisión Asesora del Desarrollo Eléctrico
CDEC:	Centro de Despacho Económico de Carga
CNE:	Comisión Nacional de Energía
DFL:	Decreto con Fuerza de Ley
ENRC:	Energía Renovable No Convencional
e-seia:	Expedientes web de proyectos ingresados al SEIA
GN:	Gas Natural
KDD:	Knowledge Discovery in Databases
LGSE:	Ley General de Servicios Eléctricos
PMG:	Pequeño Medio de Generación
RCA:	Resolución de Calificación Ambiental
SEA:	Servicio de Evaluación Ambiental
SEIA:	Sistema de Evaluación de Impacto Ambiental
SIC:	Sistema Interconectado Central
SING:	Sistema Interconectado del Norte Grande
STT:	Sistema de Transmisión Troncal

# Capítulo 1 – Introducción

## 1.1. Motivaciones

Ante el aumento de la demanda de energía eléctrica proyectado para los próximos veinte años, es que se ha hecho creciente el interés de los privados por invertir en el sector. Desde 1982 el desarrollo del sector de la generación de energía eléctrica está en manos de los inversionistas privados, dejando a los gobiernos una labor fiscalizadora y creadora de políticas públicas que encaucen el desarrollo de esta área. En tanto, por el lado de la transmisión troncal, hasta el año 2005 su planificación se basaba en la interacción entre la oferta y demanda de energía, pero a partir de la Ley corta II esto cambió, dejando en manos de un proceso regulado la planificación de las obras, las que se licitan en caso de ser nuevas.

Por su parte, por distintas razones, se constata un atraso tanto en obras de generación como transmisión. Lo anterior genera incertidumbres en el sector, lo que se traduce en una volatilidad de los precios en el sector.

Como se muestra en el informe definitivo que publicó la Comisión Asesora para el Desarrollo Eléctrico [1] (en adelante CADE), el retraso en un año de todos los proyectos que se encuentran en construcción y los que escoge el modelo utilizado<sup>1</sup>, implican un aumento del 25% del precio monómico en un horizonte de evaluación que va desde el 2012 hasta el 2030. Pese a que el atraso en un año fue un supuesto hecho para evaluar su eventual impacto, éste no se aleja de la realidad debido a que, como la experiencia en Chile lo indica, la mitad de los anuncios que se realizan en los informes de precio de nudo difieren en al menos un semestre de la fecha de entrada en operación real.

La incertidumbre presente en los retrasos de las centrales dificulta la toma de decisiones por parte de la autoridad, la que debe elaborar planes y políticas que vayan en el sentido de prever posibles problemas de suficiencia y competencia en el sector (y solucionarlos si es que ya existen). Estos problemas pueden truncar o hacer más difícil el desarrollo de otros sectores productivos (donde el caso más representativo es la minería), implicando un alza de tarifas o bien

---

<sup>1</sup> El modelo utiliza el concepto de BAU (Business As Usual), que busca representar el desarrollo eléctrico según la política vigente [3].

conducir a racionamientos especialmente cuando se está frente a años con hidrología seca. Sumado a lo anterior, también resulta importante poder establecer un juicio en base a la estadística sobre si se podrá cumplir con desafíos tan grandes como tener un 20% de generación anual en base a energías renovables para el 2020 considerando las tendencias históricas. En resumen, son necesarios análisis de riesgos de los proyectos eléctricos que consideren sus características.

Dada la naturaleza de los proyectos eléctricos y del mercado energético en general, es intuitivo cuestionarse si los retrasos de los proyectos anunciados o la decisión de construcción siguen tendencias o patrones que se abstraen de las condiciones del mercado o de otras variables típicas consideradas en los análisis de riesgos hechos por las entidades financieras. Sin embargo y como una primera aproximación a un problema muy complejo, resulta interesante el hecho de poder extraer conocimiento de la información que se maneja, utilizando la metodología empleada en el proceso de KDD (Knowledge Discovery in Databases)<sup>2</sup>. Esta metodología es usada en muchas áreas del saber, incluso en temas relacionados con los sistemas eléctricos, como por ejemplo la estimación de la demanda eléctrica horaria utilizando redes neuronales artificiales [2]. Hay que tener en cuenta que tanto la implementación de la metodología, como los algoritmos que se utilizan, dependen del problema que se esté enfrentando y del tipo de información que se esté manejando.

Para poder obtener de manera oportuna toda la información requerida para la construcción de los modelos, es de suma importancia que la autoridad disponga de los medios necesarios para un eficaz seguimiento de la evolución del sector eléctrico.

Finalmente, y como un aspecto igual de importante a lo ya mencionado, resulta la visualización de los resultados obtenidos junto con la información que pueden entregar otras representaciones gráficas estadísticas, como lo son los histogramas o diagramas de cajas. Éstos permiten tener una referencia rápida y clara de la información que se encuentra almacenada y de los resultados de aplicar distintos modelos predictivos al conjunto de datos en estudio.

---

<sup>2</sup> Descubrimiento de conocimiento en las bases de datos.

## **1.2. Objetivos**

### **1.2.1. Objetivo general**

El objetivo general de este trabajo es sistematizar, a partir de la información pública disponible, los grados de avance de proyectos de generación y transmisión en un sistema eléctrico, de manera de poder establecer un juicio sobre la materialización de los proyectos eléctricos o si los anuncios sobre las fechas de puesta en servicio presentarán retrasos en función de las características del proyecto en cuestión. La información disponible permite identificar características importantes, que tienen implicancia sobre la fecha esperada de puesta en servicio, la que puede diferir de los anuncios públicos respectivos. Se busca un análisis que se sustente en información estadística de proyectos pasados que permita determinar la probabilidad de que un anuncio se cumpla o presente atrasos y además dar juicios sobre si un proyecto se logra materializar.

Para establecer los juicios se utilizará la metodología de extracción de conocimiento KDD, la que incluye una visualización con herramientas gráficas y su integración al sistema de seguimiento de proyectos.

### **1.2.2. Objetivos específicos**

Entre los objetivos específicos de este trabajo se encuentran:

- Desarrollar un esquema semi-automático de extracción y transformación de la información.
- Mejorar la estructura en la que se almacena la información de modo que se ajuste a los requerimientos de la metodología propuesta.
- Proponer una metodología que permita seleccionar las variables relevantes que posibilitan explicar los atrasos de obras o la materialización de los proyectos, utilizando para ello algoritmos de selección.
- Construir un modelo predictivo a partir de las distintas fuentes de información públicas disponibles, que permita pronosticar el comportamiento de los proyectos que se encuentran en desarrollo y los que están en evaluación ambiental.

- Validar el modelo obtenido calculando su porcentaje de aciertos y utilizarlo para realizar predicciones. Comparar con otros modelos predictivos.
- Desarrollar herramientas de visualización que permitan graficar los resultados de la aplicación del modelo y otras variables estadísticas relacionadas con los atrasos de los proyectos.

### **1.3. Alcances**

El alcance del trabajo incluye la definición de los siguientes aspectos:

- Sistema semiautomático de adquisición y transformación de datos.
- Selección de los atributos relevantes que permitan explicar el fenómeno en estudio.
- Modelo de regresión que permita estimar las fechas de puesta en servicio utilizando las fuentes de información disponibles y definir si una central se construye o no.
- Comparación con otros modelos.
- Validación de la propuesta.
- Visualización de los resultados.
- Propuesta de futuras mejoras.

### **1.4. Estructura del documento**

En el capítulo 2 se realiza una revisión de los sistemas de monitoreo de mercado en el mundo, además de lo que aparece en la literatura especializada. También se describe el estado actual del sistema de seguimientos de proyectos implementado en Chile y se describe el problema el cual motiva al seguimiento de las obras asociadas a instalaciones eléctricas.

En el capítulo 3 se expone la metodología que se utiliza para el análisis de datos, poniendo especial énfasis en las mejoras que se proponen para el sistema de seguimiento de proyectos, de manera que la adquisición de datos pase de un esquema manual a uno semi-automático. Luego de ello, se plantean los modelos matemáticos que se usarán para seleccionar y clasificar los proyectos en función de sus atributos.

En el capítulo 4 se describe como se implementará la metodología para este caso de estudio, además de visualizar los resultados de aplicar la metodología con los datos disponibles públicamente.

En el capítulo 5 se discuten aspectos como la metodología, los modelos utilizados y la información disponible, además de un análisis de los resultados obtenidos. Se hace hincapié en que se pueden mejorar los resultados en caso de disponer de una base de datos más robusta.

Finalmente, en el capítulo 6 se plasman las conclusiones de este trabajo.



# Capítulo 2 –Antecedentes

## 2.1. Mercado eléctrico chileno

Desde 1982 el sistema eléctrico en Chile se rige por el DFL 1/82 [3], con los subsecuentes cambios hechos por el DFL 4/2007 y otras modificaciones dictadas por otras leyes. De acuerdo a esto, la LGSE declara que los sistemas eléctricos interconectados funcionan como un mercado tipo *pool* descentralizado, desintegrado verticalmente, dejando la regulación y la fiscalización en manos de la autoridad. Los tres sectores que componen este mercado son: Generación, Transmisión y Distribución, en donde se reconoce que al ser las dos últimas monopolios naturales se regulan por la autoridad, dejando la primera a la libre competencia.

Es por esta última razón, que el desarrollo y crecimiento del sector de generación eléctrica depende casi exclusivamente de las iniciativas de los privados, donde el estado solo cumple una función reguladora, fiscalizadora y generadora de políticas públicas que encaucen el desarrollo del sector. Hasta el año 2004, la expansión de la transmisión también se encontraba determinada por las iniciativas privadas, pero debido a que se estaban presentando dificultades, la Ley corta I dictaminó que la expansión de los sistemas de transmisión troncal debía ser planificada por la autoridad. De este modo se determinó que las obras nuevas se licitarían su construcción, mientras que las obras correspondientes a ampliaciones debían ser efectuadas por los dueños de las instalaciones.

## 2.2. Demanda futura de energía

Según los resultados del modelo de proyección de la demanda energética global de largo plazo para Chile planteado por PROGEA [4], el SIC y el SING, tendrán el 2030 un requerimiento de potencia máxima de 21.893 [MW] y 3.767 [MW] y un consumo anual de energía de 145.752.157 [GWh] y 23.760.614 [GWh], respectivamente. Estos grandes volúmenes de energía y potencia, se deben fundamentalmente al crecimiento que experimentará la industria chilena, en particular la minería, junto con el crecimiento de la demanda vegetativa (residencial y comercial).

Para poder satisfacer estos futuros requerimientos de energía, es que últimamente ha crecido el interés de los inversionistas, tanto nacionales como extranjeros, por invertir en proyectos energéticos en general, y del sector eléctrico en particular. Lo anterior, sumando a que con los cambios introducidos por la ley de Energías Renovables no Convencionales (ERNC) [5] se requerirá que al año 2024 el 10% de las ventas anuales de energía de cada empresa provengan de fuentes renovables no convencionales, han incentivado a que nuevos actores apuesten a integrarse al mercado eléctrico nacional.

Este creciente interés se ve reflejado claramente en el número de proyectos ingresados de generación al SEIA en los últimos años, que a la fecha de este trabajo sumaban casi 300. El problema que se presenta, es que muchos de estos proyectos no se logran concretar (aproximadamente un 40% del total, según datos obtenidos del SEIA, la CNE y los CDEC).

## **2.3. Institucionalidad**

Dada la cantidad de información que se tiene que manejar (y que va en aumento), es que se hace necesario un monitoreo constante del desarrollo del sector eléctrico, ya que de esto depende también el desarrollo de los otros sectores productivos.

Es por esta razón que a continuación se hará un breve repaso de cuáles son las autoridades responsables encargadas del monitoreo del sector y cuáles son los objetivos y funciones relacionadas con este tema.

### **2.3.1. Ministerio de Energía**

Como lo declara en su portal de internet [6], el objetivo de este ministerio es elaborar y coordinar los planes, políticas y normas para el buen funcionamiento y desarrollo del sector, velar por su cumplimiento y asesorar al Gobierno en todas aquellas materias relacionadas con la energía. Y además de ello, y como se menciona en el mismo sitio web, las funciones relativas al tema del monitoreo del sector son:

1. Estudiar y preparar las proyecciones de la demanda y oferta nacional de energía que deriven de la revisión periódica de los planes y políticas del sector.
2. Elaborar, coordinar, proponer y dictar según corresponda, las normas aplicables al sector energía que sea necesario dictar, para el cumplimiento de los planes y políticas energéticas de carácter general, así como para la eficiencia energética, la seguridad y

adecuado funcionamiento y desarrollo del sistema en su conjunto. Al efecto, podrá requerir la colaboración de las instituciones y organismos que tengan competencia normativa, de fiscalización o ejecución en materias relacionadas con la energía.

#### **2.3.1.1. Comisión Nacional de Energía (CNE)**

Como se declara en su portal de internet [7] el objetivo de la CNE es: “ser el organismo técnico encargado de analizar precios, tarifas y normas técnicas a las que deben ceñirse las empresas de producción, generación, transporte y distribución de energía, con el objeto de disponer de un servicio suficiente, seguro y de calidad, compatible con la operación más económica”. En particular, una de las funciones específicas que atañen a la temática que se desarrolla en este trabajo es “Monitorear y proyectar el funcionamiento actual y esperado del sector energético, y proponer al Ministerio de Energía las normas legales y reglamentarias que se requieran, en las materias de su competencia”.

Esta última función es la que se realiza semestralmente, en la elaboración del informe de precio de nudo de corto plazo que se publica los meses de abril y octubre de cada año y en los cuales se elabora un plan de obras con los proyectos de generación y transmisión recomendados por la misma comisión y con el estado de las obras que se encuentran en construcción.

#### **2.3.1.2. Centros de Despacho Económicos de Carga (CDEC)**

Los CDEC, son los organismos responsables de coordinar la operación de las instalaciones eléctricas de los concesionarios que operen interconectados entre sí [3]. En Chile existen dos CDEC: el CDEC-SIC y el CDEC-SING. El primero coordina las instalaciones que van desde Tal-tal hasta Chiloé, y el segundo las instalaciones presentes en el Norte Grande (XV, I y II región), que principalmente abastecen el sector minero e industrial. Existen otros dos sistemas medianos (Aysén y Magallanes), pero éstos son operados por los dueños de las instalaciones.

Particularmente y dentro de lo que respecta a esta memoria, una de las funciones de dichas instituciones declaradas en sus portales web [8] es: “Planificar la operación de corto plazo del sistema eléctrico, considerando su situación actual y la esperada para el mediano y largo plazo”. Por este motivo es que en el CDEC-SIC realiza un seguimiento de las obras que están prontas a entrar y las considera en la programación de largo plazo (PLP), que es la que define los valores

del agua en las centrales que operan con embalses y las que están en serie hidráulica. Este seguimiento se ve reflejado en los datos de entrada al PLP.

### **2.3.2. Ministerio del Medio Ambiente**

Después de que se promulgó la Ley general del medio ambiente en 1994, el organismo a cargo de hacer cumplir la legislación fue la Comisión Nacional del Medio Ambiente (en adelante CONAMA), la que tenía sus sedes regionales en las Comisiones Regionales del Medio Ambiente (en adelante COREMA). Por mucho tiempo estas instituciones eran las que en la práctica hacían las políticas públicas en materias ambientales, hasta que en el 2010 se legisló una nueva institucionalidad del medio ambiente, en la que el recién creado Ministerio del Medio Ambiente toma especial relevancia en la elaboración y propuesta de políticas públicas. En su página web [9] se declara que el ministerio “Es el órgano del Estado encargado de colaborar con el presidente de la República en el diseño y aplicación de políticas, planes y programas en materia ambiental, así como en la protección y conservación de la diversidad biológica y de los recursos naturales renovables e hídricos, promoviendo el desarrollo sustentable, la integridad de la política ambiental y su regulación normativa.” Además en esta nueva normativa se crean el Servicio de Evaluación Ambiental y la Superintendencia del Medio Ambiente.

#### **2.3.2.1. Servicio de Evaluación Ambiental (SEA)**

Como lo declaran en su web [10] es un organismo público descentralizado. Su función central es tecnificar y administrar el instrumento de gestión ambiental denominado “Sistema de Evaluación de Impacto Ambiental” (SEIA), cuya gestión se basa en la evaluación ambiental de proyectos ajustada a lo establecido en la norma vigente, fomentando y facilitando la participación ciudadana en la evaluación de los proyectos. Los proyectos se aprueban cuando la RCA es favorable.

#### **2.3.2.2. Superintendencia del Medio Ambiente**

En el marco de la nueva institucionalidad ambiental, este organismo de servicio público descentralizado y sometido al ministerio del medio ambiente es al que le corresponde -de forma exclusiva-, ejecutar, organizar y coordinar el seguimiento y fiscalización de las RCA, de las medidas de los Planes de Prevención y/o de Descontaminación Ambiental, del contenido de las

Normas de Calidad Ambiental y Normas de Emisión, y de los Planes de Manejo, cuando corresponda, y de todos aquellos otros instrumentos de carácter ambiental que establezca la ley [11].

## **2.4. Estado del arte de los sistemas de monitoreo**

### **2.4.1. Literatura científica**

De una revisión de la literatura internacional disponible en el sitio de la IEEE<sup>3</sup> sobre los sistemas de monitoreo de mercados energéticos, se puede observar que la mayoría apunta a la supervisión de los parámetros que tienen un carácter instantáneo (costos marginales, precio de los combustibles, congestiones, volúmenes de los embalses, etc.), junto con el cálculo de indicadores determinísticos que utilizan información histórica o del mercado spot (CPMI, CPD, HHI, market share, etc) [12]. Un poco más cerca del problema del seguimiento de la oferta es lo que ilustra el indicador de margen de reserva, aunque éste se enfoca en lo que ocurre en tiempo real sin hacerse cargo del futuro.

Sin embargo, Gao [13] se hace cargo del problema de largo plazo de la oferta de energía mediante el planteamiento de un escenario base junto con un escenario pesimista y un escenario optimista de crecimiento de la capacidad instalada. Pero la definición de escenarios futuros resulta confusa, sin dejar claro los supuestos utilizados en la construcción de éstos.

### **2.4.2. Índice de credibilidad de las inversiones**

En [14], se define este índice que permite informar acerca de la credibilidad de las compañías generadoras en relación a sus anuncios sobre los proyectos de instalación de centrales generadoras. Se compara la consistencia entre los anuncios realizados por una empresa generadora, respecto de la construcción de nuevas centrales, y lo efectivamente realizado. A medida que se va retrasando la construcción y entrada en servicio de una nueva central respecto a su fecha de anuncio, el índice comienza a reducir su valor, pudiendo variar entre 0% y 100%. Para su cálculo se requiere conocer la potencia instalada que se encuentra en operación y la capacidad anunciada de cada empresa generadora (EGen) para el horizonte de tiempo bajo análisis. El índice se define de la siguiente manera:

---

<sup>3</sup> IEEE: Institute of Electrical and Electronic Engineers

$$ICI_i = \frac{\sum_{t=ID}^{FD} CI_{i,t}}{FD - ID + 1} \cdot 100\%$$

Dónde,

$$CI_{i,t} = \begin{cases} \frac{Ejecutadas_t - Ejecutadas_1}{Anunciadas_t - Ejecutadas_1} & \text{si } Anunciadas_t \neq Ejecutadas_1 \\ 1 & \text{si } Anunciadas_t = Ejecutadas_1 \end{cases}$$

ID Fecha inicial escogida para el cálculo del ICI

FD Fecha final escogida para el cálculo del ICI

*Ejecutadas<sub>t</sub>* Suma de potencia de las centrales en estado ejecutada, pertenecientes a la empresa seleccionada, en el período t.

*Anunciadas<sub>t</sub>* Suma de potencia de las centrales en estado anunciada, pertenecientes a la empresa seleccionada, en el período t.

El índice de credibilidad se calcula en cada instante de tiempo como la razón de potencia de las centrales ejecutadas y anunciadas en el instante t, sin importar lo ocurrido antes de la fecha de inicio (por esta razón es que se restan los valores de potencia de las centrales antes del período de inicio). Este dato se integra de modo que su valor represente todo el intervalo de tiempo.

Su valor dependerá de la fecha de inicio seleccionada por el usuario para el cálculo del índice. Es importante que para que los valores sean comparables, el IC calculado sobre distintas empresas debe tener el mismo período de inicio.

### 2.4.3. Casos de otros países

#### 2.4.3.1. Perú

El caso de Perú es muy similar al chileno, debido a que su legislatura también establece la presencia de una autoridad medioambiental ante la cual se tienen que presentar los proyectos, para que ésta los evalúe. Sin embargo, a diferencia del caso chileno, existe un registro de las

empresas consultoras que pueden presentar estudios de impacto ambiental<sup>4</sup> y además se les exige a las empresas llevar un monitoreo constante de las obras de construcción, junto con las medidas de mitigación para compensar los efectos negativos del proyecto.<sup>5</sup> Más aún, el incumplimiento de la normativa se traduce en multas acumulativas. Lo mismo ocurre para las concesiones de electricidad y construcción, en donde debe existir un informe de una clasificadora de riesgos clasificada y además el Ministerio recibe un 1% de la inversión con el fin de que se cumplan los plazos previstos.<sup>6</sup>

## 2.5. Sistema de monitoreo de los mercados energéticos en Chile

Basado en el prototipo de sistema de monitoreo planteado en los trabajos de Reid [14], Escobar [15] y Mancilla [16], se instaló en los servidores del Ministerio de energía un sistema de monitoreo de mercado energético, el cual consiste en una plataforma web que contiene múltiples secciones, cada una con un fin bien específico. Las secciones de este sistema son: Visualizador, indicadores, traspaso de archivos y seguimiento de proyectos.

El visualizador es una herramienta basada en el software DeepEdit, que permite representar la información que aparece en las programaciones de corto plazo de la operación de los dos sistemas interconectados más grandes de Chile. Además posee una extensión para ver los archivos de salida de las simulaciones que realiza el software OSE 2000<sup>7</sup>.

La sección de indicadores calcula los indicadores planteados en la base teórica del proyecto [17], los que reciben la información necesaria desde los archivos descargados automáticamente, desde el sistema de traspaso de archivos y de la base de datos del sistema de seguimiento de proyectos.

El sistema de seguimiento de proyectos consiste en un conjunto de expedientes web llamados *fichas de proyectos*, en donde se consideran tanto los proyectos de generación (diferenciados por tipos de tecnologías), transmisión troncal (líneas y subestaciones) e instalaciones asociadas al sector de hidrocarburos (gasoductos, oleoductos, poliductos y centros

---

<sup>4</sup> Resolución Ministerial N° 580-98-EM/VMM .- Dictan Normas Referidas al Registro de Entidades Autorizadas a realizar Estudios de Impacto Ambiental en el Sector Energía y Minas.

<sup>5</sup> Decreto Supremo N° 029-94-EM.-aprueban el Reglamento de Protección Ambiental en las Actividades Eléctricas.

<sup>6</sup> Artículos 3º, 25º y 38º de la Ley de Concesiones Eléctricas, Decreto Ley N° 25844, artículos 37º y 66º de su Reglamento y Artículo 8º de la Ley 16053

<sup>7</sup> Software que utiliza de manera estándar la autoridad para realizar las planificaciones de largo plazo.

de almacenamiento). Sin embargo, el llenado de la información de esta ficha es totalmente manual, lo que conlleva a que existan dificultades para actualización permanente de la información. Este último punto es de suma importancia, ya que según la información que se obtiene directamente desde la página del sistema de evaluación de impacto ambiental, cada año aumentan los proyectos que se ingresan al sistema, siendo de suma importancia una herramienta que permita dar un juicio sobre la aprobación de un proyecto, tiempo de tramitación, construcción de un proyecto y tiempo que toma un proyecto desde su tramitación ambiental y operación en el respectivo sistema interconectado, dadas las características que este proyecto posee.

## **2.6. Etapas de un proyecto eléctrico**

A continuación se describen las principales etapas de un proyecto, haciendo especial énfasis en lo correspondiente a los proyectos eléctricos.

