



**UNIVERSIDAD TÉCNICA DE COTOPAXI**  
**FACULTAD DE CIENCIAS DE LA INGENIERÍA Y APLICADAS**  
**INGENIERÍA ELÉCTRICA EN SISTEMAS ELÉCTRICOS DE**  
**POTENCIA.**

**PROYECTO DE INVESTIGACIÓN**

**ESCENARIOS DE PRODUCCIÓN DE ENERGÍA ELÉCTRICA EN UNA PEQUEÑA  
CENTRAL HIDROELÉCTRICA UTILIZANDO SERIES SINTÉTICAS.**

Proyecto de Titulación presentado previo a la obtención del Título de Ingeniero  
Eléctrico en Sistemas Eléctricos de Potencia.

**Autor:**

Carrillo Carrillo Manuel Francisco

Tapia Guerrero Jipson Danilo

**Tutor:**

Ing. Salazar Achig Roberto Ms.C

**Latacunga-Ecuador**

**Febrero 2023**



## DECLARACIÓN DE AUTORÍA

Nosotros **CARRILLO CARRILLO MANUEL FRANCISCO** y **TAPIA GUERRERO JIPSON DANILO** declaramos ser autores del presente proyecto de investigación: **“ESCENARIOS DE PRODUCCIÓN DE ENERGÍA ELÉCTRICA EN UNA PEQUEÑA CENTRAL HIDROELÉCTRICA UTILIZANDO SERIES SINTÉTICAS.”**, siendo el Ing. **SALAZAR ACHIG ROBERTO** el tutor del presente trabajo; y eximo expresamente a la Universidad Técnica de Cotopaxi y a sus representantes legales de posibles reclamos o acciones legales.

Además, certifico que las ideas, conceptos, procedimientos y resultados vertidos en el presente trabajo investigativo, son de mi exclusiva responsabilidad.

Carrillo Carrillo Manuel Francisco

CC: 060435195-7

Tapia Guerrero Jipson Danilo

CC: 050357143-2



## AVAL DEL TUTOR DE PROYECTO DE TITULACIÓN

En calidad de Tutor del Trabajo de Investigación sobre el título:

**“ESCENARIOS DE PRODUCCIÓN DE ENERGÍA ELÉCTRICA EN UNA PEQUEÑA CENTRAL HIDROELÉCTRICA UTILIZANDO SERIES SINTÉTICAS”**, de los estudiantes **CARRILLO CARRILLO MANUEL FRANCISCO** y **TAPIA GUERRERO JIPSON DANILO**, de la carrera de ingeniería en electricidad, considero que dicho informe investigativo cumple con los requerimientos metodológicos y aportes científico-técnicos suficientes para ser sometido a la evaluación del Tribunal de Validación de proyecto que el Consejo Directivo de la Facultad de Ciencias de la Ingeniería y aplicadas de la Universidad Técnica de Cotopaxi designe, para su correspondiente estudio y calificación.

Latacunga, febrero, 2023

El tutor

Ing. Salazar Achig Roberto Ms.C.

C.C. 050284761-9



## APROBACIÓN DEL TRIBUNAL DE TITULACIÓN

En calidad de Tribunal de Lectores, aprueban el presente Informe de Investigación de acuerdo a las disposiciones reglamentarias emitidas por la Universidad Técnica de Cotopaxi, y por la **FACULTAD DE CIENCIAS DE LA INGENIERÍA Y APLICADAS**; por cuanto, los postulantes: **CARRILLO CARRILLO MANUEL FRANCISCO**, con cédula de ciudadanía No. **060435195-7**, y **TAPIA GUERRERO JIPSON DANILO**, con cédula de ciudadanía No. **050357143-2**, con el título de Proyecto de titulación: **ESCENARIOS DE PRODUCCIÓN DE ENERGÍA ELÉCTRICA EN UNA PEQUEÑA CENTRAL HIDROELÉCTRICA UTILIZANDO SERIES SINTÉTICAS** han considerado las recomendaciones emitidas oportunamente y reúne los méritos suficientes para ser sometido al acto de Sustentación de Proyecto.

Por lo antes expuesto, se autoriza realizar los empastados correspondientes, según la normativa institucional.

Latacunga, febrero, 2023

Para constancia firman:

**Lector 1 (Presidente)**

Ing. Guamán Cuenca Wilian Patricio Ms.C

CC: 060357895-6

**Lector 2**

Ing. Castillo Fiallos Jessica Nataly Ms.C

CC: 060459021-6

**Lector 3**

Ing. Pesantez Palacios Gabriel Ms.C

CC: 030189388-9

## **AGRADECIMIENTOS**

En primer lugar, a agradezco a Dios y mi virgen María por brindarme salud y sabiduría para poder culminar mi carrera universitaria de la mejor manera, mi mayor agradecimiento a la Universidad Técnica de Cotopaxi por darme la oportunidad de ser parte de ella y prepararme para ser un buen profesional.

Gracias a mi familia por su apoyo incondicional, en especial a mi padre por ser mi ejemplo de lucha, esfuerzo y superación, a mi madre por demostrarme que no hay límites para luchar que frente a las adversidades.

A mi compañero y amigo de tesis Danilo Tapia y su familia por abrirme las puertas como mi segunda familia.

Finalmente, un gran agradecimiento al Ing. Roberto Salazar, por su paciencia prestada y compartir sus conocimientos los cuales fueron fundamentales para culminar este proceso académico.

*Carrillo Carrillo Manuel Francisco*

## **AGRADECIMIENTOS**

En primer lugar, a agradezco a Dios y mi negrita que siempre me cuidan espiritualmente, físicamente y saben guiarme por el sendero del bien.

A mi madre Silvia Guerrero y mi padre Marcos Tapia quienes fueron mi apoyo incondicional para mi crecimiento personal y profesional con sus cuidados, amor, esfuerzo y consejos que nunca faltan hasta hoy en día.

A mi hija Sofia y Fernanda quienes son un pilar fundamental en el desarrollo de mi vida.

A mis hermanos William, Brayan, David y hermanas Jazmin, Irma, Victoria que de una u otra manera siempre han estado a mi lado con sus consejos y apoyo incondicional desde el inicio de mi carrera hasta el final y un agradecimiento especial a mi hermana mayor Ximena, mi segunda madre quien me enseñó el don del trabajo y ser mejor cada día, y Carlos quien ha sido un ejemplo y motivación para crecer profesionalmente.

A mi compañero de tesis Francisco, gracias a su ayuda y trabajo, se logró culminar nuestra carrera universitaria.

Al Ingeniero Roberto Salazar, tutor de la tesis que gracias a su ayuda, consejos y apoyo en todo en momento se logró la culminación del proyecto de investigación y consecuente la terminación de la carrera universitaria.

Al Alma mater la Universidad Técnica de Cotopaxi, por abrirme las puertas y desarrollar mis estudios de ingeniería eléctrica. Un Sincero agradecimiento a los docentes de la carrera quienes, con su esfuerzo y compromiso, contribuyeron en mi formación personal y profesional.

*Jipson Danilo Tapia Guerrero*

## **DEDICATORIA**

Dedico este proyecto de investigación p a Dios a mis padres por ser quienes han estado incondicionalmente apoyándome en mi etapa universitaria y todos los ámbitos de mi vida.

A mis abuelos quienes me enseñaron valores y la dedicación para seguir adelante y poder alcanzar las metas propuestas.

*Carrillo Carrillo Manuel Francisco*

## **DEDICATORIA**

Dedico este proyecto de investigación principalmente a Dios y mi negrita, por darme fuerzas para continuar en los momentos más difíciles por brindarme salud, calma y sabiduría cuando más lo necesite en el transcurso de la carrera.

Le dedico al motor de mi vida mi hija Sofia quien se convirtió en el impulso para crecer profesionalmente y sobre todo brindarle un ejemplo y una mejor vida.

Para mi familia que siempre he contado con su apoyo y amor incondicional.

*Jipson Danilo Tapia Guerrero*



# ÍNDICE GENERAL

## Contenido

1.	INFORMACIÓN GENERAL.....	1
2.1	EL PROBLEMA.....	2
2.1.1	Situación Problemática .....	2
2.1.2	Formulación del Problema.....	2
2.2	OBJETO Y CAMPO DE ACCIÓN.....	2
2.3	BENEFICIARIOS.....	3
2.3.1	Beneficiarios Directos.....	3
2.3.2	Beneficiarios indirectos .....	3
2.4	JUSTIFICACIÓN .....	3
2.5	HIPÓTESIS.....	4
2.6	OBJETIVOS .....	4
2.6.1	Objetivo general.....	4
2.6.2	Objetivos Específicos.....	4
2.7	SISTEMA DE TAREAS.....	5
3	FUNDAMENTACIÓN TEÓRICA .....	6
3.1	ANTECEDENTES .....	6
3.2	CENTRAL ILLUCHI II .....	10
3.3	PREDICCIÓN DE LA PRODUCCIÓN DE ENERGÍA ELÉCTRICA.....	11
3.3.1	Curvas de carga diaria.....	12
3.3.2	Curvas de carga diaria .....	13
3.3.3	Curvas de carga anual .....	13
3.4	PROCESO ESTOCÁSTICO .....	14
3.4.1	Procesos estocásticos estacionarios.....	15

3.5	SERIES TEMPORALES.....	15
3.5.1	Análisis preliminar de una serie.....	16
3.5.2	Componentes de una serie temporal.....	16
3.5.2.1	Tendencia y nivel de la serie.....	16
3.5.2.2	Estacionalidad.....	17
3.5.2.3	Dependencia entre variabilidad y nivel.....	17
3.5.2.4	Ruido blanco.....	18
3.5.2.5	Comportamientos anómalos.....	18
3.6	Modelos a utilizar para predicción de series temporales.....	18
3.7	Modelos ARIMA.....	18
3.8.1	Identificación práctica del modelo.....	19
3.8	MODELOS AUTORREGRESIVOS AR(P).....	20
3.9	Modelo de medias móviles $Ma(q)$ .....	21
3.10	Modelos Arma (p, q).....	23
3.11	Modelos ARIMA (p, d, q).....	24
3.12.1	Dickey-Fuller aumentado (ADF).....	25
3.13	Metodología modelo ARIMA.....	25
3.13.1	Primera fase.....	25
3.13.2	Segunda fase.....	26
3.13.3	Tercera fase.....	26
3.13.4	Cuarta fase.....	26
3.14	Modelo estacional autorregresivo integrado de medias móviles (SARIMA).....	26
3.15	REDES Neuronales Artificiales.....	27
3.15.1	Rol de una neurona.....	27
3.16	Modelo matemático de la red neuronal artificial.....	28

