

PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA INFORMÁTICA

RECONOCIMIENTO DE PATRONES Y PRONÓSTICO DE CONSUMO ELÉCTRICO

GABRIEL NATÁN PIZARRO HERRERA

Profesor guía: Rodrigo Alfaro Arancibia

INFORME FINAL DEL PROYECTO
PARA OPTAR AL TÍTULO PROFESIONAL DE
INGENIERO CIVIL EN INFORMÁTICA

MARZO 2017

Dedicatoria

A Dios, por darme las capacidades y paciencia para culminar este trabajo; por estar presente siempre, en cualquier horario o circunstancia.

A mis padres y familia por su apoyo incondicional; por respaldarme afectiva, espiritual y económicamente; por interesarse en conocer mi trabajo, aun cuando les era difícil de comprender. Sin ellos, este proceso habría sido menos llevadero.

A mis amigos, compañeros de universidad, y cuerpo docente de la escuela de ingeniería informática; por ofrecerme el apoyo y las herramientas para crecer y formarme como profesional.

Índice

Resumen	iv
Abstract	iv
Lista de figuras	v
Lista de tablas	vi
1 Introducción.....	1
2 Objetivos.....	2
2.1 Objetivo general	2
2.2 Objetivos específicos	2
3 Problema	3
3.1 Contexto: Actividades del sector eléctrico	3
3.2 Enfoque desde las empresas distribuidoras de energía eléctrica	4
3.2.1 Clientes	4
3.2.2 Tarifas	4
3.2.3 Desafíos	5
3.2.3.1 Hábitos de consumo desconocidos.....	5
3.2.3.2 Incertidumbre del consumo futuro	5
3.3 Avances.....	6
3.3.1 Medición inteligente (Smart Metering).....	6
3.3.1.1 Medidores inteligentes	6
3.3.1.2 Redes de sensores inalámbricas (WSN).....	7
3.3.1.3 Aplicaciones.....	8
4 Marco teórico	9
4.1 Patrones de consumo.....	9
4.1.1 Lineamientos de investigación	9
4.1.1.1 Detección de patrones fraudulentos	9
4.1.1.2 Detección de patrones de consumo considerando condiciones climáticas	9
4.1.1.3 Detección incremental de patrones de consumo.....	10
4.1.1.4 Identificación de artefactos electrónicos a través del análisis de consumo	10
4.1.1.5 Extracción de patrones de comportamiento humano.....	10
4.1.1.6 Detección de ocupación a partir de datos de consumo	10
4.1.2 Técnicas de Clustering	10
4.1.2.1 Algoritmo K-means.....	11
4.1.2.2 Agrupamiento jerárquico.....	12
4.1.2.3 Evaluación de desempeño: Silhouette	12
4.2 Pronóstico de consumo eléctrico.....	13
4.2.1 Lineamientos de investigación	13

4.2.1.1 Pronóstico de consumo eléctrico integrando Redes neuronales, series de tiempo y ANOVA [19]	13
4.2.1.2 Un algoritmo de red neuronal basada en la simulación para la predicción de la energía eléctrica el consumo en Irán [20]	13
4.2.1.3 Revisión de aplicaciones de ANN y SVM para la construcción de pronósticos de consumo eléctrico [21]	14
4.2.2 Modelos de pronóstico	14
4.2.2.1 Perceptrón Multicapa de regresión (MLPR)	14
4.2.2.2 Bosque aleatorio (Random Forest)	15
4.2.2.3 Regresión lineal (Linear Regression)	15
4.2.3 Métricas de evaluación de desempeño	16
4.2.3.1 Error cuadrático medio	16
4.2.3.2 Error absoluto medio	16
4.2.3.3 Error porcentual absoluto medio	16
4.2.3.4 Coeficiente de determinación	17
5 Datos	18
5.1 Bases de datos de consumo eléctrico	18
5.1.1 Estructura de bases de datos de consumo	20
5.2 Base de datos de clientes	21
5.2.1 Estructura base de datos de clientes	21
5.3 Pre-procesado de datos	24
6 Experimentos	25
6.1 Herramientas y librerías utilizadas	25
6.1.1 Spyder 3	25
6.1.2 Pandas	25
6.1.3 NumPy	25
6.1.4 SciPy: fftpack	25
6.1.5 Matplotlib Pyplot (Generación de gráficos)	26
6.1.6 Scikit-learn	26
6.2 Clustering	26
6.2.1 Representación de patrones de consumo	26
6.2.1.1 Patrón de consumo diario	27
6.2.1.2 Patrón de consumo semana (lunes a viernes)	30
6.2.1.3 Patrón de consumo fin de semana (sábado y domingo)	34
6.2.1.4 Patrón de consumo con transformada rápida de Fourier	38
6.2.2 Aplicación	40
6.3 Pronóstico mediante modelos predictivos	41
6.3.1 Análisis de la serie	41
6.3.2 Preparación de los datos	42
6.3.3 Aplicación de modelos	43

7 Resultados	44
7.1 Clustering.....	44
7.1.1 Clusters patrón diario de consumo	44
7.1.1.1 Clusters representativos - Análisis completo de horario	44
7.1.1.2 Cluster representativos - análisis de horario punta	45
7.1.2 Clusters patrón de consumo semana (lunes a viernes)	47
7.1.2.1 Clusters representativos - análisis completo de horario	47
7.1.2.2 Análisis de horario punta.....	49
7.1.3 Clusters patrón de consumo fin de semana (sábado-domingo)	50
7.1.3.1 Análisis completo de horario.....	50
7.1.3.2 Análisis de horario punta.....	52
7.1.4 Clustering representación transformada de Fourier	53
7.1.4.1 Análisis completo de horario.....	53
7.1.4.2 Análisis horario punta	55
7.2 Pronósticos.....	56
7.2.1 Regresión lineal	56
7.2.2 Regresión bayesiana.....	57
7.2.3 Bosque aleatorio.....	58
7.2.4 Perceptrón multicapa	59
8 Conclusiones.....	60
9 Referencias	61
Anexos.....
A: Gráficos y tablas de resumen para análisis estadístico del consumo eléctrico en horario punta	
B: Gráficos y tablas de resultados - Clustering de consumo eléctrico	

Resumen

Una de las mayores problemáticas en el sector de la distribución de energía eléctrica es la dificultad para conocer con detalle el comportamiento del consumo en el tiempo. Por ello, se han incorporado nuevas tecnologías, como los medidores inteligentes, que permiten registrar y monitorear de manera constante las fluctuaciones del consumo eléctrico. A partir de estos datos, las empresas distribuidoras pueden obtener conocimiento relevante para la toma de decisiones.

El presente trabajo de investigación se ha enfocado, por una parte, en descubrir patrones comunes entre clientes, a partir de datos recolectados por medidores inteligentes. Se elaboraron distintas representaciones de consumo, y a partir de éstas se agrupó a clientes similares, aplicando un algoritmo de clustering. Por otra parte, se abordó también el entrenamiento y evaluación de modelos de pronóstico para el consumo eléctrico, utilizando intervalos pequeños de tiempo.

Los resultados obtenidos a partir del clustering, permitieron reconocer diversos segmentos de clientes, los cuales, no siendo representativos en cantidad, aportaban un alto porcentaje en el consumo total. En cuanto a los modelos de pronóstico evaluados, se obtuvo un buen desempeño, superior al 99% en el coeficiente de correlación.

Palabras-claves: consumo eléctrico, medidores inteligentes, patrones de consumo, clustering, modelos predictivos.

Abstract

One of the major problems in the area of electric power distribution is the difficulty in knowing in detail the behavior of electric consumption over time. For this reason, new technologies, such as smart meters, have been incorporated, allowing the constant measure of the fluctuations in electricity consumption. Based on these data, the distribution companies can generate relevant knowledge for decision making.

The focus of the present research is, on the one hand, the discovery of common patterns of customers, based on smart meters data. For this reason, several consumer representations were elaborated, from which similar clients were grouped, applying a clustering algorithm. This work also involved the training and evaluation of forecasting models for electricity consumption, using small time intervals.

The results obtained from the clustering allowed the recognition of different customer segments, which, although not representative in terms of quantity, contributed a high percentage of the total consumption. Regarding the forecast models evaluated, a good performance was obtained, above 99% in the correlation coefficient.

Keywords: electric consumption, Smart meters, consumption patterns, clustering, predictive models

Lista de figuras

Figura 3.1 Actividades del sector eléctrico en Chile	3
Figura 3.2 Comunicación Sistema de Medición Inteligente.....	6
Figura 3.3 Medidores Inteligentes	7
Figura 4.1 Representación gráfica de Clustering	11
Figura 4.2 Clustering con K-Means	11
Figura 4.3 Dendograma de agrupamiento Jerárquico.....	12
Figura 4.4 Arquitectura MLP	14
Figura 5.1 Cantidad de mediciones por fecha, año 2014	19
Figura 5.2 Cantidad de mediciones por fecha, año 2015	20
Figura 5.3 Estructura BDD de consumo.....	20
Figura 5.4 Estructura BDD de clientes.....	21
Figura 6.1 Mapa conceptual de experimentos clustering	27
Figura 6.2 Resumen general - PD FULL.....	28
Figura 6.3 Promedio general – PD FULL	29
Figura 6.4 Resumen por estaciones – PD FULL	29
Figura 6.5 Promedio estaciones – PD FULL.....	30
Figura 6.6 Resumen general – SEM FULL.....	31
Figura 6.7 Promedio general - SEM FULL	32
Figura 6.8 Resumen por estaciones - SEM FULL.....	33
Figura 6.9 Promedio estaciones - SEM FULL	34
Figura 6.10 Resumen general - FDS FULL.....	35
Figura 6.11 Promedio general - FDS FULL.....	36
Figura 6.12 Resumen por estaciones - FDS FULL.....	36
Figura 6.13 Promedio estaciones - FDS FULL	37
Figura 6.14 N° mediciones por fecha, año 2015 - Serie temporal patrón de Fourier.....	38
Figura 6.15 Resumen general - PF FULL	40
Figura 6.16 Serie temporal suma de consumo 2015.....	41
Figura 7.1 Clusters representativos - PD FULL General.....	44
Figura 7.2 Clusters representativos - PD HP General.....	46
Figura 7.3 Clusters representativos - SEM FULL General.....	47
Figura 7.4 Clusters representativos - SEM HP General	49
Figura 7.5 Clusters representativos - FDS FULL GEN.....	50
Figura 7.6 Clusters representativos - FDS HP GEN.....	52
Figura 7.7 Clusters representativos - PF FULL GEN.....	53
Figura 7.8 Clusters representativos - PF HP GEN.....	55
Figura 7.9 Pronóstico Regresión lineal	56
Figura 7.10 Pronóstico Regresión Bayesiana.....	57
Figura 7.11 Pronóstico Bosque Aleatorio	58
Figura 7.12 Pronóstico Perceptrón Multicapa	59

Lista de tablas

Tabla 5.1 Detalles BDD Consumo 2014	18
Tabla 5.2 Detalles BDD Consumo 2015	19
Tabla 5.3 N° de clientes por tipo de tarifa.....	22
Tabla 5.4 N° de clientes por tarifa.....	23
Tabla 5.5 N° de clientes por ubicación	23
Tabla 5.6 N° de clientes por cte. de energía.....	23
Tabla 6.1 Indicadores descriptivos - PD FULL General	28
Tabla 6.2 Indicadores descriptivos - PD FULL Estaciones.....	30
Tabla 6.3 Indicadores descriptivos - SEM FULL General	32
Tabla 6.4 Indicadores descriptivos - SEM FULL Estaciones.....	33
Tabla 6.5 Indicadores descriptivos - FDS FULL General	35
Tabla 6.6 Indicadores descriptivos - FDS FULL Estaciones.....	37
Tabla 6.7 Resultados test Dickey-Fuller - Serie temporal patrón de Fourier	39
Tabla 6.8 Resultados test Dickey Fuller - Serie temporal 2015	42
Tabla 7.1 Indicadores descriptivos, Clusters representativos - PD FULL General	45
Tabla 7.2 Tarifas, Clusters representativos - PD FULL General.....	45
Tabla 7.3 Indicadores descriptivos, Clusters representativos - PD HP General.....	46
Tabla 7.4 Tarifas, Clusters representativos - PD HP General	47
Tabla 7.5 Indicadores descriptivos, Clusters representativos - SEM FULL General	48
Tabla 7.6 Tarifas, Clusters representativos - SEM FULL General.....	48
Tabla 7.7 Indicadores descriptivos, Clusters representativos - SEM HP GEN	49
Tabla 7.8 Indicadores descriptivos, Clusters representativos – SEM HP GEN.....	50
Tabla 7.9 Indicadores descriptivos, Clusters representativos – FDS FULL GEN.....	51
Tabla 7.10 Tarifas, Clusters representativos – FDS FULL GEN	51
Tabla 7.11 Indicadores descriptivos, Clusters representativos – FDS HP GEN	52
Tabla 7.12 Tarifas, Clusters representativos – FDS HP GEN	53
Tabla 7.13 Indicadores descriptivos, Clusters representativos – PF FULL GEN	54
Tabla 7.14 Tarifas, Clusters representativos – PF FULL GEN	54
Tabla 7.15 Indicadores descriptivos, Clusters representativos – PF HP GEN	55
Tabla 7.16 Tarifas, Clusters representativos – PF HP GEN	56
Tabla 7.17 Desempeño Regresión lineal.....	57
Tabla 7.18 Desempeño Regresión Bayesiana	57
Tabla 7.19 Desempeño Bosque aleatorio	58
Tabla 7.20 Desempeño Perceptrón Multicapa	59

1 Introducción

La energía eléctrica es un recurso que ha cobrado vital importancia en los últimos años; El avance de la tecnología y su llegada masiva a las ciudades y hogares ha generado una alta demanda que va en aumento cada día. Desde el alumbrado público, hasta la iluminación y artefactos del hogar, tales como refrigeradores y televisores, necesitan ser abastecidos constantemente de electricidad para mantenerse en funcionamiento.

Esta fuerte demanda, ha llevado a concentrar los esfuerzos de investigación en mejorar la eficiencia de la generación, distribución y consumo de energía eléctrica en hogares e industrias. Con el mismo objetivo, las empresas distribuidoras de energía eléctrica han estado incorporando nuevas tecnologías que les permitan obtener información completa y en tiempo real del consumo de sus clientes. Es así, como se ha adoptado un nuevo concepto llamado “Smart Metering” o medición inteligente, que da paso a nuevos tipos de medidores que interactúan en un sistema interconectado. Esto permitirá realizar análisis sobre los datos, hacer pronósticos certeros y detectar patrones de consumo que serán relevantes en la toma de decisiones.

En esta instancia, la disciplina que ha tenido mayor auge es la minería de datos (Data Mining), ya que, en primer lugar, provee una serie de técnicas que permiten agrupar y caracterizar de manera efectiva patrones de consumo eléctrico; en segundo lugar, el aprendizaje de máquina permite aplicar modelos de pronóstico certeros. Estos análisis no tan solo están orientados a favorecer y mejorar las operaciones de distribución por parte de las empresas, sino también tienen la finalidad de ofrecer al consumidor un análisis de su propio consumo, y así promover un uso consciente que resguarde la eficiencia energética.

En este proyecto de investigación se estudiará el reconocimiento de patrones y la aplicación de modelos predictivos para el consumo eléctrico, a partir de datos provistos por medidores inteligentes. Para lo anterior se abordarán los siguientes tópicos principales:

- Avances y aplicaciones en Medición Inteligente
- Lineamientos de investigación en reconocimiento de patrones y pronóstico de consumo eléctrico
- Técnicas de minería de datos aplicables al reconocimiento de patrones y pronóstico de consumo eléctrico.
- Pre-procesado, representación y clustering de patrones de consumo.
- Series de tiempo y modelos de pronóstico.
- Aplicación y evaluación de modelos de pronóstico para consumo eléctrico.

2 Objetivos

En esta sección se presentan los objetivos del presente trabajo de investigación, cuyos enunciados guiarán el desarrollo del proyecto durante todas sus etapas. Se especifica objetivo general, seguido de los objetivos específicos.

2.1 Objetivo general

Reconocer patrones y crear pronósticos de consumo eléctrico, aplicando técnicas de Minería de Datos (Data Mining) sobre un conjunto de datos reales obtenidos desde medidores inteligentes.

2.2 Objetivos específicos

- Comprender el concepto de Smart Metering y su impacto en el sector eléctrico y el análisis de datos.
- Analizar representaciones de patrones de consumo eléctrico.
- Aplicar técnicas de clustering.
- Aplicar y evaluar desempeño de modelos de pronósticos aplicados al consumo eléctrico.
- Implementar la solución en algún lenguaje de alto nivel (Python, JULIA)

3 Problema

En el sector de la energía eléctrica existen distintos tipos de actividades, desde las cuales puede abordarse el problema del consumo de manera distinta. En las secciones siguientes, se hará un breve repaso de las actividades del sector eléctrico; luego, se definirá la problemática desde la perspectiva o enfoque del presente proyecto. Finalmente, se mencionarán avances y aplicaciones actuales.

3.1 Contexto: Actividades del sector eléctrico

En el ámbito de la energía eléctrica, pueden distinguirse tres tipos de actividades claves, las cuales son realizadas por distintos tipos de empresas que se dedican exclusivamente a ellas. Estas actividades son generación, transmisión y distribución.

En primer lugar, se tiene la actividad de generación, en donde se concentra todo el espectro de empresas generadoras de energía eléctrica. En segundo lugar, está la actividad de transmisión, en donde empresas se dedican únicamente a transportar la energía generada por las centrales eléctricas a distintos puntos y estaciones. En tercer lugar, se encuentra la actividad de distribución, la cual está directamente ligada a los usuarios finales, pues se encarga de proveer de energía eléctrica a los consumidores o clientes.

En la Figura 3.1 se muestra un esquema que explica el funcionamiento del sistema eléctrico en Chile.

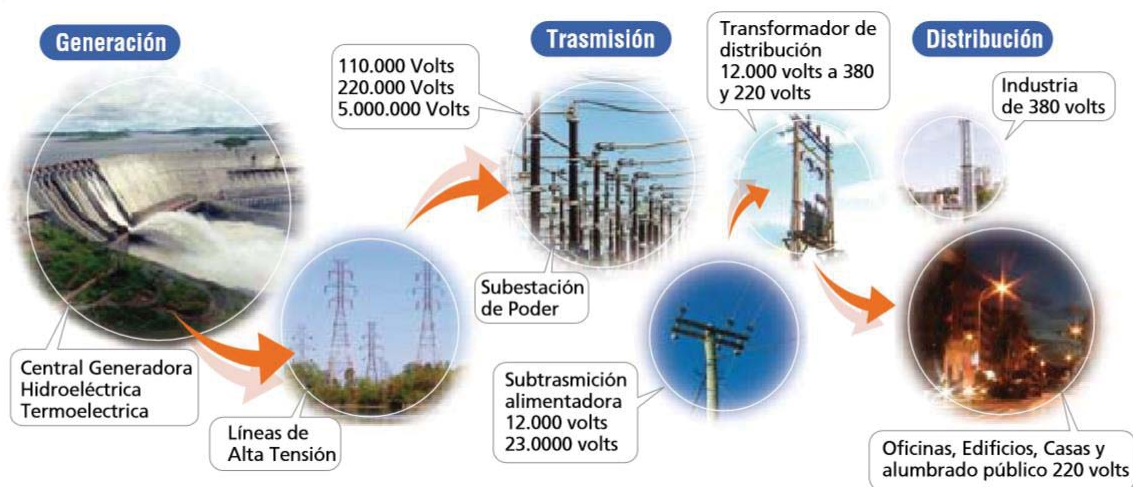


Figura 3.1 Actividades del sector eléctrico en Chile

La problemática abordada en este trabajo podría asumir distintos enfoques por cada una de las actividades del sector eléctrico mencionadas previamente. Por lo mismo, y con el fin de trazar una dirección concreta en el desarrollo de la investigación, se ha determinado orientar el problema desde la perspectiva de las empresas distribuidoras de energía eléctrica.

3.2 Enfoque desde las empresas distribuidoras de energía eléctrica

Las empresas distribuidoras en Chile cubren zonas geográficas específicas, que son licitadas y reguladas por la Superintendencia de Electricidad y Combustibles. Actualmente, existen 31 empresas de este tipo, las cuales, bajo un régimen de concesión de servicio público, prestan el servicio de distribución a consumidores finales por medio de equipos, líneas y subestaciones. Entre las empresas más destacadas en éste sector se encuentran Chilectra, Grupo Emel y Chilquinta [1].

Chilquinta tiene un rango de servicio de 11.496 km² en la región de Valparaíso, y cuenta con 535.707 clientes atendidos al año 2013. Chilectra por su parte, abarca 2.065 km² en la región Metropolitana, sumando un número de 1.693.948 clientes (2013). Finalmente, el grupo Emel cuenta con 22.123 km de líneas de distribución, comprendidas entre la I y VIII región.

3.2.1 Clientes

Los consumidores de energía eléctrica en Chile pueden ser clasificados, según la Comisión Nacional de Energía (CNE), en dos tipos: regulados y libres. Los clientes de empresas distribuidoras de energía eléctrica son del tipo regulados, es decir, tienen una potencia conectada inferior o igual a 2.000 kilowatts (kW) y además están sujetos a tarifas fijadas por el Estado [2]. Dentro de este grupo se encuentran mayormente consumidores de carácter residencial, comercial e industrial. Por otra parte, los clientes libres son aquellos que poseen una potencia conectada igual o superior a 2.000 (kW), y tienen plena libertad para negociar sus tarifas directamente con las empresas generadoras.

3.2.2 Tarifas

Como se mencionó antes, los clientes regulados están sujetos a tarifas que son fijadas por el Estado. Cada cliente tiene la libertad de escoger la tarifa que desea contratar, por lo cual las empresas concesionarias de distribución eléctrica están obligadas a aceptar la opción tarifaria de cada cliente. Tales opciones se han estructurado de acuerdo con diversas formas para el consumo (sólo energía; potencia máxima leída o contratada; y potencia leída o contratada horariamente), bajo dos categorías de clientes: en alta tensión (AT) y en baja tensión (BT).

La inclusión en una u otra categoría depende de si el usuario está conectado con su empalme a líneas de voltaje superiores o inferiores a 400 volts. Así, las opciones tarifarias para los clientes en baja tensión son [3]:

- BT1:** Medición de energía cuya potencia conectada sea inferior a 10 kW o la demanda sea limitada a 10 kW (residencial)
- BT2:** Medición de energía y contratación de potencia (comercial y alumbrado público)
- BT3:** Medición de energía y medición de demanda máxima
- BT4:** Medición de energía y alguna de las siguientes modalidades

- BT4.1:** Contratación de demanda máxima de potencia en horas de punta y de la demanda máxima de potencia
- BT4.2:** Medición de demanda máxima de potencia en horas de punta y contratación de la demanda máxima de potencia
- BT4.3:** Medición de demanda máxima de potencia en horas de punta y de la demanda máxima de potencia suministrada

Cada una de estas tarifas tiene su equivalente paralelo para clientes con modalidad de alta tensión.

3.2.3 Desafíos

Como se vio en 3.2.1, cada empresa distribuidora posee una gran cantidad de clientes, con distintos perfiles, a los cuales debe entregar su servicio de manera satisfactoria. Esta satisfacción dependerá de diversos factores, entre los cuales se destacan la fijación conveniente de tarifas, información detallada del consumo, prevención de caídas de servicio, entre otros. Además de esto, la empresa debe preocuparse de mejorar continuamente sus procesos internos de distribución y monitoreo de consumo, para reducir al mínimo la pérdida de eficiencia energética.

Para afrontar de mejor manera estos desafíos, se presentan dos grandes dificultades. Primero, existe un desconocimiento de los hábitos de consumo de los clientes; y segundo, las empresas no tienen mayor certeza sobre el comportamiento del consumo a futuro. Estas dificultades se abordan más ampliamente a continuación.

3.2.3.1 Hábitos de consumo desconocidos

Las empresas de distribución poseen un conocimiento muy escaso sobre los hábitos de consumo de sus clientes, debido a las limitaciones de su sistema de medición. Actualmente, la gran mayoría de los equipos instalados para medir el consumo eléctrico son de funcionamiento mecánico, y actúan como un contador que va acumulando el consumo en KWh del cliente. Para que este registro llegue a la empresa, es necesario anotar presencialmente y de forma manual el consumo, proceso que se realiza una vez al mes.