### **2.6.1. Pre-factibilidad**

En esta etapa se estudia la factibilidad técnica y económica de los proyectos de manera de descartar alternativas y estudiar la rentabilidad de las distintas opciones de inversión de capitales. En esta etapa los proyectos se ingresan al SEIA como Estudio de Impacto Ambiental (en adelante EIA) o Declaración de Impacto Ambiental (DIA), dependiendo de sus características y su eventual impacto en el entorno. Además, en esta etapa se tramitan la mayor parte de los permisos sectoriales y se negocian las servidumbres.

### **2.6.2. Factibilidad**

Una vez que se obtienen todos los permisos sectoriales respectivos y la RCA favorable, se revisan los cambios que fueron solicitados para cumplir con la normativa y se re-evalúa el proyecto para verificar si sigue siendo viable económica y técnicamente.

Además, en esta etapa los impulsores del proyecto comienzan a conseguirse los contratos de suministro para así disminuir el riesgo de la inversión y poder obtener financiamiento a través de inversionistas o mediante la solicitud de un crédito en alguna entidad financiera. Generalmente la tasa de retorno que exigen los bancos es menor a las tasa de retorno que exige un inversionista, por lo que generalmente la inversión resulta ser mixta.



Debido al riesgo que significa no poseer contratos de suministro de largo plazo, es que una vez que un proyecto ha obtenido todos sus permisos no siempre logra concretarse.

### 2.6.3. Inversión y construcción

Una vez aprobados todos los permisos ambientales y sectoriales, adquiridos los contratos de suministro y conseguido el financiamiento, se procede a la construcción de la central. Usualmente esta etapa tiene retrasos que son típicos en este tipo de obras, pero que sin embargo, se deben considerar al momento de enviar una fecha de la puesta en operación de la central a la CNE.

Los atrasos en esta fase corresponden a imprevistos (accidentes o cosas no contempladas en el diseño que fueron detectadas en el camino) o bien a causas de fuerza mayor, como lo son los terremotos o desastres naturales. Aunque lo más probable es que el proyecto se logre concluir, se ve afectado el tiempo en el que se podrá disponer la central, lo cual influye en el óptimo funcionamiento del mercado.

### 2.6.4. Operación

En esta etapa el proyecto ya se encuentra operativo y en condiciones de abastecer la demanda del sistema donde se encuentra conectado.

Finalmente las etapas de un proyecto se resumen en la Figura 2.1.



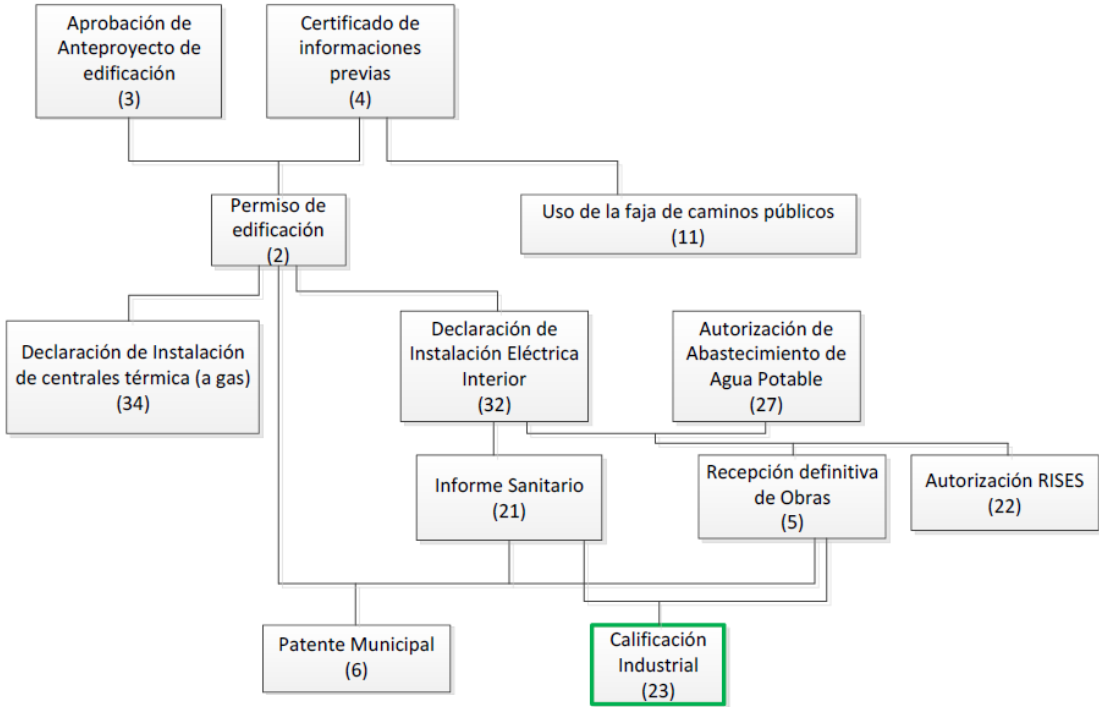
Figura 2.1 Etapas de un proyecto y lugar donde se ubica el ingreso al SEIA

## 2.7. Permisos aplicables a los proyectos

Además de la tramitación en el SEIA, los proyectos de generación y transmisión de energía requieren de múltiples permisos sectoriales que se tramitan en diferentes instituciones del Estado. Según la Consultora Estudios de Medio Ambiente y Gestión S.A. [18], la cantidad de permisos totales asciende a 56, de los cuales 39 se tramitan fuera del SEIA. Algunos permisos sectoriales se tienen que tramitar antes del ingreso al SEIA, debido a que son requisitos de algún permiso

ambiental, mientras que otros se tienen que tramitar después, puesto que tienen como requisito la aprobación de algún permiso ambiental (que viene a la par con la calificación favorable de la RCA). El resto se puede tramitar de manera paralela. Cabe hacer notar que los permisos dependen del tipo de proyecto, por lo que la cantidad de permisos que requiere un proyecto es siempre menor a 56.

En la Figura 2.2 se aprecian los permisos sectoriales que dependen de la aprobación de otros permisos para poder tramitarse. Estos diagramas se obtienen de analizar la información sistematizada en la base de datos que venían como anexo al informe [18].



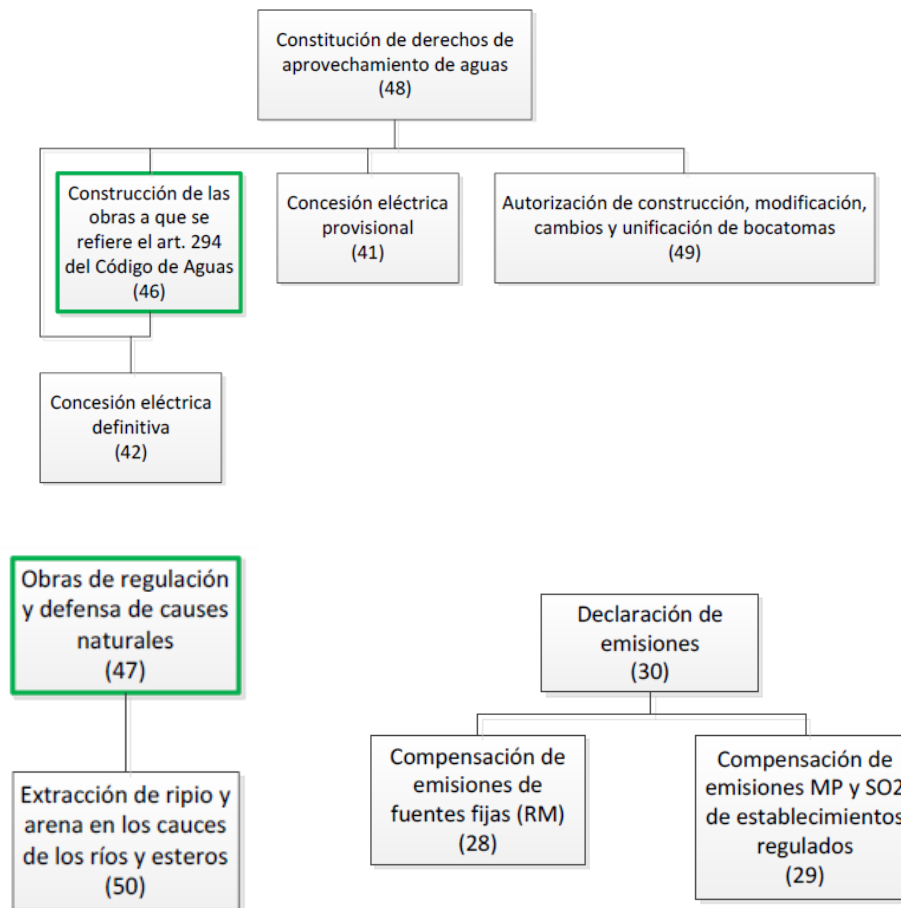


Figura 2.2 Permisos sectoriales que tienen como pre-requisitos otros permisos. Elaboración propia a partir de la información obtenida de [18].

Por simple lógica, la demora en alguno de estos permisos implica la demora del tiempo total del proyecto. Generalmente, de los 56 permisos requeridos para los distintos tipos de proyectos, son unos pocos los que presentan dificultades en su tramitación que se traducen en largos tiempos de espera y en el riesgo que ello conlleva en la decisión de los inversionistas en invertir en este negocio.

### 2.7.1. Permisos que presentan dificultades e instituciones responsables

Pese a la gran cantidad de permisos que se deben solicitar para llevar a cabo un proyecto de generación o transmisión, sólo algunos presentan problemas en el tiempo que toma este trámite.

Las trabas dependen fuertemente del tipo de tecnología utilizada para la construcción de la central, además de la ubicación en la cual se emplaza el proyecto.

A continuación se presenta un detalle de cuáles son las principales trabas identificadas en el estudio [18], agrupándolas por institución a cargo de gestionar estos permisos.

#### **2.7.1.1. Dirección General de Aguas (DGA)**

Este es el organismo que administra las aguas territoriales en Chile. Las trabas asociadas a este organismo están relacionadas a la determinación de los caudales ecológicos y a la constitución de derechos de agua, especialmente lo no-consuntivos, que son los que se ocupan para la construcción de centrales hidroeléctricas de cualquier tipo.

En el caso de la determinación de los caudales ecológicos, el problema se basa en la incompatibilidad que en ocasiones existe con respecto a los requerimientos de caudal que necesitan los proyectos y al caudal ecológico determinado por el SEA.

El caso más crítico se encuentra en la constitución de derechos de aprovechamientos de agua, ya que sin él no existe proyecto y no se puede hacer absolutamente nada. Esto afecta directamente a los proyectos hidroeléctricos, pero también puede afectar a otro tipo de proyectos que necesiten de agua para su funcionamiento, como lo pueden llegar a ser las centrales geotérmicas. El problema con estos derechos es el largo tiempo de tramitación, el que se explica principalmente por la falta de personal técnico existente en la institución. Además de lo anterior, y dada la continuidad hidrológica de las fuentes superficiales, se necesitan tener resueltas las solicitudes anteriores en el mismo sector.

Por todo lo antes mencionado, se hace difícil a una entidad que no cuenta con los recursos necesarios poder coordinar los derechos que están en tramitación, los nuevos derechos y las solicitudes de cambios de puntos, para así poder seguir con la continuidad hidrológica de una determinada cuenca.

Otro permiso asociado a las aguas superficiales, pero que está a cargo de la Dirección de Obras Hidráulicas (DOH), es el de obras hidráulicas mayores, que se tramita en el SEIA, pero que se consigue ulteriormente a la obtención de la RCA y que, además de necesitar del derecho de aprovechamiento de aguas, es prerrequisito de las concesiones eléctricas definitivas.

### **2.7.1.2. Dirección General del Territorio Marítimo (DIRECTEMAR)**

Este es el organismo encargado de entregar las concesiones marítimas necesarias para utilizar las aguas de mar. Si sumamos el limitado personal con el que cuenta esta institución con el hecho de ser éste un permiso de carácter estratégico y que responde a un tema ligado con el ordenamiento y la planificación territorial, tenemos como resultado que la tramitación de este permiso toma en promedio más de 650 días en ser otorgado.

Esta concesión es clave si se quiere construir una central a carbón, debido a que, como éste es importado, se requiere que la central esté cerca de un puerto, los que son escasos en este país (además de que no existen muchos lugares aptos en la costa chilena para la instalación de puertos). De este modo se utiliza el agua proveniente del mar para la refrigeración de las calderas, por la proximidad que tienen las centrales a las costas.

### **2.7.1.3. Ministerio de Bienes Raíces**

Este es el ministerio responsable de administrar todo tipo de territorios fiscales. El principal problema que tiene este organismo son los altos tiempos que demora en vender o arrendar territorios necesarios para la construcción de centrales de todo tipo. Particularmente, los proyectos más afectados por la burocracia son las que requieren más cantidad de hectáreas, como por ejemplo: los parques eólicos, las centrales fotovoltaicas y las líneas de transmisión.

### **2.7.1.4. Ministerio de Energía**

Además de ser el organismo encargado del sector energético, es el responsable directo de otorgar las licitaciones de concesiones de exploración y explotación geotérmica. Estas licitaciones, junto con el arriendo y compra de bienes fiscales, pueden tomar tiempos de tramitaciones que superan fácilmente los 800 días. Este largo de tiempo se debe a que se requieren muchas consultas a otros organismos del Estado, además de tener compatibilidad con la planificación territorial.

Así, sumado al gran riesgo que conlleva la exploración geotérmica se suma el riesgo de ganarse la licitación, lo que aumenta la incertidumbre para este tipo de iniciativas que son altamente rentables desde el punto de vista económico (si se deja afuera el riesgo asociado a la exploración).

### **2.7.1.5. Superintendencia de Electricidad y Combustibles (SEC)**

Este es el organismo encargado de tramitar las concesiones eléctricas provisionales, que son las necesarias para que las líneas de transmisión, subestaciones y centrales hidroeléctricas puedan realizar sus estudios previos. Luego de la concesión provisoria, para lograr establecer las servidumbres, se hace necesario que el Ministerio de energía (en representación del Presidente de la República) entregue las concesiones definitivas.

### **2.7.1.6. Otras trabas**

Además de los permisos antes mencionados, existen otras dos trabas que son de carácter territorial y que tienen que ver con la nueva Ley de bosque nativo, las normativas de calidad del aire y el cumplimiento del convenio 169, firmado por el estado chileno y que indica que se deben hacer consultas a la población indígena sobre los nuevos proyectos que se emplacen en territorios que sean de su propiedad o patrimonio histórico.

El cumplimiento de la Ley de bosques nativos traba los desarrollos de líneas de transmisión, además de otros proyectos que se proyectan en el sur del país, debido a que uno de los requisitos es tener una concesión para establecer servidumbres, que como ya se revisó, constituyen uno de los mayores problemas para los proyectos.

### **2.7.1.7. Judicialización**

Finalmente, un problema que ha empezado a surgir este último tiempo, tienen relación con la judicialización de los proyectos. Esto se materializa en recursos tanto administrativos como judiciales que hacen tanto personas naturales como instituciones con personalidad jurídica en contra de una determinada iniciativa. En relación a los recursos administrativos, la mayoría tiene relación con la validez de la RCA por diferentes motivos, mientras que en el caso de la judicialización el recurso más utilizado es el Recurso de protección.

Como se concluye en un estudio elaborado por la Facultad de Derecho de la Universidad de Chile [19], uno de las principales motivaciones de este fenómeno se relaciona con que un conjunto de la sociedad siente que no es escuchada y que los mecanismos de participación ciudadana no son suficientes y tampoco influyen en la forma en que se llevarán adelante los proyectos.

## **2.8. Fuentes de información**

Para la extracción de la información necesaria para el desarrollo de este trabajo, se recurrió a las distintas fuentes que se describirán a continuación.

### **2.8.1. Prensa especializada**

En ocasiones, las empresas anuncian posibles futuras inversiones en el rubro de las instalaciones energéticas en la prensa especializada, como lo es, por ejemplo, el portal de la Revista Electricidad [20]. En estos medios también se informa sobre el avance de las tramitaciones, los recursos administrativos y judiciales interpuestos, el avance de las construcciones, entradas en operación, nuevas estrategias comerciales, etc.

Si bien, estos medios pueden entregarnos información interesante, ésta es poco detallada y en ocasiones muy superficial.

### **2.8.2. Sistema de Evaluación de Impacto Ambiental**

Desde que en 1994 se promulgó la Ley General del Medio Ambiente, los proyectos de generación eléctrica que poseen más de 3 MW de capacidad total, tienen la obligación de tramitarse ambientalmente en el Sistema de Evaluación Ambiental, también conocido como SEA. Dependiendo del impacto de los proyectos, estos se ingresan al sistema como Evaluación de impacto ambiental (EIA) o como Declaración de impacto ambiental (DIA). Al ingresarse al sistema, se genera un expediente en que se almacena la información relevante y los hitos importantes en la tramitación, como lo es por ejemplo la resolución de calificación ambiental, mediante la cual un proyecto es aprobado.

Es por esta razón, que generalmente la primera fuente de información sobre los proyectos que están en las carteras de los inversionistas, se halla en esta plataforma de información. Si bien, parte de la información que aparece cuando se ingresa un proyecto al sistema cambia en el tiempo, a partir de este punto se puede empezar a construir una ficha de seguimiento de proyecto con las características relevantes que permitan ir dando cuenta del estado de avance de un proyecto de inversión.

En este sistema también se realiza un seguimiento de los recursos administrativos que son interpuestos por personas o agentes, con el fin de denunciar malas prácticas administrativas en la

tramitación del proyecto. Aquí aparece tanto la interposición del Recurso como de las resoluciones por parte de los agentes responsables.

Finalmente, y luego que la RCA ha resultado favorable para el proyecto, se realiza un seguimiento de las obras. Este seguimiento en ocasiones no existe o no se realiza periódicamente.

### **2.8.3. Informe de Precio de nudo de corto plazo**

Para el cálculo semestral del precio de nudo de cada uno de los sistemas interconectados, la legislación [3] entrega a la CNE la labor de elaborar un plan de obras óptimo de expansión de cada sistema, para un horizonte de evaluación de 10 años, en donde debe considerar un set de alternativas tecnológicas técnica y económicamente factibles que cubran distintas fuentes de energía.

Para determinar el plan óptimo se minimiza una función objetivo mediante un proceso iterativo, en donde va comparando distintas opciones de desarrollo. La función objetivo es:

$$\text{Min } \left\{ \sum \text{Inv} + \text{CO\&M} + \text{Cvar} - \text{Resid} \right\}$$

s.a. Restricciones de demanda, limitaciones del sistema de transmisión, restricciones de riesgo, potencias máximas de centrales generadoras, variabilidad hidrológica, etc.

Donde:

*Inv*: Valor actualizado de todas las inversiones futuras a optimizar.

*CO&M*: Valor actualizado de todos los costos de operación y mantenimiento de las nuevas instalaciones, que en caso de las centrales a gas natural incluye el costo fijo anual de transporte de gas. Los valores de CO&M anual previos a su actualización se consideran al final de cada año.

*Cvar*: Costo de operación y falla futuro actualizado del sistema.

*Resid*: Valor actualizado del monto residual de todas las inversiones futuras

Con el fin de que el plan se ajuste lo más posible a la realidad, se consideran como puntos de partida las obras que se encuentran en construcción en donde la fecha de entrada en operación es la que informan los dueños de éstas. Sin embargo, y después de una revisión de todos los informes publicados desde 1990 a la fecha, se puede apreciar que algunos proyectos específicos se consideran dentro de las obras recomendadas por la CNE. Este juicio es hecho luego de cruzar



los planes de obras con otras fuentes de información (por ejemplo el SEIA) y con informes posteriores, teniendo especial cuidado de no confundir los proyectos con módulos genéricos utilizados por la CNE no atribuibles a ninguna empresa en particular.

El gran problema de estos planes de obras, es que las fechas de entrada en operación de las centrales que proponen distan mucho de la fecha real de entrada en operación de los proyectos. Prueba de ello es un resumen gráfico de las postergaciones que se han ido realizando en los distintos informes de precio de nudo de algunas centrales que tienen características en común y que se grafican en la Figura 2.3, Figura 2.4, Figura 2.5, Figura 2.6, Figura 2.7, Figura 2.8 y Figura 2.9.

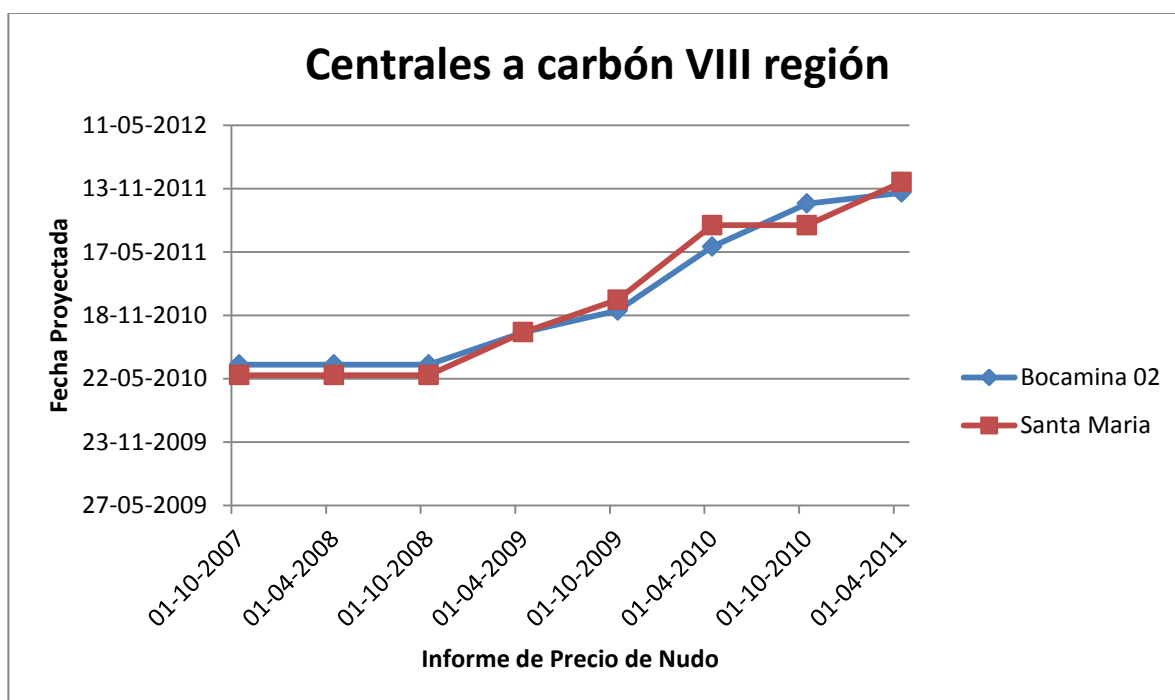


Figura 2.3 Atrasos que han presentados los anuncios de las centrales a carbón que se están construyendo en la región del Bío Bío

