

UNIVERSIDAD DE LOS ANDES
FACULTAD DE INGENIERÍA Y CIENCIAS APLICADAS



MODELOS DE PRONÓSTICO CON MACHINE LEARNING
PARA APOYAR LA TOMA DE DECISIONES

JUAN PABLO KARMY DIBÁN

TESIS PARA OPTAR AL GRADO DE
DOCTORADO EN CIENCIAS DE LA INGENIERÍA

PROFESOR GUÍA: JUAN PÉREZ, PhD
PROFESOR CO-GUÍA: CARLA VAIRETTI, PhD

SANTIAGO, AGOSTO DE 2023

Dedicatoria

Este trabajo está dedicado a mi familia, especialmente a mi señora, Laly, y a mis tres hijos que nacieron mientras cursaba este doctorado. A mis padres, mis hermanos, mis abuelos, a los Morgan Lavín (mi segunda familia), y a cada uno de mis amigos, que me acompañan cada día: José Manuel, Berni, Trini, Benja, Nico, Rai, David, Nicovas, Alex, Vito, Joaco, Santiago, Enzo, Ignacio, María José, Andrés, Seba, Poul, Gonzalo, y a cada uno que no nombro en esta lista.

En memoria del Huito y la Rita, me hubiese encantado que me acompañen este día.



We were here...

Resumen

El análisis de series de tiempo es una tarea fundamental dentro de la ciencia de datos. Ser capaces de almacenar, procesar, y entender los datos históricos nos permite utilizarlos como insumo para el proceso de toma de decisiones. Normalmente las series de tiempo poseen, ya sea implícita o explícitamente, características que permiten realizar su análisis en base a una estructura jerárquica. Esta tesis se centra en el estudio de estas series de tiempo, y en la posibilidad de obtener mejores estimaciones de las mismas, mediante el estudio de sus jerarquías. Este trabajo presenta cuatro investigaciones. Dos de ellas poseen una contribución científica aplicada, mediante la predicción de series de tiempo jerárquicas (HTS) en la industria del Travel Retail y en un Contact Center, respectivamente. Las otras dos investigaciones entregan una contribución metodológica, mediante la incorporación de la estructura jerárquica de las series en un algoritmo de Machine Learning, buscando generar nuevos modelos especializados en la estimación de este tipo de series temporales.

En la primera contribución aplicada, se generan estimaciones de venta de productos en la industria del Travel Retail mediante el enfoque clásico ε -SVR, pero adaptado a tres propuestas de algoritmos para su utilización en HTS. En este caso, nuestra metodología propuesta, SVR-BU, obtiene los mejores resultados, superando incluso a las técnicas clásicas ARIMA y Holt-Winters en términos de *mean absolute percentage error* (MAPE). En la segunda contribución aplicada, se realiza la estimación del volumen de llamadas entrantes en un Contact Center mediante los mismos algoritmos propuestos en el trabajo anterior, y nuevamente nuestra metodología es la que obtiene los mejores desempeños en términos de MAPE. En este caso, la contribución corresponde a la utilización de estos resultados en un modelo de optimización estocástica de *staffing* propuesto.

Las dos contribuciones metodológicas corresponden a nuestras propuestas θ -SVR y KAT-SVR. Ambos métodos pensados como una extensión de ε -SVR para HTS. La idea es agregar información a través de los niveles jerárquicos, previniendo que las series de los niveles inferiores se desvíen mucho de las series de los niveles superiores. Buscamos construir un modelo capaz de lidiar con el ruido intrínseco de los niveles inferiores. La gran diferencia entre ambas propuestas radica en que KAT-SVR incorpora métodos de kernel a los modelos previamente presentados, con el objetivo de otorgarles mayor flexibilidad, mediante la posibilidad de modelar de mejor manera problemas más complejos, con características no lineales. Los modelos presentados en ambos trabajos, θ -SVR y KAT-SVR, obtienen los mejores desempeños en términos de MAPE al ser comparados con ε -SVR tradicional, ARIMA, y Holt-Winters.

Abstract

Time series analysis is a fundamental task within data science. Being able to store, process and understand historical data allows us to use it as an input for decision-making. Normally, time series have, either implicitly or explicitly, characteristics that allow their analysis to be carried out based on a hierarchical structure. This Thesis focuses on the study of these time series, and on the possibility of obtaining better estimates of them by studying their hierarchies. This Thesis presents four investigations: two of them have an applied scientific contribution through prediction of hierarchical time series (HTS) in the Travel Retail industry and in a Contact Center, respectively. The other two investigations provide a methodological contribution, by incorporating the hierarchical structure of the series in a Machine Learning algorithm, seeking to generate new models specialized in estimating this type of time series.

In the first applied contribution, sales forecasting of products in the Travel Retail industry are generated using the classical ε -SVR approach, but adapted to three proposed algorithms for its use in HTS. In this case, our proposed SVR-BU methodology obtains the best results, outperforming the classical ARIMA and Holt-Winters techniques in terms of mean absolute percentage error (MAPE). In the second applied contribution, the estimation of the volume of incoming calls in a Contact Center is carried out using the same algorithms proposed in the previous work, and again our methodology is the one that obtains the best performance in terms of MAPE. In this case, the contribution corresponds to the use of these results in a proposed staffing stochastic optimization problem.

The two methodological contributions correspond to our proposals θ -SVR and KAT-SVR. Both methods consist on a novel ε -SVR approach for HTS. The idea is to pool information across the hierarchical levels, preventing the bottom-level series from deviating much with respect to the series at the upper levels. The motivation behind this approach is to build robust bottom-level models that can deal with the intrinsic noise present at this level. The biggest difference between both proposals is that KAT-SVR incorporates kernel methods to the previously presented models, with the aim of giving them greater flexibility, through the possibility of better modeling more complex problems, with non-linear patterns. The models presented in both works, θ -SVR and KAT-SVR, obtain the best performance in terms of MAPE when compared to traditional ε -SVR, ARIMA, or Holt-Winters.

Agradecimientos

En primer lugar, quiero agradecer a ANID y a la Universidad de los Andes, por las becas que me entregaron para financiar mis estudios de doctorado.

Además, esta tesis fue parcialmente financiada por el Instituto de Sistemas Complejos de Ingeniería, ISCI (ICM-FIC: P05-004-F, CONICYT: FB0816), por los proyectos Fondecyt 1160738, 1160894, 1200221 y 1201403, y por los programas de ANID PIA/BASAL AFB180003 y STIC-AMSUD 22-STIC-09.

También quiero agradecer a todos los investigadores que colaboraron en esta investigación, especialmente a mi tutor, Juan Pérez, a Sebastián Maldonado, y a Julio López.

Finalmente, quiero agradecer a las empresas Generation Research AB, CrossNet S.A., y Asociación Chilena de Seguridad (ACHS), por la confianza depositada en esta investigación, y por permitirnos utilizar sus datos para generar estudios de casos reales.

Índice general

Dedicatoria	II
Resumen	III
Abstract	IV
Agradecimientos	V
1. Introducción	1
1.1. Hipótesis	1
1.2. Objetivos	2
1.3. Contexto de Investigación	3
1.4. Estructura del informe	4
2. Revisión Bibliográfica	5
2.1. Proceso de descubrimiento de conocimiento en bases de datos	6
2.2. Series de tiempo	8
2.3. Series de tiempo jerárquicas	9
2.4. Enfoques predictivos para análisis de series de tiempo jerárquicas	11
2.5. Aprendizaje Supervisado	14
2.6. Support Vector Regression (SVR) para análisis de series de tiempo	15
2.7. Estimación de demanda y políticas de <i>staffing</i> en Contact Centers	18
3. Metodología	21
3.1. SVR para la predicción de series de tiempo jerárquicas	21
3.1.1. Algoritmo SVR-BU	22
3.1.2. Algoritmo SVR-TD	23
3.1.3. Algoritmo SVR-MO	23
3.1.4. Selección del modelo y procedimiento de validación	24
3.2. θ -SVR: Un enfoque novedoso de HTS basado en agregaciones de información	25
3.3. Incorporación de kernel al modelo propuesto	29

3.3.1.	Formulación del modelo basado en kernel	35
3.3.2.	Aplicación a series de tiempo jerárquicas	36
3.4.	Machine Learning como <i>input</i> para <i>staffing</i>	38
3.4.1.	Obtención de Intervalos de Confianza a partir de métodos de Machine Learning	39
3.4.2.	VARIABLES DE DECISIÓN Y OTROS PARÁMETROS	41
3.4.3.	Función Objetivo	42
3.4.4.	Configuraciones Experimentales	43
4.	Resultados y Análisis	44
4.1.	Predicción de HTS de venta de productos de la industria del Travel Retail vía SVR	45
4.2.	θ -SVR: mejoras al modelo SVR tradicional	47
4.3.	KAT-SVR: la inclusión de métodos kernel a θ -SVR	53
4.4.	Predicción de volumen de llamadas en un Contact Center	60
5.	Conclusiones	62
5.1.	Conclusiones teóricas y prácticas	62
5.2.	Investigación futura	65
	Nomenclatura	67
	Referencias	68
	Anexos	74
A.	Descripción de las bases de datos utilizadas	74
A.1.	Datos <i>benchmark</i>	74
A.2.	Datos Travel Retail	77
A.3.	Datos Contact Center	81
B.	Configuraciones Experimentales	83
C.	Desempeño de cada conjunto de datos	86
D.	Mejores parámetros para los modelos de ε-SVR	91

Índice de figuras

2.1. Proceso KDD.	7
2.2. Ejemplo de jerarquía.	10
3.1. Diferencia entre los métodos KAT-SVR_J^0 y KAT-SVR_J^{J-1}	37
3.2. Ejemplo de jerarquía para un Contact Center.	38
3.3. Resultado de los modelos de estimación de demanda.	39
A.1. Estructura jerárquica de la base de datos ‘Austourism’.	74
A.2. Estructura jerárquica de la base de datos ‘Departures’.	75
A.3. Estructura jerárquica de la base de datos ‘Htseg1’.	76
A.4. Estructura jerárquica de la base de datos ‘Htseg2’.	76
A.5. Estructura jerárquica de la base de datos ‘Melsyd’.	77
A.6. Serie de tiempo total para los datos de Generation Research.	78
A.7. Análisis descriptivo de la primera y segunda segmentación.	78
A.8. Análisis descriptivo de la tercera y cuarta segmentación.	79
A.9. Análisis descriptivo de la quinta y sexta segmentación.	80
A.10. Ventas totales para todas las compañías del Travel Retail en Europa.	81

Índice de cuadros

4.1.	Resumen de resultados de la primera investigación.	46
4.2.	Resultados detallados de θ -SVR para la base de datos ‘Htseg1’.	48
4.3.	Resultados detallados de θ -SVR para la base de datos ‘Htseg2’.	49
4.4.	Resultados detallados de θ -SVR para la base de datos ‘Austourism’.	49
4.5.	Resultados detallados de θ -SVR para la base de datos ‘Melsyd’.	50
4.6.	Resultados detallados de θ -SVR para la base de datos ‘Departures’.	51
4.7.	Resumen global de resultados para θ -SVR.	52
4.8.	Mejores parámetros para las variantes de la metodología θ -SVR.	53
4.9.	Configuración óptima de hiperparámetros para el método KAT-SVR $_j^0$	54
4.10.	Configuración óptima de hiperparámetros para el método KAT-SVR $_j^{J-1}$	54
4.11.	Resumen de los resultados globales para todas las bases de datos y métodos, utilizando MAPE como métrica de desempeño.	55
4.12.	Resultados detallados de KAT-SVR para la base de datos ‘Austourism’.	56
4.13.	Resultados detallados de KAT-SVR para la base de datos ‘Departures’.	57
4.14.	Resultados detallados de KAT-SVR para la base de datos ‘Htseg1’.	58
4.15.	Resultados detallados de KAT-SVR para la base de datos ‘Htseg2’.	58
4.16.	Resultados detallados de KAT-SVR para la base de datos ‘Melsyd’.	59
4.17.	Resumen de los resultados globales para todas las bases de datos y métodos, utilizando RMSE como métrica de desempeño.	60
4.18.	Resultados detallados para la base de datos de Contact Center.	61
A.1.	Cuadro resumen de los conjuntos de datos <i>benchmark</i>	77
A.2.	Todas las series de tiempo jerárquicas de Generation Research.	81
A.3.	Los 18 servicios considerados para la base de ACHS con sus respectiva cantidad de RP que nacen de ellos.	82
B.1.	Proceso de Validación.	84
B.2.	Configuraciones experimentales para cada base de datos.	84

C.1. Resultados detallados de ε -SVR para la base de datos ‘Htseg2’.	86
C.2. Resultados detallados de ε -SVR para la base de datos ‘Departures’.	87
C.3. Resultados detallados de ε -SVR para la base de datos ‘Melsyd’.	87
C.4. Resultados detallados de ε -SVR para la base de datos ‘Austourism’.	88
C.5. Resultados detallados de ε -SVR para la base de datos de Generation Research, bajo la estructura LuxuryC.	88
C.6. Resultados detallados de ε -SVR para la base de datos de Generation Research, bajo la estructura LuxuryS.	88
C.7. Resultados detallados de ε -SVR para la base de datos de Generation Research, bajo la estructura Company.	89
C.8. Resultados detallados de ε -SVR para la base de datos de Generation Research, bajo la estructura Market.	89
C.9. Resultados detallados de ε -SVR para la base de datos de Generation Research, bajo la estructura Other.	90
C.10. Resultados detallados de ε -SVR para la base de datos de Generation Research, bajo la estructura Gender.	90
D.1. Mejores parámetros para las variantes de ε -SVR.	92

Capítulo 1

Introducción

La generación de grandes volúmenes de datos y las inmensas posibilidades de análisis a partir de los mismos han provocado que la ciencia de datos se convierta en un aliado fundamental para los tomadores de decisiones, quienes ya no se conforman solo con el conocimiento básico de los negocios, sino que ven la posibilidad de ir más allá gracias a la gran disponibilidad de data que hay hoy en el mundo. Los datos son cada vez más fáciles de generar, gracias a nuevas tecnologías que surgen día tras día, y además son cada vez más baratos de almacenar, lo que ha llevado a las empresas a comprender los datos como uno de sus principales activos.

La posibilidad de utilizar los datos a nuestro favor, sin duda viene no sólo a respaldar la toma de decisiones, sino que también a encaminarla, y liderarla. El ser capaces no sólo de ver el pasado, sino que de predecir el futuro, se vuelve una tarea fundamental a la hora de explorar nuevas posibilidades para los tiempos que vienen.

Es por esto que mi principal foco de estudio corresponde a modelos que permitan predecir lo que va a suceder en el futuro, en base a lo evidenciado en el pasado. Con esto, no sólo se busca predecir la realidad, sino que entender que estas estimaciones no son mas que una manera de enriquecer las fuentes de información, para con ellas, tomar mejores decisiones. Ahí está el verdadero objetivo. En que el tomador de decisiones sea capaz de tomar cada vez más y mejores decisiones.

1.1. Hipótesis

Cuatro hipótesis fueron consideradas en el desarrollo de esta investigación:

- I. Los algoritmos de Machine Learning adaptados para realizar *business analytics*, como predicciones de series de tiempo, obtienen una precisión mayor y un mejor desempeño predictivo que las técnicas clásicas de minería de datos.
Se demostró empíricamente la eficacia de los métodos propuestos en términos de indicadores de desempeño como MAPE o RMSE.
- II. Al realizar predicciones aprovechando los niveles jerárquicos implícitos en las series de tiempo, se puede mejorar el desempeño predictivo.
Se demostró empíricamente para las series de tiempo con jerarquía implícita que mediante enfoques de bottom-up, top-down, middle-out, se mejora el desempeño predictivo en términos de indicadores de desempeño como MAPE o RMSE.
- III. Se puede adaptar modelos de SVR para series de tiempo jerárquicas para prevenir que las series de los niveles inferiores se desvíen mucho de las series superiores, mediante agrupaciones de información en todos los niveles.
Se demostró empíricamente que nuestro enfoque propuesto de SVR posee un desempeño predictivo superior a SVR clásico con series de tiempo jerárquicas, en términos de MAPE y RMSE.
- IV. Para *business analytics*, predicciones de series de tiempo jerárquicas mediante algoritmos de Machine Learning, combinadas con modelos de optimización, mejoran la interpretación de la solución, apoyando de mejor forma a reducir la incertidumbre y realizar mejores decisiones de negocio.
Considerando un problema de predicción del volumen de llamadas en un Contact Center, y posterior incorporación de un modelo de optimización para encontrar la dotación de agentes óptima necesaria, se entregan herramientas de recomendación para los tomadores de decisiones. Esta hipótesis se puede medir en base a la cantidad de reglas de decisión que se extraen del modelo.

1.2. Objetivos

El objetivo general de esta investigación es generar modelos predictivos con Machine Learning, y que además incluyan otro tipo de herramientas que apoyen de forma integral

al mejoramiento de la toma de decisiones de las empresas mediante la reducción de la incertidumbre.

Los objetivos específicos de esta investigación son:

- a) Analizar los principales problemas metodológicos presentes en aplicaciones de *business analytics* para predicciones, como el análisis de variables relevantes y *big-data*.
- b) Incorporación de investigación de operaciones en la toma de decisiones para distintos casos de estudio: predicción de ventas, predicción de flujo de llamadas de un Contact Center y análisis de dotación necesaria.
- c) Aplicar la metodología propuesta en la predicción de venta de productos en la industria del Travel Retail, y su extensión a otros problemas de predicción reales.
- d) Aplicar la metodología propuesta en el cálculo de la dotación de agentes necesarios para operar en un Contact Center de una empresa chilena.
- e) Estudiar el desempeño de diferentes estrategias de predicción de series de tiempo con SVR en aplicaciones de *business analytics*.
- f) Ampliar las ideas introducidas en trabajos anteriores de Karmy y Maldonado sobre predicciones de series de tiempo jerárquicas vía SVR (Karmy & Maldonado, 2019).

1.3. Contexto de Investigación

Como parte de esta investigación, tres trabajos fueron publicados en importantes revistas científicas, Q1 del *Journal Citation Report* (JCR), y un cuarto trabajo está en proceso de desarrollo.

- Karmy, J.P., Maldonado, S., Hierarchical Time Series Forecasting via Support Vector Regression in the European Travel Retail Industry. *Expert Systems with Applications*, 137, 59-73.

- Karmy, J.P., Maldonado, S., López, J., Simultaneous model construction and noise reduction for hierarchical time series via Support Vector Regression. *Knowledge-Based Systems*, 232, 17492.
- Karmy, J.P., Maldonado, S., López, J., Pooling information across levels in hierarchical time series forecasting via Kernel methods. *Expert Systems with Applications*, 213A, 118830.

Además, parte de esta investigación fue presentada en la versión N°30 de la Conferencia Europea en Investigación de Operaciones (Euro2019) en Dublín, Irlanda. También se presentó en la versión N°23 de la Conferencia de la Federación Internacional de Sociedades de Investigación de Operaciones (IFORS2023) en Santiago de Chile.

1.4. Estructura del informe

El presente informe se estructura como sigue:

- a) Capítulo 1: Introducción.
- b) Capítulo 2: Revisión bibliográfica general y específica para esta investigación.
- c) Capítulo 3: Metodología propuesta para esta investigación. Aquí se presentan tres propuestas de modelos predictivos: la adaptación de las máquinas de soporte vectorial para regresiones, ε -SVR, y su utilización para series de tiempo jerárquicas; una variante del mismo para mejorar la predicción de series de tiempo jerárquicas, θ -SVR y una tercera variante, que incluye máquinas kernel para entregar mayor flexibilidad al modelo al enfrentarse a problemas complejos, KAT-SVR. Finalmente se introduce un modelo de optimización estocástica, que recibe como *inputs* los resultados de nuestros modelos predictivos, para hacer gestión del *staffing* de una empresa.
- d) Capítulo 4: Presentación y análisis de los principales resultados de esta investigación, para cada uno de los métodos presentados en la sección anterior.
- e) Capítulo 5: Principales conclusiones de esta investigación, y propuestas de trabajo futuro para su continuidad.

Capítulo 2

Revisión Bibliográfica

En este capítulo se presenta una revisión bibliográfica de los diversos temas a tratar en esta tesis. Cada temática va acompañada de una revisión de trabajos en los que ha sido estudiado. El primer concepto que se introduce corresponde al del proceso de descubrimiento de conocimiento en bases de datos, un proceso que enmarca a un proyecto de *analytics*, definiendo cada una de sus etapas. En este marco se desarrollan todas las investigaciones presentadas posteriormente. Luego, se introduce el concepto de series de tiempo, para su posterior extensión a series de tiempo jerárquicas. A esto le sigue una sección que explica los diversos enfoques predictivos que se le pueden dar a las series de tiempo jerárquicas, en el marco de un proyecto de *business analytics*, y se mencionan algunos trabajos disponibles en la literatura sobre este tema. Luego, se presenta el concepto de Aprendizaje Supervisado y el cómo estas técnicas están teniendo cada vez mayor importancia, para luego explicar uno de estos métodos en detalle: el modelo Support Vector Regression (SVR) como una adaptación de las tradicionales máquinas de soporte vectorial (SVM) para el análisis de series de tiempo. Finalmente, se presenta una sección relacionada a cómo realizar estimación de demanda en un Contact Center, y se explican los métodos tradicionales que se usan para ello.

En los diversos trabajos de investigación que componen a esta tesis, se busca resolver una serie de problemáticas relacionadas a la predicción de series de tiempo jerárquicas (HTS) que no han sido tratadas con anterioridad en la literatura. Dos de las investigaciones presentadas tienen un carácter aplicado, mientras que los otros dos corresponden a un aporte metodológico. En particular, el primer trabajo considera la utilización de la técnica clásica de SVR como motor predictivo, presentando novedosos algoritmos que

lo complementan con los enfoques tradicionales de HTS, con una contribución científica aplicada. Un caso similar ocurre en el cuarto trabajo, en el que se genera un modelo de optimización estocástica que utiliza como *inputs* los resultados de modelos de Machine Learning, y busca generar una asignación de *staffing*, aplicado a un caso real. Por otro lado, el segundo y tercer trabajo presentan una contribución científica metodológica, mediante propuestas de nuevos algoritmos de Machine Learning basados en SVR; buscando mejorar los desempeños predictivos mediante la resolución de problemas intrínsecos de SVR con HTS.

2.1. Proceso de descubrimiento de conocimiento en bases de datos

El proceso de descubrimiento de conocimiento en bases de datos, conocido por las siglas KDD (Fayyad, 1996), derivadas del nombre en inglés *Knowledge Discovery in Databases*, corresponde a un proceso que enmarca las distintas etapas de un proyecto de *business analytics*. En este se consideran las etapas iniciales de recolección de la información desde las fuentes de datos, procesos de consolidación para generar las fuentes de datos consolidadas, y que posteriormente son sometidas a una fase de pre-procesamiento, en la que se ejecutan una serie de manipulaciones o transformaciones a los datos; de tal modo que puedan ser utilizados de manera adecuada como *input* de los diversos modelos matemáticos. Es recién en esta fase, una vez que ya se tiene la información procesada, que entran los modelos de minería de datos, con el objetivo de obtener resultados que nos permitan enriquecer las fuentes de información, y con ello, extraer conocimiento desde las bases de datos, permitiéndonos utilizar este conocimiento para reducir la incertidumbre del problema, y con esto, tomar nuevas y mejores decisiones. La idea es implementar las etapas de forma consecutiva. La figura 2.1 muestra un diagrama de las distintas etapas y procesos involucrados en esta metodología.

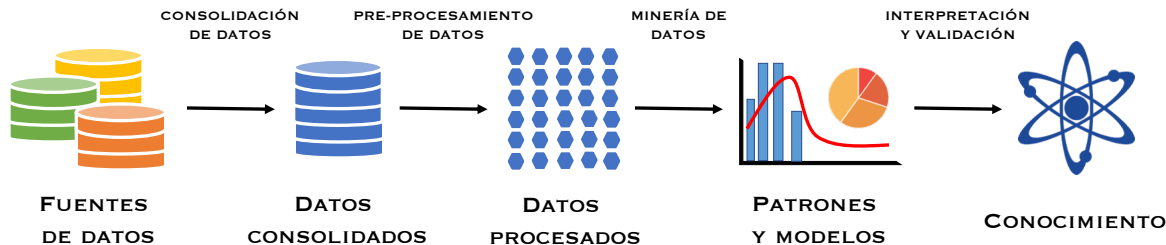


Figura 2.1: Proceso KDD.

Dentro de las etapas del proceso KDD, se considera la recolección de los datos. Se debe tener en cuenta que si se obtendrán subconjuntos de datos para el estudio, estos deben estar contruidos de forma adecuada en términos de tamaño, para que contengan todos los patrones y características necesarias para el estudio. Una vez consolidados los datos, estos son pre-procesados, para obtener datos útiles para los diversos modelos que se quieran construir. Si la data es de mala calidad, los resultados de los modelos también lo serán. En esta etapa es donde los datos son manipulados y transformados para hacerlos más accesibles (Pyle, 1999). Las principales tareas de la etapa de pre-procesamiento consisten en la limpieza de los datos, purificando inconsistencias y tratando de manera adecuada los valores perdidos que puedan existir. Por otro lado, se transforman las variables que requieran otro formato para su utilización en los modelos, y se eliminan los datos que no se necesitan.

Una vez que los datos fueron procesados, viene la fase de la construcción de los modelos y patrones. Aquí es donde se ejecutan las tareas, por ejemplo, de Machine Learning. A esta etapa se le denomina minería de datos o *analytics*. Una vez contruidos los modelos, viene una de las etapas más importantes del proceso: la interpretación y la validación de los modelos. Sin esta última etapa nos será imposible obtener buenas conclusiones desde los datos, y por ende, tomar buenas decisiones. La interpretación y validación consiste en medir el desempeño de los modelos, y la técnica a utilizar dependerá exclusivamente de la naturaleza del problema y de los datos. Una de las estrategias más utilizadas para validación corresponde a la división de los datos en varias muestras, para luego utilizar sólo algunas de ellas para la generación de los modelos, y el resto para la medición del error. Este tipo de estrategias permiten reducir el riesgo de sobreajuste de los modelos. Una de las particiones más conocidas se denomina *holdout* (Yang & Liu,

2002), que consiste en dividir los datos en tres subconjuntos, utilizando dos de ellos para el entrenamiento, y el tercero para la medición del error. Otra opción, muy famosa para problemas de clasificación, corresponde a la validación cruzada. Este tipo de validación fue diseñada buscando evitar pérdidas de información, y consiste en dividir los datos en k partes de tamaño similar, y realizar iteraciones, dejando uno de los conjuntos para validación, y el resto para el entrenamiento. Este tipo de técnica permite disminuir el sesgo que se podría generar con estos modelos.