En vista de lo anterior, la empresa sólo posee un dato mensual como indicador del consumo del cliente, lo que es insuficiente para determinar sus hábitos de consumo. Como consecuencia de esto, se hace necesario la incorporación de infraestructura y equipamiento tecnológico que permita a las empresas reconocer patrones de consumo en sus clientes, y de esta manera mejorar sustancialmente sus operaciones comerciales.

3.2.3.2 Incertidumbre del consumo futuro

Como se vio anteriormente, las empresas no cuentan con información detallada del comportamiento de sus clientes. Esto también obstaculiza el pronóstico del consumo eléctrico; ya que, para aplicar un modelo que pronostique correctamente el comportamiento de los

clientes, es necesario contar con una base de datos rica en información, a partir de la cual se puedan obtener resultados más precisos.

Como este no es el caso, la falta de pronósticos confiables deriva en un grado de incertidumbre que dificulta la toma de decisiones, y afecta directamente en la planificación y costos proyectados por la empresa.

3.3 Avances

En pos de combatir las problemáticas antes descritas, se han logrado grandes avances que permiten acercarse hacia el reconocimiento efectivo de patrones de consumo eléctrico y un correcto pronóstico de éste. Dentro de los avances más significativos se encuentra la medición inteligente (Smart Metering), la cual se describe a continuación.

3.3.1 Medición inteligente (Smart Metering)

Es una solución compuesta por medidores inteligentes o sensores, infraestructura de telecomunicaciones y sistemas centrales que permite una gestión remota y automática de la red, así como un flujo bidireccional de información y energía, permitiendo optimizar el funcionamiento de la red [4]. Este sistema de medición permite tener un control en línea del consumo eléctrico y todos los eventos asociados a éste. En la Figura 3.2 se muestra el proceso de comunicación llevado a cabo en un sistema de medición inteligente.



Figura 3.2 Comunicación Sistema de Medición Inteligente

3.3.1.1 Medidores inteligentes

Los medidores inteligentes añaden nuevas funcionalidades a los medidores electromecánicos [5]. Este tipo de aparato es capaz de recoger información diversa del consumo del cliente, tal como la energía activa y reactiva, potencia activa y reactiva, voltaje y corriente eficaz y potencia media cada 15 minutos [6]. Una de sus mayores ventajas, es que puede transmitir de manera automática la información del consumo hacia las centrales de las empresas.



Figura 3.3 Medidores Inteligentes

Como principales características se tiene [7]:

- Integrar comunicación bidireccional
- Entregar señales en caso de ser intervenido o si se trató de intervenir Ser configurado para medir distintas tarifas al mismo tiempo
- Entregar información de la calidad de energía,
- Realizar teledatada, es decir, ser leídos a distancia con distintas opciones de comunicación, y muchas veces, estos pueden interactuar con el usuario.
- Integrarse con otros servicios como gas y agua.
- Integrarse a aplicaciones que permitan al cliente observar y administrar la información a través de otros dispositivos inteligentes.
- Permiten la desconexión y reconexión remota de clientes.
- Permiten la limitación remota de la potencia de suministro.
- Permiten la detección de robo de energía.
- Permiten la optimización de redes y disminución de pérdidas técnicas.
- Consumidores disponen de mayor información sobre su consumo.
- Consumidores pueden responder ante estímulos de precio u otros incentivos.
- Permiten la incorporación de micro-generación o generación distribuida

3.3.1.2 Redes de sensores inalámbricas (WSN)

Una red de sensores inalámbrica (Wireless Sensor Network -WSN) es una red Ad Hoc compuesta por un conjunto de dispositivos electrónicos provistos de sensores y transmisores que intercambian información pertinente entre ellos, sin necesidad de un cableado dedicado [8]. Las redes de sensores también pueden ser incorporadas al sistema de medición inteligente, como alternativa a la implementación de medidores inteligentes. Sin embargo, no es una tecnología muy explotada en el área de la medición de variables de energía eléctrica.

3.3.1.3 Aplicaciones

La incorporación de medición inteligente no es algo común en Chile. Sólo en los últimos años algunas empresas de distribución eléctrica han iniciado planes de prueba en algunas ciudades del país. No obstante, esta tecnología es común en otros países, especialmente en Europa, donde ya se han incorporado con éxito hace algunos años. Dado los efectos positivos que ha generado la medición inteligente en las ciudades europeas, esta tecnología no tardará mucho en establecerse completamente en países de América Latina.

En Italia, la empresa Enel ha instalado una cantidad de cuarenta millones de medidores inteligentes, los cuales permiten, entre otras cosas, realizar operaciones relacionadas con el suministro como cambios de comercializador, cambios en las características del contrato, altas o bajas de los clientes de manera remota y en menos de 24 horas en la mayoría de los casos.

En España, la empresa Endesa ya ha alcanzado el número de siete millones de medidores inteligentes instalados en viviendas. Se espera que para el 2017 se complete la instalación masiva de medidores, que en total dan un número de 11,6 millones de dispositivos.

4 Marco teórico

En este capítulo se hará una revisión del estado del arte relativo al reconocimiento de patrones y pronóstico de consumo eléctrico. Se comenzará haciendo un repaso de las investigaciones existentes al respecto, indicando su enfoque y el proceso general que se utiliza para obtener resultados. Luego se verán las técnicas más comunes que se aplican para tales fines.

4.1 Patrones de consumo

La investigación sobre los datos de consumo eléctrico ha sido un área de estudio muy concurrente en el último tiempo. La eficiencia energética es un tema que ha tomado gran relevancia, y el reconocimiento de patrones de consumo apunta directamente a promover un consumo y distribución eficientes de la electricidad. Por lo tanto, se han desarrollado múltiples trabajos de investigación, enfocados especialmente en el estudio del comportamiento del consumo residencial.

4.1.1 Lineamientos de investigación

A continuación, se exponen los principales lineamientos de las investigaciones consideradas como referencia para el descubrimiento de patrones de consumo eléctrico.

4.1.1.1 Detección de patrones fraudulentos

Una de las constantes preocupaciones de las empresas distribuidoras de energía eléctrica es el robo o utilización fraudulenta de los servicios que proveen. Las pérdidas generadas por este tipo de eventos no son menores. Por lo tanto, muchas investigaciones apuntan a la detección de patrones fraudulentos a partir del análisis del consumo otorgado por medidores inteligentes.

En [9] los autores aseveran que en la actualidad las aplicaciones de minería de datos en el análisis de consumo eléctrico no se enfocan lo suficientemente en los subgrupos cuyos patrones varían significativamente de los patrones presentes en el conjunto de datos. Para enfrentar esta insuficiencia, los autores proponen la utilización del algoritmo “subgroup discovery” para detectar patrones inusuales, y a partir de ellos generar sus modelos predictivo y descriptivo.

4.1.1.2 Detección de patrones de consumo considerando condiciones climáticas

Diversos autores consideran de gran relevancia complementar los datos de consumo entregados por los medidores inteligentes, con datos externos tal como las condiciones climáticas o la temperatura. En [10] los autores afirman que hasta un 80% del consumo eléctrico en los hogares puede ser explicado analizando estas variables en conjunto.

En [11] los autores concluyeron mediante la aplicación de regresión, que existe una gran relación entre la temperatura externa y el consumo, el cual tiende a ser mayor en periodos de clima frío.

4.1.1.3 Detección incremental de patrones de consumo

Una de las principales consideraciones que se deben tener en cuenta al momento de analizar la información recogida por los sistemas de medición inteligente, es el gran volumen de datos que se va generando minuto tras minuto. La información tiene un comportamiento dinámico, por lo cual, un análisis dinámico podría tener un mayor grado de exactitud.

En [12] el autor propone un marco de trabajo llamado “Incremental Summarization and Pattern Characterization” (Resumen incremental y caracterización de patrones), para el análisis incremental de patrones de consumo eléctrico. Este marco de trabajo tiene como principal objetivo acumular nuevo conocimiento basándose en patrones ya detectados en el pasado. El autor argumenta que las técnicas convencionales no son aplicables en su totalidad al gran volumen de datos que se va renovando constantemente.

4.1.1.4 Identificación de artefactos electrónicos a través del análisis de consumo

Distintas investigaciones se enfocan en la identificación de la utilización de artefactos eléctricos, a través del análisis de patrones de consumo eléctrico. Actualmente, esta información sólo puede ser obtenida a través de sensores. Sin embargo, en [13] los autores realizaron un estudio en donde reconocieron patrones de uso y la cantidad de energía usada cuando un artefacto está encendido. Demostraron que algunos artefactos pueden detectarse con un alto grado de certeza; no obstante, otros fueron difíciles de identificar con los datos estudiados.

4.1.1.5 Extracción de patrones de comportamiento humano

En [14] los investigadores muestran un método por el cual pueden determinar la actividad que una persona está realizando, sólo a través de los datos de consumo eléctrico de los artefactos con una certeza del 98%.

4.1.1.6 Detección de ocupación a partir de datos de consumo

En [15] los autores concluyen que se puede detectar eficazmente la ocupación de una casa usando métodos comunes de clasificación, aplicados sobre datos de consumo obtenidos desde medidores digitales de electricidad.

4.1.2 Técnicas de Clustering

El Clustering es una de las técnicas de Data Mining más utilizadas en los estudios de reconocimiento de patrones. Los algoritmos de Clustering son capaces de dividir conjuntos de datos en grupos de objetos con características similares. Esto es muy útil, especialmente en la detección de patrones. A pesar que al hacer clustering se pierden algunos detalles de la información, éste simplifica la visualización y el análisis posterior [16].

En la Fig. 4.1 se puede observar una representación gráfica de la aplicación de algún algoritmo de clustering sobre un conjunto de datos. Se ve claramente la clasificación de los datos en 5 clusters representados por colores: verde, azul, rojo, celeste, rosado.

4.1.2.2 Agrupamiento jerárquico

Este tipo de clustering busca construir una jerarquía de grupos, y, a diferencia del K-means, no se fija el número K. El criterio utilizado por esta técnica puede interpretarse como: “La similitud entre dos objetos viene dada por la “altura” del nodo común más cercano” [17].

Puede adoptar dos estrategias: Aglomerativas y divisivas. La estrategia aglomerativa consiste en un acercamiento ascendente, en donde se parte con cada caso como cluster individual, pero luego se van combinando los pares de clusters más cercanos hasta que sólo quede uno. La estrategia divisiva, al contrario, comienza con un único cluster que engloba todos los casos del conjunto de datos; luego, en cada paso se debe dividir un cluster hasta que cada uno contenga un único caso.

En la Fig. 4.3 se puede observar la representación en un dendograma del proceso de agrupamiento jerárquico sobre un conjunto de datos.

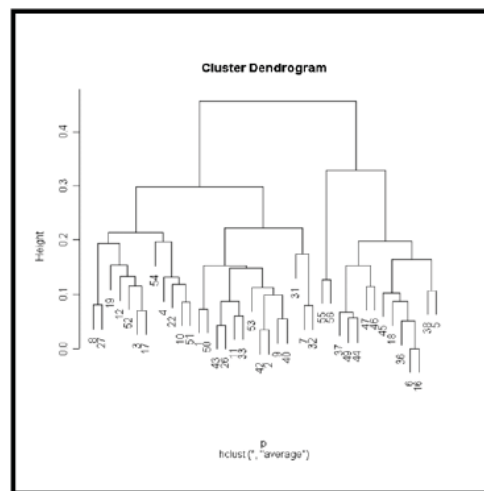


Figura 4.3 Dendograma de agrupamiento Jerárquico

4.1.2.3 Evaluación de desempeño: Silhouette

La métrica de evaluación Silhouette permite medir el desempeño del clustering desde una perspectiva interna. Cuando no se sabe o no se tiene las etiquetas o la agrupación esperada, no se puede medir la precisión del clustering. Sin embargo, con la implementación de Silhouette se obtiene un coeficiente interno que permite identificar la precisión del modelo de clustering. Este coeficiente se define para cada muestra y se compone de dos indicadores [18]:

- a) La distancia media entre una muestra y otros puntos de su misma clase
- b) La distancia media entre una muestra y todos los puntos de la clase más cercana.

El coeficiente de Silhouette para una muestra individual está dado por:

$$s = \frac{b - a}{\max(a, b)} \quad (4.1)$$

4.2 Pronóstico de consumo eléctrico

Para las empresas distribuidoras de energía eléctrica, es muy importante hacer proyecciones del consumo total en periodos futuros. Estas proyecciones les permiten tener una mejor base para la toma de decisiones, y así disminuir el riesgo que implica la incertidumbre. Dependiendo de la calidad de los datos, es posible realizar pronósticos con muy alta precisión, para lo cual, existen diversos modelos y también investigaciones que se dedican a aplicar y evaluar dichos modelos.

En las siguientes secciones, se mostrarán algunas investigaciones sobre pronóstico de consumo eléctrico y aplicación de modelos; también se hará una descripción de los modelos más utilizados y las métricas que se emplean para evaluar su desempeño.

4.2.1 Lineamientos de investigación

En este apartado se muestran algunas investigaciones con respecto al pronóstico de consumo eléctrico. Se hará una breve descripción de cada una de ellas, mostrando principalmente los modelos utilizados.

4.2.1.1 Pronóstico de consumo eléctrico integrando Redes neuronales, series de tiempo y ANOVA [19]

En esta investigación, los autores plantean la dificultad de modelar el problema usando métodos convencionales, debido a los constantes cambios del comportamiento del consumo eléctrico entre meses y estaciones. Para ello, se propone un modelo de pronóstico basado en redes neuronales (ANN) aplicadas a series de tiempo. Se argumenta la superioridad de las redes neuronales con respecto a los métodos convencionales de regresión mediante el análisis de varianza ANOVA.

4.2.1.2 Un algoritmo de red neuronal basada en la simulación para la predicción de la energía eléctrica el consumo en Irán [20]

Este trabajo de investigación es similar al anterior, pero a diferencia de este, incorpora simulación computacional para generar variables aleatorias de consumo eléctrico mensual, usando procedimientos estocásticos. Esto tiene por finalidad prever los efectos de la distribución probabilística del consumo eléctrico mensual. Además, el algoritmo se encarga de comparar ANN con otros modelos de series de tiempo, escogiendo al mejor mediante el análisis de ANOVA, MAPE y Duncan method (DMRT).

Finalmente, la investigación concluye que el algoritmo es capaz de identificar al método convencional de series de tiempo como el mejor modelo para el pronóstico del consumo

eléctrico, en desmedro de estudios anteriores que señalaban que ANN siempre arrojaba resultados más precisos.

4.2.1.3 Revisión de aplicaciones de ANN y SVM para la construcción de pronósticos de consumo eléctrico [21]

Este trabajo hace una revisión del uso de métodos de inteligencia artificial, como redes neuronales (ANN) y máquinas de soporte vectorial (SVM), para el pronóstico de consumo eléctrico. Además, aborda el uso de modelos de pronósticos híbridos, con el fin de obtener una mayor precisión. Acerca de estos últimos, se estudia el uso de un modelo híbrido llamado GLSSVM, que combina el método de agrupación para el manejo de datos (GMDH) con el modelo de mínimos cuadrados de máquinas de soporte vectorial (LSSVM).

4.2.2 Modelos de pronóstico

En esta sección se hará una breve descripción de algunos modelos de regresión, que son comúnmente utilizados para el pronóstico de series de tiempo, y, por lo tanto, también aplicables al problema del consumo eléctrico. Los modelos a revisar son: Red neuronal Perceptron multicapa, Bosque aleatorio y Regresión Lineal.

4.2.2.1 Perceptrón Multicapa de regresión (MLPR)

El perceptrón multicapa es uno de los tipos de red neuronal más comunes. Su funcionamiento es unidireccional, y está basado en una red más simple denominada perceptrón; pero a diferencia de ésta, es capaz de resolver problemas que no son linealmente separables. Ésta capacidad se debe a que implementa múltiples capas, como se aprecia en el esquema de su arquitectura, en la figura 4.4 a continuación. [22].

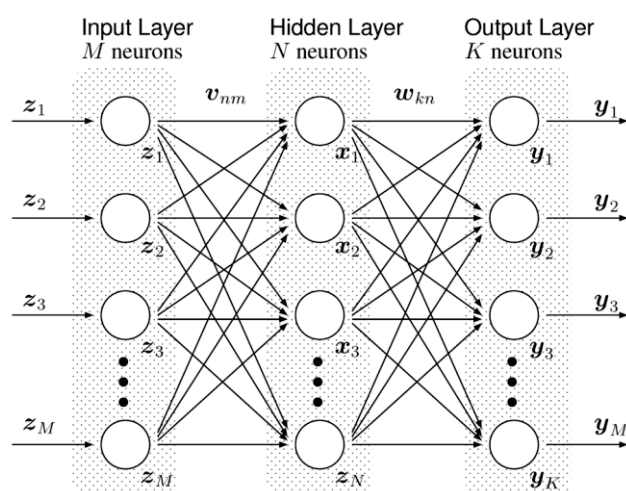


Figura 4.4 Arquitectura MLP

Las neuronas de la capa oculta usan como regla de propagación la suma ponderada de las entradas con los pesos sinápticos w_{ij} y sobre esa suma ponderada se aplica una función de transferencia de tipo sigmoïdal, que es acotada en respuesta. El aprendizaje que se suele usar en este tipo de redes recibe el nombre de retro propagación del error (backpropagation). Como función de coste global, se usa el error cuadrático medio [23].

4.2.2.2 Bosque aleatorio (Random Forest)

Random Forest es una combinación de árboles predictores, tal que cada árbol depende de los valores de un vector aleatorio probado independientemente y con la misma distribución para cada uno de estos. Es una modificación sustancial de bagging que construye una larga colección de árboles no correlacionados y luego los promedia [24].

La idea esencial del bagging es promediar muchos modelos ruidosos, pero aproximadamente imparciales, y por tanto reducir la variación. Los árboles son los candidatos ideales para el bagging, dado que ellos pueden registrar estructuras de interacción compleja en los datos, y si crecen suficientemente profundo, tienen relativamente baja parcialidad. Producto de que los árboles son notoriamente ruidosos, ellos se benefician grandemente al promediar.

Cada árbol es construido usando el siguiente algoritmo:

1. Sea N el número de casos de prueba, M es el número de variables en el clasificador.
2. Sea m el número de variables de entrada a ser usado para determinar la decisión en un nodo dado; m debe ser mucho menor que M
3. Elegir un conjunto de entrenamiento para este árbol y usar el resto de los casos de prueba para estimar el error.
4. Para cada nodo del árbol, elegir aleatoriamente m variables en las cuales basar la decisión. Calcular la mejor partición a partir de las m variables del conjunto de entrenamiento.

Para la predicción un nuevo caso es empujado hacia abajo por el árbol. Luego se le asigna la etiqueta del nodo terminal donde termina. Este proceso es iterado por todos los árboles en el ensamblado, y la etiqueta que obtenga la mayor cantidad de incidencias es reportada como la predicción. [25]

4.2.2.3 Regresión lineal (Linear Regression)

El análisis de regresión permite determinar el grado de dependencia de las series de valores X e Y , prediciendo el valor “ y ” estimado que se obtendría para un valor “ x ” que no esté en la distribución. Existen distintos tipos de regresión lineal, y entre los principales se encuentran la regresión simple y múltiple [26].

La regresión lineal simple se basa en estudiar los cambios en una variable, no aleatoria, que afectan a una variable aleatoria, en el caso de existir una relación funcional entre ambas variables que puede ser establecida por una expresión lineal, es decir, su representación gráfica es una línea recta. En otras palabras, se está en presencia de una regresión lineal simple cuando una variable independiente ejerce influencia sobre otra variable dependiente; y puede expresarse de la siguiente manera:

$$Y = f(x) \quad (4.2)$$

La regresión lineal múltiple permite trabajar con una variable a nivel de intervalo o razón. De esta forma, se puede comprender la relación de dos o más variables, permitiendo relacionarlas mediante ecuaciones. La regresión lineal múltiple se aplica cuando dos o más variables independientes influyen sobre una variable dependiente; y se puede expresar de la siguiente manera:

$$Y = f(x, w, z) \quad (4.3)$$

4.2.3 Métricas de evaluación de desempeño

Para poder evaluar y escoger entre aplicar uno u otro modelo, es necesario emplear diversas métricas que expliquen la precisión del pronóstico realizado. En esta parte del trabajo, se mostrarán las métricas más utilizadas en la evaluación de modelos de pronóstico.

4.2.3.1 Error cuadrático medio

En estadística, el error cuadrático medio (MSE o mean square error en inglés) es una forma de evaluar la diferencia entre un estimador y el valor real de la cantidad que se quiere calcular. El MSE mide el promedio del cuadrado del "error", siendo el error el valor en la que el estimador difiere de la cantidad a ser estimada [27].

Se caracteriza por penalizar mucho a valores de error grandes, y a quitar importancia a valores de error bajo cero.

$$ECM = \frac{\sum_{t=1}^n (x_t - \hat{x}_t)^2}{n} \quad (4.4)$$

4.2.3.2 Error absoluto medio

Mide el promedio absoluto del error, penalizando a valores altos o pequeños en la misma magnitud.

$$MAE = \frac{\sum_{t=1}^n |x_t - \hat{x}_t|}{n} \quad (4.5)$$

4.2.3.3 Error porcentual absoluto medio

Expresa la exactitud como un porcentaje del error. Como este número es un porcentaje, puede ser más fácil de entender que los otros indicadores. Por ejemplo, si el MAPE es 5, en promedio, el pronóstico está errado en un 5% [28]. La ecuación es:

$$MAPE = \frac{\sum_{t=1}^n |(x_t - \hat{x}_t)/x_t|}{n} * 100, (x_t \neq 0) \quad (4.6)$$

4.2.3.4 Coeficiente de determinación

El coeficiente de determinación es una medida estadística de la bondad del ajuste o fiabilidad del modelo estimado a los datos. Se representa por R^2 e indica cuál es la proporción de la variación total en la variable dependiente (Y), que es explicada por el modelo de regresión estimado, es decir, mide la capacidad explicativa del modelo estimado [29].

$$R^2 = \frac{S_{xy}^2}{S_x^2 S_y^2} \quad (4.7)$$

5 Datos

En este capítulo se abordará todo lo relacionado a los datos sobre los cuales se realizarán los experimentos de la investigación. Se hará un análisis por cada base de datos, mostrando tablas que resuman aspectos de tamaño y estructura. Se hará también una descripción de sus columnas o atributos, y se mostrarán gráficos que enseñan el comportamiento de los datos.

Por último, se explicará el pre-procesado de los datos, a fin de preparar la información necesaria para la correcta aplicación de los experimentos.

5.1 Bases de datos de consumo eléctrico

Los datos que se utilizaron en la experimentación fueron obtenidos a partir de medidores inteligentes de consumo eléctrico. Estos medidores son propiedad de una empresa distribuidora de energía eléctrica en la Quinta región de Valparaíso, Chile. La base de datos de consumo eléctrico consiste en dos archivos (.csv), que contienen mediciones registradas entre los años 2014 y 2015.

Las mediciones están expresadas en KWh, y son grabadas en intervalos de 15 minutos, que comienzan a desde las 00:00:00 hrs. De esta manera, se espera que cada medidor registre 96 datos de consumo diariamente, completando así 24 horas de consumo en intervalos de 15 minutos. A continuación, las tablas 5.1 y 5.2 muestran detalles de los archivos de bases de datos.