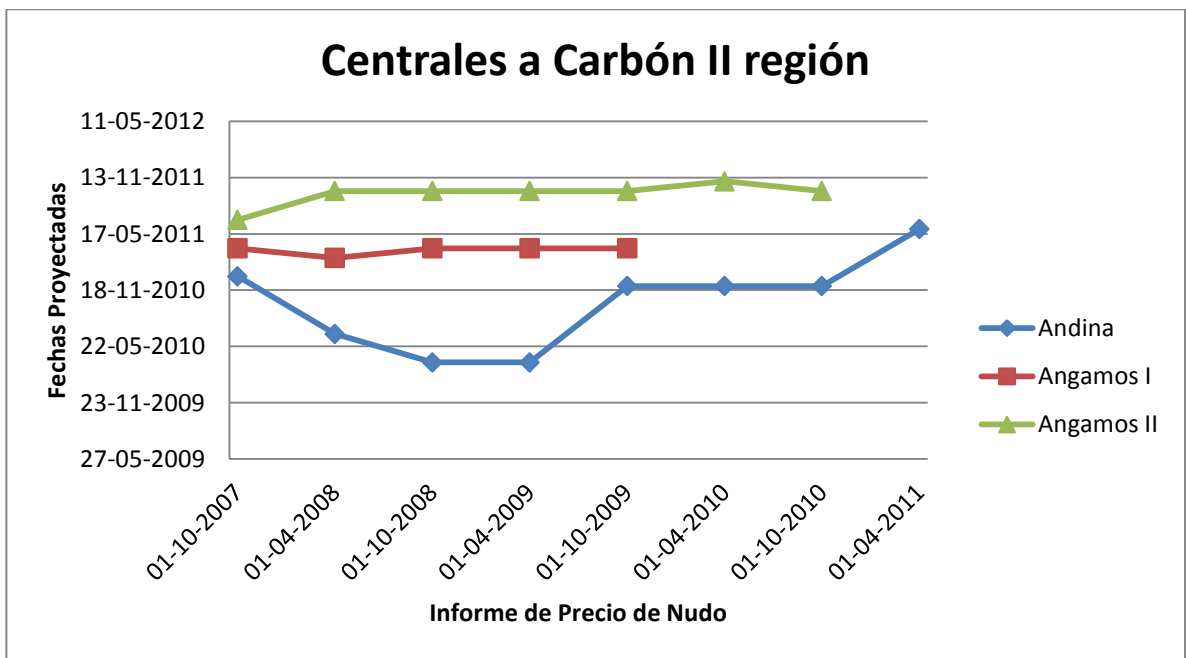


Figura 2.4 Atraso en los anuncios que han presentado las últimas centrales a carbón construidas en el SING

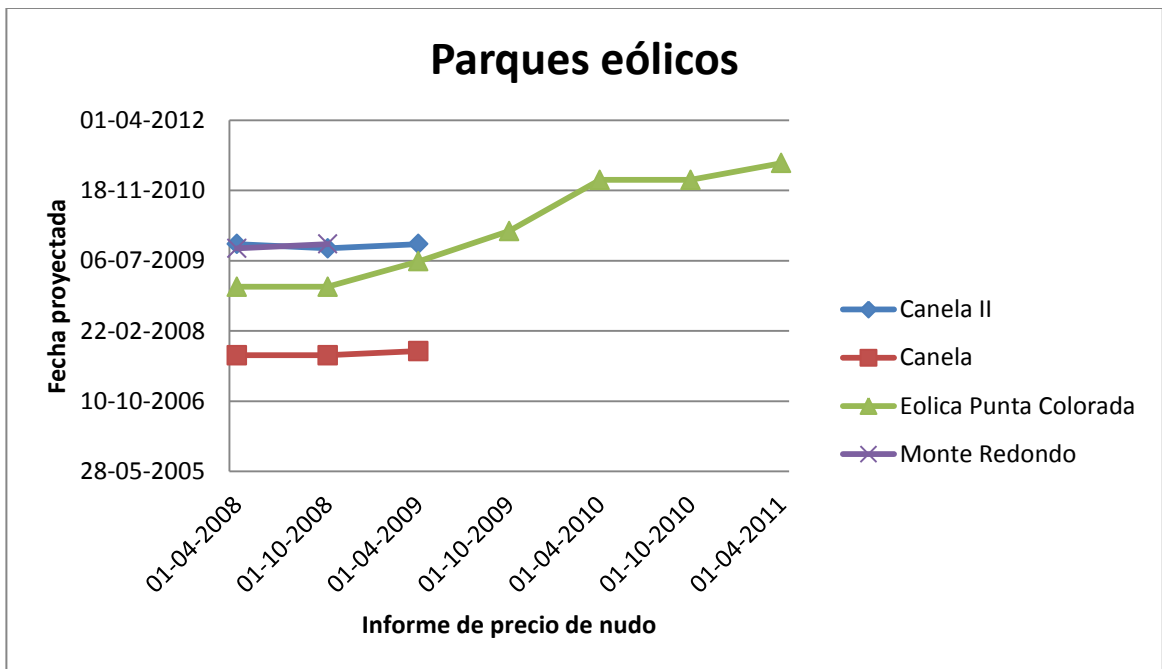


Figura 2.5 Atraso en los anuncios que han presentado los parques eólicos construidos en el SIC

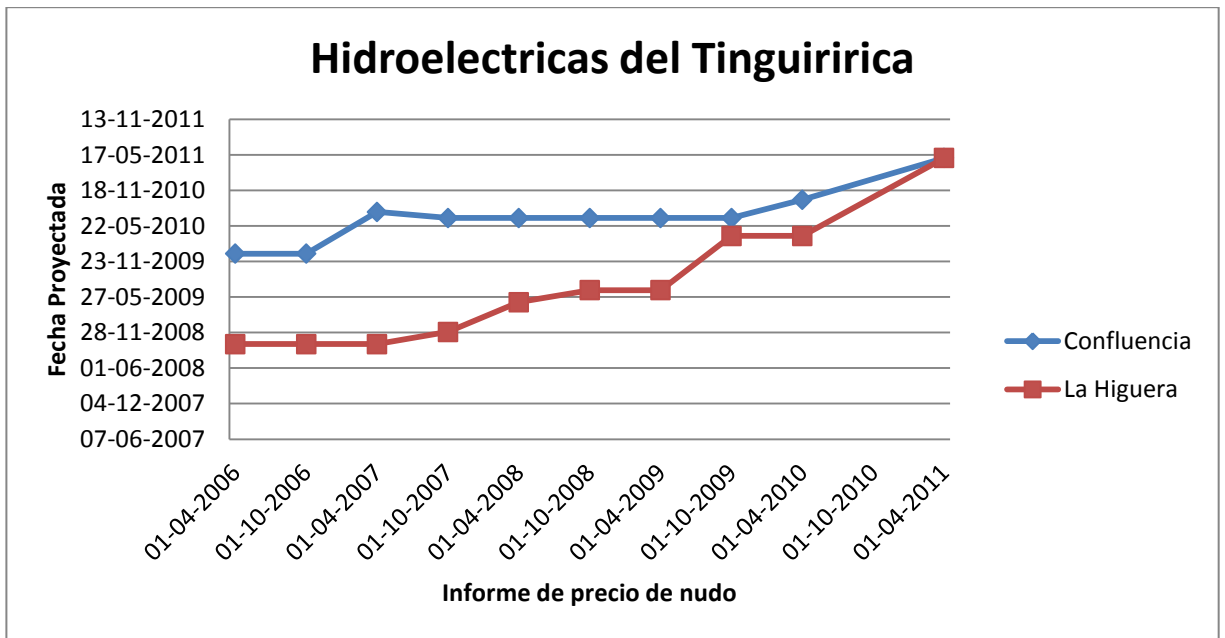


Figura 2.6 Atraso en los anuncios que han presentado dos grandes proyectos de Tinguiririca S.A.

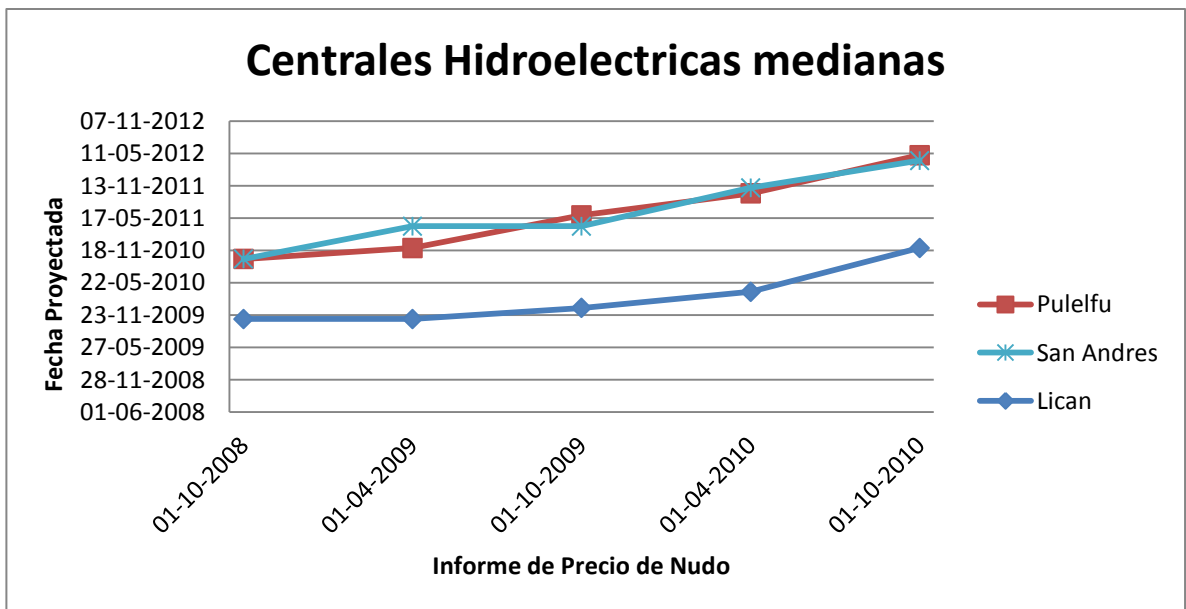


Figura 2.7 Atraso en los anuncios que han presentado centrales hidroeléctricas de menos de 100 MW en el SIC

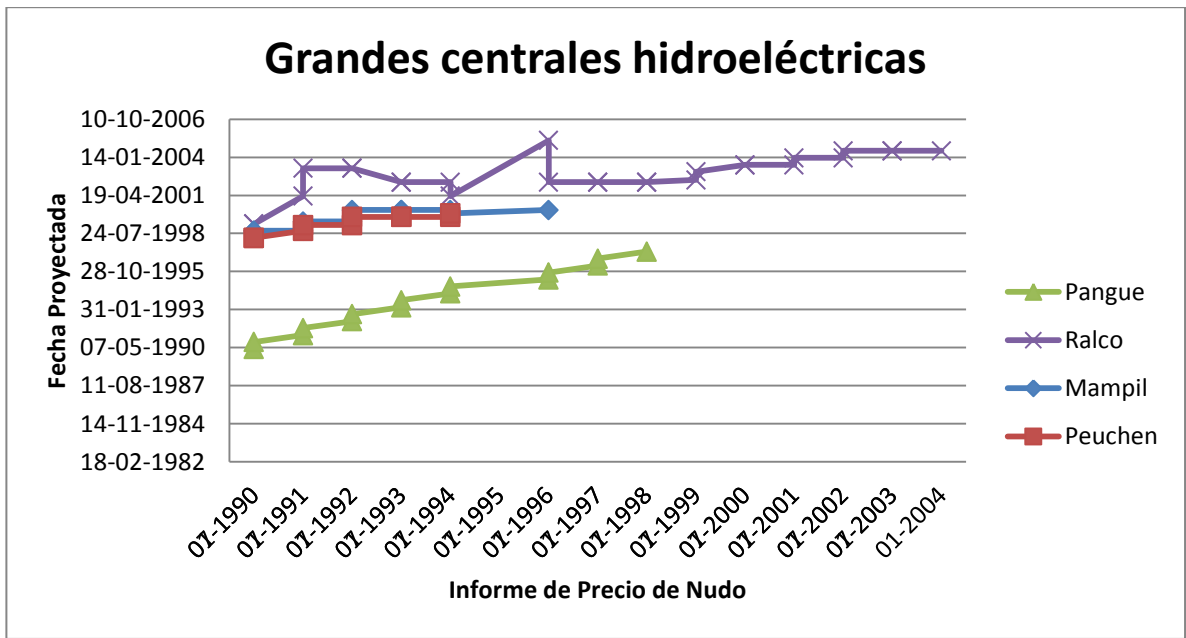


Figura 2.8 Atraso que han presentado grandes proyectos hidroeléctricos de grandes empresas en el SIC

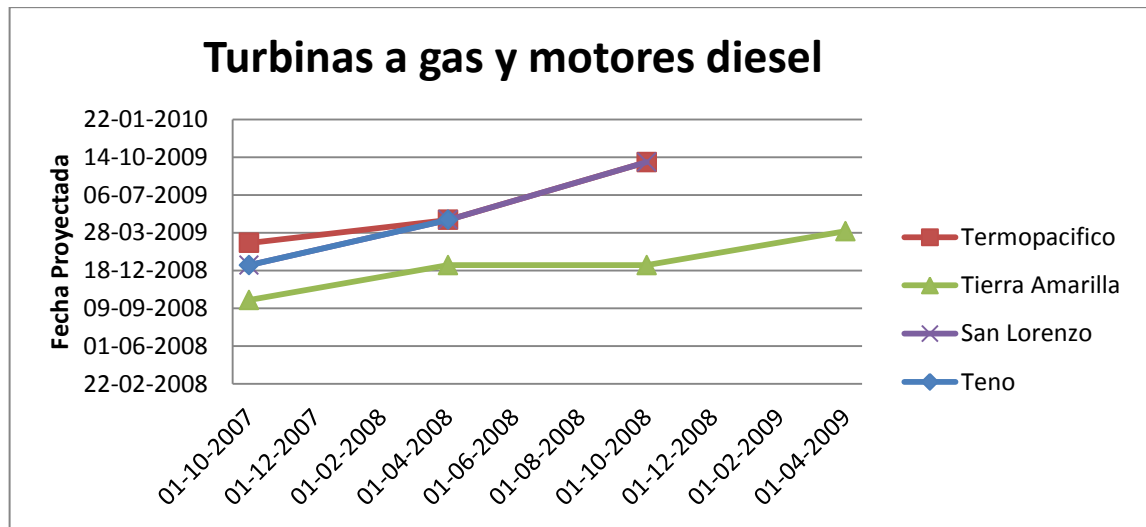


Figura 2.9 Atrasos que han presentado turbinas de respaldo para el SIC

Nótese que dadas ciertas características en común, los atrasos en los anuncios parecieran seguir tendencias, que justamente son los que se pretenden identificar y caracterizar.

Finalmente en [21] Cerda. P. realiza un análisis histórico del plan de obras real y del plan de obras de la CNE, junto con la comparación entre ambos separando los proyectos por tipos de

tecnología. Además se hace una revisión de la historia del SIC desde 1990, en donde se relata el efecto de las nuevas leyes y las condiciones de mercado tanto internas como externas presentes a la hora de evaluar el futuro de la matriz. En este trabajo se concluye que las condiciones del mercado como la legislación producen efectos en el desarrollo de la matriz energética del país.

#### **2.8.4. Centros Económicos de Despacho de Carga (CDEC)**

En los sitios web de los CDEC se encuentran los inventarios detallados con toda la infraestructura eléctrica presente en los sistemas interconectados que por Ley están sujetos a coordinación (CDEC-SIC y CDEC-SING). En estas planillas se puede encontrar información sobre la capacidad instalada, el tipo de tecnologías y los combustibles que utilizan las distintas centrales para la generación de electricidad, así como mapas geográficos de donde se puede sacar una idea de la ubicación geográfica de las centrales. Junto con todo lo anterior, se encuentra la variable más importante para este trabajo, que es la fecha en entrada en operación de las centrales que se encuentran conectadas a cada uno de los sistemas. Esta información, o aparece explícita o bien, se puede desprender al revisar las planillas con la generación real mensual.

### **2.9. Implicancias de los atrasos en el sistema**

Existen instancias de corto y mediano plazo en las cuales los atrasos de los proyectos afectan la operación del sector, transformando el problema de los retrasos en un problema de largo plazo cuando estos se repite constantemente en el tiempo, agregando bastante aleatoriedad en lo que vaya a ocurrir en el futuro.

A continuación se explican algunas de las implicancias que tienen los retrasos en la operación, que son particularmente notorias en el caso del SIC.

#### **2.9.1. Programación de Largo Plazo (PLP)**

El CDEC-SIC realiza semanalmente una programación de corto plazo en donde se determina el punto de operación horario de cada central para la semana. Es decir, determina la potencia tanto activa como reactiva y el perfil de tensión de las distintas unidades que conforman el sistema. Para ello, cuenta con un programa llamado PCP, al cual se le introducen datos como las potencias mínimas y máximas, los afluentes de los ríos y el volumen de los embalses, costos variables, mantenimientos, valor del agua, entre otros.

Sin embargo, para determinar el valor del agua embalsada es necesario realizar una simulación de largo plazo, en la que se considera un periodo de aproximadamente 21 meses, en donde se toman en cuenta las centrales que supuestamente deben estar operativas en esta ventana de tiempo. Si una obra se retrasa, el valor actual del agua se ve afectado, ya que considera que un nuevo bloque de energía entrará a competir, lo que implica que los embalses pueden ocupar más agua en el presente. Si esto se repite en el tiempo, además de transformar un problema de corto plazo en uno de largo plazo, se podría llegar a una situación en la cual los embalses no cuentan con los volúmenes suficientes y se corre el riesgo de racionamiento, en caso de presentarse una hidrología seca.

### **2.9.2. Estudio de Transmisión Troncal (ETT)**

En este estudio la CNE realiza una expansión del Sistema de Transmisión Troncal (en adelante STT), para lo cual utiliza como insumo 3 escenarios de expansión de la generación<sup>8</sup>, que en la práctica son tres variaciones del último informe de precio de nudo publicado por la misma entidad. De este modo el plan de obras que se proponga para robustecer el sistema de transmisión troncal se encuentra acoplado a los escenarios de generación que se definan. Es por ello que el cumplimiento de al menos las centrales que se encuentran en construcción, es un factor importante en la definición de escenarios lo más realista posible. Pese a que el atraso de una central en uno o dos años pareciese un problema de corto plazo, si esto se repite en el tiempo y con muchos proyectos, el problema pasa a ser de largo plazo; de ahí la importancia del cumplimiento al menos de la fecha de entrada en operación de las centrales que se encuentran en construcción. Más crítico es el caso del incumplimiento de las obras en transmisión que se encuentran en construcción, ya que esto se ocupa como punto de partida para la expansión de los STT.

Un sistema de transmisión con deficiencias puede transformarse en una barrera de entrada, especialmente cuando se estudian proyectos de gran tamaño, teniendo incidencia directa en la evaluación técnico-económica de los proyectos.

Finalmente, se debe considerar que bastantes estudios utilizan como dato casi determinístico los planes de obras publicado por la CNE, manejando la incertidumbre mediante la

---

<sup>8</sup> Según el último estudio de transmisión troncal referido al periodo 2010-2014

construcción de escenarios que se basan en la entrada o no de ciertos proyectos, pero muy pocas veces moviendo las fechas de entrada en operación de las centrales.

### **2.9.3. Precio Básico de la Energía**

En cada informe de precio de nudo de corto plazo emitido por la CNE semestralmente, los meses de Abril y de Octubre de cada año, se calcula el precio básico de la energía y de la potencia, considerando un horizonte de simulación de 48 meses partiendo desde el mes de la fecha de emisión. Para ello se simula la operación considerando las obras en construcción que están prontas a entrar en los siguientes meses. Sin embargo, según la recopilación de información que se hace en este trabajo, las obras que se encuentran en etapa de construcción también presentan retrasos importantes en la fecha de entrada en operación que comunican a la CNE. Este último hecho provoca distorsiones en los valores calculados, dando señales de precio erróneas al mercado.

# Capítulo 3 – Metodología propuesta

A continuación se detalla la metodología propuesta en este trabajo. Ésta se aplicará utilizando la información pública disponible, teniendo en cuenta sus limitaciones y considerando que se pueden construir modelos más exactos si se dispusiese de una base de datos más robusta.

## 3.1. Descripción general de la metodología

En este trabajo se usará una metodología similar a la del proceso de extracción de conocimiento. El esquema se ilustra en la Figura 3.1.

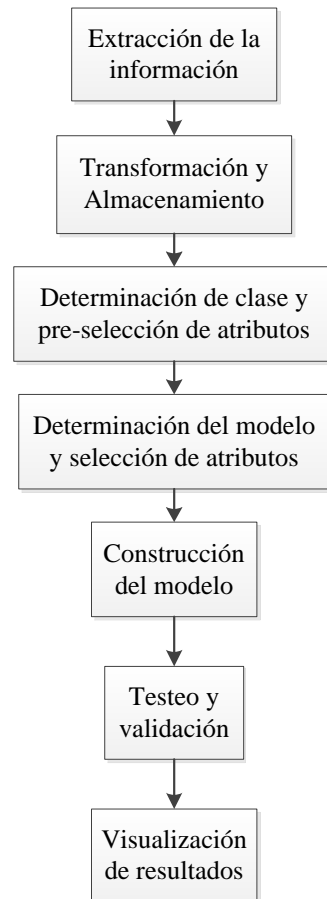


Figura 3.1 Esquema de la metodología propuesta



En los siguientes subcapítulos se explica de modo detallado en qué consiste cada fase del procedimiento. Sin embargo, a continuación se explicará de modo general cada etapa y cómo se tratará en este trabajo.

Primero se parte por la extracción de la información, que se recopilará de modo manual debido a que como se explicó en los antecedentes, no existe una herramienta que permita obtener la información necesaria de un modo automático o semiautomático. Para ello, se recurrió a las fuentes de información descritas en el capítulo 2.8.

Luego de extraída la información, se transforma y almacena en las bases de datos que fueron diseñadas para las fichas de proyectos mencionada en el capítulo 2.5.

En el proceso de transformación se procedió a realizar modificaciones que permitieran generar nuevas variables a partir de la información disponible. Así se calcularon diferencias de fechas, se agruparon variables numéricas en intervalos y categorizaron las variables binarias, debido a que los modelos utilizados trabajan con variables cualitativas.

Para la etapa de almacenamiento se hicieron algunos ajustes a las bases de datos del sistema de monitoreo, a causa de los requerimientos surgidos de la aplicación de la metodología. El cambio más importante fue crear tablas donde estuviera condensada toda la información sobre los proyectos, debido principalmente a que el software de análisis estadístico utilizado [22] no permite trabajar con bases de datos multirelacionales.

Una vez construidas las tablas a analizar, se procedió a ubicar las variables en estudio o las “Clases” en las últimas columnas. Esto debido a que como el software es orientado a objetos, ve a una tabla como un arreglo de vectores, donde cada uno tiene definida una posición dentro de la matriz; esto facilita la automatización del procedimiento, ya que se define que por *default* la clase está ubicada en el vector  $n$ , con  $n$  el número de columnas de la tabla. En este estudio, las clases o variables en estudio son “Atraso de la construcción” y “Construcción”. Al resto de las variables se les llamará “Atributos”.

En la tercera etapa se partió realizando una pre-selección de los atributos más significativos que permitieran explicar las variables en estudios. Para ello se hicieron  $n - 1$  test de chi-cuadrado entre la clase y los  $n - 1$  atributos, filtrando los atributos en los cuales el test arrojara un p-valor menor a 0.05. Para la selección final de los atributos se aplicó el método de forward, el cual elimina la redundancia que pueda existir entre los atributos preseleccionados.

Una vez seleccionados los atributos se procedió a aplicar el modelo de regresión logística multinomial. También se aplicaron otros modelos predictivos al conjunto de datos con el fin de comparar los porcentajes de aciertos de cada uno.

Debido a la escasa información obtenida no se hizo una validación cruzada de los datos<sup>9</sup>, ya que al seleccionar un subconjunto de los datos se podía perder información relevante. Pese a que la validación *leave-one-out* (dejar un dato afuera para construir el modelo con los datos restantes, para luego testear el modelo en el dato excluido) podía arrojar resultados más precisos, se descartó por los altos requerimientos computacionales que requiere. Por esto, la validación consistió en analizar la tasa de aciertos y observar donde existían problemas.

## **3.2. Detalle de etapas**

Como se mencionó anteriormente, este es el procedimiento general propuesto para el problema en estudio. Este trabajo se enfocará principalmente en el sector de la generación, utilizando como insumo principal los informes de precio de nudo y el sistema de evaluación de impacto ambiental.