2.2. Series de tiempo

Las series de tiempo corresponden a secuencias de objetos o_1, o_2, \dots, o_n , declarados en momentos sucesivos de tiempo t_1, t_2, \dots, t_n (Steffen, Sarda, Artz-Beielstein, Zaefferer & Strok, 2015). Con el objetivo de generalizar los patrones actuales para predecir valores futuros, las series de tiempo se pueden descomponer en diversas componentes, como lo son su tendencia, ciclo, componente estacional y componente aleatoria. (Hyndman & Athanasopoulos, 2013). Los dos métodos más conocidos para hacer estimaciones con series de tiempo corresponden al Modelo Autoregresivo Integrado de Media Móvil, mejor conocido como ARIMA, por sus siglas en inglés, y los modelos Holt-Winters (Hyndman & Athanasopoulos, 2013). El modelo ARIMA puede ser formulado como:

$$D_t = -(\Delta^d D_t - D_t) + \sum_{i=1}^p \alpha_i \cdot \Delta^d D_{t-i} + \sum_{j=1}^q \beta_j \cdot e_{t-j} + e_t, \quad (2.1)$$

donde α_i ($i = 1, \dots, p$) y β_j ($j = 1, \dots, q$) son los coeficientes relacionados a los procesos auto-regresivos y media móvil respectivamente, y e_t corresponde al término del error, también conocido como ruido. La expresión $\Delta^i D_t$ es la diferencia entre la observación t y su predecesor ($D_t - D_{t-1}$).

Por otro lado, Holt-Winters es una extensión de la suavización exponencial, que incluye dentro del modelamiento la tendencia y estacionalidad. La estimación se realiza mediante las siguientes formulaciones:

$$A_t = \alpha \frac{D_t}{C_{t-L}} + (1 - \alpha)(A_{t-1} - T_{t-1}), \quad \alpha \in [0, 1] \quad (2.2)$$

$$T_t = \beta(A_t - A_{t-1}) + (1 - \beta)T_{t-1}, \quad \beta \in [0, 1] \quad (2.3)$$

$$C_t = \gamma \frac{D_t}{A_t} + (1 - \gamma)C_{t-L}, \quad \gamma \in [0, 1] \quad (2.4)$$

$$F_{t+k} = (A_t + kT_t)C_{t-L+((k-1) \bmod L)} \quad (2.5)$$

En este caso, el parámetro L representa la frecuencia de la estacionalidad; por ejemplo, $L = 4$ corresponde a datos semestrales, mientras que $L = 12$ a datos mensuales. Los parámetros de suavización, α , β y γ , son usados para controlar el *trade-off* entre el nivel, tendencia y estacionalidad, respectivamente.

En esta tesis, se abordan una serie de alternativas para la predicción de series de tiempo asociadas a la estimación de demanda. Un proyecto de estimación de demanda es muy importante a la hora de tomar decisiones (Franses & Legerstee, 2011; Huber, Gossmann & Stuckenschmidt, 2017; Jiang, Chin, Wang, Qu & Tsui, 2017; Martínez, Frías, Pérez-Godoy & Rivera, 2018), ya que el ser capaces de conocer con cierto grado de certeza lo que ocurrirá en el futuro, nos lleva, sin duda, a prepararnos de mejor forma para enfrentarlo. Un ejemplo de esto podría ser la gestión de productos en un retail, o un caso particular, en el Travel Retail, como lo que se trabaja en la primera investigación conducente a esta tesis. A nivel de productos, el ser capaces de conocer la demanda previamente, nos puede ayudar a tomar mejores decisiones tanto a nivel operativo, táctico y estratégico, como la selección de productos a reponer o vender, o incluso la decisión de abrir nuevas tiendas (Wood & Tasker, 2008). Existen muchos contextos en los que se pueden aplicar predicciones de series de tiempo, siendo la venta de productos solo un ejemplo de esto. Se puede generar predicciones en el contexto energético (Valente & Maldonado, 2020; Zhang, Wang, Liao, Xiao & Fu, 2022), en temas de *pricing* (Liu, Wang, Chen & Zhu, 2022), en salud (Jiang, Chin, Wang, Qu & Tsui, 2017), entre otros.

2.3. Series de tiempo jerárquicas

Una serie de tiempo puede implícitamente ser descompuesta en base a ciertas características que la representan. Por ejemplo, si la serie de tiempo que estamos estudiando

corresponde a la venta diaria de cierto producto, es razonable pensar que dicho producto puede tener otras características, que nos permitan separar este análisis en nuevas series de tiempo, como por ejemplo, su color. Si nuestra serie de tiempo contempla la venta total del producto X , podríamos extraer, por ejemplo, dos nuevas series: la venta del producto X de color rojo (X^{rojo}) y las ventas del producto X de color azul (X^{azul}). En el caso de que estos fueran los únicos colores disponibles para el producto X , es directo pensar que la venta total del producto X corresponderá a la suma de las ventas de su versión roja, más la suma de las ventas de su versión azul. Por lo tanto, nuestra serie de tiempo total para un instante de tiempo t se podría definir como

$$X_t = X_t^{\text{rojo}} + X_t^{\text{azul}}. \tag{2.6}$$

Así mismo, se podría utilizar cualquier característica de la serie de tiempo en cuestión para su estudio y desagregación. Esto es lo que denominamos serie de tiempo jerárquica, o HTS por sus siglas en inglés. Hyndman y Athanasopoulos (Hyndman & Athanasopoulos, 2013) formalizaron la idea de estimaciones relacionadas a series de tiempo conformadas por una estructura de agregación jerárquica. Los ejemplos más comunes son estimación de ventas que pueden ser desagregadas por tipo de producto, o por ubicación geográfica.

La figura 2.2 presenta un ejemplo de estructura jerárquica (Hyndman & Athanasopoulos, 2013), donde la serie de tiempo “Total” es desagregada en un primer nivel en las series A y B , las cuales a su vez son desagregadas nuevamente, generando una estructura jerárquica de tres niveles, obteniendo dos nodos en el primer nivel de desagregación, y cinco nodos en el último nivel.

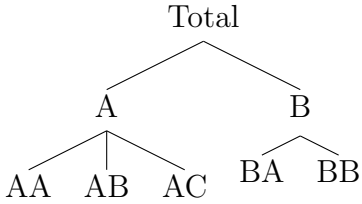


Figura 2.2: Ejemplo de jerarquía.

Con esto, podemos observar que en lugar de una única serie de tiempo “Total”, ahora tenemos una jerarquía de ocho nodos, donde cada uno de ellos corresponde a una serie de

tiempo individual. Podemos notar que algunas de estas series de tiempo nacen de otras, lo que nos lleva a poder nombrar algunos nodos como “Padre” o “Hijo”. En el ejemplo de la figura 2.2, el nodo “Total” es el padre tanto de los nodos A como B , quienes a su vez son padres de los nodos AA , AB y AC y BA y BB respectivamente. A la serie “Total” la denominaremos y_t , y a cualquier serie asociada al nodo n la denominaremos $y_{n,t}$. Con esta nomenclatura, notemos que la serie del ejemplo la podemos formalizar como:

$$\begin{bmatrix} y_t \\ y_{A,t} \\ y_{B,t} \\ y_{AA,t} \\ y_{AB,t} \\ y_{AC,t} \\ y_{BA,t} \\ y_{BB,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_{AA,t} \\ y_{AB,t} \\ y_{AC,t} \\ y_{BA,t} \\ y_{BB,t} \end{bmatrix}. \quad (2.7)$$

2.4. Enfoques predictivos para análisis de series de tiempo jerárquicas

Existen una serie de enfoques que se pueden utilizar a la hora de generar estimaciones con series de tiempo jerárquicas. Entre ellos, los más conocidos corresponden a los enfoques bottom-up (BU), top-down (TD) y middle-out (MO). El enfoque bottom-up se centra en ejecutar la predicción en el último nivel de la jerarquía, para cada uno de sus nodos, y luego agregar estos resultados basados en la estructura jerárquica del problema para obtener las predicciones de los niveles superiores. Denominaremos $\hat{y}_{n,h}$ como la estimación de h pasos para la serie de tiempo del nodo n . Siguiendo el ejemplo de la figura 2.2, una estimación de h pasos utilizando el enfoque bottom-up se puede llevar a cabo para cada nodo n basado en la predicción de los nodos del último nivel, AA , AB , AC , BA y BB . Para el ejemplo, las estimaciones se calcularían como sigue:

$$\hat{y}_{A,h} = \hat{y}_{AA,h} + \hat{y}_{AB,h} + \hat{y}_{AC,h}, \quad (2.8)$$

$$\hat{y}_{B,h} = \hat{y}_{BA,h} + \hat{y}_{BB,h}, \quad (2.9)$$

$$\hat{y}_h = \hat{y}_{A,h} + \hat{y}_{B,h}. \quad (2.10)$$

Este método tiene la ventaja de prácticamente no tener pérdidas de información en las agregaciones, pero eso no quiere decir que los datos presentes en algunas series de tiempo de la jerarquía, especialmente en los niveles inferiores, puedan ser ruidosos debido a falta de información, pudiendo arrastrar con ellos malos resultado predictivos (Hyndman & Athanasopoulos, 2013).

Con respecto al enfoque top-down, al contrario del anterior, se busca realizar la predicción en el nivel “Total” de la jerarquía (\hat{y}_h). Luego, este resultado se puede desagregar en los diferentes nodos, basados en una proporción definida previamente. Lo más común para esta proporción es utilizar la distribución promedio histórica para cada nodo j (Hyndman & Athanasopoulos, 2013). Esto se puede expresar mediante la siguiente fórmula:

$$p_j = \frac{1}{T} \sum_{t=1}^T \frac{y_{j,t}}{y_t} \quad (2.11)$$

Si volvemos nuevamente al ejemplo de la figura 2.2, las predicciones para cada nodo, basados en la estimación de la serie “Total” \hat{y}_h , estarían dadas por:

$$\hat{y}_{A,h} = p_A \cdot \hat{y}_h, \quad (2.12)$$

$$\hat{y}_{B,h} = p_B \cdot \hat{y}_h, \quad (2.13)$$

$$\hat{y}_{AA,h} = p_{AA} \cdot \hat{y}_{A,h}, \quad (2.14)$$

$$\hat{y}_{AB,h} = p_{AB} \cdot \hat{y}_{A,h}, \quad (2.15)$$

$$\hat{y}_{AC,h} = p_{AC} \cdot \hat{y}_{A,h}, \quad (2.16)$$

$$\hat{y}_{BA,h} = p_{BA} \cdot \hat{y}_{B,h}, \quad (2.17)$$

$$\hat{y}_{BB,h} = p_{BB} \cdot \hat{y}_{B,h}. \quad (2.18)$$

Una de las principales desventajas de este método es la pérdida de información potencial que se produce en el proceso de desagregación, y con cada nueva repetición de este proceso, para generar nuevas desagregaciones, estas pérdidas de información pueden incrementar (Hyndman & Athanasopoulos, 2013).

El tercer enfoque clásico corresponde a middle-out. Este enfoque se centra en ejecutar la predicción en algún nivel intermedio de la jerarquía, ni el “Total” ni el último nivel. Una vez definido el nivel inicial, en el que se realiza la predicción, se utilizan ambos enfoques mencionados anteriormente para la estimación del resto de los nodos del árbol jerárquico. En este caso, los nodos superiores al nivel inicial serían estimados mediante el enfoque bottom-up, mientras que los nodos inferiores mediante el enfoque top-down (Hyndman & Athanasopoulos, 2013).

Existe un cuarto enfoque que ha ganado fuerza en ciertos contextos, y que fue formalizado por Hyndman y Athanasopoulos el año 2014. Este enfoque es denominado *Optimal Reconciliation* (R. J. Hyndman, 2014). Lo que se busca es encontrar una combinación óptima para todas las predicciones independientes en todos los niveles jerárquicos, mediante promedios ponderados. Todos estos métodos fueron presentados empíricamente en Athanasopoulos et al. (Athanasopoulos, Hyndman, Kourentzes & Petropoulos, 2017), y fueron implementados en un paquete de R (Hyndman, Athanasopoulos & Shang, 2012).

Para el caso de series de tiempo jerárquicas, Hyndman y Athanasopoulos recientemente introdujeron una alternativa para la estimación de demanda sujeta a productos de muchas categorías (Hyndman & Athanasopoulos, 2013). La estructura jerárquica para aprovechar las características intrínsecas del problema puede ser muy beneficioso de cara a la estimación de demanda a nivel de productos (Athanasopoulos, Hyndman, Kourentzes & Petropoulos, 2017). Por otro lado, si bien los enfoques más utilizados corresponden a bottom-up, top-down y middle-out, existe potencial para explorar los métodos de reconciliación. Franca Lila et al. (Franca Lila, Meira & Cyrino Oliveira, 2022) realizan tareas de predicción de series de tiempo jerárquicas usando un enfoque robusto de reconciliación. Las metodologías de reconciliación, introducidas en los trabajos de Hyndman y Athanasopoulos, también han sido estudiadas en detalle en el trabajo de Panagiotelis et al. (Panagiotelis, Athanasopoulos, Gamakumara & Hyndman, 2021).

2.5. Aprendizaje Supervisado

En el contexto de la ciencia de datos, la Inteligencia Artificial se ha convertido en un tema de extremo interés. Machine Learning es la parte de la Inteligencia Artificial que ha desarrollado una serie de técnicas que le permiten a los computadores “aprender” (Mitchell, 1997), a través de un proceso de inducción del conocimiento, capaz de modelar comportamientos desde ejemplos.

Una de las partes más importantes dentro de los modelos de Machine Learning, corresponde al proceso de aprendizaje que se utilice. Existen una serie de tipos de aprendizaje, siendo los más conocidos el aprendizaje supervisado, el aprendizaje no supervisado y el aprendizaje reforzado.

En este trabajo se abordan una serie de problemas de aprendizaje supervisado, enfocados en problemas de regresión. Esto quiere decir que se entrenan modelos con datos históricos, para que esa base de conocimiento se utilice como ejemplo para la predicción de nuevas observaciones.

Una de las técnicas más conocidas de aprendizaje supervisado corresponde a las máquinas de soporte vectorial, o SVM por sus siglas en inglés. Este tipo de modelos son muy populares en el contexto de problemas de clasificación, debido a su gran capacidad de construir funciones no lineales complejas, y obtener desempeños predictivos muy positivos, mostrando en general mejores resultados que otros modelos de clasificación (Vapnik, 1998). Esta técnica posee una adaptación para regresiones, denominada Support Vector Regression (SVR) o ε -SVR, ya que introduce la variable ε (Drucker, Burges, Kaufman, Smola & Vapnik, 1997; Smola & Scholkopf, 1998), que corresponde a la desviación máxima permitida de la función $f(x)$ buscada con respecto a sus objetivos y_i , buscando que a la vez, esta función sea lo más “plana” posible. Esto quiere decir que el método admite errores sin penalización, siempre y cuando estos errores no superen este parámetro ε (Smola & Scholkopf, 2004). Este método está explicado en detalle en la próxima sección, ya que corresponde a la principal herramienta que se utiliza como modelo predictivo a lo largo de nuestras investigaciones.

En la literatura, la estimación de demanda generalmente ha sido abordada por los métodos tradicionales de series de tiempo, como ARIMA o Holt-Winters (Makridakis & Wheelwright, 1977), pero en el último tiempo, los métodos relacionados a Machine

Learning están siendo cada vez más estudiados, debido principalmente a su capacidad de modelar patrones no lineales, demostrando ser capaces de obtener muy buenos resultados en contextos de datos reales (Crone, Lessmann & Pietsch, 2006; Hansen, McDonald & Nelson, 2006; Ahmed, Atiya, El-Gayar & El-Shishiny, 2010).

2.6. Support Vector Regression (SVR) para análisis de series de tiempo

En esta sección se presenta el método ε -SVR (Drucker, Burges, Kaufman, Smola & Vapnik, 1997; Smola & Scholkopf, 1998), que corresponde a una extensión de las máquinas de soporte vectorial (SVM) para regresiones. Este método se utiliza para generar estimaciones de series de tiempo.

Dado un conjunto de datos de entrenamiento que abarca una variable dependiente continua, $y_i \in \mathbb{R} \forall i = 1, \dots, m$, y las covariables $\mathbf{x}_i \in \mathbb{R}^p, i = 1, \dots, m$, que corresponden a rezagos obtenidos desde la variable dependiente, SVR minimiza la función objetivo del siguiente problema de programación cuadrática:

$$\begin{aligned} \min_{\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\xi}^*} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^m (\xi_i + \xi_i^*) \\ \text{s.t.} \quad & y_i - (\mathbf{w}^\top \mathbf{x}_i + b) \leq \varepsilon + \xi_i, \quad i = 1, \dots, m, \\ & (\mathbf{w}^\top \mathbf{x}_i + b) - y_i \leq \varepsilon + \xi_i^*, \quad i = 1, \dots, m, \\ & \xi_i, \xi_i^* \geq 0, \quad i = 1, \dots, m, \end{aligned} \tag{2.19}$$

La introducción de las variables de holgura, $\boldsymbol{\xi}, \boldsymbol{\xi}^* \in \mathbb{R}^m$, permite que los objetos permanezcan fuera del “tubo” de radio ε , pero los penaliza en la función objetivo. El hiperparámetro C nos ayuda a controlar el *trade-off* entre el ajuste del modelo y la reducción de complejidad (Smola & Scholkopf, 1998). Por otro lado, el uso de la regularización l_2 (la norma euclídeana de \mathbf{w}), garantiza que la función de regresión, $f(\mathbf{x})$, sea lo más plana posible (Smola & Scholkopf, 1998).

La formulación (2.19) también puede ser representada por su forma dual, lo que lleva a la incorporación del método del kernel para la construcción de superficies no lineales.

En este caso, la función kernel, $K(\mathbf{x}_i, \mathbf{x}_s)$, se introduce para mapear los datos a un espacio de mayor dimensión (Schölkopf & Smola, 2002). La representación dual de la formulación (2.19) es la siguiente:

$$\begin{aligned}
\max_{\alpha, \alpha^*} \quad & \sum_{i=1}^m y_i(\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^m (\alpha_i + \alpha_i^*) - \frac{1}{2} \sum_{i,s=1}^m (\alpha_i - \alpha_i^*)(\alpha_s - \alpha_s^*)K(\mathbf{x}_i, \mathbf{x}_s) \\
\text{s.t.} \quad & \sum_{i=1}^m (\alpha_i - \alpha_i^*) = 0, \\
& 0 \leq \alpha_i \leq C, \quad i = 1, \dots, m, \\
& 0 \leq \alpha_i^* \leq C, \quad i = 1, \dots, m.
\end{aligned} \tag{2.20}$$

Con esto, para una nueva observación, \mathbf{x} , la regla de decisión, $f(\mathbf{x})$, corresponde a:

$$f(\mathbf{x}) = \sum_{i=1}^m (\alpha_i - \alpha_i^*)K(\mathbf{x}, \mathbf{x}_i) + b \tag{2.21}$$

Existen una serie de alternativas para la función kernel. En este trabajo se utilizan dos de ellas: el kernel lineal y el kernel gaussiano, también conocido como *Radial Basis Function* (RBF). Ambos tipos de kernel son de los más utilizados en la literatura. El kernel lineal, por su parte, es una función de menor complejidad, que permite modelar problemas más simples. Por otro lado, el kernel gaussiano, nos permite modelar problemas no lineales de alta complejidad, con muy buenos resultados.

El kernel lineal es de la forma $K(\mathbf{x}_i, \mathbf{x}_s) = \mathbf{x}_i^T \mathbf{x}_s$, mientras que el kernel gaussiano se define como $K(\mathbf{x}_i, \mathbf{x}_s) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_s\|^2)$, donde $\gamma > 0$ corresponde al hiperparámetro que controla el ancho del kernel (Schölkopf & Smola, 2002).

La utilización de Machine Learning para la predicción de series de tiempo se ha estudiado bastante, como por ejemplo, en los trabajos de He et al. (He, Wu & Si, 2022), quienes realizan estimaciones de venta para gran cantidad de datos utilizando redes neuronales, o Sohrabpour et al. (Sohrabpour, Oghazi, Toorajipour & Nazarpour, 2021), quienes proponen un novedoso marco causal basado en inteligencia artificial para modelar y pronosticar ventas de exportación.

En particular, SVR se ha utilizado ampliamente para el análisis de series de tiempo. Se puede encontrar una encuesta relacionada en Sapankevych y Sankar (Sapankevych &

Sankar, 2009), quienes afirman que SVR puede funcionar mejor que otros métodos en el pronóstico de series de tiempo, principalmente debido a su naturaleza no lineal y la existencia de un único mínimo global.

Crone et al. (Crone, Guajardo & Weber, 2006) analizan la capacidad de SVR para pronosticar patrones básicos de series de tiempo, comparando ambas funciones de kernel. En su trabajo, afirman que SVR con kernel gaussiano supera a otros métodos para datos sin tendencia. Para series de tiempo con componentes de tendencia, encontraron que el kernel lineal era mejor que la versión gaussiana.

Además, Levis y Papageorgiou (Levis & Papageorgiou, 2009), realizaron estimación de demanda con un algoritmo de tres etapas, buscando lidiar con patrones no lineales de forma efectiva, y lograron una muy buena precisión predictiva en todos los casos reportados. Por otro lado, Chi-Jie Lu (Lu, 2014) utiliza SVR para predicción de ventas, combinando el método con *splines* de regresión adaptativa multivariante. Otro trabajo en esta línea es el de Du, et al. (Du, Leung, Zhang & Lai, 2013), quienes utilizan SVR para la predicción de demanda para productos agrícolas. Otro enfoque fue presentado por Chi-Jie Lu (Lu & Wang, 2010), quien desarrolló un método de estimación de demanda con SVR, en el que se utilizan análisis de componentes independientes y mapas autoorganizados como herramienta de preprocesamiento. Así, podemos encontrar una serie de estudios que utilizan SVR como método predictivo de series de tiempo, incluso en contextos muy diferentes a la predicción de venta, como son el caso de la estimación de la velocidad del viento o de la carga eléctrica (Hu, Bao & Xiong, 2013; Ren, Suganthan & Srikanth., 2016; Valente & Maldonado, 2020).

Por otro lado, es posible encontrar en la literatura algunas modificaciones de SVR, buscando mejorar su desempeño predictivo. Por ejemplo, Liu et al. (Liu, Wang, Li & Xiao, 2021) realizan una adaptación de SVR para procesamiento de señales con funciones multidominio. Otra adaptación de SVR se encuentra en Sabzekar et al. (Sabzekar & Hossein Hasheminejad, 2021) quienes buscan mayor robustez en fuentes de datos complejas y ruidosas. En su modelo, las desigualdades de las restricciones de SVR relacionadas a ε se cambian a desigualdades difusas sin ninguna variación en su función de pérdida.

2.7. Estimación de demanda y políticas de *staffing* en Contact Centers

Los Contact Center, antiguamente conocidos como Call Center ya que solo consideraban el canal de voz, incorporan una serie de aspectos técnicos y de recursos humanos para concretar la comunicación entre dos partes (tradicionalmente empresa-cliente). Un Contact Center puede recibir cientos de llamadas en un solo día, por lo que ser capaces de comprender el problema de la cantidad de llamadas entrantes resulta muy importante a la hora de la toma de decisiones, ya sea operacionales, como la distribución de los turnos de los ejecutivos, o más estratégicas, como la determinación de la dotación óptima o cantidad de licencias del software de Contact Center que se debe contratar. Tradicionalmente se han estudiado dos formas de modelar los parámetros de un Contact Center: fórmulas Erlang y modelos de Markov.

Las llamadas que ingresan a la plataforma cuando no hay ejecutivos disponibles para atenderlas caen a una cola de atención, y son asignadas a los agentes a medida que ellos se desocupan. Si la cola de espera está vacía, el ejecutivo está disponible, y esperará a que entre la próxima llamada.

Erlang C es una fórmula definida por dos variables: el número de agentes N y la carga A , definida como la cantidad de llamadas promedio por unidad de tiempo λ dividido por la intensidad del tráfico μ . Así, se puede definir η como la carga promedio por agente, dividiendo A entre N (Chromy, Misuth & Kavacky, 2011). Con esto, se puede calcular la probabilidad P_c de que una llamada caiga a una cola de espera como:

$$P_c(N, \eta) = \frac{\frac{(N\eta)^N}{N!(1-\eta)}}{\sum_{i=0}^{N-1} \left\{ \frac{(N\eta)^i}{i!} + \frac{(N\eta)^N}{N!(1-\eta)} \right\}}. \quad (2.22)$$

Se puede demostrar que el modelo Erlang C es equivalente a un modelo de Markov $M/M/m/\infty$ (Chromy, Misuth & Kavacky, 2011). Además, existen otros parámetros que se pueden monitorear, como por ejemplo el tiempo de espera en la cola, que corresponde a una variable aleatoria que se puede describir por la siguiente función de distribución (Bolch, Greiner, De Meer & Trivedi, 2006):

$$F_W(\tau) = 1 - P_C \cdot e^{-\mu(N-A)\tau}, \quad (2.23)$$

siendo $\tau > 0$. Con esto, es posible calcular W , el tiempo promedio en una cola de espera, como:

$$W = \frac{P_C}{\mu(N-A)}. \quad (2.24)$$

Aplicando un teorema y las fórmulas anteriores (Bolch, Greiner, De Meer & Trivedi, 2006) se puede obtener Q , la cantidad promedio de llamadas en espera, la cuál está dada por la siguiente fórmula:

$$Q = \lambda A = \frac{A}{(N-A)} \cdot P_C. \quad (2.25)$$

En un Contact Center, uno de los indicadores más importantes corresponde al nivel de servicio NS . Este indicador corresponde al porcentaje de las llamadas que son contestadas antes del tiempo aceptable de espera t^* . En general, t^* se usa como 15 o 20 segundos. Se puede obtener una fórmula para el nivel de servicio NS utilizando la definición general de función de distribución y aplicando sus propiedades (Riečanová, Horváth, Olejček & Riečan, 1987). La fórmula que se obtiene para el nivel de servicio es la siguiente:

$$NS = 1 - P_C \cdot e^{-\mu(N-A)t^*}. \quad (2.26)$$

Con esto, la utilización promedio de los agentes η se puede calcular con la siguiente expresión (Unčovský, 1980):

$$\eta = 1 - \frac{\sum_{k=0}^{N-1} \left(\frac{N-k}{n \cdot k!} \right) A^k}{\sum_{i=0}^{N-1} \frac{A^i}{i!} + \frac{A^N N}{N!(N-A)}}. \quad (2.27)$$

Uno de los grandes supuestos de la fórmula Erlang C para el análisis de Contact Centers es que se asume que las colas de espera son infinitas (Chromy, Misuth & Kavacky, 2011). Actualmente, muchas empresas utilizan esta formulación para sus decisiones, no necesariamente debido a la precisión de sus resultados, sino que por la holgura que ellos

le dan, entregando resultados que con una confianza bastante alta son peores que la realidad, pudiendo tomar decisiones conservadoras en base a ellos.