3.17 Células LSTM.....	30
3.18 Medidas de centralización.....	31
3.18.1 La media aritmética.....	31
3.18.2 Mediana.....	32
3.19 CRITERIOS PARA LA VALIDACIÓN DE LOS RESULTADOS APLICANDO UN MÉTODO DE ANÁLISIS.....	32
3.19.1 Error medio cuadrático (MSE – mean square error).....	32
3.19.2 Raíz del Error Medio Cuadrático (Root Mean Squared Error – RMSE).....	32
3.19.3 Error Medio Absoluto (Mean Absolute Error MAE).....	33
3.19.4 Error Medio Absoluto Porcentual (Mean Absolute Percentage Error - MAPE).....	33
3.19.5 Error residual (EMM).....	33
3.19.6 Desviación media absoluta (MAD).....	34
3.19.7 El criterio de información de Akaike (AIC).....	34
3.19.8 Desviación estándar.....	34
3.19.9 Varianza.....	34
3.19.10 Función de correlación automática (ACF).....	35
3.19.11 Función de autocorrelación parcial (PACF).....	35
3.20 Lenguaje de programación Python.....	36
3.20.1 STATSMODELS en Python.....	36
3.21 Librerías necesarias para el uso del software de Python.....	36
3.21.1 Librería Numpy.....	36
3.21.2 Librería Pandas.....	36
3.21.3 Librería Matplotlib.....	37
3.21.4 Librería Scikit-Learn.....	37
3.21.5 Librería Keras.....	37

3.21.6	Librería Tensorflow.....	37
3.21.7	Uso de variables exógenas: ARMAX, ARIMAX y SARIMAX.....	38
4	METODOLOGÍA .....	39
4.1	TIPO DE INVESTIGACIÓN.....	39
4.1.1	Investigación descriptiva.....	39
4.1.2	Investigación diagnóstica .....	39
4.1.3	Investigación aplicada .....	39
4.2	MÉTODO DE INVESTIGACIÓN .....	40
4.2.1	Metodología de la investigación bibliográfica .....	40
4.2.2	Inductivo – deductivo.....	40
4.3	TÉCNICAS E INSTRUMENTOS.....	41
4.3.1	Pasos a seguir mediante un diagrama de flujo para realizar la predicción de la producción eléctrica de la central eléctrica ILLUCHI II mediante el método ARIMA. ....	41
4.3.2	Estructura de la red neuronal artificial para el desarrollo del algoritmo de predicción. 57	
5	ANÁLISIS DE RESULTADOS .....	59
5.1	ANÁLISIS DE LA CENTRAL HIDROELÉCTRICA ILLUCHI II .....	59
5.2	ANÁLISIS DE CORRELACIÓN DE LA INFORMACIÓN DE LA CENTRAL HIDROELÉCTRICA ILLUCHI II.....	60
5.3	PREPROCESAMIENTO DE LA INFORMACIÓN DE LA CENTRAL HIDROELÉCTRICA ILLUCHI II.....	60
5.3.1	Depuración de datos de la CENTRAL HIDROELÉCTRICA ILLUCHI II.....	61
5.3.2	Integración de datos de la central HIDROELÉCTRICA ILLUCHI II.....	62
5.3.3	Transformación de datos de la central ILLUCHI II.....	63
5.4	SELECCIÓN DE LAS VARIABLES DE ENTRADA PARA EL DESARROLLO DEL ALGOTIRMO DE PREDICCIÓN.....	64

5.5	VALIDACIÓN Y PREPROCESAMIENTO DE DATOS DE ENTRADA PARA EL DESARROLLO DEL ALGORITMO DE PREDICCIÓN.....	64
5.5.1	Desarrollo del algoritmo de predicción en Python.....	65
5.5.2	Preprocesado de los datos en Python .....	66
5.5.3	Descomposición de los datos de series temporales a su tendencia, estacionalidad y componentes residuales. ....	68
5.6	DEFINIR LA ESTRUCTURA para que la serie sea estacionaria .....	69
5.7	ESCENARIOS APLICADOS CON EL MODELO ARIMA.....	71
	Límite superior de control.....	71
	Límite inferior de control:.....	71
5.7.1	Escenario 1 .....	72
5.7.2	Escenario 2 .....	72
5.8	COMPARACIÓN DEL MODELO ARIMA FRENTE A REDES NEURONALES ARTIFICIALES (RNA) .....	76
5.8.1	Pronóstico de la producción de energía mediante modelo matemático ARIMA.....	76
5.8.1.1	MAD (Desviación Media Absoluta) .....	77
5.8.2	PRONÓSTICO DE LA PRODUCCIÓN DE ENERGIA MEDIANTE RNA-LSTM	78
5.9	COMPORTAMIENTO DE PRODUCCIÓN kW PARA SU VALIDACIÓN .....	80
6	CONCLUSIONES Y RECOMENDACIONES. ....	84
6.1	CONCLUSIONES .....	84
6.2	RECOMENDACIONES .....	85
7	REFERENCIAS.....	86
	INDICE DE ANEXOS .....	91
	Anexo A: INFORME ANTI PLAGIO PROYECTO DE TITULACION.....	92
	ANEXO b. Base de datos desde el año 2018 hasta el año 2022.....	92
	ANEXO E. Modelo AUTO-ARIMA.....	104

ANEXO F. Estructura del método ARIMA en los Escenarios..... 108

## ÍNDICE DE FIGURAS

<b>Figura 3.1.</b> Consumo energético por habitante (BEP/hab) [3]. .....	7
<b>Figura 3.1.</b> Consumo eléctrico por habitante (kWh/hab) [3]. .....	8
<b>Figura 3.1</b> Tanque de captación perteneciente a la central hidroeléctrica ILLUCHI II. ....	10
<b>Figura 3.3</b> Consumo diario eléctrico instantánea y promedio [5]. .....	12
<b>Figura 3. 6.</b> Curvas de carga diaria obtenidas – junio 2007 Ecuador [2]......	13
<b>Figura 3. 7.</b> Curvas de carga anual [6]......	14
<b>Figura 3. 9.</b> Gráfica de series de tiempo de las muestras simuladas [10] .....	15
<b>Figura 3. 10.</b> Procedencia de una serie temporal observada [11]. .....	16
<b>Figura 3. 11.</b> Tendencias y nivel de la serie [12]......	17
<b>Figura 3. 12.</b> La gráfica muestra un ruido blanco con media cero y varianza constante e igual a uno [13]......	18
<b>Figura 3. 12.</b> Gráficas Autorregresivos en margen de sus tendencias [8]. .....	21
<b>Figura 3. 13.</b> Gráfico Autorregresivos en margen de ACF y ACFP [8]. .....	21
<b>Figura 3. 14.</b> Gráficas Medias Móviles en margen de su tendencia [8] .....	22
<b>Figura 3. 15.</b> Gráfico Medias Móviles en margen de ACF y ACFP [8] .....	23
<b>Figura 3. 16.</b> Arquitectura de una neurona artificial [19]. .....	27
<b>Figura 3. 17.</b> Esquema de una red neuronal artificial [20]. .....	30
<b>Figura 3. 18.</b> Estructura interna de una célula LSTM [21].....	30
<b>Figura 3.20.</b> Gráfica de función de correlación automática [30]. .....	35
<b>Figura 3.21.</b> Gráfica de función de autocorrelación parcial [30]......	35
<b>Figura 4.1.</b> Flujograma metodológico empleado para la predicción de la producción de la energía eléctrica utilizando el modelo ARIMA.....	41
<b>Figura 4.2.</b> Ubicación del Tanque de almacenamiento central de generación ILLUCHI II [37]......	42

<b>Figura 4.3.</b> Tubería de presión perteneciente a la central hidroeléctrica ILLUCHI II.....	42
<b>Figura 4.4.</b> Tubería de presión perteneciente a la central hidroeléctrica ILLUCHI II [37]. ....	43
<b>Figura 4.5.</b> Producción media diaria anual de la base de datos. ....	43
<b>Figura 4.6.</b> Librerías empleadas en Python.....	44
<b>Figura 4.7.</b> Validación y preprocesamiento de la información procedente de una base de datos.....	45
<b>Figura 4.8.</b> Información de nuestro archivo data. ....	46
<b>Figura 4.9.</b> Gráfica de los datos reales para determinar la tendencia. ....	46
<b>Figura 4.10.</b> Gráfica de los datos reales para determinar la tendencia. ....	47
<b>Figura 4.11.</b> Gráfica de la descomposición de los datos.....	47
<b>Figura 4.12.</b> Librería empleada para el modelo.....	48
<b>Figura 4.13.</b> Librería empleada para la aplicación de la prueba de DICKEY-FULLER.....	48
<b>Figura 4.14.</b> Gráfica de los resultados al aplicar la prueba de DICKEY-FULLER. ....	49
<b>Figura 4.15.</b> Gráfica ACF. ....	49
<b>Figura 4.16.</b> Gráfica PACF.....	50
<b>Figura 4.17.</b> Partición de los datos en subconjuntos para la predicción. ....	51
<b>Figura 4.18.</b> Partición de los datos en subconjuntos para la predicción. ....	51
<b>Figura 4.19.</b> Ejecución del modelo automático de ARIMA. ....	52
<b>Figura 4.20.</b> Resultados modelo AUTO-ARIMA.....	52
<b>Figura 4.21.</b> Escenario 1 con nuestros datos inferiores. ....	53
<b>Figura 4.22.</b> Escenario 2 con nuestros datos inferiores. ....	53
<b>Figura 4.23.</b> Gráfico de línea de errores residuales. ....	53
<b>Figura 4.24.</b> Gráfico de los residuos. ....	54
<b>Figura 4.25.</b> Gráfico de la predicción. ....	55
<b>Figura 4.26.</b> Gráfica de la comparación de la potencia actual con la predicción. ....	55



<b>Figura 4.27.</b> Gráfica del Análisis del error. ....	56
<b>Figura 4.28.</b> Gráfica de las librerías del comportamiento de los datos.....	56
<b>Figura 4.29.</b> Gráfica del comportamiento de los datos. ....	57
<b>Figura 4.30:</b> Estructura de la red neuronal artificial implementado en el algoritmo de predicción. ....	57
<b>Figura 5.1.</b> Predicción 2023-2028 con base de datos desde el 2018. ....	59
<b>Figura 5.2.</b> Variables de datos obtenidos por ELEPCO S.A. ....	60
<b>Figura 5.2.</b> Secuencia de pasos a seguir para el preprocesamiento de la información procedente de una base de datos [39].....	61
<b>Figura 5.3.</b> Muestra de la base de datos del proceso de generación de la CENTRAL HIDROELÉCTRICA ILLUCHI II cada media hora enero 2018.....	62
<b>Figura 5.4.</b> Integración de la información hacia una nueva base de datos [36].....	63
<b>Figura 5.5.</b> Muestra de la base de datos depurada año 2018 del proceso de generación de la CENTRAL HIDROELÉCTRICA ILLUCHI II. ....	63
<b>Figura 5.6.</b> Diagrama de flujo en base del análisis de Datos para el modelo ARIMA [41]. ...	65
<b>Figura 5.7.</b> Aplicación de las librerías de Python para implementar el algoritmo de predicción.....	66
<b>Figura 5.8.</b> Datos creados por el nuevo Dataframe para la implementación del algoritmo de predicción. ....	67
<b>Figura 5.9.</b> Serie Temporal de la predicción junio2018-junio2023.....	67
<b>Figura 5.10.</b> Desviación Estándar por el intervalo de tiempo 2018-2023. ....	68
<b>Figura 5.11.</b> Serie temporal deducida por los datos modelados 2018-2023. ....	68
<b>Figura 5.12.</b> Tendencia que sigue la serie temporal desde el 2018-2023. ....	69
<b>Figura 5.13.</b> La estacionalidad de la serie temporal datos del 2018-2023.....	69
<b>Figura 5.14.</b> Los residuos de la serie temporal datos del 2018-2023.....	69
<b>Figura 5.15.</b> Estructura para el procesamiento de la prueba Dickey Fuller Aumentada. ....	70