### **3.2.1. Extracción de la información**

En el marco del sistema de monitoreo de mercado energético, se propone un sistema semiautomático de extracción de la información, ya que actualmente todo se ingresa de modo manual. Cabe destacar que el proceso no puede ser completamente automático, debido a la forma en la cual se presenta la información en las distintas fuentes y a la dificultad permanente de homologar la información para su sistematización y almacenamiento en bases de datos estructuradas.

#### **3.2.1.1. Sistema semiautomático de extracción**

Para poder lidiar con el constante problema que significa la actualización de la información es que aquí se propone un sistema semiautomático de extracción de la información.

Como se revisó en los antecedentes, si bien existen algunas iniciativas que se enuncian en la prensa, estos anuncios son poco específicos en cuanto a las características que tendrá el proyecto, enfocándose más en los requerimientos de capital que presenta la iniciativa.

---

<sup>9</sup> *Cross-validation*

Es por esta razón, que se utilizará la base de datos disponible en la página web del sistema de evaluación de impacto ambiental (e-seia), ya que en esta instancia los proyectos ya pasaron la ingeniería conceptual y se encuentran en la etapa de pre-factibilidad o factibilidad técnica. Además, gran parte de los atributos que podrían ser de interés ya se encuentran definidos a esta altura.

Se propone que para automatizar el proceso de detección del ingreso de un nuevo proyecto al sistema, se pueden utilizar método de extracción automática de archivos propuestos en la memoria de Felipe Mansilla [16] y establecer un procedimiento que vaya descargando la base de datos del SEIA y comparándola con la descargada el día anterior; de este modo, si la rutina hecha con éste propósito detecta una diferencia, entonces se envía una alerta para el encargado de llenar la ficha, de que existe un nuevo expediente en el SEIA.

Existen varias razones por las cuales no se puede llenar la ficha de proyectos directamente desde la base de datos del SEIA, entre ellas se encuentran:

- Al sistema sólo se ingresan las centrales de generación de electricidad mayores a 3 MW de potencia. Si bien las centrales cuya potencia es menor a este número deben conseguirse el resto de los permisos sectoriales, la información se encuentra dispersa entre los distintos servicios encargados de tramitar los permisos.
- La mayoría de la información no se encuentra sistematizada en la base de datos disponible para la descarga, lo que dificulta el llenado de la ficha. La información que se encuentra homologada y estructurada son la región, la inversión, el nombre del proyecto, la empresa, el tipo de tramitación en el SEIA y la fecha de ingreso del proyecto. El resto de los atributos, como lo son por ejemplo, el tipo de tecnología, el combustible a utilizar, la potencia instalada, la construcción por etapas, la existencia previa de alguna obra en el lugar donde se emplaza el proyecto, si es ampliación o es nuevo, etc. se desprende de la lectura minuciosa de los expedientes.
- Al SEIA se ingresan proyectos cuyo uso es exclusivo de la empresa que solicita la aprobación, no teniendo ninguna intención de conectar la central a algún sistema interconectado. En general se trata de motores diesel o calderas de biomasa o cogeneración, en donde las empresas utilizan los residuos para generar la energía necesaria para abastecer sus propios procesos o parte de ellos. Hay que notar que algunas centrales sí se conectan a un sistema interconectado, pero inyectan

solamente los excedentes de energía. Por esta razón se necesita de una lectura minuciosa del expediente y aplicar criterio.

- Existen otras instancias para ingresar centrales, como por ejemplo, cuando es necesaria la construcción de algún embalse. Aquí se prioriza el hecho de la construcción del embalse, por lo que hay que revisar esta sección. Sin embargo, las centrales que necesitan embalses son de tal envergadura, que son anunciadas en distintos medios, por lo que queda en manos de la persona encargada estar al tanto de las novedades del sector.
- Existen centrales cuyo fin no es operar en base y con ello obtener ganancias en la venta de energía, sino que servir como respaldo en caso de existir alguna emergencia que justifique su operación. Es importante esta diferencia, ya que la evaluación del proyecto no se realiza de la misma forma o se consideran otros tipos de parámetros para evaluar la conveniencia.
- Existen centrales menores a 9 MW que se conectan como PMGD a los sistemas de distribución, por lo cual los CDEC no llevan un registro exhaustivo de las instalaciones, ya que estas centrales se pueden auto-despachar. Esto dificulta el seguimiento del proyecto después de obtenida la evaluación ambiental.

### **3.2.1.2. Sistema de alarma**

Como el SEIA no es el único trámite que deben realizar los proyectos para materializarse, es que en base a los permisos identificados como conflictivos según el estudio [18], se propone un sistema de alarmas que vaya preguntando a las instituciones encargadas o personas responsables dentro de ellas el estado de estos permisos. Este consiste en un sistema de envío de correos automáticos con consultas sobre el estado de los permisos que se activa una vez terminado el plazo legal para la resolución por parte de la autoridad. En caso de no existir un plazo establecido en la normativa aplicable, se pueden enviar consultas cada cierto intervalo de tiempo predefinido.

Para la implementación de este sistema es necesaria la base de datos con los antecedentes de contacto de las personas encargadas en cada institución, además de un servidor de correos que utilice el estándar SMTP para la transferencia de mensajes. Como este sistema está montado en

una plataforma web que utiliza el framework MVC Cake PHP, entonces se utilizarán las funcionalidades de este framework para la automatización de estos mails.

### **3.2.2. Almacenamiento y transformación**

#### **3.2.2.1. Atributos**

En estadística, los atributos son variables que no se pueden cuantificar o expresar mediante expresiones numéricas. En otras palabras, son las características cualitativas que describen una determinada clase. En el caso de los proyectos eléctricos, ejemplos de atributos pueden ser el combustible que utiliza la central para operar, el tipo de turbina, la zona donde está ubicada, la empresa propietaria de las instalaciones, etc.

Algunos atributos se pueden obtener directamente desde la base de datos que genera la ficha de proyectos, sin embargo existen otros atributos que son generados a partir de aplicar funciones o transformaciones a un subconjunto de variables. Hay que volver a hacer hincapié, que como se estudiará un modelo de regresión logística, se deberán transformar las variables de cualquier tipo (binarias, booleanas, numéricas, fechas, etc.) en atributos. Así estas funciones o transformaciones pueden ser producto de alguna definición hecha en la normativa o en la legislación, pueden ser algoritmos de agrupación de variables numéricas en intervalos, manejo de fechas, etc., todo con el fin de agrupar un subconjunto de datos que cumplen características similares en relación a un parámetro.

#### **3.2.2.2. Clases**

Las clases son las variables en estudio y de las que se desea establecer un juicio sobre su comportamiento. Estas clases pueden ser tanto numéricas como nominales, pero para efectos de este trabajo las clases sólo serán de carácter nominal, ya que el enfoque estará puesto en los resultados del modelo de regresión logística.

En caso de existir variables numéricas que se requieran estudiar por la vía de la regresión logística, ésta se tendrá que agrupar en rangos siguiendo algún algoritmo (aunque sea simple) de agrupación. Esto se logra mediante una rutina de transformación de la variable numérica en variable nominal.

### **3.2.3. Obtención de modelos**

#### **3.2.3.1. Análisis exploratorio**

Con el fin de poder cuantificar de mejor manera el efecto que tienen los distintos atributos sobre las clases en estudio, es que se tiene que realizar un análisis exploratorio de datos. En este análisis se intenta establecer regularidades, singularidades y patrones que siguen los objetos en estudio en función de los atributos que se pueden obtener directamente y de otros atributos que se pudiesen generar a partir de los primeros. Junto con lo anterior, se nominalizan las variables binarias, se hacen diferencias entre las fechas y se asocian las variables numéricas en intervalos que tengan sentido práctico con el fin de nominalizarlas.

Como parte del análisis exploratorio, fueron utilizadas las herramientas de visualización que se proponen en el título 3.2.4, ya que por ejemplo, los árboles de decisión como el J48 admiten tanto variables numéricas como nominales, permitiendo establecer los puntos de corte de las variables numéricas que definen en qué categoría cae la clase en estudio.

Además de todo lo mencionado, hay que tener especial cuidado en que las categorías de cada atributo engloben a un número significativo de datos, debido a que de no ser así, se pueden cometer errores tanto en la selección de atributos, como en la construcción del modelo.

#### **3.2.3.2. Selección de atributos**

Una vez lista la tabla con la mayor cantidad de atributos que pueden obtener directa o indirectamente de la información disponible y que eventualmente puedan ser de aporte a la explicación del fenómeno en estudio, se procede a definir cuáles son las clases y sus categorías o instancias.

El siguiente paso consiste en seleccionar los atributos que logren explicar la clase definida, en función del modelo que se quiere obtener. Como el modelo de regresión logística utiliza variables nominales, se usará el test chi-cuadrado para validar la hipótesis de dependencia entre las clases y los atributos y el método forward, para determinar el set final de atributos que serán utilizados para la construcción del modelo de regresión.

#### 3.2.3.2.1. Definición de clases

Como fue descrito con anterioridad, las clases son las variables en estudio para las cuales el modelo entregará una predicción. En este caso, las variables de las cuales se quiere entregar un juicio son: la materialización de un proyecto y el retraso que presentaría un anuncio de puesta en servicio.

Junto con definir las clases se debe definir cuáles son las instancias que pueden tomar. Si la cantidad de categorías de una clase es igual a dos, entonces se le llamará variable dicotómica y se utilizará un modelo de regresión simple. Si la cantidad de categorías es mayor que dos, entonces se utilizará la extensión multivariante de la regresión logística clásica, también llamada regresión logística multinomial.

#### 3.2.3.2.2. Test de hipótesis $\chi^2$

Para el caso de las variables cualitativas se utiliza el test de chi-cuadrado ( $\chi^2$ ), en donde la hipótesis nula es la independencia que existe entre dos variables que se comparan. Si el p-valor es menor a una cierta cantidad previamente determinada (y que generalmente se fija en  $p = 0.05$ ), entonces se rechaza la hipótesis nula y se concluye que existe una dependencia entre las dos variables que se estudian.

Para ello se tienen que construir  $n$  tablas de contingencia, con  $n$  la cantidad de variables que definen los proyectos en estudio. Una vez construidas estas tablas se procede a aplicar el test chi-cuadrado y se ordenan las variables desde la que tiene el p-valor más pequeño hasta la que tiene el p-valor más grande, en orden creciente. Como se quiere determinar cuáles son las variables de las que dependen las clases, es que sólo se consideran las que entregan un p-valor menor a 0,05.

#### 3.2.3.2.3. Método Forward

Una vez que las variables están ordenadas de menor a mayor p-valor, entonces se ingresan a un método de selección de atrás hacia adelante (forward), el cual definirá cuáles son las variables que se utilizarán para la construcción del modelo definitivo.

### 3.2.3.3. Obtención del modelo

#### 3.2.3.3.1. Modelo de regresión logística

Estos modelos de regresión parten de la idea que existen un conjunto de variables que permiten explicar el comportamiento de una variable específica. En el caso de este proyecto, se asume que existen un conjunto de factores que explican cuan plausible es un proyecto eléctrico de desarrollarse o no.

Los modelos de regresión Logit y Probit se distinguen del modelo de regresión lineal general en que la variable a explicar corresponde a una variable dicotómica (proyecto se construye/no se construye), lo que se refleja tanto en la forma del modelo paramétrico como en sus supuestos. Los modelos Logit y Probit suponen que existe una variable aleatoria latente continua,  $u$ , que puede ser explicada a partir de variables predictoras  $x_1, x_2, \dots, x_{p-1}$  a través de la relación:

$$u = \beta_0 + \beta_1 x_1 + \dots + \beta_{p-1} x_{p-1} + \epsilon$$

donde la aleatoriedad viene dada por la componente de error  $\epsilon$ , siendo los componentes restantes determinísticos. Ambos modelos suponen que al observarse  $u > 0$ , se activa la función que determina uno de los resultados denominado como éxito, en este caso, la construcción del proyecto.

Los modelos Logit y Probit se diferencian en la especificación de la distribución de probabilidad del error,  $\epsilon$ . El modelo Logit, o de Regresión Logística, supone que la componente de error sigue una distribución Logística, con función de distribución:

$$F(u) = \frac{1}{1 + e^{-u}}$$

donde  $u$  es un número real. De este modo, bajo el modelo Logit, la probabilidad de éxito es calculada con la siguiente expresión:

$$P(\text{éxito}) = \frac{1}{1 + e^{-u}} = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \epsilon}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \epsilon}}$$

Por otra parte, el modelo Probit supone que la componente de error sigue una distribución Normal estándar con función de distribución  $\Phi(u)$ , donde  $u$  es un número real. De este modo, bajo el modelo Probit, la probabilidad de éxito es calcula con la siguiente expresión:



$$P(\text{éxito}) = \Phi(u) = \Phi(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)$$

donde  $\Phi$  corresponde a la función de distribución acumulada de la distribución Normal estándar.

### 3.2.3.4. Validación

#### 3.2.3.4.1. Validación del modelo

Una primera forma de validar los resultados entregados por el modelo obtenido, es aplicarlo sobre los mismos datos que lo construyeron para así obtener el porcentaje de aciertos. Además, para medir el rendimiento de los algoritmos se construye la llamada matriz de confusión, que es una tabla de contingencia en donde se comparan las instancias reales de la clase con las que predice el modelo.

Existen otras formas de validación como la validación cruzada<sup>10</sup>, la cual deja un porcentaje de datos fuera, se construye el modelo con el resto y luego se prueba el modelo con los valores que se dejaron apartados. Esto puede ser útil cuando se maneja un conjunto grande de datos, puesto que agiliza los cálculos, pero en los casos de estudio dejar afuera un subconjunto de datos podría implicar perder información importante. Ante este problema, se puede optar por la validación *leave-one-out*, en la cual se deja afuera un solo dato, se construye el modelo con el resto y se prueba con el dato apartado. Sin embargo, para efectos de este trabajo, se optó por no aplicar esta solución debido a los altos requerimientos de hardware que necesita.

#### 3.2.3.4.2. Test con set de proyectos en tramitación y construcción

Una vez obtenido el modelo y validado con los datos que lo construyeron, se procede a aplicarlo en un set de datos nuevos para verificar que las predicciones tengan sentido. Estas predicciones luego serán visualizadas con las herramientas que se describen a continuación.

### 3.2.4. Formas de visualización

Para ilustrar los resultados obtenidos de la aplicación del modelo propuesto, se visualizarán los resultados en tablas y gráficos de barras apilados para así, integrarlos al sistema de monitoreo de mercado actualmente operativo.

---

<sup>10</sup> Más conocida en la literatura científica como *k-fold cross-validation*

También se proponen indicadores estadísticos simples que permitan visualizar la información estructurada disponible de una manera más ilustrativa y de este modo apoyar la toma de decisiones. Entre los indicadores estadísticos más comunes se encuentran: la media, mediana, desviación estándar, etc. y otros test paramétricos. Sin embargo, se dará énfasis a otras herramientas de visualización como lo son: los Diagramas de Cajas, Histogramas y Gráficas de Medias.

Por último, se utilizarán los árboles de decisión, que se obtienen de aplicar otros modelos de predicción al conjunto de datos en estudio para así comparar estos resultados con los del modelo de regresión propuesto.

#### **3.2.4.1. Tablas y gráficos de barras apilados**

Como este modelo asigna una probabilidad a cada nivel de la variable clase, para mostrar los resultados se utilizará un gráfico de barras apilado, en donde todas las barras tendrán altura 1, que por definición es el valor de la suma de todas las probabilidades.

Para entregar la información un poco más clara se desplegará una tabla con todos los proyectos y los niveles de las clases, colocando un 0 si la probabilidad es menor que 0.5 y un 1 si la probabilidad es mayor o igual que 0.5. En el caso de que todas las probabilidades sean menores a 0.5 entonces se pondrán sólo ceros, lo cual indica la incerteza que existe sobre la predicción.

#### **3.2.4.2. Diagrama de cajas**

Estas herramientas gráficas permiten dar cuenta de cómo se concentran los datos, entregando información sobre los valores máximo, mínimo, los cuartiles  $Q_1$ ,  $Q_2$  (o mediana) y  $Q_3$ , la existencia de valores atípicos y la simetría de la distribución.

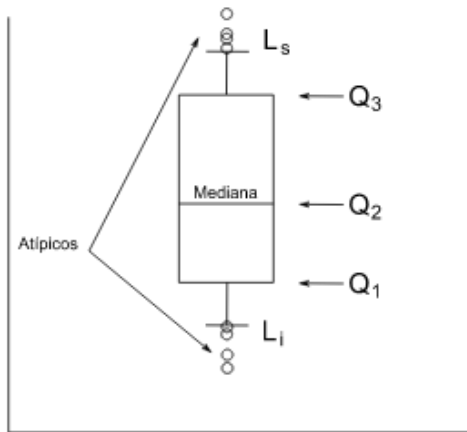


Figura 3.2 Diagrama de caja

### 3.2.4.3. Histogramas

Son representaciones gráficas de una variable en forma de barras, donde la superficie de cada barra es proporcional a la frecuencia de los valores representados. Estos histogramas pueden ser suavizados mediante el uso de Kernerls, sin embargo no se hará esto en este trabajo, ya que no se busca una distribución de probabilidades de los atrasos.

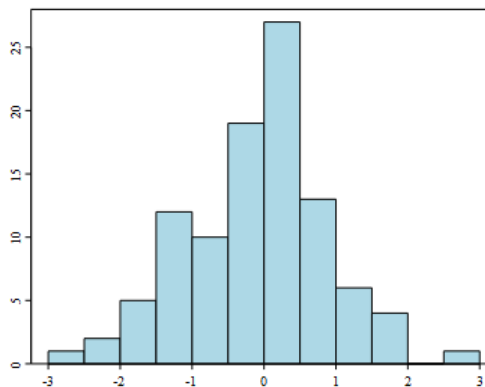


Figura 3.3 Histograma

### 3.2.4.4. Árboles de decisión

Aunque los árboles de decisión no permiten establecer una probabilidad de ocurrencia de las clases en estudio, son herramientas de visualización muy poderosas que nos entregan un juicio, de por ejemplo, si un proyecto se construye o no.

Al ser una herramienta gráfica que utiliza los datos provenientes de la información que se encuentra estructurada en la base de datos, es que podría ser de utilidad dentro de la suite de herramientas de visualización para el sistema de monitoreo de mercado.

### **3.3. Aspectos de la implementación**

#### **3.3.1. Softwares utilizados**

Para la implementación del procedimiento, y dado el contexto académico y de extensión del sistema de monitoreo de mercado ya instalado, se utilizará software de libre distribución en cada una de sus etapas.

- Para el proceso de extracción, transformación y almacenamiento de la información, se utilizará el software de Inteligencia de Negocios Kettle, de la suite Pentaho Data Integration [23]. Las transformaciones se pueden ordenar de manera secuencial mediante archivos llamados Jobs, los cuales se pueden ejecutar en segundo plano con Kitchen, el cual viene incluido dentro de la suite.
- El sistema de gestión de base de datos multi-relacionales que se utilizará será MySQL.
- Para la selección de atributos y la obtención de los modelos se utilizará el software estadístico R [22], el cual posee como característica importante, ser orientado a objetos y poder ejecutarse mediante script. Lo último da la opción de correr procedimientos en segundo plano y sin la necesidad de abrir ningún tipo de interfaz para ello.
- Para comparar los distintos métodos de clasificación utilizados en minería de datos, además de generar los árboles de decisión, se utilizará el software de minería de datos Weka [24].
- Finalmente para visualizar los resultados de aplicar los modelos predictivos a un conjunto de datos se utilizará el software Microstrategy, debido a que éste es el que actualmente se utiliza para la visualización de los indicadores del sistema de monitoreo de mercado.

Tanto Kettle como R, poseen la enorme ventaja de que se pueden correr en segundo plano, mediante la construcción de archivos por lotes ejecutables (\*.bat en Windows o \*.sh Linux) que contengan las líneas de código necesarias para llamar al programa y al archivo que se quiere ejecutar. Esto permite y facilita la automatización de la metodología, actualizando la información y ejecutando los procesos necesarios en momentos en los cuales los servidores se encuentran más desocupados. Esta flexibilidad se da, ya que la información necesaria no es actualizada tan frecuentemente, lo que permite holguras en este aspecto.

Tal como se propone en el trabajo de Mansilla [16], para automatizar los procesos y las rutinas de cálculo, se utilizará el programa Cron, el que se puede programar para correr archivos cada ciertos intervalos de tiempo.

Otra ventaja es que al ser los software multiplataforma, se puede testear su comportamiento en un computador de escritorio con Windows 7, para luego migrar los archivos a un servidor Linux (realizando los ajustes necesarios), que es donde se haya alojado el sistema de monitoreo.

### **3.3.2. Esquema del procesamiento**

El esquema general que se utilizará para el procesamiento de la información es el mostrado en la Figura 3.4. Este esquema sólo considera el procesamiento de los datos y la secuencia en la que se harán los pasos. Para representar los softwares que se utilizaron en cada etapa, se utilizaron los logos disponibles en sus páginas web oficiales.

Pese a que no aparece en el esquema, para automatizar todos estos cálculos propuestos, se utilizarán los archivos por lotes ejecutables, los cuales se correrán automáticamente gracias a la ayuda del software Cron, el cual tiene versiones tanto para testear con Windows como para implementar definitivamente en el servidor Linux.

Otro detalle que no está presente es la extracción de la información que se almacena en la base de datos, puesto que este esquema se propone en el título 3.2.1.1 de este trabajo.

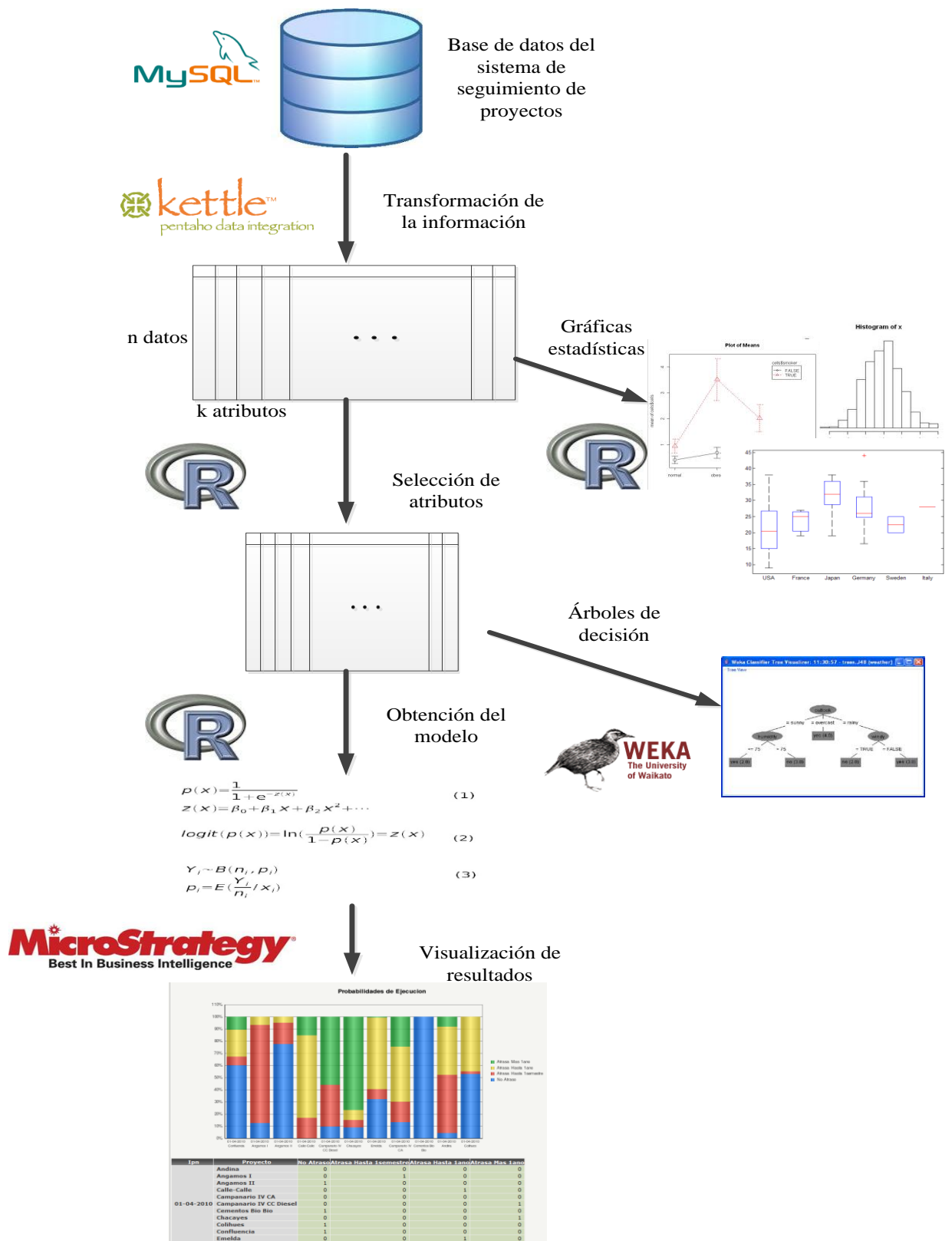


Figura 3.4 Esquema de implementación del procesamiento de la información propuesta

## Capítulo 4 – Resultados

A continuación se presentan los resultados de aplicar la metodología descrita en el capítulo anterior a los casos de estudio que se proponen. Se partirá por detallar los datos que se utilizaron, las transformaciones que se le hicieron a estos, además de describir las agrupaciones que se hicieron a las variables numéricas y explicar el criterio utilizado.