Si bien la tarea de la estimación de demanda en un Contact Center generalmente es abordada por la metodología Erlang C, explicada anteriormente, existen una serie de estudios que realizan predicciones de volumen de llamadas en un Contact Center, como por ejemplo Kanthanathan et al. (Kanathanathan, Carty, Raja & Ryan, 2020), quienes utilizan una serie de técnicas derivadas de redes neuronales para la estimación. Por otro lado, Abut et al. (Abut, Tartuk, Nurdag, Acar, Erdem & Akay, 2022) utilizan *Extreme Gradient Boosting* (XGBoost) y comparan su desempeño con modelos basados en la técnica de media móvil, como el caso de ARIMA, superándolos en términos de desempeño predictivo para todos los experimentos reportados.

Con respecto a la predicción de llamadas en Contact Centers, y su posterior utilización en modelos de *staffing*, existen una serie de trabajos que lo abordan, pero hasta donde pudimos averiguar, a la fecha no hay ningún trabajo que considere modelos de pronóstico con Machine Learning y que utilice los intervalos de confianza de sus resultados para la generación de escenarios de un modelo de optimización estocástica, que permita generar una distribución óptima de los agentes del Contact Center. Lo más cercano podría ser Chernichenko et al. (Chernichenko, Kapkanets & Panchenko, 2016), quienes usan modelos básicos para la predicción de demanda en un Contact Center, y proponen un modelo para resolver la problemática de *staffing*. Fuera del contexto de Contact Center, se puede encontrar algunos trabajos que consideran modelos predictivos como *input* de modelos de optimización, como por ejemplo Mizan et al. (Mizan & Taghipour, 2022), quienes buscan reducir tiempo de espera de los pacientes mediante la distribución de recursos de radiología. Proponen una solución de tres fases, que integra métodos de Machine Learning con un modelo de optimización.

Capítulo 3

Metodología

3.1. SVR para la predicción de series de tiempo jerárquicas

Si bien antes de este trabajo existían una serie de trabajos relacionados a modelos predictivos con SVR, hasta el momento nadie había implementado un algoritmo para lidiar con series de tiempo jerárquicas. Una de las principales ventajas del trabajo con series de tiempo jerárquicas es que permite explotar de mejor forma las características propias de las series de tiempo, utilizándolas a nuestro beneficio para mejorar el desempeño predictivo de los modelos.

En esta sección se formaliza una serie de algoritmos para estimaciones de series de tiempo jerárquicas, utilizando ε -SVR como modelo predictivo. El principal objetivo es usar la estructura jerárquica de los problemas para obtener mejores predicciones, mediante el correcto uso de los algoritmos con métodos de kernel que permitan el modelamiento de patrones no lineales, construyendo funciones de regresiones flexibles. Hasta donde sabemos, el uso de ε -SVR para series de tiempo jerárquicas aún no se ha informado en la literatura, siendo este el primer trabajo en el tópico.

Para poder formalizar los diversos enfoques de series de tiempo jerárquicas utilizando SVR como modelo predictivo, lo primero es definir los parámetros y conjuntos de datos que se utilizan:

- $j = (0, \dots, J)$: niveles de la jerarquía, donde el nivel 0 representa el nodo raíz o “Total”, y el nivel J corresponde al que contiene todos los nodos del último nivel.

- $n = (1, \dots, N)$: conjunto de nodos de la jerarquía.
- A_j : subconjunto de nodos del nivel j .
- S_n : subconjunto de nodos que nacen desde el nodo n .
- D_n : nodo padre para el nodo n .
- p_n : proporción histórica del nodo n .
- F_n : predicción para el nodo n .

Con esto, podremos definir tres algoritmos, uno para cada enfoque de HTS: SVR-BU, algoritmo que utiliza el enfoque bottom-up; SVR-TD, el que utiliza top-down; y finalmente SVR-MO, el que utiliza middle-out.

3.1.1. Algoritmo SVR-BU

Este método entrena un modelo de ε -SVR para cada nodo del último nivel de la jerarquía (formulación (2.20)), y obtiene sus respectivas predicciones, F_n , ($\forall n \in A_{|J|}$). Con estas estimaciones, se procede a calcular las predicciones de los niveles superiores de la jerarquía, mediante un proceso de agregación utilizando $F_n = \sum_{i \in S_n} F_i$, es decir, la suma de todas las predicciones de todos los nodos hijos que nacen desde el nodo n . El algoritmo de este método es el siguiente:

- a) Calcular predicciones, F_n , $\forall n \in A_J$:
 - I) Para cada nodo del último nivel, entrenar ε -SVR resolviendo la formulación (2.20) con una configuración de hiperparámetros dada para obtener como solución la tupla (α, α^*) .
 - II) Obtener las predicciones, F_n , en el último nivel utilizando la formulación (2.21): $F_n(\mathbf{x}) = \sum_{i=1}^m (\alpha_i - \alpha_i^*)K(\mathbf{x}, \mathbf{x}_i) + b$.
- b) Para cada nivel, $j = (0, \dots, J-1)$, calcular las predicciones, F_n , $\forall n \in A_j$, mediante $F_n = \sum_{i \in S_n} F_i$, es decir, la suma de todas las predicciones de los nodos que nacen desde el nodo n .

3.1.2. Algoritmo SVR-TD

El método SVR-TD opera de la siguiente manera: primero se debe entrenar ε -SVR en el nodo raíz (formulación (2.20)), obteniendo las predicciones F_0 . Luego, dicha predicción se desagrega hacia los niveles inferiores de la jerarquía utilizando las proporciones históricas de cada nodo. Este algoritmo es el siguiente:

- a) Calcular las predicciones F_0 en el nodo raíz:
 - i) Para el nivel superior de la jerarquía, entrenar ε -SVR resolviendo la formulación (2.20) con una configuración de hiperparámetros dada para obtener como solución la tupla (α, α^*) .
 - ii) Obtener las predicciones en el nivel raíz, F_0 , mediante la formulación (2.21):

$$F_0(\mathbf{x}) = \sum_{i=1}^m (\alpha_i - \alpha_i^*) K(\mathbf{x}, \mathbf{x}_i) + b$$
- b) Para cada nivel $j = (1, \dots, J)$, obtener las predicciones, $F_n, \forall n \in A_j$, mediante $F_n = p_n \cdot F_{D_j}$, es decir, desagregando las predicciones del nodo raíz siguiendo la proporcionalidad histórica de la jerarquía.

3.1.3. Algoritmo SVR-MO

En el caso del enfoque middle-out, lo primero que se debe hacer es definir el nivel inicial del método $0 < j < J$, en el que se efectuarán las predicciones mediante el entrenamiento de un modelo ε -SVR para cada nodo perteneciente a dicho nivel. Las predicciones, $F_n, \forall n \in A_j$, se obtienen mediante la formulación (2.21). Una vez que se tienen las estimaciones del nivel inicial, se procede a agregarlas hacia los niveles superiores, $k < j$, siguiendo el enfoque bottom-up: para cada nivel k , las predicciones, $F_n, \forall n \in A_k$, se obtienen mediante $F_n = \sum_{i \in S_n} F_i$, es decir, la suma de las predicciones asociadas a los nodos que nacen desde el nodo n . Para el caso de los niveles $k > j$, que están bajo en nivel j en la jerarquía, se utiliza el enfoque top-down, desagregando las predicciones hacia los niveles inferiores en base a sus proporciones históricas. El algoritmo formalizado sería el siguiente:

- a) Definir el nivel inicial, $0 < j < J$.
- b) Calcular las predicciones, $F_n, \forall n \in A_j$, en este nivel inicial intermedio:

- i) Para cada nodo de la jerarquía j , entrenar ε -SVR mediante la formulación (2.20) con una configuración de hiperparámetros dada, para obtener como solución la tupla (α, α^*) .
- ii) Calcular las predicciones del nivel intermedio, F_n , mediante la formulación (2.21):
$$F_n(\mathbf{x}) = \sum_{i=1}^m (\alpha_i - \alpha_i^*) K(\mathbf{x}, \mathbf{x}_i) + b$$
- c) Para cada nivel superior, $k < j$, calcular las predicciones, $F_n, \forall n \in A_k$, utilizando $F_n = \sum_{i \in S_n} F_i$ (método bottom-up).
- d) Para cada nivel inferior, $k > j$, calcular las predicciones, $F_n, \forall n \in A_k$, utilizando $F_n = p_n \cdot F_{D_k}$ (método top-down).

3.1.4. Selección del modelo y procedimiento de validación

Con respecto a la selección de los hiperparámetros, se utiliza un procedimiento de validación cruzada tradicional para series de tiempo, denominado *rolling forecasting origin* (Hyndman & Athanasopoulos, 2013). Dado un conjunto de entrenamiento y testeo, el conjunto de entrenamiento se separa nuevamente en un subconjunto de entrenamiento y uno de validación. Lo más importante es que el subconjunto de entrenamiento no incluya muestras que ocurran después de las incluidas en el subconjunto de validación. Para obtener la selección de hiperparámetros para cada modelo ε -SVR se sigue el siguiente proceso de validación:

- a) Seleccionar la función kernel a utilizar (lineal o gaussiano).
- b) Separar el conjunto de entrenamiento en un subconjunto de entrenamiento y un subconjunto de validación.
- c) Definir un rango para cada hiperparámetro de SVR:
 - ε y C para la versión de kernel lineal.
 - ε, C y γ para la versión de kernel gaussiano.
- d) Entrenar el modelo utilizando el subconjunto de entrenamiento para cada combinación de hiperparámetros, resolviendo la formulación (2.20) para obtener (α, α^*) .

- e) Calcular la predicción para el nodo en el subconjunto de validación, utilizando la tupla, (α, α^*) y la formulación (2.21).
- f) Calcular el error asociado a la predicción, comparando la solución del modelo con los valores reales. Como fórmula de error se puede utilizar tanto el error porcentual absoluto medio, o MAPE, por sus siglas en inglés, como también la raíz del error cuadrático medio, o RMSE, por sus siglas en inglés.
- g) Guardar la configuración de hiperparámetros que obtienen el menor MAPE o RMSE en el subconjunto de validación.

3.2. θ -SVR: Un enfoque novedoso de HTS basado en agregaciones de información

Si bien con el primer trabajo publicado se pudo observar que los enfoques clásicos de ε -SVR para la predicción de series de tiempo jerárquicas tienen resultados muy alentadores, y que pueden fácilmente superar los resultados de los métodos clásicos de la literatura, como ARIMA y Holt-Winters; en esta sección presentamos un nuevo enfoque para el uso de SVR, en el que buscamos construir regresores en el último nivel de forma simultánea, en un único problema de optimización, e incorporando un objetivo adicional: que los pesos de cada hiperplano no se desvíen mucho de los obtenidos por los regresores en los niveles superiores. De esta forma, podemos agrupar información entre jerarquías de manera óptima, en lugar de simplemente desagregar el pronóstico obtenido de un nivel superior proporcionalmente a sus nodos hijos.

Para la metodología propuesta se definen dos variantes. La primera estrategia agrupa información directamente desde el nivel superior, mientras que la segunda variante minimiza la desviación entre los hiperplanos del último nivel (nivel J), con respecto a los nodos que están en el nivel $J-1$, es decir, el nivel inmediatamente superior al último nivel. Cabe destacar que ambos enfoques son equivalentes en el caso de un problema jerárquico de sólo dos niveles, pero serán distintos entre ellos siempre que tengamos más de dos niveles de jerarquía.

A esta novedosa propuesta la denominamos θ -SVR, y corresponde a una extensión de la metodología bottom-up, en el sentido de que estamos estimando todas las funciones

para el nivel J . A la variante que relaciona información directamente desde el nivel 0, construyendo un hiperplano en el último nivel, la denominaremos θ -SVR $_{\{J,0\}}$, mientras que a la variante que relaciona los niveles J y $J-1$, construyendo regresores para ambos niveles, y minimizando la desviación entre las series de tiempo del nivel $J-1$ y sus respectivos nodos hijos, lo denominaremos θ -SVR $_{\{J,J-1\}}$.

Dado que ambas variantes poseen la misma estructura, podemos formalizarlas de la siguiente manera:

Sean $\mathbf{x}_{in} \in \text{Re}^p$ las covariables del problema de regresión, es decir, los rezagos relacionados a las series de tiempo del último nivel. $n \in A_J$ y $y_{in} \in \text{Re}$ corresponden a sus respectivos *outputs*, siendo $i = 1, \dots, N_n$ cada entrada relacionada con el nodo n , es decir, los diferentes periodos de tiempo utilizados para el entrenamiento.

Nuestra propuesta estima una función lineal en los nodos del último nivel de la forma:

$$f_n(\mathbf{x}) := \mathbf{w}_n^\top \mathbf{x} + b_n, \quad n \in A_J, \quad (3.1)$$

donde la información es agrupada desde los niveles superiores.

Para la primera variante, se define $\bar{\mathbf{w}}_0$ como un vector de peso auxiliar que se relaciona con el nodo raíz ya que considera la información de todos los nodos. La idea es construir los hiperplanos $f_n(\mathbf{x})$ de tal forma que sus pesos no se desvíen mucho de $\bar{\mathbf{w}}_0$. Esto se incluye en el problema como un tercer objetivo, además del ajuste del modelo y la minimización de la norma euclidiana. Como este nuevo objetivo relaciona todas las estimaciones para los nodos del último nivel, la formulación propuesta debe ser escrita en un único problema de optimización, como:

$$\begin{aligned} \min_{\substack{\bar{\mathbf{w}}_0, \mathbf{w}_n, b_n \\ \boldsymbol{\xi}_n, \boldsymbol{\xi}_n^*}} \quad & \frac{1}{2} \sum_{n \in A_J} (\|\mathbf{w}_n\|^2 + \theta \|\mathbf{w}_n - \bar{\mathbf{w}}_0\|^2) \\ & + C \sum_{n \in A_J} \mathbf{e}_n^\top (\boldsymbol{\xi}_n + \boldsymbol{\xi}_n^*) \\ \text{s.t.} \quad & y_{in} - (\mathbf{w}_n^\top \mathbf{x}_{in} + b_n) \leq \varepsilon + \xi_{in}, \quad \xi_{in} \geq 0, \\ & i = 1, \dots, N_n, \quad n \in A_J, \\ & (\mathbf{w}_n^\top \mathbf{x}_{in} + b_n) - y_{in} \leq \varepsilon + \xi_{in}^*, \quad \xi_{in}^* \geq 0, \\ & i = 1, \dots, N_n, \quad n \in A_J, \end{aligned} \quad (3.2)$$

donde $\varepsilon > 0$ define el tubo insensible a ε y θ, C corresponden a parámetros positivos que controlan el *trade-off* entre los tres objetivos: regularización, ajuste del modelo, y control de la heterogeneidad a lo largo de las series de tiempo. Podemos notar que el parámetro θ es sumamente importante en nuestros modelos, ya que nos entrega los pesos para el control de la heterogeneidad a lo largo de las series. Es por esto que la propuesta recibe la denominación de θ -SVR. Notemos que la configuración del parámetro $\theta = 0$ corresponde a la estrategia bottom-up utilizando ε -SVR.

Para garantizar que nuestra propuesta posee una única solución óptima, se debe cumplir que nuestro modelo sea estrictamente convexo. Esto se demostrará a continuación. Primero, denotamos

$$\widetilde{\mathbf{w}} = [\widetilde{\mathbf{w}}_0^\top, \mathbf{w}_1^\top, \dots, \mathbf{w}_N^\top]^\top \in \text{Re}^{p(N+1)}$$

el vector de pesos conjunto, y

$$\mathcal{Q}(\theta) = \begin{bmatrix} N\theta I_p & -\theta I_p & -\theta I_p & \cdots & -\theta I_p \\ -\theta I_p & (1+\theta)I_p & 0 & \cdots & 0 \\ -\theta I_p & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ -\theta I_p & 0 & \cdots & 0 & (1+\theta)I_p \end{bmatrix}, \quad (3.3)$$

una matriz de $p(N+1) \times p(N+1)$, donde I_p representa la matriz identidad de orden p . Con esto, el término cuadrático en la función objetivo del problema (3.2) puede expresarse como:

$$\sum_{n \in A_J} (\|\mathbf{w}_n\|^2 + \theta \|\mathbf{w}_n - \widetilde{\mathbf{w}}_0\|^2) = \widetilde{\mathbf{w}}^\top \mathcal{Q}(\theta) \widetilde{\mathbf{w}}. \quad (3.4)$$

Proposición 3.2.1. *Para cada $\theta > 0$, la matriz $\mathcal{Q}(\theta)$ es simétrica definida positiva. Además,*

$$\mathcal{Q}(\theta)^{-1} = \frac{1}{N} \begin{bmatrix} \frac{\theta+1}{\theta} I_p & I_p & I_p & \cdots & I_p \\ I_p & \frac{\theta+N}{\theta+1} I_p & \frac{\theta}{\theta+1} I_p & \cdots & \frac{\theta}{\theta+1} I_p \\ I_p & \frac{\theta}{\theta+1} I_p & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \frac{\theta}{(\theta+1)} I_p \\ I_p & \frac{\theta}{\theta+1} I_p & \cdots & \frac{\theta}{(\theta+1)} I_p & \frac{\theta+N}{\theta+1} I_p \end{bmatrix}. \quad (3.5)$$

Demostración. Es claro que la matriz $\mathcal{Q}(\theta)$ es simétrica definida positiva (ver ecuación (3.4)). Ahora, definimos $F_i \in \text{Re}^{p \times p(N+1)}$ y $C_i \in \text{Re}^{p(N+1) \times p}$ el bloque correspondiente

a la fila i y la columna i de $\mathcal{Q}(\theta)$ y $\mathcal{Q}(\theta)^{-1}$, respectivamente. Entonces,

$$F_1 C_1 = I_p, \quad F_1 C_i = \frac{1}{N} \left(N\theta - \frac{N\theta + \theta^2 N}{\theta + 1} \right) = 0,$$

para $i = 2, \dots, N + 1$, y

$$F_i C_1 = 0, \quad F_i C_i = I_p, \quad F_i C_j = \frac{1}{N} (\theta - \theta) = 0, i \neq j.$$

Por lo tanto, queda demostrado. \square

Nuestra segunda variante agrupa información entre los niveles J y $J - 1$, y resuelve $|A_{J-1}|$ modelos de optimización diferentes en lugar de un único problema de optimización, uno para cada nodo del nivel $J - 1$. Esto es, para cada nodo, $n \in A_{J-1}$, se resuelve:

$$\begin{aligned} \min_{\substack{\bar{\mathbf{w}}_n, \mathbf{w}_k, b_k \\ \boldsymbol{\xi}_k, \boldsymbol{\xi}_k^*}} \quad & \frac{1}{2} \sum_{k \in S_n} (\|\mathbf{w}_k\|^2 + \theta \|\mathbf{w}_k - \bar{\mathbf{w}}_n\|^2) \\ & + C \sum_{k \in S_n} \mathbf{e}_k^\top (\boldsymbol{\xi}_k + \boldsymbol{\xi}_k^*) \\ \text{s.t.} \quad & y_{ik} - (\mathbf{w}_k^\top \mathbf{x}_{ik} + b_k) \leq \varepsilon + \xi_{ik}, \quad \xi_{ik} \geq 0, \\ & i = 1, \dots, N_k, \quad k \in S_n, \\ & (\mathbf{w}_k^\top \mathbf{x}_{ik} + b_k) - y_{ik} \leq \varepsilon + \xi_{ik}^*, \quad \xi_{ik}^* \geq 0, \\ & i = 1, \dots, N_k, \quad k \in S_n. \end{aligned} \tag{3.6}$$

Se puede notar que la formulación (3.6) posee la misma estructura que la primera variante, por lo tanto, tal como se presenta en la Proposición 1, es estrictamente convexa. Los *outputs* de cada variante corresponden a las funciones de regresión, $f_n(\mathbf{x})$, para las series de tiempo del último nivel ($n \in A_J$). Una vez que se obtienen las estimaciones de estas funciones, se utiliza el enfoque bottom-up de series de tiempo jerárquicas, presentado anteriormente.

3.3. Incorporación de kernel al modelo propuesto

En la sección anterior, presentamos una novedosa metodología para realizar predicciones de series de tiempo jerárquicas, mediante la construcción de un único modelo de optimización que permite aprovechar la información proveniente de diversos niveles de la jerarquía. En esta sección, se propone una extensión al enfoque anterior, mediante la construcción de máquinas kernel diseñadas para construir todas las superficies de decisión requeridas, en un único problema de optimización. Al igual que en el caso anterior, los datos son agregados a través de las jerarquías, mediante la incorporación de un nuevo objetivo en la formulación de SVR: la minimización de las desviaciones entre los parámetros de los regresores en el último nivel y los niveles superiores. Este método lo denominamos KAT-SVR, proveniente de las siglas en inglés para *kernel-based all-together SVR for HTS*.

Recordemos que el método θ -SVR resuelve el siguiente problema de programación cuadrático:

$$\begin{aligned} \min_{\mathbf{w}_0, \mathbf{w}_i, b_i, \xi_i, \xi_i^*} \quad & \frac{1}{2} \sum_{i=1}^N (\|\mathbf{w}_i\|^2 + \theta \|\mathbf{w}_i - \mathbf{w}_0\|^2) + C \sum_{i=1}^N \mathbf{e}_i^\top (\boldsymbol{\xi}_i + \boldsymbol{\xi}_i^*) \\ \text{s.t.} \quad & y_j - (\mathbf{w}_i^\top \mathbf{x}_{ij} + b_i) \leq \varepsilon + \xi_{ij}, \quad j = 1, \dots, M_i, \quad i = 1, \dots, N, \\ & (\mathbf{w}_i^\top \mathbf{x}_{ij} + b_i) - y_j \leq \varepsilon + \xi_{ij}^*, \quad j = 1, \dots, M_i, \quad i = 1, \dots, N, \\ & \xi_{ij}, \xi_{ij}^* \geq 0, \quad j = 1, \dots, M_i, \quad i = 1, \dots, N, \end{aligned} \quad (3.7)$$

donde $\boldsymbol{\xi}_i = (\xi_{i1}, \dots, \xi_{iM_i}) \in \text{Re}^{M_i}$, $\boldsymbol{\xi}_i^* = (\xi_{i1}^*, \dots, \xi_{iM_i}^*) \in \text{Re}^{M_i}$, y $\mathbf{e}_i \in \text{Re}^{M_i}$ denota un vector de unos. Y se consideran los siguientes parámetros positivos: ε (tamaño del tubo), θ y C .

Y siendo $\mathbf{x}_{ij} \in \text{Re}^n$, $y_j \in \text{Re}$ para $j = 1, \dots, M_i$ y $i = 1, \dots, N$. Se busca un *set* de funciones lineales de la forma

$$f_i(\mathbf{x}) := \mathbf{w}_i^\top \mathbf{x} + b_i, \quad i = 1, \dots, N, \quad (3.8)$$

Para obtener un método kernel, lo primero que haremos será construir el problema dual de la formulación (3.7). Para esto, necesitamos el siguiente Lema técnico:

Lema 3.3.1. *Para $\theta > 0$, sea $\tilde{\mathcal{Q}}(\theta) = NI_{nN} + \theta\mathcal{J} \in \text{Re}^{nN \times nN}$, donde I_{nN} denota a la*

matriz identidad de orden nN y

$$\mathcal{J} = \begin{bmatrix} I_n & \cdots & I_n \\ \vdots & \ddots & \vdots \\ I_n & \cdots & I_n \end{bmatrix} \in \text{Re}^{nN \times nN}.$$

Entonces, existe una matriz simétrica $\tilde{\mathcal{Q}}(\theta)^{1/2}$ que satisface $(\tilde{\mathcal{Q}}(\theta)^{1/2})^2 = \tilde{\mathcal{Q}}(\theta)$.

Demostración. Sea

$$\tilde{\mathcal{Q}}(\theta)^{1/2} = \sqrt{N}I_{nN} - \frac{1 - \sqrt{1 + \theta}}{\sqrt{N}}\mathcal{J}. \quad (3.9)$$

Notemos que $\mathcal{J}^2 = Nn$. Entonces,

$$(\tilde{\mathcal{Q}}(\theta)^{1/2})^2 = NI_{nN} - 2(1 - \sqrt{1 + \theta})\mathcal{J} + \frac{(1 - \sqrt{1 + \theta})^2}{N}(N\mathcal{J}) = NI_{nN} + \theta\mathcal{J}.$$

□

Proposición 3.3.2. *La formulación dual del problema (3.7) está dado por*

$$\begin{aligned} \min_{\boldsymbol{\alpha}, \boldsymbol{\alpha}^*} & \frac{1}{2N(\theta + 1)} \|\tilde{\mathcal{Q}}(\theta)^{1/2} \mathbf{A}^\top (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)\|^2 - \mathbf{Y}^\top (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) + \varepsilon \mathbf{e}^\top (\boldsymbol{\alpha} + \boldsymbol{\alpha}^*) \\ \text{s.t.} & \mathbf{E}^\top (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) = \mathbf{0}, \quad \mathbf{0} \leq \boldsymbol{\alpha}, \boldsymbol{\alpha}^* \leq C\mathbf{e}, \end{aligned} \quad (3.10)$$

donde $\mathbf{Y} = [\mathbf{y}_1^\top, \mathbf{y}_2^\top, \dots, \mathbf{y}_N^\top]^\top$ con $\mathbf{y}_i = (y_1, y_2, \dots, y_{M_i}) \in \text{Re}^{M_i}$, para cada $i = 1, \dots, N$, \mathbf{A} y \mathbf{E} son definidos en la ecuación (3.15) y (3.16) más adelante, respectivamente.