Se considera:

N° esperado de registros por medidor = $96 * N^{\circ}$ días

N° esperado de registros = N° medidores * N° esperado de registros por medidor

Tabla 5.1 Detalles BDD Consumo 2014

Archivo BDD 2014	
Tamaño (GB)	1,87
Rango lecturas (inicio/fin)	01-04-2014 / 31-12-2014
N° meses	9
N° días	275
N° medidores	1.271
N° esperado de registros totales	33.554.400
N° real de registros totales	15.060.868
N° esperado de registros por medidor	26.400
N° real promedio de registros por medidor	11.850
N° medidores con registros sobre el promedio	672
N° medidores con registros bajo el promedio	599

Tabla 5.2 Detalles BDD Consumo 2015

Archivo BDD 2015	
Tamaño (GB)	4,6
Rango lecturas (inicio/fin)	01-01-2015 / 01-10-2015
N° meses	9
N° días	273
N° medidores	1.542
N° esperado de registros totales	40.412.736
N° real de registros totales	37.011.674
N° esperado de registros por medidor	26.208
N° real promedio de registros por medidor	24.002
N° medidores con registros sobre el promedio	1.233
N° medidores con registros bajo el promedio	309

En conjunto, los dos archivos registran el consumo de 1.555 clientes, sumando un total de 52.072.542 mediciones.

Cabe mencionar que, conforme a los datos expresados en la tabla 5.1, existe una deficiencia de registros en el año 2014. Sólo el 52,87% de los medidores en ese periodo presenta un número de mediciones superior al promedio real de registros. Por otra parte, en el año 2015 el porcentaje asciende a un 80%, como puede observarse en la tabla 5.2. Las figuras 5.1 y 5.2 expresan de mejor manera la distribución de la cantidad de registros en ambos años.

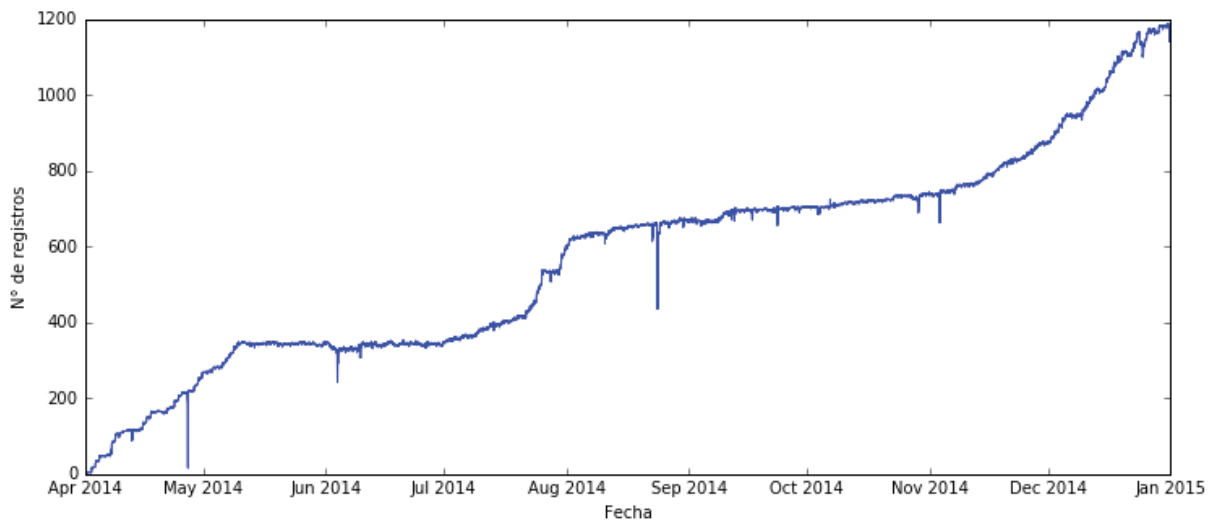


Figura 5.1 Cantidad de mediciones por fecha, año 2014

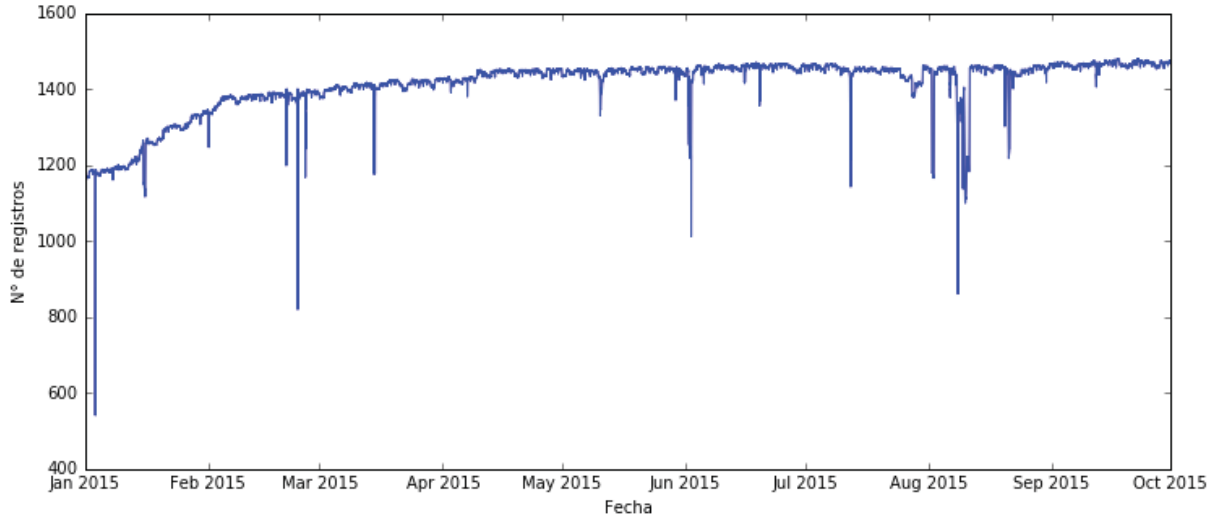


Figura 5.2 Cantidad de mediciones por fecha, año 2015

5.1.1 Estructura de bases de datos de consumo

Para poder trabajar correctamente sobre los datos de consumo, es necesario conocer cómo están estructurados. En la figura 5.3 se muestra un extracto de la base de datos de consumo.

Reading Date	CIM Code	Reading Type Description	Value	Reading Quality Code	Received Date	Mac Address	Meter Serial
10/01/2015 00:00:00 GMT	2.0.4.1.0.12.0.0.0.3.72	Delivered 15 min Energy kWh	0.048	DST Flag	10/01/2015 00:17:44 GMT	0x412b0	10004474
09/30/2015 23:45:00 GMT	2.0.4.1.0.12.0.0.0.3.72	Delivered 15 min Energy kWh	0.048	DST Flag	10/01/2015 00:17:44 GMT	0x412b0	10004474
09/30/2015 23:30:00 GMT	2.0.4.1.0.12.0.0.0.3.72	Delivered 15 min Energy kWh	0.047	DST Flag	09/30/2015 23:34:05 GMT	0x412b0	10004474
09/30/2015 23:15:00 GMT	2.0.4.1.0.12.0.0.0.3.72	Delivered 15 min Energy kWh	0.050	DST Flag	09/30/2015 23:34:05 GMT	0x412b0	10004474
09/30/2015 23:00:00 GMT	2.0.4.1.0.12.0.0.0.3.72	Delivered 15 min Energy kWh	0.095	DST Flag	09/30/2015 23:34:05 GMT	0x412b0	10004474
09/30/2015 22:45:00 GMT	2.0.4.1.0.12.0.0.0.3.72	Delivered 15 min Energy kWh	0.097	DST Flag	09/30/2015 22:48:04 GMT	0x412b0	10004474
09/30/2015 22:30:00 GMT	2.0.4.1.0.12.0.0.0.3.72	Delivered 15 min Energy kWh	0.098	DST Flag	09/30/2015 22:48:04 GMT	0x412b0	10004474
09/30/2015 22:15:00 GMT	2.0.4.1.0.12.0.0.0.3.72	Delivered 15 min Energy kWh	0.097	DST Flag	09/30/2015 22:48:04 GMT	0x412b0	10004474
09/30/2015 22:00:00 GMT	2.0.4.1.0.12.0.0.0.3.72	Delivered 15 min Energy kWh	0.096	DST Flag	09/30/2015 22:48:04 GMT	0x412b0	10004474
09/30/2015 21:45:00 GMT	2.0.4.1.0.12.0.0.0.3.72	Delivered 15 min Energy kWh	0.096	DST Flag	09/30/2015 22:48:04 GMT	0x412b0	10004474
09/30/2015 21:30:00 GMT	2.0.4.1.0.12.0.0.0.3.72	Delivered 15 min Energy kWh	0.095	DST Flag	09/30/2015 22:48:04 GMT	0x412b0	10004474
09/30/2015 21:15:00 GMT	2.0.4.1.0.12.0.0.0.3.72	Delivered 15 min Energy kWh	0.084	DST Flag	09/30/2015 22:48:04 GMT	0x412b0	10004474

Figura 5.3 Estructura BDD de consumo

Como se observa en la Figura 6, la base de datos de consumo eléctrico posee distintos atributos, los cuales son: Reading Date, CIM Code, Reading Type Description, Value, Reading Quality Code, Received Date, Mac Address y Meter Serial. Cada atributo se describe brevemente a continuación.

Reading Date: Fecha y hora de lectura.

Reading Type Description: Descripción del tipo de dato, siendo “Delivered 15 min Energy KWh”, la energía activa suministrada al cliente.

Value: Energía consumida (sin considerar la constante), en un intervalo de 15 min.

Reading Quality Code: Se refiere a la clasificación o calidad de la lectura enviada por el medidor. Ésta puede tomar los siguientes valores: Configuration Changed, Overflow, DST Flag, Partial Interval, LP Recording Stopped, Clock Set Backward, Test Data, Clock Set Forward, Long Interval, Data Valid, Power Fail, Skipped Interval. La mayor parte de los registros posee el atributo “Data Valid” y “DST Flag”, indicando de esta manera una lectura correcta del consumo.

Received Date: fecha en la que se recibió la información.

Mac Address: Identificador de red del medidor.

Meter Serial: Número serial del medidor

5.2 Base de datos de clientes

Además de los datos de consumo, también se dispone de una base de datos de clientes. En ella puede identificarse la zona a la cual pertenece cada medidor y su tipo de tarifa. A diferencia los archivos detallados anteriormente, esta base de datos es muy pequeña y tiene un peso de tan sólo 69 KB. En su interior contiene los datos de 1.831 clientes pertenecientes a la V región de Valparaíso. En la sección siguiente se muestra la estructura de los datos.

5.2.1 Estructura base de datos de clientes

La estructura de la base de datos de clientes se muestra, como un extracto, en la figura a continuación.

Mac Address	Ubicación	Tipo Tarifa	Tarifa	Cte.
0x0004098F	QUILPUE	Normal	BT-1A	1
0x0004089E	QUILPUE	Normal	BT-1A	1
0x00040FD9	QUILPUE	Normal	BT-1A	1
0x00040F9F	QUILPUE	Normal	BT-1A	1
0x00040EC3	Quilpue	Normal	BT-1A	1
0x00040E65	Quilpue	Normal	BT-1A	1
0x00040F31	Quilpue	Normal	BT-1A	1
0x00040E03	Quilpue	Normal	BT-1A	1
0x00040D39	QUILPUE	Normal	BT-1A	1
0x00040886	QUILPUE	Normal	BT-1A	1
0x00040DB6	QUILPUE	Normal	BT-1A	1
0x00040CCD	QUILPUE	Normal	BT-1A	1
0x00040BD3	VALPARAISO	Normal	BT-1A	1
0x000405D4	VALPARAISO	Normal	BT-1A	1

Figura 5.4 Estructura BDD de clientes

Como se muestra en la figura 5.4, cada cliente posee cuatro atributos que son: Ubicación, Tipo Tarifa, Tarifa y Cte. Éstos se describen a continuación.

Ubicación: Se refiere a la zona geográfica en que se encuentra ubicado el medidor.

Tipo Tarifa: Este atributo varía dependiendo de la tarifa de cada cliente. Las tarifas BT-1, BT-2, AT-1 y AT-2, son de tipo Normal. Mientras que tarifas BT-3, BT-4, AT-3 y AT-4, son de tipo Demanda. El resto de tarifas se clasifica como Sin facturación.

Tarifa: Indica la tarifa que tiene por contrato el cliente.

Cte.: Indica la constante de energía, valor que se multiplica sobre el monto final a cobrar, dependiendo del tipo de instalación.

Los atributos descritos pueden adoptar los siguientes valores:

Ubicación	Los Andes	Valparaíso	Tipo tarifa	Normal
	Quillota	Viña del Mar		Demanda
	Quilpué			Sin facturación

Tarifa	AT-3PP	BT-1A	BT-3PP	BT-3PPP Prov	Micromedición
	AT-3PPP	BT-2PP	BT-3PP Prov	BT-4.1	NetBilling
	AT-4.3	BT-2PPP	BT-3PPP	BT-4.3	

Cte.	$0 < \text{Cte.} < 2000$
-------------	--------------------------

Del total de 1.555 clientes que presentan registros de consumo, 1504 de ellos pueden identificarse en esta base de datos. La distribución de atributos queda de la siguiente manera:

Tabla 5.3 N° de clientes por tipo de tarifa

Tipo tarifa	N° clientes
Demanda	517
Normal	935
Sin facturación	52

La tabla 5.3, muestra que la mayoría de clientes en estudio posee un tipo tarifa Normal, cuyo cobro se calcula en base al consumo registrado durante el mes. También existe un importante número de clientes con tipo de tarifa por Demanda, cuyo cobro se calcula en base a la demanda máxima de potencia durante el mes. Por último, solo 52 clientes presentan un tipo de tarifa sin facturación.

La tabla 5.4 que se muestra a continuación, indica que la tarifa con mayor cantidad de clientes en estudio es BT-1A. Cabe recordar que esta tarifa se clasifica como Normal. Es seguido por la tarifa BT-3PPP, con un total de 278 clientes, cuya clasificación de cobro es por Demanda. La tercera tarifa con mayoría de clientes es la BT-4.3, cuya clasificación es también de Demanda.

Tabla 5.4 N° de clientes por tarifa

Tarifa	N° clientes
AT-3PPP	10
AT-4.3	5
BT-1A	903
BT-2PP	21
BT-2PPP	8
BT-3PP	88
BT-3PPP	278
BT-4.1	3
BT-4.3	136
Micro medición	51
NetBilling	1

Tabla 5.5 N° de clientes por ubicación

Ubicación	N° clientes
Quillota	5
Quilpué	873
Valparaíso	626

En la tabla 5.5, se observa que la mayoría de clientes se concentra entre las ciudades de Quilpué y Valparaíso, con 873 y 626 clientes respectivamente. En último lugar queda la ciudad de Quillota, que solo cuenta con 5 clientes.

Tabla 5.6 N° de clientes por cte. de energía

Cte.	N° clientes
0	52
1	1166
10	60
20	65
30	9
40	89
60	38
80	13
100	2
120	10

La tabla 5.6, indica que 1166 clientes tienen asignada una constante de energía con valor 1. Para los otros valores de constante también existen clientes, pero en mucho menor cantidad.

5.3 Pre-procesado de datos

Antes de comenzar con el curso de los experimentos, y luego de haber analizado las estructuras de las bases de datos a utilizar, es necesario hacer algunos ajustes que permitan un desarrollo correcto de la investigación.

Primero, se verifica la existencia de valores nulos dentro de los registros de consumo. Cada fila que presente esta condición deberá ser removida. Luego, se debe eliminar la duplicidad de registros; cuando hay duplicidad, el descubrimiento de patrones y el análisis de la serie temporal se ven altamente afectados por valores anómalos de consumo registrados en un instante de tiempo determinado. Por lo tanto, se verifica por cada cliente que no presente más de un registro por cada fecha e instante de tiempo.

Además, se deben filtrar todos los intervalos incorrectos que no se correspondan con el orden de las mediciones cada 15 minutos, tomando como inicio las 00:00:00 hrs. de cada día. Cuando hay existencia de intervalos intermedios, por ejemplo, registros a las 00:03:00 hrs. se produce una desconfiguración de la serie temporal, lo que afectará directamente al desempeño de los modelos de pronósticos.

Por último, se realiza un filtro del atributo “Reading Quality Code”, que corresponde a la calidad de la lectura. Para este atributo sólo se tomarán en cuenta aquellos registros con valores “Data Valid” o “DST Flag”, debido a que otros tipos de calidad de lectura son en su mayoría indicadores de error.

6 Experimentos

En el presente capítulo se explicarán las metodologías y experimentos realizados. Se darán a conocer primero las herramientas y librerías necesarias para el desarrollo del proyecto; luego se detallará el proceso de experimentación para el descubrimiento de patrones de consumo mediante Clustering; finalmente se abordará el pronóstico de consumo eléctrico a través de modelos predictivos.

6.1 Herramientas y librerías utilizadas

El proyecto será desarrollado utilizando el lenguaje de programación Python, en su versión 3.5. La vasta cantidad de librerías especializadas para el análisis de datos hacen de este lenguaje el más apropiado para la realización de los experimentos. En los siguientes párrafos se nombran las librerías y herramientas más relevantes que serán utilizadas.

6.1.1 Spyder 3

Spyder es un entorno de desarrollo (IDE) para Python, orientado especialmente para el uso científico. Sus principales ventajas son la exploración de variables, que permite monitorear el valor de éstas durante la ejecución del programa, y las funciones de dibujo de gráficas integradas, que permiten visualizar directamente los gráficos en la consola. Es un software multiplataforma, y su descarga está dispuesta gratuitamente desde su sitio web oficial.

6.1.2 Pandas

Pandas es una librería de análisis de datos, que provee estructuras útiles para el manejo de los mismos, tales como series y dataframes. En el caso de los dataframes, éstos se asemejan a tablas de bases de datos relacionales, que permiten cargar en memoria, manipular y procesar fácilmente grandes cantidades de datos. Una de las mayores ventajas de esta librería, es que ofrece una función dedicada para la carga de archivos con extensión *.csv; de esta forma, se hace muy sencilla la utilización de archivos como los mostrados en el capítulo 5.

6.1.3 NumPy

Ésta es una librería dedicada especialmente al manejo de vectores y matrices. Provee una gran cantidad de funciones matemáticas que pueden aplicarse directamente sobre los arreglos de datos, aportando así mayor flexibilidad en la obtención de indicadores estadísticos y cálculos complejos.

6.1.4 SciPy: fftpack

SciPy es una librería que ofrece herramientas y algoritmos matemáticos para un cálculo numérico de gran capacidad. Para el proyecto se utilizará el paquete fftpack, que contiene funciones especiales para la aplicación de la transformada rápida de Fourier. Su utilización se explicará más adelante en la sección 6.2.1.4.

6.1.5 Matplotlib Pyplot (Generación de gráficos)

Todos los gráficos del proyecto se realizarán mediante el uso de Pyplot. Esta librería ofrece una gran cantidad de funciones que permiten generar gráficos completamente personalizados y de alta calidad.

6.1.6 Scikit-learn

Para realizar el proceso de clustering y pronósticos se utilizará una librería llamada Scikit-learn. Esta librería se usa frecuentemente en “Machine Learning”, pues provee una serie de herramientas simples y eficientes que permiten realizar, entre otras cosas, procesos de: Clasificación, Regresión, Clustering, Reducción dimensional, Selección de modelo y Pre-procesado.

Para este proyecto, se aplicará el algoritmo de clustering “K-means”, que se caracteriza por su simplicidad y eficacia en el descubrimiento de patrones. Tal como fue descrito previamente en la sección 4.3, el K-means es uno de los algoritmos más utilizados, ya que permite al analista de datos definir la cantidad de clusters que se desea obtener.

Para los experimentos de pronóstico se utilizarán las funciones de regresión lineal, regresión bayesiana, bosque aleatorio y perceptrón multicapa de regresión; todos estos modelos se encuentran incluidos en esta librería.

6.2 Clustering

En esta parte del documento, se detallará la experimentación y el proceso de aplicación de algoritmos clustering para el descubrimiento de patrones de consumo eléctrico. Se explicará con detención las representaciones propuestas para caracterizar el consumo de cada cliente. Luego, se darán algunos detalles de la implementación del clustering, como algoritmos utilizados y parámetros establecidos.

6.2.1 Representación de patrones de consumo

Para realizar el clustering y agrupar clientes con patrones de consumo similares, es necesario representar de alguna manera el consumo individual de cada cliente. En esta investigación se crearon cuatro tipos de representación. La primera consiste en un patrón de consumo diario; la segunda es un patrón de consumo por semana (lunes a viernes); la tercera describe un patrón de consumo para fines de semana (viernes y domingo); la cuarta es un patrón basado en la transformada rápida de Fourier.

Por cada representación se hizo un análisis completo y otro basado en horas punta. El análisis completo considera todas las mediciones de consumo en cualquier horario y fecha. Mientras que, el análisis de horario punta se realiza sobre todas aquellas mediciones registradas entre las 18:00 y las 23:00 horas durante los meses de abril, mayo, junio, julio, agosto y septiembre [30]. Es importante incluir este análisis, ya que, en dicho periodo se incorporan cargos como energía adicional de invierno para clientes con tarifa BT1; además, el cobro por potencia consumida es aproximadamente 10 veces mayor que el cobro por potencia en horas fuera de punta para clientes con tarifa BT4 o AT4.

Para ambos tipos de análisis, se agruparon los datos en dos modalidades: resumen general y resumen por estación del año. Cada resumen contiene el patrón de consumo de todos los clientes en estudio. El resumen general considera todo el rango de fechas posibles de acuerdo al tipo de análisis; mientras que el resumen por estaciones del año restringe fechas de acuerdo a cada estación para su posterior análisis. A continuación, se muestra un mapa conceptual que resume la metodología de experimentación para cada patrón de consumo.

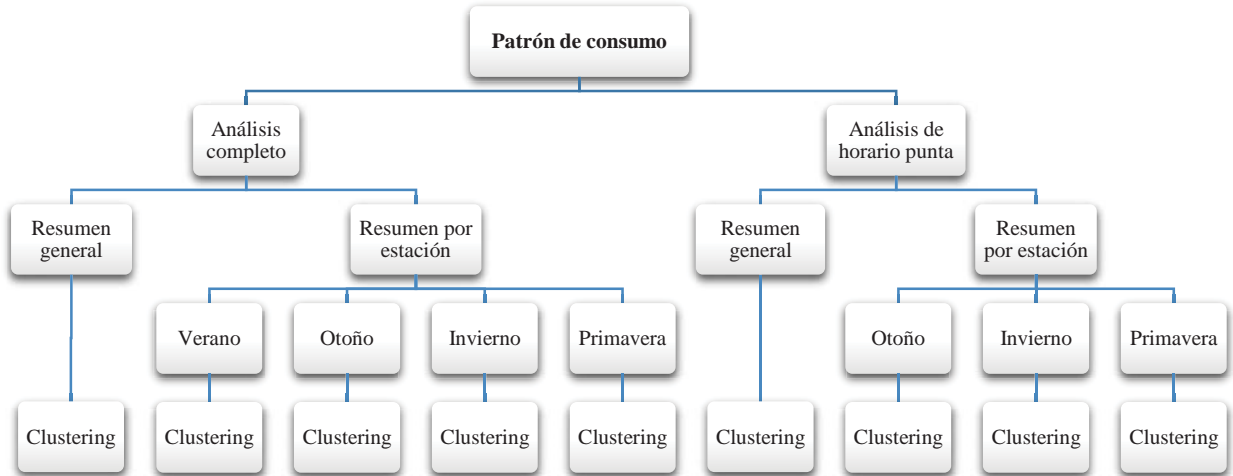


Figura 6.1 Mapa conceptual de experimentos clustering

En los siguientes apartados, se explicará detalladamente cada patrón de consumo, y se hará un estudio descriptivo de los resúmenes generales y por estaciones, con análisis completo de horario. Los gráficos y tablas, para el estudio descriptivo con análisis de horario punta, se encuentran en el Anexo A adjunto a este documento.

6.2.1.1 Patrón de consumo diario

El patrón de consumo diario de un cliente consiste en un vector de N mediciones continuas en intervalos de 15 minutos; cada elemento del vector corresponde al promedio de consumo en cada intervalo de tiempo. El valor N depende del tipo de análisis de horario que se utilice. En el caso del análisis completo de horario, N tomará un valor de 96, lo que equivale a un lapso de 24 horas continuas, registradas desde las 00:00 hrs. a las 23:45 hrs. En el caso del análisis de horario punta, N tomará un valor de 20, lo que equivale a un lapso de 4 horas continuas, registradas entre las 18:00 y las 23:00 hrs.