### **4.1. Aplicación de la metodología**

A continuación se hará una descripción de cómo se aplicó la metodología a los dos casos de estudio presentados en esta memoria, utilizando la información pública disponible.

#### **4.1.1. Atraso de las obras**

##### **4.1.1.1. Extracción de la información**

Para la construcción del modelo predictivo que permita dar un juicio sobre el atraso que podría presentar un anuncio de puesta en operación de una determinada obra, se utilizará como principal fuente de información los informes de precio de nudo que elabora la CNE semestralmente. Para ello se realizó una revisión de los planes de obras de todos los informes de precio de nudo publicados desde abril de 1990 hasta octubre de 2011 en el SIC. Para el caso del SING solo se revisaron los informes disponibles desde abril de 2004 hasta octubre de 2011, teniendo en cuenta que algunos planes de obras aparecen sin información debido a que hubo un periodo de tiempo en el cual este sistema se encontraba sobre-instalado. De estos documentos se obtiene directamente el nombre del proyecto, la potencia instalada, la fecha proyectada de entrada en operación y la fecha de emisión del informe de precio de nudo.

Así se recopilaron un total de 413 anuncios de proyectos claramente identificados en los informes de precio de nudo, que corresponden a 103 proyectos (lo que da un promedio de 4 anuncios por proyecto). En tanto para el caso de los proyectos ingresados al seia se consideraron 289 proyectos hasta la fecha. Con estos datos ya se logra apreciar que gran cantidad de proyectos

no se termina anunciando en los informes de precio de nudo, porque o bien no lograron financiamiento o eran especulaciones.

Finalmente, para obtener el resto de los atributos de los proyectos (región, empresa, tipo de tecnología, combustibles utilizados, fuente de energía primaria, etc.), se realizó un cruce de información con los catastros disponibles en las páginas de los CDEC y con la generación real de energía para obtener la fecha de entrada en operación real de los proyectos. En caso de que algo fuese ambiguo, se procedió a hacer el cruce de información con los expedientes web del SEIA.

#### 4.1.1.2. Transformación de los datos

La primera transformación que se hizo fue relativa a las fechas, debido a que estas por sí solas, no entregan información relevante. Por ello se parte definiendo las siguientes variables:

$$Atraso_{i,t} = FechaOp_{i,real} - FechaOp_{i,t}$$

$$Proyección_{i,t} = FechaOp_{i,t} - t$$

Donde:

$FechaOp_{i,real}$  Fecha de entrada en operación efectiva del proyecto  $i$ .

$t$  Fecha de publicación del informe de precio de nudo.

$FechaOp_{i,t}$  Fecha proyectada de entrada en operación del proyecto  $i$ , en el informe de precio de nudo publicado en  $t$ .

El resultado de estas diferencias se obtiene en días, sin embargo el pasarlo a meses, semestres o años es directo. Como los atributos tienen que ser nominales, se agruparon las diferencias de la siguiente manera:

Atraso:

- No se atrasa<sup>11</sup>
- Se atrasa hasta un año
- Se atrasa entre uno y dos años

---

<sup>11</sup> Se considera atraso cuando un anuncio se retrasa 4 meses o más de lo anunciado con respecto a lo real.



- Se atrasa más de dos años

Proyección:

- Para el mismo año
- Para el año siguiente
- Para dos años más
- De tres a cinco años más
- Para seis años o más

Como existen proyectos en construcción y recomendados que aún no están operativos, para considerar el efecto que ha presentado el retraso en sus anuncios, se consideró como fecha de entrada en operación la fecha proyectada en el último informe de precio de nudo correspondiente a Octubre de 2011. Por este último hecho se dejó fuera de los datos considerados en la construcción del modelo, la información sobre las centrales que aparecieron en el último informe de precio de nudo y que a la fecha aún no han entrado en servicio.

Otra variable numérica que se tuvo que agrupar, fue la potencia instalada de los proyectos. La agrupación hecha y el criterio fueron los siguientes:

- Mega-proyectos: Proyectos de capacidad instalada mayor a 300 MW
- Grandes centrales: Proyectos de capacidad instalada entre 100 y 300 MW
- Centrales medianas: Proyectos de capacidad instalada entre 40 y 100 MW
- Pequeñas centrales: Centrales menores a 40 MW

Otra agrupación que se hizo fue la que establece la normativa de PMG, en donde se agrupó los proyectos de potencia menor a 9 MW como PMG y los otros como NO PMG.

En el mismo sentido se creó la variable “ERNC”, en donde se clasificaron los proyectos como energía renovable no convencional, si cumplían con lo que establece la normativa al respecto. El resto de los proyectos se clasificaron como convencionales.

Ahora, como la variable empresa podía tomar varios valores y al ser los sistemas casi totalmente hidráulicos y térmicos, se clasificaron las empresas por tipo según el siguiente criterio:

- Grandes empresas (más de un 5% del mercado, considerando a las filiales).
- Empresas hidroeléctricas (empresas con una o más centrales hidroeléctricas).

- Otras empresas (Principalmente empresas que tienen motores diesel, turbinas a gas o parques eólicos).

En el caso de las regiones, se las consideró tal cual y además se agruparon en macro-zonas (sur, centro y norte) y como comúnmente se las agrupa en geografía (norte grande, norte chico, centro, centro-sur, sur y austral). El detalle de las agrupaciones por zona y macro-zona se encuentra en la Tabla 8.1.

Además se consideró un atributo que hiciera referencia a la existencia de otra central en el lugar donde se pretende construir el nuevo proyecto o no, junto con un atributo que dijese si una obra es una ampliación o es totalmente nueva ya que cuando ya existe infraestructura puede resultar más simple hacer ampliaciones que partir de cero. También se hizo alusión a si el anuncio estaba hecho en el plan de obras recomendado o estaba en la lista de centrales en construcción que aparecen en los informes de precio de nudo.

Otro atributo que se consideró fue uno que hiciera alusión al combustible principal que utiliza una central. De este modo se diferenció por Carbón, GN y Petróleo, además de Eólico, Biomasa, Pasada y Embalse (se hizo la distinción entre los últimos dos, dadas las diferencias que existen en su operación y las obras asociadas). Junto con ellos se creó un atributo que se denominó fuente, y que hace referencia a si la fuente es hidráulica, térmica o no convencional, ya que esta clasificación es muy usada en el sector.

Finalmente, se agruparon los proyectos por tecnologías. Dentro de las tecnologías que utilizan las centrales aquí en Chile tenemos que actualmente hay instalados aerogeneradores, turbinas hidráulicas, turbinas a vapor, turbinas a gas, motores y calderas de biomasa o cogeneración.

En resumen, la lista con todos los atributos que se utilizarán para construir la tabla que será procesada por el software estadístico, junto con todas las instancias de cada atributo, se muestra en la Tabla 4.1.

Tabla 4.1 Atributos candidatos a describir la Clase “Atraso”, junto con los niveles que pueden tomar.

<b>Atributo</b>	<b>Número de instancias</b>	<b>Instancias</b>
<b>Proyección</b>	5	Mismo año, Próximo año, año subsiguiente, de 3 a 5 años más, más de 6 años.
<b>Potencia segmentada</b>	4	Pequeñas centrales, centrales medianas, grandes centrales, mega-proyectos
<b>PMG</b>	2	PMG, No PMG
<b>ERNC</b>	2	ERNC, Convencional
<b>Combustible</b>	7	Eólico, Embalse, Pasada, Carbón, GN, Petróleo, Biomasa
<b>Fuente</b>	3	Hidráulica, Térmica, No Convencional
<b>Tecnología</b>	7	Aerogenerador, Caldera de biomasa, Turbina Vapor, Turbina Gas, Turbina Hidráulica, Motores
<b>Tipo de Empresa</b>	3	Grandes empresas, Empresa hidroeléctrica, Otras empresas
<b>Región</b>	15	XV,I,II,III,IV,V,VI,VII,VIII,IX,XIV,X,XI,XII,RM
<b>Zonas clásicas</b>	6	Norte Grande, Norte Chico, Centro, Centro-Sur, Sur, Austral
<b>Macro-zonas</b>	3	Norte, Centro, Sur
<b>Clasificación</b>	2	Ampliación, Nuevo
<b>Existencia</b>	2	No existía, Existía
<b>Construcción o Recomendada</b>	2	En Construcción, Recomendada
<b>Atraso</b>	5	No atraso, Atraso hasta un semestre, Atraso entre un semestre y un año, Atraso entre uno y dos años, Atraso más de dos años

#### 4.1.1.3. Selección de atributos

De la aplicación del test de chi-cuadrado los atributos que arrojaron un p-valor fueron los mostrados en la Tabla 4.2.

Tabla 4.2 p-valor resultante de los test chi-cuadrados realizados entre la clase atraso y los atributos pre-seleccionados.

<b>Orden de significancia</b>	<b>Atributo</b>	<b>P-valor</b>
<b>1</b>	Proyección	3.16E-12
<b>2</b>	Región	8.37E-11
<b>3</b>	Combustible	2.31E-09
<b>4</b>	Construcción o Recomendado	5.92E-08
<b>5</b>	Tipo de empresa	5.19E-06
<b>6</b>	Existencia	0.0002519
<b>7</b>	Potencia segmentada	0.00036107
<b>8</b>	Tecnología	0.00097712
<b>9</b>	Fuente	0.00770207
<b>10</b>	Zonas clásicas	0.01492127
<b>11</b>	Macro-zonas	0.02935191
<b>12</b>	Clasificación	0.0457385

Pese a que forma parte del análisis, nótese la fuerte dependencia que existe entre el atraso de un proyecto y la lejanía de la fecha proyectada con respecto a la fecha en la cual se hace el anuncio.

Luego de aplicar el método forward, las variables seleccionadas que explican de mejor manera, y sin existir redundancia entre ellas, el atraso de los proyectos utilizando el método de regresión logística multinomial son las mostradas en la Tabla 4.3. Nótese que el método elimina la redundancia entre las variables pre-seleccionadas que pueden estar correlacionadas.

Tabla 4.3 Atributos seleccionados para el modelo de regresión multinomial.

<b>Orden</b>	<b>Atributo</b>
<b>1</b>	Proyección
<b>2</b>	Región
<b>3</b>	Combustible
<b>4</b>	Construcción o Recomendado
<b>5</b>	Tipo de empresa
<b>6</b>	Existencia
<b>7</b>	Potencia segmentada
<b>8</b>	Tecnología

#### 4.1.1.4. Modelo de regresión logística multinomial

Debido a que las categorías que puede tomar la clase “Atraso de la construcción” son un número mayor que dos<sup>12</sup>, entonces se utilizará la extensión multivariante de la regresión logística binaria<sup>13</sup>.

Dado que la cantidad de atributos seleccionados es numerosa, el escribir la expresión para la probabilidad que toma cada uno de los niveles de la clase resulta engorroso y no tiene mayor importancia para efectos de lo que se quiere ilustrar. Sin embargo, en la Tabla 4.4 se muestra la matriz de confusión, que consiste en una tabla de contingencia, contiene los valores reales que toma la clase en estudio y las predicciones hechas por el modelo. En la diagonal se encuentran marcados con negritas los aciertos hechos por el modelo.

Tabla 4.4 Matriz de confusión del modelo de regresión logística multinomial obtenido.

	Predicción	No se atrasa	Hasta 1 año	Entre 1 y 2 años	Más de 2 años
Real					
No se atrasa		<b>222</b>	16	11	2
Hasta 1 año		55	<b>22</b>	8	6
Entre 1 y 2 años		23	2	<b>24</b>	2
Más de 2 años		8	0	3	<b>9</b>

#### 4.1.1.5. Validación

El porcentaje de aciertos del modelo predictivo fue de un 70%, tomando en cuenta que se consideró como predicción el valor de la clase que tenía la mayor probabilidad. De una revisión más minuciosa, se constató que en los errores de la predicción muchas veces el valor real de la clase era el que tenía la segunda mayor probabilidad, donde en algunos casos la segunda probabilidad más alta era muy parecida a la mayor probabilidad.

De este último hecho radica la importancia de no sólo entregar la predicción, sino que también la probabilidad de que esa predicción sea cierta, ya que esta variable entrega más información sobre la proyección real del atraso considerando la historia pasada.

---

<sup>12</sup> A este tipo de variables se les llama también politómicas

<sup>13</sup> Conocida en la literatura como modelo de regresión logística multinomial

#### 4.1.1.6. Comparación con otros modelos.

Una vez programado el procedimiento general, la implementación de otro tipo de modelos predictivos es directa. Es por ello que se ocuparon los paquetes *nnet* y *rpart* para implementar una red neuronal artificial de una capa oculta y un árbol de decisiones, respectivamente.

A diferencia de la regresión logística que un tipo de regresión lineal generalizada, los árboles de decisiones y las redes neuronales son modelos no lineales que buscan relaciones e interacciones entre los atributos, para así lograr explicar la variable clase. Existen otros modelos, como los inspirados en la teoría de Bayes, que no son implementados debido a que no forma parte de la finalidad de este trabajo.

Dada la naturaleza del problema, y de lo difícil de establecer relaciones lineales entre la clase y los atributos, es que las predicciones hechas por los árboles de decisión y la red neuronal arrojaron modelos con errores de predicción mucho menores.

Tabla 4.5 Comparación de los resultados de los distintos modelos aplicados al caso de los atrasos de los anuncios

<b>Modelo</b>	<b>Porcentaje de aciertos</b>	<b>Atributos seleccionados</b>
<b>Regresión logística</b>	70%	Construcción/Recomendada, Región, Tipo Empresa, Existencia, Combustible, Potencia y Proyección.
<b>Red neuronal de una capa</b>	69%	Construcción/Recomendada, Región, Tipo Empresa, Existencia, Combustible y Potencia.
<b>Árbol de decisión</b>	68%	Construcción/Recomendada, Zona, Proyección, Fuente

#### 4.1.2. Materialización de la construcción

##### 4.1.2.1. Extracción de la información

El principal insumo utilizado para este caso de estudio fue la base de datos que se encuentra disponible en el sistema de evaluación de impacto ambiental. A partir de los expedientes se puede obtener la potencia, la inversión, la región, la empresa, el tipo de tecnología, la entrada al sistema, la fecha de la RCA, el estado de la tramitación y el sistema al cual se conectan las centrales.

Para verificar la entrada en operación de las centrales se consultó a los inventarios de los CDEC, además de la información que está disponible en las memorias de estas entidades. En caso de que una central aún no se encuentre operativa, se consultará el plan de obras de los últimos informes de precio de nudo.

#### 4.1.2.2. Transformación de los datos

Al igual que el caso de estudio anterior, se partirá transformando las fechas, ya que éstas por sí solas, contienen información que será de utilidad. Así se definirán las siguientes variables:

$$Tramitación = Fecha_{RCA} - Fecha_{Ingreso SEIA}$$

$$Retraso = \begin{cases} \max \{Tramitación - 60, 0\} & \text{si se tramita por DIA} \\ \max \{Tramitación - 120, 0\} & \text{si se tramita por EIA} \end{cases}$$

Donde:

$Fecha_{RCA}$  Fecha en la que el proyecto obtuvo la resolución de calificación ambiental (RCA).

$Fecha_{Ingreso SEIA}$  Fecha en la que se ingresaron los expedientes del proyecto al sistema de evaluación de impacto ambiental (SEIA).

De este modo se agruparán los atrasos como: Atraso hasta dos meses, Atraso entre dos y seis meses y Atraso de más de seis meses.

Las variables numéricas presentes en los expedientes del SEIA como lo son la potencia instalada y el valor de la inversión, se segmentaron para transformarlas en atributos. Se utilizó la misma clasificación de potencias empleado en el caso anterior (Pequeñas centrales, centrales medianas, grandes centrales y mega centrales) y el nivel de inversión se segmentó en tres: Pequeña inversión, inversión moderada y gran inversión.

También se utilizaron agrupaciones definidas en la legislación, como por ejemplo el atributo ERNC y el atributo PMG. Si un proyecto cumplía con la definición de ERNC que aparece en la Ley, entonces tenía la categoría ERNC; en caso contrario, se le asignó la categoría convencional. Se hizo algo análogo para el caso de PMG.

Al igual que el caso anterior también se definieron los atributos de localización como lo son la región del proyecto, la macro-zona y la zona a la cual pertenece.

Otra información que se puede rescatar de los expedientes web es el tipo de tecnología que utiliza cada central para su operación. De este modo se clasificaron las centrales como: solares,

eólicas, de pasada, de embalse, turbina a vapor, turbina a gas y motor diesel. Como se ha ingresado sólo una central geotérmica y ésta aún está en calificación, entonces no se consideró esta instancia del atributo.

Debido a la gran cantidad de titulares que han ingresado proyectos al SEIA, se realizó una clasificación basada en la intuición. Así se agruparon las grandes empresas, las empresas hidroeléctricas, otras empresas (fundamentalmente nuevas empresas de energía solar y eólica y de motores diesel), agro-forestal (centrales a biomasa y de respaldo diesel), industrial (centrales de respaldo diesel o fuel oil), mineras (del mismo tipo que las industriales más proyectos renovables) y personas particulares (especialmente proyectos mini-hidro).

La clase en estudio en este caso sólo toma dos valores, que son “No construido” y “Construido/En Construcción”.

Finalmente todos los atributos considerados junto con las instancias de cada uno de ellos se resumen en la Tabla 4.6.

Tabla 4.6 Atributos para el caso de estudio sobre la materialización de los proyectos.

<b>Atributo</b>	<b>Número de instancias</b>	<b>Instancias</b>
<b>Inversión segmentada</b>	3	Pequeña inversión, inversión moderada, gran inversión
<b>Potencia segmentada</b>	4	Pequeñas centrales, centrales medianas, grandes centrales, mega-proyectos
<b>PMG</b>	2	PMG, No PMG
<b>ERNC</b>	2	ERNC, Convencional
<b>Tecnología</b>	7	Eólica, Solar, Turbina a Gas, Motor Diesel, Pasada, Embalse, Turbina Vapor
<b>Tipo de Titular</b>	3	Grandes empresas, Empresa hidroeléctrica, Nuevas empresas, Minería, Industria, Personas Naturales, Agro-Forestal
<b>Región</b>	15	XV,I,II,III,IV,V,VI,VII,VIII,IX,XIV,X,XI,XII,RM
<b>Zonas clásicas</b>	6	Norte Grande, Norte Chico, Centro, Centro-Sur, Sur, Austral
<b>Macro-zonas</b>	3	Norte, Centro, Sur
<b>Atraso Tramitación</b>	3	Atraso de hasta dos meses, Atraso de entre dos y seis meses, atraso de más de seis meses
<b>Construcción</b>	2	No construido, Construido/En Construcción



#### 4.1.2.3. Selección de atributos

Los atributos pre-seleccionados son los que se muestran en la Tabla 4.7.

Tabla 4.7 p-valor resultante de aplicar los test chi-cuadrado entre la clase construcción y cada uno de los atributos.

Orden de significancia	Atributo	P-valor
1	ERNC	5.23E-06
2	Tecnología	9.34E-05
3	Tipo de Titular	0.00473731
4	Atraso tramitación	0.00501921
5	PMG	0.0418911

Después de aplicar el método forward se elimina el atributo que hace referencia a si una central es PMG.

Tabla 4.8 Atributos seleccionados para explicar la clase construcción.

Orden de significancia	Atributo
1	ERNC
2	Tecnología
3	Tipo de Titular
4	Atraso tramitación

Notar que después de la selección de los atributos, la tabla que se procesará en el software estadístico disminuyó su dimensionalidad a la tercera parte de lo que era en un comienzo. Esto permite que los cálculos se efectúen con mayor velocidad.

#### 4.1.2.4. Modelo de regresión logística

Debido a que la clase puede tomar sólo dos valores (variable dicotómica), entonces se utilizó un modelo de regresión logística simple.

Al igual que en el caso anterior, para mostrar los resultado se mostrará en la Tabla 4.9 la matriz de confusión, que en este caso es una matriz de 2x2 en la que aparecen los falsos negativos, falsos positivos, verdaderos negativos y verdaderos positivos.

Tabla 4.9 Matriz de confusión para el caso de la materialización de los proyectos

	Predicción	No se construye	En construcción/ Construido
Real			
No se construye		<b>83</b>	17
En construcción/ Construido		17	<b>40</b>

De la Tabla 4.9 se puede observar que el porcentaje de aciertos de este modelo es del 78.3%.

#### 4.1.2.5. Validación

En este caso el error de las predicciones fue de aproximadamente un 21%. Como la clase era dicotómica, esta podía tomar dos valores. Analizando los resultados, las predicciones incorrectas se encuentran en todo el rango de probabilidades, pero se concentran en las probabilidades cercanas al 50% donde no existe mucha certeza cual de las dos categorías toma la variable clase.

#### 4.1.2.6. Comparación con otros modelos

Al igual que en el caso anterior, se ocuparon los paquetes *nnet* y *rpart* para implementar una red neuronal artificial de una capa oculta y un árbol de decisiones, respectivamente. Hay que tener en cuenta que estos modelos son no lineales.

Dada la naturaleza de estos modelos no fueron necesarios test de chi-cuadrado para descartar variables, simplemente se utilizó el método forward para seleccionar los atributos.

Considerando que el modelo de regresión logística es un subconjunto de los modelos de regresión lineal generalizados y la naturaleza del problema, es que los datos se ajustaron de mejor manera a estos modelos no lineales que considera la interacción entre los atributos seleccionados.

A pesar de que la predicción tiene menos error, la desventaja es que sólo proporcionan la predicción, sin dar la probabilidad de que esa predicción sea correcta. En la Tabla 4.10 se presenta un cuadro comparativo con los distintos resultados de los modelos aplicados.

Tabla 4.10 Tabla que compara los distintos modelos de predicción para el caso de estudio acerca de la materialización de los proyectos eléctricos.

<b>Modelo</b>	<b>Porcentaje de aciertos</b>	<b>Atributos seleccionados</b>
<b>Regresión logística</b>	78%	ERNC, Tecnología, Tipo de Titular
<b>Red neuronal de una capa</b>	93%	Inversión, Atraso Tramitación, Región, ERNC, PMG, Tecnología, Tipo de Titular
<b>Árbol de decisión</b>	82%	Región, Tecnología, Tipo de Titular

## **4.2. Visualización**

### **4.2.1. Resultados**

#### **4.2.1.1. Atraso de las obras**

Para visualizar los resultados de aplicar el modelo predictivo obtenido a un conjunto de datos, se utilizarán los proyectos que se encuentran en etapa de construcción o en el plan de obras recomendados por la CNE en el último informe de precio de nudo, que a la fecha corresponde al publicado en Octubre de 2011.