Demostración. Definimos $\tilde{\mathbf{w}} = [\mathbf{w}_0^\top, \mathbf{w}_1^\top, \dots, \mathbf{w}_N^\top]^\top \in \text{Re}^{n(N+1)}$, y

$$\mathcal{Q}(\theta) = \begin{bmatrix} N\theta I_n & -\theta I_n & -\theta I_n & \cdots & -\theta I_n \\ -\theta I_n & (1 + \theta)I_n & 0 & \cdots & 0 \\ -\theta I_n & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ -\theta I_n & 0 & \cdots & 0 & (1 + \theta)I_n \end{bmatrix} \quad (3.11)$$

es una matriz de $n(N + 1) \times n(N + 1)$, donde I_n representa a la matriz identidad de

orden n . Esta matriz es simétrica definida positiva con matriz inversa dada por

$$\mathcal{Q}(\theta)^{-1} = \frac{1}{N} \begin{bmatrix} \frac{\theta+1}{\theta} I_n & I_n & I_n & \cdots & I_n \\ I_n & \frac{\theta+N}{\theta+1} I_n & \frac{\theta}{\theta+1} I_n & \cdots & \frac{\theta}{\theta+1} I_n \\ I_n & \frac{\theta}{\theta+1} I_n & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \frac{\theta}{(\theta+1)} I_n \\ I_n & \frac{\theta}{\theta+1} I_n & \cdots & \frac{\theta}{(\theta+1)} I_n & \frac{\theta+N}{\theta+1} I_n \end{bmatrix}. \quad (3.12)$$

Luego, el término cuadrático en la función objetivo del problema (3.7) se puede expresar como

$$\sum_{i=1}^N (\|\mathbf{w}_i\|^2 + \theta \|\mathbf{w}_i - \mathbf{w}_0\|^2) = \tilde{\mathbf{w}}^\top \mathcal{Q}(\theta) \tilde{\mathbf{w}}. \quad (3.13)$$

Además, denotemos $A_i = [\mathbf{x}_{i1}^\top; \mathbf{x}_{i2}^\top; \dots; \mathbf{x}_{iM_i}^\top] \in \text{Re}^{M_i \times n}$ y $\mathbf{y}_i = (y_1, y_2, \dots, y_{M_i}) \in \text{Re}^{M_i}$, para cada $i = 1, \dots, N$. Entonces, las restricciones del problema (3.7) se pueden expresar como

$$\begin{aligned} \mathbf{y}_i - A_i \mathbf{w}_i - b_i \mathbf{e}_i &\leq \varepsilon \mathbf{e}_i + \boldsymbol{\xi}_i, \quad \boldsymbol{\xi}_i \geq 0, \quad i = 1, \dots, N, \\ A_i \mathbf{w}_i + b_i \mathbf{e}_i - \mathbf{y}_i &\leq \varepsilon \mathbf{e}_i + \boldsymbol{\xi}_i^*, \quad \boldsymbol{\xi}_i^* \geq 0 \quad i = 1, \dots, N, \end{aligned} \quad (3.14)$$

donde $\mathbf{e}_i \in \text{Re}^{M_i}$ denota un vector de unos. Luego, consideramos la siguiente notación: $M = \sum_{i=1}^N M_i$, $\mathbf{Y} = [\mathbf{y}_1^\top, \mathbf{y}_2^\top, \dots, \mathbf{y}_N^\top]^\top$, $\boldsymbol{\xi} = [\boldsymbol{\xi}_1^\top, \dots, \boldsymbol{\xi}_N^\top]^\top$, $\boldsymbol{\xi}^* = [\boldsymbol{\xi}_1^\top, \dots, \boldsymbol{\xi}_N^\top]^\top$, $\mathbf{e} = [\mathbf{e}_1^\top, \dots, \mathbf{e}_N^\top]^\top \in \text{Re}^M$, $\mathbf{b} = (b_1, \dots, b_N) \in \text{Re}^N$, $\mathbb{A} = [\mathbf{0} \ \mathbf{A}] \in \text{Re}^{M \times N(n+1)}$ con

$$\mathbf{A} = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & A_N \end{bmatrix} \in \text{Re}^{M \times Nn}, \quad (3.15)$$

y

$$\mathbf{E} = \begin{bmatrix} \mathbf{e}_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{e}_N \end{bmatrix} \in \text{Re}^{M \times N}. \quad (3.16)$$

Luego, las restricciones (3.14) se pueden reescribir equivalentemente como

$$\mathbf{Y} - \mathbb{A}\tilde{\mathbf{w}} - \mathbf{E}\mathbf{b} \leq \varepsilon\mathbf{e} + \boldsymbol{\xi}, \quad \boldsymbol{\xi} \geq \mathbf{0}, \quad (3.17)$$

$$\mathbb{A}\tilde{\mathbf{w}} + \mathbf{E}\mathbf{b} - \mathbf{Y} \leq \varepsilon\mathbf{e} + \boldsymbol{\xi}^*, \quad \boldsymbol{\xi}^* \geq \mathbf{0}. \quad (3.18)$$

Sea $L : \text{Re}^{n(N+1)} \times \text{Re}^M \times \text{Re}^M \times \text{Re}^M \times \text{Re}^M \times \text{Re}^M \times \text{Re}^M \rightarrow \text{Re}$ la función Lagrangiana asociada a la formulación (3.7), dado por

$$\begin{aligned} L(\tilde{\mathbf{w}}, \boldsymbol{\xi}, \boldsymbol{\xi}^*, \boldsymbol{\alpha}, \boldsymbol{\alpha}^*, \boldsymbol{\eta}, \boldsymbol{\eta}^*) &= \frac{1}{2}\tilde{\mathbf{w}}^\top \mathcal{Q}(\theta)\tilde{\mathbf{w}} + \mathbf{C}\mathbf{e}^\top(\boldsymbol{\xi} + \boldsymbol{\xi}^*) - \boldsymbol{\eta}^\top \boldsymbol{\xi} - \boldsymbol{\eta}^{*\top} \boldsymbol{\xi}^* \\ &+ \boldsymbol{\alpha}^\top(\mathbf{Y} - \mathbb{A}\tilde{\mathbf{w}} - \mathbf{E}\mathbf{b} - \varepsilon\mathbf{e} - \boldsymbol{\xi}) \\ &+ \boldsymbol{\alpha}^{*\top}(\mathbb{A}\tilde{\mathbf{w}} + \mathbf{E}\mathbf{b} - \mathbf{Y} - \varepsilon\mathbf{e} - \boldsymbol{\xi}). \end{aligned} \quad (3.19)$$

Con esto, el problema (3.7) puede ser reescrito equivalentemente como

$$\min_{\tilde{\mathbf{w}}, \boldsymbol{\xi}, \boldsymbol{\xi}^*} \max_{\boldsymbol{\alpha}, \boldsymbol{\alpha}^*, \boldsymbol{\eta}, \boldsymbol{\eta}^*} \{L(\tilde{\mathbf{w}}, \boldsymbol{\xi}, \boldsymbol{\xi}^*, \boldsymbol{\alpha}, \boldsymbol{\alpha}^*, \boldsymbol{\eta}, \boldsymbol{\eta}^*) : \boldsymbol{\alpha}, \boldsymbol{\alpha}^*, \boldsymbol{\eta}, \boldsymbol{\eta}^* \geq \mathbf{0}\}.$$

Por eso, la formulación dual de (3.7) está dada por (remitimos al lector a (Bertsekas, 1982) para obtener más información sobre la teoría de la dualidad)

$$\max_{\boldsymbol{\alpha}, \boldsymbol{\alpha}^*, \boldsymbol{\eta}, \boldsymbol{\eta}^*} \min_{\tilde{\mathbf{w}}, \boldsymbol{\xi}, \boldsymbol{\xi}^*} \{L = L(\tilde{\mathbf{w}}, \boldsymbol{\xi}, \boldsymbol{\xi}^*, \boldsymbol{\alpha}, \boldsymbol{\alpha}^*, \boldsymbol{\eta}, \boldsymbol{\eta}^*) : \boldsymbol{\alpha}, \boldsymbol{\alpha}^*, \boldsymbol{\eta}, \boldsymbol{\eta}^* \geq \mathbf{0}\}. \quad (3.20)$$

La expresión anterior nos permite calcular el problema dual basándonos únicamente en los multiplicadores de Lagrange $\boldsymbol{\alpha}, \boldsymbol{\alpha}^*$. Las condiciones de primer orden del problema de minimización interno siguen:

$$\nabla_{\tilde{\mathbf{w}}} L = \mathcal{Q}(\theta)\tilde{\mathbf{w}} - \mathbb{A}^\top(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) = \mathbf{0}, \quad (3.21)$$

$$\nabla_{\mathbf{b}} L = -\mathbf{E}^\top(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) = \mathbf{0}, \quad (3.22)$$

$$\nabla_{\boldsymbol{\xi}} L = \mathbf{C}\mathbf{e} - \boldsymbol{\alpha} - \boldsymbol{\eta} = \mathbf{0}, \quad (3.23)$$

$$\nabla_{\boldsymbol{\xi}^*} L = \mathbf{C}\mathbf{e} - \boldsymbol{\alpha}^* - \boldsymbol{\eta}^* = \mathbf{0}. \quad (3.24)$$

Usando (3.21)-(3.24) en (3.19), se obtiene que

$$L = -\frac{1}{2}\|\mathcal{Q}(\theta)^{1/2}\tilde{\mathbf{w}}\|^2 + \mathbf{Y}^\top(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) - \varepsilon\mathbf{e}^\top(\boldsymbol{\alpha} + \boldsymbol{\alpha}^*).$$

Dado que $\mathcal{Q}(\theta)$ es no singular (ver proposici3n 3.2.1), sigue desde (3.21) que la expresi3n anterior puede ser escrita como

$$L = -\frac{1}{2}\|\mathcal{Q}(\theta)^{-1/2}\mathbb{A}^\top(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)\|^2 + \mathbf{Y}^\top(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) - \varepsilon\mathbf{e}^\top(\boldsymbol{\alpha} + \boldsymbol{\alpha}^*). \quad (3.25)$$

Usando la relaci3n (3.12), el lema 3.3.1 y la definici3n de \mathbb{A} , la expresi3n (3.25) se reduce a

$$L = -\frac{1}{2N(\theta + 1)}\|\tilde{\mathcal{Q}}(\theta)^{1/2}\mathbf{A}^\top(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)\|^2 + \mathbf{Y}^\top(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) - \varepsilon\mathbf{e}^\top(\boldsymbol{\alpha} + \boldsymbol{\alpha}^*). \quad (3.26)$$

Por otro lado, dado que $\boldsymbol{\eta}, \boldsymbol{\eta}^* \geq \mathbf{0}$, en (3.23) y (3.24), sigue que $\mathbf{0} \leq \boldsymbol{\alpha}, \boldsymbol{\alpha}^* \leq C\mathbf{e}$. Por tanto, la formulaci3n dual se deduce de estas desigualdades y de las expresiones (3.20), (3.22) y (3.26). \square

Ahora, relacionamos las variables primales y duales de las formulaciones (3.7) y (3.10) con la proposici3n 3.3.3.

Proposici3n 3.3.3. *Supongamos que $\boldsymbol{\alpha}, \boldsymbol{\alpha}^* \in \text{Re}^M$ son soluciones del problema (3.10). Entonces, las solucines $\mathbf{w}_0, \mathbf{w}_i$ relacionadas al problema (3.7) est3n dadas por*

$$\mathbf{w}_0 = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^{M_i} \mathbf{x}_{ij}(\alpha_{ij} - \alpha_{ij}^*), \quad (3.27)$$

y

$$\mathbf{w}_i = \frac{1}{\theta + 1} \left(\sum_{j=1}^{M_i} \mathbf{x}_{ij}(\alpha_{ij} - \alpha_{ij}^*) + \frac{\theta}{N} \sum_{r=1}^N \sum_{j=1}^{M_r} \mathbf{x}_{rj}(\alpha_{rj} - \alpha_{rj}^*) \right), \quad i = 1, \dots, N. \quad (3.28)$$

Adem3s, las funciones de regresi3n lineal obtenidas con el problema (3.7) pueden ser redefinidas en base a las variables duales de la siguiente manera:

$$f_i(\mathbf{x}) = \frac{1}{\theta + 1} \left(\sum_{j=1}^{M_i} (\alpha_{ij} - \alpha_{ij}^*) \mathbf{x}_{ij}^\top \mathbf{x} + \frac{\theta}{N} \sum_{r=1}^N \sum_{j=1}^{M_r} (\alpha_{rj} - \alpha_{rj}^*) \mathbf{x}_{rj}^\top \mathbf{x} \right) + b_i, \quad (3.29)$$

para $i = 1, \dots, N$.

Demostraci3n. Usando (3.21) y (3.12), se obtienen expresiones que relacionan las variables primales y duales de las formulaciones (3.7) y (3.10), respectivamente. Estas

expresiones están dadas por

$$\mathbf{w}_0 = \frac{1}{N} \sum_{i=1}^N A_i^\top (\boldsymbol{\alpha}_i - \boldsymbol{\alpha}_i^*), \quad \mathbf{w}_i = \frac{1}{\theta + 1} (A_i^\top (\boldsymbol{\alpha}_i - \boldsymbol{\alpha}_i^*) + \theta \mathbf{w}_0), \quad i = 1, \dots, N, \quad (3.30)$$

donde $\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_i^* \in \text{Re}^{M_i}$ para $i = 1, \dots, N$. Reemplazando el valor de A_i , tenemos que

$$\mathbf{w}_0 = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^{M_i} \mathbf{x}_{ij} (\alpha_{ij} - \alpha_{ij}^*),$$

y

$$\mathbf{w}_i = \frac{1}{\theta + 1} \left(\sum_{j=1}^{M_i} \mathbf{x}_{ij} (\alpha_{ij} - \alpha_{ij}^*) + \frac{\theta}{N} \sum_{r=1}^N \sum_{j=1}^{M_r} \mathbf{x}_{rj} (\alpha_{rj} - \alpha_{rj}^*) \right), \quad i = 1, \dots, N.$$

Finalmente, reemplazando la expresión anterior en (3.8), se obtiene la función de regresión dada en (3.29). \square

Observación 3.3.4. *Se obtiene de (3.30) que*

$$\mathbf{w}_0 = \frac{1}{N} \sum_{i=1}^N \mathbf{w}_i. \quad (3.31)$$

La siguiente observación presenta la condiciones de Karush-Kuhn-Tucker (KKT) para el problema (3.7) que resultan de las proposiciones previas.

Observación 3.3.5. *Similar al enfoque tradicional de SVR (Drucker, Burges, Kaufman, Smola & Vapnik, 1997), el desplazamiento b_i se puede reescribir en términos de las variables duales utilizando las condiciones complementarias KKT del problema (3.7).*

Para una serie de tiempo dada $i \in \{1, \dots, N\}$, estas condiciones son:

$$\begin{aligned} (y_j - \mathbf{w}_i^\top \mathbf{x}_{ij} - b_i - \varepsilon - \xi_{ij}) \alpha_{ij} &= 0, \\ (\mathbf{w}_i^\top \mathbf{x}_{ij} + b_i - y_j - \varepsilon - \xi_{ij}^*) \alpha_{ij}^* &= 0, \\ \xi_{ij} \eta_{ij} = \xi_{ij}^* \eta_{ij}^* &= 0, \\ \alpha_{ij}, \alpha_{ij}^*, \xi_{ij}, \xi_{ij}^*, \eta_{ij}, \eta_{ij}^* &\geq 0. \end{aligned}$$

Ya que se tiene que $\eta_{ij} = C - \alpha_{ij}$ y $\eta_{ij}^* = C - \alpha_{ij}^*$ (ver ecuaciones (3.23), (3.24)), tenemos que si hay una muestra $j \in \{1, \dots, M_i\}$ tal que $\alpha_{ij} \in (0, C)$, entonces $b_i = y_j - \mathbf{w}_i^\top \mathbf{x}_{ij} - \varepsilon$.

Alternativamente, si hay una muestra $j \in \{1, \dots, M_i\}$ tal que $\alpha_{ij}^* \in (0, C)$, entonces $b_i = y_j - \mathbf{w}_i^\top \mathbf{x}_{ij} + \varepsilon$.

3.3.1. Formulación del modelo basado en kernel

El método KAT-SVR propuesto resulta de la incorporación de funciones kernel en la formulación (3.10). Primero, reescribimos el término cuadrático de la función objetivo como:

$$\left\| \tilde{\mathcal{Q}}(\theta)^{1/2} \mathbf{A}^\top (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) \right\|^2 = \sum_{i=1}^N \sum_{j=1}^N (\boldsymbol{\alpha}_i - \boldsymbol{\alpha}_i^*)^\top \Phi_{ij}(\theta) (\boldsymbol{\alpha}_j - \boldsymbol{\alpha}_j^*), \quad (3.32)$$

donde $\Phi_{ij}(\theta)$ denota a una matriz de $M_i \times M_j$ definida por

$$\Phi_{ii}(\theta) = (N + \theta) A_i A_i^\top, \quad \Phi_{ij}(\theta) = \theta A_i A_j^\top, \quad \text{if } i \neq j, \quad (3.33)$$

con A_i definida por $A_i = [\mathbf{x}_{i1}^\top; \mathbf{x}_{i2}^\top; \dots; \mathbf{x}_{iM_i}^\top] \in \text{Re}^{M_i \times n}$.

Basado en las definiciones anteriores, se sigue que los elementos de las matrices $\Phi_{ij}(\theta)$ están dados por

$$[\Phi_{ij}(\theta)]_{ls} = \begin{cases} (N + \theta) \mathbf{x}_{il}^\top \mathbf{x}_{js}, & i = j \\ \theta \mathbf{x}_{il}^\top \mathbf{x}_{js}, & i \neq j \end{cases}, \quad l = 1, \dots, M_i, \quad s = 1, \dots, M_j. \quad (3.34)$$

Dado que los conjuntos de entrenamiento en la ecuación (3.34) aparecen sólo como productos internos de la forma $\mathbf{x}_{il}^\top \mathbf{x}_{js}$, podemos aplicar el truco del kernel, reemplazándolos por $K(\mathbf{x}_{il}, \mathbf{x}_{js})$, donde $K: \text{Re}^n \times \text{Re}^n \rightarrow \text{Re}$ es cualquier función que satisface la condición de Mercer (Hastie, Tibshirani & Friedman, 2009; Maldonado & López, 2016).

Luego de seleccionar una función kernel, la relación (3.34) puede ser reescrita como

$$[\Phi_{ij}^K(\theta)]_{ls} = \begin{cases} (N + \theta) K(\mathbf{x}_{il}, \mathbf{x}_{js}), & i = j \\ \theta K(\mathbf{x}_{il}, \mathbf{x}_{js}), & i \neq j \end{cases}, \quad l = 1, \dots, M_i, \quad s = 1, \dots, M_j. \quad (3.35)$$

Por lo tanto, considerando la ecuación (3.35), la formulación de KAT-SVR para HTS estará dada por el siguiente problema de programación cuadrático:

$$\begin{aligned}
\min_{\boldsymbol{\alpha}, \boldsymbol{\alpha}^*} & \frac{1}{2N(\theta + 1)} \sum_{i,j=1}^N (\boldsymbol{\alpha}_i - \boldsymbol{\alpha}_i^*)^\top \Phi_{ij}^K(\theta) (\boldsymbol{\alpha}_j - \boldsymbol{\alpha}_j^*) - \sum_{i=1}^N \mathbf{y}_i^\top (\boldsymbol{\alpha}_i - \boldsymbol{\alpha}_i^*) \\
& + \varepsilon \sum_{i=1}^N \mathbf{e}_i^\top (\boldsymbol{\alpha}_i + \boldsymbol{\alpha}_i^*) \\
\text{s.t. } & \mathbf{e}_i^\top (\boldsymbol{\alpha}_i - \boldsymbol{\alpha}_i^*) = 0, \mathbf{0} \leq \boldsymbol{\alpha}_i, \boldsymbol{\alpha}_i^* \leq C\mathbf{e}_i, \quad i = 1, \dots, N.
\end{aligned} \tag{3.36}$$

Finalmente, se obtiene desde la ecuación (3.28) que las funciones de regresión tienen la siguiente forma:

$$f_i(\mathbf{x}) = \frac{1}{\theta + 1} \left(\sum_{j=1}^{M_i} (\alpha_{ij} - \alpha_{ij}^*) K(\mathbf{x}_{ij}, \mathbf{x}) + \frac{\theta}{N} \sum_{r=1}^N \sum_{j=1}^{M_r} (\alpha_{rj} - \alpha_{rj}^*) K(\mathbf{x}_{rj}, \mathbf{x}) \right) + b_i, \tag{3.37}$$

para $i = 1, \dots, N$.

Observación 3.3.6. *El modelo clásico ε -SVR (en su versión dual, ver (2.20)) resuelve un problema de programación cuadrático de orden $2m$ y posee una complejidad temporal de $\mathcal{O}((2m)^3)$. Además, notamos que la formulación dual (3.36) posee una estructura similar a (2.20), es decir, es un problema de programación cuadrático de orden $2M$. Luego, su complejidad temporal es $\mathcal{O}((2M)^3)$, donde $M = \sum_{i=1}^N M_i$. Por otro lado, esta propuesta puede ser resuelta eficientemente, por ejemplo, mediante una optimización mínima secuencial (SMO) (Platt, 1998) o bien, por métodos de descomposición (Osuna, Freund & Girosi, 1997).*

3.3.2. Aplicación a series de tiempo jerárquicas

Existen varios contextos en los que se podría utilizar la formulación (3.36) para predicción de series de tiempo jerárquicas. Al igual que θ -SVR, KAT-SVR corresponde a una variación del enfoque bottom-up, en el que cada función de regresión $f_i(\mathbf{x})$, para $i = 1, \dots, N$, entrega una estimación para una serie de tiempo del último nivel de la jerarquía. Para agrupar información a través de los niveles de la jerarquía, debemos elegir un nivel superior de pivote entre 0 y $J - 1$.

Nosotros consideraremos dos variantes para el enfoque KAT-SVR, al igual que lo hici-

mos con θ -SVR. La primera variante, KAT-SVR_J^0 , usa el nodo raíz como pivote (nivel 0). La segunda variante, KAT-SVR_J^{J-1} , al contrario, usa el nivel $J - 1$ como pivote. Con esto, KAT-SVR_J^0 construye todas las funciones en el último nivel, y una superficie de decisión para el nivel 0. Por otro lado, KAT-SVR_J^{J-1} se enfoca en una rama específica del árbol jerárquico, usando un nodo en el nivel $J - 1$ y construyendo todas las funciones de decisión en el nivel J que se derivan de este nodo. En estricto rigor, la versión KAT-SVR_J^{J-1} no construye todas las funciones de decisión de forma simultánea, y utiliza la formulación (3.36) en cada nodo del nivel $J - 1$. Esto lo lleva a dividir la complejidad en problemas más pequeños. De todas formas, los modelos HTS tienden a ser pequeños en términos de tamaño de muestra, lo que lleva a un entrenamiento rápido independientemente del enfoque utilizado.

Volviendo al ejemplo 2.2, en que se muestra una estructura de ejemplo para una serie de tiempo jerárquica, KAT-SVR_J^0 considera la serie “Total” (nodo raíz) como pivote, resolviendo un único problema de optimización y obteniendo las cinco funciones de regresión del último nivel. Por otro lado, KAT-SVR_J^{J-1} es entrenado utilizando ambos nodos del pen-último nivel (A y B), teniendo que resolver dos problemas de optimización. Esta diferencia se ilustra en la figura 3.1.

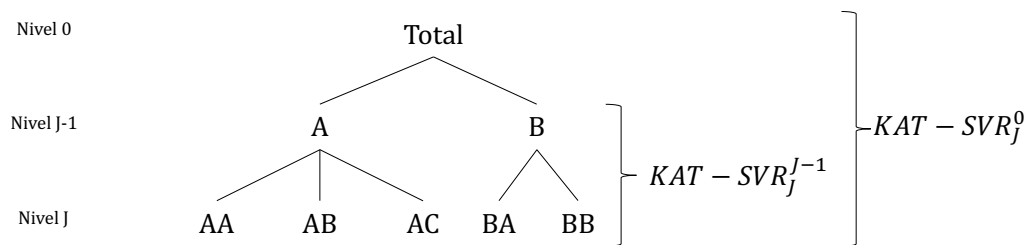


Figura 3.1: Diferencia entre los métodos KAT-SVR_J^0 y KAT-SVR_J^{J-1} .

Cabe destacar que ambos modelos generan predicciones para el último nivel, y que las predicciones de los niveles superiores vienen dadas por una agregación tradicional, similar a lo estudiado en el algoritmo SVR-BU. En otras palabras, las funciones pivote construidas por los métodos KAT-SVR_J^0 y KAT-SVR_J^{J-1} no se consideran con fines predictivos.

3.4. Machine Learning como *input* para *staffing*

La generación de modelos predictivos permite el enriquecimiento de las fuentes de información disponibles, y la generación de nuevos indicadores, que apoyen a los tomadores de decisiones a tomar decisiones con menor incertidumbre. El siguiente trabajo de esta tesis, corresponde a un trabajo en el que se generan predicciones para series de tiempo jerárquicas mediante modelos de SVR. En este caso, se cuenta con datos de una empresa proveedora de plataformas de Contact Center, y los datos corresponden a la cantidad de llamadas de teléfono entrantes (*inbound*) hacia los Contact Center de sus clientes (empresas), en periodos definidos de tiempo.

La estructura jerárquica de estos datos viene dada por características propias de las empresas, como por ejemplo, sus distintas áreas o departamentos. Un ejemplo de esto se presenta en la figura 3.2.

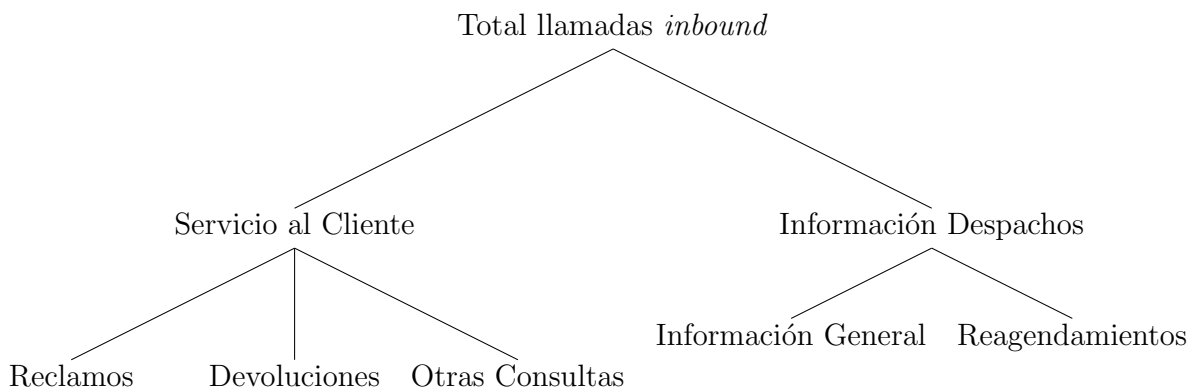


Figura 3.2: Ejemplo de jerarquía para un Contact Center.

Por otro lado, además de los rezagos propios de cada serie de tiempo, existen otras variables interesantes para entender los problemas de la cantidad de llamadas que recibe un Contact Center. Por ejemplo, si la empresa corresponde a un retail, habrá más llamadas relacionadas a ventas para un *Cyberday*, o si las llamadas ingresan a un servicio de urgencias médicas, sin duda se producirá un *peak* de llamadas en medio de una pandemia.