<b>Figura 5.16.</b> Resultados de la prueba Dickey Fuller Aumentada en Python. ....	70
<b>Figura 5.17.</b> En caso de no ser estacionarios se realiza la diferenciación entre los datos. ....	70
<b>Figura 5.18.</b> Partes de un gráfico de control [42]. ....	71
<b>Figura 5.19.</b> Programación de la gráfica de control escenario 1 en Python. ....	72
<b>Figura 5.20.</b> Programación de la gráfica de control escenario 2 en Python. ....	72
<b>Figura 5.21.</b> Errores residuales escenario 1. ....	73
<b>Figura 5.22.</b> Los Errores residuales escenarios 2. ....	74
<b>Figura 5.23.</b> Relación entre la predicción del escenario 1 frente al actual análisis mensual 2022. ....	74
<b>Figura 5.24.</b> Relación entre la predicción del escenario 2 frente al actual análisis mensual 2022. ....	75
<b>Figura 5.25.</b> Gráfica de producción actual y la predicción con el método ARIMA 2022. ....	77
<b>Figura 5.27.</b> Producción Real frente al método ARIMA 2022. ....	78
<b>Figura 5.28.</b> Gráfica del MAD con RNA-LSTM 2022. ....	79
<b>Figura 5.29.</b> Gráfica de producción actual y la predicción con el método RNA 2022. ....	80
<b>Figura 5.30.</b> Gráfica de la producción actual con la predicción con el método RNA-LSTM y ARIMA enero-agosto 2022. ....	81
<b>Figura 5.31.</b> Gráfica de la producción actual con la predicción con el método RNA-LSTM y ARIMA 2022-2023. ....	82
<b>Figura 5.32.</b> Mapa de calor de la producción de la central ILLUCHI II. ....	83
<b>Figura 5.33.</b> Proyección de producción de la central ILLUCHI II desde el 2023-2028. ....	83
<b>Figura B.1.</b> Librerías ejecutadas en Python. ....	95
<b>Figura B.2.</b> Librerías para determinar errores ejecutadas en Python. ....	95
<b>Figura C.1.</b> Llamado de datos depurados. ....	96
<b>Figura C.1.</b> Información de los datos vacíos y espacios. ....	96
<b>Figura C.2.</b> Serie Temporal de la predicción junio2018-junio2023. ....	97

<b>Figura C.3.</b> Desviación Estándar por el intervalo de tiempo 2018-2023. ....	97
<b>Figura C.4.</b> Gráfica de la descomposición de los datos.....	98
<b>Figura D.1.</b> Librería empleada para la aplicación de la prueba de DICKEY-FULLER.....	99
<b>Figura D.2.</b> Resultados de la prueba de Dickey-Fuller.....	99
<b>Figura D.3.</b> Diferenciación de la producción KWH-DIA 2018-2023. ....	100
<b>Figura D.4.</b> Descomposición de la serie temporal 2018-2023.....	101
<b>Figura D.5.</b> Autocorrelación simple de la serie temporal. ....	102
<b>Figura D.6.</b> Autocorrelación parcial de la serie temporal.....	102
<b>Figura D.7.</b> División de datos de entrenamiento y testeo. ....	103
<b>Figura E.1.</b> Modelamiento AUTO-ARIMA. ....	104
<b>Figura E.2.</b> Búsqueda paso a paso para minimizar AIC.....	104
<b>Figura E.3.</b> Resultados de Testeo modelo AUTO-ARIMA. ....	105
<b>Figura E.4.</b> Implementación del modelo. ....	106
<b>Figura E.5.</b> Gráfico de línea de residuos AIC.....	106
<b>Figura E.6.</b> Modelo de diagnóstico.....	106
<b>Figura E.7.</b> Comparación del modelo ARIMA frente a la predicción análisis mensual. ....	106
<b>Figura E.7.</b> Predicción con el Modelo ARIMA análisis anual. ....	107
<b>Figura F.1.</b> Librerías ejecutadas en Python. ....	108
<b>Figura F.2.</b> Llamado de datos depurados.....	108

## ÍNDICE DE TABLAS

<b>Tabla 3.2.</b> Características de la central ILLUCHI II generador I Y II [4]. .....	11
<b>Tabla 3.3.</b> Características de las funciones de autocorrelación [16]. .....	20
<b>Tabla 4.1</b> Selección preliminar de las variables para la predicción de la producción de energía eléctrica.45	
<b>Tabla 4.2.</b> Selección preliminar de las variables para la predicción de la producción de energía eléctrica.58	
<b>Tabla 5.1.</b> Selección final de las variables de entrada para la predicción de la producción de energía eléctrica. ....	64
<b>Tabla 5.1.</b> Pronóstico de la producción de energía mediante modelo matemático ARIMA 2022.....	76
<b>Tabla 5.2.</b> Pronóstico de la demanda de energía mediante modelo RNA-LSTM 2022.....	79
<b>Tabla 5.3.</b> Comparación de la producción Real con el modelo ARIMA y RNA-LSTM 2022.81	
<b>Tabla 5.3.</b> Comparación de errores modelo ARIMA y el modelo RNA-LSTM 2022. ....	82

## RESUMEN

**Tema:** “ESCENARIOS DE PRODUCCIÓN DE ENERGÍA ELÉCTRICA EN UNA PEQUEÑA CENTRAL HIDROELÉCTRICA UTILIZANDO SERIES SINTÉTICAS”.

**Autores:**

Carrillo Carrillo Manuel Francisco

Tapia Guerrero Jipson Danilo

El presente trabajo de investigación propone el uso de un modelo matemático para la proyección de la producción de energía eléctrica de la central hidroeléctrica ILLUCHI II, basándose en la importancia que trae consigo el estudio del comportamiento futuro de la energía para la población local, para ello se emplea el modelo ARIMA, con los datos proporcionados por la empresa eléctrica de Cotopaxi (ELEPCO S.A). Los datos históricos abordan la producción energética desde el año 2018 hasta el año 2022, los cuales permiten generar una base datos en Excel depurada, consiguiendo dar inicio a la creación del algoritmo y su codificación en Jupiter Notebook, mediante el llamado de las diferentes librerías que brinda dicho software. En la aplicación del modelo existen ciertas pruebas como lo es la prueba de Dickey-Fuller aumentada, la cual permite la comprobación de la estacionariedad de la base datos de tal manera que permitió el modelamiento. La comprobación de la efectividad del modelo ARIMA se realizó mediante el análisis de las métricas de aprendizaje y la respectiva observación e interpretación de las gráficas que genera el propio algoritmo.

Finalmente, se concluye que este tipo de modelos son de gran utilidad, tanto para describir como para proyectar el comportamiento de la producción de energía eléctrica de ILLUCHI II.

**Palabras Clave:** ARIMA, energía, producción, proyección, algoritmo.

## **ABSTRACT**

**Theme: “ELECTRICITY PRODUCTION SCENARIOS IN A SMALL HYDROELECTRIC POWER PLANT USING SYNTHETIC SERIES”.**

**Authors:**

Carrillo Carrillo Manuel Francisco

Tapia Guerrero Jipson Danilo

The present research work proposes the use of a mathematical model for the projection of the production of electrical energy of the ILLUCHI II hydroelectric power station, based on the importance that the study of the future behavior of energy for the local population brings with it. The ARIMA model is used, with the data provided by the Cotopaxi electric company (ELEPCO S.A). The historical data address the energy production from the year 2018 to the year 2022, which allow generating an already refined Excel database, consequently starting the creation of the algorithm in Jupiter Notebook, by calling the different libraries that said software simultaneously with the ARIMA model. In the application of the model there are certain tests such as the Dickey-Fuller test increases which allows the verification of the stationarity of our database in such a way that it allows to continue with the modeling. The verification of the effectiveness of the ARIMA model is carried out through the analysis of the learning metrics and the respective observation and interpretation of the graphs generated by the algorithm itself.

Finally, it is concluded that this type of models are very useful, both to describe and to predict the behavior of the electrical energy production of ILLUCHI II.

**Keywords:** ARIMA, energy, production, data, algorithm.



## 1. INFORMACIÓN GENERAL

**Título:** ESCENARIOS DE PRODUCCIÓN DE ENERGÍA ELÉCTRICA EN UNA PEQUEÑA CENTRAL HIDROELÉCTRICA UTILIZANDO SERIES SINTÉTICAS

**Fecha de inicio:** octubre 2022

**Fecha de finalización:** febrero 2023

**Lugar de ejecución:** No aplica

**Facultad que auspicia:** Facultad de Ciencias de la Ingeniería y Aplicadas.

**Carrera que auspicia:** Carrera de Ingeniería Eléctrica.

**Proyecto Macro Asociado:** Desarrollo de Sistemas Eléctricos eficientes a nivel residencial, comercial e industrial.

**Equipo de Trabajo:**

Grupo de investigación: Energía y Tecnologías Apropriadas

- **Tutor de Titulación:**

Ing. Salazar Achig Roberto Ms.C

- **Estudiante:**

Carrillo Carrillo Manuel Francisco, Tapia Guerrero Jipson Danilo

**Área de Conocimiento:** 07 Ingeniería, Industria y Construcción / 071 Ingeniería y Profesiones Afines / 0713 Electricidad y Energía.

**Línea de investigación:** Energías Alternativas y Renovables, Eficiencia Energética y Protección Ambiental.

**Sublíneas de investigación de la Carrera:** Control y optimización en el uso de la energía del sector industrial, comercial y residencial.



## **INTRODUCCIÓN**

En este capítulo introductorio se presentan las bases a desarrollar la investigación de series temporales con el modelo ARIMA utilizando la base de datos de la central ILLUCHI II, se enfocará en la producción. Iniciando con la sección 2.1 el problema. En las secciones 2.1 y 2.2 se expone el campo de acción y los beneficiarios directos e indirectos. En la sección 2.4 se presentan las razones principales para la predicción de la central IILLUCHI II, en la sección 2.5 se plantea la hipótesis sobre el tema. Por último, en las secciones 2.6 y 2.7 se presentan los objetivos y el sistema de estudio de la investigación sobre el modelo ARIMA.