El patrón de consumo diario de un cliente 'S' queda expresado de la siguiente manera:

$$\text{consumo_diario}^S(N) = [\bar{C}^S(h_1), \bar{C}^S(h_2), \bar{C}^S(h_3), \dots, \bar{C}^S(h_N)], \text{ donde}$$

$$\bar{C}^S(h_i) = \frac{1}{n} \sum_{k=1}^n C_k^S(h_i) \quad (6.1)$$

h_i intervalo de tiempo, con $i = \{1, 2, \dots, N\}$

$n =$ cantidad de registros de cliente S en h_i

$C_k^S(h_i) = k - \text{ésimo registro de cliente } S \text{ en } h_i$

$N=96$

$N=20$

$h_1 = 00:00 \text{ hrs.}$

$h_1 = 18:00 \text{ hrs.}$

$h_2 = 00:15 \text{ hrs.}$

$h_2 = 18:15 \text{ hrs.}$

$h_3 = 00:30 \text{ s.}$

$h_3 = 18:30 \text{ s.}$

\vdots

\vdots

$h_{96} = 23:45 \text{ hrs.}$

$H_{20} = 23:00 \text{ hrs.}$

Para aplicar esta representación, se consideró solamente los días en que el cliente posee un registro completo de 96 intervalos. De esta manera, el promedio de consumo en cada horario (h_i) se calculará siempre sobre un mismo valor n .

En la siguiente figura se muestra un gráfico de resumen general, con el patrón de consumo diario (PD) de todos los clientes, usando un análisis completo (FULL).

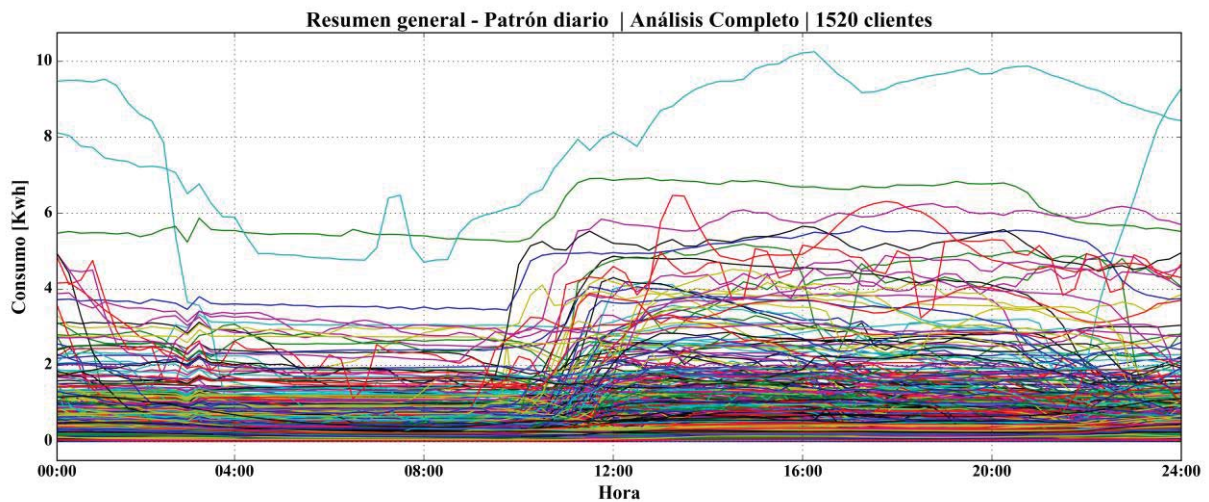


Figura 6.2 Resumen general - PD | FULL

La Figura 6.2 revela que existen clientes con un consumo muy superior al resto, alcanzando registros incluso superiores a 10 KWh. También se observa un aumento en los niveles de consumo desde cerca de las 12:00 hrs. en adelante. En la siguiente tabla se muestran los indicadores descriptivos correspondientes al gráfico.

Tabla 6.1 Indicadores descriptivos - PD | FULL | General

Promedio (KWh)	Desv. estándar (KWh)	Suma total (KWh)	Valor máximo (KWh)	Valor mínimo (KWh)
0,264	0,603	38079,266	10,282	0,000

Como muestra la Tabla 6.1, el promedio de consumo es de 0,264 KWh, y se aleja considerablemente del valor máximo de consumo registrado de 10,282 KWh. La desviación estándar es pequeña, lo que da a entender que los valores se concentran mayormente en

cantidades bajas de consumo, entre 0 y 1 KWh. Esto es congruente con la cantidad de clientes con tarifa BT1, que es el grupo mayoritario dentro de los medidores estudiados.

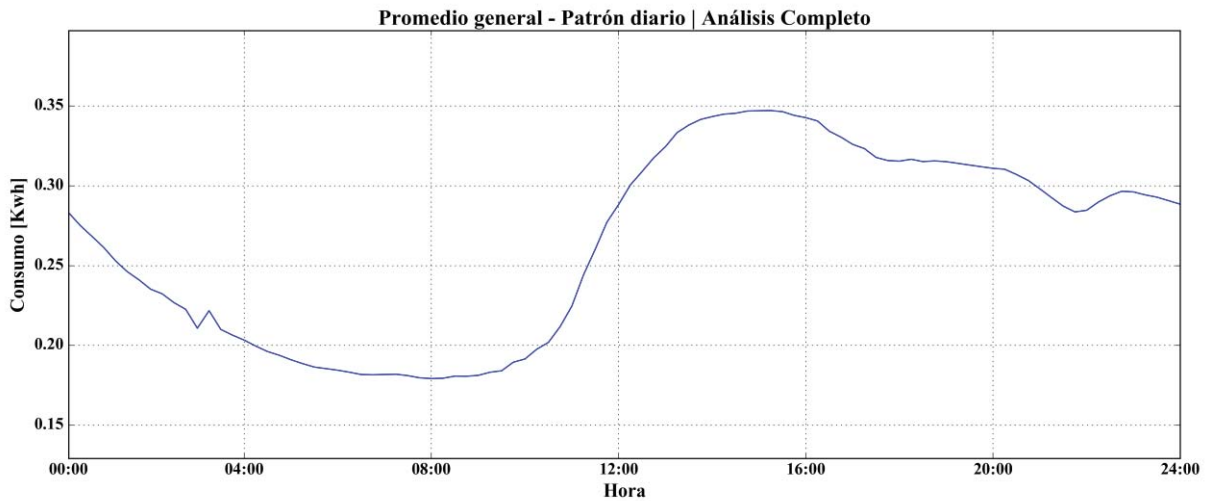


Figura 6.3 Promedio general – PD | FULL

La Figura 6.3, muestra el promedio general de consumo en cada horario del día. Se puede observar que los niveles decrecen desde la medianoche, alcanzando el promedio más bajo a eso de las 08:00 hrs. El promedio más alto se alcanza entre las 12:00 y las 16:00 hrs

A continuación, se muestran los gráficos de resumen por estaciones del año, para el patrón diario con análisis completo de horario.

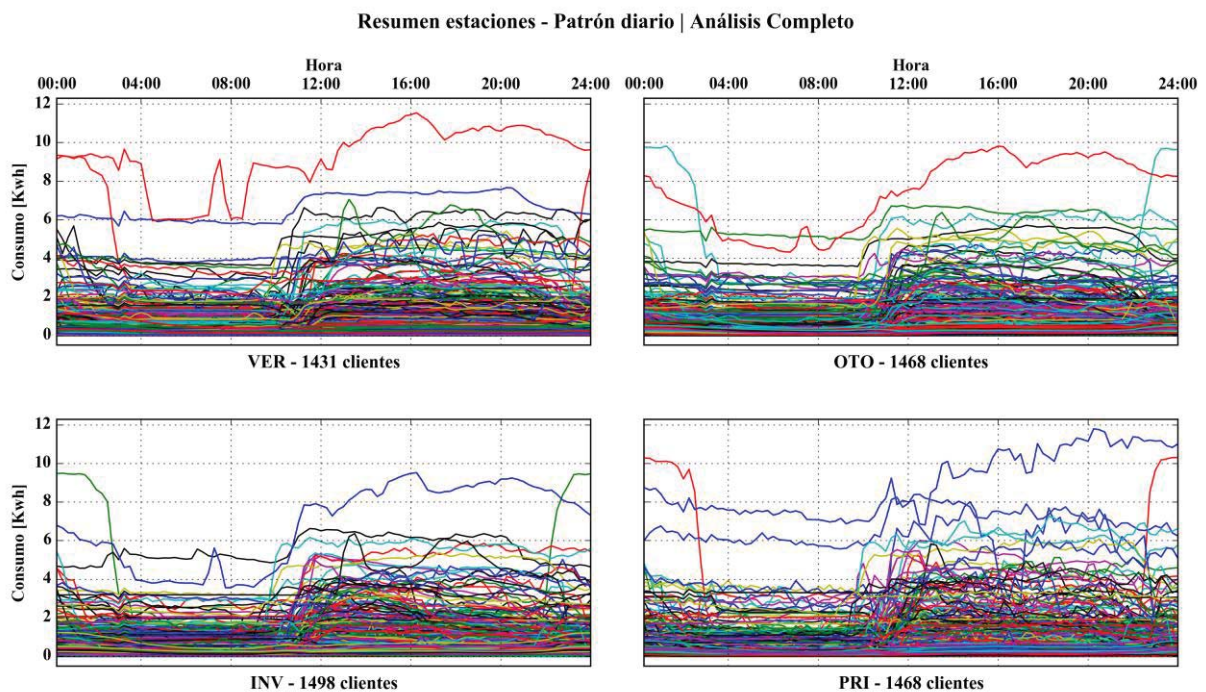


Figura 6.4 Resumen por estaciones – PD | FULL

La figura 6.4 permite notar que tanto en verano como en primavera existe un cliente que presenta un consumo cercano a los 12 KWh. Mientras que, en otoño e invierno, los clientes no registran niveles tan altos, los cuales bordean los 10 KWh. La tabla a continuación entrega los estadísticos correspondientes al gráfico, y corrobora el análisis realizado.

Tabla 6.2 Indicadores descriptivos - PD | FULL | Estaciones

Estación	Promedio (KWh)	Desv. estándar (KWh)	Suma total (KWh)	Valor máximo (KWh)	Valor mínimo (KWh)
Verano	0,266	0,650	36172,877	11,650	0,000
Otoño	0,263	0,600	36662,283	9,819	0,000
Invierno	0,262	0,587	37357,318	9,533	0,000
Primavera	0,265	0,652	36980,490	11,799	0,000

También, la tabla 6.2 indica que la suma total de consumo es mayor durante el invierno, dando a entender que el consumo eléctrico aumenta en dicha estación. En la siguiente figura se muestra el promedio por estaciones, para cada horario del día.

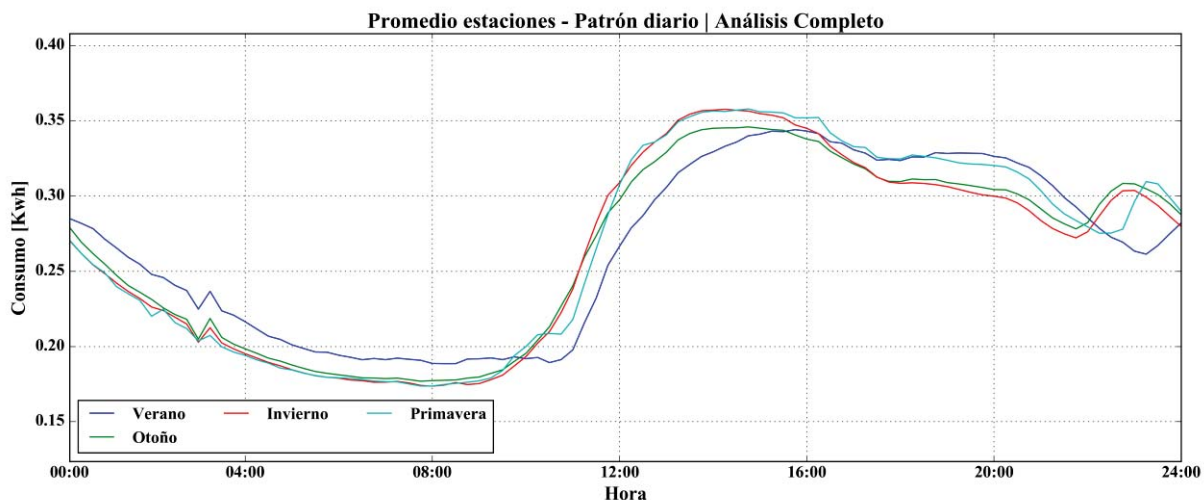


Figura 6.5 Promedio estaciones – PD | FULL

Las curvas promedio, mostradas en la figura 6.5, son similares para cada estación, variando levemente en los niveles de consumo. No obstante, se puede observar una pequeña variación en el comportamiento durante la estación de verano. Mientras que en las demás estaciones se presenta un aumento pronunciado al final del día, en verano el consumo tiene una baja sostenida, incrementando levemente al final.

6.2.1.2 Patrón de consumo semana (lunes a viernes)

Para modelar de manera más detallada el comportamiento de cada cliente, se diseñó una representación de consumo eléctrico semanal. Este patrón considera los días lunes a viernes, y promedia el consumo en cada intervalo de tiempo, de manera separada por cada día de la semana. De esta forma, se genera una matriz de dimensiones 5xN que modela el consumo del cliente. Al igual que el patrón diario, el valor de N depende del tipo de análisis a realizar, siendo

96 y 20 los valores para el análisis completo y horario punta, respectivamente. La representación para un cliente 'S' queda expresada de la siguiente manera:

$$\text{consumo_semana}^S[5][N] = \begin{matrix} \bar{C}_1^S(h_1) & \bar{C}_1^S(h_2) & \dots & \bar{C}_1^S(h_N) \\ \vdots & \vdots & & \vdots \\ \bar{C}_5^S(h_1) & \bar{C}_5^S(h_2) & \dots & \bar{C}_5^S(h_N) \end{matrix}, \text{ donde}$$

$$C_d^S(h_i) = \frac{1}{n} \sum_{k=1}^n C_k^{d,S}(h_i) \quad (6.2)$$

$d = \text{día de la semana} = \{1,2,3,4,5\} = \{\text{Lun}, \text{Mar}, \text{Mie}, \text{Jue}, \text{Vie}\}$
 $n = \text{cantidad de días "d" registrados por el cliente S}$
 $h_i = \text{intervalo de tiempo, con } i = \{1, 2, \dots, N\}$

$N=96$	$N=20$
$h_1 = 00:00 \text{ hrs.}$	$h_1 = 18:00 \text{ hrs.}$
$h_2 = 00:15 \text{ hrs.}$	$h_2 = 18:15 \text{ hrs.}$
$h_3 = 00:30 \text{ s.}$	$h_3 = 18:30 \text{ s.}$
\vdots	\vdots
$h_{96} = 23:45 \text{ hrs.}$	$H_{20} = 23:00 \text{ hrs.}$

Al igual que el anterior, para cada cliente se consideraron solamente aquellos días en que registraron un total de 96 mediciones. A continuación, se muestra en la figura 6.6 un gráfico general con el patrón de consumo de semana (SEM) de todos los clientes, usando un análisis completo.

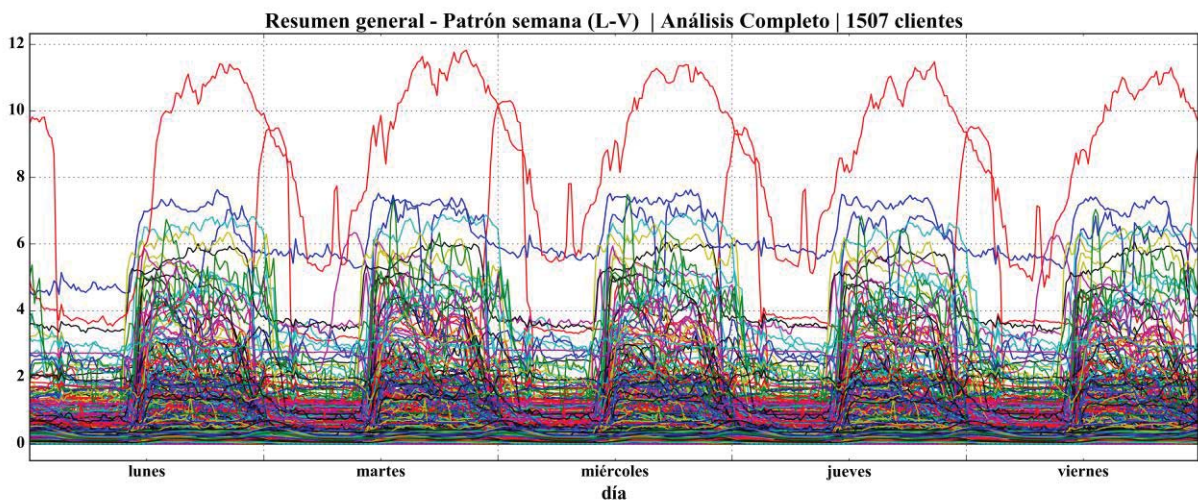


Figura 6.6 Resumen general – SEM | FULL

La figura 6.6 permite ver que el comportamiento del consumo durante los días de semana tiene un patrón repetitivo. Se logra apreciar el aumento en el nivel de consumo cerca del mediodía, registrando los niveles más bajos durante la madrugada. También, se observa un par de clientes que sobrepasan los 9 KWh en gran parte del día. Los demás clientes presentan en general niveles de consumo inferiores a 6 KWh.

Tabla 6.3 Indicadores descriptivos - SEM | FULL | General

Promedio (KWh)	Desv. estándar (KWh)	Suma total (KWh)	Valor máximo (KWh)	Valor mínimo (KWh)
0,284	0,665	204697,273	11,819	0,000

La tabla 6.3, indica una suma total de consumo de 204.697 KWh, que corresponde a la suma acumulada desde el día lunes hasta el viernes. En tanto, el promedio calculado es de 0,284 KWh. El valor máximo registrado alcanza los 11,819 KWh, mientras que el valor mínimo es 0 KWh. La desviación estándar es cercana a cero, lo que indica una concentración mayor de clientes con consumos entre 0 KWh y 1 KWh.

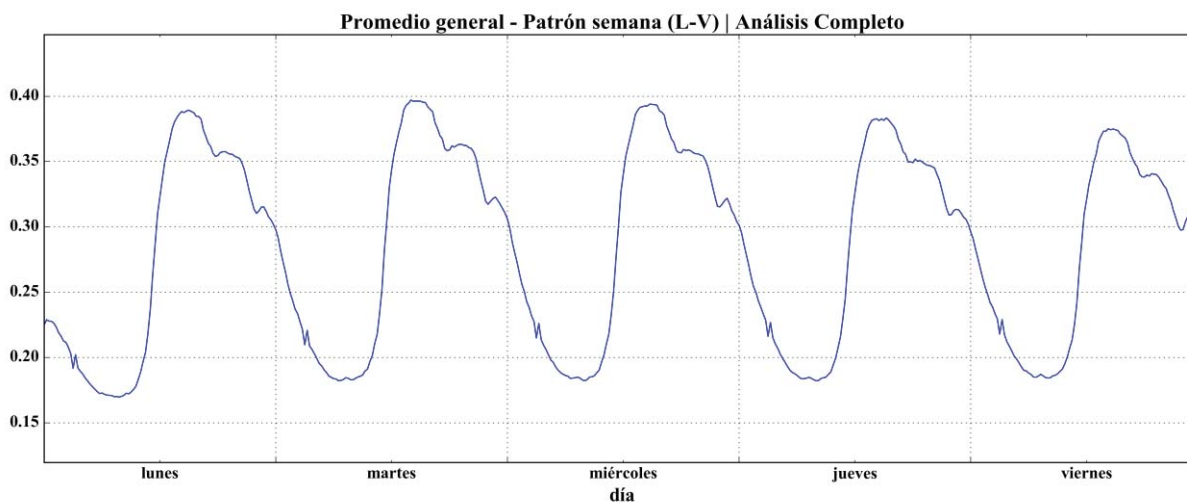


Figura 6.7 Promedio general - SEM | FULL

En la figura 6.7, se puede apreciar que el promedio describe un patrón prácticamente idéntico entre lunes y viernes. Sin embargo, el gráfico deja ver una leve disminución en el promedio de consumo máximo en los días jueves y viernes. Los valores promedio más altos de consumo se registran pasadas las 12:00 hrs. de los días martes y miércoles. Por otra parte, los valores promedio más bajos se registran durante la madrugada del día lunes.

En la siguiente figura se muestra el resumen de clientes por estación, con un análisis completo de horario.

Resumen estaciones - Patrón semana (L-V) | Análisis Completo

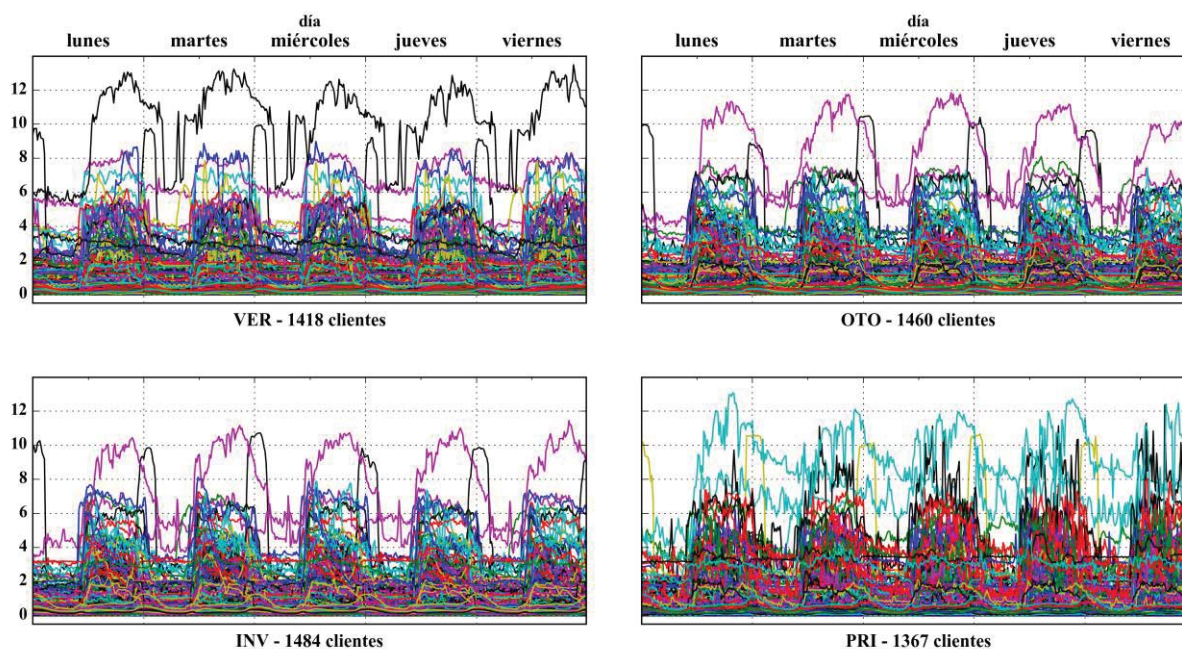


Figura 6.8 Resumen por estaciones - SEM | FULL

La figura 6.8 permite destacar rápidamente que, durante las estaciones de verano y primavera, se tiene el registro de un cliente con promedios de consumo superiores a los 12 KWh. A modo general, los niveles de consumo se sitúan mayormente entre los 0 y 4 KWh.

En la tabla 6.4 se muestran estadísticos descriptivos de cada estación.

Tabla 6.4 Indicadores descriptivos - SEM | FULL | Estaciones

Estación	Promedio (KWh)	Desv. estándar (KWh)	Suma total (KWh)	Valor máximo (KWh)	Valor mínimo (KWh)
Verano	0,284	0,706	193085,535	13,481	0,000
Otoño	0,285	0,675	199606,928	11,843	0,000
Invierno	0,285	0,664	202554,488	11,429	0,000
Primavera	0,276	0,698	180780,189	13,067	0,000

La tabla 6.4 muestra que el promedio de consumo es prácticamente el mismo, salvo en primavera, en donde se registra un promedio general un poco más bajo. Los valores máximos confirman el análisis gráfico, al mostrar que durante las estaciones de verano y primavera se registraron valores por sobre los 13 KWh; en tanto, otoño e invierno registran valores máximos inferiores a 12 KWh. Por último, se tiene que en invierno se acumula la mayor suma total de consumo, por sobre los 202.500 KWh, de lunes a viernes.