Lo que buscan las empresas es ser capaces de encontrar la distribución óptima de sus ejecutivos, persiguiendo metas de niveles de atención y de servicio mensuales. Es muy importante para ellos, por ejemplo, que la menor cantidad de llamadas posible abandonen en una cola de atención antes de ser atendidas por un ejecutivo; o por otro

lado, que sus clientes esperen el menor tiempo posible en una cola de atención antes de ser atendidos.

A continuación, se busca incorporar investigación de operaciones a un modelo predictivo, buscando mejorar la interpretación de las soluciones, apoyando a reducir la incertidumbre, y con ello, realizar mejores decisiones de negocio.

3.4.1. Obtención de Intervalos de Confianza a partir de métodos de Machine Learning

Consideremos entonces el problema de la cantidad de llamadas entrantes (*inbound*) que recibe una empresa. Mediante un modelo predictivo, como los presentados en las secciones anteriores, se puede realizar una estimación de demanda, para luego utilizar este resultado como *input* de un modelo de *staffing*. Notemos que la demanda estimada, al corresponder a una estimación, tendrá un margen de error, por lo que utilizaremos los intervalos de confianza de nuestras soluciones para la generación de escenarios de un modelo de optimización estocástico.

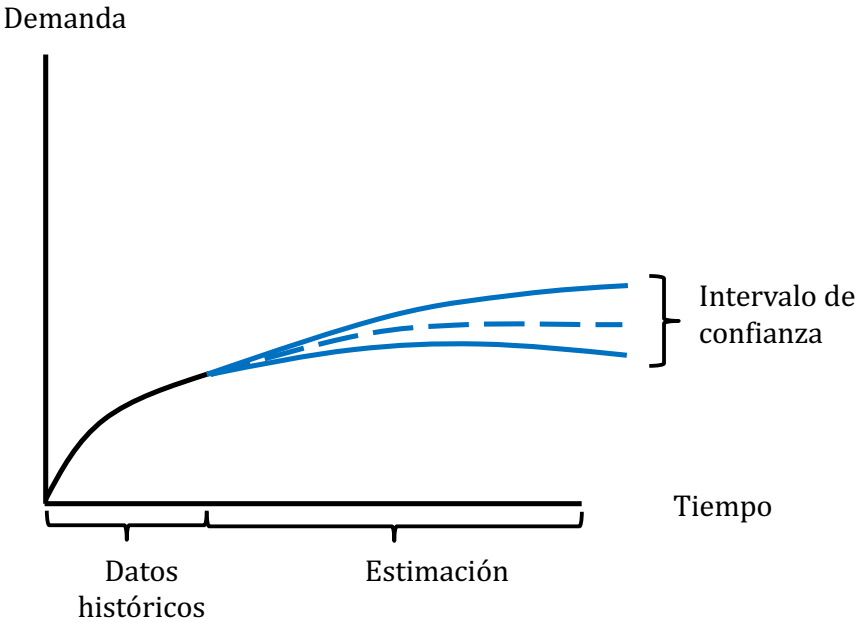


Figura 3.3: Resultado de los modelos de estimación de demanda.

En el caso de los modelos ARIMA, la obtención de los intervalos de confianza es mucho más directa que en los problemas de Machine Learning, dado que la metodología ARIMA asume que el error respeta una distribución normal de probabilidades. Con ello, se puede estimar un intervalo de confianza basado en dicha distribución de probabilidad. Este supuesto no se puede tomar en el caso de los métodos de Machine Learning, ya que la distribución de los errores no respeta necesariamente una distribución normal. Para la obtención de estos intervalos en modelos de Machine Learning se proponen dos metodologías. La primera considera la utilización de los residuos de los procesos de *bootstrap*, y la segunda utilizando modelos de regresión cuantil (Hyndman & Athanasopoulos, 2021). Cabe destacar que la obtención de intervalos de confianza para metodologías de HTS debe ser estudiada en detalle. Si bien es clara su utilización para las metodologías que utilizan el enfoque bottom-up, los cuales corresponden a los modelos que mejor desempeño obtuvieron en nuestra investigación. Sería importante tener en cuenta el cómo funciona la propagación de los intervalos en la estructura jerárquica para enfoques como top-down y middle-out.

Intervalos de confianza a partir de los residuos de los procesos de *bootstrap*

El error e_t asociado a la estimación de una observación futura se puede estimar como:

$$e_t = y_t - \hat{y}_t, \quad (3.38)$$

lo que es equivalente a decir que el valor real y_t corresponde a la estimación basada en el periodo anterior \hat{y}_t más un término de error, como sigue:

$$y_t = \hat{y}_t + e_t. \quad (3.39)$$

Si se asume que el error pasado explica a los errores futuros, la idea es, iterativamente, predecir la siguiente observación, generando una serie de predicciones y sus respectivos errores, que representan la varianza esperada en el proceso predictivo. Con esto, los intervalos de confianza pueden ser calculados al estimar los percentiles $\alpha/2$ y $1 - \alpha/2$ para las predicciones simuladas (Hyndman & Athanasopoulos, 2021).

La principal ventaja de este método es que sólo se necesita un modelo de Machine Learning para estimar cualquier intervalo. Sin embargo, este tipo de algoritmos puede

resultar bastante costoso en términos computacionales (Hyndman & Athanasopoulos, 2021).

Intervalos de confianza a partir de modelos de regresión cuantil

La regresión por cuantiles tiene como objetivo estimar los cuantiles condicionales de la variable de respuesta. Para una función de distribución continua, el cuantil α , $Q_\alpha(x)$, se define de tal forma que la probabilidad de que Y sea menor que $Q_\alpha(x)$ para un valor de $X = x$ dado sea igual a α .

En este caso, se puede construir un intervalo mediante la combinación de las predicciones de dos regresiones por cuantiles. En este caso, por ejemplo, un intervalo de confianza del 80 % estará dado por las predicciones de los modelos para los cuantiles 0,1 y 0,9 respectivamente (Hyndman & Athanasopoulos, 2021).

Si comparamos esta metodología con el algoritmo basado en *bootstrap*, la gran desventaja del algoritmo de cuantiles es que no es directamente aplicable a cualquier metodología de Machine Learning (Hyndman & Athanasopoulos, 2021).

3.4.2. Variables de decisión y otros parámetros

Lo primero que debemos hacer es definir las variables de decisión involucradas en este problema. En este caso se considerará $X_t \in \mathbb{Z}_+$, $\forall t \in \{1, \dots, T\}$, como la cantidad de posiciones de Contact Center que se contratarán en el periodo t . La otra variable que consideraremos, dado que estamos generando un modelo para una empresa proveedora de plataformas y servicios de Contact Center, es el precio en que podrán vender cada posición de la plataforma en el periodo t , y la definiremos como $P_t \in \mathbb{R}_+$, $\forall t$.

Además, definiremos las ventas en el periodo t como $V_t \in \mathbb{Z}_+$. Notemos que estas ventas las podemos definir como una función de la demanda, como sigue:

$$V_t = d_t + \min \{0, X_t - d_t\}, \quad (3.40)$$

donde d_t corresponde a la demanda en el periodo t .

Por otro lado, sabemos que cada posición de Contact Center que contratemos tendrá un costo, al cuál denominaremos $c_t \in \mathbb{R}_+$, correspondiendo al costo de la posición de Contact Center en el periodo t .

Otro costo que se debe considerar es el de no satisfacer la demanda. Para esto, definiremos la demanda que no es abastecida en el periodo t como $y_t \in \mathbb{Z}_+$, la cual puede expresarse como

$$y_t = \max \{0, d_t - X_t\}. \quad (3.41)$$

Al costo por no satisfacer la demanda en el periodo t lo denominaremos $\mu_t \in \mathbb{R}_+$.

3.4.3. Función Objetivo

Notemos que la función objetivo de nuestro problema vendrá dada por la maximización de las utilidades de la empresa, que se puede expresar como la diferencia entre los ingresos y los costos.

Los ingresos I los podemos definir como la multiplicación de las ventas por su respectivo precio

$$I = \sum_t P_t V_t, \quad (3.42)$$

y por otro lado, los costos corresponden a la suma de todos los costos asociados a la operación. En este caso, los costos C corresponden a

$$C = \sum_t \{c_t X_t + \mu_t y_t\}. \quad (3.43)$$

Así, podremos definir nuestra función objetivo como

$$\max_{X_t, P_t} \mathbb{E}_{d_t} \left\{ \sum_t (P_t V_t - c_t X_t - \mu_t y_t) \right\}, \quad (3.44)$$

donde los escenarios se construyen en base a los intervalos de confianza generados con las estimaciones de demanda. Notemos que hasta este punto tenemos un problema separable en el tiempo $t \in 1, \dots, T$. Por lo tanto, tendremos T problemas $P^{[t]}$ de la misma estructura:

$$P^{[t]} : \max_{X_t, P_t} \mathbb{E}_{d_t} \{P_t V_t - c_t X_t - \mu_t y_t\}. \quad (3.45)$$

El caso separable es equivalente a la resolución de T problemas del tipo *Newsvendor Problem*. Esto no se cumpliría en el caso de la existencia de una restricción presupuesta-

ria, dejando sin validez la formulación (3.45). En el caso de la existencia de presupuesto, los problemas se enlazan intertemporalmente. Esto se puede abordar con enfoques de control automático.

3.4.4. Configuraciones Experimentales

El principal objetivo del modelo propuesto anteriormente corresponde a la definición de la cantidad de posiciones de Contact Center a contratar para un periodo t , dado que se conoce la estimación de la demanda d_t para dicho periodo, y que, por otro lado, se construyen intervalos de confianza de dichas estimaciones, para poder manejar escenarios en nuestro modelo de optimización propuesto.

La demanda d_t proviene, en este caso, de las predicciones de un modelo ε -SVR para las series de tiempo jerárquicas, cuyas jerarquías están dadas en base a distintas áreas de la empresa y sus respectivas líneas de entrada asociadas (ver más detalles de los datos en apéndice A). En este caso, se tienen datos en intervalos cada 15 minutos, pero por la naturaleza de la decisión, se decide usar agregaciones diarias para las estimaciones y posterior definición de la variable de decisión. Con esto, los periodos $t \in (1, \dots, T)$ corresponden a días. La predicción se hace para 30 días futuros, con el objetivo de que nuestro modelo genere políticas de *staffing* de forma mensual.

Cabe destacar que, si bien puede existir el caso de que alguna empresa no posea restricción presupuestaria para la contratación de posiciones de Contact Center, lo más común será que esta restricción si esté activa, por lo que se deben plantear ambos problemas. En el caso de no existir restricción presupuestaria, el problema de optimización será más sencillo, y separable en el tiempo, como fue planteado en la formulación (3.45), pero en el caso de existir dicha restricción, se deberá resolver directamente la formulación (3.44).

Por otro lado, ciertas empresas podrían tener otro tipo de restricciones, distintas a las restricciones presupuestarias, como por ejemplo una diferenciación por tipo de llamada, dependiendo de la línea de entrada que esta posea, o bien un nivel de servicio mínimo aceptado (tiempo máximo que se puede tener a los clientes esperando en colas de atención) para ciertas áreas de la empresa. Todo este tipo de restricciones complejizan el modelo, evitando que este sea separable en el tiempo.

Capítulo 4

Resultados y Análisis

Lo primero es definir la medida de desempeño que utilizaremos para evaluar los modelos predictivos. En este caso, principalmente se utilizan dos medidas: el Error Porcentual Absoluto Medio (MAPE) y la Raíz del Error Cuadrático Medio (RMSE). Estas medidas se utilizan para determinar el desempeño de cada nodo, es decir, para cada serie de tiempo de la jerarquía. Además, se computa un MAPE o RMSE promedio para cada nivel, que considera el error promedio obtenido para todos los nodos de dicho nivel. Para cada conjunto de datos, se genera un ranking de 1 (mejor) a M (peor), donde M corresponde a la cantidad total de modelos a comparar. Este ranking se genera en relación al desempeño del modelo predictivo para cada nivel, con lo que finalmente el desempeño global del modelo se estima como el promedio de los rankings obtenidos en todos los niveles de la jerarquía. Este ranking promedio para cada modelo m lo denominaremos \bar{R}_m , y se calcula como:

$$\bar{R}_m = \sum_{l=1}^L \frac{R_{m,l}}{L}, \quad (4.1)$$

donde $R_{m,l}$ corresponde al ranking obtenido por el modelo m en el nivel l y L es el número total de niveles de dicha jerarquía. El ranking global se estima como el promedio de todos los rankings \bar{R}_m obtenidos por el modelo m en todos los conjuntos de datos.

4.1. Predicción de HTS de venta de productos de la industria del Travel Retail vía SVR

En esta sección se presentan en detalle los resultados obtenidos por los diferentes algoritmos de SVR para HTS propuestos, para la predicción en algunos conjuntos de datos de prueba, como también para datos reales de la industria del Travel Retail. Se puede consultar una descripción detallada de estos datos en el apéndice A. Por otro lado, en el apéndice B se encuentran todas las configuraciones experimentales consideradas para cada conjunto de datos.

En este primer trabajo, para cada conjunto de datos, se probaron los tres enfoques clásicos de HTS: bottom-up, top-down y middle-out. Cada uno de ellos se utilizó en complemento del modelo predictivo, pudiendo ser ε -SVR (metodología propuesta), ARIMA o Holt-Winters. Con esto, se reportan resultados de los siguientes modelos:

- SVR-BU: El mejor modelo entre SVR con kernel lineal o gaussiano, con el enfoque bottom-up.
- SVR-TD: El mejor modelo entre SVR con kernel lineal o gaussiano, con el enfoque top-down.
- SVR-MO: El mejor modelo entre SVR con kernel lineal o gaussiano, con el enfoque middle-out y el mejor entre todos los niveles iniciales posibles.
- ARIMA-BU: El modelo construido con el método tradicional ARIMA y el enfoque bottom-up.
- ARIMA-TD: El modelo construido con el método tradicional ARIMA y el enfoque top-down.
- ARIMA-MO: El mejor modelo entre todos los construidos con el método tradicional ARIMA y el enfoque middle-out, con todos los niveles de inicio posibles.
- HW-BU: El modelo construido con el método tradicional Holt-Winters y el enfoque bottom-up.
- HW-TD: El modelo construido con el método tradicional Holt-Winters y el enfoque top-down.

- HW-MO: El mejor modelo entre todos los construidos con el método tradicional Holt-Winters y el enfoque middle-out, con todos los niveles de inicio posibles.

Para el primer trabajo se cuenta con cuatro conjuntos de datos jerárquicos *benchmark*, y seis de la empresa sueca Generation Research AB, que consideran la venta de productos en la industria del Travel Retail. Para cada uno de estos conjuntos de datos se construyeron los nueve modelos mencionados anteriormente, calculando su respectivo MAPE por nivel, para obtener sus respectivos rankings. El desempeño detallado de cada modelo para cada uno de los diez conjuntos de datos se encuentra disponible en el apéndice C. El ranking global para esta investigación se presenta en el cuadro 4.1, en donde se puede apreciar el ranking promedio por modelo y por conjunto de datos. El ranking global se estima como el promedio de todos los rankings \bar{R}_m obtenidos por el modelo m en todos los conjuntos de datos.

Modelo	Austourism	Htseg2	Departures	Melsyd	Other	Gender	Market	Company	LuxuryS	LuxuryC	Ranking Global
1 SVR-BU	1	4.75	4	3.33	2.33	3.33	2.8	2.8	2.67	2.83	2.99
2 HW-BU	3	3.75	1.67	8.67	2.33	2.33	3.6	1.8	2.33	1.67	3.12
3 HW-MO	4.67	3.75	2	7.67	5.67	5.67	6	2	4.83	2	4.43
4 SVR-TD	5.33	8.25	7	4.33	2.67	3	2.6	4.4	2.83	4.5	4.49
5 SVR-MO	4	4.5	6.33	6.33	5	4.33	3.2	5.6	4.5	5.33	4.91
6 ARIMA-BU	6	4	4	4.67	4.67	4	5	6	5	6	4.93
7 ARIMA-MO	5.33	2.25	6.33	5.67	5.67	5.67	6.4	6	5.83	6.17	5.53
8 ARIMA-TD	9	6.25	7.67	1.33	7.67	7.67	7	7.6	8	7.67	6.99
9 HW-TD	6.67	7.5	6	3	9	9	8.4	8.8	9	8.83	7.62

Cuadro 4.1: Resumen de resultados de la primera investigación.

Se puede observar que, en promedio, el mejor desempeño lo alcanza el modelo ε -SVR con el enfoque bottom-up, seguido de Holt-Winters con el mismo enfoque. Si realizamos una comparación de todos los enfoques de HTS, se ve que ε -SVR no sólo es mejor con la estrategia bottom-up, sino que también es superior a ARIMA y Holt-Winters utilizando el enfoque top-down. En el caso de middle-out, se aprecia como Holt-Winters es levemente superior al resto de los modelos, especialmente a SVR-MO. Con esto se demuestran las ventajas del modelamiento de patrones no lineales para la estimación de series de tiempo. También podemos notar que en promedio, la estrategia bottom-up funciona mejor que los enfoques top-down y middle-out, para la mayoría de los modelos. Los mejores parámetros reportados para los modelos de ε -SVR para cada conjunto de datos se encuentran disponibles en el apéndice D.

4.2. θ -SVR: mejoras al modelo SVR tradicional

En la segunda investigación, se busca mejorar el desempeño de los modelos ε -SVR mediante nuestra nueva propuesta: θ -SVR. En este caso, las dos variantes del modelo que introdujimos en el capítulo anterior se incluyen a los nueve modelos trabajados en la investigación previa, teniendo ahora un total de once modelos para la comparación. Los nuevos modelos que se incorporan son:

- θ -SVR $_{\{J,0\}}$: Variante que relaciona los niveles 0 y J .
- θ -SVR $_{\{J,J-1\}}$: Variante que relaciona los niveles $J - 1$ y J .

Para este caso, se utilizaron cinco bases de datos *benchmark*, y se consideró el MAPE como medida de desempeño para generar los rankings. A continuación se reportan cinco cuadros, cada uno con los resultados de los once modelos para una de las bases de datos. Notemos que dependiendo de la estructura jerárquica de cada conjunto de datos, la cantidad de niveles en cada base puede variar, denominando L0 al nivel 0, L1 al nivel 1, L2 al nivel 2, y L3 al nivel 3. El detalle de cada conjunto de datos y la respectiva configuración de parámetros para los modelos propuestos se presenta en los apéndices A y B respectivamente.

El cuadro 4.2 muestra los resultados de la base de datos ‘Htseg1’. Podemos notar que el mejor modelo en este caso es nuestra propuesta θ -SVR $_{\{J,0\}}$, seguido de ARIMA con la estrategia middle-out. En general, para este caso el desempeño del resto de los enfoques de SVR no es muy bueno, y esto se puede explicar por que este conjunto de datos contiene datos simulados, creados para probar métodos lineales, como ARIMA o Holt-Winters. Esto nos lleva a concluir que la ventaja esperada de usar modelos no lineales sea insignificante.

	Modelo	MAPE			RANKING			\bar{R}_m
		L0	L1	L2	L0	L1	L2	
1	θ -SVR $_{\{J,0\}}$	0.12	0.45	0.48	1	1	1	1
2	ARIMA-MO	0.51	0.52	0.72	4	2	2	2.67
3	SVR-BU	0.2	1.16	1.24	2	6	4	4
4	ARIMA-BU	0.59	0.6	0.77	6	3	3	4
5	ARIMA-TD	0.51	0.62	1.62	5	4	6	5
6	SVR-MO	0.22	1.34	1.74	3	7	7	5.67
7	SVR-TD	0.82	0.88	1.35	7	5	5	5.67
8	θ -SVR $_{\{J,J-1\}}$	0.85	1.57	1.78	8	8	10	8.67
9	HW-MO	1.8	1.81	1.78	10	9	8	9
10	HW-BU	1.8	1.81	1.78	11	10	9	10
11	HW-TD	1.8	1.87	2	9	11	11	10.33

Cuadro 4.2: Resultados detallados de θ -SVR para la base de datos ‘Htseg1’.

La siguiente base de datos *benchmark*, ‘Htseg2’, tiene resultados similares al caso anterior, ya que SVR no demuestra superioridad al ser comparado contra los métodos tradicionales de la literatura, principalmente debido a las características lineales del conjunto de datos. Para este caso se puede observar que el mejor modelo corresponde a ARIMA con el enfoque middle-out, seguido de cerca por nuestro modelo propuesto θ -SVR $_{\{J,0\}}$. Estos resultados se pueden ver en el cuadro 4.3.

Modelo		MAPE				RANKING				\bar{R}_m
		L0	L1	L2	L3	L0	L1	L2	L3	
1	ARIMA-MO	0.23	2.52	2.34	7.58	2	3	2	4	2.75
2	θ -SVR $_{\{J,0\}}$	0.73	3.24	2.48	4.5	4	5	3	1	3.25
3	HW-BU	10.11	2.45	2.07	8.36	8	2	1	7	4.5
4	HW-MO	2.24	2.65	2.59	8.22	5	4	4	6	4.75
5	ARIMA-BU	0.23	3.78	3.05	7.82	3	7	5	5	5
6	SVR-MO	4.9	1.59	8.53	29.21	6	1	8	8	5.75
7	SVR-BU	12.63	3.49	3.53	6.49	9	6	6	2	5.75
8	θ -SVR $_{\{J,J-1\}}$	8.22	6.1	8.18	6.96	7	8	7	3	6.25
9	ARIMA-TD	0.23	318.3	355.3	448.3	1	10	10	10	7.75
10	HW-TD	19.41	236.7	266.4	340.1	11	9	9	9	9.5
11	SVR-TD	17.26	600.2	650.6	785.9	10	11	11	11	10.75

Cuadro 4.3: Resultados detallados de θ -SVR para la base de datos ‘Htseg2’.

Para los datos ‘Austourism’, cuyos resultados están presentes en el cuadro 4.4, el mejor modelo es nuestra propuesta, con sus dos variantes. A estas propuestas le siguen en ranking promedio el algoritmo presentado en el trabajo anterior: ϵ -SVR, con el enfoque bottom-up.

Modelo		MAPE			RANKING			\bar{R}_m
		L0	L1	L2	L0	L1	L2	
1	θ -SVR $_{\{J,0\}}$	2.58	5.62	7.15	2	1	1	1.33
2	θ -SVR $_{\{J,J-1\}}$	2.57	6.14	7.67	1	3	3	2.33
3	SVR-BU	2.78	6.09	7.28	4	2	2	2.67
4	HW-BU	5.12	6.89	8.99	7	4	4	5
5	SVR-MO	3.19	7.46	10	5	6	7	6
6	HW-MO	5.93	7.4	9.48	9	5	5	6.33
7	SVR-TD	2.72	10.2	10.64	3	9	9	7
8	ARIMA-BU	5.64	7.71	9.52	8	7	6	7
9	HW-TD	5.09	10.62	11.38	6	10	10	8.67
10	ARIMA-MO	6.52	8.54	10.37	11	8	8	9
11	ARIMA-TD	6.26	11.25	11.83	10	11	11	10.67

Cuadro 4.4: Resultados detallados de θ -SVR para la base de datos ‘Austourism’.

La cuarta base de datos en estudio corresponde al conjunto ‘Melsyd’, en el cual nues-

tra metodología propuesta θ -SVR $_{\{J,0\}}$ nuevamente presenta superioridad al compararse contra los otros métodos. En este caso le sigue ARIMA con el enfoque top-down. Estos resultados se detallan en el cuadro 4.5.

	Modelo	MAPE			RANKING			\bar{R}_m
		L0	L1	L2	L0	L1	L2	
1	θ -SVR $_{\{J,0\}}$	9.22	30.29	30.92	2	5	1	2.67
2	ARIMA-TD	9.99	24.46	32.21	3	1	4	2.67
3	θ -SVR $_{\{J,J-1\}}$	7.33	30.66	31.68	1	6	3	3.33
4	SVR-BU	15.94	29.37	31.09	8	3	2	4.33
5	HW-TD	13.73	27.02	32.47	6	2	5	4.33
6	SVR-TD	15.15	29.56	34.82	7	4	6	5.67
7	ARIMA-BU	11.04	41.48	48.57	4	8	8	6.67
8	ARIMA-MO	12.46	42.55	51.23	5	9	9	7.67
9	SVR-MO	24.99	36.07	35.2	11	7	7	8.33
10	HW-MO	16.61	48.53	57.71	9	10	10	9.67
11	HW-BU	17.77	48.73	57.93	10	11	11	10.67

Cuadro 4.5: Resultados detallados de θ -SVR para la base de datos ‘Melsyd’.

Por último, la quinta base de datos corresponde a ‘Departures’, en la cuál nuevamente la metodología θ -SVR $_{\{J,0\}}$ presenta el mejor desempeño predictivo (ver cuadro 4.6). En este caso, Holt-Winters muestra muy buenos resultados, incluso superando a los modelos de ARIMA y ε -SVR.

	Modelo	MAPE			RANKING			\bar{R}_m
		L0	L1	L2	L0	L1	L2	
1	θ -SVR _{J,0}	2.28	3.4	4.95	3	1	1	1.67
2	HW-BU	1.92	5	6.12	1	4	4	3
3	HW-MO	2.32	3.72	5.15	5	2	2	3
4	θ -SVR _{J,J-1}	2.41	4.02	5.51	8	3	3	4.67
5	SVR-BU	2.29	5.89	7.26	4	6	7	5.67
6	ARIMA-BU	2.34	5.67	6.57	7	5	5	5.67
7	HW-TD	2.02	19	64.17	2	10	10	7.33
8	SVR-MO	2.9	5.95	41.71	10	7	8	8.33
9	ARIMA-MO	2.97	6.42	6.92	11	8	6	8.33
10	SVR-TD	2.72	10.31	42.86	9	9	9	9
11	ARIMA-TD	2.33	19.44	64.74	6	11	11	9.33

Cuadro 4.6: Resultados detallados de θ -SVR para la base de datos ‘Departures’.

A continuación, se presenta el cuadro 4.7, en el que se comparan los desempeños de todos los modelos utilizados para cada una de las bases de datos, en función del ranking global, $\bar{\bar{R}}_m$, que se calcula como el promedio de todos los rankings individuales \bar{R}_m . En el caso de ε -SVR tradicional, en sus variantes SVR-BU, SVR-TD y SVR-MO, el ranking del cuadro es subrayado cuando el mejor desempeño se reporta con kernel gaussiano. Para los casos que no se subrayan en el cuadro, significa que los mejores desempeños se obtuvieron con kernel lineal.

