



UNIVERSIDAD CATÓLICA DEL MAULE
FACULTAD DE CIENCIAS DE LA INGENIERÍA
ESCUELA INGENIERÍA CIVIL INFORMÁTICA

Desarrollo de un método de predicción de la demanda eléctrica mediante técnicas de aprendizaje automático, para el estudio de ciclos de combustible nuclear.

JAVIER ALONSO SOTO LETELIER
PROFESOR GUÍA: DRA. XAVIERA LOPEZ
PROFESOR CO-GUÍA: DR. IVÁN MERINO

Proyecto de título para optar al Título Profesional de Ingeniero Civil Informático.

TALCA, MAYO 2023



UNIVERSIDAD CATÓLICA DEL MAULE
FACULTAD DE CIENCIAS DE LA INGENIERÍA
ESCUELA INGENIERÍA CIVIL INFORMÁTICA

Proyecto de título para optar al Título Profesional de Ingeniero Civil Informático.

PROYECTO DE TÍTULO
JAVIER ALONSO SOTO LETELIER

COMISIÓN EXAMINADORA

FIRMA

Profesor de comisión interna
Dr. Ruber Hernández
Universidad Católica del Maule

Profesor de comisión externa
Dr. Alejandro Prada Valverde
Universidad Católica del Maule

Profesora Guía
Dra. Xaviera Lopez
Universidad Católica del Maule

NOTA FINAL EXAMEN DE TÍTULO

TALCA, MAYO 2023

Agradecimientos

Este proyecto de titulación no habría sido posible sin todas las personas que me rodean: amigos, mi papá, mi hermano Sebastián, mi abuela materna, como también mi mamá, que lamentablemente durante este último año me ha dejado, siendo mi inspiración para seguir estudiando y lograr siempre conseguir lo mejor de mí. Aunque mi mayor enemigo siempre sea yo mismo, agradezco a las personas que conocí en mi pasantía en la SCK-CEN, que me enseñaron que siempre se puede aprender más. También doy gracias a mis profesores, que me han enseñado a aprender, al profesor Iván Merino, que me dio la oportunidad de viajar y seguir aprendiendo, al profesor Ruber Hernández, que me mostró que siempre hay más cosas para aprender en Informática, a la profesora Xaviera López que hizo que esta tesis tuviera sentido y dirección, como también a mi equipo de estudio y trabajo, el equipo 12: Christian, Marcelo y Felipe. Aunque hayamos tenido nuestras discusiones, hemos aprendido juntos y de cada uno.

Javier Soto Letelier.

Mayo, 2023

Abstract

« When we think about what electricity production entails, we don't see how much is actually spent in terms of resources and money, since this involves a large number of expenses to produce what is necessary for the country and/or region. That is why it is important to think before the total demand exceeds the total generation. This is where Machine Learning comes into play, preparing for future cases that will arise in the next years or decades and preventing them from becoming an unpleasant surprise.»

Keywords: MACHINE LEARNING, PROCESAMIENTO DE DATOS, ELECTRICIDAD, GENERACIÓN, PREDICCIÓN, RENOVABLES, ENERGÍAS LIMPIAS, DATASET, REGRESIÓN LINEAL.

Índice general

1. Introducción	1
1.1. Hipótesis	2
1.2. Problema u oportunidad	2
1.2.1. Contexto del proyecto	2
1.3. Objetivos	3
1.3.1. Objetivos Generales	3
1.3.2. Objetivos Específicos	3
1.4. Resultados Esperados	4
1.5. Estructura del Documento	4
2. Marco Teórico	5
2.1. Introducción	6
2.2. Machine Learning	6
2.3. Regresión Lineal	6
2.4. Series de Tiempo	7
2.5. Bases de Datos	7
2.6. Conclusiones	8
3. Estado del Arte	9
3.1. Energía Nuclear	10
3.2. Trabajos Relacionados	11
4. Metodología	13
4.1. Introducción	14
4.2. Recursos utilizados	14
4.3. Ruta de trabajo	14
4.4. Origen de los datos	15
4.5. Los datos	15
4.6. Preprocesamiento de los datos	16
4.6.1. Formato de los datos	16
4.7. Análisis de los datos	16
4.7.1. Skewness y Kurtosis	16
4.7.2. Tipos de datos	17
4.8. Transformación a Datos Estacionarios	18
4.8.1. Transformación logarítmica	18
4.9. Tendencias	20

4.9.1. Hidráulica	20
4.9.2. Eólica	20
4.9.3. Solar	21
4.10. Elección de Modelo	21
4.10.1. Prophet by Facebook	22
4.11. Archivo de Entrada	22
4.12. Entrenamiento del Modelo	24
4.12.1. HiperParámetros	24
4.12.1.1. Hiperparámetros para Prophet	26
4.13. Módulo de patrones	26
4.13.1. Obtención de variaciones diarias	27
4.14. Módulo de Predicción	28
4.14.1. Procesamiento Matemático	28
4.15. Módulo de generación de bases de datos zonales	29
5. Planificación	31
5.1. Plan de trabajo	32
5.2. Carta Gantt Proyecto de Titulación	32
6. Resultados	33
6.1. Introducción	34
6.2. Archivo de entrada utilizado	34
6.3. Resultados	37
6.3.1. Resultados producción energía nuclear	37
6.3.2. Caso Simple	39
6.3.3. Caso Complejo	39
6.3.4. Producción eléctrica	41
6.3.5. Análisis de casos	42
6.4. Producto de la tesis	44
7. Discusión y conclusiones	45
7.1. Discusión y conclusión	46
7.2. Trabajo futuro	46
Bibliografía	48
A. Anexo	50

Índice de Figuras

1.1. Generación Eléctrica en Chile por tipo de energía.	3
4.1. Diagrama de flujo.	15
4.2. Tendencia producción Hidráulica.	20
4.3. Tendencia producción Eólica.	21
4.4. Tendencia producción Solar.	21
4.5. Flujo de trabajo Prophet.	22
4.6. Diagrama de flujo Módulo de Patrones.	26
4.7. Diagrama de flujo Módulo de Predicción.	28
5.1. Carta Gantt	32
6.1. Caso Simple: Generación Energía Nuclear 2022-2050.	37
6.2. Caso Complejo: Generación Energía Nuclear 2022-2050.	38
6.3. Caso Simple: Producción Total a través de los años 2022-2050.	39
6.4. Caso Complejo: Producción Total a través de los años 2022-2050.	40
6.5. Generación Energía Eólica 2021-2022 Datos originales Vs Predicción.	42
A.1. Tendencia producción Solar,Hidráulica y Eólica, Zona Norte, Centro y Sur. .	50

Índice de Tablas

4.1. Recursos utilizados.	14
4.2. Ejemplo del formato usado.	16
4.3. Skewness y Kurtosis para las principales energía Belgas.	17
4.4. Skewness y Kurtosis para las principales energía Chilenas.	17
4.5. p-Value para las principales energía Chilenas.	18
4.6. p-Value para las principales energía Belgas.	18
4.7. Variables de Entrada.	23
4.9. Hiperparámetros Prophet.	26
4.10. Porcentajes Creados para Hidráulica,Eólica,Solar.	27
4.12. Modificación de las Variables de Entrada.	30
6.1. Variables de Entrada.	35
6.3. Primeros diez Resultados Pipeline Chile.	41
6.4. Primeros diez Resultados Pipeline Chile 2.	41
6.5. Fechas Finales Predicción.	42

Índice de Código

4.8. Diccionario Contribución y crecimiento Energías.	23
4.11. Código Prophet Obtención variaciones Diarias Y Mensuales	27
4.13. Definición diccionario macrozonas	30
6.2. Diccionario Contribución y crecimiento Energías.	35

Capítulo 1

Introducción

En este capítulo se presenta el tema, problema y oportunidad, se situará en un contexto donde se describirán los objetivos (general y específicos) y lo que se espera del trabajo o contribuciones de este.

1.1. Hipótesis

Chile camina hacia un futuro sostenible donde las energías renovables reemplazarán los combustibles fósiles. Así, para el año 2050, se espera que la matriz energética sea carbono neutral, es decir, que la emisión de gases de cambio climático sea equivalente a lo que puedan ser absorbidos por la naturaleza. Sin embargo, la particularidad de las energías renovables por las que apuesta Chile, principalmente solar y eólica, son de carácter variable.

Dentro de unos de estos objetivos planteados a nivel país como a nivel mundial es la eliminación de la generación eléctrica basada en energías fósiles, para ello es necesario reemplazarlas por fuentes renovables, pero para ello primero es necesario conocer cuáles son los tipos de energías más empleados y como va aumentando la demanda de cada país para conocer a que ritmo es necesario realizar la transición energética, un ejemplo en la eliminación de toda fuente de carbón en la generación eléctrica en Chile para el año 2050, desafío planteado por diversos gobiernos, el cual tomando en cuenta que el 30 % de toda la energía producida viene de esta fuente, siendo esta una de las principales generadoras, el cual significaría un enorme gasto público para el cual es necesario realizar los estudios correspondientes para generar la electricidad necesaria y ser económicamente responsable, otro caso es el escenario Belga, puesto que aunque no exista una gran producción eléctrica a base de carbón si se utiliza otras fuentes fósiles como son la generación eléctrica en base de Gas y combustibles, que representan un 40 % de la matriz energética, pero que a contrario de Chile la generación utilizando energías renovables ha venido creciendo de forma sostenida en forma de energía eólica y con una fuerte base gracias la energía Nuclear considerado una energía limpia por la Unión Europea.

1.2. Problema u oportunidad

Este trabajo se enmarca en el Proyecto de Investigación Colaborativa “Evaluación Técnica y Optimización de Sistemas Híbridos de Energía Nuclear-Renovable” de la Organización Internacional de Energía Atómica (OIEA), enfocándose en el desarrollo de técnicas de Machine Learning para la predicción de la demanda eléctrica de un país o región. Se utilizarán los datos de la matriz energética chilena obtenidos de la Web de la CNE (Comisión Nacional de Energía), los cuales serán procesados y utilizados en el método de predicción final. La finalidad del desarrollo es utilizar la información obtenida, para establecer la base para simulaciones realistas de desarrollo energético de un país o región. Así, con el desarrollo de este proyecto, se crearán los inputs necesarios para el estudio de diferentes estrategias de desarrollo nuclear mediante el código de simulación de ciclos de combustible ANICCA.

1.2.1. Contexto del proyecto

Según la Comisión Nacional de Energía de Chile (CNE), en su Balance Energético anual del año 2020, alrededor del 67 % de la energía total consumida en Chile se produjo a partir de combustibles fósiles, de los cuales, el 40 % se destinó a la producción de energía eléctrica, lo que representa el 55 % de la potencia actual. Cabe destacar, que Chile importa la totalidad de este tipo de combustibles, los que muestra una fuerte dependencia energética. Adicionalmente, la segunda fuente del total de energía consumida (después del petróleo) es la biomasa, con un 24 %. Esta fuente representa un significativo 25 % de la generación eléctrica y es muy utilizada

en procesos térmicos industriales y de calefacción residencial[1]. Ambos casos se esperan que sean electrificados parcialmente o alimentados con fuentes limpias. Por otro lado, la solar y la eólica representan solo el 3.5 % de la energía total y cerca del 20 % de la electricidad, tal como muestra la Figura 1.1

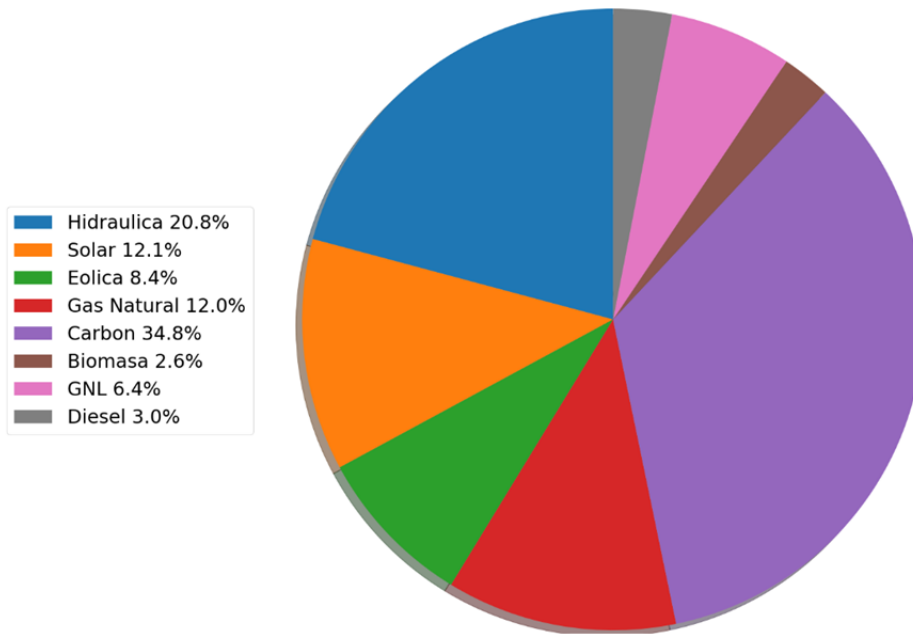


Figura 1.1: Generación Eléctrica en Chile por tipo de energía.

1.3. Objetivos

1.3.1. Objetivos Generales

Implementar un Pipeline para la predicción de la generación eléctrica mediante la inclusión de la producción de energía nuclear, utilizando modelos de aprendizaje automático.