### **2.1 EL PROBLEMA**

#### **2.1.1 Situación Problemática**

Para mantener el servicio energético para posteriores años se debe tener un control y estudio de la producción de la central ILLUCHI II, de forma que se pueda predecir la generación energética debido al aumento poblacional, al paso de los años se eleve el consumo mayoritario, por eso es necesario contar con los datos registrados anualmente de la central ILLUCHI II, pues esto es sumamente importante para el diseño e implementación de métodos matemáticos.

Entre los métodos matemáticos de los datos obtenidos son los que permitirá determinar la predicción de los próximos años que conlleva a la producción energética de la central ILLUCHI II.

#### **2.1.2 Formulación del Problema**

Se requiere del análisis del comportamiento de los datos de producción mediante un modelo ARIMA para proyectar escenarios de producción de energía en la central ILLUCHI II.

### **2.2 OBJETO Y CAMPO DE ACCIÓN**

El objeto de estudio son los sistemas de distribución.

El campo de acción para el proyecto: 330000 ciencias Tecnológicas / 3306 Ingeniería y Tecnología Eléctricas / 3306.99 Transmisión y Distribución.

## **2.3 BENEFICIARIOS**

### **2.3.1 Beneficiarios Directos**

ELEPCO S.A, al no contar con un programa que permita conocer la proyección de la producción de energía eléctrica por parte de ILLUCHI II, los estudiantes de Ingeniería Eléctrica de la Universidad Técnica de Cotopaxi se beneficiarán al tratarse de un tema netamente estadísticos con el tratamiento de una base de datos y el uso de herramientas nuevas y un modelo matemático como lo es ARIMA.

### **2.3.2 Beneficiarios indirectos**

Los estudiantes, investigadores que se encuentren realizando estudios y algoritmos para la predicción relacionado con la producción de energía eléctrica.

## **2.4 JUSTIFICACIÓN**

La presente investigación se basa en la Agenda del 2030 y los Objetivos de Desarrollo Sostenibles específicamente en el objetivo siete que indica garantizar el acceso a una energía asequible, fiable, sostenible y moderna para todos, tomando como base de estudio, sobre todo las consecuencias que trajo consigo la misma, específicamente en la parte de producción energética en donde se debe garantizar que la población tenga acceso a la electricidad.

La producción se base en los datos históricos, es importante en la operación de sistemas eléctricos, debido a que es primordial en la planeación de la cantidad de potencia eléctrica que se debe generar para cubrir eficientemente, sin ella se desperdiciarían los recursos utilizados por las centrales generadoras de energía eléctrica.

Para ello mediante la simulación, se podrá analizar de mejor manera la producción de energía posterior.

Este análisis forma parte del proyecto macro formativo titulado:” Desarrollo de sistemas eléctricos eficiente a nivel residencial, comercial e industrial” y del grupo de Investigación de Energías y Tecnologías Apropriadas de Potencia de la Universidad.

## **2.5 HIPÓTESIS**

El análisis de producción energética anual de la central ILLUCHI II de ELEPCO S.A, permitirá conocer posibles escenarios de generación de energía.

## **2.6 OBJETIVOS**

### **2.6.1 Objetivo general**

Generar series sintéticas de la producción de energía de la central hidroeléctrica ILLUCHI II en base a los históricos de generación.

### **2.6.2 Objetivos Específicos**

- Revisar el estado del arte de las técnicas utilizada en la generación de series sintéticas de producción de energía.
- Generar series sintéticas de producción de energía a partir de datos reales utilizando el modelo ARIMA.
- Validar la serie sintética de producción de energía mediante la comparación de históricos y estadísticos de la serie de los años anteriores y actuales.

## 2.7 SISTEMA DE TAREAS

Objetivos específicos	Actividades	Resultados esperados	Técnicas, Medios e Instrumentos
<p><b>Revisar el estado del arte de las técnicas utilizadas en la generación de series sintéticas de producción de energía.</b></p>	<ul style="list-style-type: none"> <li>- Recopilación de información y modelos matemáticos y estadísticos.</li> <li>- Relacionar los datos que se dispone acorde de las necesidades del modelo para generar un sistema eficiente.</li> </ul>	<ul style="list-style-type: none"> <li>- Registro de datos recolectados evaluando el nivel y preferencias óptimas acorde a los modelos estadísticos.</li> <li>- Organizar de manera adecuada los datos en base al modelo a escoger.</li> </ul>	<ul style="list-style-type: none"> <li>- Libros</li> <li>- Artículos</li> <li>- Tesis</li> <li>- Central ILLUCHI II</li> <li>- Elepco S.A</li> </ul>
<p><b>Generar series sintéticas de producción de energía a partir de datos reales utilizando el modelo ARIMA</b></p>	<ul style="list-style-type: none"> <li>- Realizar una base de datos día mes de los datos proporcionados por la central hidroeléctrica.</li> <li>- Usar el modelo Arima para la generación de series sintéticas.</li> </ul>	<ul style="list-style-type: none"> <li>- Documentos con bases para la generación de las series compatibles entre diferente software.</li> <li>- Gráficas acorde a la base de datos que certifiquen la eficiencia del modelo.</li> </ul>	<ul style="list-style-type: none"> <li>- Excel</li> <li>- Python</li> </ul>
<p><b>Validar la serie sintética de producción de energía mediante la comparación de históricos y estadísticos de la serie de los años anteriores y actuales.</b></p>	<ul style="list-style-type: none"> <li>- Comparación de la serie de datos e históricos con la base original.</li> <li>- Evaluación de la serie generada en función de la metodología establecida y la producción de energía.</li> </ul>	<ul style="list-style-type: none"> <li>- Visualización y manejo de los datos entre del programa generado.</li> <li>- Asertividad al momento de la comparación de la nueva serie sintética con la serie original.</li> </ul>	<ul style="list-style-type: none"> <li>- Software especializado</li> </ul>

### **3 FUNDAMENTACIÓN TEÓRICA**

En este capítulo se presenta bases de la teoría para el desarrollo escenarios de producción de energía utilizando series sintéticas en la sección 3.1 presenta los hechos que se ha producido (antecedentes) y tienen relación con el tema, destacando sus aspectos más importantes, en la Sección 3.2 presenta los fundamentos las series sintéticas, el modelo ARIMA y otros elementos que servirá para el procedimiento que se desarrollará en los capítulos siguientes para conseguir el modelo.

#### **3.1 ANTECEDENTES**

La producción eléctrica es la parte esencial del consumo energético, debido al crecimiento poblacional y distintos escenarios en orden social influyen directamente en la diversidad de consumidores provocando el aumento de la energía eléctrica para satisfacer a los usuarios, dependiendo la producción en los intervalos de tiempo la central de generación deberá cumplir.

Según la información recopilada por Agencia de Regulación y Control de Electricidad (ARCONEL) [1], la producción de electricidad para el país se basa en el aporte proveniente de centrales hidroeléctricas, termoeléctricas, eólicas y fotovoltaicas. Entre esta clasificación participan las empresas generadoras, autogeneradoras, sistemas no incorporados e interconexiones internacionales se estableció en la que la producción eléctrica del 2015 correspondió en un 49,49 % a las centrales hidráulicas, 46,52 % a las centrales termoeléctricas, 1,93 % a las interconexiones internacionales, 1,54 % a la generación con biomasa (bagazo de caña), 0,37 % a las centrales eólicas y 0,14 % a las centrales fotovoltaicas.

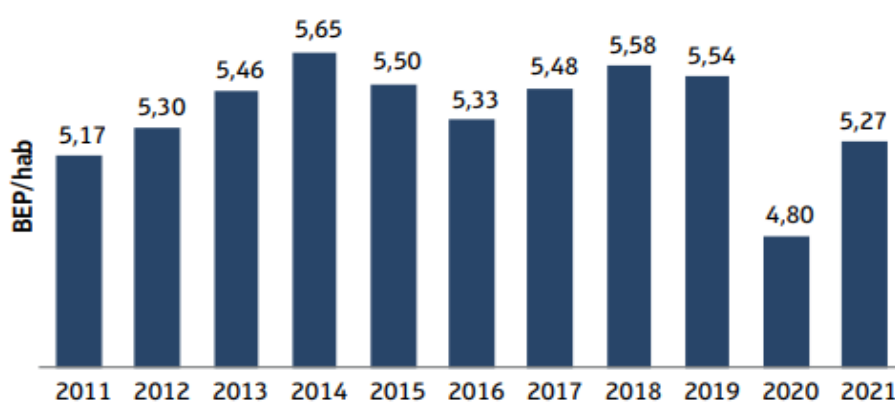
Salazar G. y Panchi B. describen en su documento [2], la revisión de la situación energética del Ecuador, se confirma la necesidad de la introducción de proyectos de eficiencia energética a gran escala que se dio a partir del 2008 y se investiga la influencia. Contando con datos históricos se proponen dos métodos para la predicción éstos se basan en el crecimiento tendencial, pero se diferencian al momento de considerar la evolución. Por otra parte, con base en la predicción obtenida, se analiza la influencia de cada proyecto de eficiencia energética ejecutado.

Por parte del documento Plan Nacional de Eficiencia Energética 2016-2035, publicado por el Ministerio de Electricidad y Energía Renovable (MEER) [3], establece que la capacidad instalada

de generación eléctrica ha evolucionado en la última década, pasando de 4.070 MW, en 2006, a 8089 MW, en 2016. Este incremento en la oferta de generación de electricidad fue posible gracias a nuevas centrales hidroeléctricas, las cuales se incorporaron al Sistema Nacional Interconectado (SNI). Las centrales Coca Codo Sinclair (1.500 MW), Sopladora (487 MW), Manduriacu (65 MW), Minas San Francisco (275 MW) y Mazar (160 MW), entre otras, permitieron incrementar la generación hidroeléctrica.

Mientras en el año 2020 el informe anual del Centro Nacional de Control de Energía (CENACE) [4], establece que el país alcanzó una producción neta total de energía de 26.979,96 GWh distribuida de la siguiente manera: 24.168,66 GWh generación hidroeléctrica; 2.159,64 GWh generación termoeléctrica; 400,86 GWh generación no convencional y 250,79 GWh importación desde Colombia, no se presentó importación desde Perú. En comparación al 2019, año en la que la producción neta fue 27.532,24 GWh distribuida de la siguiente manera: 24.458,10 GWh generación hidroeléctrica; 2.665,69 GWh generación termoeléctrica; 402,62 GWh generación no convencional y 5,83 GWh importación desde Colombia, sin presentarse importación desde Perú.