Modelo	Htseg1	Htseg2	Austourism	Melsyd	Departures	\bar{R}_m
θ -SVR $_{\{J,0\}}$	1	3.25	1.33	2.67	1.67	1.98
SVR-BU	4	5.75	<u>2.67</u>	<u>4.33</u>	5.67	4.48
θ -SVR $_{\{J,J-1\}}$	8.67	6.25	2.33	3.33	4.67	5.05
ARIMA-BU	4	5	7	6.67	5.67	5.67
ARIMA-MO	2.67	2.75	9	7.67	8.33	6.08
HW-MO	9	4.75	6.33	9.67	3	6.55
HW-BU	10	4.5	5	10.67	3	6.63
SVR-MO	5.67	<u>5.75</u>	<u>6</u>	8.33	8.33	6.82
ARIMA-TD	5	7.75	10.67	2.67	9.33	7.08
SVR-TD	5.67	10.75	<u>7</u>	<u>5.67</u>	9	7.62
HW-TD	10.33	9.5	8.67	4.33	7.33	8.03

Cuadro 4.7: Resumen global de resultados para θ -SVR.

Se puede ver claramente como nuestra metodología propuesta, θ -SVR $_{\{J,0\}}$, obtiene los mejores resultados en términos de ranking por desempeño predictivo, con un ranking promedio de 1.98. Se puede ver como el método es superior a todas las otras metodologías planteadas, obteniendo no sólo el mejor ranking global, sino que también el mejor ranking en casi todas las bases de datos individuales.

El segundo mejor modelo, con un desempeño notoriamente inferior al primer modelo, le corresponde a la metodología clásica ε -SVR, con el enfoque bottom-up, la misma metodología que propusimos en el trabajo anterior. El ranking global que obtuvo este método es de 4.48, y es seguido de cerca por la segunda variante propuesta en este trabajo, θ -SVR $_{\{J,J-1\}}$, que obtuvo un ranking global de 5.05. Esto demuestra las virtudes de nuestra propuesta, y su ventaja en comparación a otros métodos clásicos.

En general, es el enfoque bottom-up el que promedió los mejores resultados, al comparar con los otros enfoques de HTS. Esto se podría explicar por los buenos resultados predictivos obtenidos en los niveles inferiores. De hecho, el enfoque que obtuvo los peores resultados corresponde al top-down, que corresponde al modelo que menos aprovecha los niveles inferiores para la predicción.

Los resultados no demuestran una ganancia significativa al utilizar métodos no lineales. Si bien SVR-BU se desempeñó claramente mejor que los otros enfoques, se puede

ver que en el caso de ARIMA, tanto con los enfoques top-down y middle-out, tuvo resultados promedio incluso mejores que ε -SVR con las mismas estrategias. También se puede observar una inferioridad de los modelos Holt-Winter al ser comparados con las otras metodologías. Por otro lado, si comparamos los desempeños de ε -SVR en sus dos versiones, considerando kernel lineal y gaussiano, podemos ver que solo en dos de cinco ocasiones el kernel gaussiano fue superior.

Mientras los mejores parámetros obtenidos para todas las variantes de ε -SVR se encuentran disponibles en el apéndice D, en el siguiente cuadro se reportan los mejores parámetros obtenidos para las variantes de ε -SVR:

		θ -SVR $_{\{J,0\}}$			θ -SVR $_{\{J,J-1\}}$		
		ε	C	θ	ε	C	θ
<i>ds1</i>	Austourism	0.1	2^{-5}	2^{12}	0.1	2^{-4}	2^{10}
<i>ds2</i>	Departures	0.1	2^{-12}	2^{15}	0.1	2^{-13}	2^5
<i>ds3</i>	Htseg1	0.1	2^6	2^{10}	0.1	2^{12}	2^2
<i>ds4</i>	Htseg2	0.1	2^{15}	2^6	0.1	2^{15}	2^{12}
<i>ds5</i>	Melsyd	0.1	2^{-7}	2^{-6}	0.1	2^{-6}	2^{-10}

Cuadro 4.8: Mejores parámetros para las variantes de la metodología θ -SVR.

4.3. KAT-SVR: la inclusión de métodos kernel a θ -SVR

La incorporación de métodos kernel, sin duda, le da mayor robustez a nuestra propuesta anterior, haciéndola más flexible, y pudiendo modelar de mejor forma problemas no lineales. Para este trabajo, se consideró un enfoque adicional de HTS para los modelos ARIMA y Holt-Winters, que no había sido considerado en los trabajos anteriores, el enfoque de reconciliación óptima (OR). Además, para este caso, se reemplazarán los modelos anteriores (variante θ -SVR) por nuestra nueva variante KAT-SVR, en sus dos versiones: KAT-SVR $_J^0$ y KAT-SVR $_J^{J-1}$, ya que θ -SVR corresponde a un caso particular de la nueva propuesta. Con esto, los nuevos modelos que se considerarán para los rankings en este trabajo son los siguientes:

- KAT-SVR $_J^0$: Variante que relaciona los niveles 0 y J e incorpora máquinas kernel.
- KAT-SVR $_J^{J-1}$: Variante que relaciona los niveles $J - 1$ y J e incorpora máquinas kernel.
- ARIMA-OR: El modelo construido con el método tradicional ARIMA y el enfoque optimal reconciliation.
- HW-OR: El modelo construido con el método tradicional Holt-Winters y el enfoque optimal reconciliation.

Antes de presentar los resultados de esta metodología, los cuadros 4.9 y 4.10 presentan la configuración de hiperparámetros óptima utilizada en las variantes de KAT-SVR.

Base de datos		Kernel lineal			Kernel gaussiano			
		ε	C	θ	ε	γ	C	θ
ds1	Austourism	0.1	2^{-5}	2^{12}	0.1	2^6	2^{-8}	2^{12}
ds2	Departures	0.1	2^{-12}	2^{18}	0.1	2^2	2^{-12}	2^7
ds3	Htseg1	0.1	2^6	2^{10}	0.1	2^2	2^{20}	2^{12}
ds4	Htseg2	0.1	2^{23}	2^6	0.1	2^4	2^{20}	2^6
ds5	Melsyd	0.1	2^{-7}	2^{-6}	0.1	2^6	2^{14}	2^{-10}

Cuadro 4.9: Configuración óptima de hiperparámetros para el método KAT-SVR $_J^0$.

Base de datos		Kernel lineal			Kernel gaussiano			
		ε	C	θ	ε	γ	C	θ
ds1	Austourism	0.1	2^{-4}	2^{10}	0.1	2^0	2^{-4}	2^6
ds2	Departures	0.1	2^{-13}	2^5	0.1	2^6	2^{12}	2^{16}
ds3	Htseg1	0.1	2^{12}	2^2	0.1	2^4	2^{10}	2^1
ds4	Htseg2	0.1	2^{24}	2^{12}	0.1	2^{-1}	2^{24}	2^{11}
ds5	Melsyd	0.1	2^{-6}	2^{-10}	0.1	2^8	2^0	2^{-10}

Cuadro 4.10: Configuración óptima de hiperparámetros para el método KAT-SVR $_J^{J-1}$.

Se puede observar que los hiperparámetros varían principalmente entre 2^{-13} y 2^{24} . En este caso estamos generando una grilla para ajustar los parámetros C , γ y θ mediante la estrategia de validación mencionada en el capítulo anterior, mientras que ε lo mantenemos fijo en 0.1.

El cuadro 4.11 presenta el resumen del ranking global obtenido por los distintos modelos con esta metodología, bajo el mismo principio de los rankings presentados en secciones anteriores. Al igual que en los estudios previos, para las variantes de los métodos de SVR se considera tanto el kernel lineal, como su versión gaussiana, reportando para cada caso el que obtenga el mejor desempeño. Nuevamente los resultados que consideran kernel gaussiano se encuentran subrayados en el cuadro, mientras los restantes consideraron kernel lineal.

Modelo	Austourism	Departures	Htseg1	Htseg2	Melsyd	\bar{R}_m
KAT-SVR _J ⁰	1.33	2.00	<u>1.00</u>	3.75	<u>3.33</u>	2.28
ε -SVR-BU	<u>2.67</u>	7.00	4.67	7.25	<u>5.00</u>	5.32
KAT-SVR _J ^{J-1}	2.33	5.33	10.00	7.50	<u>2.33</u>	5.50
ARIMA-BU	8.00	7.00	4.67	6.25	7.67	6.72
ARIMA-OR	9.67	8.00	4.33	5.50	6.67	6.83
ARIMA-MO	11.00	10.33	2.67	3.25	8.67	7.18
HW-OR	6.00	2.67	11.00	6.25	10.67	7.32
HW-MO	7.33	3.33	10.00	5.75	11.67	7.62
HW-BU	5.00	4.00	12.00	5.00	12.67	7.73
ε -SVR-MO	<u>6.67</u>	10.33	6.33	<u>7.00</u>	9.00	7.87
ARIMA-TD	12.67	11.00	5.67	9.25	2.67	8.25
ε -SVR-TD	<u>8.33</u>	11.00	6.67	12.75	<u>6.00</u>	8.95
HW-TD	10.00	9.00	12.00	11.50	4.67	9.43

Cuadro 4.11: Resumen de los resultados globales para todas las bases de datos y métodos, utilizando MAPE como métrica de desempeño.

Se puede apreciar como el método propuesto KAT-SVR_J⁰ obtiene los mejores desempeños globales al compararlo con los otros doce métodos, con un ranking promedio de 2.28. El modelo que le sigue, al igual que en la investigación anterior, corresponde a ε -SVR con el enfoque bottom-up, que a su vez es seguido de cerca por la segunda variante propuesta, KAT-SVR_J^{J-1}.

Dado los rankings de los modelos propuestos, se refuerza la hipótesis de que agrupar información a través de los niveles de jerarquía es beneficioso para el pronóstico de HTS. Por otro lado, es claro que la variante KAT-SVR_J⁰ fue muy superior a KAT-SVR_J^{J-1}, lo que nos indica que es preferible agrupar la información desde el nodo raíz. Esto también refuerza nuestra teoría de las ventajas de la construcción de un único modelo de optimización.

Notemos además, que los tres primeros lugares del ranking lo ocupan métodos de Machi-

ne Learning, que se benefician de la flexibilidad de la incorporación de métodos kernel, más allá de que el kernel gaussiano no siempre mejora los resultados para las diferentes bases de datos, al compararlo con la versión lineal. De todas formas, su incorporación entrega la flexibilidad en caso de querer lidiar con bases de datos de una naturaleza más compleja.

Los primeros cuatro modelos del ranking, por su parte, corresponden a estrategias del tipo bottom-up, demostrando las ventajas de realizar las predicciones en el último nivel de la jerarquía, para luego agregar los resultados hacia los niveles superiores. Además, la incorporación de métodos con el enfoque de reconciliación óptima nos permite concluir que mientras más desagregada esté la data para el momento del entrenamiento, mejores son los resultados. Esto se aprecia puesto que los resultados de este enfoque tienden a ser mejores que los de enfoques como top-down o middle-out.

Si analizamos los resultados para cada base de datos, vamos a encontrar, por ejemplo, que ‘Austourism’ es un conjunto de datos en el que la metodología propuesta obtiene muy buenos resultados, y donde los tres primeros lugares del ranking muestran un desempeño muy superior al resto de los modelos. Esto se puede apreciar en el cuadro 4.12.

Modelo	MAPE			RANKING			\bar{R}_m
	L0	L1	L2	L0	L1	L2	
KAT-SVR $_J^0$	2.58	5.62	7.15	2	1	1	1.33
KAT-SVR $_J^{J-1}$	2.57	6.14	7.67	1	3	3	2.33
ε -SVR-BU	2.78	6.09	7.28	4	2	2	2.67
HW-BU	5.12	6.89	8.99	7	4	4	5.00
HW-OR	5.44	6.92	9.19	8	5	5	6.00
ε -SVR-MO	3.19	7.46	10.00	5	7	8	6.67
HW-MO	5.93	7.40	9.48	10	6	6	7.33
ARIMA-BU	5.64	7.71	9.52	9	8	7	8.00
ε -SVR-TD	2.72	10.20	10.64	3	11	11	8.33
ARIMA-OR	6.11	7.87	10.12	11	9	9	9.67
HW-TD	5.09	10.62	11.38	6	12	12	10.00
ARIMA-MO	6.52	8.54	10.37	13	10	10	11.00
ARIMA-TD	6.26	11.25	11.83	12	13	13	12.67

Cuadro 4.12: Resultados detallados de KAT-SVR para la base de datos ‘Austourism’.

La importancia de este resultado radica en que esta base de datos no corresponden a data simulada, sino que a datos reales. La metodología propuesta, en este caso, ocupa los puestos 1 y 2 del ranking, con valores promedio de 1.33 y 2.33. Les sigue de cerca

SVR-BU, con un ranking de 2.67. Sin duda estos tres modelos lograron capturar el comportamiento de estos datos de mucho mejor forma que todos los demás modelos presentados, ya que el cuarto modelo en el ranking ya baja considerablemente su desempeño, con un ranking promedio de 5.00. Si bien no se consideró kernel gaussiano para la metodología propuesta en esta base de datos, éste si mejoró los resultados de ε -SVR. La tendencia de KAT-SVR $_J^0$ como el mejor modelo se mantiene para la base de datos ‘Departures’, con un ranking global de 2.00. En este caso Holt-Winters si logra anteponerse a los otros métodos, tomando el segundo y tercer lugar del ranking, con los enfoques de reconciliación óptima y middle-out. En este caso el kernel lineal es superior a la versión gaussiana para los métodos de ε -SVR. Estos resultados se presentan en el cuadro 4.13.

Modelo	MAPE			RANKING			\bar{R}_m
	L0	L1	L2	L0	L1	L2	
KAT-SVR $_J^0$	2.28	3.40	4.95	4	1	1	2.00
HW-OR	1.91	4.24	5.45	1	4	3	2.67
HW-MO	2.32	3.72	5.15	6	2	2	3.33
HW-BU	1.92	5.00	6.12	2	5	5	4.00
KAT-SVR $_J^{J-1}$	2.41	4.02	5.51	9	3	4	5.33
ARIMA-BU	2.34	5.67	6.57	8	6	7	7.00
ε -SVR-BU	2.29	5.89	7.26	5	7	9	7.00
ARIMA-OR	2.42	5.90	6.56	10	8	6	8.00
HW-TD	2.02	19.00	64.17	3	12	12	9.00
ε -SVR-MO	2.90	5.95	41.71	12	9	10	10.33
ARIMA-MO	2.97	6.42	6.92	13	10	8	10.33
ARIMA-TD	2.33	19.44	64.74	7	13	13	11.00
ε -SVR-TD	2.72	10.31	42.86	11	11	11	11.00

Cuadro 4.13: Resultados detallados de KAT-SVR para la base de datos ‘Departures’.

En el caso de los datos simulados ‘Htseg1’, nuevamente el mejor modelo es KAT-SVR $_J^0$, ahora con kernel gaussiano, obteniendo el incuestionable ranking de 1. Para este caso, le sigue ARIMA con el enfoque middle-out (ver cuadro 4.14).

En el caso de los datos simulados ‘Htseg2’, tenemos la primera ocasión en que un modelo de Machine Learning no ocupa el primer lugar del ranking. En este caso se impone ARIMA, con el enfoque middle-out, aunque seguido de cerca por nuestro método propuesto KAT-SVR $_J^0$. El buen desempeño de ARIMA tanto para ‘Htseg1’ como para ‘Htseg2’ lo podemos explicar por la naturaleza simulada de estos datos, dados sus buenos resultados para estas dos bases de datos, siendo que poseen un ranking global bastante

Modelo	MAPE			RANKING			\bar{R}_m
	L0	L1	L2	L0	L1	L2	
KAT-SVR $_J^0$	0.19	0.31	0.45	1	1	1	1
ARIMA-MO	0.51	0.52	0.72	4	2	2	2.67
ARIMA-OR	0.54	0.55	0.77	6	3	4	4.33
ε -SVR-BU	0.20	1.16	1.24	2	7	5	4.67
ARIMA-BU	0.59	0.60	0.77	7	4	3	4.67
ARIMA-TD	0.51	0.62	1.62	5	5	7	5.67
ε -SVR-MO	0.22	1.34	1.74	3	8	8	6.33
ε -SVR-TD	0.82	0.88	1.35	8	6	6	6.67
KAT-SVR $_J^{J-1}$	0.85	1.57	1.78	9	9	12	10
HW-MO	1.80	1.81	1.78	11	10	9	10
HW-OR	1.80	1.81	1.78	12	11	10	11
HW-BU	1.80	1.81	1.78	13	12	11	12
HW-TD	1.80	1.87	2.00	10	13	13	12

Cuadro 4.14: Resultados detallados de KAT-SVR para la base de datos ‘Htseg1’.

deficiente al agregar las otras fuentes de información a la comparación. También podemos notar que para ambos conjuntos de datos, rara vez se utilizó kernel gaussiano, lo que se puede deber a la naturaleza lineal intrínseca de estos datos simulados para análisis econométrico. Los resultados de ‘Htseg2’ se presentan en el cuadro 4.15.

Modelo	MAPE				RANKING				\bar{R}_m
	L0	L1	L2	L3	L0	L1	L2	L3	
ARIMA-MO	0.23	2.52	2.34	7.58	2	3	3	5	3.25
KAT-SVR $_J^0$	0.73	3.24	2.48	4.50	4	6	4	1	3.75
HW-BU	10.11	2.45	2.07	8.36	9	2	1	8	5
ARIMA-OR	1.14	3.34	2.84	7.58	5	7	6	4	5.5
HW-MO	2.24	2.65	2.59	8.22	6	5	5	7	5.75
ARIMA-BU	0.23	3.78	3.05	7.82	3	9	7	6	6.25
HW-OR	10.43	2.58	2.08	8.74	10	4	2	9	6.25
ε -SVR-MO	4.90	1.59	8.53	29.21	7	1	10	10	7
ε -SVR-BU	12.63	3.49	3.53	6.49	11	8	8	2	7.25
KAT-SVR $_J^{J-1}$	8.22	6.10	8.18	6.96	8	10	9	3	7.5
ARIMA-TD	0.23	318.29	355.29	448.31	1	12	12	12	9.25
HW-TD	19.41	236.75	266.43	340.11	13	11	11	11	11.5
ε -SVR-TD	17.26	600.18	650.61	785.90	12	13	13	13	12.75

Cuadro 4.15: Resultados detallados de KAT-SVR para la base de datos ‘Htseg2’.

Finalmente, en el cuadro 4.16 se presentan los resultados de la base de datos ‘Melsyd’, en el cual nuevamente se ve la presencia de nuestra metodología KAT-SVR $_J^0$ en la cima del ranking (2.33). Le siguen ARIMA con el enfoque top-down, y luego nuestra segunda variante, KAT-SVR $_J^{J-1}$. En este caso si se nota una mejoría en los resultados al utilizar la versión de kernel gaussiano en lugar del lineal, estando presente en cuatro de los

cinco modelos posibles. Con esto, cuatro de las primeras seis posiciones del ranking consideran la utilización de kernel gaussiano dentro de sus algoritmos.

Modelo	MAPE			RANKING			\bar{R}_m
	L0	L1	L2	L0	L1	L2	
KAT-SVR_J^{J-1}	6.03	29.70	30.44	1	5	1	2.33
ARIMA-TD	9.99	24.46	32.21	3	1	4	2.67
KAT-SVR_J^0	6.91	30.13	31.00	2	6	2	3.33
HW-TD	13.73	27.02	32.47	7	2	5	4.67
ε -SVR-BU	15.94	29.37	31.09	9	3	3	5.00
ε -SVR-TD	15.15	29.56	34.82	8	4	6	6.00
ARIMA-OR	10.91	37.68	45.75	4	8	8	6.67
ARIMA-BU	11.04	41.48	48.57	5	9	9	7.67
ARIMA-MO	12.46	42.55	51.23	6	10	10	8.67
ε -SVR-MO	24.99	36.07	35.20	13	7	7	9.00
HW-OR	16.52	48.43	57.58	10	11	11	10.67
HW-MO	16.61	48.53	57.71	11	12	12	11.67
HW-BU	17.77	48.73	57.93	12	13	13	12.67

Cuadro 4.16: Resultados detallados de KAT-SVR para la base de datos ‘Melsyd’.

Para este trabajo, se decidió replicar los experimentos, pero considerando RMSE en lugar de MAPE como medida de desempeño. Un resumen de estos resultados se presenta en el cuadro 4.17. Se puede notar que el primer lugar del ranking no varía, manteniéndose KAT-SVR_J^0 en la cima del ranking global. Esto demuestra la robustez de nuestro método propuesto, incluso ante variaciones de la métrica de error seleccionada.

Modelo	Austourism	Departures	Htseg1	Htseg2	Melsyd	\bar{R}_m
KAT-SVR $_j^0$	1.67	6.67	<u>1.33</u>	3.75	<u>2.33</u>	3.15
ARIMA-OR	9.00	4.67	4.67	6.25	4.00	5.72
ARIMA-BU	8.00	6.00	5.00	7.00	5.67	6.33
ε -SVR-BU	<u>3.00</u>	8.00	4.33	7.75	<u>8.67</u>	6.35
ARIMA-MO	11.00	10.67	2.67	2.50	6.67	6.70
KAT-SVR $_j^{J-1}$	2.67	9.00	9.00	9.50	<u>3.67</u>	6.77
HW-OR	6.00	1.33	11.67	6.50	9.33	6.97
ARIMA-TD	12.33	7.33	5.33	9.25	1.67	7.18
HW-BU	5.00	1.67	12.67	5.25	12.00	7.32
HW-MO	8.33	4.00	10.67	5.00	10.33	7.67
ε -SVR-MO	<u>6.33</u>	11.67	6.00	<u>4.00</u>	13.00	8.20
HW-TD	10.00	9.00	11.00	11.50	4.00	9.10
ε -SVR-TD	<u>7.67</u>	11.00	6.67	12.75	<u>9.67</u>	9.55

Cuadro 4.17: Resumen de los resultados globales para todas las bases de datos y métodos, utilizando RMSE como métrica de desempeño.

4.4. Predicción de volumen de llamadas en un Contact Center

En esta sección se presentan los resultados obtenidos hasta el momento en la última investigación relacionada a esta tesis doctoral. El principal objetivo es ser capaces de utilizar algún método robusto de estimación de demanda, para a su vez, utilizar sus resultados como *input* de un modelo de optimización estocástico, que permita dimensionar la dotación de ejecutivos y posiciones de Contact Center requeridas por una empresa para atender sus llamados telefónicos.

Para esto, se generan predicciones de series de tiempo jerárquicas para la cantidad de llamadas entrantes (*inbound*) que una empresa recibe. En primera instancia, el método predictivo a utilizar corresponderá a el clásico ε -SVR, que será contrastado con las versiones de ARIMA y Holt-Winters, utilizando los tres principales enfoques de HTS: bottom-up, top-down y middle-out.

Con respecto a los datos, la empresa elegida para los primeros resultados predictivos corresponde a la Asociación Chilena de Seguridad (ACHS), y sus datos de cantidad de llamadas están distribuidos en dos niveles bajo la serie total. En el primer nivel, las llamadas se desagregan entre 18 áreas de la empresa, las cuáles reciben las siguientes denominaciones: APC, C19, CAP, CEM, COM, EME, EPR, GES, INF, PPE, PRE,

SEL, SSA, SUB, SWE, TRA, URG y OTR.

A su vez, cada una de estas áreas puede ser desagregada nuevamente, en base a sus puntos de ruteo o entradas telefónicas. Cada una de estas áreas puede tener entre 1 y 5 entradas asociadas. Cabe destacar que se cuenta con datos de 4 años, desde septiembre de 2017 en adelante. El apéndice A presenta más detalles de este conjunto de datos.

En el cuadro 4.18 se presentan los resultados para los modelos utilizados. Cabe destacar que en este caso la medida de desempeño utilizada para obtener estos resultados fue RMSE.

Modelo	RMSE			RANKING			\bar{R}_m
	L0	L1	L2	L0	L1	L2	
1 SVR-BU	1271.29	90.44	42.89	1	1	1	1
2 SVR-MO	1283.24	91.08	44.26	2	2	3	2.33
3 HW-BU	1453.68	93.15	43.92	6	3	2	3.66
4 SVR-TD	1409.51	98.10	43.72	3	5	5	4.33
5 HW-MO	1423.51	94.42	46.58	5	4	4	4.33
6 HW-TD	1412.35	157.90	78.28	4	8	8	6.66
7 ARIMA-BU	1664.94	104.28	48.44	9	6	6	7
8 ARIMA-MO	1613.08	104.82	49.58	8	7	7	7.33
9 ARIMA-TD	1525.25	161.63	79.88	7	9	9	8.33

Cuadro 4.18: Resultados detallados para la base de datos de Contact Center.

Se puede apreciar como ε -SVR con el enfoque bottom-up posee por amplia ventaja los mejores resultados, registrando el mejor ranking en cada nivel individual, y por lo tanto, el mejor ranking global.

Capítulo 5

Conclusiones

5.1. Conclusiones teóricas y prácticas

Esta tesis presenta resultados de una serie de estudios, que se enmarcan en la generación de modelos de pronóstico utilizando tecnología de Machine Learning, buscando como principal objetivo que sus resultados permitan enriquecer las fuentes de información disponibles, para que los tomadores de decisiones cuenten con mayor contexto a la hora de realizar el análisis de los datos, reduciendo la incertidumbre de las distintas decisiones que ellos deben tomar.

Es evidente que con la capacidad de conocer el futuro se pueden tomar muchísimas decisiones de bajo riesgo de cara a cualquier negocio. Como nadie puede conocer con certeza el futuro, la generación de modelos que lo puedan predecir con un bajo margen de error se vuelve relevante; ya que entregan, con cierta probabilidad, datos que de otra forma no se conocerían. Es por esto que hablamos del enriquecimiento de las fuentes de información, lo que nos permitirá no sólo generar nuevos reportes para los tomadores de decisiones, sino que definir lineamientos estratégicos con mayor exactitud, reduciendo la incertidumbre del “¿qué pasará?”.

Si queremos reducir la incertidumbre, es fundamental que los modelos que busquemos construir como representaciones de la realidad, sean lo más cercanos posibles a esta realidad. Es por esto que la correcta construcción de estos modelos es clave para la obtención de buenos resultados. Es fundamental contar con los datos, pero por otro lado, también es muy importante contar con expertiz en el negocio que se está pensando

modelar, ya que la inclusión o no inclusión de ciertas variables puede resultar una tarea crucial a la hora de construir los modelos. Por otro lado, es muy importante considerar la problemática del *big-data*. Las empresas cada vez tienen más datos, y cada vez le están dando más importancia al cuidado y mantención de estos, entendiéndolos como uno de sus principales activos. Es por esto que muchos modelos o técnicas clásicas de la literatura pueden no ser los más eficientes hoy en día, y es fundamental que los métodos se construyan teniendo en cuenta este contexto de los datos.