1.3.2. Objetivos Específicos

- Recopilar y analizar la información del historial de producción eléctrica por fuente del país o región a estudiar.
- Aplicar técnicas de Machine Learning para la predicción a largo plazo de la producción eléctrica, considerando los objetivos de producción del país o región en estudio.
- Desarrollar una herramienta que genere entradas de producción eléctrica para el código de simulación ANICCA a partir de los resultados de predicción, permitiendo perturbaciones en los datos de entrada.
- Realizar predicciones utilizando fuentes de energía nuclear como una nueva fuente de generación.

1.4. Resultados Esperados

La realización del proyecto de titulación actual permitirá la formulación de una estrategia energética a corto, mediano y largo plazo, lo cual mejorará la toma de decisiones en cuanto a futuras políticas públicas y privadas en el ámbito de la generación eléctrica.

1.5. Estructura del Documento

«Esta tesis esta conformada por 6 secciones que están señaladas a continuación.»

- En el Capítulo 2 se presenta el marco teórico, el cual contiene los conceptos que serán de ayuda para tener una mejor comprensión del proyecto abordado.
- En el Capítulo 3 se presenta el estado del arte que lleva relación con el tema principal del proyecto de titulación.
- En el Capítulo 4 se presenta la metodología utilizada en el desarrollo del proyecto.
- En el Capítulo 5 se presenta la planificación del proyecto por medio de una carta Gantt.
- En el Capítulo 6 se presentan los resultados del proyecto.
- En el Capítulo 7 se presentan las discusiones, conclusiones, limitaciones y trabajos futuros propuestos del proyecto.

Capítulo 2

Marco Teórico

En este capítulo, se presenta el Marco teórico que sustenta los elementos para resolver la problemática

2.1. Introducción

En este capítulo se presenta el marco teórico relacionado con el tema del proyecto de titulación. Se presentan teorías y conceptos que permiten abordar el problema correctamente.

2.2. Machine Learning

El aprendizaje automático es un subconjunto de la inteligencia artificial que implica la creación de algoritmos que pueden aprender patrones en datos y hacer predicciones o decisiones sin ser programados explícitamente. [2]

Los algoritmos de aprendizaje automático se pueden dividir en dos categorías principales: supervisados y no supervisados. Los algoritmos de aprendizaje supervisado se entrenan en un conjunto de datos etiquetados, donde se conoce la salida deseada. El algoritmo aprende a hacer predicciones sobre nuevos datos basados en los patrones que ha aprendido del conjunto de entrenamiento. Los algoritmos de aprendizaje no supervisado, por otro lado, trabajan con un conjunto de datos no etiquetado e intentan encontrar patrones o relaciones dentro de los datos.[3]

Uno de los principales beneficios del aprendizaje automático es que puede automatizar tareas que serían laboriosas y propensas a errores para los humanos. Por ejemplo, los algoritmos de aprendizaje automático se pueden entrenar para categorizar automáticamente imágenes en diferentes clases o detectar fraudes en transacciones financieras.

Los algoritmos de aprendizaje automático se evalúan a menudo en función de su capacidad para generalizar a nuevos datos no vistos. Esto significa que los algoritmos deberían ser capaces de hacer predicciones precisas sobre datos que no han visto durante el proceso de entrenamiento.

En general, el aprendizaje automático es un campo en rápida evolución que tiene el potencial de revolucionar muchas industrias y ya ha llevado a avances significativos en áreas como la salud, las finanzas y el transporte.[4]

2.3. Regresión Lineal

La regresión lineal es un algoritmo de aprendizaje automático supervisado que se utiliza para modelar la relación entre una variable dependiente y una o más variables independientes. Supone que la relación entre las variables es lineal y tiene como objetivo ajustar una línea a los datos que mejor captura esta relación.

En la regresión lineal simple, solo hay una variable independiente, mientras que en la regresión lineal múltiple hay dos o más variables independientes. La línea de mejor ajuste se encuentra minimizando la suma de las diferencias al cuadrado entre los valores observados y predichos de la variable dependiente.

La regresión lineal se puede utilizar tanto para tareas de regresión (predicción de un resultado continuo) como de clasificación (predicción de un resultado categórico), aunque se utiliza con mayor frecuencia para la regresión.[5]

2.4. Series de Tiempo

Una serie tiempo es una secuencia de observaciones, medidos en determinados momentos del tiempo, ordenados cronológicamente y, espaciados entre sí de manera uniforme[6]. Los datos usualmente son dependientes entre sí y pueden incluir diferentes componentes, como:

- **Tendencia:** Es la dirección general en la que se mueve la serie de tiempo, es decir, si aumenta o disminuye a lo largo del tiempo.
- **Estacionalidad:** Es el patrón que se repite cada cierto período de tiempo, como por ejemplo la estacionalidad de las ventas durante el año.
- **Ciclo:** Es un patrón que se repite a plazos más largos que la estacionalidad, como los ciclos económicos.
- **Ruido:** Son los valores aleatorios o irregulares que no siguen un patrón evidente y pueden ser causados por factores externos o aleatorios.
- **Tendencia a corto plazo:** Son los cambios a corto plazo en la serie de tiempo que no siguen un patrón evidente y pueden ser causados por eventos puntuales o factores externos.

En conclusión, las series de tiempo son una secuencia de observaciones medidas en determinados momentos del tiempo y ordenadas cronológicamente. Estas observaciones pueden tener diferentes componentes, como tendencia, estacionalidad, ciclo, ruido y tendencia a corto plazo, y pueden ser utilizadas para hacer predicciones y tomar decisiones basadas en patrones y tendencias en los datos. Es importante tener en cuenta que cada serie de tiempo es única y puede tener diferentes características y componentes que deben ser considerados al analizarla.

2.5. Bases de Datos

Las bases de datos son una forma de almacenar y gestionar grandes cantidades de información de manera organizada y estructurada. Hay muchas razones por las que es importante guardar la información en bases de datos:

- 1 Permite un acceso rápido y sencillo a la información: Las bases de datos permiten hacer consultas a la información de manera rápida y sencilla, lo que hace que sea fácil encontrar y obtener los datos que se necesitan.
- 2 Ayuda a mantener la información actualizada y precisa: Las bases de datos permiten hacer seguimiento y controlar los cambios que se hacen a la información, lo que asegura que la información siempre sea precisa y actualizada.
- 3 Facilita la toma de decisiones: Las bases de datos permiten analizar la información de manera sencilla y rápida, lo que puede ayudar a tomar decisiones más informadas y efectivas.
- 4 Ayuda a ahorrar tiempo y recursos: Al tener la información organizada y estructurada en una base de datos, es más fácil y rápido acceder a ella, lo que puede ahorrar tiempo y recursos.

- 5 Mejora la seguridad de la información: Las bases de datos suelen incluir medidas de seguridad para proteger la información y evitar su pérdida o acceso no autorizado.
- 6 Si los datos son los esperados y dentro de los valores, es necesario guardar el modelo para futuras pruebas, y evitar estar entrenando el modelo cada vez que se inicia el programa.

En resumen, guardar la información en bases de datos es importante porque permite un acceso rápido y sencillo a la información, ayuda a mantenerla actualizada y precisa, facilita la toma de decisiones, ahorra tiempo y recursos, y mejora la seguridad de la información.

2.6. Conclusiones

La selección de las series de tiempo como también la forma en que estas son guardadas influyen enormemente en la eficiencia y efectividad del modelo de predicción, puesto que estas son las columnas vertebrales de cualquier entrenamiento y predicción realizadas con los mismos. Es de suma importancia priorizar este aspecto antes de efectuar cualquier otra actividad para realizar[7].

Capítulo 3

Estado del Arte

En este capítulo se explicara la ruta de trabajo como la explicación de los elementos utilizados en la creación y manipulación de los datos trabajados como tambien los modelos utilizados y la conclusión de por que estos fueron elegido.

3.1. Energía Nuclear

La energía nuclear es una fuente de energía que ha sido ampliamente utilizada en la producción de electricidad y en aplicaciones industriales y médicas. Según la Agencia Internacional de Energía Atómica (IAEA), en 2021 había 443 reactores nucleares en operación en todo el mundo, y la energía nuclear representó aproximadamente el 10 % de la generación de electricidad global en 2020 [8].

Las plantas nucleares utilizan uranio enriquecido como combustible para generar calor a través de reacciones nucleares controladas. El calor generado se utiliza para producir vapor, que a su vez mueve las turbinas y genera electricidad. Hay varios tipos de reactores nucleares, como los de agua ligera, agua pesada y de gas, que utilizan diferentes métodos para producir la reacción nuclear [8].

Además de la generación de electricidad, la energía nuclear también se utiliza en la medicina y la industria. En la medicina, se utiliza en la radioterapia para tratar el cáncer y en la producción de radioisótopos para diagnóstico y tratamiento médico [8]. En la industria, se utiliza en la esterilización de alimentos y productos médicos [9].

Sin embargo, la energía nuclear también presenta ciertos riesgos y desafíos. La posibilidad de accidentes nucleares es uno de los mayores riesgos, como se evidenció en los desastres nucleares de Chernobyl y Fukushima. Además, la gestión de los residuos radiactivos es un desafío importante que requiere soluciones a largo plazo [8]. Por lo tanto, es importante utilizar la energía nuclear de manera responsable y segura, y tomar medidas para minimizar los riesgos asociados con su uso.

3.2. Trabajos Relacionados

Dentro del campo de la predicción de generación y el uso de técnicas de aprendizaje automático en el ámbito de la energía nuclear, se han llevado a cabo numerosos estudios previos que han sido fundamentales para el desarrollo de investigaciones en este ámbito. En consecuencia, se ha procedido a la recopilación de diversas referencias y usos anteriores que se han utilizado como punto de partida y como referencia en la realización del proyecto de titulación que se presenta en este momento.

Es importante destacar que estos estudios previos han permitido explorar y comprender de manera más profunda los mecanismos y procesos asociados a la predicción de generación y el uso de técnicas de aprendizaje automático en la energía nuclear. De este modo, los resultados obtenidos a través de las investigaciones previas han sido esenciales para la elaboración del proyecto de titulación, proporcionando una base sólida y coherente para el diseño de la metodología y la estrategia que se ha utilizado en la investigación actual.

Application of machine learning to nuclear energy: A review[10]

Resumen: Como forma de energía limpia, la energía nuclear tiene ventajas únicas en comparación con otras fuentes de energía en la era actual, en la que se están promoviendo ampliamente políticas de bajo carbono. El crecimiento exponencial de la tecnología de inteligencia artificial (IA) en las últimas décadas ha dado lugar a nuevas oportunidades y desafíos en términos de mejorar la seguridad y la economía de los reactores nucleares.

Nuclear data evaluation augmented by machine learning[11]

Resumen: La precisión de las herramientas de modelado y simulación de neutrones depende en gran medida de la calidad de los datos nucleares. Las bibliotecas de datos se generan mediante la combinación de códigos de modelos basados en la física y datos experimentales por parte de los evaluadores. En muchos casos, los datos experimentales no están disponibles, no se informan rigurosamente o son discordantes. En tales casos, los evaluadores necesitan hacer un juicio experto exponiendo los datos generados a sesgos humanos y grandes incertidumbres.

Machine learning based renewable energy generation and energy consumption forecasting[12]

Resumen: La generación de energía renovable es una necesidad urgente para mitigar los desafíos de los combustibles fósiles, los impactos ambientales, las grandes pérdidas en la transmisión y distribución, y los patrones de consumo variables. La energía renovable, como la generación de energía solar fotovoltaica o eólica, es la generación más impredecible y depende de las limitaciones climáticas. Una predicción prometedora de la generación es una solución adecuada para este problema. Se propone un algoritmo de red neuronal convencional basado en aprendizaje automático para predecir la generación y calcular los errores en la predicción de la generación. La energía generada se suministra a los consumidores, pero los consumidores consumen electricidad por su elección y se requiere un modelo de gestión de energía en el extremo del consumidor para suministrar la generación predicha disponible y desarrollar una microrred autosostenible.

Adaptive Power Transformer Lifetime Predictions Through Machine Learning and Uncertainty Modeling in Nuclear Power Plants[12]

Resumen: El tiempo de vida útil restante (RUL) del papel de aislamiento del transformador está determinado en gran medida por la temperatura del punto caliente del devanado (HST). Con frecuencia, la HST no se monitorea directamente y se infiere a partir de otras mediciones. Sin embargo, los errores de medición afectan los modelos de predicción y si no se tienen en cuenta las variables inciertas, esto puede llevar a decisiones de mantenimiento incorrectas. Además, los modelos analíticos existentes para el cálculo de la HST no siempre son precisos porque no pueden generalizar las propiedades de los transformadores que operan en diferentes contextos.

Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders[13]

Resumen: El consumo de energía en los edificios está aumentando debido al desarrollo social y la urbanización. La predicción del consumo de energía en los edificios es esencial para mejorar la eficiencia energética y el desarrollo sostenible, y así reducir los costos de energía y el impacto ambiental. Esta investigación presenta una revisión exhaustiva de las técnicas de aprendizaje automático (ML) para la predicción de series temporales de consumo de energía utilizando datos reales. Se recopilieron datos en tiempo real de una red inteligente instalada en un edificio experimental y se utilizaron para evaluar la eficacia y efectividad de las técnicas estadísticas y de aprendizaje automático.