En la siguiente figura, se muestran las curvas de promedios de consumo por estación.

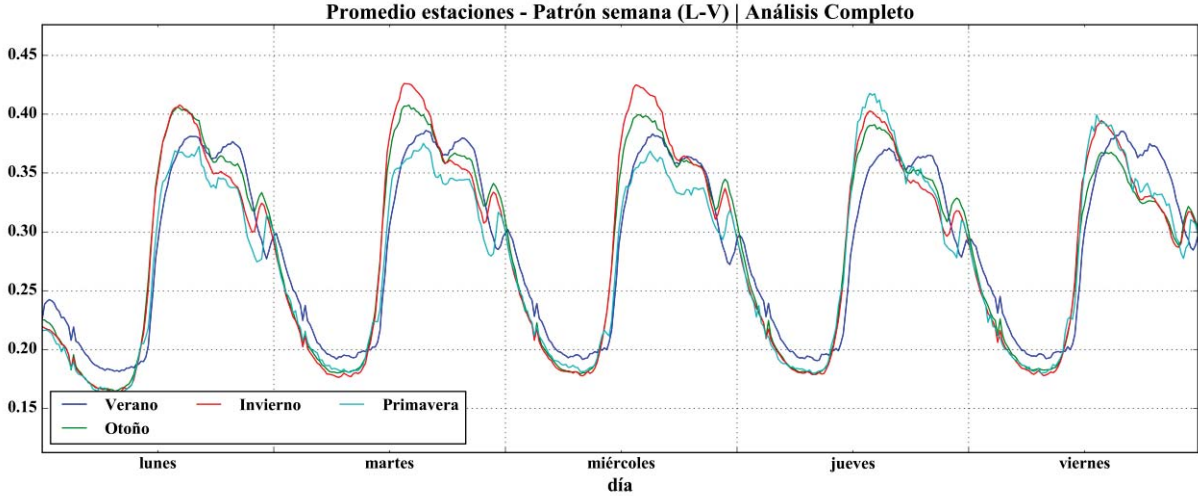


Figura 6.9 Promedio estaciones - SEM | FULL

La figura 6.9 complementa el análisis antes realizado. Se puede observar que en primavera se registra el promedio de consumo más bajo entre los días lunes y miércoles; en tanto, en el día jueves es la estación con mayor promedio de consumo. En cuanto a la estación de invierno, se tiene que entre los días lunes y miércoles presenta el promedio más alto. En verano, se registra el promedio de consumo más alto durante la madrugada.

6.2.1.3 Patrón de consumo fin de semana (sábado y domingo)

Esta representación de consumo es una variación del anterior. En vez de considerar los días lunes a viernes, ahora se busca representar el consumo durante los días sábado y domingo. De esta forma, el patrón de fin de semana queda representado por una matriz de dimensiones $2 \times N$, con $N=96$ para un análisis completo de horario, y $N=20$ para análisis de horario punta. Para un cliente 'S', el patrón se expresa de la siguiente manera.

$$\text{consumo_fin_de_semana}^S[2][96] = \begin{matrix} \bar{C}_6^S(h_1) & \bar{C}_6^S(h_2) & \dots & \bar{C}_6^S(h_N) \\ \bar{C}_7^S(h_1) & \bar{C}_7^S(h_2) & \dots & \bar{C}_7^S(h_N) \end{matrix}, \text{ donde}$$

$$C_d^S(h_i) = \frac{1}{n} \sum_{k=1}^n C_k^{d,S}(h_i) \quad (6.2)$$

$d = \text{día de la semana} = \{6,7\} = \{\text{Sáb}, \text{Dom}\}$
 $n = \text{nº de días "d" registrados por el cliente 'S'}$
 $h_i \text{ intervalo de tiempo, con } i = \{1, 2, \dots, N\}$

$N=20$	$N=96$
$h_1 = 18:00 \text{ hrs.}$	$h_1 = 00:00 \text{ hrs.}$
$h_2 = 18:15 \text{ hrs.}$	$h_2 = 00:15 \text{ hrs.}$
$h_3 = 18:30 \text{ s.}$	$h_3 = 00:30 \text{ s.}$
\vdots	\vdots
$H_{20} = 23:00 \text{ hrs.}$	$h_{96} = 23:45 \text{ hrs.}$

A continuación, se muestra un gráfico como resumen general del patrón de consumo de fin de semana (FDS), usando un análisis completo.

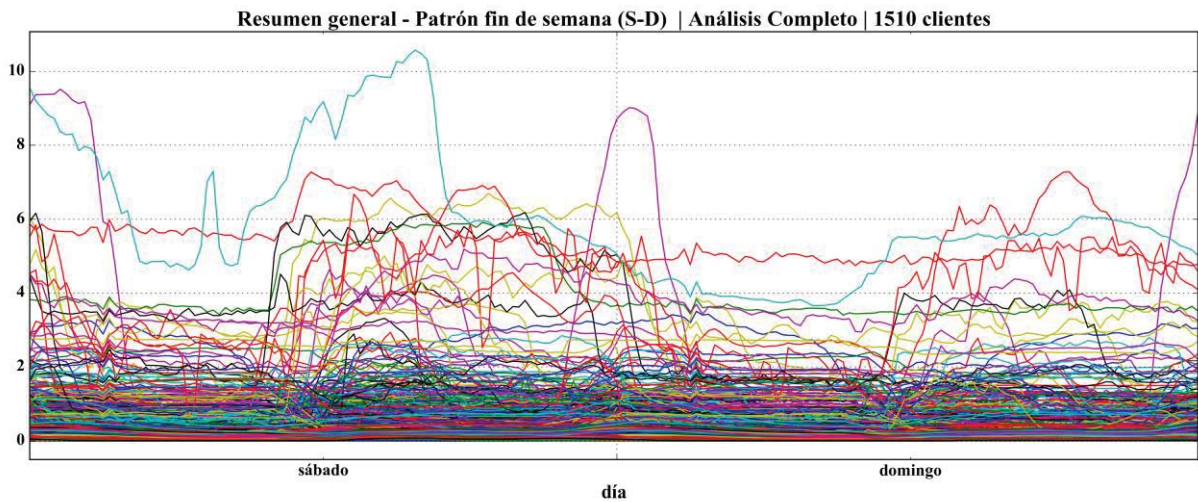


Figura 6.10 Resumen general - FDS | FULL

La figura 6.10 permite ver claramente que durante el día sábado existe un cliente con un importante registro de consumo, por sobre los 10 KWh. En tanto, dicho cliente no presenta los mismos niveles de consumo durante el día domingo. También se puede apreciar un segundo cliente que registra promedios de consumo por sobre los 9 KWh, y a diferencia del primero, éste repite su consumo en ambos días. Los demás clientes tienen promedios por debajo de los 8 KWh, y se concentran mayormente en rangos bajos de 0 a 1 KWh.

Tabla 6.5 Indicadores descriptivos - FDS | FULL | General

Promedio (KWh)	Desv. estándar (KWh)	Suma total (KWh)	Valor máximo (KWh)	Valor mínimo (KWh)
0,210	0,503	60609,084	10,573	0,000

La tabla 6.5 indica que el valor máximo registrado en este patrón de consumo es de 10,574 KWh, mientras que el valor mínimo es de 0 KWh. El promedio es cercano a cero, confirmando que la mayoría de clientes tiene un promedio de consumo entre 0 y 1 KWh. En cuanto al consumo acumulado, la tabla indica que durante el fin de semana se registra una suma total de consumo de 60.609,084 KWh.

En la figura a continuación, se muestra un gráfico con la curva promedio de consumo durante el fin de semana.

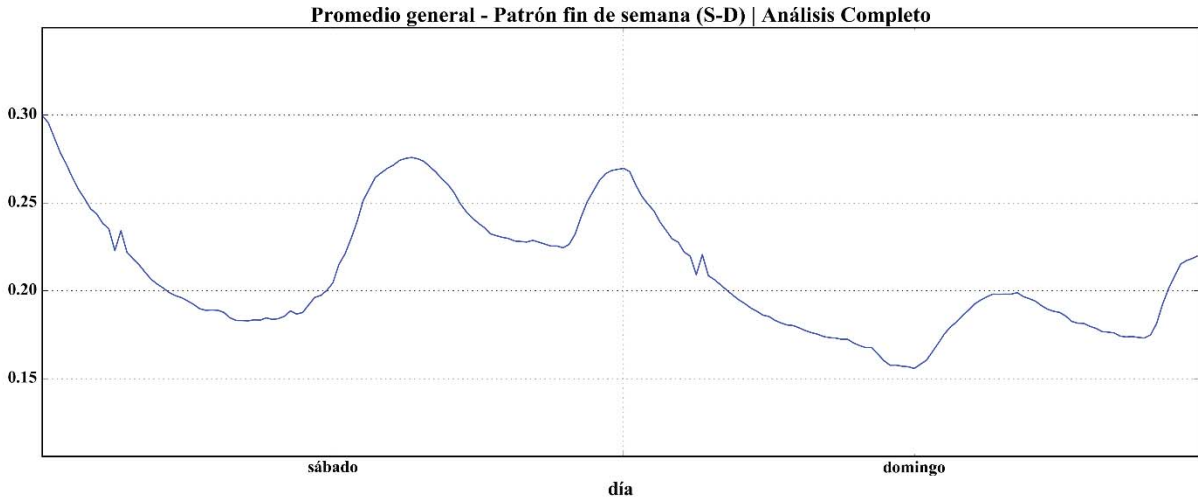


Figura 6.11 Promedio general - FDS | FULL

La figura 6.11 revela una gran diferencia en los niveles de consumo entre los días sábado y domingo. A partir del mediodía, el consumo aumenta en ambos días, sin embargo, durante el día domingo ese incremento es mucho menor.

La figura a continuación, muestra el resumen por estaciones del patrón de fin de semana, con análisis completo de horario.

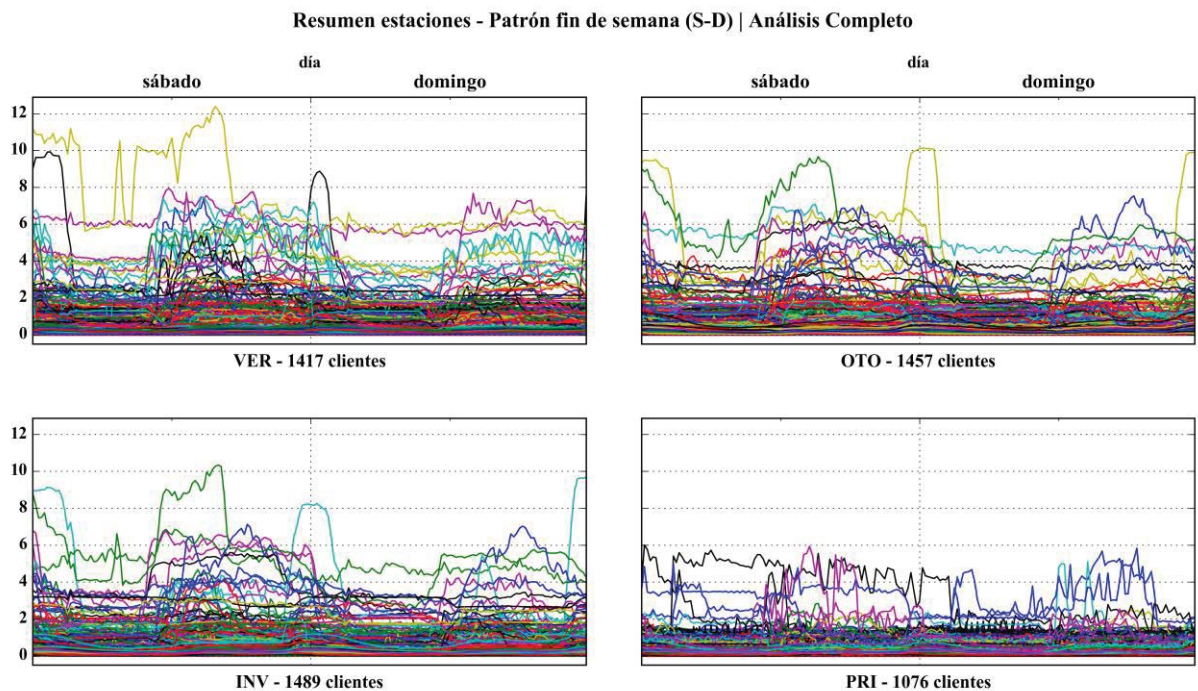


Figura 6.12 Resumen por estaciones - FDS | FULL

La figura 6.12 permite observar que, durante los meses de primavera, no existen clientes con registros superiores a 6 KWh. En tanto, durante las demás estaciones, se registran valores de consumo superiores a los 10 KWh. También, se observa una actividad mayor de consumo

durante el día sábado, a excepción de la primavera, en donde se ve un consumo equilibrado entre sábado y domingo. La tabla 6.6 muestra valores estadísticos descriptivos del consumo para los gráficos de la figura 6.12.

Tabla 6.6 Indicadores descriptivos - FDS | FULL | Estaciones

Estación	Promedio (KWh)	Desv. estándar (KWh)	Suma total (KWh)	Valor máximo (KWh)	Valor mínimo (KWh)
Verano	0,221	0,570	59906,150	12,388	0,000
Otoño	0,207	0,498	57554,964	10,123	0,000
Invierno	0,201	0,472	57203,287	10,350	0,000
Primavera	0,134	0,304	27538,037	6,010	0,000

Como muestra la tabla 6.6, y para confirmar el análisis previo, el promedio de consumo es más bajo en los meses de primavera, alcanzando los 0,134 KWh. La suma total de consumo es también considerablemente menor, dando un total de 26.538 KWh, en comparación a las demás estaciones que suman un consumo superior a los 57.000 KWh. El valor máximo de consumo en primavera también es muy inferior a las demás estaciones, teniendo un registro de 6,010 KWh; en tanto, durante el verano el valor máximo es de 12,388 KWh.

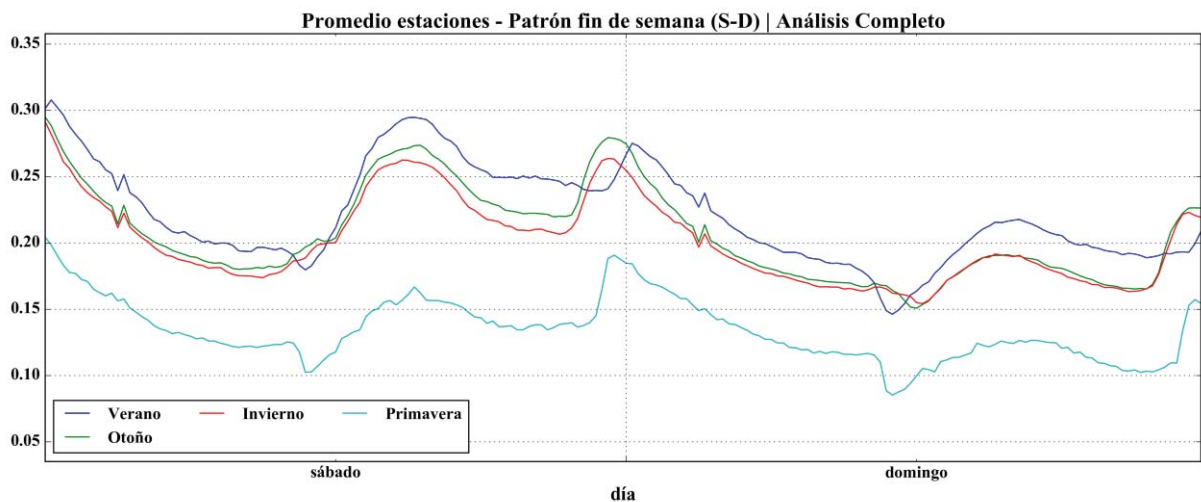


Figura 6.13 Promedio estaciones - FDS | FULL

Como muestra la figura 6.13, el comportamiento de consumo es similar en cada estación, pero varía considerablemente en el nivel registrado, especialmente durante los meses de primavera. Este valor inferior de consumo puede explicarse debido a que el análisis realizado sobre la estación de primavera considera una cifra mucho menor de clientes (1076), mientras que en las demás estaciones la cifra de clientes supera los 1400.

La baja cifra de clientes en primavera se explica con las restricciones de la elaboración misma del patrón de consumo. Se dijo anteriormente, que solo se considerarían aquellos días en que el cliente presentara una lectura total de consumo, es decir, 96 mediciones. En este caso, solamente 1076 clientes registran días completos de lecturas durante los fines de semana en los meses de primavera.

6.2.1.4 Patrón de consumo con transformada rápida de Fourier

A diferencia de las anteriores, esta representación no considera un promedio de las mediciones por cada cliente, sino utiliza la transformada rápida de Fourier sobre una serie temporal de consumo. Esta transformación permite descomponer la señal en sus componentes sinusoidales de diferentes frecuencias, en otras palabras, es considerada una técnica matemática para transformar el punto de vista de una señal desde la base de tiempo a la base frecuencia.

La transformada rápida de Fourier es una función reversible, es decir, puede aplicarse inversamente para obtener la serie original en el dominio del tiempo. Además, para efectos de procesamiento, el número de muestras de entrada debe ser potencia de dos; de esta manera, el tiempo de cálculo se reduce considerablemente.

En esta investigación, se utilizará un número de 2048 muestras para construir el patrón de consumo con análisis completo de horario. Para el análisis de horario punta, el número de muestras a utilizar será de 512. Ambas cantidades de muestras considerarán una serie temporal continua de 21 y 25 días respectivamente.

Para determinar la serie de tiempo a utilizar, se hizo un análisis gráfico del número de lecturas a lo largo del año 2015. A partir de esto, se logró identificar un periodo comprendido entre el 10 de abril y el 10 de mayo, que posee un número estable de registros de medidores superior a 1400, como se muestra en la figura 6.14. Esto permitirá integrar la mayor cantidad de medidores posible en la representación del patrón de Fourier.

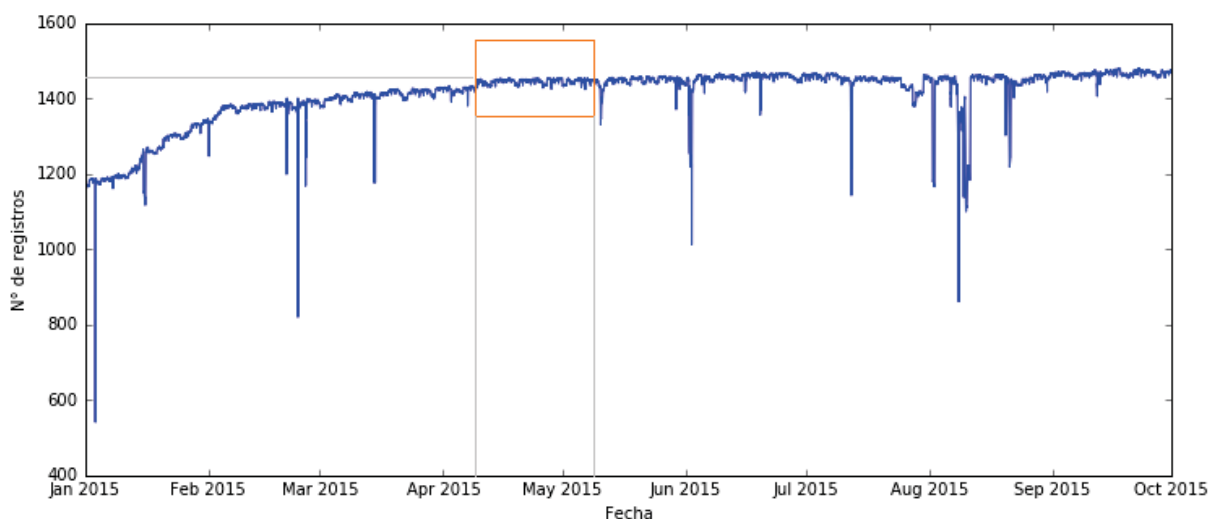


Figura 6.14 N° mediciones por fecha, año 2015 - Serie temporal patrón de Fourier

Para que la representación sea efectiva, es necesario considerar dos principales puntos. Primero, cada cliente debe tener la totalidad de registros de consumo posibles a lo largo de la serie temporal escogida. Segundo, la serie temporal debe comportarse de manera estacionaria para cada cliente.

Para cumplir el primer punto, se diseñó un algoritmo con el objetivo de filtrar aquellos clientes que no presenten la totalidad de registros posibles en el intervalo de tiempo, según la

cantidad de muestras utilizadas. De esta manera, el número de clientes que cumplen con la restricción se redujo a 1083.

En cuanto al segundo punto, se realizó la comprobación de estacionariedad de la serie mediante una prueba llamada Dicker-Fuller aumentado. Ésta prueba es utilizada comúnmente en estadística y econometría; su propósito es evaluar la hipótesis nula de que la serie posee una raíz unitaria [31], característica de los procesos que evolucionan en el tiempo, y que dificulta la aplicación de modelos predictivos. Cuando se detecta una raíz unitaria, entonces se dice que la serie de tiempo se comporta de manera no estacionaria. Por el contrario, de no hallarse raíz unitaria, se dice que la serie se comporta de manera estacionaria.

Para rechazar la hipótesis nula, es necesario que el valor estadístico del test sea menor a los puntos críticos evaluados en distintos niveles de confianza. Al aplicar la prueba, se obtuvieron los siguientes resultados.

Tabla 6.7 Resultados test Dickey-Fuller - Serie temporal patrón de Fourier

Resultados Dickey-Fuller Test	
Test estadístico	-9,547
Valor p	0
# Lags usados	28
Nº observaciones	2.947
Valor crítico (1%)	-2,567
Valor crítico (5%)	-2,863
Valor crítico (10%)	-3,433

Como se observa en la tabla 6.7, el valor del test estadístico es menor a los valores críticos evaluados. Por consiguiente, se afirma que la serie de tiempo analizada es de carácter estacionaria. Por ahora, se asumirá que cada cliente también se comportará de manera estacionaria en ese periodo.

Finalmente, el patrón de Fourier para un cliente ‘S’ queda expresado de la siguiente manera:

$$patrón_fourier^S [N] = [A_{F_1}, A_{F_2}, A_{F_3}, \dots, A_{F_N}], \text{ donde}$$

N = número de muestras

A_{F_i} = amplitud de la señal en la i – ésima frecuencia

i = valor de la frecuencia = $\{0, 1, 2, \dots, N\}$

En la imagen a continuación se muestra un gráfico general del patrón de consumo basado en la transformada de Fourier (PF) para todos los clientes, con análisis completo de horario.

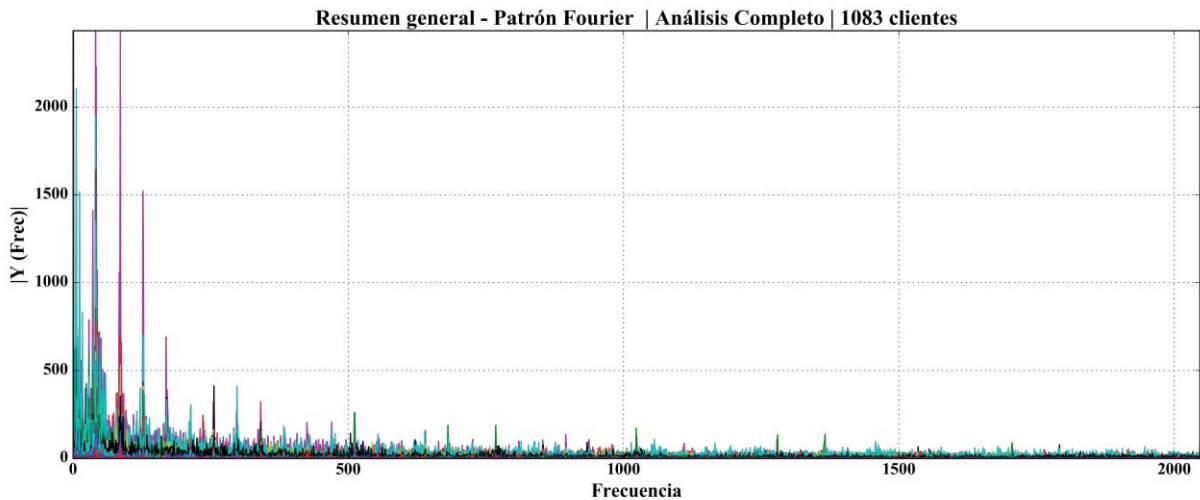


Figura 6.15 Resumen general - PF | FULL

La figura 6.15 muestra los valores de consumo transformados a un dominio de frecuencias. La amplitud de la señal aumenta en valores cercanos a cero, mientras que el valor disminuye acercándose a cero en frecuencias más altas. No es necesario realizar un análisis más acabado en el plano de frecuencias, dado que luego de aplicar el algoritmo de clustering sobre los datos, se analizarán los resultados recuperando las series originales en el plano temporal.