Dado esto, resulta clave que los modelos predictivos a desarrollar sean capaces de representar la realidad de la forma más precisa posible. Es por eso que uno de los primeros objetivos de esta investigación corresponde a ser capaces de generar mejores modelos que los que existen actualmente, que obtengan una precisión mayor, y por lo tanto, generen un dato que reduzca la incertidumbre de los tomadores de decisión, apoyándolos íntegramente en este proceso.

Pudimos apreciar a lo largo de esta investigación como con modelos de Machine Learning fuimos capaces de mejorar los resultados de los modelos clásicos de la literatura para la predicción de series de tiempo; en particular, para series de tiempo jerárquicas. En la primera parte de esta investigación vimos como la sola incorporación de una técnica denominada Support Vector Regression obtuvo resultados muy alentadores, mejorando los pronósticos de dos de los métodos más reconocidos para estas tareas, como lo son ARIMA y Holt-Winters. En este caso, vimos como en el pronóstico de series de tiempo jerárquicas, tanto de *benchmark*, como de un contexto de ventas de la industria del Travel Retail de una empresa real, nuestra metodología propuesta obtuvo los mejores rankings asociados al menor error a la hora de la prueba de los modelos. En particular, la versión ε -SVR con la metodología de agregación de series jerárquicas bottom-up, obtuvo, en promedio, los mejores resultados al compararlo tanto con otros enfoques de HTS, como también con otras técnicas predictivas. Por otro lado, no sólo con el enfoque bottom-up se obtuvieron buenos resultados. Si comparamos entre enfoques de HTS, para el enfoque top-down nuestra metodología también mostró superioridad a la hora de compararla con ARIMA o Holt-Winters. En el único enfoque donde Holt-Winters fue levemente superior a la metodología propuesta, fue en middle-out.

Otra conclusión importante que podemos desprender de este primer trabajo, es que la inclusión del kernel gaussiano resultó, en promedio, ser mejor que el kernel lineal,

sobretudo para las bases de datos reales, con datos no simulados. Esto se puede explicar debido a que normalmente este tipo de datos poseen patrones de comportamiento no lineales, los cuales pueden ser modelados de mejor forma con un kernel gaussiano. Esto le entrega mucha flexibilidad a nuestra propuesta.

Otro punto fundamental de esta tesis corresponde a la ventaja de la utilización de series de tiempo con estructura jerárquica. Muchas veces, para una serie de tiempo, tendremos una serie de características que nos permitirán generar una estructura jerárquica en base a ella. Es más, vimos con las bases de datos de la empresa Generation Research AB, como a partir de la misma serie de tiempo (ventas de productos en la industria Travel Retail), se obtuvieron seis series de tiempo jerárquicas distintas, solo modificando el cómo utilizamos sus características para su construcción y estudio. Pudimos ver, mediante la utilización de los enfoques clásicos de HTS (bottom-up, top-down y middle-out) cómo se aprovecha esta jerarquía para mejorar los desempeños predictivos. Es más, fue claro que el enfoque que mejores resultados obtuvo corresponde a bottom-up, lo que nos dice que al realizar predicciones en los niveles más bajos de la jerarquía, podemos aprovechar sus características a nuestro favor para mejorar las predicciones. Por otro lado, vimos que al relacionar estos nodos del último nivel, con niveles superiores de la jerarquía, se mejora aun más el desempeño predictivo. Esto lo pudimos demostrar mediante nuestro segundo y tercer trabajo, donde se introducen los modelos θ -SVR y KAT-SVR respectivamente, y donde vimos que la variante que relaciona el último nivel con el nivel raíz obtuvo los mejores resultados predictivos, incluso comparando con estrategias tradicionales de Machine Learning, como ε -SVR.

Con estos dos trabajos, se demostró que es posible adaptar Support Vector Regression para series de tiempo jerárquicas, de tal forma de prevenir que las series de los niveles inferiores se desvíen mucho de las series superiores, mediante agrupaciones con otros niveles de la jerarquía.

Finalmente, el último proyecto presentado en esta tesis, busca aprovechar el poder predictivo de los modelos presentados anteriormente, con el fin de utilizar estos resultados, que entendemos poseen un margen de error muy bajo, y que permiten la reducción de incertidumbre en la toma de decisiones. La forma de aprovechar estos resultados es utilizándolos como *input* de un problema de optimización clásico de *staffing*, pero que en su versión clásica no considera un insumo tan valioso como lo son las estimaciones de

demanda futuras con un bajo nivel de error, y sus respectivos intervalos de confianza. Para esto, se generan predicciones con ε -SVR para las series de tiempo jerárquicas asociadas a las llamadas telefónicas entrantes que recibe una empresa en Chile, pudiendo, con estos resultados, anteponernos de mejor forma a la demanda, reducir la incertidumbre, y utilizar esto para estimar de forma precisa cuántos puestos de Call Center necesitará la empresa para atender a sus clientes.

5.2. Investigación futura

Sin lugar a dudas este trabajo resuelve varias interrogantes, pero también nos lleva a nuevas preguntas, que requieren estudios posteriores para poder responderse. En esa línea, esta sección presenta una serie de propuestas de investigación futura, que podrían complementar, completar, o bien enriquecer esta investigación.

Con respecto a series de tiempo jerárquicas, uno de los principales desafíos sería el aprovechar de la mejor forma posible la estructura jerárquica de las series. En ese sentido, nuestro trabajo se podría extender con el uso de enfoques más sofisticados, en lugar de bottom-up, top-down y middle-out. Por ejemplo, se podría implementar SVR con el enfoque de reconciliación óptima propuesto por Rob Hyndman en 2014 (R. J. Hyndman, 2014), o incluso se podría generar alguna variante inteligente del enfoque top-down, que en lugar de utilizar proporciones históricas, considere la evolución de dichas proporciones en el tiempo.

Otro estudio que podríamos evaluar, es la utilización de otras tecnologías de Machine Learning en lugar de SVR, como por ejemplo redes neuronales diseñadas para HTS. Por otro lado, dado que estamos buscando representar la realidad, y definiendo las variables que mejor la modelan, sería interesante incluir estrategias de selección de atributos (ver (Maldonado & Weber, 2010; Dai, Shao & Lu, 2013)), para generar modelos dinámicos, que encuentren atributos relevantes dentro del algoritmo.

Otro tema interesante a investigar corresponde a estudiar la robustez de los modelos de Machine Learning cuando se enfrentan a bases de datos pequeñas, lo que suele ser muy común en series de tiempo jerárquicas. Una solución para esto es el desarrollo de modelos de optimización robusta, generados para garantizar cierto nivel de ajuste mínimo (López, Maldonado & Carrasco, 2019). Estos conceptos se podrían extender al

análisis de series de tiempo jerárquicas.

Un problema importante de nuestras metodologías θ -SVR y KAT-SVR es que la complejidad del modelo de optimización puede crecer considerablemente bajo ciertas circunstancias, y aquí entra el desafío del tratamiento de datos de alta frecuencia para predicciones de corto plazo (Valente & Maldonado, 2020). Otro problema se produce al enfrentarse a HTS con gran cantidad de nodos en el último nivel como ocurre en Karmy & Maldonado (Karmy & Maldonado, 2019). Proponemos entonces estudiar más en profundidad las complejidades de las series de tiempo jerárquicas reales, considerando esquemas de optimización eficientes que puedan agrupar información a través de niveles de jerarquía, y que permitan lidiar con esta problemática del *big-data*.

Finalmente, con respecto al modelo de *staffing* propuesto, se sugiere el estudio de las distintas reglas de negocio que podrían incorporarse en el modelo para abarcar más casos reales de distintas empresas. Por otro lado, con respecto a los intervalos de confianza para los métodos de Machine Learning en predicciones de series de tiempo jerárquicas, se recomienda estudiar la propagación de dichos intervalos a lo largo de la jerarquía para enfoques distintos a bottom-up, como por ejemplo top-down, middle-out, u optimal reconciliation.

Nomenclatura

ACHS: Asociación Chilena de Seguridad.

ARIMA: Autoregressive Integrated Moving Average.

BU: Bottom-up.

HTS: Hierarchical Time Series.

JCR: Journal Citation Report.

KAT: Kernel all-together.

KDD: Knowledge Discovery in Databases.

KKT: Karush-Kuhn-Tucker.

MAPE: Mean absolute percentage error.

MO: Middle-out.

OR: Optimal reconciliation.

RBS: Radial Basis Function.

RMSE: Root-mean-square error.

RP: Routing point.

SMO: Sequential minimal optimization.

SVM: Support Vector Machine.

SVR: Support Vector Regression.

TD: Top-down.

Referencias

- Abut, F., Tartuk, M., Nurdag, T., Acar, V., Erdem, S., & Akay, F. (2022). Forecasting call center arrivals using xgboost combined with consecutive and periodic lookback. *Eastern Anatolian Journal of Science*, 8(1), 20–25.
- Ahmed, N. K., Atiya, A. F., El-Gayar, N., & El-Shishiny, H. (2010). An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, 29(5-6), 594–621.
- Athanasopoulos, G., Hyndman, R. J., Kourentzes, N., & Petropoulos, F. (2017). Forecasting with temporal hierarchies. Working Paper.
- Bertsekas, D. (1982). *Constrained Optimization and Lagrange Multiplier Methods*. New York: Academic Press.
- Bolch, G., Greiner, S., De Meer, H., & Trivedi, K. (2006). *Queueing Networks and Markov Chains*. John Wiley & Sons, Inc.
- Chernichenko, K., Kapkanets, A., & Panchenko, T. (2016). Contact center load forecasting and operator schedule planning. *Problems in programming*, 2-3, 227–236.
- Chromy, E., Misuth, T., & Kavacky, M. (2011). Erlang c formula and its use in the call centers.
- Crone, S., Guajardo, J., & Weber, R. (2006). Artificial intelligence in theory and practice, ed. m. bramer, (boston: Springer). *TFTP International Federation for Information Processing*, 217, 149–158.
- Crone, S., Lessmann, S., & Pietsch, S. (2006). Forecasting with computational intelligence - an evaluation of support vector regression and artificial neural networks for time series prediction. In *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, (pp. 3159–3166).
- Dai, W. S., Shao, Y. E., & Lu, C. (2013). Incorporating feature selection method into support vector regression for stock index forecasting. *Neural Computing and Applications*, 23(6), 1551–1561.

- Drucker, H., Burges, C., Kaufman, L., Smola, A., & Vapnik, V. (1997). Support vector regression machines. In *Advances in Neural Information Processing Systems*, volume 9, (pp. 155–161). MIT Press.
- Du, X. F., Leung, S. C. H., Zhang, J. L., & Lai, K. K. (2013). Demand forecasting of perishable farm products using support vector machine. *International Journal of Systems Science*, 44(33), 556–567.
- Fayyad, U. (1996). Data mining and knowledge discovery- making sense out of data. *IEEE Expert-Intelligent Systems and Their Applications*, (11), 20–25.
- Franca Lila, M., Meira, E., & Cyrino Oliveira, F. (2022). Forecasting unemployment in brazil: A robust reconciliation approach using hierarchical data. *Socio-Economic Planning Sciences*, 82, Part B.
- Franses, P. H. & Legerstee, R. (2011). Combining sku-level sales forecasts from models and experts. *Expert Systems with Applications*, 38(3), 2365–2370.
- Hansen, J., McDonald, J., & Nelson, R. (2006). Some evidence on forecasting time-series with support vector machines. *Journal of the Operational Research Society*, 57(9), 1053–1063.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.
- He, Q.-Q., Wu, C., & Si, Y.-W. (2022). Lstm with particle swam optimization for sales forecasting. *Electronic Commerce Research and Applications*, 51, 101118.
- Hu, Z., Bao, Y., & Xiong, T. (2013). Electricity load forecasting using support vector regression with memetic algorithms. *The Scientific World Journal*, 2013, 10.
- Huber, J., Gossmann, A., & Stuckenschmidt, H. (2017). Cluster-based hierarchical demand forecasting for perishable goods. *Expert Systems with Applications*, 76, 140–151.
- Hyndman, R. & Athanasopoulos, G. (2021). *Forecasting: principles and practice, 3rd edition*. OTexts: Melbourne, Australia.

- Hyndman, R. J. (2015). Package fpp.
- Hyndman, R. J. & Athanasopoulos, G. (2013). *Forecasting: principles and practice*. O Texts.
- Hyndman, R. J., Athanasopoulos, G., & Shang, H. L. (2012). hts: An r package for forecasting hierarchical or grouped time series.
- Jiang, S., Chin, K.-S., Wang, L., Qu, G., & Tsui, K. L. (2017). Modified genetic algorithm-based feature selection combined with pre-trained deep neural network for demand forecasting in outpatient department. *Expert Systems with Applications*, 82, 216–230.
- Kanthanathan, C., Carty, G., Raja, M., & Ryan, C. (2020). Recurrent neural network based automated workload forecasting in a contact center. In *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*, (pp. 1423–1428).
- Karmy, J. P. & Maldonado, S. (2019). Hierarchical time series forecasting via support vector regression in the european travel retail industry. *Expert Systems with Applications*, 137, 59–73.
- Levis, A. A. & Papageorgiou, L. G. (2009). Customer demand forecasting via support vector regression analysis. *Chemical Engineering Research and Design*, 83(8), 1009–1018.
- Liu, J., Wang, P., Chen, H., & Zhu, J. (2022). A combination forecasting model based on hybrid interval multi-scale decomposition: Application to interval-valued carbon price forecasting. *Expert Systems with Applications*, 191, 116267.
- Liu, Q., Wang, F., Li, J., & Xiao, W. (2021). A hybrid support vector regression with multi-domain features for low-velocity impact localization on composite plate structure. *Mechanical Systems and Signal Processing*, 154(1), 107547.
- Liu, Z., Yan, Y., Yang, J., & Hauskrecht, M. (2015). Missing value estimation for hierarchical time series: A study of hierarchical web traffic. In *2015 IEEE International Conference on Data Mining*, (pp. 895–900).

- López, J., Maldonado, S., & Carrasco, M. (2019). Robust nonparallel support vector machines via second-order cone programming. *Neurocomputing*, *364*, 227–238.
- Lu, C.-J. (2014). Sales forecasting of computer products based on variable selection scheme and support vector regression. *Neurocomputing*, *128*, 491–499.
- Lu, C.-J. & Wang, Y.-W. (2010). Combining independent component analysis and growing hierarchical self-organizing maps with support vector regression in product demand forecasting. *International Journal of Production Economics*, *128*(2), 603–613.
- Makridakis, S. & Wheelwright, S. C. (1977). Forecasting: Issues & challenges for marketing management. *Journal of Marketing*, *41*(4), 24–38.
- Maldonado, S. & López, J. (2016). Robust kernel-based multiclass support vector machines via second-order cone programming. *Applied Intelligence*, *46*, 983–992.
- Maldonado, S. & Weber, R. (2010). Feature selection for support vector regression via kernel penalization. In *2010 International Joint Conference on Neural Networks Proceedings*, (pp. 1973–1979).
- Martínez, F., Frías, M. P., Pérez-Godoy, M. D., & Rivera, A. J. (2018). Dealing with seasonality by narrowing the training set in time series forecasting with knn. *Expert Systems with Applications*, *103*, 38–48.
- Mitchell, T. (1997). *Machine Learning*. McGraw Hill.
- Mizan, T. & Taghipour, S. (2022). Medical resource allocation planning by integrating machine learning and optimization models. *Artificial Intelligence in Medicine*, *134*.
- Osuna, E., Freund, R., & Girosi, F. (1997). An improved training algorithm for support vector machines. In *Neural Networks for Signal Processing VII. Proceedings of the 1997 IEEE Signal Processing Society Workshop*, (pp. 276–285).
- Panagiotelis, A., Athanasopoulos, G., Gamakumara, P., & Hyndman, R. (2021). Forecast reconciliation: A geometric view with new insights on bias correction. *International Journal of Forecasting*, *37*(1), 343–359.

- Platt, J. (1998). Sequential minimal optimization: A fast algorithm for training support vector machines. Technical Report MSR-TR-98-14, Microsoft.
- Pyle, D. (1999). *Data preparation for data mining*. Morgan Kaufmann Publishers.
- R. J. Hyndman, G. A. (2014). Optimally reconciling forecasts in a hierarchy. *Fore-sight: The International Journal of Applied Forecasting, International Institute of Forecasters*, 35, 42–48.
- Ren, Y., Suganthan, P. N., & Srikanth., N. (2016). A novel empirical mode decomposition with support vector regression for wind speed forecasting. *IEEE Transactions on Neural Networks and Learning Systems*, 27(8), 1793–1798.
- Riečanová, Z., Horváth, J., Olejček, V., & Riečan, B. (1987). Numerické metody a matematická štatistika (numerical methods and mathematical statistics).
- Sabzekar, M. & Hossein Hasheminejad, S. (2021). Robust regression using support vector regressions. *Chaos, Solitons & Fractals*, 144(110738).
- Sapankevych, N. & Sankar, R. (2009). Time series prediction using support vector machines: a survey. *IEEE Computational Intelligence Magazine*, 4, 24–38.
- Schölkopf, B. & Smola, A. J. (2002). *Learning with Kernels*. MIT Press, Cambridge, MA, USA.
- Smola, A. & Scholkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, 14(3), 199–222.
- Smola, A. J. & Scholkopf, B. (1998). A tutorial on support vector regression. Technical report, NeuroCOLT Technical Report NC-TR-98-030, Royal Holloway College, University of London, UK.
- Sohrabpour, V., Oghazi, P., Toorajipour, R., & Nazarpour, A. (2021). Export sales forecasting using artificial intelligence. *Technological Forecasting and Social Change*, 163, 120480.
- Steffen, M., Sarda, A., Artz-Beielstein, T., Zaefferer, M., & Strok, J. (2015). Comparison of different methods for univariate time series imputation in r. arXiv:1510.03924.

- Unčovský, L. (1980). Stochastické modely operačnej analýzy. *Bratislava: Alfa*, 416 p.
- Valente, J. M. & Maldonado, S. (2020). Svr-ffs: A novel forward feature selection approach for high-frequency time series forecasting using support vector regression. *Expert Systems with Applications*, 160, 113729.
- Vapnik, V. (1998). *Statistical Learning Theory*. John Wiley and Sons.
- Wood, S. & Tasker, A. (2008). The importance of context in store forecasting: The site visit in retail location decision-making. *Journal of Targeting, Measurement and Analysis for Marketing*, 16(2), 139–155.
- Yang, J. & Liu, G. (2002). The evaluation of classification models for credit scoring. *Arbeitsbericht*, (02-2002). Institut für Wirtschaftsinformatik, Georg-August-Universität Göttingen.
- Zhang, S., Wang, C., Liao, P., Xiao, L., & Fu, T. (2022). Wind speed forecasting based on model selection, fuzzy cluster, and multi-objective algorithm and wind energy simulation by betz's theory. *Expert Systems with Applications*, 193, 116509.

Apéndice A

Descripción de las bases de datos utilizadas

A.1. Datos *benchmark*

La primera base de datos *benchmark* corresponde a ‘Austourism’ (Hyndman, 2015) y contiene datos trimestrales de turismo en Australia, correspondiente a la cantidad de noches que los pasajeros se quedan de visita en el país entre 1998 y 2011 para varias regiones. Este conjunto de datos tiene la siguiente estructura (ver figura A.1):

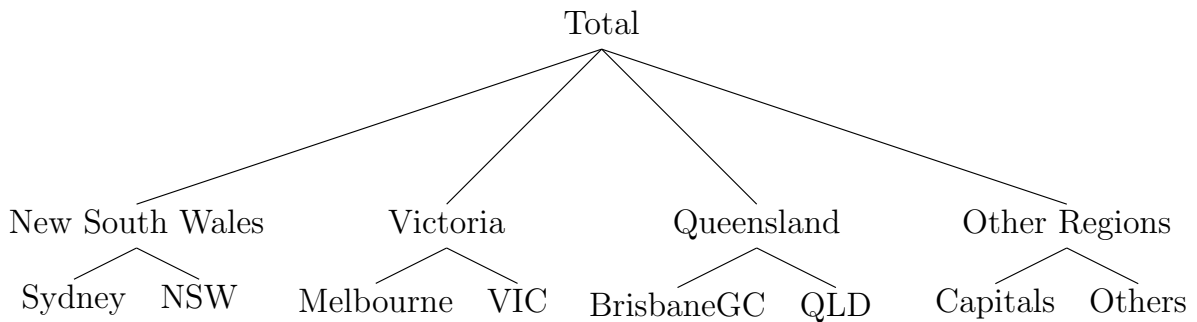


Figura A.1: Estructura jerárquica de la base de datos ‘Austourism’.

donde,

- Sydney: el área metropolitana de Sydney.
- NSW: New South Wales sin considerar Sydney.
- Melbourne: el área metropolitana de Melbourne.

- VIC: Victoria sin considerar Melbourne.
- BrisbaneGC: la áreas Brisbane y Gold Coast.
- QLD: Queensland sin considerar las áreas Brisbane y Gold Coast.
- Capitals: las otras cinco ciudades capitales (Adelaide, Hobart, Perth, Darwin y Canberra).
- Other: Otras áreas de Australia.

La segunda base de datos corresponde a ‘Departures’, que estudia a nivel mensual la cantidad de salidas de Australia hacia otros países. Su estructura es la siguiente:

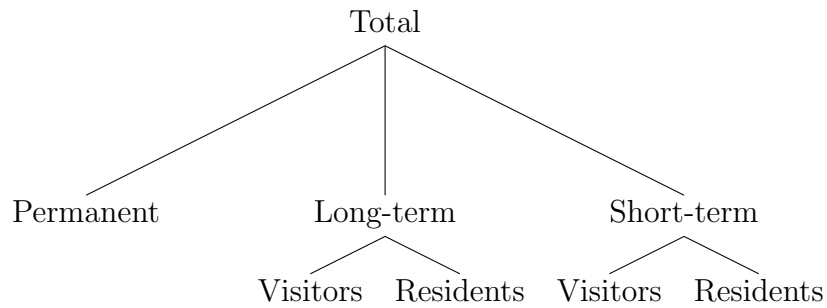


Figura A.2: Estructura jerárquica de la base de datos ‘Departures’.

donde,

- Permanent: salidas permanentes desde Australia.
- Long-term Residents: residentes de largo plazo que se van de Australia.
- Long-term Visitors: visitantes de largo plazo que se van de Australia.
- Short-term Residents: residentes de corto plazo que se van de Australia.
- Short-term Visitors: visitantes de corto plazo que se van de Australia.

La tercera base de datos se denomina ‘Htseg1’, y considera datos simulados para estudios estadísticos. Considera tres niveles, con un total de cinco series de tiempo en el último nivel, dos en el nivel intermedio, y la serie de tiempo total. Esta serie se ilustra en la figura A.3.

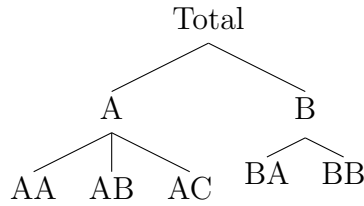


Figura A.3: Estructura jerárquica de la base de datos ‘Htseg1’.

El cuarto conjunto de datos *benchmark* se llama ‘Htseg2’. Considera una jerarquía de cuatro niveles, con un total de 17 series de tiempo individuales. Este conjunto de datos tiene dos series temporales en el primer nivel: A y B . Cada nodo, A y B , se desagrega en dos nuevas series, de las cuales nacen los diez nodos del nivel más bajo. Esto se muestra en la figura A.4.

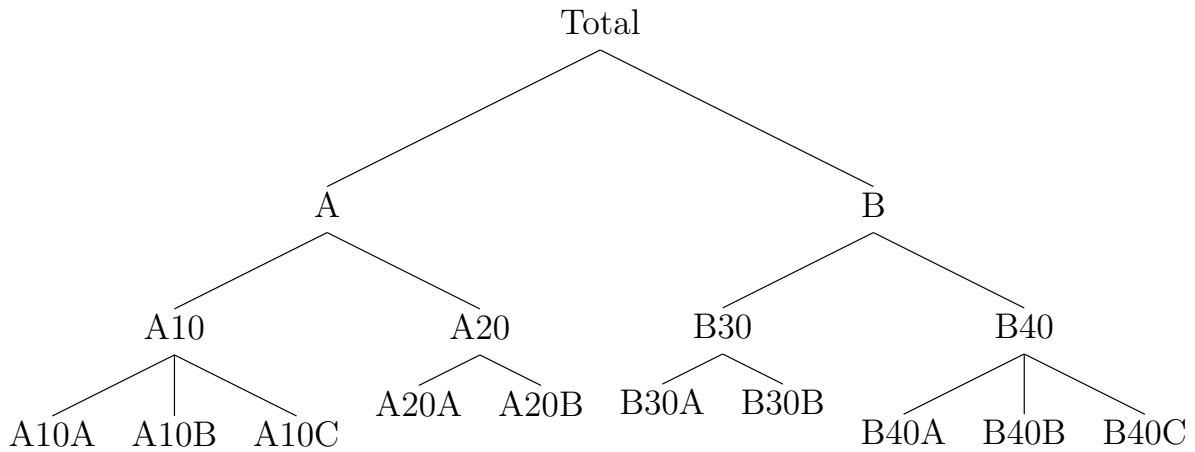


Figura A.4: Estructura jerárquica de la base de datos ‘Htseg2’.

La última base de datos *benchmark* corresponde a ‘Melsyd’, que contiene vuelos entre Melbourne y Sydney para la aerolínea Ansett. Estos datos tienen agregación semanal, y se presentan en la figura A.5. La estructura jerárquica está definida por el tipo de boleto aéreo, considerando las clases económica, business y primera clase.

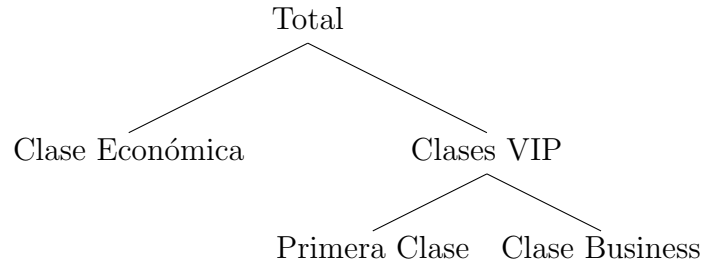


Figura A.5: Estructura jerárquica de la base de datos ‘Melsyd’.

El cuadro A.1 muestra un resumen de todas las bases de datos *benchmark*.