En conclusión, el análisis y la comprensión de las referencias y usos anteriores en la predicción de generación y el uso de técnicas de aprendizaje automático en la energía nuclear, han sido cruciales para el desarrollo de la investigación actual, permitiendo la elaboración de una estrategia efectiva y coherente que ha permitido obtener resultados significativos y relevantes para el campo de la energía nuclear.

Capítulo 4

Metodología

Dentro de toda creación de modelos para Machine Learning la parte esencial son los datos, sin ellos nada de esto sería inclusive posible, por ende, es necesario ocupar la mayor cantidad del tiempo en que los datos sean, Confiables y ajustados a la hipótesis principal.

4.1. Introducción

Dentro de este capítulo, se presenta de forma detallada cada uno de los pasos y procedimientos que se llevarán a cabo para resolver el problema presentado por el proyecto. En el se entrega la descripción de los aspectos necesarios para llevar a cabo la solución al mismo.

4.2. Recursos utilizados

Recurso	Nombre
Sistema operativo	Windows 11 Pro
Procesador	AMD Ryzen 7 5800H 3.20GHz
RAM	8 GB

Tabla 4.1: Recursos utilizados.

4.3. Ruta de trabajo

Dentro de la ruta de trabajo realizado se encuentran los siguientes pasos:

- 1 Obtención y guardado de los datos de las fuentes investigadas y seleccionadas.
- 2 Confirmación del formato a usar, en este caso debe corresponder a un formato previamente diseñado a lo cual cualquier otro tipo de formato debe ser transformado antes de usar el programa.
- 3 Al confirmar que los datos a utilizar están en el formato correspondiente se vuelven a guardar si es que esto es necesario y se continua con el siguiente paso.
- 4 EL modelo elegido es entrenado con los datos previamente guardados, seleccionando la cantidad de datos requerida para el entrenamiento como de validación.
- 5 Se estiman y analizan los resultados del entrenamiento del modelo utilizando los datos de validación y revisando si el error obtenido cumple las expectativas si estos valores se encuentran en dentro de los esperados se pasa al siguiente paso, si esto no es así, el modelo es vuelto a ser entrenado cambiando los valores de los hyperparametros y inclusive la cantidad de los datos de entrenamiento como de validación.
- 6 Si los datos son los esperados y dentro del los valores, es necesario guardar el modelo para futuras pruebas, y evitar estar entrenando el modelo cada vez que se inicia el programa.
- 7 Se realiza la predicción de X valores donde X es la cantidad de valores requeridos, los cuales pueden ser X Dias, X Meses, X Años o cualquier cantidad de tiempo que se especifique dentro de los datos usado en el entrenamiento.
- 8 Los datos generados por la predicción se guardan para su futuro analisis por parte del usuario en el formato que este deseé.

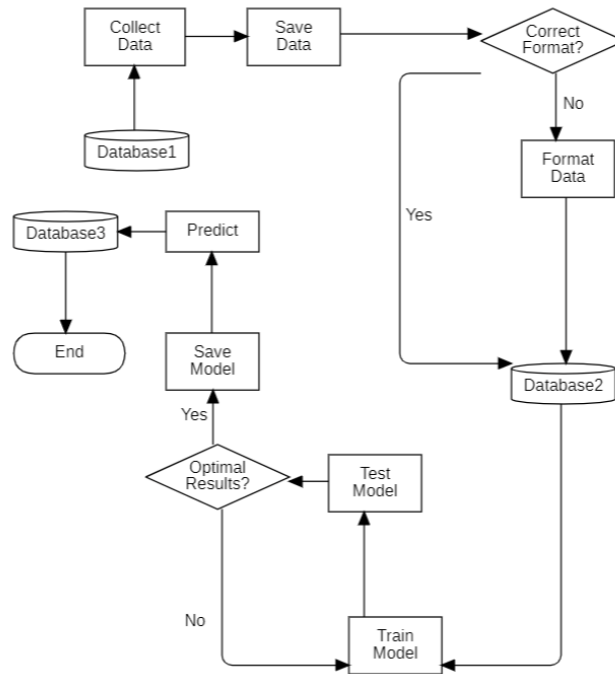


Figura 4.1: Diagrama de flujo.

4.4. Origen de los datos

La fuente de los datos como se indico anteriormente son la CNE y ELIA, siendo estas de Chile y Belgica respectivamente, para una mayor fiabilidad de los resultados y del analisis es importante recalcar que estos son los datos oficiales con lo que cuenta cada país y son de libre obtención. Estos datos se embarcan en las politicas de transparencia de los sistemas publicos tantos Chilenos como Belgas.

4.5. Los datos

Los datos utilizados para la prueba de la predicción y que son usados aquí han sido recolectados de las páginas web de la CNE(Comisión Nacional de energía) y la ELIA(El sistema de transmisión de electricidad Belga), dentro del marco de la política de transparencia que disponen ambos países para los datos de la matriz eléctrica propia, ha de indicar que estos solo han dispuesto la información desde cierto años siendo la fecha histórica belga desde el 2015 y la chilena desde el 2014, en cuanto a los datos chiles, si se presenta otra cantidad de datos históricos llegando incluso a los sistemas anteriores (SIC y SING), enmarcando los años 1998 hasta 2010, luego de esto se presento la interconexión de estos sistema dando como resultado el sistema SEN que son los datos utilizados actualmente. En base a ser utilizado en modelos de Machine Learning es importante indicar el tipo de dato, y otras variables que se analizarán en la siguiente sección.

4.6. Preprocesamiento de los datos

«Es de suma importancia para cualquier implementación de modelo de Machine Learning que la información que se usa para el entrenamiento y validación posean el formato a utilizar planteado en la recopilación de los datos y hipótesis, debido a que el procesamiento de estos es la parte más importante en cualquier modelo, por esto es siempre imperativo ocupar la mayor cantidad del tiempo en esta labor, para lograr tener resultados acorde a la investigación, aunque existes diversas tecnicas de manejo de datos unas de las más utilizadas es utilizando la librería Pandas de Python[14][15], la cual permite la modificación,eliminación o selección de los datos que se quiera, en el TFG como los datos venían de diferentes formas se utilizaron los metodos de edición que se explicaran en esta sección.»

4.6.1. Formato de los datos

«Para la generación de los dataset para el entrenamientos de los datos se utilizaron los datasets entregados por la CNE (Comisión Nacional de energía) para Chile y los datasets entregado por la ELIA (El operador del sistema de transmisión eléctrica Belga), Estos datos diferían en el formato necesario, por lo cual se realizó un Script para generar un nuevo dataset que cumpla con los requisitos,["Tipo de combustible "," Date "," Generación "], para convertir estos datos en una Serie de tiempo compatible con los modelos de ML utilizados. Para mayor adecuación y generar una menor presión en el cómputo de los datos, usaremos los datos por Día, utilizando la librería Pandas para transformar los datos, Cabe recalcar que los datos en si utilizados eran guardados en formato de 15 minutos (Datos Belgas) y en formato de Día (Datos Chilenos)»

Fecha	Tipo de combustible	Generación
01-01-2015	Solar	64.5
02-01-2015	Solar	44.1
03-01-2015	Solar	54.3
04-01-2015	Solar	74.5
05-01-2015	Solar	64.1
06-01-2015	Solar	34.3

Tabla 4.2: Ejemplo del formato usado.

4.7. Análisis de los datos

4.7.1. Skewness y Kurtosis

- **Skewness:** es una medida de la simetría de una distribución de datos. Se refiere a la forma en que los datos se distribuyen alrededor de la media. Si los datos tienen una distribución simétrica alrededor de la media, se dice que tienen un Skewness cero. Si los datos tienen una distribución sesgada hacia la derecha, se dice que tienen un Skewness positivo. Si los datos tienen una distribución sesgada hacia la izquierda, se dice que tienen un Skewness negativo.

$$Skew(x) = \sum \left[\frac{(X - \mu)^3}{\sigma} \right] \quad (4.1)$$

- **Kurtosis:** es una medida de la concentración de datos alrededor de la media. Se refiere a la forma en que los datos se concentran alrededor de la media y cómo se extienden hacia los valores más altos y más bajos. Si los datos tienen una distribución "apuntada", con valores muy altos y muy bajos, se dice que tienen un Kurtosis alto. Si los datos tienen una distribución "plana", con valores más dispersos alrededor de la media, se dice que tienen un Kurtosis bajo.

$$Kurtosis(x) = \sum \left[\frac{(X - \mu)^4}{\sigma} \right] \quad (4.2)$$

La Skewness y el Kurtosis son importantes porque pueden afectar la forma en que se analizan y procesan los datos y cómo se hacen predicciones basadas en ellos. Por ejemplo, si los datos tienen un Skewness muy alto o un Kurtosis muy alto, es posible que sea necesario aplicar transformaciones especiales para poder trabajar con ellos de manera efectiva. Además, la Skewness y el Kurtosis pueden afectar la validez de ciertas suposiciones estadísticas, por lo que es importante tener en cuenta estos factores al analizar los datos.

Energía	Skewness	Kurtosis
Coal	2.85	6.67
Gas	0.48	-0.68
Nuclear	-0.56	-0.28
Wind	1.31	1.17

Tabla 4.3: Skewness y Kurtosis para las principales energía Belgas.

Energía	Skewness	Kurtosis
Coal	-1.89	3.99
Gas	1.55	2.30
Hydro	0.72	0.58
Wind	0.07	-0.91

Tabla 4.4: Skewness y Kurtosis para las principales energía Chilenas.

4.7.2. Tipos de datos

Unos de los requisitos para la gran mayoría de los modelos predictivos es que los datos a utilizar sean datos estacionarios, es decir, datos que muestren una tendencia y no presenten una gran varianza entre sus propios datos que podrían generar ruido dentro del modelo de Machine learning, existen varias formas de arreglar o transformar estos datos, como también es importante reconocerlos debidos a los problemas que podrían generar a futuro dentro del modelo predictivo a realizar.

El valor de p es una medida estadística comúnmente utilizada que proporciona información sobre la significancia de un resultado. En la prueba de hipótesis, el valor de p es la probabilidad

de observar una estadística de prueba tan extrema o más extrema que la observada, asumiendo que la hipótesis nula es verdadera. Cuanto menor sea el valor de p , más evidencia hay en contra de la hipótesis nula.

Un umbral comúnmente utilizado para determinar la significancia estadística es un valor de p de 0,05, lo que significa que hay un 5% de probabilidad de observar un resultado tan extremo o más extremo que el observado por casualidad, asumiendo que la hipótesis nula es verdadera. Si el valor de p es menor que 0,05, el resultado se considera estadísticamente significativo y se rechaza la hipótesis nula.

En resumen, el valor de p proporciona una medida de evidencia en contra de la hipótesis nula y es una herramienta comúnmente utilizada para tomar decisiones e inferencias estadísticas en la investigación científica y el análisis de datos.[16]

Energía	p-Value
Carbón	9.271E-17
Gas	1.71E-20
Hidráulica	2.839E-20
Eólica	4.475E-17

Tabla 4.5: p-Value para las principales energía Chilenas.

Energía	p-Value
Carbón	0.006
Gas	2.820E-06
Nuclear	0.004
Eólica	0.0001

Tabla 4.6: p-Value para las principales energía Belgas.

4.8. Transformación a Datos Estacionarios

Una parte fundamental de las predicciones es la realización de estas mismas sobre datos estacionarios, muchos de estos datos que se obtienen tanto en los datos de mercados y producción han de ser en su mayoría datos No-Estacionarios, esto causado por la gran varianza que estos datos pueden mostrar, debido a esto los modelos de predicción tanto de regresión lineal como secuenciales pueden llegar a tener problemas con los resultados finales, por la misma razón el Modelo LSTM realiza una transformación antes de realizar el entrenamiento, una solución a esto es la transformación de los datos a partir de diferentes formas. Un ejemplo de lo ultimo mencionado es la transformación logarísmica que transforma los datos a valores con mucha menor variación, generado que al modelo le sea mucho más sencillo el entrenamiento y mostrar unos resultados menos contaminados y con menos ruido causado por la gran varianza que estos presentarían.

4.8.1. Transformación logarítmica

La transformación logarítmica, un método ampliamente utilizado para abordar datos sesgados, es una de las transformaciones más populares utilizadas en la investigación biomédica

y psicosocial. Debido a su facilidad de uso y popularidad, la transformación logarítmica está incluida en la mayoría de los paquetes de software estadístico importantes, incluyendo SAS, Splus y SPSS.[17].

4.9. Tendencias

Dentro de las energías aca analizadas se puede mostrar que muchas de estas , y mas especificamente las energías Renovables presentan grandes tendencias estacionarias, es decir, presentan un incremento o decremento en función de la estación del año en que se encuentren operativa, algunas de éstas son las que se analizan mas a profundidad a continuacion

4.9.1. Hidráulica

Como se puede obsevar en la figura 4.2 la producción de energía a traves de centrales Hidráulica se ve disminuida en los primeros meses del año entre Enero y Mayo, en hasta un 26 %. Y aumentando en gran medida en los meses de Octubre y Diciembre llegando hasta un pico de 42 %, esto confirma que estas centrales presentan una severa estacionalidad que puede afectar en la producción general de energía.