Las visualizaciones son hechas con el software Microstrategy, para así poder integrarlas al sistema de monitoreo de mercado. Se muestran gráficos de barra con la probabilidad que tiene cada instancia de la clase atraso. Además se pone debajo una tabla que muestra la predicción del modelo.

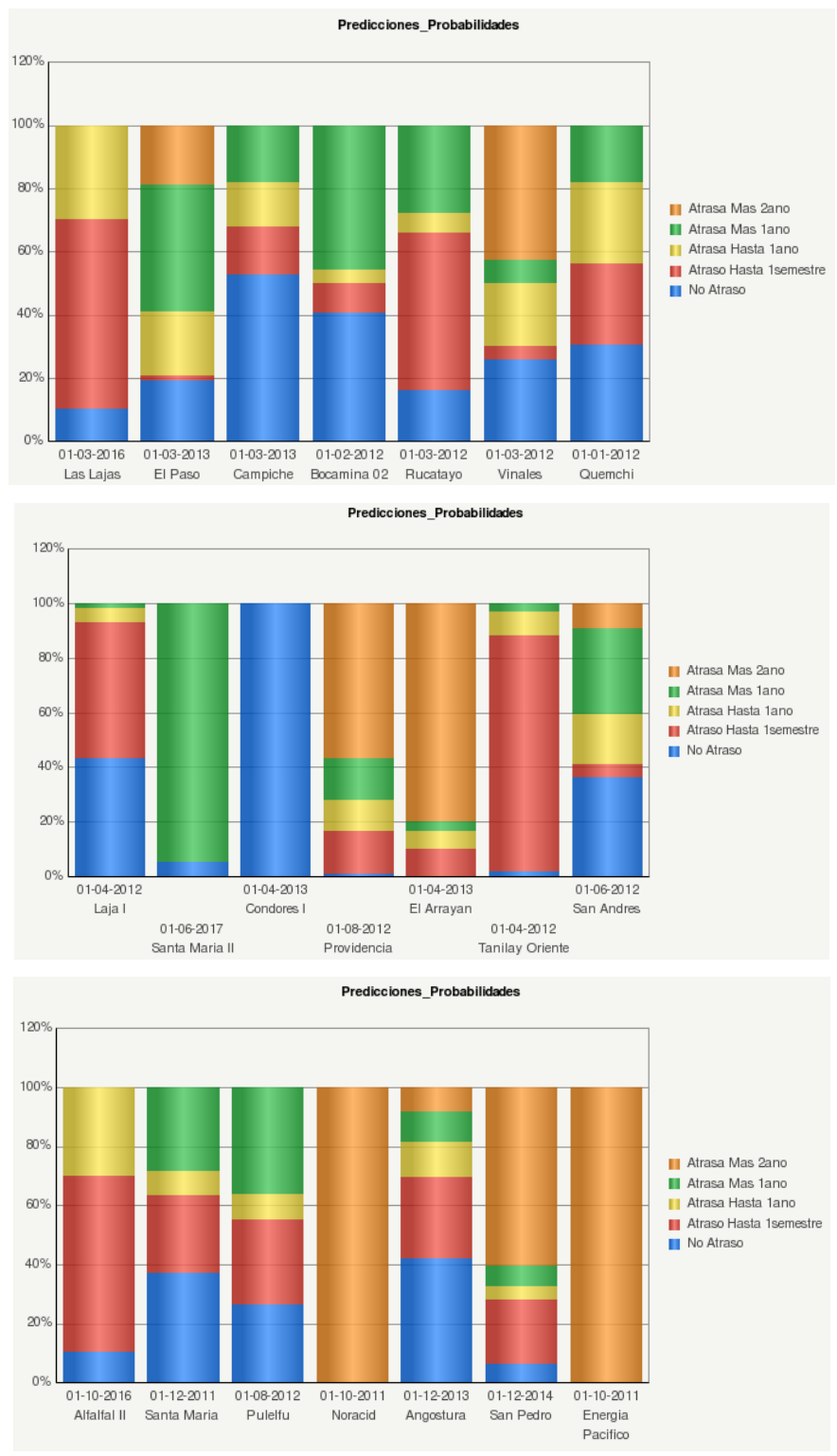


Figura 4.1 Resultados de aplicar el modelo de regresión logística a las obras que se encuentran en construcción en el Informe de Precio de Nudo de Octubre de 2011

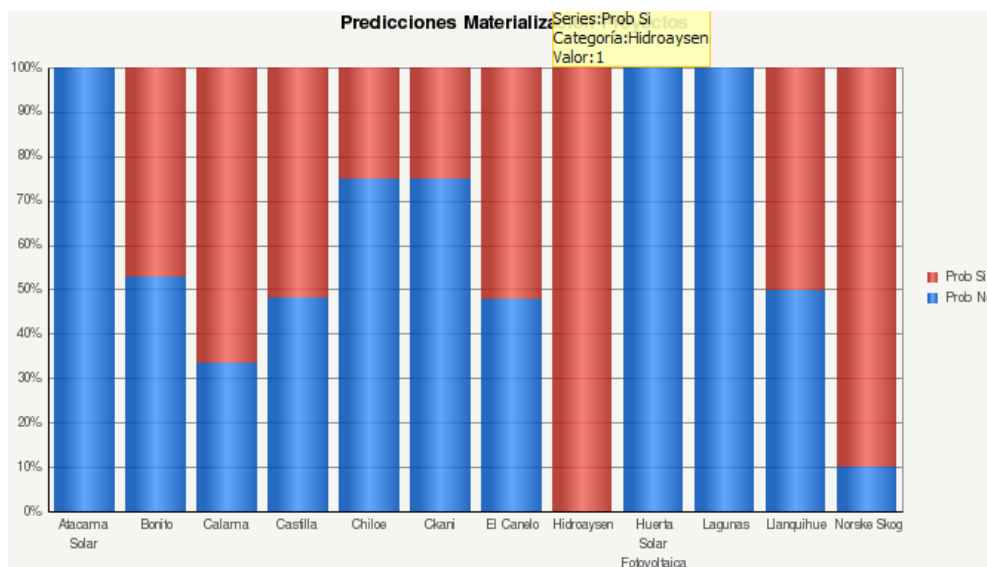
Fecha Proyectada	Proyecto	No Atraso	Atraso Hasta 1 semestre	Atrasa Hasta 1 año	Atrasa Mas 1 año	Atrasa Mas 2 años
01-01-2012	Quemchi	0	0	0	0	0
01-02-2012	Bocamina 02	0	0	0	0	0
01-03-2012	Rucatayo	0	0	0	0	0
	Vinales	0	0	0	0	0
01-03-2013	Campiche	1	0	0	0	0
	El Paso	0	0	0	0	0
01-03-2016	Las Lajas	0	1	0	0	0
01-04-2012	Laja I	0	0	0	0	0
	Tanilay Oriente	0	1	0	0	0
01-04-2013	Condores I	1	0	0	0	0
	El Arrayan	0	0	0	0	1
01-06-2012	San Andres	0	0	0	0	0
01-06-2017	Santa Maria II	0	0	0	1	0
01-08-2012	Providencia	0	0	0	0	1
	Pulelfu	0	0	0	0	0
01-10-2011	Energia Pacifico	0	0	0	0	1
	Noracid	0	0	0	0	1
01-10-2016	Alfalfal II	0	1	0	0	0
01-12-2011	Santa Maria	0	0	0	0	0
01-12-2013	Angostura	0	0	0	0	0
01-12-2014	San Pedro	0	0	0	0	1

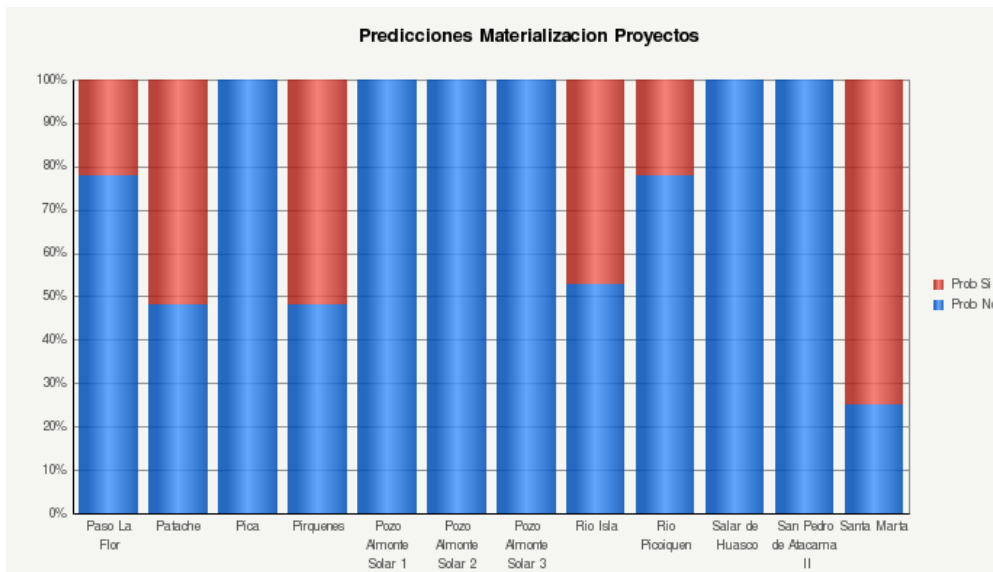
Figura 4.2 Tabla que muestra las predicciones hechas por el modelo, entregando un 1 si la mayor probabilidad es superior al 50%

#### 4.2.1.2. Materialización de la construcción

Para visualizar los resultados de aplicar el modelo predictivo obtenido a un conjunto de datos, se utilizarán los proyectos aprobados ambientalmente de Enero de 2011 a la fecha.

Las visualizaciones son hechas con el software Microstrategy, para así poder integrarlas al sistema de monitoreo de mercado. Se muestran gráficos de barra con la probabilidad que tiene cada instancia de la clase atraso. Además se pone debajo una tabla que muestra la predicción del modelo.





Nombre	Prediccion	Prob No	Prob Si
Atacama Solar	No se construye	1	0
Bonito	No se construye	1	0
Calama	Se construye	0	1
Castilla	Se construye	0	1
Chiloe	No se construye	1	0
Ckani	No se construye	1	0
El Canelo	Se construye	0	1
Hidroaysen	Se construye	0	1
Huerta Solar Fotovoltaica	No se construye	1	0
Lagunas	No se construye	1	0
Llanquihue	Se construye	0	1
Norske Skog	Se construye	0	1
Paso La Flor	No se construye	1	0
Patache	Se construye	0	1
Pica	No se construye	1	0
Pirquenes	Se construye	0	1
Pozo Almonte Solar 1	No se construye	1	0
Pozo Almonte Solar 2	No se construye	1	0
Pozo Almonte Solar 3	No se construye	1	0
Rio Isla	No se construye	1	0
Rio Picoiquen	No se construye	1	0
Salar de Huasco	No se construye	1	0
San Pedro de Atacama II	No se construye	1	0
Santa Marta	Se construye	0	1

Figura 4.3 Visualización de aplicar el modelo de regresión logística al caso de la materialización de los proyectos

#### 4.2.2. Árboles de decisión

En la Figura 4.4 se presenta el ejemplo de un árbol de decisión J48 aplicado al caso de la materialización de los proyectos, ya que presenta menos atributos significativos además de que la clase tiene solo dos categorías. Este caso simple permitirá ilustrar la utilidad que puede llegar a tener esta herramienta gráfica.

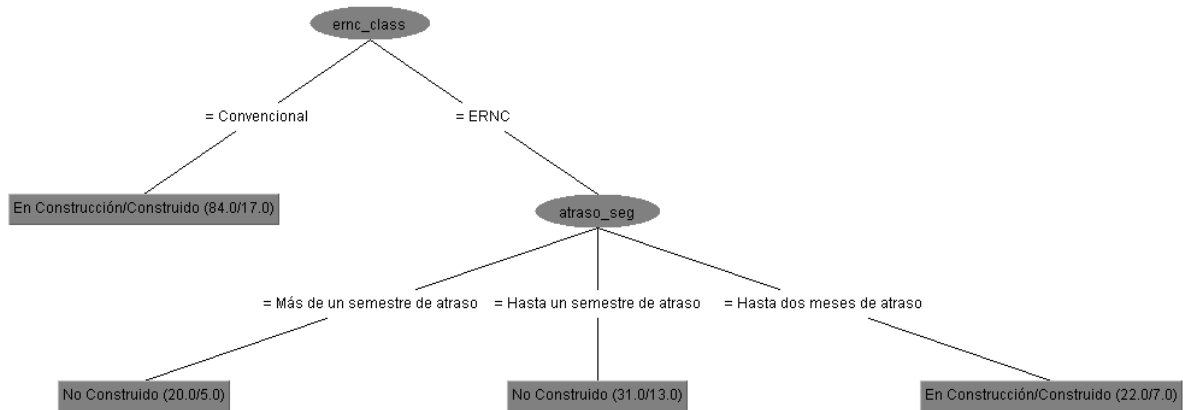


Figura 4.4 Árbol de decisión J48 aplicado al caso que trata la materialización de los proyectos

A modo de explicación este árbol de decisión primero pregunta si el proyecto corresponde a tecnologías convencionales o de ERNC. En base a la respuesta determina si sigue preguntando más características del proyecto o bien entrega un juicio de si el proyecto se termina construyendo o no. En el caso de los proyectos ERNC, además de su naturaleza también tiene relevancia para el modelo el atraso que presentó su estudio de impacto ambiental. Según la estadística, si el estudio resultó expedito entonces lo más probable es que el proyecto se construya.

Pese a que esta clasificación solo tiene un 70% de aciertos, lo cual es muy distante al porcentaje de aciertos que entregaron los otros modelos utilizados (79%, 83% y 89%), este ejemplo permite mostrar como a partir de reglas simples se puede generar una clasificación de modo de que sea visualmente atractivo y fácil de entender para el usuario que utiliza este sistema.

### 4.2.3. Diagramas de cajas

Como la única variable numérica son los atrasos en los anuncios que son realizados en los informes de precio de nudo, los cuales se pueden expresar en meses, en las Figura 4.5, Figura 4.6, Figura 4.7, Figura 4.8 y Figura 4.9 se muestran los diagramas de cajas de los atrasos con respecto a los atributos más significativos resultante de la etapa de selección. En la sección de Anexos 8.4 se encuentra un mayor detalle con todos los atributos definidos para esa clase.

Cabe hacer notar que estas gráficas se generan automáticamente utilizando el paquete *ggplot2* del software R.

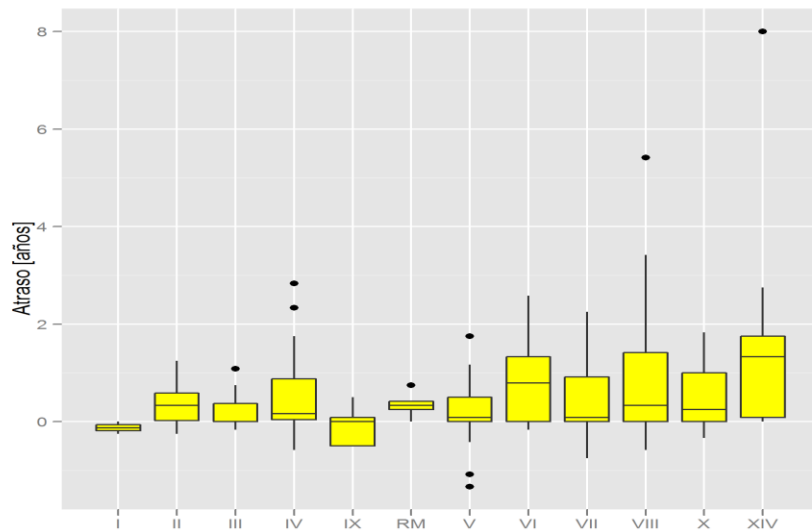


Figura 4.5 Boxplot del atraso con respecto a las distintas regiones

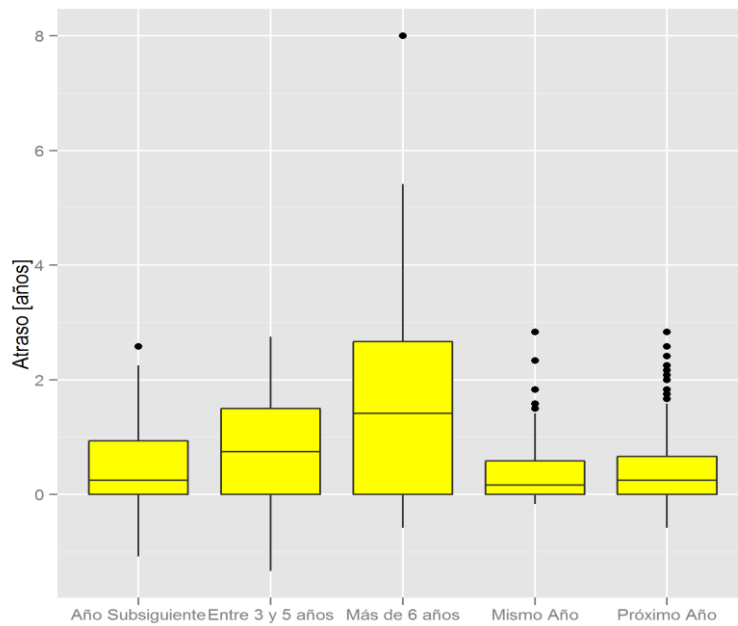


Figura 4.6 Boxplot del atraso con respecto a la proyección o lejanía de la fecha proyectada vs la fecha en la que se hizo el anuncio



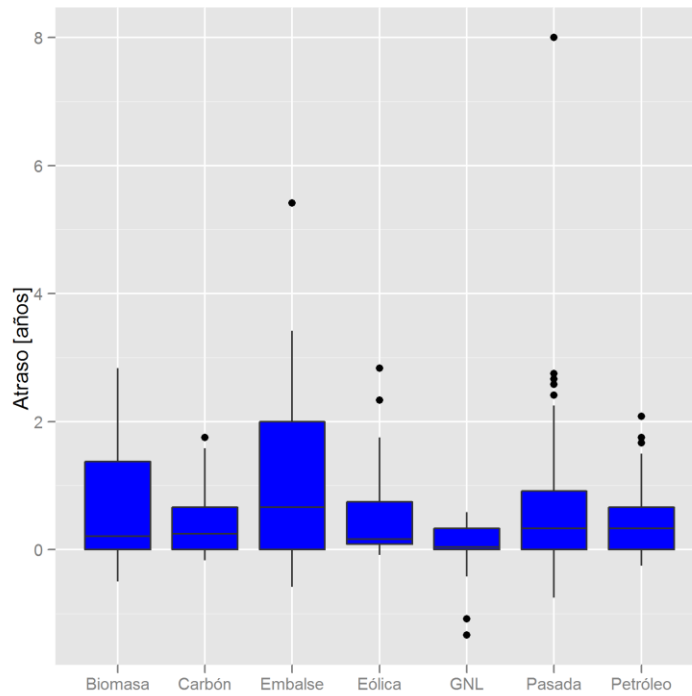


Figura 4.7 Boxplot del atraso con respecto al combustible principal que utiliza la central

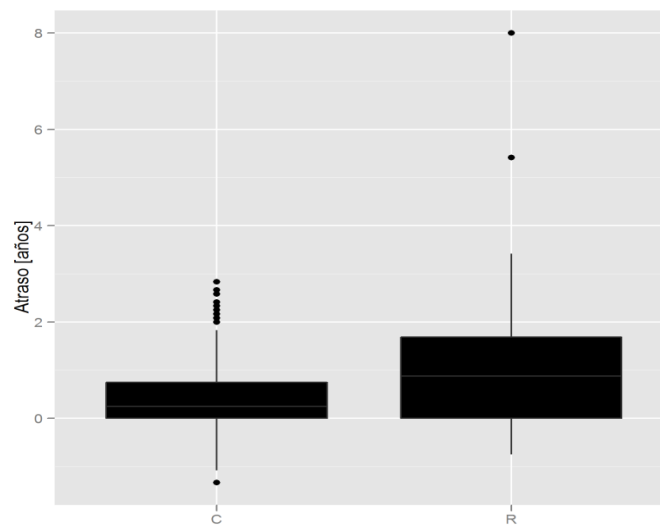


Figura 4.8 Boxplot del atraso con respecto a si aparece como en construcción o recomendado en el plan de obras de la CNE

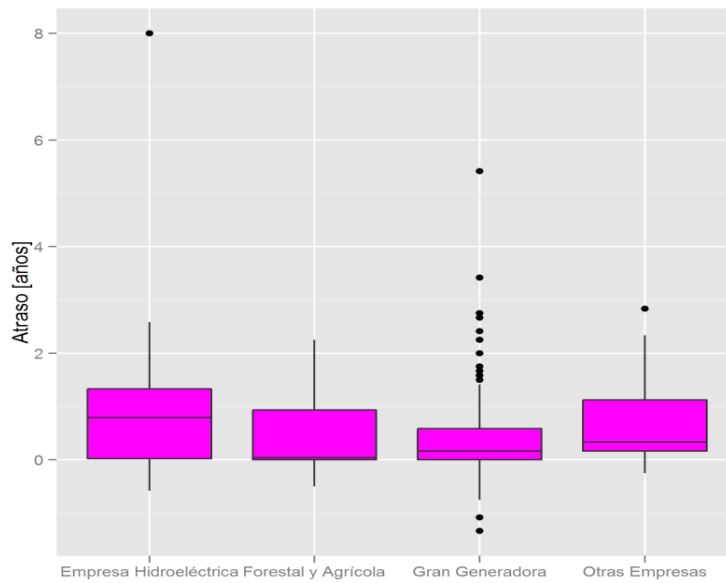


Figura 4.9 Boxplot del atraso con respecto al tipo de empresa dueña de la central

Luego de una primera revisión a los resultados de los diagramas de caja, resulta útil poder tener una idea de las variables o las categorías de las variables que pudieran estar generando distorsiones en los resultados. Un caso de este estilo es lo que ocurre con los atrasos de las grandes empresas, que presentan una gran dispersión y aleatoriedad, ya sea cuando una obra está en construcción o está como recomendada en el plan de obras.

#### 4.2.4. Histogramas

Con el fin de tener una idea de cómo se distribuyen los atrasos para todos los anuncios, es que se genera automáticamente un histograma el cual da cuenta de el porcentaje de cada intervalo de atraso. En la Figura 4.10 se puede observar que los atrasos tienen un comportamiento similar al de una distribución de probabilidades decreciente, como es el caso por ejemplo, de una distribución exponencial.

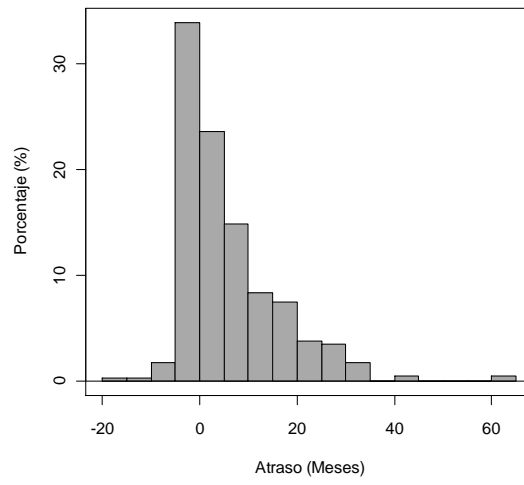


Figura 4.10 Histograma de los atrasos

# Capítulo 5 – Análisis y discusiones

## 5.1. Análisis de resultados

### 5.1.1. Importancia de la selección de atributos

Antes de analizar los resultados de aplicar la metodología a los dos casos de estudios en los que se enfoca este trabajo, se considera importante mencionar la importancia que tiene la selección y pre-selección de los atributos que se utilizarán para construir los modelos. Esto no sólo con el fin de acelerar los cálculos reduciendo la dimensionalidad de la base de datos, sino para que el modelo entregue resultados más precisos y se ajuste a la realidad con un mayor porcentaje de aciertos.