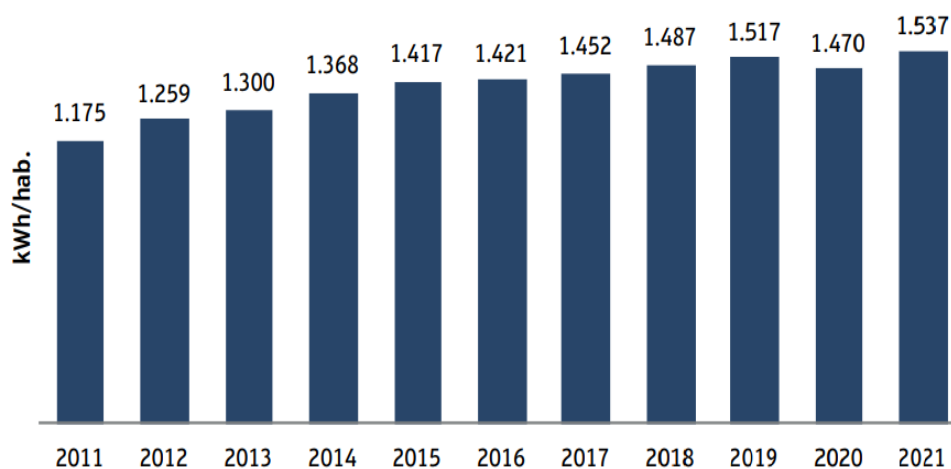
Según El Balance Energético Nacional 2021 [5], dice que entre 2011 y 2021, se observó un incremento en el consumo energético por habitante del país de 1,9%, pasando de 5,17 BEP/hab. a 5,27 BEP/hab. En 2020 se produjo una reducción marcada en el consumo energético total del país, debido al impacto generado por las medidas tomadas para mitigar los efectos de la pandemia de COVID-19; por esta razón se observa una recuperación del consumo por habitante en 2021 con respecto a 2020 de 9,8% que se muestra en la Figura 3.1.



**Figura 3.1.** Consumo energético por habitante (BEP/hab) [3].

El consumo de energía eléctrica por habitante aumentó un 30,9% entre 2011 y 2021, pasando de 1.175 kWh por persona a 1.537 kWh por persona. Igualmente se puede apreciar una recuperación

de este indicador en 2021 con respecto a 2020, al pasar de 1.470 kWh por habitante a 1.537 kWh por habitante, lo que representa el 4,6%, como se muestra en la Figura 3. 2.



**Figura 3.1.** Consumo eléctrico por habitante (kWh/hab) [3].

La Agencia de Regulación, Control de Energía y Recursos Naturales No Renovables (ARCERNNR) menciona que el 2021 [6], que a nivel nacional se registraron 8.734,41 MW de potencia nominal y 8.100,68 MW de potencia efectiva, las mismas consideran las capacidades registradas por los generadores, autogeneradores y distribuidores con generación. Cabe mencionar que la mayoría de las empresas petroleras corresponden a los autogeneradores.

Establece S. Trujillo [7], que una central hidroeléctrica se utiliza energía hidráulica para la generación de energía eléctrica. En general, estas centrales aprovechan la energía potencial que posee la masa de agua de un cauce natural en virtud de un desnivel, también conocido como salto geodésico. El agua en su caída entre dos niveles 3 del cauce se hace pasar por una turbina hidráulica la cual transmite la energía a un generador donde se transforma en energía eléctrica.

El trabajo desarrollado por Detzel, Bessa, Rita, y Mine [8], menciona que el uso de modelos para generar series de tiempo sintéticas es ampliamente aceptado en varias áreas. En los que se limita el uso de información histórica, entre ellos la planificación de los recursos hídricos. Depende de la aplicación, sin embargo, el uso de todos los escenarios sintéticos puede ser computacionalmente costoso, por lo que es necesaria la simplificación de los sistemas utilizados. En este contexto, se propone un estudio muestral de series sintéticas, con la intención de reducir el número de escenarios generados sin perder la representatividad obtenida con el modelo.

La investigación de las redes neuronales artificiales [9], se remonta a 1943, año en el que Warren McCulloch y Walter Pitts presentaron el primer modelo de redes artificiales. Desde ese hito se han

realizado propuestas nuevas y más sofisticadas. Las redes neuronales presentan una estructura multicapa que permite que la información se transmita no solo entre las neuronas vecinas sino a unidades más distantes. Una red neuronal artificial no necesita una programación, los parámetros de la red son encontrados adaptativamente. Aunque las redes neuronales artificiales son potencialmente más eficientes que la programación convencional en ciertas áreas de aplicaciones, al momento de su concepción no estaba lista para diversas aplicaciones. La teoría necesaria aún se encuentra en desarrollo y se prevé continúe por muchos años.

El control estadístico de la producción consiste en la aplicación de los modelos ARIMA.

Los modelos ARIMA son una familia de modelos lineales para el análisis y pronóstico de series temporales para datos de series temporales tanto estacionarios como no estacionarios, estacionales y no estacionales [10].

Los autores [11], proponen el pronóstico de carga a corto plazo usando el método de media móvil integrado autorregresivo (ARIMA) y red neuronal artificial (RNA) basado en carga no lineal. Este método de pronóstico utilizando redes neuronales artificiales, basado en el algoritmo de propagación de retroalimentación.

Según Francisco Reyes [12], menciona la posibilidad modelar los consumos de energía eléctrica utilizando los métodos ARIMA, dado que el comportamiento de estas series no es completamente aleatorio y se pueden describir como series de tiempo con una alta probabilidad de éxito en el modelado a través de dicha metodología. Mientras el modelado mediante redes neuronales arroja también valores relativamente precisos, incluso ofreciendo resultados virtualmente idénticos a los proporcionados por el modelo ARIMA. No obstante, la presencia de ciertas oscilaciones en las estimaciones hace recordar la “desconexión” existente entre la red neuronal y la realidad.

Según Rehan Jamil en su artículo “Pronóstico de consumo de energía hidroeléctrica para Pakistán utilizando modelos ARIMA y análisis para el año 2030” [13], presenta un estudio de pronóstico del consumo de energía hidroeléctrica en Pakistán basado en los datos históricos de los últimos 53 años utilizando el promedio móvil integrado autorregresivo (ARIMA) modelado. Con base en la ecuación de pronóstico desarrollada, se pronosticó el consumo de energía hidroeléctrica hasta el año 2030. Para validar la confiabilidad de los datos pronosticados, los resultados se compararon con los valores reales que se muestra un buen ajuste con una desviación mínima. Los valores pronosticados de consumo hidroeléctrico revelaron un incremento promedio anual de 1,65% con un aumento acumulado de 23,4% hasta el año 2030. Los resultados fueron comparados con los planes de generación hidroeléctrica del Gobierno de Pakistán para su efectividad. También



se realizó un análisis de sensibilidad para estudiar la relación del consumo de energía hidroeléctrica con la tasa de crecimiento anual de la población y el PIB del país. La investigación es significativamente útil para una mejor planificación y gestión de los recursos hídricos de Pakistán.

### 3.2 CENTRAL ILLUCHI II

La Central Hidroeléctrica ILLUCHI II, es una central de caudal de paso y está compuesta de dos turbinas marca Bell, tipo Pelton, que accionan dos generadores marca BBC que funcionan a un voltaje de 2400 voltios. La potencia total instalada es de 6.5 MVA distribuida en dos grupos de 3250 kVA a 720 rpm. La subestación de elevación está formada por un transformador de 6.5 MVA con un voltaje de 2.4/13.8 kV y que se enlazan al sistema nacional interconectado, mediante una línea trifásica de transmisión de 13.8 kV, la misma que está conformada por conductores N° 466 MCM ACSR, estructuras de hormigón tipo 3 CP y tiene una longitud de 10 km desde la subestación de la central hasta la S/E El Calvario [14].

**Tabla 3.1.** Características de la central ILLUCHI II.

Ubicación	Parroquia	Juan Montalvo, Pusuchusí Alto
	<b>Cantón</b>	Latacunga
	<b>Coordenadas</b>	772916,60 E 9896516,29 N
Fecha de puesta en servicio	<b>Año</b>	1984
Niveles de tensión que opera	<b>Voltaje de Generación</b>	2400 V nominal
	<b>Voltaje ingreso Subestación</b>	2400 V nominal
	<b>Voltaje salida Subestación</b>	13800 V (13.8 KV)
	<b>Voltaje servicio de auxiliares</b>	220VAC / 125 VDC



**Figura 3.1** Tanque de captación perteneciente a la central hidroeléctrica ILLUCHI II.

**Tabla 3.2.** Características de la central ILLUCHI II generador I Y II [4].

<b>Generadores 1 y 2</b>	3250	kVA	Potencia aparente
	0,8	-	Factor de potencia
	60	Hz	Frecuencia
	720	rpm	Velocidad
	10	-	Polos
	2400±5%	V	Voltaje
	781.8	A	Corriente
<b>Generador 1</b>	1,51*	p.u.	Impedancia ( $X_d$ )
	1,057*	p.u.	Impedancia ( $X_q$ )
<b>Generador 2</b>	1,69*	p.u.	Impedancia ( $X_d$ )
	1,183*	p.u.	Impedancia ( $X_q$ )
<b>Turbina 1 y 2</b>	2675	kW	Potencia activa
	950	l/s	Caudal
	720/1320	rpm	Velocidad
<b>Transformador Principal</b>	2400/13800	V	Voltaje

### 3.3 PREDICCIÓN DE LA PRODUCCIÓN DE ENERGÍA ELÉCTRICA

La predicción de la producción es un aspecto importante en el desarrollo de cualquier para la planificación de un sistema eléctrico. Las predicciones a corto plazo son fundamentales para el control y programación de los sistemas de energía y ellas son necesarias como dato entrada para los algoritmos de planificación, generación y transmisión de electricidad [16].

En general se los puede clasificar en tres principales grupos [16]:

1. Modelos estocásticos.
2. Inteligencia artificial.

3. Aquellos en base al criterio humano.

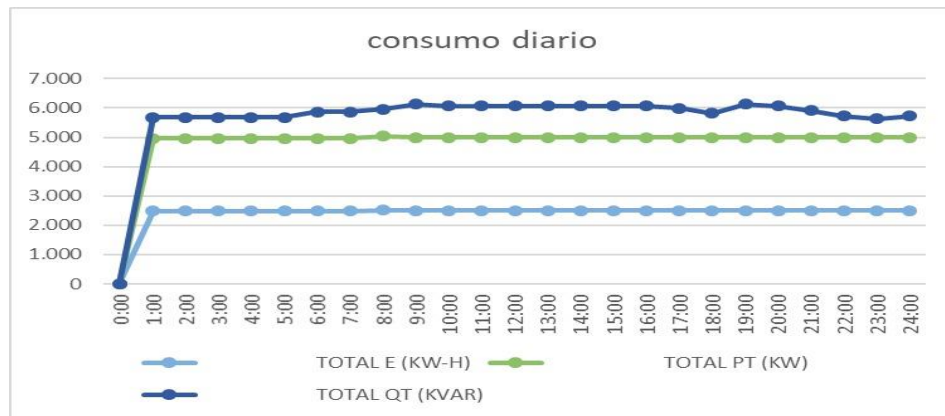


Figura 3.3 Consumo diario eléctrico instantánea y promedio [5].