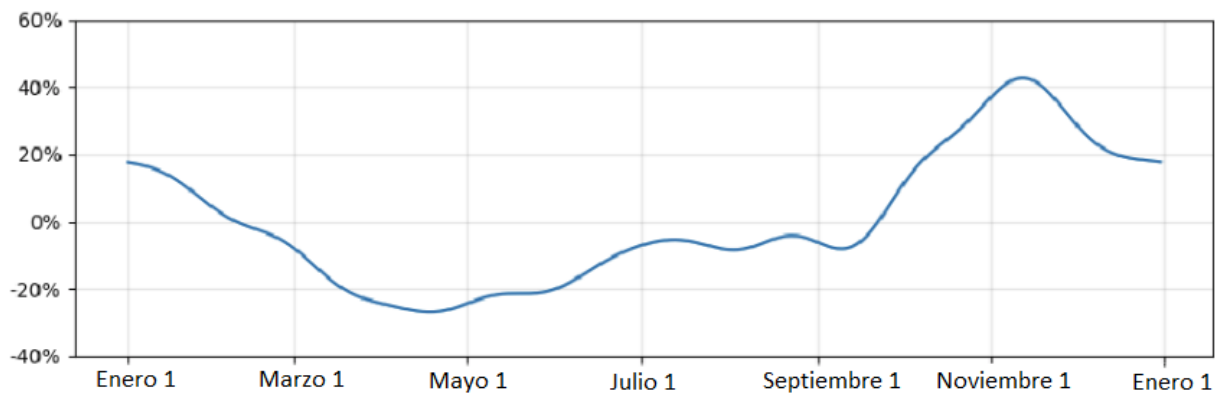


Figura 4.2: Tendencia producción Hidráulica.

4.9.2. Eólica

Se observa en la figura 4.3 que la producción de energía a través de centrales eólicas presenta una marcada disminución en los primeros meses del año, específicamente entre febrero y agosto, alcanzando un máximo de reducción del 14 %. Por otro lado, se puede constatar un incremento significativo en la producción de energía eólica durante los meses de septiembre a enero, alcanzando un pico del 15 %. Este patrón de producción de energía eólica confirma la presencia de una severa estacionalidad que puede incidir negativamente en la producción general de energía.

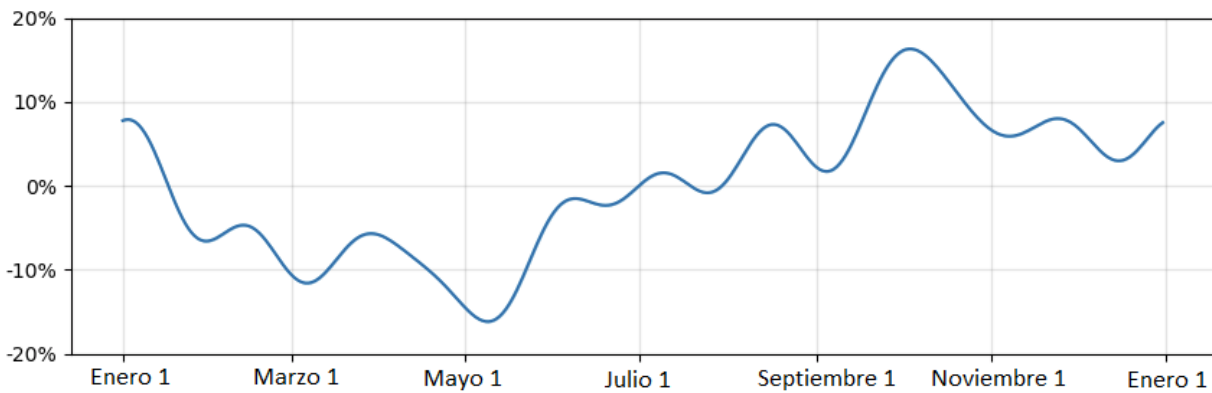


Figura 4.3: Tendencia producción Eólica.

4.9.3. Solar

Como se puede observar en la figura 4.4 la producción de energía a través de centrales Solares se ve disminuida en los primeros meses del año entre Marzo y Agosto hasta un 22 %. Y aumentando en gran medida en los meses de Septiembre y Enero llegando hasta un pico de 19 %, esto confirma que estas centrales presentan una severa estacionalidad que puede afectar en la producción general de energía.

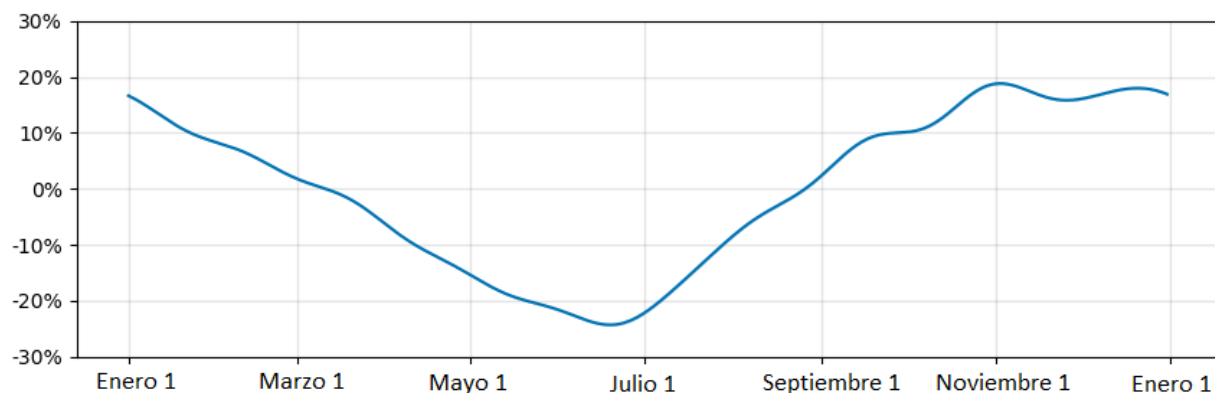


Figura 4.4: Tendencia producción Solar.

4.10. Elección de Modelo

Entre todos los modelos probados para este TFG, el con mejor resultado y el que más se acercaba a los requisitos y el objetivo del mismo era el modelo desarrollado por Facebook, Prophet el cual se explicara más a profundidad en la siguiente sección. Siendo un modelo relativamente nuevo siendo publicado su paper en Septiembre 27 del 2017.

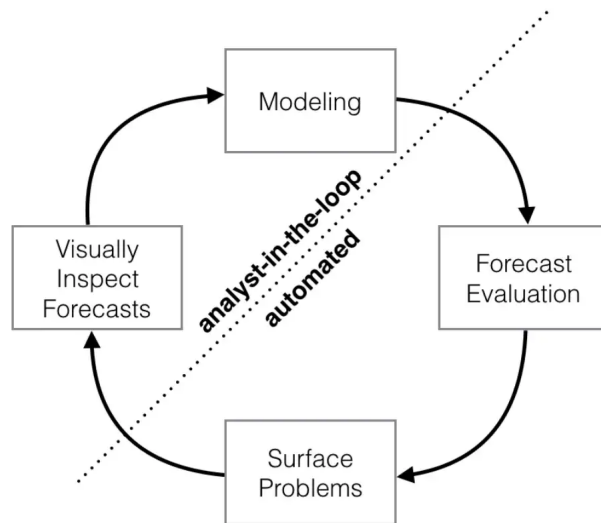


Figura 4.5: Flujo de trabajo Prophet.

4.10.1. Prophet by Facebook

PROPHET es un procedimiento para la previsión de datos de series temporales creado por el equipo de Ciencia de Datos Central de Facebook. Su objetivo es ser capaz de hacer pronósticos "a gran escala", es decir, PROPHET busca ser la herramienta de previsión automatizada, que brinde mayor facilidad de uso en la sintonización de los métodos de series temporales y permita a los analistas de cualquier campo o personas con poco o (posiblemente) ningún conocimiento previo en previsión, poder hacer pronósticos con éxito.[18]

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \quad (4.3)$$

Aquí, $g(t)$ es la función de tendencia que modela los cambios no periódicos en el valor de la serie temporal, $s(t)$ representa los cambios periódicos (por ejemplo, estacionalidad semanal y anual) y $h(t)$ representa los efectos de las vacaciones que ocurren en horarios potencialmente irregulares durante uno o más días. El término de error ε_t representa cualquier cambio idiosincrático que no está acomodado por el modelo.[19]

4.11. Archivo de Entrada

Dentro del Pipeline, es imprescindible ingresar algunas variables en un archivo de formato YAML para garantizar su correcto funcionamiento. En esta sección se detallarán cada una de ellas y su función dentro del proceso.

Es importante tener en cuenta que estas variables son esenciales para la ejecución del Pipeline, por lo que cualquier error o falta de información en su configuración puede afectar significativamente el resultado final.

Por tanto, se describirá detalladamente cada variable y su función para que su uso sea correcto y efectivo dentro del proceso. De esta forma, se garantizará que el Pipeline funcione de manera

adecuada y eficiente.

Variable	Entrada	Descripción
data	String	Ubicación del archivo que contenga la información a predecir
Create_Percentage	True False	Utilizando TRUE se crea el archivo que guarda los crecimientos de las energías mientras que al utilizar FALSE , el pipeline utiliza un archivo previamente creado
freq	M=Month D=Day 1H= 1Hour 15m=15 minutes	Indica la frecuencia que se va a utilizar para la creación de la predicción
date	String	Nombre de la columna que contenga la fechas de la base de datos
generation	String	Nombre de la columna que contenga la generación de la base de datos
tecnology	String	Nombre de la columna que contenga el tipo de energía de la base de datos
pattern	String	Nombre del archivo que guardara los crecimientos de las energías
objective_date	Date	Fecha objetivo a la que se quiere predecir
groupby	M=Month D=Day 1H= 1Hour 15m=15 minutes	Agrupar los datos de la base de datos por diferentes frecuencias
target_production	Int	Producción que se quiere lograr en la fecha objetivo
plot	True False	Mostrar crecimientos Anuales, Mensuales y Semanales
energy: Predict	Diccionario	Diccionario que contiene las energías que se quieren predecir, junto con su contribución y su crecimiento

Tabla 4.7: Variables de Entrada.

```

1  energy: # Energies to predict
2  Predict: # PUT HERE THE NAME OF THE ENERGY YOU WANT TO PREDICT
3      - Hidraulica de Pasada :
4          - 0.1 # PERCENTAGE OF THE ENERGY TO PREDICT
5          - linear
6      - Eolica :
7          - 0.15
8          - linear # NAME OF THE ENERGY : PERCENTAGE
9      - Solar Fotovoltaica :
10         - 0.15
11         - linear # NAME OF THE ENERGY : PERCENTAGE
12     - Nuclear:
13         - 0.2
14         - linear

```


Tabla 4.8: Diccionario Contribución y crecimiento Energías.

4.12. Entrenamiento del Modelo

Para lograr una captura y presentación de datos con la menor tasa de error posible, es fundamental encontrar los parámetros y métodos de validación que mejor se adapten al problema y a los datos. En este sentido, la búsqueda de hiperparámetros óptimos es de suma importancia para la optimización del modelo de aprendizaje automático.

El proceso de entrenamiento de un modelo de aprendizaje automático se lleva a cabo a través de iteraciones, donde el modelo se ajusta gradualmente a los datos de entrenamiento para mejorar su rendimiento en la predicción de nuevos datos. Durante el entrenamiento, se utilizan diferentes técnicas y algoritmos para optimizar los parámetros del modelo y mejorar su capacidad para generalizar a nuevos datos.

La validación cruzada y la búsqueda de hiperparámetros son dos técnicas comunes utilizadas para optimizar los modelos de aprendizaje automático. La validación cruzada se utiliza para evaluar la capacidad de generalización del modelo mediante la división de los datos en conjuntos de entrenamiento y validación. Por otro lado, la búsqueda de hiperparámetros se refiere al proceso de ajustar los valores de los parámetros del modelo para obtener el mejor rendimiento posible.

En resumen, para entrenar un modelo de aprendizaje automático es necesario seguir un proceso de iteración constante para ajustar los parámetros y mejorar la capacidad de generalización del modelo. La validación cruzada y la búsqueda de hiperparámetros son técnicas esenciales para obtener los mejores resultados posibles en la predicción de nuevos datos.

4.12.1. HiperParámetros

En el contexto del aprendizaje automático, los hiperparámetros son valores que se establecen antes del entrenamiento del modelo y que influyen en su rendimiento y capacidad de generalización. A diferencia de los parámetros del modelo, que se aprenden durante el entrenamiento, los hiperparámetros deben establecerse de antemano y ajustarse mediante ensayo y error.

Algunos ejemplos comunes de hiperparámetros son la tasa de aprendizaje, el número de capas ocultas en una red neuronal, el tamaño del lote y el número de árboles en un modelo de bosque aleatorio. Estos valores pueden variar significativamente dependiendo del problema y del conjunto de datos.

Para encontrar los hiperparámetros óptimos, se utiliza la validación cruzada, que implica dividir el conjunto de datos en varias partes y realizar múltiples entrenamientos y validaciones del modelo en diferentes subconjuntos de datos. Este proceso se repite para cada combinación de hiperparámetros hasta encontrar la configuración que mejor se adapte al problema y minimice el error de predicción[20].

En resumen, la elección adecuada de los hiperparámetros es esencial para el éxito de un modelo de aprendizaje automático. La validación cruzada es una técnica valiosa para encontrar

los hiperparámetros óptimos mediante la evaluación exhaustiva de múltiples combinaciones de valores.