6.2.2 Aplicación

Para efectos de clustering, a excepción del patrón de Fourier, se utilizaron las bases de datos de 2014 y 2015. A pesar del déficit de registros en el año 2014, que fue comentado en la sección 5.1, se consideró relevante incluir dicha información para complementar el historial de consumo de los clientes.

Luego de generar los distintos patrones de consumo por cada cliente, ahora se procede a la aplicación del algoritmo de clustering para agrupar aquellos clientes que presenten un comportamiento similar. El algoritmo escogido es el K-means.

Con el fin de obtener un resultado más certero, se evaluó el número de clusters óptimo a utilizar mediante la métrica Silhouette. Para ello, se aplicó el algoritmo K-means varias veces, utilizando distinta cantidad de clústers. La cantidad de clusters a evaluar se definió entre 4 y 10, basándose en la cantidad de tarifas con mayor número de clientes en estudio. Luego, se evaluó cada resultado con Silhouette, y de acuerdo a ello se estableció el número de clusters óptimo en 10, para cada representación de consumo.

Finalmente, se aplica por última vez el algoritmo K-means con el número de clusters determinado en el paso anterior. Los parámetros utilizados en la implementación mediante la librería sklearn son los siguientes:

```
init='k-means++', n_init=10, max_iter=300, tol=0.0001, precompute_distances='auto', verbose=0,
random_state=None, copy_x=True, n_jobs=1, algorithm='auto'
```


Una descripción más completa y detallada de los parámetros de la función k-means puede encontrarse en [32].

6.3 Pronóstico mediante modelos predictivos

En esta sección, se mostrarán los experimentos realizados para la obtención de pronósticos de consumo eléctrico. Primero se abordará el análisis de la serie temporal; posteriormente, se mostrará la preparación de los datos; y finalmente, se detallará la aplicación de los modelos utilizados, especificando ajustes y parámetros establecidos.

6.3.1 Análisis de la serie

Para realizar los experimentos de pronóstico, es necesario primero escoger cuidadosamente la serie de tiempo a utilizar. De acuerdo a la información recabada en el capítulo 5, los datos correspondientes al año 2014 no reúnen las condiciones para construir una adecuada para el pronóstico. Sin embargo, los datos registrados durante el año 2015 sí permiten construir una serie de tiempo con más valor para el pronóstico.

La construcción de la serie temporal para pronóstico, consiste en sumar todos los valores de consumo en cada instante de tiempo. De esta manera, se tendrá un valor único que representará la suma total de consumo de todos los clientes. Por consiguiente, éste será el valor a pronosticar mediante los modelos predictivos.

Las características más importantes de la serie están dadas, en una primera instancia, por la cantidad de registros que se mantiene regular durante todo el periodo, como se vio en el capítulo 5, a excepción de algunos lapsos puntuales de tiempo. En segunda instancia, pero no menos importante, la serie aparenta ser estacionaria, lo cual es fundamental para aplicar correctamente los modelos de pronóstico.

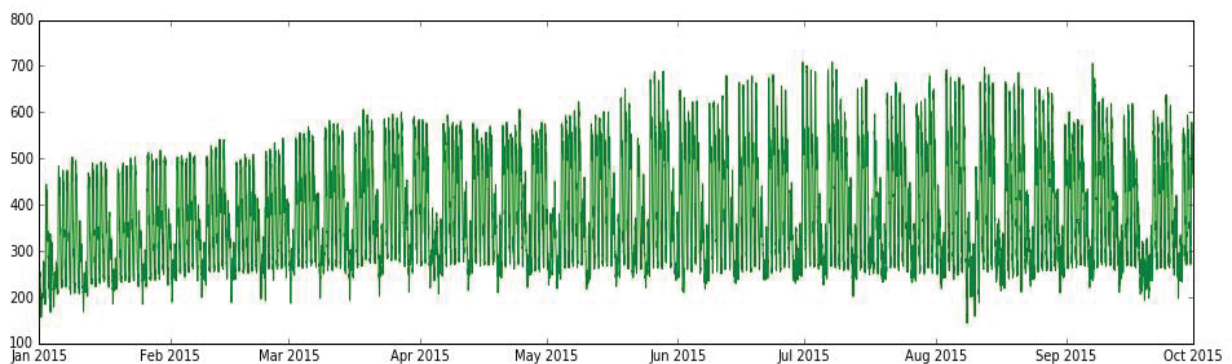


Figura 6.16 Serie temporal suma de consumo 2015

Para confirmar que la serie contemplada desde el 01-01-2015 hasta el 01-10-2015 sea estacionaria, fue sometida al “Dickey-Fuller Test”. De acuerdo a los criterios del test, la serie presenta características estacionarias, pues el Test estadístico tiene un valor inferior a los valores

críticos, lo que terminaría por confirmar su uso para la aplicación de modelos de pronóstico. Los resultados del test de Dickey-Fuller se muestran en la tabla 6.8 a continuación.

Tabla 6.8 Resultados test Dickey Fuller - Serie temporal 2015

Resultados Dickey-Fuller Test	
Test estadístico	-20,08
Valor p	0
# Lags usados	49
Nº observaciones	26.155
Valor crítico (1%)	-3,43
Valor crítico (5%)	-2,86
Valor crítico (10%)	-2,56

6.3.2 Preparación de los datos

Luego de comprobar que la serie es estacionaria, ahora corresponde preparar los datos para ingresarlos a los modelos de pronóstico. El primer paso es generar un nuevo set de datos a partir de la serie temporal. Este nuevo set de datos estará estructurado con lags o retardos, a fin de separar las variables dependientes de la variable independiente, que es la que se pretende pronosticar.

Por ejemplo, la primera fila de un set de datos, con un lag o retardo igual a 3, estaría representada de la siguiente manera: $X_1, X_2, X_3 | Y$; donde X_1, X_2, X_3 son variables independientes; el valor de la variable Y se calcula en base a los valores de X , por lo cual se le llama variable dependiente.

En esta investigación, la variable dependiente que se desea pronosticar, es un registro de consumo que representa un intervalo de 15 minutos; este valor depende de las lecturas de consumo previas. A modo de ejemplo, si se define el número de retardos en 3, se buscará predecir el consumo que habrá a las 16:45 hrs, basándose en los valores registrados a las 16:00, 16:15 y 16:30 hrs. respectivamente.

El segundo paso en la preparación de los datos, es definir un porcentaje para asignar como datos de entrenamiento para los modelos de pronóstico. En el presente trabajo, se definió como porcentaje del set de entrenamiento un 80%, mientras que para el Testing se deja el 20% restante.

Un último paso antes de comenzar con la evaluación de modelos de pronóstico, es normalizar los datos a una escala entre 0 y 1. Esto permitirá a los modelos evitar procesar números grandes y disminuir así el rendimiento del proceso o inferir negativamente en el resultado pronosticado.

6.3.3 Aplicación de modelos

Habiendo finalizado la preparación de los datos, se puede dar inicio a la aplicación de los modelos. En el presente estudio, se realizaron pruebas con 4 modelos distintos. Éstos son: Regresión lineal, Perceptrón multicapa de regresión, Bosque aleatorio y Regresión bayesiana.

Cada uno de estos modelos será entrenado con la misma serie de tiempo, que corresponde al tiempo comprendido entre el 01-01-2015 hasta el 01-10-2015. Como se dijo ya con anterioridad, se utilizará un 80% de los datos para entrenar cada modelo, mientras que el otro 20% será utilizado para realizar y medir el pronóstico. A continuación, se indican los parámetros que fueron utilizados para cada modelo con la librería sklearn.

Regresión lineal: fit_intercept=True, normalize=False, copy_X=True, n_jobs=1

Regresión bayesiana: n_iter=300, tol=0.001, alpha_1=1e-06, alpha_2=1e-06, lambda_1=1e-06, lambda_2=1e-06, compute_score=False, fit_intercept=True, normalize=False, copy_X=True, verbose=False

Bosque aleatorio: n_estimators=10, criterion='mse', max_depth=None, min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_features='auto', max_leaf_nodes=None, min_impurity_split=1e-07, bootstrap=True, oob_score=False, n_jobs=1, random_state=None, verbose=0, warm_start=False

Perceptrón multicapa: layers= [Layer ("Sigmoid", units=30), Layer("Linear")],

learning_rate=0.001, n_iter=40

En la aplicación del modelo Perceptrón multicapa, se utilizó un número de 3 retardos o lags. Para los demás modelos de pronóstico se usó un número de 96 lags. El número de lags se definió evaluando cada modelo y ensayando con distintos valores de retardos hasta lograr el mejor desempeño.

En [33] se puede hallar una descripción completa de los parámetros y retornos de cada función o modelo de regresión.

7 Resultados

En este capítulo se muestran los resultados obtenidos tanto en el descubrimiento de patrones utilizando el algoritmo K-means, como en el pronóstico de consumo usando diversos modelos predictivos de regresión.

7.1 Clustering

En esta sección se analizarán los resultados obtenidos en la aplicación de clustering. Como resumen, por cada patrón de consumo sólo se mostrarán 8 clusters representativos; 4 de ellos con análisis completo de horario, y los 4 restantes con análisis de horario punta, todos ellos obtenidos del resumen general de clientes. Los clusters correspondientes a los resúmenes de estaciones, así como la totalidad de los resultados, se encuentran en el Anexo B de este proyecto.

7.1.1 Clusters patrón diario de consumo

A continuación, se muestran los clusters representativos del patrón diario de consumo, obtenidos del resumen general de clientes. En primer lugar, se muestran los clusters para el análisis completo de horario, posteriormente, los clusters para el análisis de horario punta.

7.1.1.1 Clusters representativos - Análisis completo de horario

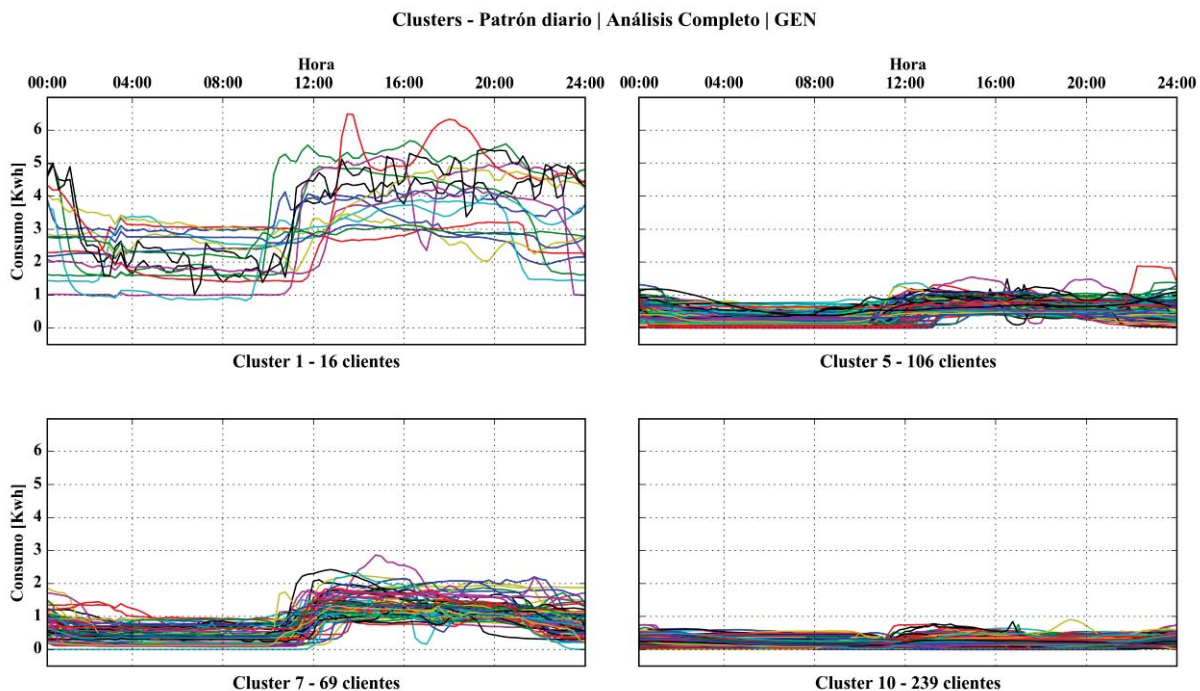


Figura 7.1 Clusters representativos - PD | FULL | General

En la figura 7.1 se muestran 4 tipos de clientes con comportamientos distintos en cuanto a los niveles de consumo. En el cluster 1 se tienen 16 clientes cuyo consumo está casi siempre sobre 1 KWh, y alcanza sus niveles más altos poco antes de las 12:00 hrs en adelante. Los demás presentan niveles de consumo mucho más bajos, bordeando los 2 KWh desde el mediodía, en el caso del cluster 5 y 7. Por último el cluster 10 no supera el consumo de 1 KWh en ningún horario del día.

Tabla 7.1 Indicadores descriptivos, Clusters representativos - PD | FULL | General

Cluster	Suma consumo (KWh)	Valor máximo (KWh)	Valor mínimo (KWh)	Promedio (KWh)	Porcentaje medidores (%)	Porcentaje consumo (%)
1	4723,820	6,492	0,835	3,075	1,052	12,265
5	4858,884	1,881	0,000	0,477	6,969	12,615
7	5570,996	2,865	0,003	0,841	4,536	14,464
10	4870,556	0,904	0,000	0,212	15,713	12,646

En la tabla 7.1, se puede observar que, pese a ser un grupo reducido de clientes, cada cluster representa más del 12% del consumo total. En el caso del cluster 1, los clientes representan solo el 1,052 % del total de medidores, pero en porcentaje de consumo, significan un 12,265% del total. Para ratificar lo visto en la figura 7.1, el cluster 1 presenta un promedio de consumo muy superior al resto, con un valor de 3,075 KWh.

Tabla 7.2 Tarifas, Clusters representativos - PD | FULL | General

Cluster	Tarifa										
	AT-3PPP	AT-4.3	BT-1A	BT-2PP	BT-2PPP	BT-3PP	BT-3PPP	BT-4.1	BT-4.3	Micro medición	Net Billing
1	1	0	0	0	0	4	6	0	4	0	0
5	0	0	42	0	0	6	40	0	13	2	0
7	0	0	8	0	0	11	34	0	14	0	0
10	3	0	124	11	7	9	37	2	17	25	0

La tabla 7.2, muestra los tipos de tarifa por cada cluster de clientes. El cluster 5 reparte los clientes entre las tarifas BT-1, BT-3 y BT-4; esto resulta llamativo debido a que los clientes BT-1 tienen un cobro normal por consumo en KWh, mientras que las otras tarifas tienen un cobro basado en demanda máxima de potencia. Esto ayuda a inferir que a pesar que los clientes tienen un comportamiento de consumo muy parecido, es posible que su cuenta a pagar sea muy diferente.

7.1.1.2 Cluster representativos - análisis de horario punta

A continuación, la figura 7.2 muestra a 4 grupos de clientes que se comportan de manera similar durante el horario punta, esto es, desde las 18:00 a las 23:00 hrs. El cluster más grande es el número 8, que está compuesto por 70 clientes; el cluster 5, a su vez, es el más pequeño con

10 clientes. Éste último, a pesar de ser el menor en número, agrupa a los clientes con mayores niveles de consumo durante el periodo de hora punta.

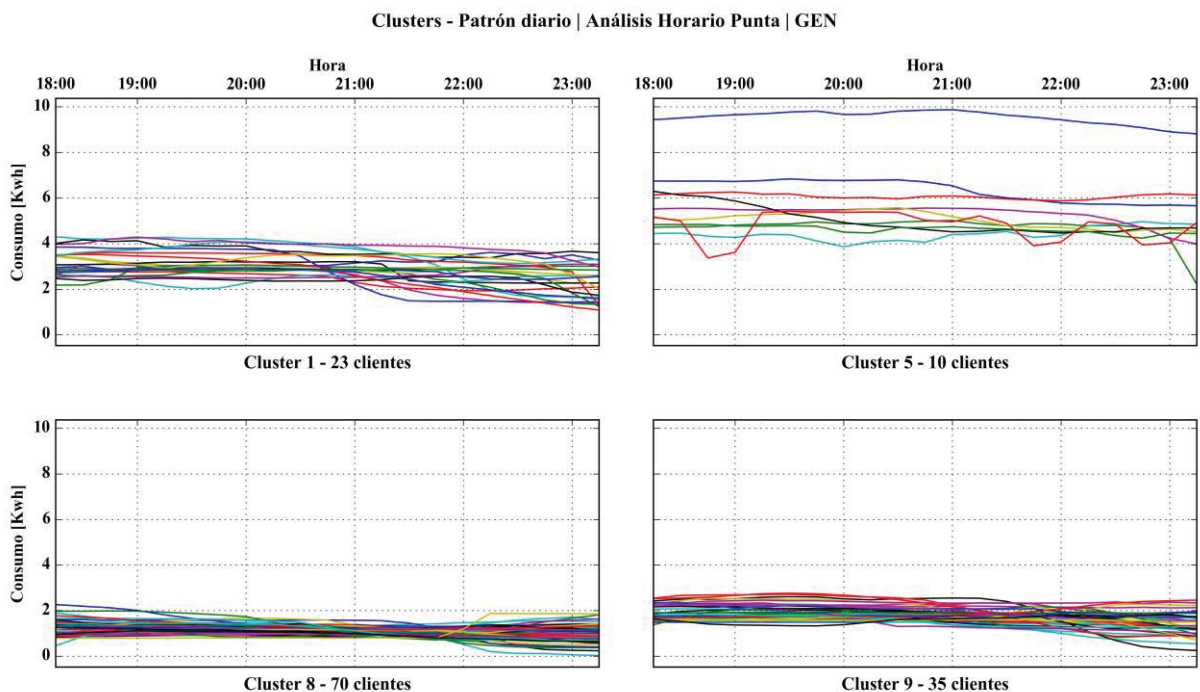


Figura 7.2 Clusters representativos - PD | HP | General

Tabla 7.3 Indicadores descriptivos, Clusters representativos - PD | HP | General

Cluster	Suma consumo (KWh)	Valor máximo (KWh)	Valor mínimo (KWh)	Promedio (KWh)	Porcentaje medidores (%)	Porcentaje consumo (%)
1	1396,399	4,298	1,092	2,891	1,512	14,441
5	1172,423	9,875	2,220	5,583	0,657	12,125
8	1655,719	2,266	0,028	1,126	4,602	17,123
9	1316,087	2,766	0,252	1,791	2,301	13,611

La tabla 7.3 muestra que, en términos de promedios, los clusters 8 y 9 son similares, aunque el último registra un nivel de consumo levemente superior. En cambio, los clusters 1 y 5 tienen un promedio más elevado que los anteriores, y además difieren entre sí. El cluster 1, que cuenta con 23 clientes, tiene un consumo promedio de 2,891 KWh; en tanto, el cluster 5 presenta un consumo promedio de 5,583 KWh, y los clientes en él alcanzan valores hasta de 9,875 KWh cerca de las 21:00 hrs.

Como se ve en la tabla 7.4 a continuación, el tipo de tarifa que más se repite en estos clusters son las de tipo BT-3, dando a entender que, en horario punta, son las tarifas que aportan más al monto total de consumo, pues los clientes que se encuentran agrupados en los clusters mostrados, reúnen entre ellos más del 57 % del consumo total en horario punta.

Tabla 7.4 Tarifas, Clusters representativos - PD | HP | General

Cluster	Tarifa										
	AT-3PPP	AT-4.3	BT-1A	BT-2PP	BT-2PPP	BT-3PP	BT-3PPP	BT-4.1	BT-4.3	Micro medición	Net Billing
1	0	0	0	0	0	3	15	0	5	0	0
5	1	0	0	0	0	6	2	0	0	0	0
8	0	0	13	0	0	13	30	0	12	0	0
9	0	0	2	0	0	6	20	0	7	0	0

7.1.2 Clusters patrón de consumo semana (lunes a viernes)

A continuación, se muestran los clusters representativos del patrón de consumo semanal (lunes a viernes), obtenidos del resumen general de clientes. Primero, se mostrarán los resultados con análisis completo de horario, y luego en horario punta.

7.1.2.1 Clusters representativos - análisis completo de horario

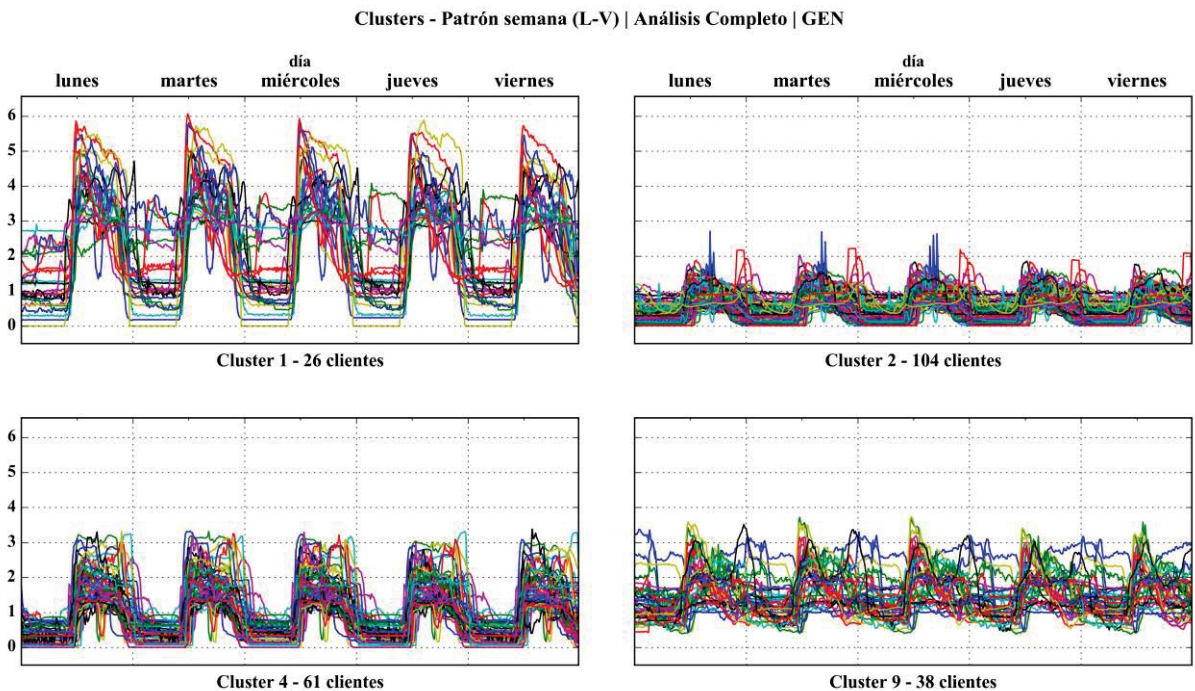


Figura 7.3 Clusters representativos - SEM | FULL | General

La figura 7.3 muestra a 4 grupos de clientes que aportan un gran porcentaje al consumo total que se registra entre lunes y viernes. El cluster 1 registra los valores más altos, y se compone de 26 clientes. Los demás clusters de clientes tienen una curva similar con variaciones en los niveles de consumo. A continuación, la tabla 7.5 muestra un detalle estadístico que permitirá diferenciar de mejor manera los perfiles de clientes.