Base	# Series	Periodicidad	# Observaciones	Periodo	Tipo
Austourism	13	Trimestral	56	1998 - 2011	Datos reales
Departures	8	Mensual	434	1976 - 2012	Datos reales
Htseg1	17	Anual	16	1992 - 2007	Datos simulados
Htseg2	17	Anual	16	1992 - 2007	Datos simulados
Melsyd	5	Semanal	283	1987 - 1992	Datos reales

Cuadro A.1: Cuadro resumen de los conjuntos de datos *benchmark*.

A.2. Datos Travel Retail

Las bases de datos de la industria del Travel Retail fueron entregadas por la empresa sueca Generation Research AB, y considera datos de la sub-industria de productos de belleza. Se consideran datos trimestrales de la venta de productos de belleza reportados en todos los canales de venta y ubicaciones de Europa entre el primer trimestre de 2013 y el segundo trimestre de 2016 (catorce trimestres). En base a estos datos, se generan seis series de tiempo jerárquicas.

La figura A.6 presenta la serie de tiempo total para los datos de Generation Research. En este caso se tiene la venta de 20,802 productos de belleza distintos en la industria del Travel Retail, reportadas por 82 empresas en Europa entre 2013 y 2016.

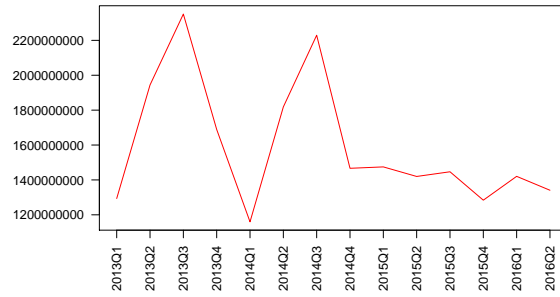
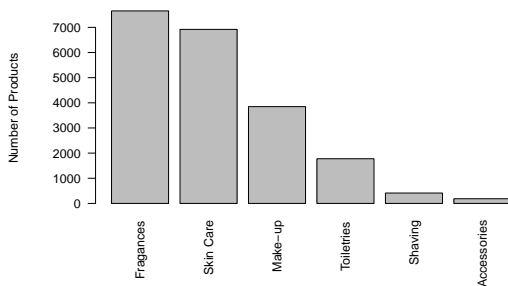


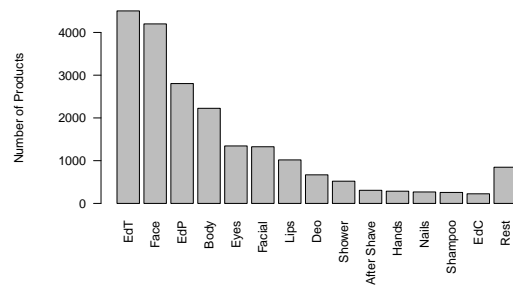
Figura A.6: Serie de tiempo total para los datos de Generation Research.

Las categorías utilizadas para la construcción de las seis series de tiempo jerárquicas corresponden a:

- Company: La empresa que reporta la venta.
- Brand: La marca del producto.
- Product line: Descripción única del producto, relacionada a su SKU.
- First segmentation: Primera segmentación basada en la categoría del producto. Las categorías identificadas son: *fragrances*, *skin-care*, *make-up*, *toiletries*, *shaving* y *accessories*. En la figura A.7(a) se presenta la distribución de frecuencia para estas categorías.



(a) First Segmentation



(b) Second Segmentation

Figura A.7: Análisis descriptivo de la primera y segunda segmentación.

- Second segmentation: La segunda segmentación define nuevas categorías entre los segmentos de la primera segmentación. Los productos de *shampoo*, *perfumes*, *eye*, *facial* y *lip* son algunos de los ejemplos de esta segmentación. *Rest* contiene todas las categorías que presentan menos de 225 productos. En este caso, el gráfico de frecuencia está disponible en la figura A.7(b).
- Third segmentation: La tercera segmentación está basada en el formato de los productos. Algunos ejemplos son *cream*, *stick*, *spray*, o *powder*. La figura A.8(a) presenta su respectivo gráfico de frecuencia.

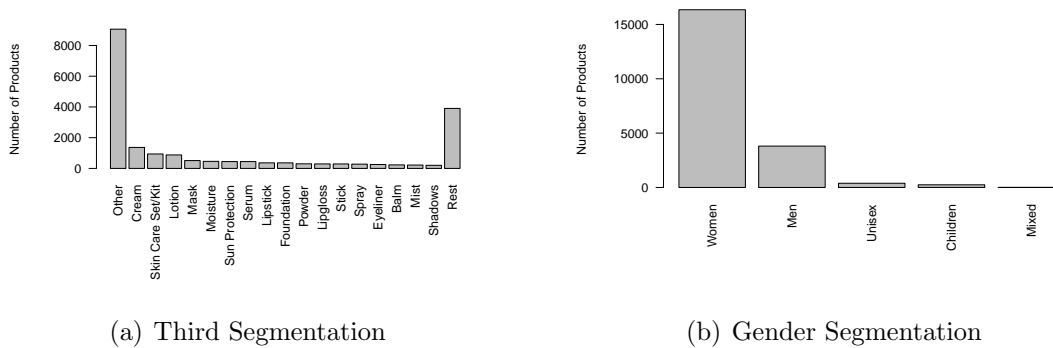


Figura A.8: Análisis descriptivo de la tercera y cuarta segmentación.

- Gender segmentation: Segmentación basada en el género, que incluye las categorías: *Women*, *Men*, *Unisex*, *Children* y *Mixed*. Un gráfico de frecuencia para esta segmentación se presenta en la figura A.8(b).
- Target market segmentation: Segmentación que representa si el producto corresponde a mercado premium (*premium market*) o masivo (*massive market*). El gráfico de frecuencia respectivo se presenta en la figura A.9(a).

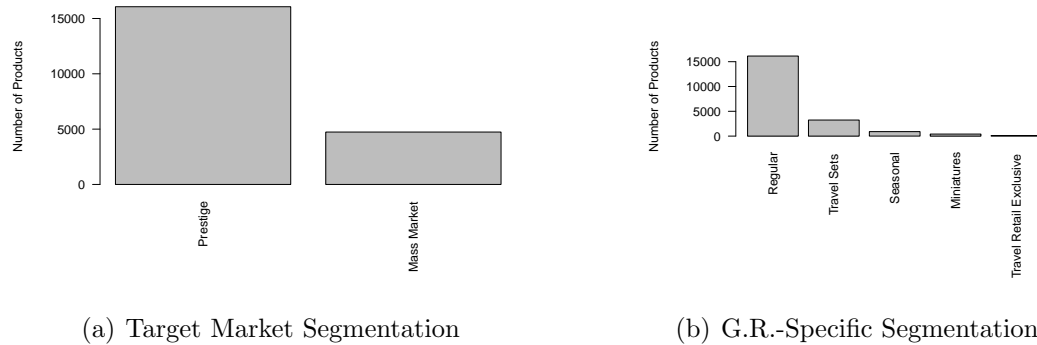


Figura A.9: Análisis descriptivo de la quinta y sexta segmentación.

- G.R.-specific segmentation: Esta es una clasificación definida por la empresa Generation Research, y contiene información adicional de cada producto. Las categorías incluídas en esta segmentación corresponden a: *Travel Set*, *Seasonal Product*, *Miniatures* y *Travel Retail Exclusives*. Su gráfico de frecuencia se reporta en la figura A.9(b).

La información de las empresas que venden los productos no está disponible, puesto que se considera información confidencial. La información con la que se dispone es:

- 15 empresas venden el 80 % de los productos (16,706). Los restantes 4,605 productos son ofrecidos por 66 empresas distintas.
- En total hay 359 marcas de productos. El 80 % de los productos están asociados solo a 69 marcas.
- El 80 % de los productos contienen 2,855 product lines diferentes, mientras el el total de product lines es 6,217.

La base de datos considera 20,801 productos diferentes, cuyas ventas fueron registradas en todos o alguno de los 14 periodos de tiempo estudiados. La figura A.10 representa el total de ventas en dólares para cada periodo estudiado.

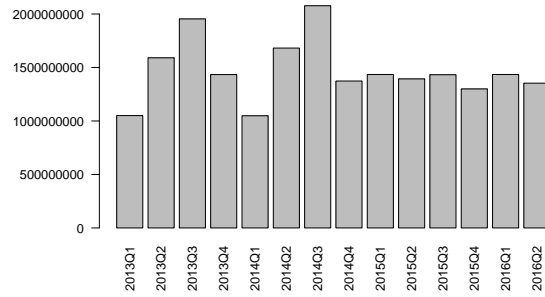


Figura A.10: Ventas totales para todas las compañías del Travel Retail en Europa.

En base a esta información, se generan las seis estructuras jerárquicas con los datos de Generation Research, y se presentan en el cuadro A.2.

Nombre Referencial	Nivel 1	Nivel 2	Nivel 3	Nivel 4	Nivel 5
1 Company	Company	Brand	Product Line	Product	-
2 Market	First Segmentation	Second Segmentation	Third Segmentation	Product	-
3 Gender	Gender	Product	-	-	-
4 LuxuryS	Target Market Segmentation	First Segmentation	Second Segmentation	Third Segmentation	Product
5 LuxuryC	Company	Target Market Segmentation	Brand	Product Line	Product
6 Other	G.R.-specific segmentation	Product	-	-	-

Cuadro A.2: Todas las series de tiempo jerárquicas de Generation Research.

A.3. Datos Contact Center

Para los datos de Contact Center, se cuenta con información de negocio entregada por la empresa CrossNet S.A., quienes son una empresa proveedora de plataformas de Contact Center. Dentro de sus clientes, se encuentran grandes empresas de Latinoamérica. A su vez, algunos de los clientes de CrossNet nos facilitan información para esta investigación. Una de las empresas que entrega la información corresponde a la Asociación Chilena de Seguridad (ACHS), cuyos datos fueron los utilizados para la cuarta investigación de esta tesis.

Los datos de ACHS que se consideran corresponden a series de tiempo con cantidad de llamadas entrantes en rangos de 15 minutos, entre septiembre de 2017 y diciembre de 2021. La estructura jerárquica es basada en las líneas de entrada de sus llamadas telefónicas, separando en un primer nivel el total de llamadas en alguno de sus 18

servicios asociados, y posteriormente, cada servicio, en uno o más puntos de ruteo (RP) específicos. En total, se tienen 41 series de tiempo en el último nivel y 18 series en el segundo nivel.

Servicio	# Puntos de ruteo
APC	1
C19	1
CAP	5
CEM	2
COM	4
EME	3
EPR	2
GES	1
INF	5
PPE	1
PRE	2
SEL	1
SSA	4
SUB	2
SWE	2
TRA	2
URG	2
OTR	1

Cuadro A.3: Los 18 servicios considerados para la base de ACHS con sus respectiva cantidad de RP que nacen de ellos.

Apéndice B

Configuraciones Experimentales

Este apéndice detalla los procesos de pre-procesamiento de datos y técnicas de validación utilizadas a lo largo de la tesis.

Para el primer trabajo, relacionado a la predicción de series de tiempo jerárquicas vía SVR en la industria del Travel Retail, además de la prueba en cuatro bases de datos *benchmark*, se puede resumir lo siguiente:

Se puede apreciar que las bases de datos *benchmark* utilizadas en estos estudios no poseen valores perdidos, pero tanto las bases de Generation Research, como las bases de ACHS poseen una serie de valores perdidos a lo largo de sus periodos. Basados en el trabajo de Liu et al. (Liu, Yan, Yang & Hauskrecht, 2015), se exploraron cuatro métodos de imputación de valores perdidos, correspondientes a el último valor transportado hacia adelante y hacia atrás, interpolación, máquina de media predictiva y *random forest*. De estos enfoques, *random forest* fue el que presentó los mejores resultados, por lo que fue el elegido para el estudio.

Con respecto a la validación y selección de modelos, se dividieron los datos en conjuntos de entrenamiento y testeo, para luego dividir nuevamente el conjunto de entrenamiento en subconjuntos de entrenamiento y validación. El diseño de estos conjuntos asegura que el subconjunto de entrenamiento contiene solo observaciones previas a las presentes en el subconjunto de validación, y éste, a su vez, solo posee obervaciones previas al conjunto de testeo (Hyndman & Athanasopoulos, 2013). Esta estrategia se ilustra en el cuadro B.1.

En los experimentos de los datos *benchmark* y de Generation Research, se utilizó la última observación del conjunto de entrenamiento para la validación, mientras que en

- $\varepsilon \in \{0, 0, 1, 0, 2, \dots, 1\}$.
- $\gamma \in \{0, 0, 1, 0, 2, \dots, 1\}$.
- $C \in \{2^{-15}, 2^{-14}, 2^{-13}, \dots, 2^{15}\}$.
- $\theta \in \{2^{-15}, 2^{-14}, 2^{-13}, \dots, 2^{15}\}$.

Mientras que para el tercer trabajo (KAT-SVR), la configuración de las grillas se hizo en base a los rangos de hiperparámetros mostrados a continuación:

- $C \in \{2^{-25}, 2^{-24}, 2^{-23}, \dots, 2^{25}\}$.
- $\varepsilon \in \{0, 0, 1, 0, 2, \dots, 1\}$.
- $\gamma \in \{0, 0, 1, 0, 2, \dots, 1\}$.
- $\theta \in \{2^{-25}, 2^{-24}, 2^{-23}, \dots, 2^{25}\}$.

Apéndice C

Desempeño de cada conjunto de datos

Este anexo presenta los resultados detallados para cada una de las bases de datos utilizadas en el primer trabajo, relacionado a la predicción de HTS via SVR para el Travel Retail.

Modelo	MAPE Promedio				RANKING				Ranking Promedio
	Total	Nivel 1	Nivel 2	Nivel 3	Total	Nivel 1	Nivel 2	Nivel 3	
1 ARIMA-MO	0.23	2.52	2.34	7.58	2	3	2	2	2.25
2 HW-MO	2.24	2.66	2.59	8.22	4	4	3	4	3.75
3 HW-BU	10.12	2.45	2.07	8.36	7	2	1	5	3.75
4 ARIMA-BU	0.23	3.78	3.05	7.82	3	6	4	3	4
5 SVR-MO	4.9	1.6	8.53	27.98	5	1	6	6	4.5
6 SVR-BU	12.63	3.49	3.53	6.49	8	5	5	1	4.75
7 ARIMA-TD	0.232	318.288	355.287	448.31	1	8	8	8	6.25
8 HW-TD	19.41	236.75	266.43	340.11	9	7	7	7	7.5
9 SVR-TD	8.87	600.18	650.61	785.9	6	9	9	9	8.25

Cuadro C.1: Resultados detallados de ε -SVR para la base de datos ‘Htseg2’.

Modelo	MAPE Promedio			RANKING			Ranking Promedio
	Total	Nivel 1	Nivel 2	Total	Nivel 1	Nivel 2	
1 HW-BU	1.92	5	6.12	1	2	2	1.67
2 HW-MO	2.32	3.72	5.15	4	1	1	2
3 SVR-BU	2.29	5.89	7.26	3	4	5	4
4 ARIMA-BU	2.34	5.67	6.58	6	3	3	4
5 HW-TD	2.02	18	64.17	2	8	8	6
6 SVR-MO	2.9	5.95	34.49	8	5	6	6.33
7 ARIMA-MO	2.97	6.42	6.92	9	6	4	6.33
8 SVR-TD	2.72	10.31	39.09	7	7	7	7
9 ARIMA-TD	2.33	19.44	64.74	5	9	9	7.67

Cuadro C.2: Resultados detallados de ε -SVR para la base de datos ‘Departures’.

Modelo	MAPE Promedio			RANKING			Ranking Promedio
	Total	Nivel 1	Nivel 2	Total	Nivel 1	Nivel 2	
1 ARIMA-TD	9.99	24.46	32.21	1	1	2	1.33
2 HW-TD	13.73	27.02	32.47	4	2	3	3
3 SVR-BU	15.94	29.37	31.09	6	3	1	3.33
4 SVR-TD	15.15	29.56	34.82	5	4	4	4.33
5 ARIMA-BU	11.05	41.48	48.57	2	6	6	4.67
6 ARIMA-MO	12.46	42.55	51.23	3	7	7	5.67
7 SVR-MO	18.62	36.07	35.2	9	5	5	6.33
8 HW-MO	16.66	48.53	57.71	7	8	8	7.67
9 HW-BU	17.77	48.73	57.93	8	9	9	8.67

Cuadro C.3: Resultados detallados de ε -SVR para la base de datos ‘Melsyd’.

Modelo	MAPE Promedio			RANKING			Ranking Promedio
	Total	Nivel 1	Nivel 2	Total	Nivel 1	Nivel 2	
1 SVR-BU	2.6	6.28	7.42	1	1	1	1
2 HW-BU	5.2	6.78	8.81	5	2	2	3
3 SVR-MO	3.32	6.98	9.86	3	3	6	4
4 HW-MO	5.89	7.27	9.24	7	4	3	4.67
5 SVR-TD	2.98	10.45	10.7	2	7	7	5.33
6 ARIMA-MO	5.72	7.99	9.75	6	6	4	5.33
7 ARIMA-BU	6.1	7.36	9.81	8	5	5	6
8 HW-TD	5.08	10.61	11.37	4	8	8	6.67
9 ARIMA-TD	6.26	11.25	11.83	9	9	9	9

Cuadro C.4: Resultados detallados de ε -SVR para la base de datos ‘Austourism’.

Modelo	MAPE Promedio						RANKING						Ranking Promedio
	Total	Nivel 1	Nivel 2	Nivel 3	Nivel 4	Nivel 5	Total	Nivel 1	Nivel 2	Nivel 3	Nivel 4	Nivel 5	
1 HW-BU	6.12	647.15	571.2	753.91	2444.67	3847.16	3	1	1	2	2	1	1.67
2 HW-MO	6.96	651.99	575.5	734.82	2444.64	3893.16	4	2	2	1	1	2	2
3 SVR-BU	4.45	816.15	717.29	813.35	4246.08	6160.91	2	3	3	3	3	3	2.83
4 SVR-TD	0.69	833.68	733.47	1084.45	8145.13	8763.1	1	5	5	4	6	6	4.5
5 SVR-MO	8.32	832.09	732.65	1125.61	9171.24	9852.88	5	4	4	5	7	7	5.33
6 ARIMA-BU	9.64	947.13	831.67	1309.65	5757.59	6753.8	6	6	6	9	4	5	6
7 ARIMA-MO	10.01	948.65	833.13	1249.62	5757.66	6677.7	7	7	7	7	5	4	6.17
8 ARIMA-TD	18.26	967.62	850.51	1230.84	10708.3	11238.4	8	8	8	6	8	8	7.67
9 HW-TD	22.11	1001.55	880.36	1272.57	11058.47	11605.78	9	9	9	8	9	9	8.83

Cuadro C.5: Resultados detallados de ε -SVR para la base de datos de Generation Research, bajo la estructura LuxuryC.

Modelo	MAPE Promedio						RANKING						Ranking Promedio
	Total	Nivel 1	Nivel 2	Nivel 3	Nivel 4	Nivel 5	Total	Nivel 1	Nivel 2	Nivel 3	Nivel 4	Nivel 5	
1 HW-BU	6.12	4.27	11.98	36.7	692.58	3847.16	4	3	2	3	1	1	2.33
2 SVR-BU	4.45	3.55	12.76	33.49	850.3	6160.91	3	2	4	1	3	3	2.67
3 SVR-TD	0.69	2.31	10.85	34.87	934.6	8763.39	1	1	1	2	6	6	2.83
4 SVR-MO	0.88	4.85	12.7	37.32	969.67	8862.74	2	4	3	4	7	7	4.5
5 HW-MO	17.57	12.05	14.99	38.56	698.25	4211.69	7	7	6	5	2	2	4.83
6 ARIMA-BU	9.64	7.16	13.58	45.81	900.59	6753.8	5	5	5	7	4	4	5
7 ARIMA-MO	15.14	11.45	15.04	45.69	904.46	7204.24	6	6	7	6	5	5	5.83
8 ARIMA-TD	18.26	16.23	21.21	49.29	1115.57	11238.4	8	8	8	8	8	8	8
9 HW-TD	22.11	20.02	24.35	52.55	1153.23	11605.78	9	9	9	9	9	9	9

Cuadro C.6: Resultados detallados de ε -SVR para la base de datos de Generation Research, bajo la estructura LuxuryS.

Modelo	MAPE Promedio					RANKING					Ranking Promedio
	Total	Nivel 1	Nivel 2	Nivel 3	Nivel 4	Total	Nivel 1	Nivel 2	Nivel 3	Nivel 4	
1 HW-BU	6.12	647.15	753.91	2444.67	3847.16	3	1	2	2	1	1.8
2 HW-MO	6.96	651.99	743.4	2444.64	3893.16	4	2	1	1	2	2
3 SVR-BU	4.45	816.15	813.35	4246.08	6160.91	2	3	3	3	3	2.8
4 SVR-TD	0.72	833.38	1084.06	8142.21	8759.99	1	5	4	6	6	4.4
5 SVR-MO	8.32	832.09	1125.61	9171.24	9852.88	5	4	5	7	7	5.6
6 ARIMA-MO	10.01	948.65	1258.43	5757.66	6677.7	7	7	7	5	4	6
7 ARIMA-BU	9.64	947.13	1309.65	5757.59	6753.8	6	6	9	4	5	6
8 ARIMA-TD	18.26	967.62	1230.84	10708.3	11238.4	8	8	6	8	8	7.6
9 HW-TD	22.11	1001.55	1272.57	11058.47	11605.78	9	9	8	9	9	8.8

Cuadro C.7: Resultados detallados de ε -SVR para la base de datos de Generation Research, bajo la estructura Company.

Modelo	MAPE Promedio					RANKING					Ranking Promedio
	Total	Nivel 1	Nivel 2	Nivel 3	Nivel 4	Total	Nivel 1	Nivel 2	Nivel 3	Nivel 4	
1 SVR-TD	0.69	8.41	25.11	328.22	8764	1	1	1	3	7	2.6
2 SVR-BU	4.45	11.17	28.68	357.62	6160.91	3	3	3	2	3	2.8
3 SVR-MO	1.54	12.08	27.54	357.57	8397.93	2	5	2	1	6	3.2
4 HW-BU	6.12	11.01	36.05	425.92	3847.16	4	2	5	6	1	3.6
5 ARIMA-BU	9.64	11	41.03	418	6753.8	5	4	8	4	4	5
6 HW-MO	18.33	16.03	38.2	429.91	4181.49	8	6	7	7	2	6
7 ARIMA-MO	16.13	16.9	41.88	423.54	7241.2	6	7	9	5	5	6.4
8 ARIMA-TD	18.26	17.01	35.28	479.47	11238.4	7	8	4	8	8	7
9 HW-TD	22.11	19.2	37.92	496.45	11605.78	9	9	6	9	9	8.4

Cuadro C.8: Resultados detallados de ε -SVR para la base de datos de Generation Research, bajo la estructura Market.

Modelo	MAPE Promedio			RANKING			Ranking Promedio
	Total	Nivel 1	Nivel 2	Total	Nivel 1	Nivel 2	
1 SVR-BU	4.45	10.1	6160.91	2	2	3	2.33
2 HW-BU	6.12	10.47	3847.16	3	3	1	2.33
3 SVR-TD	0.69	9.59	8763.26	1	1	6	2.67
4 ARIMA-BU	9.64	12.88	6753.8	5	5	4	4.67
5 SVR-MO	6.17	11.13	9297.06	4	4	7	5
6 ARIMA-MO	17.06	15.23	7085.53	6	6	5	5.67
7 HW-MO	20.85	15.88	4320.54	8	7	2	5.67
8 ARIMA-TD	18.26	17.05	11238.4	7	8	8	7.67
9 HW-TD	22.11	19.99	11605.78	9	9	9	9

Cuadro C.9: Resultados detallados de ε -SVR para la base de datos de Generation Research, bajo la estructura Other.

Modelo	MAPE Promedio			RANKING			Ranking Promedio
	Total	Nivel 1	Nivel 2	Total	Nivel 1	Nivel 2	
1 HW-BU	6.12	11.84	3847.16	4	2	1	2.33
2 SVR-TD	0.68	6.36	8765.33	1	1	7	3
3 SVR-BU	4.45	15.19	6160.91	3	4	3	3.33
4 ARIMA-BU	9.64	13.19	6753.8	5	3	4	4
5 SVR-MO	2.74	17.53	8143.58	2	5	6	4.33
6 ARIMA-MO	18.26	18.17	7257.12	6	6	5	5.67
7 HW-MO	21.92	18.71	4423.81	8	7	2	5.67
8 ARIMA-TD	18.26	21.12	11238.4	7	8	8	7.67
9 HW-TD	22.11	25.07	11605.78	9	9	9	9

Cuadro C.10: Resultados detallados de ε -SVR para la base de datos de Generation Research, bajo la estructura Gender.

Apéndice D

Mejores parámetros para los modelos de ε -SVR

A lo largo de las diferentes investigaciones, se utiliza como método predictivo ε -SVR, ya sea como la metodología propuesta (en el primer trabajo), o bien como punto de inicio para las otras investigaciones, en busca de mejorar los modelos. En las distintas investigaciones se utilizan diversas bases de datos. En el caso del primer trabajo, se consideran las bases *benchmark* ‘Htseg2’, ‘Austourism’, ‘Melsyd’, ‘Departures’, además de las 6 bases de datos de la empresa sueca Generation Research. Para el segundo y tercer trabajo, se utilizan solo las bases de datos de *benchmark*, incluyendo un quinto conjunto de datos: ‘Htseg1’. En el caso del cuarto trabajo, la base corresponde a llamadas de Contact Center de la empresa CrossNet, que contiene datos de las llamadas que recibe la Asociación Chilena de Seguridad (ACHS). Los mejores parámetros reportados para los modelos de ε -SVR para cada base de datos corresponden a los siguientes:

		Kernel lineal		Kernel gaussiano		
		ε	C	ε	γ	C
<i>ds1</i>	Austourism	0.3	2^0	0.1	0.1	2^0
<i>ds2</i>	Departures	0.1	2^{-3}	0.1	0.2	2^{-1}
<i>ds3</i>	Htseg1	0.2	2^0	0.1	0.1	2^3
<i>ds4</i>	Htseg2	0.3	2^4	0.1	0.1	2^6
<i>ds5</i>	Melsyd	0.5	2^{-1}	0.1	0.9	2^{-1}
<i>ds6 – 11</i>	Generation Research	0	2^{-7}	0.1	0.1	2^9
<i>ds12</i>	ACHS	0	2^3	0.1	2^{-7}	2^9

Cuadro D.1: Mejores parámetros para las variantes de ε -SVR.