4.12.1.1. Hiperparámetros para Prophet

En Prophet de Facebook, algunos de los hiperparámetros que se pueden ajustar son:

Parámetro	Selección	Descripción
growth	Logistic, Linear	Especifica la tasa de crecimiento de la tendencia (por defecto es "linear", pero también puede ser "logistic")
changepoints	NULL	Determina los puntos de cambio en la tendencia, por defecto se seleccionan automáticamente pero también se pueden establecer manualmente.
n_changepoints	NULL	El número de puntos de cambio a considerar, si no se especifican manualmente.
changepoint_range	0.8	La fracción del rango de los datos a considerar para los puntos de cambio, por defecto es 0.8.

Tabla 4.9: Hiperparámetros Prophet.

Después de realizar una función de GridSearch los mejores parámetros fueron los que se indican en la columna de Selección de la tabla 4.8. El modelo entrenado en base a estos parámetros sirve para encontrar las variaciones diarias, mensuales que se utilizarán y explicarán luego.

4.13. Módulo de patrones

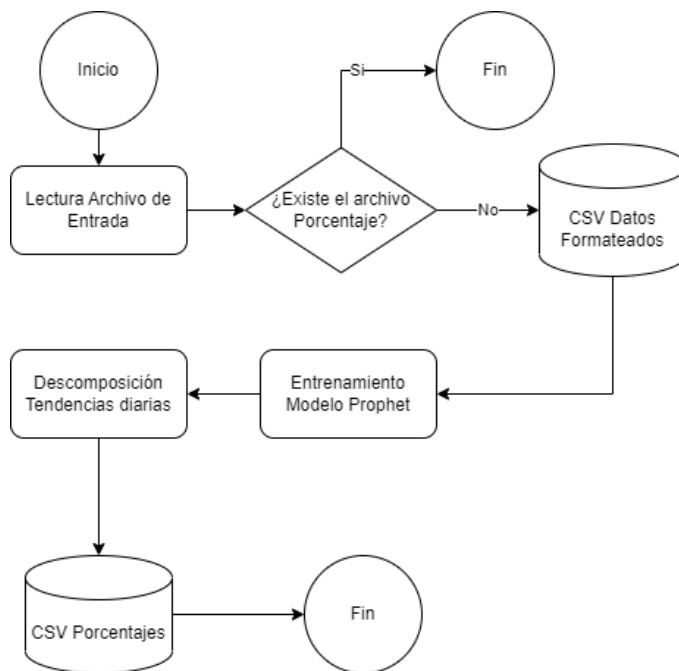


Figura 4.6: Diagrama de flujo Módulo de Patrones.

Al ejecutar el módulo de patrones desarrollado, se seguirá el flujo de ejecución descrito en la figura 4.7. Este flujo comienza con la lectura del archivo de entrada, con el fin de verificar si

el archivo que contiene los porcentajes ya ha sido creado, para evitar una sobrescritura de los datos y evitar la pérdida de información. Una vez cargados los datos, se inicia el entrenamiento del modelo de aprendizaje automático de Prophet, que se encarga de proporcionar los datos de tendencias descompuestos en las energías que se desean guardar. Finalmente, estos datos se pasan a un archivo CSV para su uso posterior.

month	day	Hidráulica de Pasada_per	Eólica_per	Solar Fotovoltaica_per
1	1	0.203	0.147	0.241
1	2	0.196	0.151	0.246
1	3	0.190	0.152	0.244
1	4	0.184	0.151	0.240

Tabla 4.10: Porcentajes Creados para Hidráulica,Eólica,Solar.

En el archivo se especifica la variación de las energías por cada día del mes, lo que permite conocer la estacionalidad que poseen, tal como se demostró en las secciones anteriores.

4.13.1. Obtención de variaciones diarias

```

1  def Decomposition(m,name,energy): #DECOMPOSITION OF SEASONALITY
2      # name = name of the seasonality component you want (yearly,monthly,etc...)
3      # m = the model object
4      # will return three arrays: days, months, and the seasonality component
5      start = pd.to_datetime('2019-01-01 0000')
6      period = m.seasonalities[name]['period']
7      end = start + pd.Timedelta(days=period)
8      plot_points = 366
9      days = pd.to_datetime(np.linspace(start.value, end.value, plot_points))
10     df_y = seasonality_plot_df(m, days)
11     seas = m.predict_seasonal_components(df_y)
12     save_dict = {}
13     save_dict["per"] = seas[name].values[:-1]
14     save_dict['day'] = df_y['ds'].dt.day[:-1]
15     save_dict['month'] = df_y['ds'].dt.month[:-1]
16     df = pd.DataFrame(save_dict)
17     return(save_dict['day'],save_dict['month'],save_dict["per"])

```

Tabla 4.11: Código Prophet Obtención variaciones Diarias Y Mensuales

La función `Decomposition` es una función que se utiliza para descomponer la estacionalidad de un modelo de series de tiempo ajustado con la librería Prophet de Facebook. Esta función toma como entrada el modelo ajustado `m`, el nombre de la componente de estacionalidad que se desea descomponer `name`, y los datos de energía `energy`.

La función devuelve tres arrays: `days`, `months` y `seas`. El array `days` contiene los días del período de tiempo que se desea analizar, el array `months` contiene los meses del período de tiempo y el array `seas` contiene los valores de la componente de estacionalidad para cada día del período de tiempo.

La función utiliza el método `seasonality_plot_df` para crear un DataFrame `df_y` con fechas espaciadas uniformemente desde el inicio del período de tiempo hasta el final, con una cantidad especificada de puntos de muestreo. A continuación, utiliza el método `predict_seasonal_components` para predecir los valores de la componente de estacionalidad para cada fecha en el DataFrame `df_y`.

Finalmente, la función devuelve los valores de día, mes y la componente de estacionalidad seas para cada día del período de tiempo analizado. Estos valores se almacenan en un diccionario `save_dict` y luego se convierten en un `DataFrame` `df` que se devuelve como resultado, el cual es guardado en una base de datos en formato CSV para su futuro uso y análisis que se efectuara tanto por el Pipeline como por cualquier otro usuario que quiera utilizar estos datos.

4.14. Módulo de Predicción

El módulo de predicción requiere de una operación de lectura sobre los archivos de entrada para verificar si la base de datos de porcentajes ya ha sido creada previamente dentro de la carpeta de ejecución correspondiente. Además, se debe realizar la lectura de los datos originales y previamente formateados con el fin de generar la recta de los datos que corresponden a los valores entre el último punto de la base de datos y el valor objetivo dentro del archivo de entrada.

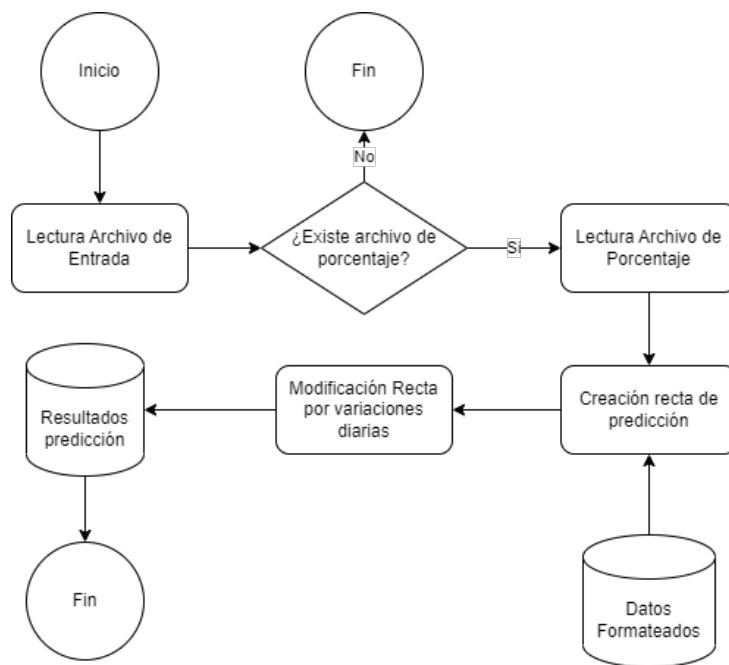


Figura 4.7: Diagrama de flujo Módulo de Predicción.

Una vez obtenidos estos datos, se utiliza el archivo de porcentaje creado previamente en el módulo de patrones para ajustar los valores a las tendencias estacionales encontradas. Finalmente, los resultados de la predicción se almacenan en una base de datos para su posterior análisis.

4.14.1. Procesamiento Matemático

Para conseguir la predicción es necesario efectuar y crear la recta que logre abarcar los objetivos, datos que se utilizan para generación de la misma.

Para lograr este objetivo, se utilizó una estrategia pensada en el objetivo final, que fue ingresado previamente en un archivo de entrada, el cual contiene información relevante y

necesaria para realizar la predicción. A partir de dicha información, se utilizó una fórmula para crear una recta que permitiera visualizar la tendencia de los datos y, de esta manera, prever su comportamiento futuro.

En particular, se utilizó el último dato guardado en la base de datos para ajustar la recta y, a partir de ella, realizar la predicción deseada. Para ello, se utilizaron las siguientes fórmulas que permitieron llevar a cabo esta tarea de manera precisa y confiable:

Primero, se calculó la pendiente de la recta utilizando la fórmula:

$$m = (y_2 - y_1)/(x_2 - x_1) \quad (4.4)$$

Donde x_2 e y_2 son las coordenadas de la fecha final y de la generación objetivo multiplicada por la contribución esperada, x_1 y y_1 son coordenadas de la ultima fecha y la generación de la misma encontrada en la base de datos.

A continuación, se calculó el término independiente de la recta utilizando la fórmula:

$$b = y_1 - m * x_1 \quad (4.5)$$

Finalmente, se utilizó la ecuación de la recta para predecir valores futuros:

$$y = mx + b \quad (4.6)$$

Donde y es el valor a predecir y x es el valor correspondiente en el eje horizontal de la recta ajustada (Fecha) y b es el término independiente de la recta.

Luego se procede a realizar una iteración por el conjunto de fechas comprendido entre la fecha inicial de predicción, definida como la última fecha registrada en la base de datos, y la fecha deseada para llevar a cabo la predicción. Esta iteración implica la ejecución de un proceso de manera secuencial para cada una de las fechas en dicho rango, con el fin de obtener las predicciones correspondientes.

4.15. Módulo de generación de bases de datos zonales

Una de las mejoras que se han incorporado al proyecto es la capacidad de generar una base de datos adicional a partir de la base de datos original, con el fin de crear y separar bases de datos específicas según las zonas definidas en la base de datos original. También se ha añadido la funcionalidad de agrupar regiones o estados en macrozonas definidas por el usuario en el archivo de entrada.

Variable	Entrada	Descripción
zones	REGIÓN	Nombre columnas que contengan las zonas definidas en la base de datos
make_macrozones	True	Activar creación por macrozonas
macrozones	Diccionario	Definición de las macrozonas por el usuario

Table 4.12 continued from previous page

Variable	Entrada	Descripción
----------	---------	-------------

Tabla 4.12: Modificación de las Variables de Entrada.

```

1  macrozones:
2  Norte:
3    - Arica y Parinacota
4    - Region de Tarapaca
5    - Region de Antofagasta
6    - Region de Atacama
7  Centro:
8    - Region de Coquimbo
9    - Region de Valparaiso
10   - Region Metropolitana
11   - Region de O'Higgins
12   - Region del Maule
13   - Region de Nuble
14  Sur:
15   - Region del Bio-Bio
16   - Region de los Rios
17   - Region de los Lagos
18   - Region de la Araucania

```

Tabla 4.13: Definición diccionario macrozonas

El diccionario al que se hizo referencia anteriormente contiene la definición y clasificación de las distintas regiones del país en diversas macrozonas para facilitar su agrupación en las diferentes bases de datos que se encuentran en la carpeta de salida denominada ".output".

Un ejemplo de la utilización de este modulo se puede observar en la figura A.1 del apéndice A del Anexo.

Capítulo 5

Planificación

En este capítulo se enseñara el plan seguido junto con las actividades que se llevaron a cabo en el proyecto de título.

5.1. Plan de trabajo

En el caso de este proyecto, el plan de trabajo se elaboró teniendo en cuenta los objetivos específicos y la metodología establecida previamente.

El programa de trabajo se dividió en diferentes fases, tareas y subtareas, y se establecieron fechas límite para cada una de ellas. De esta manera, se pudo garantizar que el proyecto avanzara de manera ordenada y eficiente.

El resultado del plan de trabajo se puede observar en la tabla 5.1, donde se detallan las diferentes fases del proyecto, las tareas y subtareas correspondientes, y las fechas límite para cada una de ellas. Gracias a este plan de trabajo, fue posible mantener el proyecto en curso y cumplir con los plazos establecidos, asegurando así el éxito del proyecto en general.