Tabla 7.5 Indicadores descriptivos, Clusters representativos - SEM | FULL | General

Cluster	Suma consumo (KWh)	Valor máximo (KWh)	Valor mínimo (KWh)	Promedio (KWh)	Porcentaje medidores (%)	Porcentaje consumo (%)
1	29436,257	6,069	0,000	2,359	1,725	14,357
2	28632,001	2,709	0,006	0,574	6,901	13,964
4	28662,525	3,387	0,000	0,979	4,048	13,979
9	27581,470	3,723	0,379	1,512	2,522	13,452

Según la tabla 7.5, es posible distinguir de manera estadística el comportamiento de los clientes en cada cluster. Como se dijo anteriormente, estos cuatro grupos de clientes aportan en conjunto más del 55 % del consumo total registrado de lunes a viernes. Es importante considerar que, en cuanto al total de medidores en estudio, estos clientes sólo representan un poco más del 15 %.

En este sentido, el cluster 1 destaca alcanzando valores máximos de consumo por sobre los 6 KWh, y promediando el valor más alto entre los otros grupos de clientes. También, este cluster aporta un 14,357 % del consumo total, a pesar de ser un grupo reducido de clientes que sólo representa al 1,725 % del total de la población estudiada.

Tabla 7.6 Tarifas, Clusters representativos - SEM | FULL | General

Cluster	Tarifa										
	AT-3PPP	AT-4.3	BT-1A	BT-2PP	BT-2PPP	BT-3PP	BT-3PPP	BT-4.1	BT-4.3	Micro medición	Net Billing
1	0	0	0	0	0	3	17	0	6	0	0
2	0	0	41	0	0	10	38	0	11	1	0
4	0	0	5	0	0	5	36	0	13	0	0
9	0	0	6	0	0	8	18	0	6	0	0

La tabla 7.6 muestra que el cluster 1, el más representativo en términos de consumo, se compone en su totalidad por clientes de tarifas BT-3 y BT-4, cuyo cobro se basa en demanda de potencia. El resto también se compone en su mayoría por clientes con estos tipos de tarifa, a excepción del cluster 2. En este último cluster existe una importante cantidad de clientes con tarifa BT-1, lo cual resulta llamativo, como se dijo antes, por la marcada diferencia que existe en el sistema de cobros para estos distintos tipos de tarifa, con respecto a tarifas BT-3 y BT-4.

7.1.2.2 Análisis de horario punta

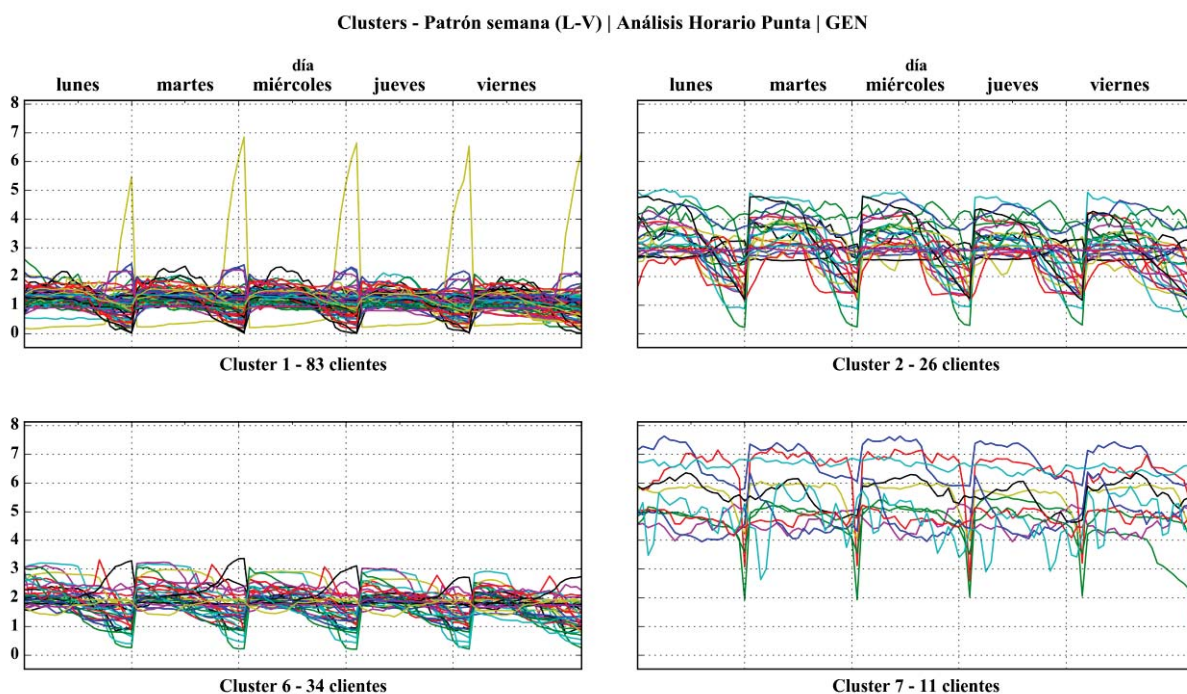


Figura 7.4 Clusters representativos - SEM | HP | General

La figura 7.4 muestra los clusters más representativos entre los días lunes a viernes, pero en horario punta. Se observa en el cluster 1, un cliente con valores altos de consumo, que se aleja evidentemente de los demás clientes del mismo cluster. En este caso, dicho cliente no fue clasificado de manera óptima, lo que se podría corregir generando un número mayor de clusters. En cuanto a los otros 3 clusters, el análisis gráfico muestra un agrupamiento bien logrado del consumo de los clientes.

Tabla 7.7 Indicadores descriptivos, Clusters representativos - SEM | HP | GEN

Cluster	Suma consumo (KWh)	Valor máximo (KWh)	Valor mínimo (KWh)	Promedio (KWh)	Porcentaje medidores (%)	Porcentaje consumo (%)
1	10307,922	6,858	0,020	1,183	5,508	19,385
2	8233,078	5,043	0,224	3,016	1,725	15,483
6	6599,787	3,372	0,208	1,849	2,256	12,412
7	6233,110	7,634	0,998	5,397	0,730	11,722

Como muestra la tabla 7.7, este conjunto de clusters representa casi un 60% del consumo total durante la semana en horario punta. El valor más alto de consumo se registra en el cluster 7, alcanzando los 7,634 KWh. A su vez, el valor más bajo se registra en el cluster 1, con 0,020 KWh.

Tabla 7.8 Indicadores descriptivos, Clusters representativos – SEM | HP | GEN

Cluster	Tarifa										
	AT-3PPP	AT-4.3	BT-1A	BT-2PP	BT-2PPP	BT-3PP	BT-3PPP	BT-4.1	BT-4.3	Micro medición	Net Billing
1	0	0	14	0	0	15	41	0	11	0	0
2	0	0	0	0	0	4	16	0	6	0	0
6	0	0	1	0	0	5	20	0	8	0	0
7	1	0	0	0	0	5	4	0	0	0	0

La tabla 7.8 muestra que los clientes representados por estos clusters, pertenecen en su mayoría a tarifas de demanda, como lo son BT-3 y BT-4. En el cluster 1, se registra un número no menor de clientes con tarifa normal BT-1.

7.1.3 Clusters patrón de consumo fin de semana (sábado-domingo)

En los siguientes apartados se mostrarán los clusters representativos para el patrón de consumo fin de semana, obtenidos desde los resúmenes generales de clientes. Primero, se mostrarán los clusters bajo un análisis completo de horario. Luego, se mostrarán los clusters bajo un análisis de horario punta.

7.1.3.1 Análisis completo de horario

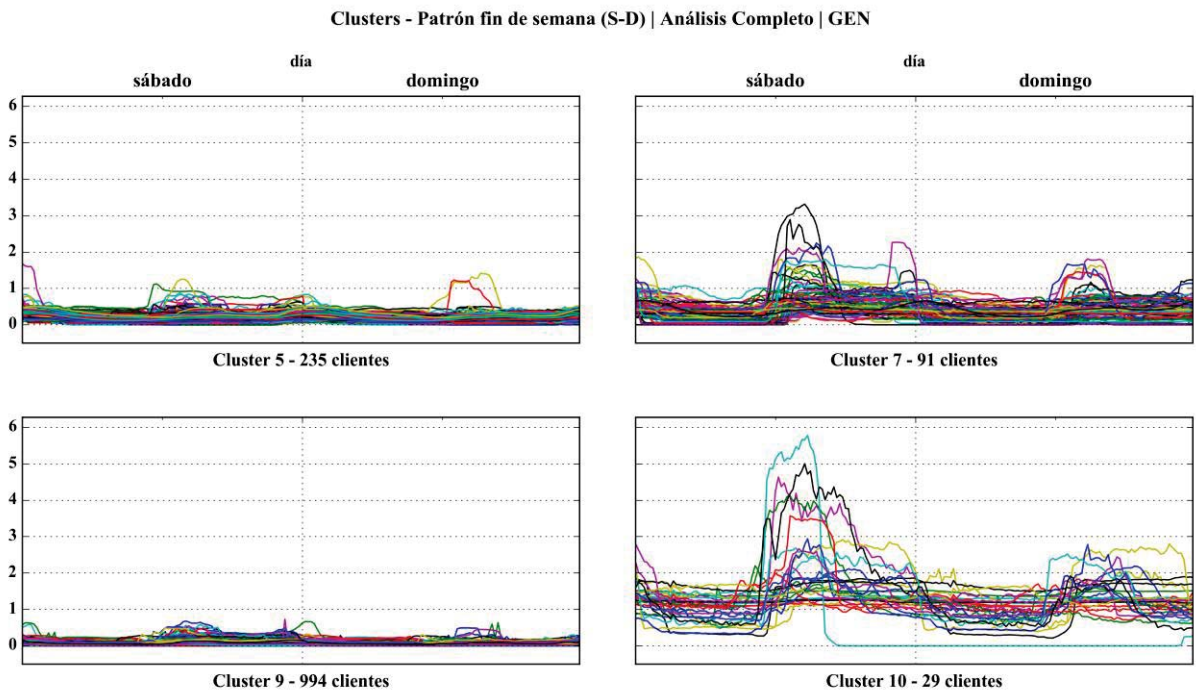


Figura 7.5 Clusters representativos - FDS | FULL | GEN

La figura 7.5, muestra los grupos de clientes más importantes en términos de consumo de energía eléctrica durante los fines de semana. El grupo mayoritario en número es el cluster 9, que agrupa 994 clientes, y sus niveles de consumo no superan 1 KWh. El gráfico también permite observar que los clusters 7 y 10 tienen un peak de consumo que se presenta a partir del mediodía del sábado, alcanzando valores superiores a 3 KWh y 5 KWh respectivamente.

Tabla 7.9 Indicadores descriptivos, Clusters representativos – FDS | FULL | GEN

Cluster	Suma consumo (KWh)	Valor máximo (KWh)	Valor mínimo (KWh)	Promedio (KWh)	Porcentaje medidores (%)	Porcentaje consumo (%)
5	8383,467	1,680	0,000	0,186	15,563	13,730
7	7370,644	3,317	0,000	0,422	6,026	12,071
9	7702,421	0,729	0,000	0,040	65,828	12,614
10	7162,475	5,781	0,003	1,286	1,921	11,730

La tabla 7.9 revela que el promedio más bajo de consumo es registrado por los clientes del cluster 10, alcanzando un valor de 0,040 KWh. No obstante, la cantidad de clientes en este cluster hacen que el porcentaje que aporta al consumo total sea de un 12,614 %. En el otro extremo, el cluster 10 presenta el promedio más alto con un valor de 1,286 KWh, registrando valores máximos de 5,781 KWh desde el mediodía del sábado.

Tabla 7.10 Tarifas, Clusters representativos – FDS | FULL | GEN

Cluster	Tarifa										
	AT-3PPP	AT-4.3	BT-1A	BT-2PP	BT-2PPP	BT-3PP	BT-3PPP	BT-4.1	BT-4.3	Micro medición	Net Billing
5	1	0	112	8	6	9	48	1	15	30	0
7	0	0	34	0	0	5	37	0	13	1	0
9	7	4	691	3	2	35	129	2	82	17	1
10	1	0	6	0	0	4	14	0	4	0	0

En la tabla 7.10, se observa que el tipo de tarifa mayoritario es BT-1 en los clusters 5 y 9. Estos números son congruentes con el nivel de consumo que se observa en dichos clusters, el cual fluctúa mayormente entre valores de 0 KWh y 1 KWh. En los clusters 7 y 10 la mayoría de clientes presenta tarifas de demanda como BT-3 y BT-4, no obstante, el número de clientes con tarifa BT-1 en dichos clusters no deja de ser menor.

7.1.3.2 Análisis de horario punta

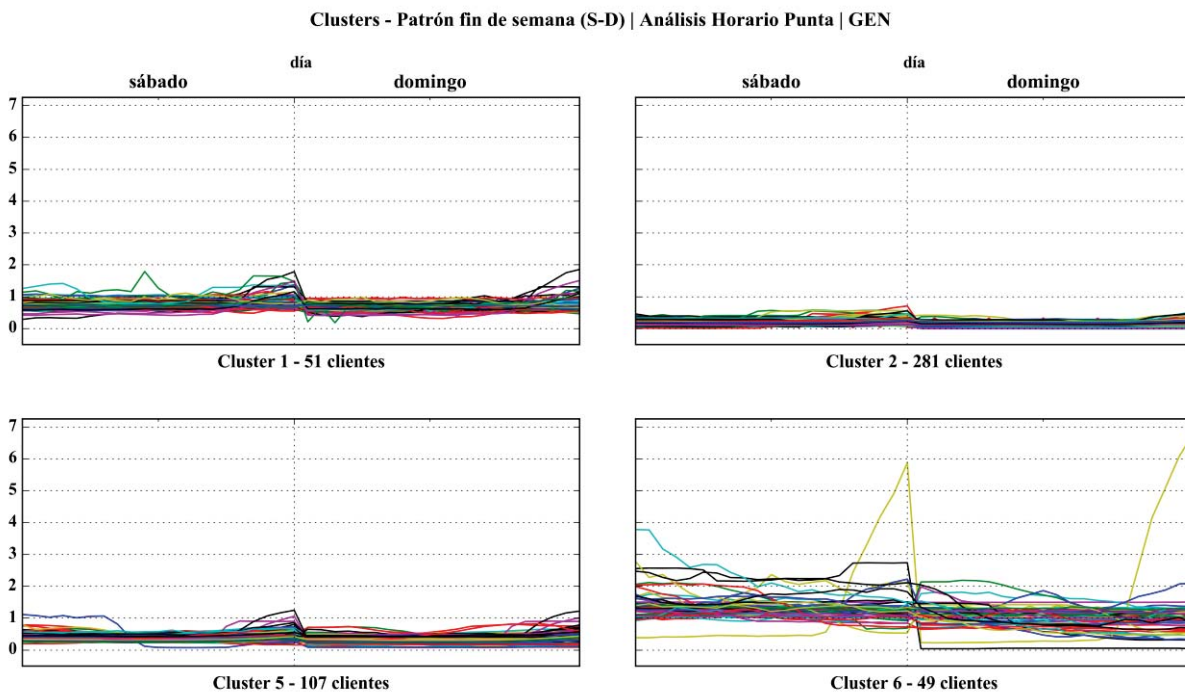


Figura 7.6 Clusters representativos - FDS | HP | GEN

Al ver la figura 7.6, se puede notar que los clusters representativos de fin de semana, en horario punta, no registran valores de consumo muy altos, concentrándose mayormente entre 0 KWh y 2 KWh. En el cluster 6 se observa un cliente con valores muy altos de consumo, los cuales difieren ampliamente del resto. Como se dijo anteriormente, este problema podría resolverse generando un número mayor de clusters. Sin embargo, este proyecto solo considera un máximo de 10 clusters por resumen de clientes.

Tabla 7.11 Indicadores descriptivos, Clusters representativos – FDS | HP | GEN

Cluster	Suma consumo (KWh)	Valor máximo (KWh)	Valor mínimo (KWh)	Promedio (KWh)	Porcentaje medidores (%)	Porcentaje consumo (%)
1	1595,991	1,859	0,186	0,745	3,377	12,065
2	1873,818	0,712	0,000	0,159	18,609	14,165
5	1669,419	1,248	0,075	0,371	7,086	12,620
6	2475,903	6,754	0,042	1,203	3,245	18,717

La tabla 7.11 muestra que el conjunto de clientes, que pertenecen a los clusters mostrados, representa más del 57 % del consumo total registrado en horario punta, durante los fines de semana. En tanto, sólo representan un 32,317 % del total de medidores en estudio.

Tabla 7.12 Tarifas, Clusters representativos – FDS | HP | GEN

Cluster	Tarifa										
	AT-3PPP	AT-4.3	BT-1A	BT-2PP	BT-2PPP	BT-3PP	BT-3PPP	BT-4.1	BT-4.3	Micro medición	Net Billing
1	0	0	16	0	0	12	16	0	6	1	0
2	2	0	148	5	6	13	57	1	18	24	0
5	1	0	31	1	0	8	38	0	16	9	0
6	0	0	11	0	0	8	20	0	9	0	0

La tabla 7.12 muestra que los tipos de tarifa con mayor presencia en los clusters mostrados son BT-1, BT-3 y BT-4. En el cluster 2, la tarifa BT-1 es mayoritaria, mientras que en los demás clusters se tiene una mayoría ligera de tarifas BT-3 y BT-4.

7.1.4 Clustering representación transformada de Fourier

En esta sección se mostrarán los clusters representativos para el patrón de Fourier, obtenidos desde los resúmenes generales de clientes. Los clusters serán analizados en el plano temporal, habiendo aplicado la función inversa de la transformada de Fourier a los resultados originales.

7.1.4.1 Análisis completo de horario

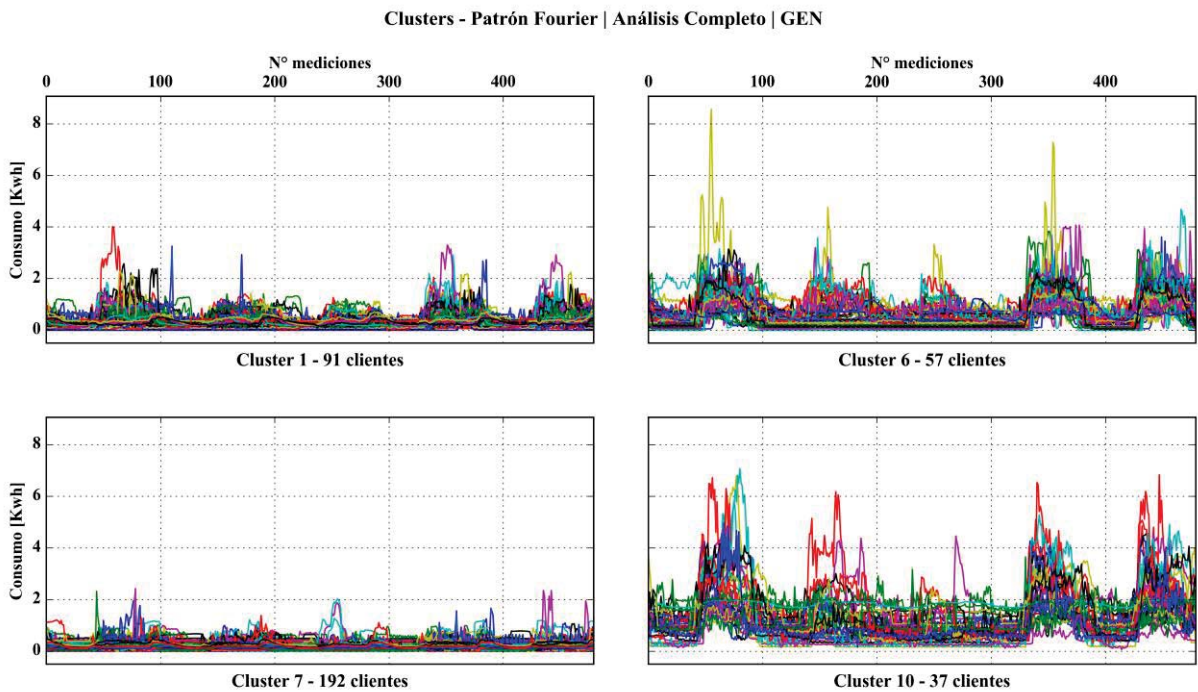


Figura 7.7 Clusters representativos - PF | FULL | GEN

En la figura 7.7, se observan los 4 clusters más representativos para el patrón de Fourier, con análisis completo de horario. Los gráficos muestran un total de 480 muestras, lo que equivale a los primeros 5 días de la serie temporal utilizada, con mediciones continuas en intervalos de 15 minutos. Al limitar el número de muestras, se logra visualizar de mejor manera las curvas de consumo de los clientes.

En los gráficos se puede observar que el cluster 7 se compone de 192 clientes con niveles bajos de consumo, cuyos valores oscilan mayormente entre 0 y 1 KWh. El resto de clusters agrupan menores cantidades de clientes, pero presentan un mayor nivel de consumo, con valores que oscilan mayormente entre 0 KWh y 4 KWh, alcanzando peaks por sobre los 6 KWh.

Tabla 7.13 Indicadores descriptivos, Clusters representativos – PF | FULL | GEN

Cluster	Suma consumo (KWh)	Valor máximo (KWh)	Valor mínimo (KWh)	Promedio (KWh)	Porcentaje medidores (%)	Porcentaje consumo (%)
1	63697,259	4,487	0,000	0,342	7,718	12,635
6	81065,078	8,569	0,000	0,694	4,835	16,080
7	62954,712	2,682	0,000	0,160	16,285	12,488
10	104218,538	8,374	0,001	1,375	3,138	20,673

En la tabla 7.13, se puede notar que los 4 clusters en conjunto representan un porcentaje cercano al 62 % del consumo total de la serie de tiempo analizada. Además, en cuanto a cantidad de clientes, representan cerca del 40 % de toda la población estudiada. También, se observa que el cluster 10 registra el mayor promedio de consumo, y a su vez aporta el porcentaje de consumo más amplio, con un 20,673 %. Cabe mencionar, además, que este cluster es el más pequeño en términos de cantidad de clientes, agrupando solo 37 clientes, que equivale a un 3,138 % del total.

Tabla 7.14 Tarifas, Clusters representativos – PF | FULL | GEN

Cluster	Tarifa										
	AT-3PPP	AT-4.3	BT-1A	BT-2PP	BT-2PPP	BT-3PP	BT-3PPP	BT-4.1	BT-4.3	Micro medición	Net Billing
1	0	0	43	1	1	3	28	1	9	4	0
6	0	0	13	0	0	5	29	0	7	1	0
7	4	0	114	7	6	9	20	1	12	19	0
10	0	0	6	0	0	6	19	0	6	0	0

La tabla 7.14 muestra los tipos de tarifa que poseen los clientes por cada cluster. Se puede observar que en los clusters 1 y 7, la tarifa que más se repite es BT-1. En tanto, para los clusters 6 y 10, la tarifa con más clientes es la BT-3.