3.3.1 Curvas de carga diaria

Las curvas de carga diaria están formadas por los picos obtenidos en intervalos de una hora para cada hora del día. Las curvas de carga diaria dan una indicación de las características de la carga en el sistema, sean estas predominantemente residenciales, comerciales o industriales y de la forma en que se combinan para producir el pico Figura 3.6. Su análisis debe conducir a conclusiones similares a las curvas de carga anual, pero proporcionan mayores detalles sobre la forma en que han venido variando durante el período histórico y constituye una base para determinar las tendencias predominantes de las cargas del sistema, permite seleccionar en forma adecuada los equipos de transformación en lo que se refiere a la capacidad límite de sobrecarga, tipo de enfriamiento para transformadores de subestaciones y límites de sobrecarga para transformadores de distribución [6].

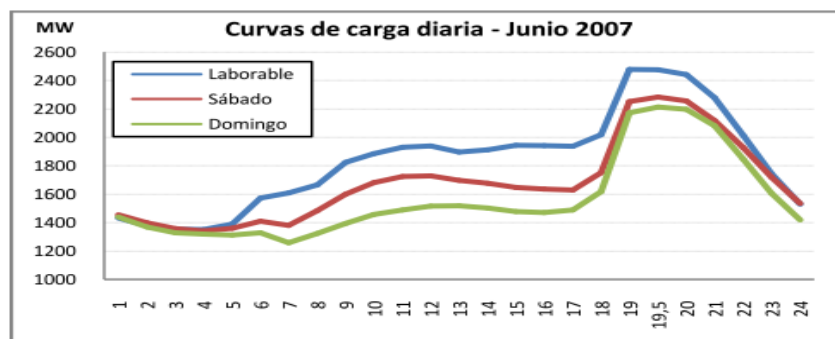


Figura 3. 5. Curvas de carga diaria obtenidas – junio 2007 Ecuador [2].

### 3.3.2 Curvas de carga diaria

Las curvas de carga diaria están formadas por los picos obtenidos en intervalos de una hora para cada hora del día. Las curvas de carga diaria dan una indicación de las características de la carga en el sistema, sean estas predominantemente residenciales, comerciales o industriales y de la forma en que se combinan para producir el pico Figura 3.6. Su análisis debe conducir a conclusiones similares a las curvas de carga anual, pero proporcionan mayores detalles sobre la forma en que han venido variando durante el período histórico y constituye una base para determinar las tendencias predominantes de las cargas del sistema, permite seleccionar en forma adecuada los equipos de transformación en lo que se refiere a la capacidad límite de sobrecarga, tipo de enfriamiento para transformadores de subestaciones y límites de sobrecarga para transformadores de distribución [6].

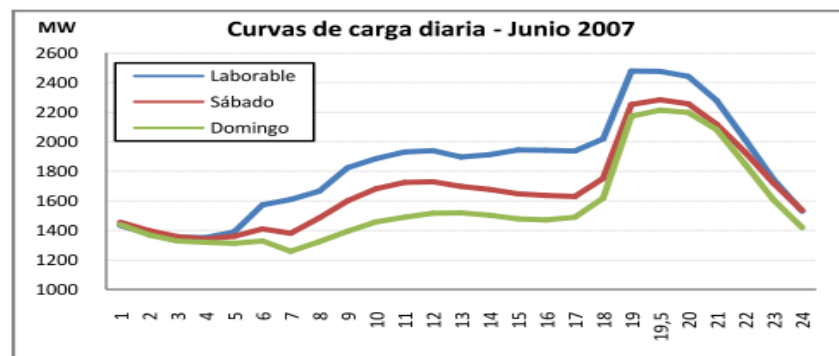
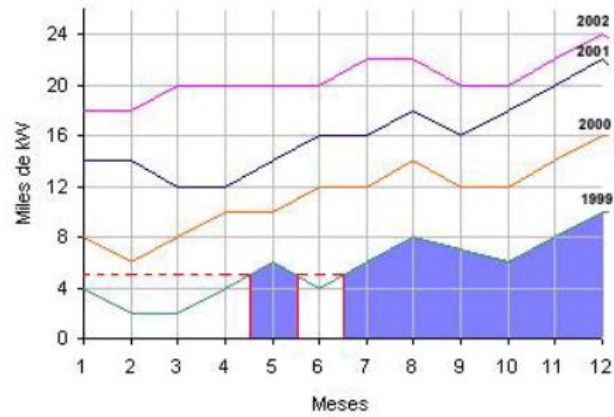


Figura 3. 6. Curvas de carga diaria obtenidas – junio 2007 Ecuador [2].

### 3.3.3 Curvas de carga anual

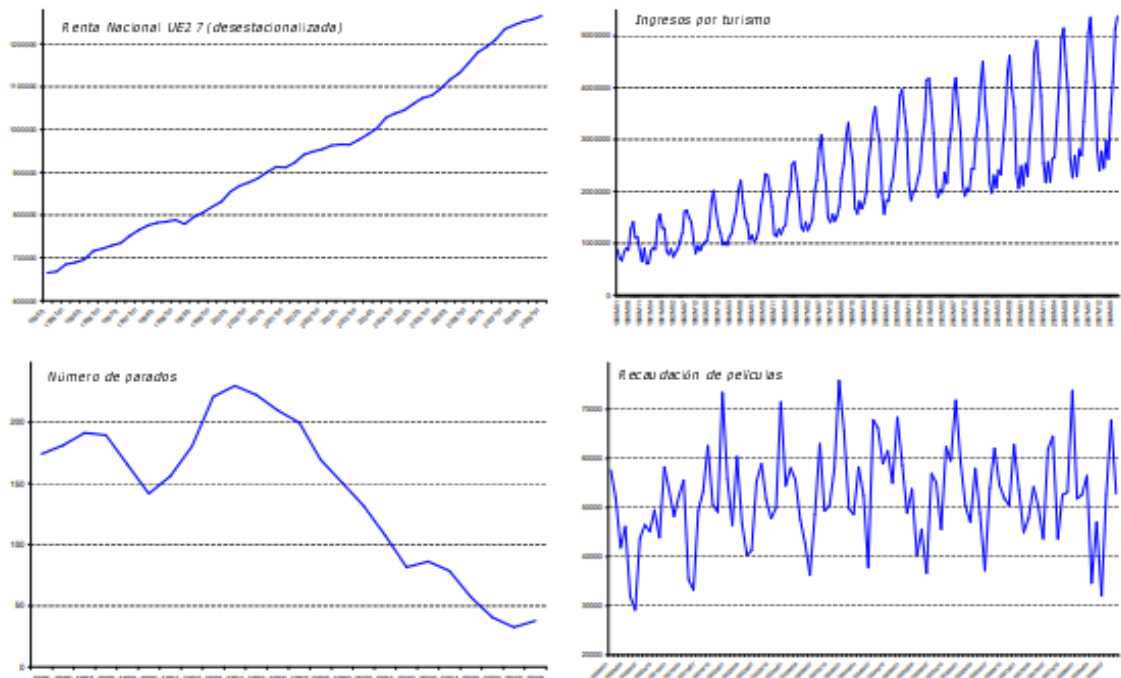
Las curvas de carga anual están formadas por los valores de la producción a la hora pico en cada mes, permiten una visualización de los crecimientos y variaciones de los picos mensuales y anuales que se puede observar en Figura 3.7. El análisis de las causas de estas variaciones debe conducir a conclusiones prácticas sobre el comportamiento del sistema y los factores que lo afectan.



**Figura 3. 7.** Curvas de carga anual [6].

### 3.4 PROCESO ESTOCÁSTICO

En la teoría de probabilidad [7], un proceso estocástico es un concepto matemático que sirve para representar magnitudes aleatorias que varían con el tiempo o para caracterizar una sucesión de variables aleatorias evolucionan en función de otra variable Figura 3.8, generalmente el tiempo.



**Figura 3. 8.** Gráficas estocásticas [7].

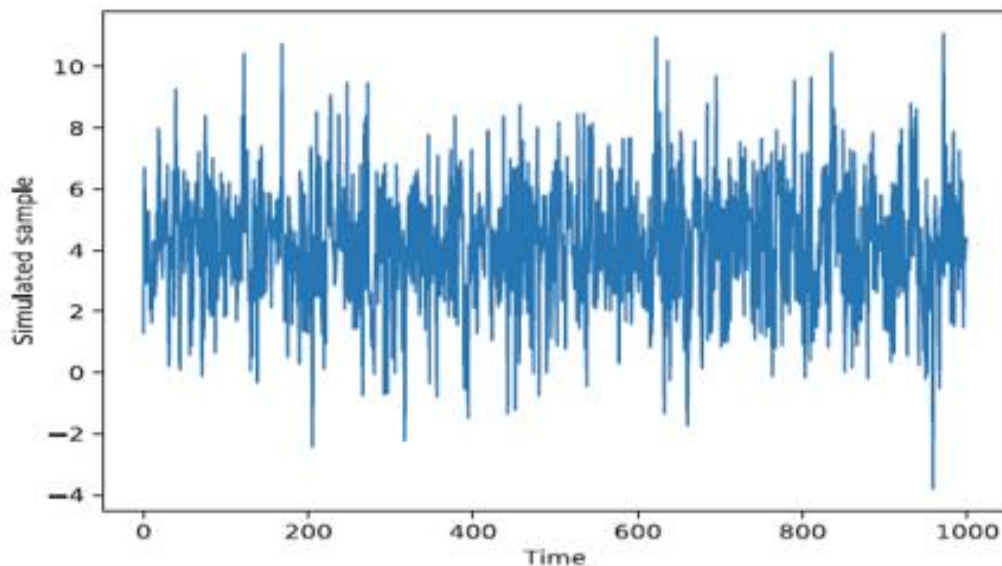
### 3.4.1 Procesos estocásticos estacionarios

En el análisis de series temporales el objetivo es utilizar la teoría de procesos estocásticos para determinar qué proceso estocástico ha sido capaz de generar la serie temporal bajo estudio con el fin de caracterizar el comportamiento de la serie y predecir en el futuro. Si se quieren conseguir métodos de predicción consistentes, no se puede utilizar cualquier tipo de proceso estocástico, sino que es necesario que la estructura probabilística del mismo sea estable en el tiempo [8].

## 3.5 SERIES TEMPORALES

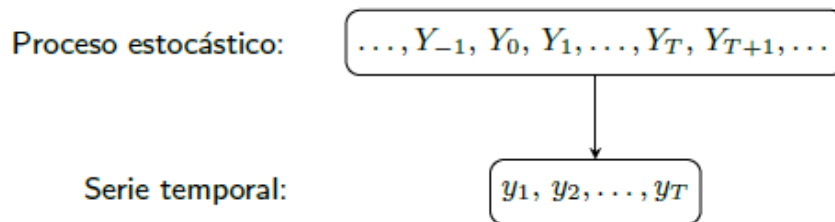
Una serie temporal es una colección de observaciones de una variable realizadas de forma secuencial en el tiempo, en las que el orden de observación es importante. Los valores de una serie temporal van ligados a instantes de tiempo, de manera que el análisis de una serie implica el manejo conjunto de dos variables; la variable en estudio propiamente dicha y la variable tiempo. Las series pueden tener una periodicidad anual, semestral, trimestral, mensual, etc., según los periodos de tiempo en los que están recogidos los datos que la componen Figura 3.9.