5.2. Carta Gantt Proyecto de Titulación

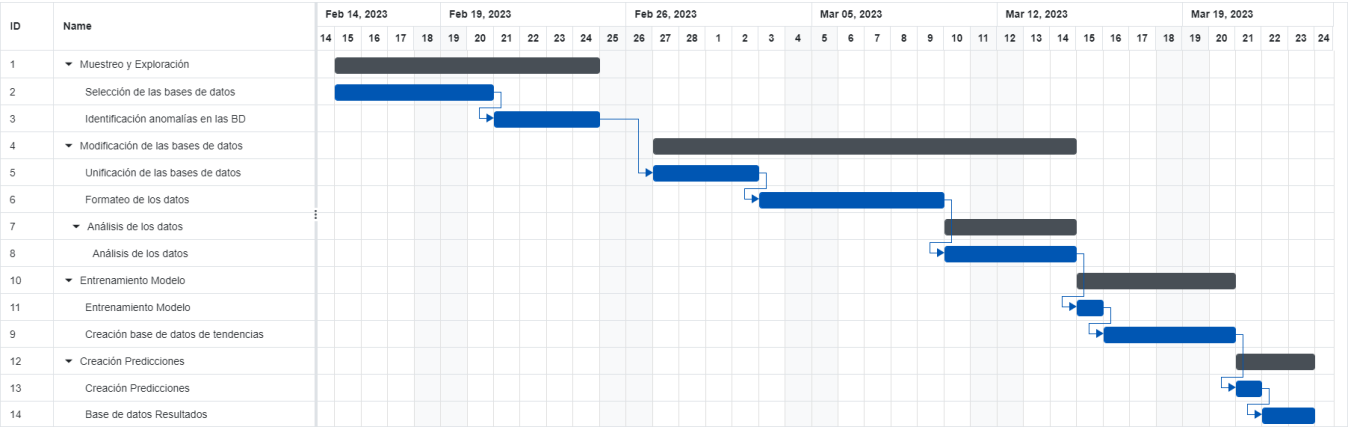


Figura 5.1: Carta Gantt

Capítulo 6

Resultados

En este capítulo, se presentan los resultados obtenidos a partir de la ejecución del Pipeline.

6.1. Introducción

En este capítulo, se presentan los resultados obtenidos a partir de la ejecución del Pipeline, el cual ha sido diseñado y ejecutado teniendo en cuenta los datos previamente mencionados en el Capítulo 4. Es importante destacar que los resultados que se presentan en este capítulo corresponden únicamente a los mejores resultados obtenidos utilizando los hiperparámetros óptimos para el conjunto de datos chilenos. Este conjunto de datos fue utilizado en todo el proyecto.

Con el fin de comprender adecuadamente los resultados obtenidos, se presentarán tanto las salidas generadas por el Pipeline como los recursos utilizados por el ordenador en el que se llevó a cabo la ejecución.

Es importante señalar que, debido a que no se ha utilizado un modelo específico, no es posible obtener un error de validación ni un error de entrenamiento. Esto se debe a que el Pipeline utiliza exclusivamente una función proporcionada por Prophet para obtener las variaciones necesarias en los datos.

6.2. Archivo de entrada utilizado

Variable	Entrada	Descripción
data	data/dataframechile.csv	Ubicación del archivo que contenga la información a predecir
Create_Percentage	True	Utilizando TRUE se crea el archivo que guarda los crecimientos de las energías mientras que al utilizar FALSE , el pipeline utiliza un archivo previamente creado
freq	D	Indica la frecuencia que se va a utilizar para la creación de la predicción
date	FECHA	Nombre de la columna que contenga la fechas de la base de datos
generation	TOTAL	Nombre de la columna que contenga la generación de la base de datos
tecnology	TIPO	Nombre de la columna que contenga el tipo de energía de la base de datos
pattern	pattern.csv	Nombre del archivo que guardara los crecimientos de las energías
objective_date	31/01/2050	Fecha objetivo a la que se quiere predecir
groupby	Day	Agrupar los datos de la base de datos por diferentes frecuencias
target_production	400000	Producción que se quiere lograr en la fecha objetivo
plot	True	Mostrar crecimientos Anuales, Mensuales y Semanales

Table 6.1 continued from previous page

Variable	Entrada	Descripción
energy: Predict	Diccionario	Diccionario que contiene las energías que se quieren predecir, junto con su contribución y su crecimiento

Tabla 6.1: Variables de Entrada.

```

1  energy: # Energies to predict
2  Predict: # PUT HERE THE NAME OF THE ENERGY YOU WANT TO PREDICT
3  - Solar :
4  - 0.4 # PERCENTAGE OF THE ENERGY TO PREDICT
5  - custom
6  - data:
7  - 2030-02-01
8  - 2040-02-02
9  - 2045-02-01
10 - values:
11 - 3000
12 - 1000
13 - -3000
14 - Eolica :
15 - 0.3
16 - otro # NAME OF THE ENERGY : PERCENTAGE
17 - data:
18 - 2030-02-01
19 - 2040-02-02
20 - values:
21 - 1000
22 - 2000
23 - Hidraulica:
24 - 0.1
25 - otro # NAME OF THE ENERGY : PERCENTAGE
26 - data:
27 - 2030-02-01
28 - 2040-02-02
29 - values:
30 - 10000
31 - 20000
32 - Carbon:
33 - 0.0
34 - linear # NAME OF THE ENERGY : PERCENTAGE
35 - Gas Natural:
36 - 0.0
37 - linear # NAME OF THE ENERGY : PERCENTAGE
38 - data:
39 - 2030-02-01
40 - 2040-02-02
41 - values:
42 - 10000
43 - 20000
44 - Nuclear:
45 - 0.2
46 - custom
47 - data:
48 - 2023-02-01
49 - 2025-02-02
50 - 2030-02-01
51 - 2034-02-02
52 - 2040-02-01
53 - 2045-02-02
54 - values:
55 - 10000
56 - 20000
57 - 30000
58 - -10000

```

59	- -20000
60	- 50000

Tabla 6.2: Diccionario Contribución y crecimiento Energías.

Dentro de la definición del archivo de entrada se incluyen las contribuciones y crecimientos correspondientes a las diferentes fuentes de energía a predecir, a saber, Carbón, Gas Natural, Nuclear, Solar, Hidráulica y Eólica. En este caso, se espera que las energías no renovables, tales como Carbón y Gas Natural, sean eliminadas y sustituidas por una mayor contribución de las fuentes de energía renovable, como Nuclear, Hidráulica, Solar y Eólica, siendo la energía solar la predominante.

En el caso de la energía nuclear, se pueden observar las diferentes centrales que se utilizarán para calcular la producción energética según lo establecido por el usuario.

6.3. Resultados

En el presente estudio, se han obtenido resultados altamente significativos mediante el uso del Pipeline, utilizando los hiperparámetros mostrados en la sección 4.13.1.1. Los datos utilizados para este análisis fueron proporcionados por la CNE después de un proceso de preprocesamiento, tal como se menciona en la sección 4.6 del presente trabajo.

Dentro del Pipeline se encuentran dos diferentes tipos de creación de predicción para la energía Nuclear, o cualquier otra, en este caso la llamaremos caso Simple y caso Complejo para diferenciarlas, definiendo al caso simple en la regresión lineal mientras que el caso complejo toma como referencia los datos de entradas que se ingresan al archivo de entrada para calcular la producción de la energía especificada con tales datos.

6.3.1. Resultados producción energía nuclear

El análisis de los datos proporcionados por el Pipeline en relación a la generación de energía nuclear esta dividida en dos casos, tanto el caso simple como el complejo el cual se explicarán mas en detalle en la siguiente sección, pero como introducción a estos se presentan las figuras 6.1 y 6.2 como el resultado de el caso simple y el complejo respectivamente.

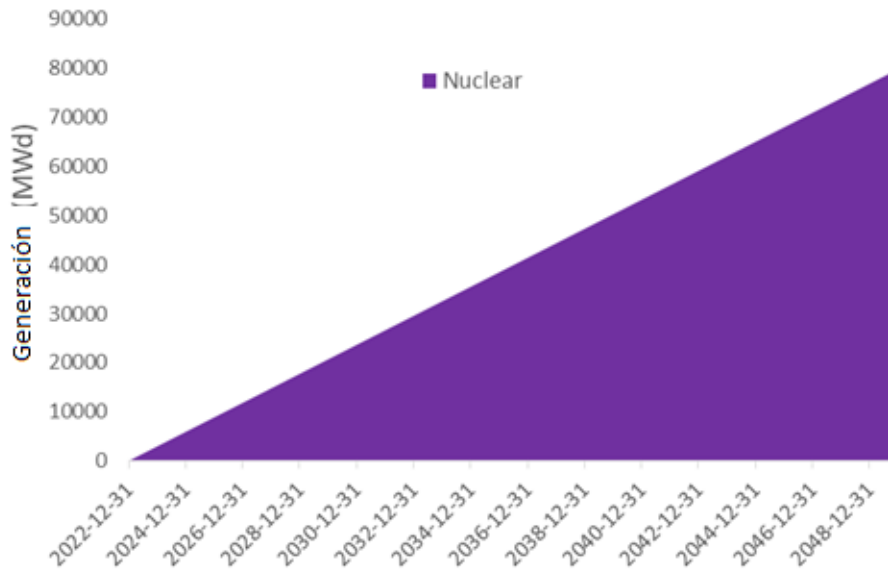


Figura 6.1: Caso Simple: Generación Energía Nuclear 2022-2050.

El caso simple, tal como se indica en la figura 6.1, presenta una tasa de crecimiento constante. Para identificar esta tendencia, se utilizó una técnica de regresión lineal, la cual permitió pronosticar la producción total que se genera a través del Pipeline.

Es importante destacar que, aunque esta solución es de naturaleza sencilla, no provee una cantidad significativa de detalles sobre la producción obtenida. En otras palabras, la solución presentada a través del uso de la regresión lineal es útil para estimar la cantidad total de producción, pero no proporciona información detallada sobre cómo se lleva a cabo dicho proceso o sobre las posibles variaciones en la producción a lo largo del tiempo.

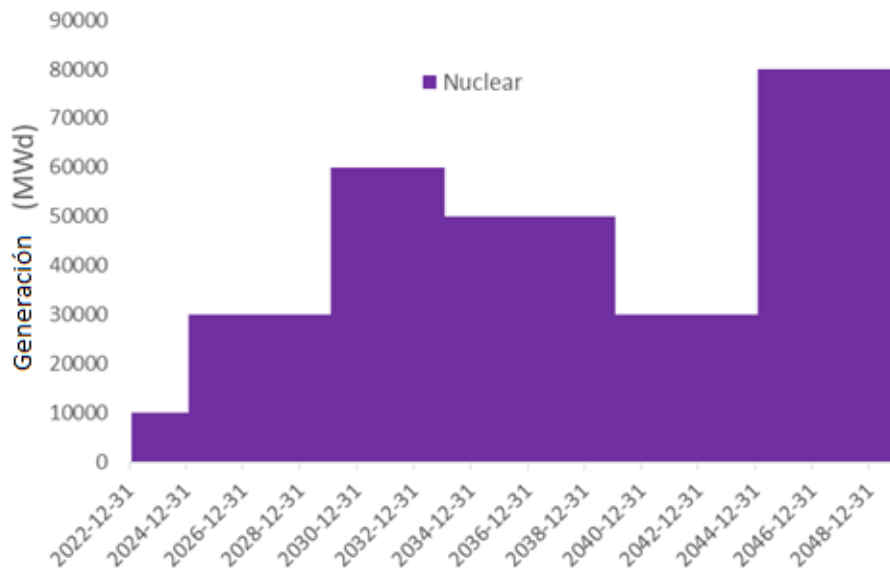


Figura 6.2: Caso Complejo: Generación Energía Nuclear 2022-2050.

En contraste con el caso simple, el caso complejo presentado en la figura 6.2 implica una serie de variaciones en la producción de energía nuclear que son impuestas por el usuario. Estas variaciones se reflejan en el archivo de entrada que se utiliza para la predicción y obtención de la producción, permitiendo al usuario tener un mayor control sobre la generación de energía.

Al permitir al usuario establecer los objetivos de producción y la forma en que se producirá la energía nuclear, se obtiene una mayor cantidad de detalles sobre el proceso en comparación con el caso simple. Además, este enfoque resuelve algunos de los problemas más recurrentes en la producción de energía nuclear, los cuales serán explicados con mayor profundidad en las secciones correspondientes a cada caso.

6.3.2. Caso Simple

El caso simple muestra una tendencia consistente con la estacionalidad de las energías que han sido estudiadas y definidas en el archivo de entrada. Este comportamiento se ha observado en los primeros años y ha logrado satisfacer la demanda energética. Sin embargo, debido al aumento en la utilización de fuentes de energía con alta estacionalidad, como la solar, eólica e hidráulica, se ha evidenciado una insuficiencia en los años posteriores.

La estacionalidad de las fuentes de energía renovable varía la producción energética a lo largo del año, lo que causa insuficiencias en momentos específicos. En este caso se utilizó la regresión lineal para obtener la producción de energía nuclear, lo que permitió un crecimiento constante hasta alcanzar la producción objetivo del 20 % de la producción total.