7.1.4.2 Análisis horario punta

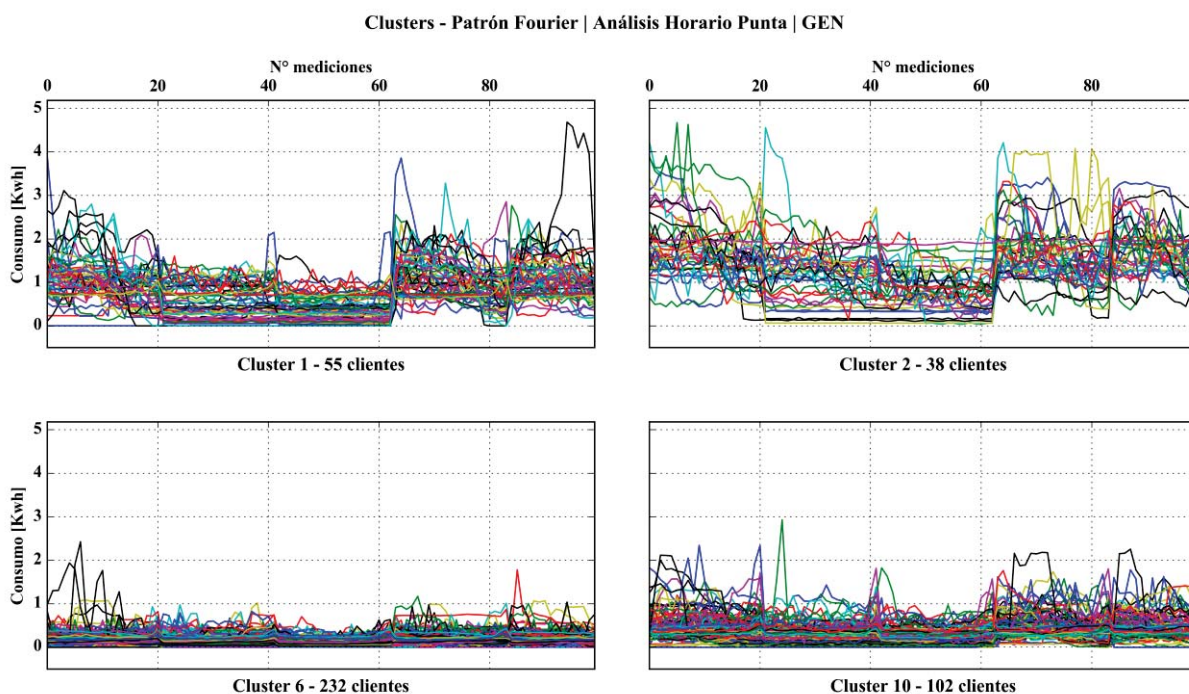


Figura 7.8 Clusters representativos - PF | HP | GEN

El gráfico 7.8 muestra los clusters más representativos para el patrón de Fourier, usando un análisis de horario punta. Los gráficos presentan un número de 100 muestras, que equivalen a los primeros 5 días de la serie temporal utilizada, con lecturas continuas en horario punta. Se puede ver que, tanto el cluster 1 como el 2, registran valores de consumo más altos que los demás. En cuanto a cantidad de clientes, el cluster 6 contiene el mayor número, con un total de 232; en cambio, el cluster 2 es el cluster más pequeño, agrupando solo 38 clientes.

Tabla 7.15 Indicadores descriptivos, Clusters representativos – PF | HP | GEN

Cluster	Suma consumo (KWh)	Valor máximo (KWh)	Valor mínimo (KWh)	Promedio (KWh)	Porcentaje medidores (%)	Porcentaje consumo (%)
1	22678,318	4,684	0,000	0,805	4,766	16,184
2	26567,578	4,674	0,000	1,366	3,293	18,960
6	19313,855	2,444	0,000	0,163	20,104	13,783
10	21633,680	3,770	0,000	0,414	8,839	15,439

En la tabla 7.15, se puede observar que, en términos de promedio de consumo, el cluster 2 registra el valor más elevado, que alcanza los 1,366 KWh. El valor máximo registrado es mayor en los clusters 1 y 2, alcanzando valores superiores a los 4,6 KWh. En cuanto al

porcentaje de consumo, el conjunto de clusters representa a un 64,375 % del total. Mientras que, en porcentaje de medidores, el conjunto de clusters representa un 37 % del total.

Tabla 7.16 Tarifas, Clusters representativos – PF |HP| GEN

Cluster	Tarifa										
	AT-3PPP	AT-4.3	BT-1A	BT-2PP	BT-2PPP	BT-3PP	BT-3PPP	BT-4.1	BT-4.3	Micro medición	Net Billing
1	0	0	19	0	0	4	26	0	4	1	0
2	0	0	7	0	0	8	16	0	7	0	0
6	5	0	140	5	6	11	25	1	16	20	1
10	1	0	46	0	1	8	31	1	9	4	0

La tabla 7.16, muestra un predominio de las tarifas BT-1 en los clusters 6 y 10. Mientras que en los clusters 1 y 2, predominan levemente las tarifas BT-3.

7.2 Pronósticos

Se exponen aquí los resultados de los modelos de pronóstico aplicados. Se muestran gráficos en donde la curva azul representa a los datos reales; la curva verde representa los datos pronosticados; y la curva negra corresponde al residuo o diferencia entre valor real y el pronosticado. Junto con el gráfico se muestra una tabla con los indicadores de desempeño.

7.2.1 Regresión lineal

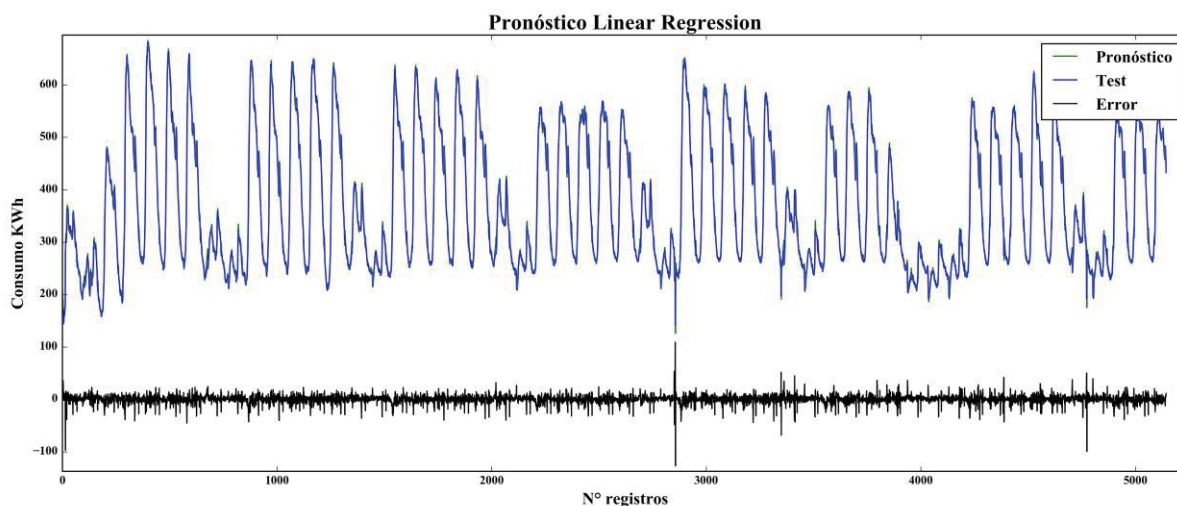


Figura 7.9 Pronóstico Regresión lineal

Tabla 7.17 Desempeño Regresión lineal

Regresión Lineal	
MSE	89,710
NMSE	0,006
MAE	6,250
MAPE	1,781 %
R ²	0,994
Training Time (s)	0,078

La tabla 7.17 muestra que el pronóstico con Regresión lineal arroja un excelente desempeño, con un MAPE del 1,781 %, y un coeficiente de determinación igual a 0,994. Además de esto, el modelo se entrena en un tiempo de 0,078 segundos, arrojando resultados casi de manera instantánea.

7.2.2 Regresión bayesiana

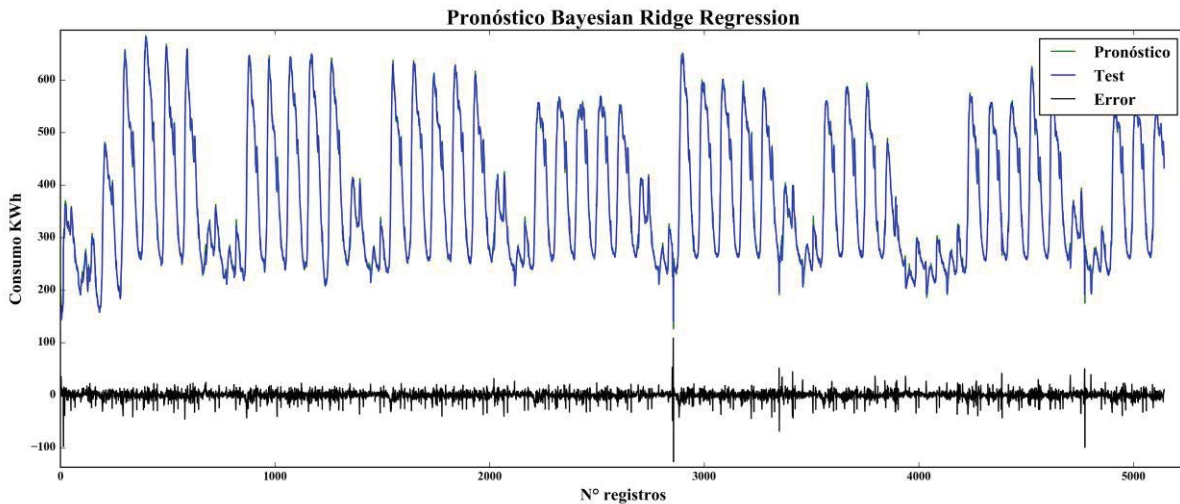


Figura 7.10 Pronóstico Regresión Bayesiana

Tabla 7.18 Desempeño Regresión Bayesiana

Regresión bayesiana	
MSE	89,749
NMSE	0,006
MAE	6,251
MAPE	1,781 %
R ²	0,994
Training Time (s)	0,219

En la tabla 7.18 se muestra el desempeño de pronóstico mediante el modelo de Regresión Bayesiana. Este modelo presenta un desempeño casi idéntico al de la regresión lineal, diferenciándose principalmente en el tiempo de entrenamiento, en donde alcanza un valor de 0,219 segundos.

7.2.3 Bosque aleatorio

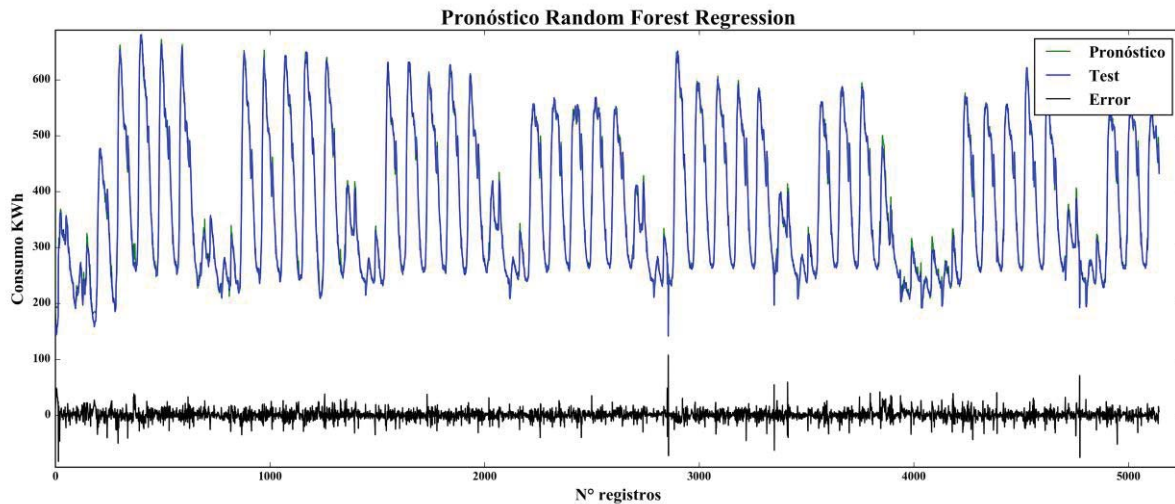


Figura 7.11 Pronóstico Bosque Aleatorio

Tabla 7.19 Desempeño Bosque aleatorio

Bosque aleatorio	
MSE	90,191
NMSE	0,006
MAE	6,463
MAPE	1,882 %
R ²	0,994
Training Time (s)	16,264

El modelo de regresión Bosque Aleatorio, presenta un desempeño un poco menor que los modelos anteriores. La tabla 7.19 muestra que alcanza un MAPE de 1,882 % y un tiempo de 16,264 segundos de entrenamiento. No obstante, presenta un coeficiente de determinación de 0,994.

7.2.4 Perceptrón multicapa

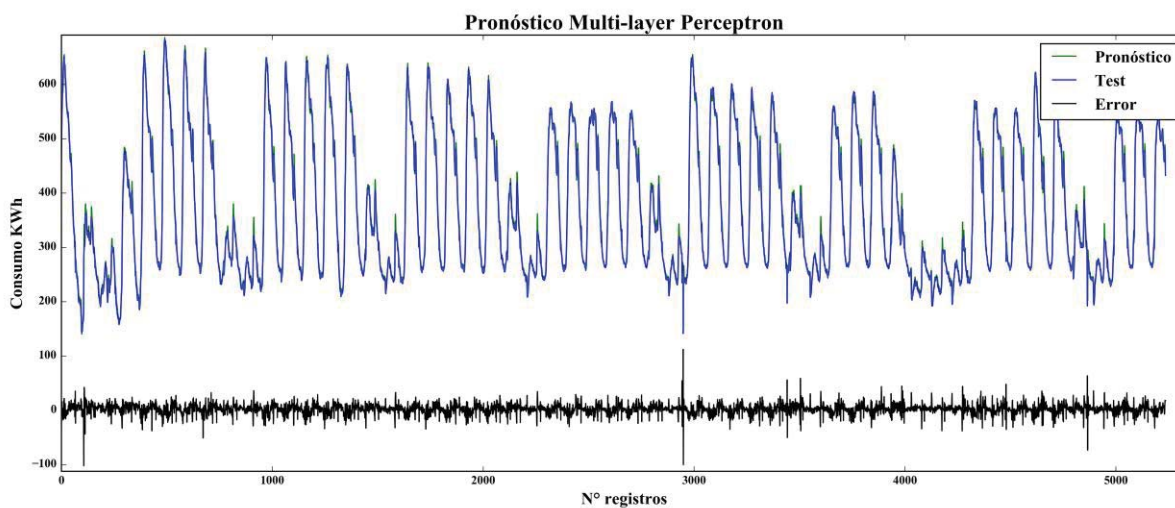


Figura 7.12 Pronóstico Perceptrón Multicapa

Tabla 7.20 Desempeño Perceptrón Multicapa

Perceptrón Multicapa	
MSE	141,351
NMSE	0,009
MAE	8,794
MAPE	2,389
R ²	0,991
Training Time (s)	669,870

El último modelo utilizado fue el Perceptrón multicapa, cuyo desempeño se muestra en la tabla 7.20. A pesar que sus resultados alcanzan un coeficiente de determinación del 0,991, el MAPE sube a un 2,389 %. En cuanto al tiempo de entrenamiento, este modelo fue el que más demoró en procesar, alcanzando una duración de 669,870 segundos.

8 Conclusiones

El reconocimiento de patrones y el pronóstico de consumo eléctrico es una medida que no se puede hacer esperar. La necesidad de elevar el ahorro y la eficiencia energética obliga en cierto modo a las empresas distribuidoras a invertir en la implementación de tecnología como lo son los medidores inteligentes. Es por ello, que hoy por hoy el sistema de medición inteligente está haciendo su ingreso a pasos agigantados al mercado o sector de la electricidad.

En el presente trabajo se abordó el descubrimiento de patrones de consumo eléctrico, a fin de conocer el comportamiento de los clientes. Esto permitiría a empresas distribuidoras crear recomendaciones o flexibilizar tarifas, buscando mejorar la eficiencia del consumo y la optimización de costos. Con este fin, se aplicó el algoritmo de clustering K-means, cuyos resultados permitieron identificar hasta 10 grupos de clientes con comportamiento similar, utilizando distintas representaciones de consumo. De esta forma, se logró identificar aquellos segmentos de clientes que tienen mayor impacto en la demanda, ya sea en horario punta, o sin restricción de horarios.

Otro ámbito abarcado en esta investigación ha sido el pronóstico del consumo eléctrico. Sin proyecciones del futuro las empresas no podrían planificar de buena manera ni tomar decisiones críticas a largo plazo reduciendo el riesgo. En la experimentación, se definió un intervalo de pronóstico mínimo, de sólo 15 minutos, debido a la escasa completitud de las muestras en estudio. No obstante, se pudo verificar el desempeño de distintos modelos de pronóstico, obteniendo resultados de alta precisión, alcanzando un valor de 0,994 en el coeficiente de determinación.

Al ser la medición inteligente una tecnología que toma cada vez más fuerza, se puede esperar que, en el mediano plazo, exista la cantidad y calidad de datos suficientes como para obtener resultados más fidedignos en el reconocimiento de patrones, y realizar pronósticos de consumo a intervalos más extensos de tiempo como semanas, meses, e incluso años.

9 Referencias

- [1] El sector eléctrico en Chile – Especial de Energía en Chile. Disponible vía web en http://www.economiaynegocios.cl/especiales/especial_energia/Sector_electrico1b.html Revisado por última vez el 29 de marzo de 2017.
- [2] Mercado de electricidad – Consumidores. Ministerio de Energía Gobierno de Chile. Disponible vía web en http://antiguo.minenergia.cl/minwww/opencms/03_Energias/Otros_Niveles/Electricidad/Mercado_electricidad/consumidores.html. Revisado por última vez el 29 de marzo de 2017.
- [3] Tarifas Suministro Eléctrico, Superintendencia de Electricidad y Combustibles. Disponible vía web en http://www.sec.cl/portal/page?_pageid=33,3417541,33_3421567&_dad=portal&_schema=PORTAL. Revisado por última vez el 29 de marzo de 2017.
- [4] Medición Inteligente (Smart Metering) – Empresas Eléctricas A.G. Disponible vía web en <http://www.electricas.cl/temas-estrategicos/medicion-inteligente/>. Revisado por última vez el 29 de marzo de 2017.
- [5] ECLARION, CITCEA, Gonzalo Carreño, Claudio Huepe Minoletti, Redes Inteligentes: Oportunidades de Desarrollo y Estrategia de implementación en Chile, Ministerio de Energía, Gobierno de Chile, octubre 2013. Disponible vía web en http://dataset.cne.cl/Energia_Abierta/Estudios/Minerg/14_Informe%20Final_Redetes%20inteligentes_%20oportunidades%20de%20desarrollo%20y%20estrategia%20de%20implementaci%C3%B3n%20en%20Chile.pdf. Revisado por última vez el 29 de marzo de 2017.
- [6] Medición Inteligente - Smart City Santiago. Disponible vía web en <http://www.smartcitysantiago.cl/medicion-inteligente>. Revisado por última vez el 29 de marzo de 2017.
- [7] Pablo Rámila y Hugh Rudnick, Medición Inteligente en Santiago de Chile, Metering International América Latina, edición 2009.
- [8] Diego Armando Herrera Chaves, Uso de las Redes de Sensores en actividades de Medición Inteligente para la promoción del uso Eficiente de la Energía, Universidad Nacional de Colombia, diciembre 2011.
- [9] Dr. V. Karpagam, et al., Interesting Pattern Generation in Smart Electricity Meter Data Using Subgroup Discovery Algorithm, International Journal of Computer Science and Mobile Computing, Vol. 4, Issue. 3, March 2015, pp. 693 – 701.
- [10] Joana M. Abreu, Francisco Camara Pereira, Paulo Ferrao, Using pattern recognition to identify habitual behavior in residential electricity consumption, Energy and Buildings, Vol. 49, June 2012, pp. 479-487, ELSEVIER.
- [11] Melissa Hart, Richard de Dear, Weather sensitivity in household appliance energy end-use, Energy and Buildings, Vol. 36, Issue 2, February 2004, pp. 161-174, ELSEVIER.

- [12] Daswin De Silva, et al., A Data Mining Framework for Electricity Consumption Analysis From Meter Data., IEEE Transactions on Industrial Informatics, Vol. 7, Issue: 3, Aug. 2011, pp. 399-407, IEEE.
- [13] Jason Lines, et al., Classification of Household Devices by Electricity Usage Profiles, Intelligent Data Engineering and Automated Learning - IDEAL 2011, pp. 403-412, Springer.
- [14] Alaa Alhamoud, et al., Extracting Human Behavior Patterns from Appliance-level Power Consumption Data, Wireless Sensor Networks pp. 52-67, European Conference on Wireless Sensor Networks 2015.
- [15] Wilhelm Kleiminger, et al., Occupancy Detection from Electricity Consumption Data, Conference: Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings, November 2013.
- [16] Clustering – EcuRed. Disponible vía web en <http://www.ecured.cu/Clustering>. Revisado por última vez el 29 de marzo de 2017.
- [17] Fernando Berzal, Clustering Jerárquico, Departamento de Ciencias de la Computación, Universidad de Granada. Disponible vía web en <http://elvex.ugr.es/idbis/dm/slides/42%20Clustering%20-%20Hierarchical.pdf>. Revisado por última vez el 29 de marzo de 2017.
- [18] Sklearn – Clustering Documentation. Disponible vía web en <http://scikit-learn.org/stable/modules/clustering.html#silhouette-coefficient>. Revisado por última vez el 29 de marzo de 2017.
- [19] A. Azadeh, et al., Forecasting electrical consumption by integration of Neural Network, time series and ANOVA, Applied Mathematics and Computation, Vol. 186, Issue 2, 15 March 2007, pp. 1753–1761, ELSEVIER.
- [20] A. Azadeh, et al., A simulated-based neural network algorithm for forecasting electrical energy consumption in Iran, Energy Policy, Vol. 36, Issue 7, July 2008, pp. 2637–2644, ELSEVIER.
- [21] A.S.Ahmad, et al., A review on applications of ANN and SVM for building electrical energy consumption forecasting, Renewable and Sustainable Energy Reviews, Vol. 33, May 2014, pp. 102–109.
- [22] El perceptron Multicapa (MLP), Herramientas en GNU/Linux para estudiantes universitarios, Capítulo 3: Tipos de redes neuronales. Disponible vía web en https://www.ibiblio.org/pub/linux/docs/LuCaS/Presentaciones/200304curso-glisa/redes_neuronales/curso-glisa-redes_neuronales-html/x105.html. Revisado por última vez el 29 de marzo de 2017.
- [23] José M^a Valls, Perceptrón multicapa, Tercer curso de Redes de Neuronas Artificiales, UC3M. Disponible vía web en <http://ocw.uc3m.es/ingenieria-informatica/redes-de-neuronas->

artificiales/transparencias/material-de-clase.-tema-3/view. Revisado por última vez el 29 de marzo de 2017.

[24] Leo Breiman, Random Forests, Machine Learning, Vol. 45, Issue 1, October 2001, pp. 5-32, Springer.

[25] Mark R. Segal, Machine Learning Benchmarks and Random Forest Regression, Division of Biostatistics, University of California, April 14 2003.

[26] Regresión lineal, EcuRed. Disponible vía web en https://www.ecured.cu/Regresi%C3%B3n_lineal. Revisado por última vez el 29 de marzo de 2017.

[27] Curtis Seubert, Definición de error cuadrático medio. Disponible vía web en http://www.ehowenespanol.com/definicion-error-cuadratico-medio-hechos_130449/. Revisado por última vez el 29 de marzo de 2017.

[28] ¿Qué es MAPE, MAD y MSD? Disponible vía web en: <http://support.minitab.com/es-mx/minitab/17/topic-library/modeling-statistics/time-series/time-series-models/what-are-mape-mad-and-msd/>. Revisado por última vez el 29 de marzo de 2017.

[29] María del Carmen García Centeno, Coeficiente de determinación. Disponible vía web en: <http://www.expansion.com/diccionario-economico/coeficiente-de-determinacion.html>. Revisado por última vez el 29 de marzo de 2017.

[30] ¿Cuál es el horario de punta vigente? Superintendencia de Electricidad y Combustibles, julio de 2013. Disponible vía web en: http://sec.custhelp.com/app/answers/detail/a_id/307/~/%C2%BFcu%C3%A1-es-el-horario-de-punta-vigente%3F. Revisado por última vez el 29 de marzo de 2017.

[31] Julio César Alonso C, Tutorial para Pruebas de Raíces Unitarias: Dickey-Fuller Aumentado y PhillipsPerron en EasyReg, Departamento de Economía, Universidad Icesi, diciembre de 2010. Disponible vía web en: <https://www.icesi.edu.co/departamentos/economia/images/at25.pdf>. Revisado por última vez el 29 de marzo de 2017.

[32] KMeans, Sklearn Documentation. Disponible vía web en: <http://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html#sklearn.cluster.KMeans>. Revisado por última vez el 31 de marzo de 2017.

[33] Supervised Learning, Sklearn Documentation. Disponible vía web en: http://scikit-learn.org/stable/supervised_learning.html#supervised-learning. Revisado por última vez el 31 de marzo de 2017.