El análisis de series temporales presenta un conjunto de técnicas estadísticas que permiten, además de estudiar y modelizar el comportamiento de un fenómeno que evoluciona a lo largo del tiempo, realizar previsiones de los valores que se alcanzarán en el futuro [9].



**Figura 3. 9.** Gráfica de series de tiempo de las muestras simuladas [10]

Una serie temporal, por tanto, está definida sobre un periodo muestral que es tan solo una porción de la historia del proceso estocástico del que dicha serie procede. Es por ello que, en la práctica, el análisis de series temporales se realiza con un conjunto finito de datos, mientras que los modelos teóricos se construyen sobre su proceso estocástico asociado, pues consideran un número infinito de observaciones [11].



**Figura 3. 10.** Procedencia de una serie temporal observada [11].

### 3.5.1 Análisis preliminar de una serie

El análisis preliminar de una serie constituye el primer paso a seguir a la hora de estudiar una serie temporal. Esta fase permite detectar las características más importantes de una serie, tales como su tendencia (creciente o decreciente), la existencia de ciclos, presencia de valores atípicos, etc.

La forma más sencilla de comenzar el análisis de una serie temporal es mediante su representación gráfica. El gráfico que se emplea para representar las series temporales es el gráfico de secuencia. Los gráficos de secuencia son diagramas de líneas en los cuales el tiempo se representa en el eje de abscisas (x), y la variable cuya evolución en el tiempo de estudio en el eje de ordenadas (y) [12].

### 3.5.2 Componentes de una serie temporal

El análisis de las series temporales considera que una serie temporal queda formada por componentes:

#### 3.5.2.1 Tendencia y nivel de la serie

El nivel de una serie es una medida local de tendencia central, como por ejemplo la media, de cada periodo de tiempo que se considera. Cuando se trabaja con un calendario (tiempo representado

en días, meses o años), no es recomendable establecer periodos de tiempo antinaturales para estudiar esta característica.

Una vez calculado el nivel de la serie se debe observar su estabilidad, ver si la medida de tendencia central elegida tiene valores similares en los periodos de tiempo en los que divide el calendario. También se tendrá que observar su tendencia, es decir, si presenta una dirección constante de cambio de nivel [12].



**Figura 3. 11.** Tendencias y nivel de la serie [12].

### 3.5.2.2 Estacionalidad

Una serie es estacional cuando se puede observar en ella un patrón sistemático que se repite periódicamente (cada año, cada mes, etc., dependiendo de las unidades de tiempo en que vengán recogidos los datos) [12].

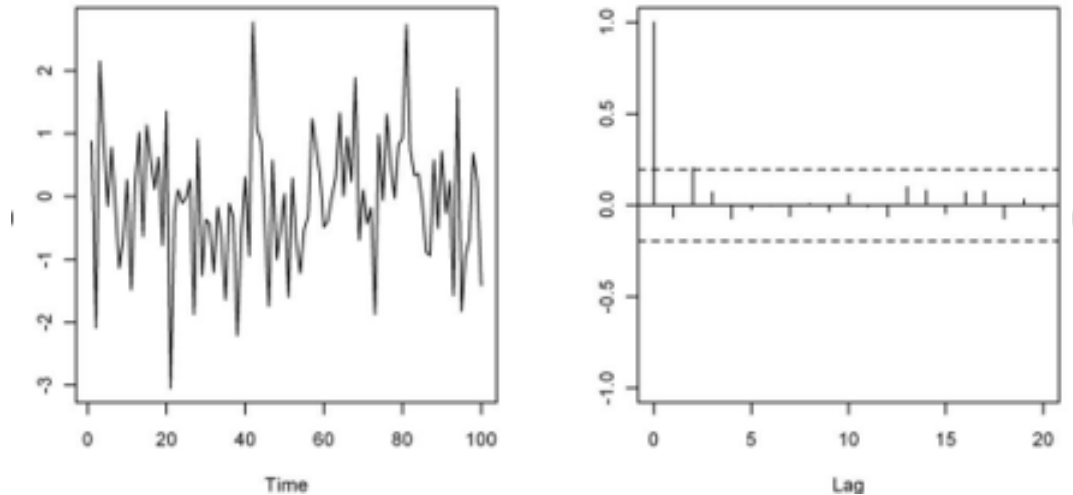
### 3.5.2.3 Dependencia entre variabilidad y nivel

Si la variabilidad [9] de una serie no depende del nivel significa que los componentes de la serie se combinan de forma aditiva, es decir, el incremento debido a la estacionalidad siempre es el mismo, aunque exista tendencia creciente o decreciente. Si la variabilidad y el nivel dependen entre sí los elementos de la serie se combinan de forma multiplicativa. Esto quiere decir que el incremento debido a la estacionalidad aumenta o disminuye conforme la tendencia crece o decrece.



### 3.5.2.4 Ruido blanco

Un ruido blanco es un caso simple de los procesos estocásticos, donde los valores son independientes e idénticamente distribuidos a lo largo del tiempo con media cero e igual varianza, se denota por  $\varepsilon_t$ .



**Figura 3. 12.** Ruido blanco con media cero y varianza constante e igual a uno [13].

### 3.5.2.5 Comportamientos anómalos

Si una serie temporal tiene valores perdidos (en un determinado momento no se han recogido datos) o valores raros, no se puede ignorarlos. Los comportamientos anómalos pueden ser de tres tipos: cambios de tendencia, subidas bruscas de nivel o aparición de valores extraños [9].

## 3.6 Modelos a utilizar para predicción de series temporales

El análisis y proceso de descomposición de una serie temporal ahora se utilizará modelos para realizar predicción de series temporales. La tendencia y estacionalidad serán un papel muy importante a la hora de la predicción y entrenamiento de los modelos; unos serán más sensibles y tratarán mejor la estacionalidad y/o tendencia, mientras que otros la ignorarán, o simplemente influirán negativamente en la predicción [14].

## 3.7 Modelos ARIMA

Son modelos paramétricos que tratan de obtener la representación de la serie en términos de la interrelación temporal de sus elementos. Este tipo de modelos que caracterizan las series como

sumas o diferencias, ponderadas o no, de variables aleatorias o de las series resultantes, fue propuesto por Yule y Slutsky en la década de los 20. Fueron la base de los procesos de medias móviles y autorregresivos que han tenido un desarrollo espectacular tras la publicación en 1970 del libro de Box-Jenkins sobre modelos ARIMA [15]. Los análisis se basan en un modelo explícito. Los modelos se conocen con el nombre genérico de ARIMA (AutoRegresive Integrated Moving Average), que deriva de sus tres componentes AR (Autoregresivo), I(Integrado) y MA (Medias Móviles) [16].

### 3.8.1 Identificación práctica del modelo

Identificar un modelo significa la utilización de los datos recogidos, así como cualquier información de cómo se generó la serie temporal objeto de estudio, para sugerir un conjunto reducido de posibles modelos, que tengan muchas posibilidades de ajustarse a los datos. Ante una serie temporal empírica, se deben encontrar los valores (p, d, q) más apropiados [16].

- Decidir si  $x_t$  necesita ser transformada para eliminar la no estacionariedad en media p en la no estacionariedad en varianza (heteroscedasticidad). Puede ser conveniente utilizar logaritmos de la serie o aplicar la transformación de Box-Cox.
- La determinación del grado de diferenciación adecuado. En general, la falta de estacionariedad se manifiesta en que los coeficientes de la función de autocorrelación estimada tienden a decrecer muy lentamente [16].

Decidir los valores de (p, q), y si existe una componente estacional, decidir los órdenes de los operadores estacionales (P, Q). Para este apartado se utilizan las funciones de autocorrelación (ACF) y autocorrelación parcial (ACFP) según el siguiente cuadro:

**Tabla 3.3.** Características de las funciones de autocorrelación [16].

<b>Proceso</b>	<b>Función de autocorrelación (ACF)</b>	<b>Función de autocorrelación parcial (ACFP)</b>
<b>MA(q)</b>	Solo los q primeros coeficientes son significativos. El resto se anulan bruscamente (coef. 0 para retardo >q)	Decrecimiento rápido exponencial atenuado u ondas sinusoidales.
<b>AR(p)</b>	Decrecimiento rápido exponencial atenuado u ondas sinusoidales.	Solo los p primeros coeficientes son significativos. El resto se anulan bruscamente (coef. 0 para retardo >q)
<b>ARIMA (p, d, q)</b>	Comportamiento irregular en los retardos (1,...,q) con q picos. Decrecimiento para retardos posteriores a q.	Decrece (aproximadamente con exponenciales atenuados y ondas sinusoidales). No cero pronto.

### 3.8 MODELOS AUTORREGRESIVOS AR(P)

Un modelo autorregresivo AR describe una clase particular de proceso en que las observaciones en un momento dado son predecibles a partir de las observaciones previas del proceso más un término de error [16]. El caso más simple es el ARIMA (1,0,0) o AR (1) o de primer orden ecuación (3.5), cuya expresión matemática.

$$AR(p) = c + \sum_{i=1}^p \Phi_i X_{t-i} + \epsilon_t \tag{3.1}$$

Donde:

$\Phi_i \dots \Phi_p$  son parámetros del modelo.

c es la constante.

$\epsilon_t$  es el ruido blanco.

$X_t$  variable relacionada con un periodo de anterior.

El proceso autorregresivo de orden en la parte pura se representa que  $Y_t$  depende solo de sus propios retrasos. Es decir  $Y_t$  es una función de rezagos.

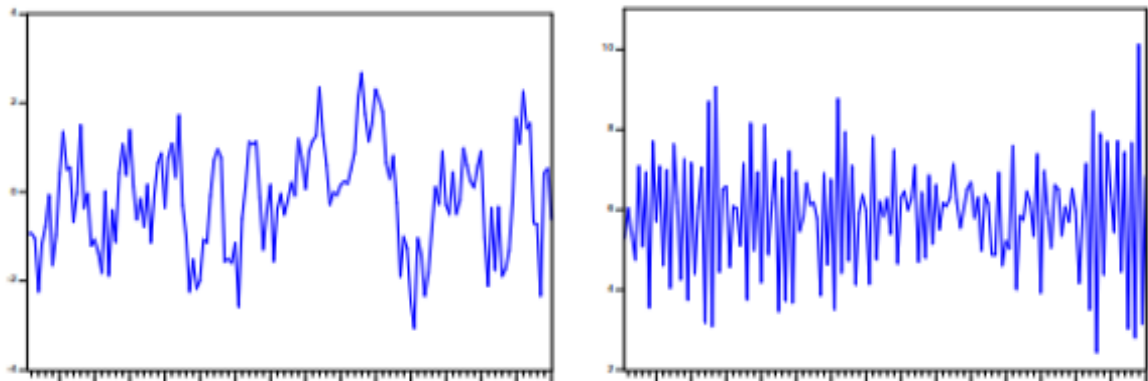
$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p} + \epsilon_1 \quad (3.2)$$

Donde:

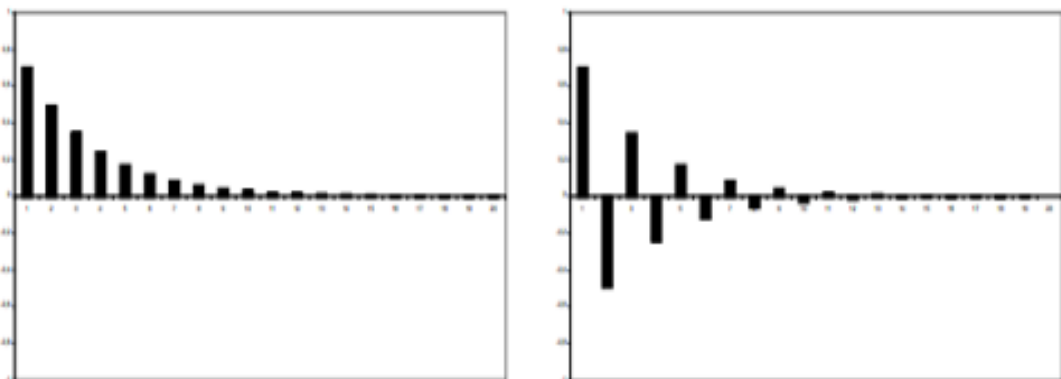
$Y_{t-1}$  es el rezago de la serie.