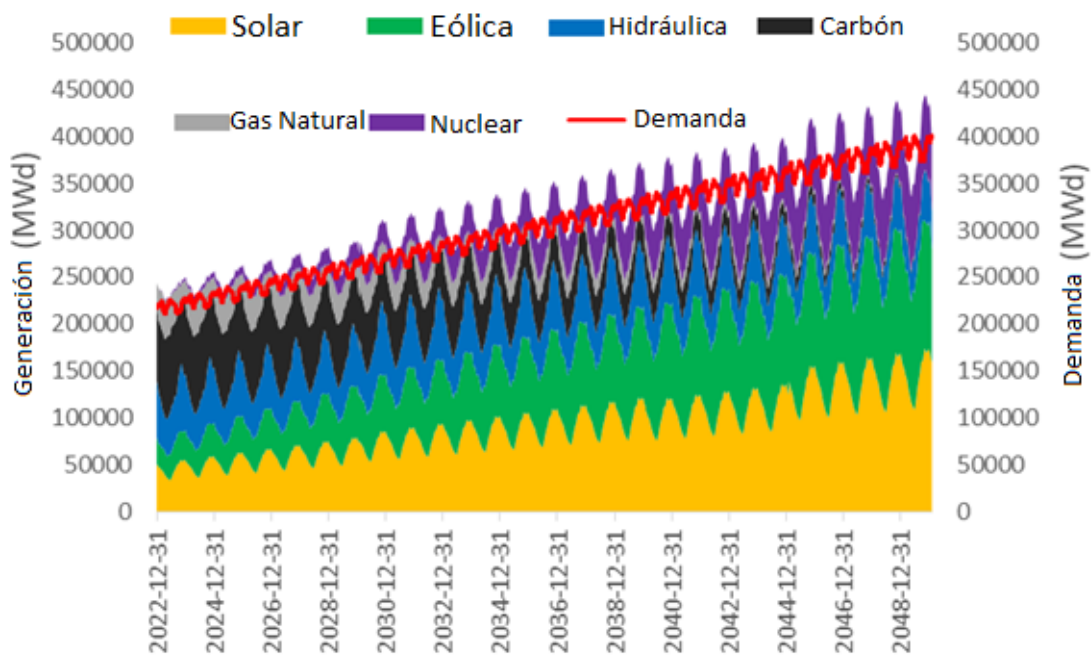


Figura 6.3: Caso Simple: Producción Total a través de los años 2022-2050.

6.3.3. Caso Complejo

El caso complejo mantienen casi las mismas características del caso simple solo cambiando en la forma de obtención de la producción de la energía nuclear, que en este caso toma en cuenta los datos ingresado por el usuario, para lograr la predicción, el cual no toma en cuenta la producción objetivo del 20 % estipulada en el caso anterior, por lo que el usuario es el que debe ingresar la cantidad de incremento o decremento que esta mantendra a lo largo de la predicción.

En el caso en cuestión, se ha logrado satisfacer la demanda energética en aquellos años en los que la producción de energía nuclear ha presentado un aumento, alcanzando así el objetivo de producción establecido. Sin embargo, cuando se ha producido una disminución

en la producción de energía nuclear en torno al año 2038, se ha observado una gran afectación en la producción energética en general, al igual que en el caso simple.

Este déficit se debe, en gran medida, a la alta estacionalidad que presentan algunas de las fuentes de energía renovable mencionadas anteriormente.

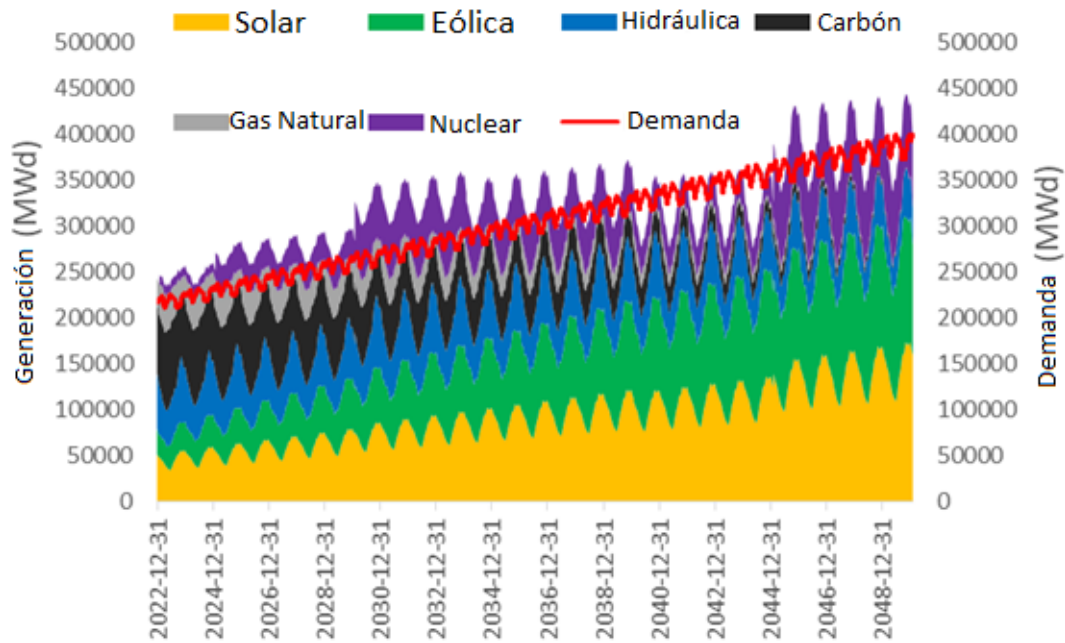


Figura 6.4: Caso Complejo: Producción Total a través de los años 2022-2050.

6.3.4. Producción eléctrica

Los resultados obtenidos se muestran en las tablas que aparecen a continuación. Estas tablas presentan información detallada sobre las diferentes variables analizadas, y muestran el impacto de los diferentes hiperparámetros utilizados en el análisis.

Fecha	Hidráulica de Pasada	Eólica	Solar Fotovoltaica	Cogeneración	Hidráulica de Embalse
2022-12-07	26209,1	39003	19850,77	566,5825	21577,87
2022-12-08	26280,84	38659,53	19780,27	581,9417	21331,3
2022-12-09	26368,03	38367,34	19726,73	598,3348	21099,69
2022-12-10	26467,59	38135,25	19693,1	615,6135	20884,25
2022-12-11	26576,03	37970,85	19681,85	633,6021	20685,58
2022-12-12	26689,65	37880,2	19694,94	652,1	20503,74
2022-12-13	26804,59	37867,67	19733,71	670,8845	20338,2
2022-12-14	26916,95	37935,85	19798,85	689,715	20187,98
2022-12-15	27022,96	38085,44	19890,34	708,337	20051,62
2022-12-16	27119,01	38315,17	20007,46	726,487	19927,33
2022-12-17	27201,84	38621,84	20148,74	743,8976	19813,02

Tabla 6.3: Primeros diez Resultados Pipeline Chile.

Fecha	Mini Hidráulica de Pasada	Biomasa	Carbón	Gas Natural	Petróleo Diesel	Nuclear	Total
2022-12-07	6611,343	4888,229	33846,02	34193,76	5704,804	0	192451,5
2022-12-08	6628,277	4948,151	34067,86	34615,77	5720,048	8,066142	192622,1
2022-12-09	6649,657	5003,241	34287,8	35012,88	5727,659	16,13228	192857,5
2022-12-10	6674,828	5052,694	34501,84	35377,85	5725,854	24,19843	193153,1
2022-12-11	6703,025	5095,844	34706,03	35704,44	5713,092	32,26457	193502,6
2022-12-12	6733,401	5132,181	34896,57	35987,6	5688,129	40,33071	193898,8
2022-12-13	6765,039	5161,363	35069,93	36223,52	5650,071	48,39685	194333,4
2022-12-14	6796,983	5183,216	35222,9	36409,81	5598,412	56,463	194797,1
2022-12-15	6828,255	5197,742	35352,69	36545,45	5533,068	64,52914	195280,4
2022-12-16	6857,882	5205,112	35456,99	36630,91	5454,389	72,59528	195773,3
2022-12-17	6884,921	5205,66	35534,03	36668,05	5363,17	80,66142	196265,8

Tabla 6.4: Primeros diez Resultados Pipeline Chile 2.

Es importante destacar que los resultados obtenidos son altamente significativos, ya que permiten comprender con mayor detalle la dinámica de las diferentes variables analizadas. Esto puede tener importantes implicaciones en diversos campos de aplicación, como en la planificación energética y en la toma de decisiones en el sector energético.

Además, los resultados obtenidos en el presente estudio pueden servir como base para futuras investigaciones, tanto en términos de la mejora del análisis realizado, como en la ampliación de la muestra de datos utilizados para el análisis. De esta manera, se puede contribuir a la generación de conocimientos más sólidos y precisos en el área de la energía.

En conclusión, los resultados obtenidos en el presente estudio son altamente significativos y demuestran la eficacia del Pipeline utilizado en el análisis de los datos proporcionados por la CNE. Los resultados obtenidos pueden tener importantes implicaciones en diversos campos de aplicación, y pueden servir como base para futuras investigaciones en el área de la energía.

Fecha	Total
2050-01-01	434484,7
2050-01-02	432681,9
2050-01-03	430599,4
2050-01-04	428268,7
2050-01-05	425726,2
2050-01-06	423012,7
2050-01-07	420172,4
2050-01-08	417251,6
2050-01-09	414298,4

Tabla 6.5: Fechas Finales Predicción.

6.3.5. Análisis de casos

Debido a la naturaleza del funcionamiento del Pipeline, no es posible comparar los resultados con los datos originales, puesto que la función principal del mismo ocupa datos fuera de los datos originales, y solo ocupa estos mismos al recabar los datos sobre las variaciones diarias o las tendencia que los datos presentan a lo largo del año,

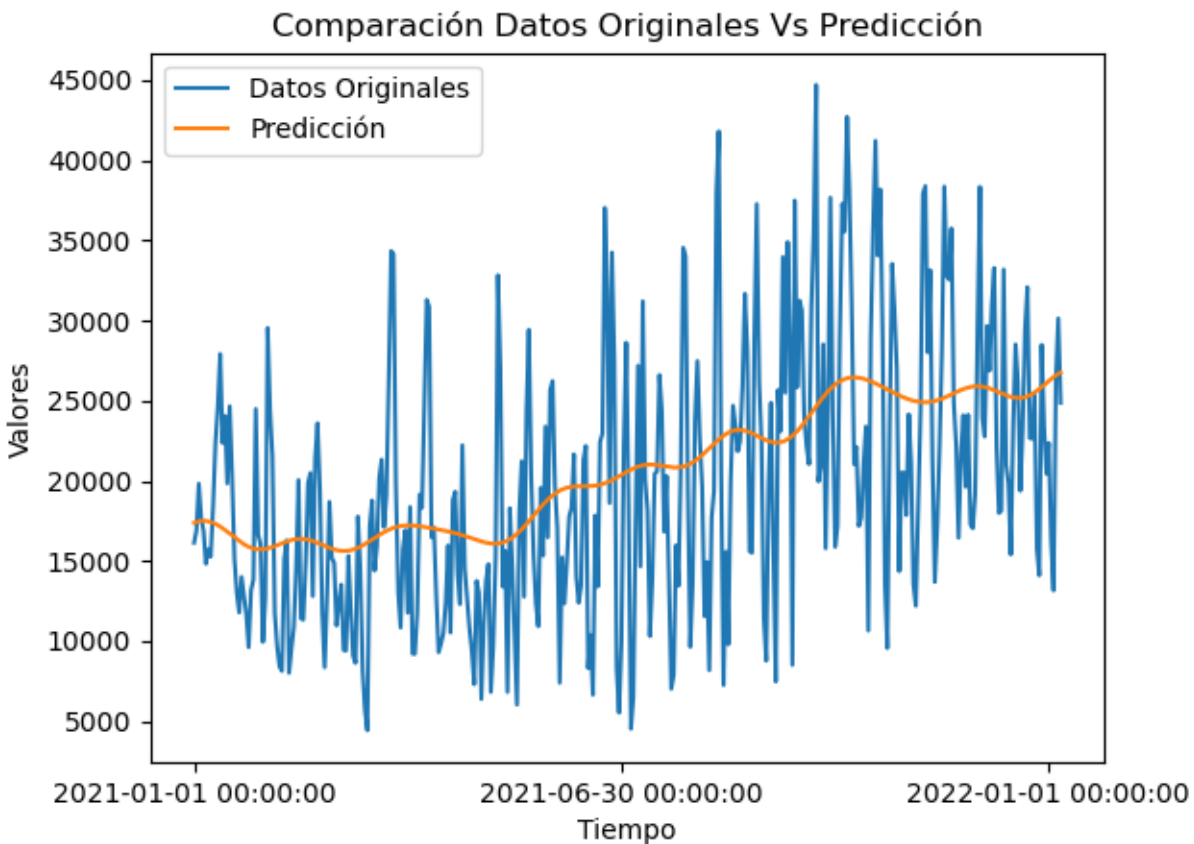


Figura 6.5: Generación Energía Eólica 2021-2022 Datos originales Vs Predicción.

Es importante tener en cuenta que, aunque la gráfica que se obtiene a través del pipeline no refleja de manera directa los datos originales, esto no significa que los datos sean manipulados o alterados. Por el contrario, el pipeline se basa en los datos originales para identificar patrones y tendencias que permitan realizar predicciones acertadas en cuanto a la evolución futura de los datos.

En definitiva, aunque resulta complicado comparar directamente los resultados obtenidos a través del pipeline con los datos originales, sí es posible utilizar la información proporcionada por ambas fuentes para obtener una visión más completa y precisa de la evolución de un determinado fenómeno. Resulta fundamental tener en cuenta la particularidad del pipeline al momento de interpretar los resultados obtenidos, para así obtener conclusiones acertadas y precisas en cuanto a la evolución de los datos.