Aunque parezca lógico que no todos los atributos considerados influyan en el comportamiento de la clase definida, resulta importante hacer hincapié en que es perjudicial utilizar toda la información, tanto desde el punto de vista del tiempo de procesamiento (debido a que cada simulación toma alrededor de 5 minutos), como desde el punto de vista de la calidad del modelo que se obtiene. En el mismo sentido es importante identificar las variables que influyen en determinados fenómenos por inspección o con ayuda experta, para poner todos los esfuerzos en definir los niveles que ayuden a cuantificar el impacto que tiene la variable sobre el fenómeno analizado.

### 5.1.2. Atraso de las obras

#### 5.1.2.1. Atributos seleccionados

De un total de 15 atributos utilizados, la selección dejó en 8 los atributos que se consideran relevantes para explicar el fenómeno en estudio, reduciendo a casi la mitad la dimensionalidad de la tabla que se procesará en el software estadístico.

Observando los resultados que se muestran en la Tabla 4.3, tanto la preselección como la selección dejaron fuera los atributos que consideraban parte de la legislación, como es el caso de los atributos ERNC y PMG. Además no se considera el hecho de si el proyecto es nuevo o una

ampliación, pero si se utiliza el hecho de que existiese otra central en el lugar físico donde se instalará el nuevo proyecto. Esto puede deberse a que muchas obras necesarias (camino, puentes, subestaciones, etc.) ya están listas, disminuyendo los riesgos de atraso.

Además, la preselección considera todas las señales de localización geográfica relevantes, eliminando la redundancia después de aplicado el método forward, en donde sólo se consideró la región, que además fue el atributo con menor p-valor.

#### **5.1.2.2. Resultados del modelo**

Analizando los resultados que arrojó el modelo, se tiene que, si se toma en cuenta que la categoría que se predijo corresponde a la que tiene una mayor probabilidad de ocurrencia, entonces el modelo tiene un porcentaje de error un tanto alto (casi un 30%). Sin embargo, si se considera la segunda probabilidad más alta de ocurrencia (que en algunos casos es bastante similar a la probabilidad más alta), entonces el modelo baja su tasa de errores significativamente. Es por esta última razón que resulta importante mostrar las probabilidades de ocurrencia de cada categoría, debido a que si sólo se toma en cuenta la de mayor probabilidad, se pierde información importante que entrega el modelo.

Además, cabe recalcar que existen casos que distorsionan el modelo, debido a que, como ya se mencionó, no siguen ninguna tendencia. Pese a lo anterior el modelo logra rescatar las tendencias de centrales similares en los resultados que arrojó el modelo.

### **5.1.3. Materialización de la construcción**

#### **5.1.3.1. Atributos seleccionados**

Pese a que fue preseleccionada como un atributo que puede explicar la materialización de un proyecto (si se hace o no), el único atributo que no fue seleccionado en la siguiente etapa fue el atributo que segmenta el atraso en la tramitación. Esto quiere decir que, a pesar de que ésta pudiera parecer una variable relevante para construir el modelo de predicción, no lo es a fin de cuentas.

A diferencia del caso anterior sobre los atrasos de los anuncios, para el caso de la materialización de los proyectos, los atributos definidos en la legislación, como lo son los atributos ERNC y PMG, fueron preseleccionados como variables que podrían explicar el

fenómeno en estudio. Pese a que sólo el atributo ERNC es seleccionado finalmente, según estos resultados, para bien o para mal, las definiciones utilizadas en las leyes han sido relevantes en este sentido.

Pese a que no se dan señales de localización, siguen siendo relevantes las variables que hacen referencia al tipo de tecnología que utiliza el proyecto, junto con el tipo de empresa que lleva a cabo la iniciativa.

En el caso de los otros modelos que fueron utilizados para comparación de resultados, se tiene que los atributos tipo de empresa y tecnología siguieron siendo seleccionados. Sin embargo, en ambos casos se agregó el atributo región, que es una variable que de alguna manera entrega señales de localización en la decisión. Además, para el caso del árbol de decisión se desechó el atributo ERNC, mientras que para el caso de la red neuronal sí se consideró, además de considerar el atraso en la tramitación del SEIA.

#### **5.1.3.2. Resultados del modelo**

El modelo utilizado para su construcción, con los atributos recién analizados, tiene un porcentaje de aciertos de un 80%.

En este caso se debe considerar que se utilizaron los proyectos cuya RCA fue emitida antes de enero de 2011, utilizando el hecho de que las resoluciones duran hasta 5 años antes de quedar caducadas e invalidadas.

Con esto se supone que los proyectos que no se han hecho hasta ahora no se construirán finalmente, lo cual podría no ser cierto. Sin embargo se considera que un año es un tiempo prudente para que un proyecto al menos de señales que se construirá en el corto plazo.

Como el modelo rescata tendencias, entonces desecha todas las centrales solares pues aún no se ha construido una. Sin embargo, acepta y rechaza otros tipos de centrales como las hidráulicas, eólicas y térmicas, considerando consistentemente lo que ha ocurrido en el pasado con centrales de características similares.

## **5.2. Discusiones**

### **5.2.1. Base de datos**

Como se aclaró en la metodología, para los casos de estudio que se analizan en este trabajo, los modelos se construyen en base al seguimiento de la información histórica de los proyectos eléctricos, proyectando las tendencias históricas, lo cual castiga en alguna medida a los proyectos que tienen similitudes con otros que en el pasado tuvieron problemas. Considerar también que la base de datos tiene casos patológicos que arruinan los modelos, ya que sus comportamientos no responden a ninguna tendencia, como lo es por ejemplo, el caso de los primeros anuncios de la central Ralco de ENDESA o los muchos anuncios de construcción del proyecto Neltume de la misma empresa. Junto a lo anterior, considerar que debido a la sobre-instalación que tiene el SING, es muy poco lo que se puede decir sobre las tendencias de futuros proyectos (además que la naturaleza del sistema es intrínsecamente para abastecer el sector minero).

No obstante, como la metodología propuesta es general, si se dispusiese de más y mejor información, entonces se podrían construir modelos predictivos que tuvieran en cuenta más variables que las consideradas en los casos de estudio planteados. En especial, sería interesante poder contar con la información estructurada y homologada que asocie los proyectos con el estado de solicitudes de derechos de agua, existencia de otros proyectos en la misma cuenca o aguas arriba, existencia de otros proyectos en la misma área urbana o zona rural, existencia de bosque nativos en las cercanías del proyecto o de las líneas necesarias para la conexión al sistema, número de recursos judiciales y administrativos interpuestos, existencia de zonas saturadas con algún contaminante en las cercanías de los proyectos, y otras trabas descubiertas y descritas en estudios elaborados con esos fines.

También sería de utilidad poder utilizar información acerca de las condiciones del mercado y de las políticas imperantes en el intervalo de tiempo en el que se encuentra un proyecto, ya que éstas influyen en el comportamiento de las variables en estudio. Sería interesante que el modelo pudiese rescatar la información presente en ciertos hitos, como por ejemplo el corte definitivo de gas argentino y la Ley corta I en el 2004 o la promulgación de la Ley de energías renovables el 2008, que son los marcos normativos que han cambiado un poco la dinámica del sector eléctrico, que partió siendo casi exclusivamente de origen hidráulico y se fue diversificando en el tiempo.

Finalmente, la decisión sobre la materialización de un determinado proyecto depende casi exclusivamente del VAN<sup>14</sup> u otro índice financiero que sirva para tener una determinada idea del retorno del proyecto. El VAN es difícil de estimar, ya que considera un gran número de variables temporales y es manejado discreta y confidencialmente por parte de los privados, sin embargo resulta necesario tener una estimación de este número actualizado a la fecha en la cual se hace la evaluación final, ya que puede dar una idea de si un proyecto sigue siendo rentable o bien la prioridad en la que se encuentra dentro de una cartera de proyectos de una determinada empresa.

### **5.2.2. Metodología**

Como ya se explicó, la aplicación de la metodología adolece de limitaciones en la cantidad y calidad de información de que se dispone y de la dificultad que existe de homologar entre las distintas fuentes. A pesar de ello, aunque se contara con una base de datos más robusta, la metodología no considera los efectos que produce la decisión de un agente en las decisiones del resto de los actores del mercado, ya que estos requieren hacer simulaciones más complejas y considerar la teoría de juegos que hay detrás de cada decisión de cada uno de los participantes en el mercado.

Además, no considera variables típicas en los análisis de riesgos, como son las hidrologías, los precios de los combustibles, las proyecciones de demanda y otras variables económicas, que generalmente utilizan los bancos para asignar una prima por riesgo a un determinado proyecto y decidir así, si le conceden un préstamo para financiarlo.

### **5.2.3. Modelos**

Los modelos son sólo aproximaciones que intentan explicar la realidad y establecer patrones de comportamiento en el fenómeno estudiado, pero que sin embargo, contienen errores intrínsecos a la naturaleza de la información. Por esta razón, no hay que considerar que los modelos explican la realidad de manera exacta, sino más bien, hay que utilizarlos como una herramienta de apoyo a la toma de decisiones, que permiten describir un gran subconjunto de la realidad, teniendo siempre presente la existencia de casos excepcionales.

Además comparando los resultados obtenidos con los distintos modelos, se tiene que en el caso de los atrasos en los anuncios, el modelo que obtuvo la menor tasa de errores fue el de

---

<sup>14</sup> VAN: Valor Actual Neto, también conocido como VPN (Valor Presente Neto).



regresión logística; sin embargo, en el caso de la materialización de los proyectos, el modelo con el mejor desempeño lo obtuvo el modelo de redes neuronales. Esto nos deja entrever que no existe mejor modelo que otro, ya que todo depende de las características del problema que se enfrenta, la información disponible y otros factores difíciles de determinar. Por ello es que no se debe descartar ningún modelo a priori y, justamente por ello, plantear una metodología lo más general posible que permita probar la mayor cantidad de modelos y adecuarse a la mayor cantidad de situaciones posibles.

## Capítulo 6 – Conclusiones

Este trabajo pretende ser una primera aproximación para desarrollar un método de clasificación de proyectos eléctricos, de un modo parecido por ejemplo, a como los bancos clasifican a las personas. Actualmente el sistema de monitoreo de mercado tiene una clasificación simple, que define a los proyectos como probables, posibles y en construcción, dependiendo de la fase en la que se encuentre dicho proyecto. Esta clasificación adolece del problema de no considerar otras variables, que dados los resultados de este proyectos, también son relevantes en la probabilidad de ejecución o atraso de un proyecto.

Pese a que las predicciones hechas por el modelo de regresión para el atraso de los proyectos tienen un error que ronda el 30%, mientras que el error del modelo para la materialización de los proyectos es de un 20%, considerando en ambos casos la categoría de la clase correspondiente que tiene la mayor probabilidad, analizando el resto de los casos en los que la predicción estuvo equivocada, la diferencia con las probabilidades de las otras categorías (o al menos con la segunda probabilidad más alta) no estuvo muy alejada. Es por ello que además de dar un juicio sobre la variable en estudio, resulta necesario entregar la probabilidad de que la predicción sea correcta, debido a que existen casos en que el modelo entrega probabilidades parecidas para dos categorías distintas.

Un aspecto relevante, considerando que esta herramienta se pudiera integrar el sistema de monitoreo de mercado que está montado como una plataforma web, es la visualización de los resultados de un modo en que sea útil, visualmente atractivo y claro para los usuarios de la plataforma.

Junto con lo anterior, el desarrollo de herramientas estadísticas visuales puede ser de gran ayuda, por ejemplo, para observar cómo se concentran los datos, los promedios, las medias, datos excepcionales, etc. Así se puede verificar visualmente cuáles son los atributos que podrían explicar el comportamiento de una determinada variable y cuales son los atributos que distorsionan los resultados de los modelos, debido a su gran dispersión y casos excepcionales.

Además, debido a que la información pública disponible recopilada para este trabajo es acotada, si se dispusiese de una base de datos más robusta, se podrían obtener modelos de predicción más sofisticados. También se podrían crear modelos que incorporasen secuencias de tiempo, para así poder incorporar el efecto que tienen ciertos hitos que han ido cambiando tendencias en lo que se refiere al desarrollo del sector.

Con respecto a los modelos, al comparar la tasa de errores que arrojaron en ambos casos de estudios, se tiene que para el caso de los atrasos el mejor modelo fue el de regresión logística; sin embargo, en el caso de la materialización o no de un proyecto, el modelo que obtuvo mejor performance fue la red neuronal de una capa. De aquí se puede concluir que no existe un modelo mejor que otro, ya que la performance depende de múltiples factores como la linealidad del problema, la naturaleza de los datos, etc. Es por ello que no se debe descartar ninguna posibilidad a priori y establecer una metodología lo más general posible que permita ir testeando cómo se ajustan los modelos a los fenómenos que se intentan explicar.

Pese a que puedan ser necesarios cambios en la base de datos, en caso de que la información requerida para extender o mejorar el modelo no se adapte a lo que ya existe, como la estructura está definida, la extensión es directa. Lo mismo ocurre si se quiere probar un nuevo modelo, ya que la metodología propuesta es general, mostrándose en este trabajo la aplicación a un conjunto de datos obtenidos manualmente.

Finalmente, como trabajo futuro se pueden mejorar los modelos considerando información sobre otras variables o en su defecto establecer funciones de probabilidad razonables que puedan explicar la naturaleza de estas variables. Un ejemplo podría ser la “percepción ciudadana de un proyecto”, o “atraso en la construcción de las obras de los clientes que son abastecidos por el proyecto en cuestión”. Si bien la percepción ciudadana es algo difícil de medir y el desarrollo de distribuciones de probabilidad puede corresponder a un trabajo de mucho tiempo, que incluyan encuestas a una población de  $n$  personas elegidas al azar, si se pueden hacer estimaciones analizando la contingencia del momento, ya que esto influye directamente en la judicialización de los proyectos. También se pueden hacer estimaciones sobre costos de desarrollos, rentabilidad aproximada de las empresas gracias al atraso de una obra, etc.

## Capítulo 7 – Bibliografía

- [1]. **Agurto C., Renato, y otros, y otros.** *Comisión Asesora para el Desarrollo Eléctrico.* Ministerio de Energía. Santiago : s.n., 2011. pág. 181.
- [2]. **González., Carlos Mallo.** *Predicción de la demanda eléctrica horaria mediante redes neuronales artificiales.* Departamento de Economía Cuantitativa. Universidad de Oviedo.
- [3]. **Ministerio de Minería.** *Decreto con Fuerza de Ley N° 1.* 1982.
- [4]. **Programa de Gestión y Economía Ambiental (PROGEA).** *Diseño de un Modelo de Proyección de Demanda Energética Global Nacional de Largo Plazo.* Universidad de Chile. Santiago, Chile : s.n., 2008.
- [5]. **Ministerio de Economía, Fomento y Reconstrucción.** *Ley 20.257.* 2008.
- [6]. **Ministerio de Energía.** Objetivo y Funciones. [En línea] [Citado el: 19 de Febrero de 2012.] <http://www.minenergia.cl/ministerio/objetivos-y-funciones.html>.
- [7]. **Comisión Nacional de Energía.** CNE- Comisión Nacional de Energía. [En línea] [Citado el: 20 de Febrero de 2012.] [http://www.cne.cl/cnewww/opencms/01\\_Institucional/Quienes\\_Somos/quienes\\_somos.html](http://www.cne.cl/cnewww/opencms/01_Institucional/Quienes_Somos/quienes_somos.html).
- [8]. **CDEC-SIC.** CDEC-SIC: Quienes Somos. [En línea] [Citado el: 20 de Febrero de 2012.] [https://www.cdec-sic.cl/contenido\\_es.php?categoria\\_id=1&contenido\\_id=000001](https://www.cdec-sic.cl/contenido_es.php?categoria_id=1&contenido_id=000001).
- [9]. **Ministerio del Medio Ambiente.** Quienes Somos - Portal del Ministerio del Medio Ambiente. [En línea] [Citado el: 20 de Febrero de 2012.] <http://www.mma.gob.cl/1257/w3-propertyvalue-15991.html>.
- [10]. **Servicio de Evaluación Ambiental.** ¿Quiénes somos? SEA - Servicio de Evaluación Ambiental. [En línea] 20 de Febrero de 2012. <http://www.sea.gob.cl/contenido/quienes-somos>.
- [11]. **Superintendencia del Medio Ambiente.** Qué es. [En línea] [Citado el: 20 de Febrero de 2012.] [http://www.sma.gob.cl/index.php?option=com\\_content&view=category&layout=blog&id=2&Itemid=20](http://www.sma.gob.cl/index.php?option=com_content&view=category&layout=blog&id=2&Itemid=20).
- [12]. **Yang, Jian.** *A market monitoring system for the Open Electricity.* s.l. : IEEE, 2001.

- [13]. **Gao, C.** *Design of the Electricity Market Monitoring System*. Nnaging, China : IEEE, 2008.
- [14]. **Reid, Tomás.** "Sistema de monitoreo de mercado para la supervisión y toma de decisiones en el sector eléctrico." *Memoria para optar al título de ingeniero civil Electricista*. Departamento de Ingeniería Eléctrica, Universidad de Chile. 2008.
- [15]. **Escobar, Sharo.** "Sistema de monitoreo de mercado para el análisis de riesgo en el segmento de generación eléctrica en sistemas hidrotérmicos" - *Tesis de magíster mención ciencias ingeniería eléctrica*. Departamento de Ingeniería Eléctrica, Universidad de Chile. Santiago, Chile : s.n., 2007.
- [16]. **Mansilla, Felipe.** "Uso de inteligencia de negocios para un sistema de monitoreo de mercado en el sistema eléctrico chileno"-*Memoria para optar al título de Ingeniero Civil de Industrias, Diploma en Ingeniería en Tecnologías de Información*. Escuela de Ingeniería, Pontificia Universidad Católica. Santiago, Chile : s.n., 2009.
- [17]. **Fundación para la Transferencia Tecnológica.** *Análisis de la Información Necesaria para el Diseño de Políticas de Seguridad del Sistema Energético y de la Eficiencia del Sector*. Santiago, Chile : s.n., 2011.
- [18]. **Consultora Estudios de Medio Ambiente y Gestión S.A.** *Identificación de dificultades en la tramitación de permisos de proyectos del sector eléctrico*. Ministerio de Energía. Santiago : s.n., 2010.
- [19]. **Facultad de Derecho y Centro de Energía, Universidad de Chile.** *Análisis de Casos de Recursos Administrativos y Judiciales Relacionados con la Tramitación de Permisos para Proyectos del Sector Eléctrico y sus Efectos en la Inversiones del Sector Energía*. 2011.
- [20]. **Revista Electricidad.** *Electricidad - La revista energética de Chile*. [En línea] [Citado el: 23 de Febrero de 2012.] [www.revistaei.cl](http://www.revistaei.cl).
- [21]. **Cerda, Pablo.** "Análisis de estrategias de inversión en generación en Mercados Competitivos. Aplicación al SIC". *Tesis para optar al grado de Magíster en Ciencias de la Ingeniería*. Escuela de Ingeniería, Pontificia Universidad Católica de Chile. Santiago de Chile : s.n., 2011.
- [22]. **R Development Core Team.** *R: A language and environment for statistical computing*. Viena : R Foundation for Statistical Computing, 2011. URL <http://www.R-project.org/>. ISBN 3-900051-07-0.

[23]. **Pentaho**. Open Source ETL designed to bridge the gap between business and IT. | Kettle Project: Pentaho Data Integration. [En línea] [Citado el: 25 de Febrero de 2012.] <http://kettle.pentaho.com/>.

[24]. **Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten**. *The WEKA Data Mining Software: An Update*. s.l. : SIGKDD Explorations, 2009. Vol. Volume 11.

## Capítulo 8 – Anexos

### 8.1. Anexo 1: Inteligencia de Negocios

También conocido en la literatura internacional como Business Intelligence, esta rama del saber hace referencia al conjunto de estrategias y herramientas enfocadas a la administración y creación de conocimiento a partir del análisis de datos para la toma de decisiones.

Las herramientas consisten en sistemas de información, los cuales son cargados mediante técnicas de “Extraction, Transformation and Loading”<sup>15</sup> (ETL), los que obtienen la información de las fuentes, la depuran y la transforman para poder cargarla en almacenes de datos<sup>16</sup>.

#### 8.1.1. Bases de datos multidimensionales

Los almacenes de datos donde se guarda la información, son un conjunto de bases de datos que se encuentran estructuradas y relacionadas entre sí mediante “*llaves*”<sup>17</sup>. Este conjunto de bases de datos se estructuran de la siguiente forma.

##### 8.1.1.1. Dimensiones

Son bases de datos que contienen los atributos o características de determinados registros de información. También se puede decir que son los catálogos con la información necesaria para una presentación más clara y contextualizada de la información.

Ejemplos de este tipo de tablas pueden ser las que hacen referencia a las zonas, regiones, fechas, combustibles.

##### 8.1.1.2. Hechos

Son las tablas de registros en las que se encuentra la información que se desea mostrar. Ejemplo de estas tablas son los registros de atrasos, el tiempo de tramitación, etc.

---

<sup>15</sup> Extracción, Transformación y Carga

<sup>16</sup> Conocido en la literatura internacional como Data Warehouse

<sup>17</sup> Más conocidas como *key*

### **8.1.2. Justificación de los softwares utilizados**

La Fundación para la Transferencia Tecnológica [17], realiza una comparación de los distintos software de Business Intelligence, Bases de datos y visualización para escoger los adecuados para el sistema de monitoreo de mercado. Entre los principales criterios para la elección de las soluciones se encontraba el atributo de software libre, además del soporte que existe en los foros web sobre posibles problemas a solucionar. Aunque los software no son los que tienen el mejor rendimiento, los requerimientos del sistema no hacen necesaria más infraestructura de software. Los software escogidos son los mismo que se utilizan para hacer los cálculos hechos en este trabajo y son:

- Business Intelligence: Pentaho Data Integration (Kettle)
- Bases de Datos: MySQL
- Visualizador: Jaspersoft para la etapa de pruebas y Microstrategy para la etapa final.

### **8.2. Anexo 2: Knowledge Discovery in Databases (KDD)**

Tal como lo dice su nombre, el proceso de KDD se refiere a la obtención de conocimiento a partir de la información almacenada en las base de datos. Esta etapa puede ser una extensión del punto anterior, en donde se hacía un repaso del almacenamiento de la información en bases de datos multi-relacionales.

Para la obtención de conocimiento se utilizan múltiples tipos de herramientas, como lo son por ejemplo, modelos estadísticos, algoritmos genéticos, lógica difusa, inteligencia computacional, web mining, etc. La herramienta a utilizar depende del problema que se desea resolver.

La metodología estándar de este tipo de procesos es la que se muestra en la Figura 8.1.