6.4. Producto de la tesis

What Hides Chile's Electricity Mix Data: a Data Science Perspective[1]

Abstract: In the context of decarbonization and fossil fuel independence, Chile has a great task ahead. The most relevant is to become carbon neutral by 2050 by replacing fossil fuels with clean energy sources, where, besides hydraulics, solar and wind power are the chosen ones. However, the commitment to rely on intermittent sources requires a variety of studies to support the roadmap and the success of this decarbonization plan. Therefore, this study aims to understand the behavior of the energy mix and its sources by analyzing the historical production data according to the geographical context. This way, as a starting point, it was possible to discover the main behaviors and patterns, such as the electric sources' trends and seasonalities that explain the use of fossil fuels. To this end, data science tools were used for the exploration of big data, their respective analysis, and subsequent projections for planning the energy matrix's decarbonization process for 2050.

Resumen:

Chile camina hacia un futuro sostenible donde las energías renovables reemplazarán a los combustibles fósiles. Así, para el año 2050, se espera que la matriz energética sea carbono neutral, es decir, que la emisión de gases de cambio climático sea equivalente a lo que puedan ser absorbidos por la naturaleza. Sin embargo, la particularidad de las energías renovables por las que apuesta Chile, principalmente solar y eólica, son de carácter variable. Esta variabilidad podría hacer que la energía no esté disponible cuando se requiera, y aunque hay intentos de almacenamiento que podrían ayudar a resolver en parte el problema, distan de la versatilidad y robustez que caracterizan a las fuentes basadas en combustibles fósiles. Por otro parte, la geografía nacional hace que se desarrollen las fuentes energéticas de acuerdo a la zona. Así, la solar sería predominante en el norte, la hidráulica en el centro y la eólica en el sur. De acuerdo a lo expuesto, y asumiendo el significativo incremento de la demanda eléctrica a futuro, el camino a la carbono neutralidad del sistema eléctrico será un camino complejo y que amerita una sólida base científica para definir las diferentes estrategias a seguir para tener una energía amigable con el medio ambiente y económicamente competitiva.

Así, este estudio pretende crear un método para analizar el histórico de la matriz eléctrica, con el fin de obtener información pertinente al contexto nacional y que pueda utilizarse en estudios posteriores de planificación energética. De esta forma, este artículo se dividirá en tres partes. Una presentación del contexto energético actual, la descripción de la metodología utilizada, la presentación de los resultados a nivel nacional y zonal, para, finalmente, se terminará con las discusiones y conclusiones.[1]

Para saber más visitar:

<https://ieeexplore.ieee.org/document/10006089>

Capítulo 7

Discusión y conclusiones

En este capítulo se realizara la discusión sobre los resultados obtenidos.

7.1. Discusión y conclusión

En el capítulo 6 del presente estudio, se han presentado los resultados de la predicción de las diversas energías estudiadas. Estos resultados son de gran importancia ya que demuestran una solución efectiva al problema inicial planteado, utilizando el enfoque propuesto en el presente trabajo.

La metodología aplicada en este estudio ha involucrado una unión de los datos, preprocesamiento y entrenamiento del modelo utilizado para la obtención de las variaciones de las diversas energías estudiadas. Esta metodología ha permitido obtener resultados precisos y confiables, lo que demuestra la eficacia de la técnica utilizada en el presente estudio.

Al analizar los resultados presentados en las figuras 6.3 y 6.4, se puede inferir que tanto el caso simple como el complejo muestran escenarios infactibles para satisfacer la demanda total. Si bien el caso complejo demuestra una capacidad superior para abastecer en gran medida la predicción, no fue suficiente para mantener la demanda cubierta en su totalidad.

En particular, con relación a las tablas 6.3, 6.4 y 6.5 presentadas en el capítulo 6, los resultados obtenidos demuestran una variación esperada utilizando los datos obtenidos por el modelo entrenado de Prophet. Estos resultados son altamente significativos y demuestran la capacidad de la técnica utilizada para predecir las variaciones de las energías estudiadas.

Es importante destacar que la predicción de las diversas energías estudiadas es de gran importancia en diversos campos de aplicación, como en la planificación energética y en la toma de decisiones en el sector energético. Por lo tanto, los resultados obtenidos en el presente estudio pueden tener importantes implicaciones en estos campos, al proporcionar información precisa y confiable para la toma de decisiones.

En conclusión, los resultados presentados en el capítulo 6 del presente estudio demuestran una solución efectiva al problema inicial utilizando el enfoque propuesto, y demuestran la capacidad de la técnica utilizada para predecir las variaciones de las diversas energías estudiadas. Estos resultados pueden tener importantes implicaciones en diversos campos de aplicación, y son altamente significativos en términos de su impacto en la planificación energética y la toma de decisiones en el sector energético.

7.2. Trabajo futuro

En lo que se refiere a las líneas de investigación futuras, se tiene previsto añadir un nuevo modelo de predicción para el software implementado en el presente trabajo. Se busca un modelo que utilice ANN como su principal herramienta de predicción, y que permita minimizar o evitar las desventajas que se han señalado brevemente en el presente TFG. La incorporación de este modelo puede ser de gran utilidad para mejorar la capacidad predictiva del software, y para obtener resultados más precisos y fiables.

Asimismo, se tiene previsto optimizar el manejo de los archivos utilizados en el software, así como incorporar una herramienta de formateo para facilitar el uso del software por parte del usuario final. Estas mejoras pueden contribuir a hacer más eficiente y fácil el manejo del software, lo que puede resultar en un mayor número de usuarios satisfechos y una mejor recepción en el mercado.

Unas de las característica que se podría agregar en el futuro es la capacidad de aprendizaje y guardado de los modelos ya entrenados para futuros datasets. Esto permitiría disminuir enormemente el tiempo de ejecución del programa, lo que puede ser de gran utilidad en aplicaciones donde el tiempo es crítico y se requiere una respuesta rápida.

Como el estudio de los resultado demostro la estacionalidad es un factor relevante en termino del estudio de la producción electrica, por lo que conocer la fluctuación en torno a las horas del día puede dar resultados y conocimientos que pueden mejorar y apoyar todo lo expuesto en el proyecto.

En términos de la predicción de la energía nuclear, una de las futuras tareas propuestas es agregar un crecimiento en relación a las centrales que se quieran construir, junto con una generación acorde a la no variación que presentan, puesto que la generación producida es prácticamente constante en todas las centrales.

Es importante destacar que todas estas líneas de investigación futuras están dirigidas a mejorar la eficiencia y la capacidad predictiva del pipeline implementado en el presente trabajo. Estas mejoras pueden tener importantes implicaciones en diversos campos de aplicación, como en la planificación energética y la toma de decisiones en el sector energético. Por lo tanto, se espera que estas mejoras contribuyan a hacer del software una herramienta más útil y valiosa para los usuarios.

Bibliografía

- [1] Javier Soto y col. «What Hides Chile's Electricity Mix Data: a Data Science Perspective». En: *2022 IEEE International Conference on Automation/XXV Congress of the Chilean Association of Automatic Control (ICA-ACCA)*. 2022, págs. 1-6. DOI: 10.1109/ICA-ACCA56767.2022.10006089.
- [2] E. Alpaydin. «Introduction to Machine Learning». En: *The MIT Press* (2010).
- [3] K.P. Murphy. *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.
- [4] Ian Goodfellow, Yoshua Bengio y Aaron Courville. *Deep learning*. MIT press, 2016.
- [5] Trevor Hastie y col. *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2. Springer, 2009.
- [6] Jhon Villavicencio. «Introducción a series de tiempo». En: *Puerto Rico* (2010).
- [7] William Stallings. *Computer Organization and Architecture*. Prentice Hall, 2006.
- [8] *PRIS - home*. en. <https://www.iaea.org/pris/>. Accessed: 2023-3-23.
- [9] <https://www.nei.org/resources/fact-sheets/nuclear-applications>. Accessed: 2023-3-23.
- [10] Qingyu Huang y col. «A review of the application of artificial intelligence to nuclear reactors: Where we are and what's next». En: *Heliyon* 9.3 (2023), e13883. ISSN: 2405-8440. DOI: <https://doi.org/10.1016/j.heliyon.2023.e13883>. URL: <https://www.sciencedirect.com/science/article/pii/S2405844023010903>.
- [11] Pedro Vicente-Valdez, Lee Bernstein y Massimiliano Fratoni. «Nuclear data evaluation augmented by machine learning». En: *Annals of Nuclear Energy* 163 (2021), pág. 108596. ISSN: 0306-4549. DOI: <https://doi.org/10.1016/j.anucene.2021.108596>. URL: <https://www.sciencedirect.com/science/article/pii/S0306454921004722>.
- [12] Akash Talwariya y col. «Machine learning based renewable energy generation and energy consumption forecasting». En: *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects* 45.2 (2023), págs. 3266-3278. DOI: 10.1080/15567036.2023.2194257. eprint: <https://doi.org/10.1080/15567036.2023.2194257>. URL: <https://doi.org/10.1080/15567036.2023.2194257>.
- [13] Jui-Sheng Chou y Duc-Son Tran. «Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders». En: *Energy* 165 (2018), págs. 709-726. ISSN: 0360-5442. DOI: <https://doi.org/10.1016/j.energy.2018.09.144>. URL: <https://www.sciencedirect.com/science/article/pii/S0360544218319145>.
- [14] The pandas development team. *pandas-dev/pandas: Pandas*. Ver. latest. Feb. de 2020. DOI: 10.5281/zenodo.3509134. URL: <https://doi.org/10.5281/zenodo.3509134>.

- [15] Wes McKinney. «Data Structures for Statistical Computing in Python». En: *Proceedings of the 9th Python in Science Conference*. Ed. por Stéfan van der Walt y Jarrod Millman. 2010, págs. 56-61. DOI: 10.25080/Majora-92bf1922-00a.
- [16] John Neter y col. «Applied linear statistical models». En: (1996).
- [17] FENG Changyong y col. «Log-transformation and its implications for data analysis». En: *Shanghai archives of psychiatry* 26.2 (2014), pág. 105.
- [18] Christophorus Beneditto Aditya Satrio y col. «Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET». En: *Procedia Computer Science* 179 (2021). 5th International Conference on Computer Science and Computational Intelligence 2020, págs. 524-532. ISSN: 1877-0509. DOI: <https://doi.org/10.1016/j.procs.2021.01.036>. URL: <https://www.sciencedirect.com/science/article/pii/S1877050921000417>.
- [19] Sean J Taylor y Benjamin Letham. «Forecasting at scale». Sep. de 2017.
- [20] Ian Goodfellow, Yoshua Bengio y Aaron Courville. *Deep Learning*. MIT Press, 2016.

Apéndice A

Anexo

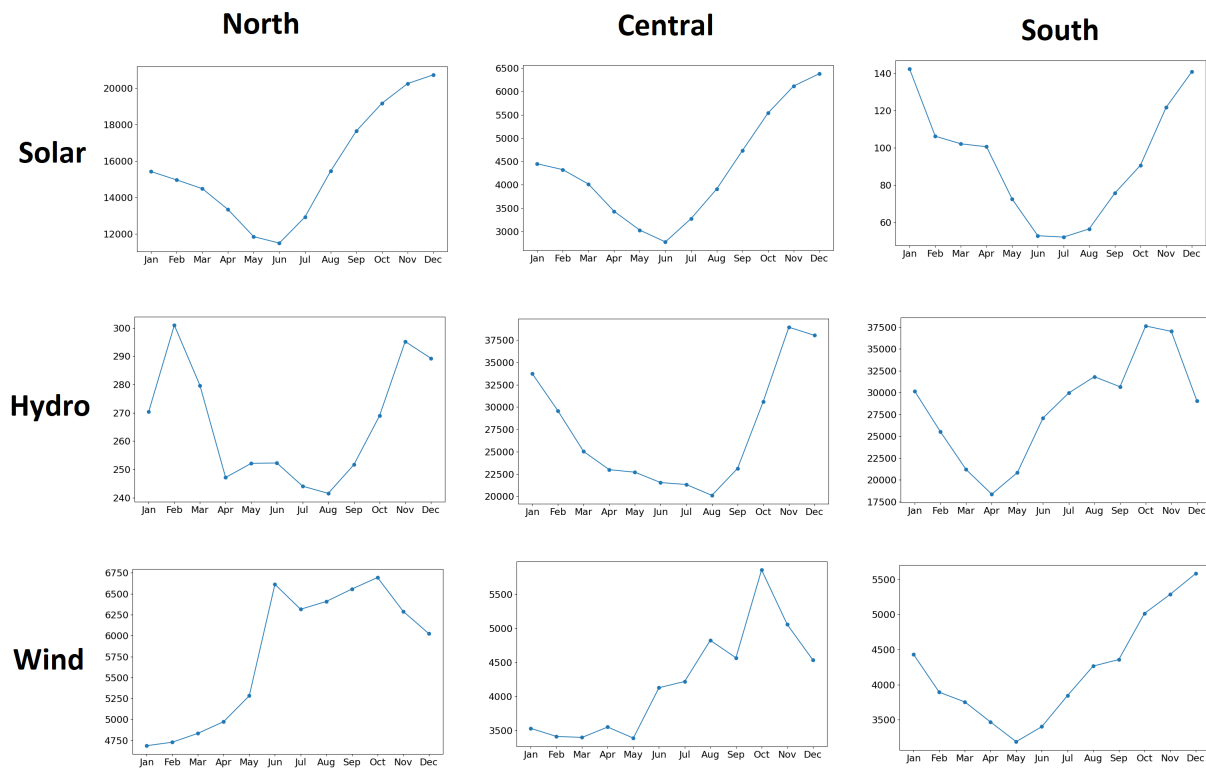


Figura A.1: Tendencia producción Solar,Hidráulica y Eólica, Zona Norte, Centro y Sur.