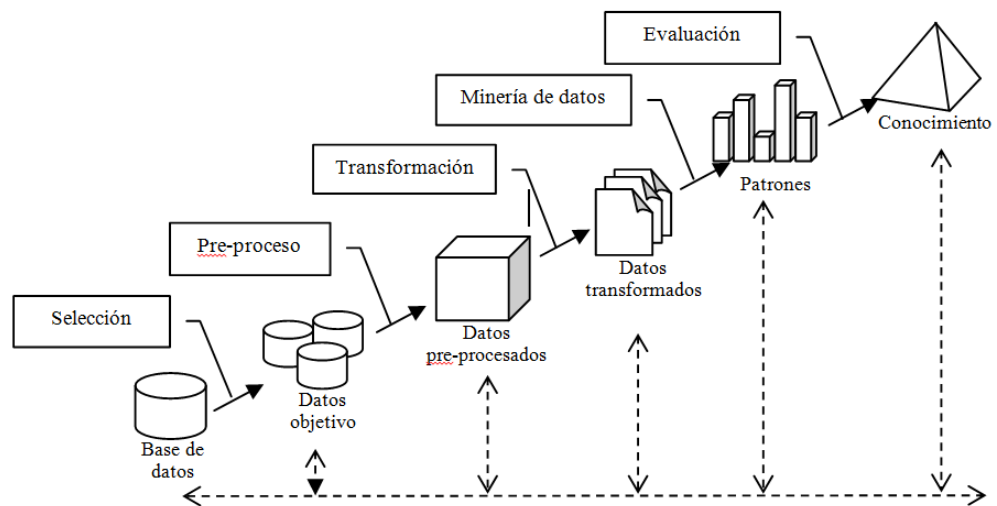


Figura 8.1 Procedimiento general del proceso de KDD

### 8.3. Anexo 3: Atributos cualitativos de los proyectos eléctricos

Tabla 8.1 Regiones y Zonas

Letra Romana	Id	Nombre Administrativo	Zona
XV	15	Arica y Parinacota	Norte Grande
I	1	Tarapacá	Norte Grande
II	2	Antofagasta	Norte Grande
III	3	Atacama	Norte Chico
IV	4	Coquimbo	Norte Chico
V	5	Valparaíso	Centro
XIII	13	Metropolitana	Centro
VI	6	Libertador Bernardo O'Higgins	Centro
VII	7	Maule	Centro-Sur
VIII	8	Biobío	Centro-Sur
IX	9	Araucanía	Sur
XIV	14	Los Ríos	Sur
X	10	Los Lagos	Sur
XI	11	Gnral. Carlos Ibáñez del Campo	Austral
XII	12	Magallanes	Austral

Tabla 8.2 Sistema interconectado

<b>Id</b>	<b>Nombre</b>
0	Autogeneración
1	SIC
2	SING
3	Aysén
4	Magallanes

### 8.4.Diagramas de cajas

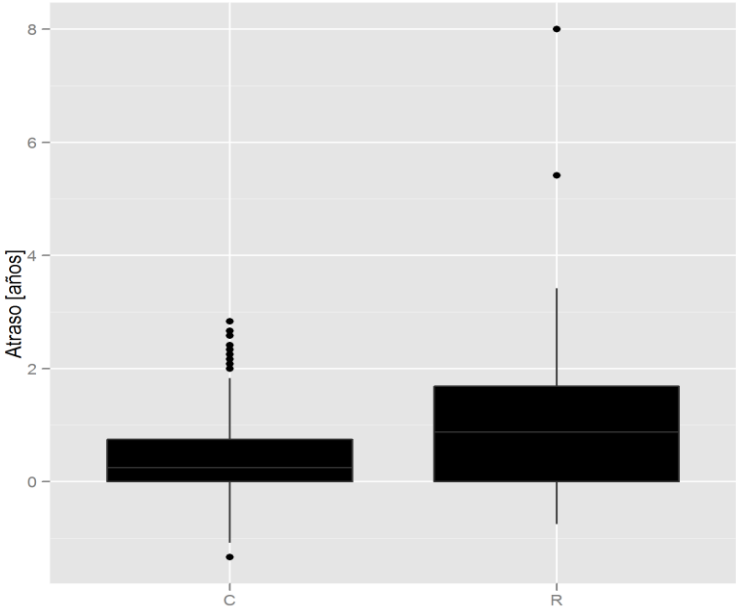


Figura 8.2 Boxplot Atraso vs Tipo de aparición en plan de obras CNE

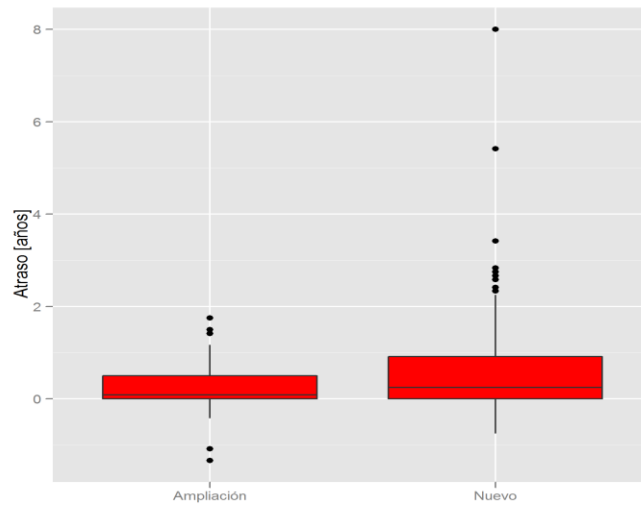


Figura 8.3 Boxplot Atraso vs Clasificación del proyecto

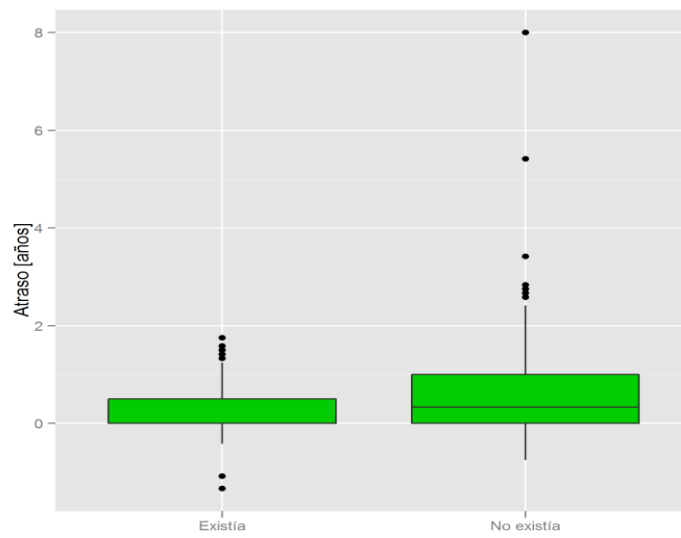


Figura 8.4 Boxplot Atraso vs Existencia previa de proyecto

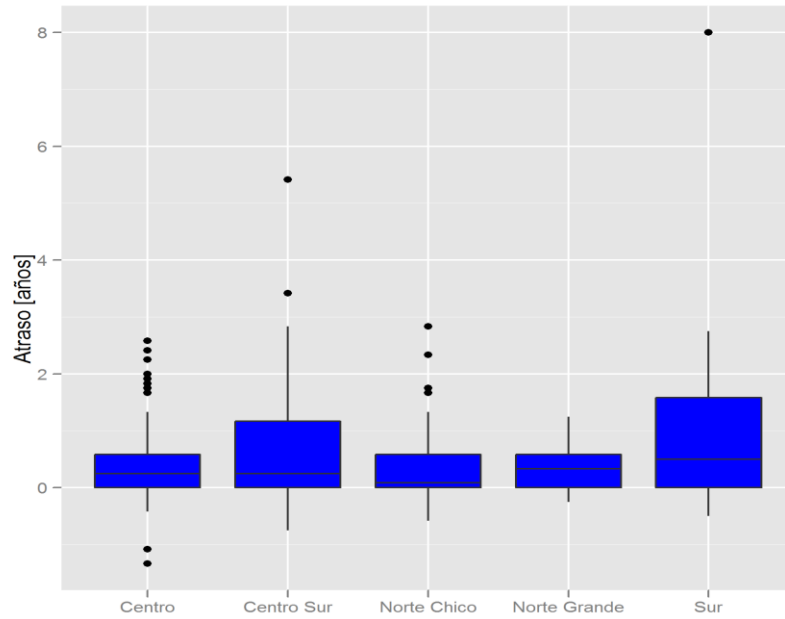


Figura 8.5 Boxplot Atraso vs Zona

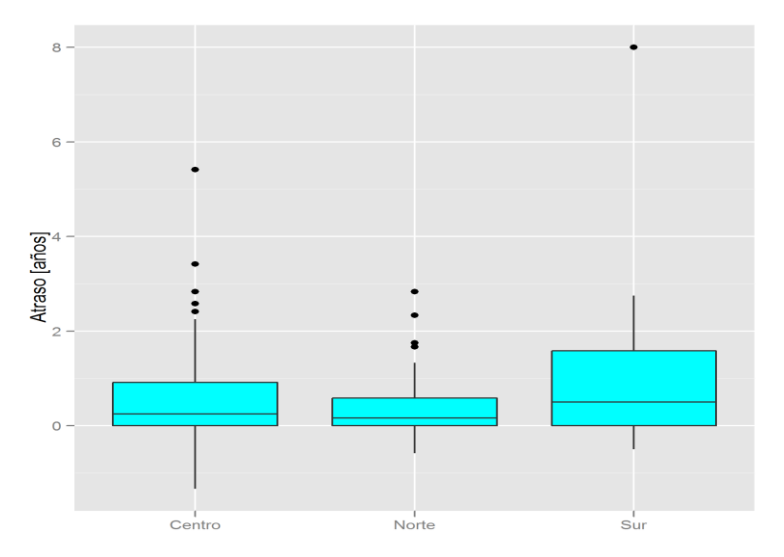


Figura 8.6 Boxplot Atraso vs Macrozona

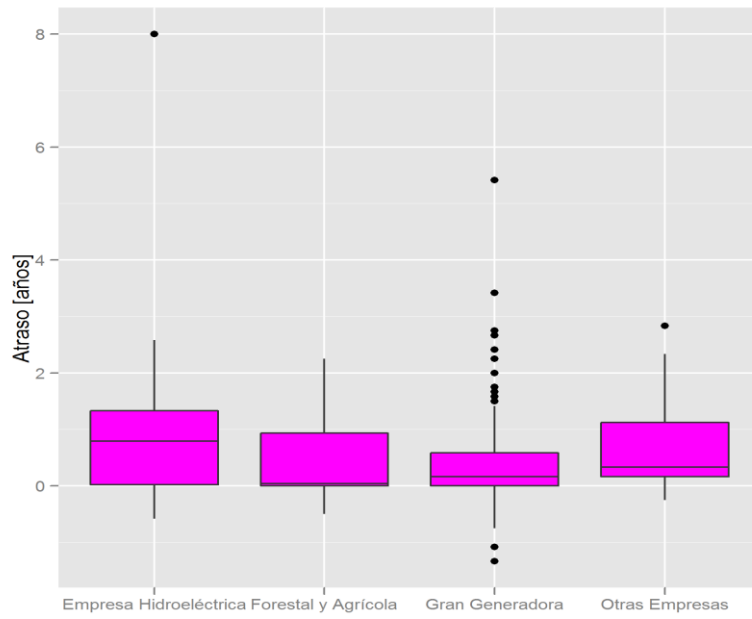


Figura 8.7 Boxplot Atraso vs Tipo de empresa

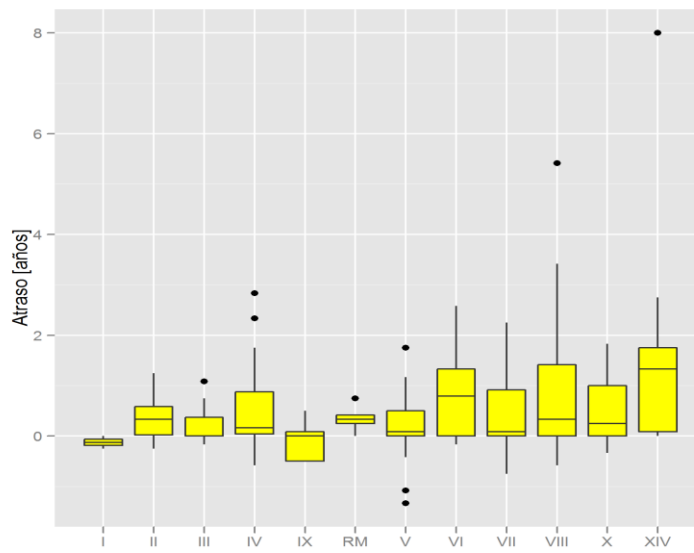


Figura 8.8 Boxplot Atraso vs Región

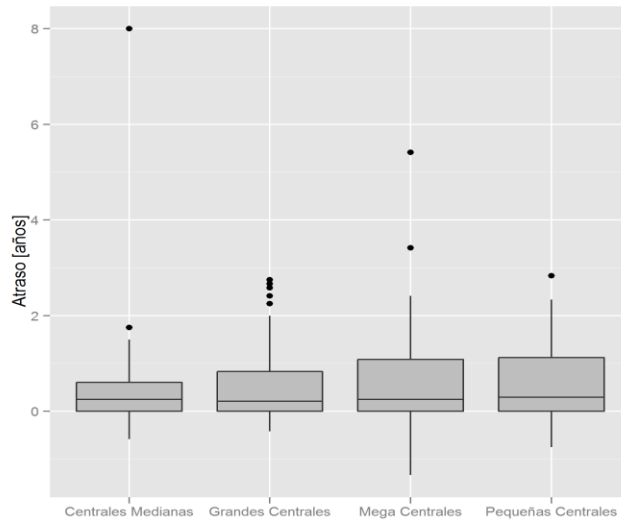


Figura 8.9 Boxplor Atraso vs Potencia

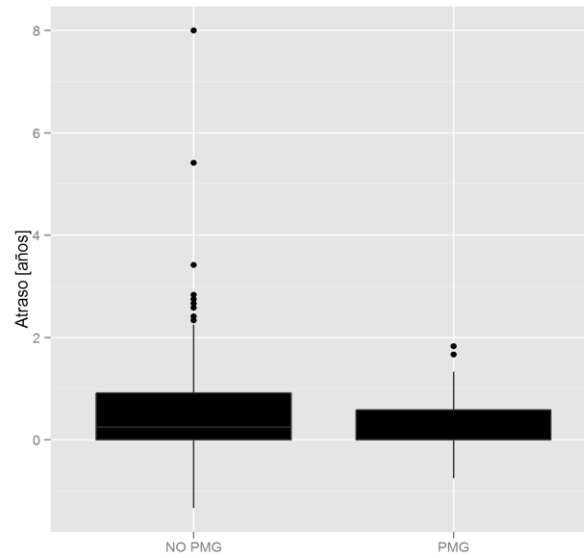


Figura 8.10 Boxplot Atraso vs atributo PMG

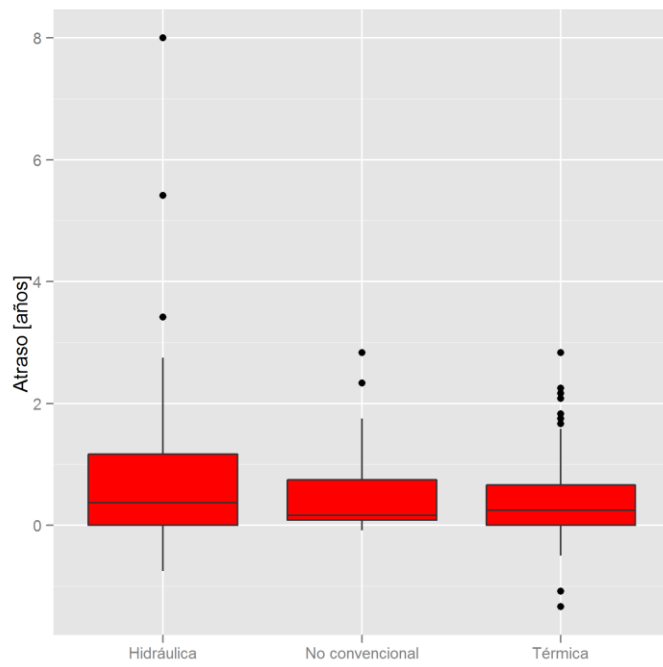


Figura 8.11 Boxplot Atraso vs Fuente

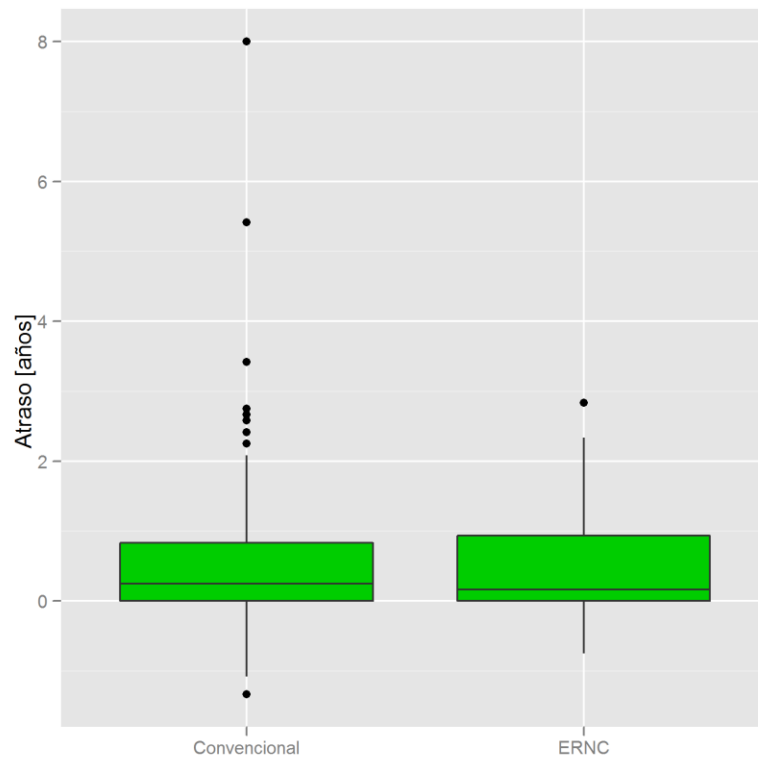


Figura 8.12 Boxplot Atraso vs Atributo ERNC

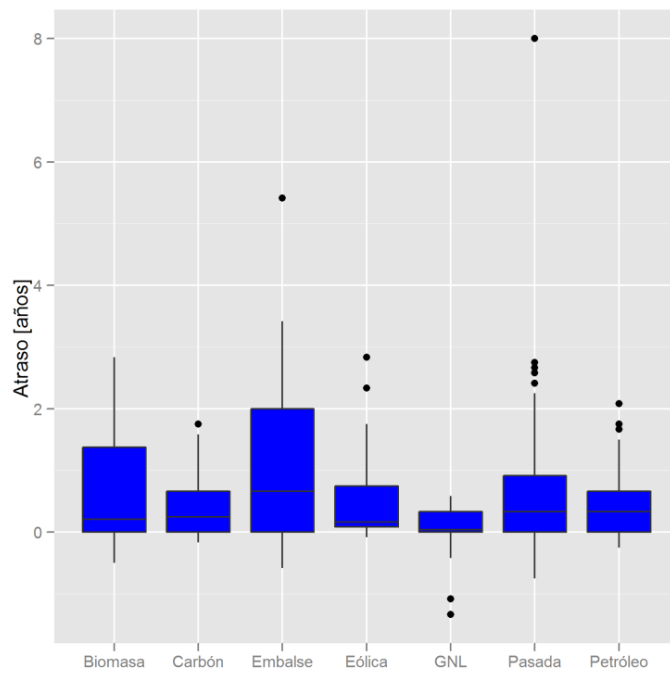


Figura 8.13 Boxplot Atraso vs Combustible principal

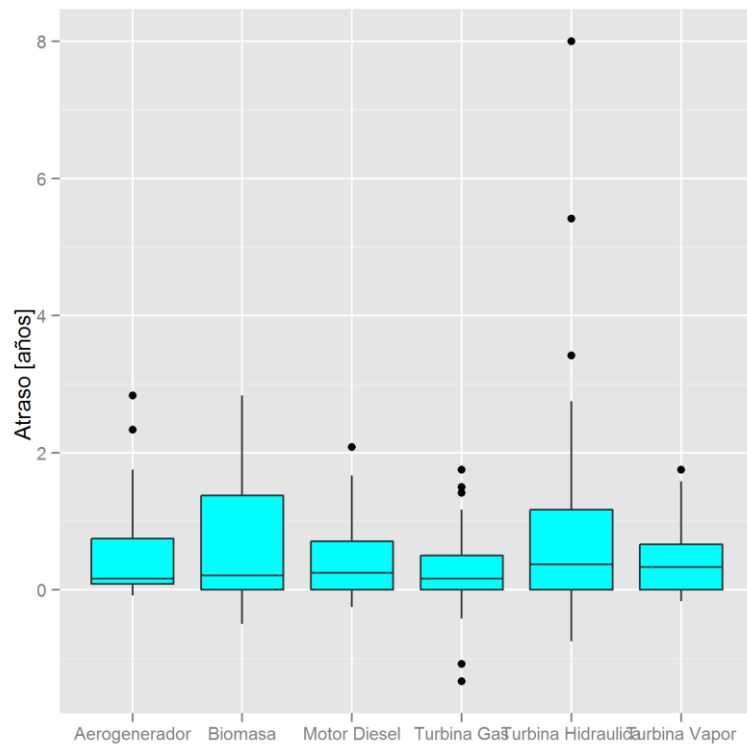


Figura 8.14 Boxplot Atrasos vs Tecnología



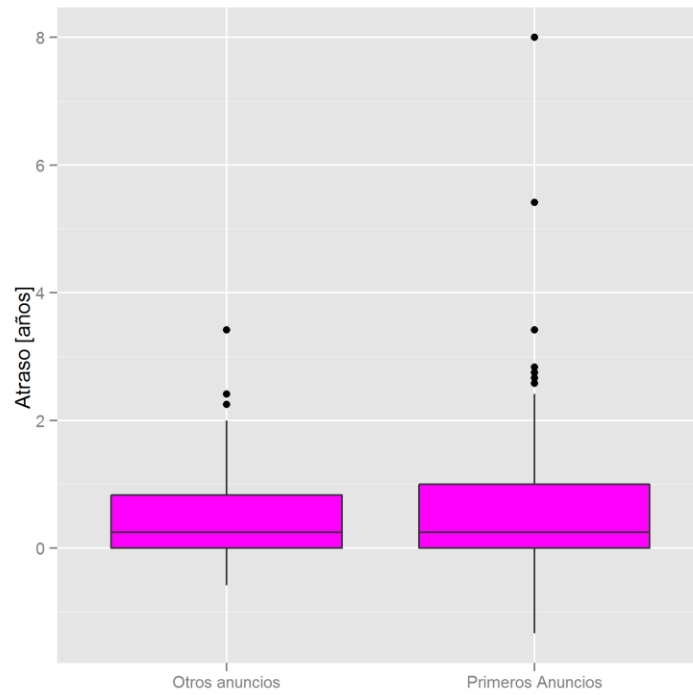


Figura 8.15 Boxplot Atrasos vs Anuncios

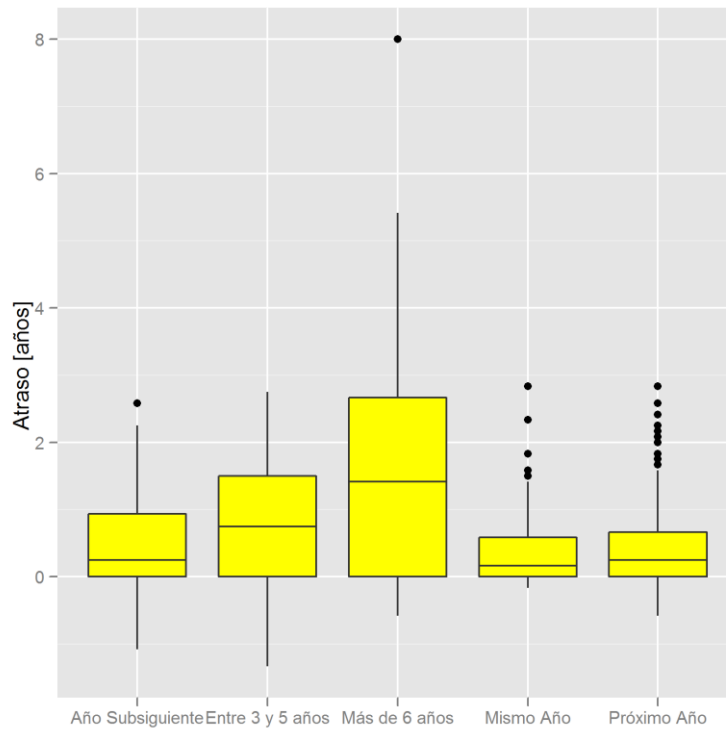


Figura 8.16 Boxplot Atraso vs Proyección