$\beta_1$  es el coeficiente de rezago que estima el modelo.

$\alpha$  es el término de la intersección.



**Figura 3. 12.** Gráficas Autorregresivos en margen de sus tendencias [8].



**Figura 3. 13.** Gráfico Autorregresivos en margen de ACF y ACFP [8].

### 3.9 Modelo de medias móviles $Ma(q)$

Un modelo de medias móviles MA describe una serie temporal estacionaria. En este modelo el valor actual puede predecirse a partir de la componente aleatoria de este momento y, en menor

medida Figura 3.17, de los impulsos aleatorias anteriores [16]. El modelo ARIMA (0,0,1), también denotado por MA (1), viene dado por la ecuación (3.12):

$$MA(q) = \sum_{i=1}^q \Phi_i \epsilon_{t-i} + \epsilon_t \quad (3.3)$$

Donde:

$\Phi_i \dots \Phi_p$  son parámetros del modelo.

$c$  es la constante.

$\epsilon_t$  es el ruido blanco.

$\epsilon_{t-1}$  son los términos de error.

El proceso para la media móvil de orden en la parte pura se representa los términos de error son los errores de los modelos autorregresivos de los respectivos rezagos [14].

$$Y_t = \alpha + \epsilon_t + \Phi_1 \epsilon_{t-1} + \Phi_2 \epsilon_{t-2} + \dots + \Phi_q \epsilon_{t-q} \quad (3.4)$$

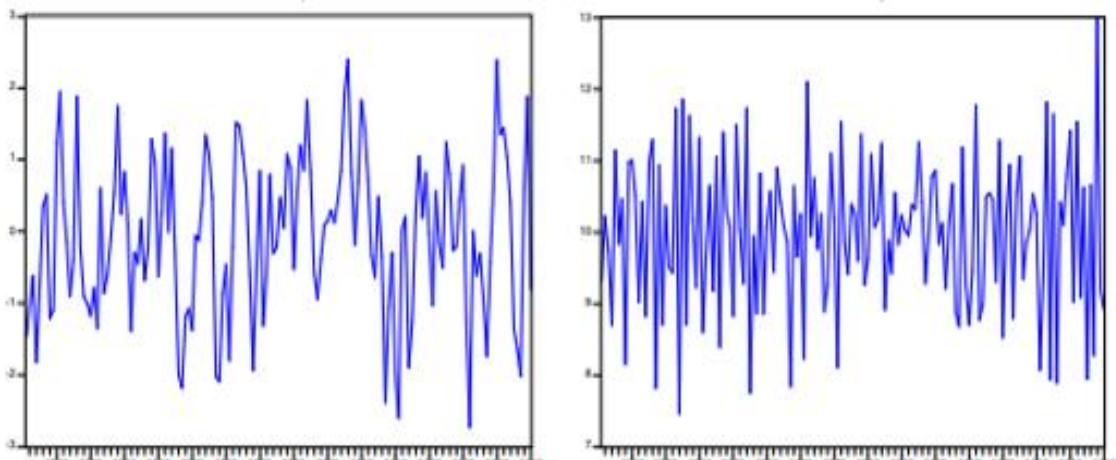
Donde:

$Y_t$  es el rezago de la serie.

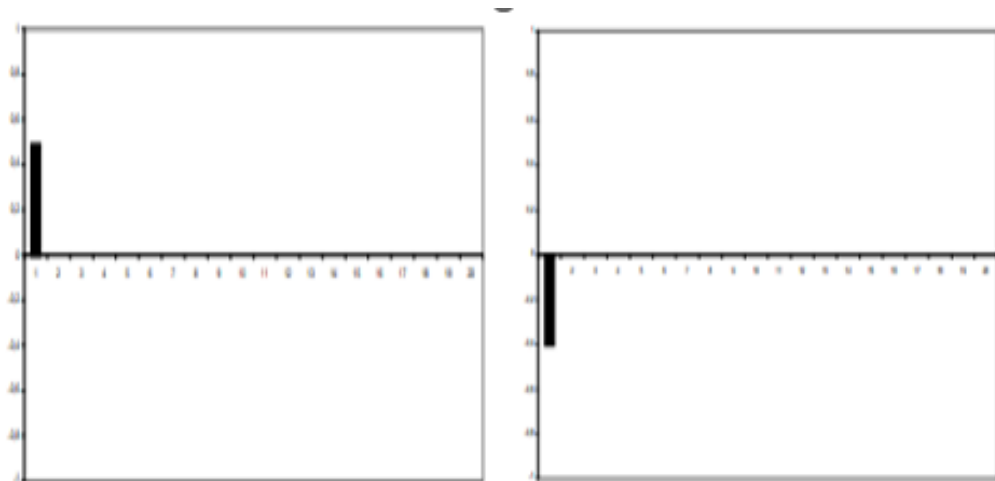
$\epsilon_{t-q}$  son errores de la serie.

$\alpha$  es el término de la intersección.

$\Phi_p$  son parámetros del modelo.



**Figura 3. 14.** Gráficas Medias Móviles en margen de su tendencia [8].



**Figura 3. 15.** Gráfico Medias Móviles en margen de ACF y ACFP [8].

### 3.10 Modelos Arma (p, q)

Es un modelo que está formado por dos partes: una autorregresiva (AR), y una parte de media móvil (MA). Habitualmente el modelo se nombra de la forma:

$$ARMA(p, q) = \sum_{i=1}^q \Phi_i \epsilon_{t-i} + \epsilon_t + \sum_{i=1}^p \Phi_i X_{t-i} + \epsilon_t \quad (3.5)$$

Donde:

$\Phi_1 \dots \Phi_p$  son parámetros del modelo.

$c$  es la constante.

$\epsilon_t$  es el ruido blanco.

$\epsilon_{t-1}$  son los términos de error.

$X_t$  variable relacionada con un periodo de anterior.

Donde p es el parámetro, u orden, que corresponde al AR y q, al MA. Notar que el modelo ARMA es también un proceso estacionario. Es decir que, si el modelo no lo es, es necesario trabajar sobre una serie diferenciada [14].

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p} + \Phi_1 \epsilon_{t-1} + \Phi_2 \epsilon_{t-2} + \dots + \Phi_q \epsilon_{t-q} \quad (3.6)$$

Donde:

$Y_t$	Predicción
$\alpha$	corresponde a una constante.
$\beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p}$	combinación lineal de retraso de Y.
$\Phi_1 \epsilon_{t-1} + \Phi_2 \epsilon_{t-2} + \dots + \Phi_q \epsilon_{t-q}$	combinación lineal de errores de pronóstico retrasado.

### 3.11 Modelos ARIMA (p, d, q)

Al modelo ARMA se añade una componente más, que es la parte integrada, que indica cuántas diferencias estacionarias son necesarias para lograr la estacionalidad [14]. Si no es necesaria ninguna, simplemente es un ARMA. La notación de un ARIMA viene definida por:

$$ARIMA(p, d, q)$$

ARIMA funcionará bien si hay estacionariedad (media y varianza constante), de lo contrario, habrá que transformar los datos para poder usar un ARIMA [14]. Un modelo ARIMA(p,d,q) viene definido como:

Por consiguiente, si debe diferenciar una serie de tiempo d veces para hacerla estacionaria y luego aplicarle el modelo ARMA (p,q), decir que la serie de tiempo original es ARIMA(p, d, q), es decir, es una serie de tiempo autorregresiva integrada de promedios móviles, donde p denota el número de términos autorregresivos, el número de veces que la serie debe diferenciarse para hacerse estacionaria y q el número de términos de promedios móviles [17].

ARIMA funcionará bien si hay estacionariedad, de lo contrario, habrá que transformar los datos para poder usar un ARIMA [14]. Un modelo ARIMA(p,d,q) viene definido como:

$$ARIMA(p, d, q) = (\Delta^d Y_t - Y_t) + \Phi_0 + \sum_{i=1}^q \Phi_i \epsilon_{t-i} + \epsilon_t + \sum_{i=1}^p \Phi_i X_{t-i} + \epsilon_t \quad (3.7)$$

Donde:

$(\Delta^d Y_t - Y_t) + \Phi_0$  parte, “d” diferencias requeridas para hacer la serie en estacionaria.

$\Phi_0$  corresponde a una constante.

$\sum_{i=1}^q \Phi_i \epsilon_{t-i} + \epsilon_t$  parte de media móvil (MA).

$\sum_{i=1}^p \Phi_i X_{t-i} + \epsilon_t$  parte autorregresiva (AR).

### 3.12.1 Dickey-Fuller aumentado (ADF)

El Test de Dickey Fuller Aumentada es una forma para determinar si existen raíces unitarias en las series de tiempo y por tanto comprobar si hay o no estacionariedad.

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \sum_{i=1}^s \alpha_i \Delta Y_{t-s} + \epsilon_t \quad (3.13)$$

Donde:

$\alpha$  constante

$Y_{t-1}$  primera diferencia de la serie.

$\epsilon_t$  termino de error de ruido blanco.

$\sum_{i=1}^s \alpha_i \Delta Y_{t-s} + \epsilon_t$  términos incrementales retardos de las variables.

## 3.13 Metodología modelo ARIMA

Esta metodología se basa en tratar de determinar cuál es el modelo probabilístico que rige el comportamiento del proceso a lo largo del tiempo. Cabe apuntar en este momento la diferencia entre proceso y modelo. Se puede decir que un proceso es lo real, es decir, el fenómeno en sí, del cual se desconoce su mecanismo generador. Por el otro lado un modelo es solo la imitación o representación del proceso [16].

### 3.13.1 Primera fase

Consiste en identificar el posible modelo ARIMA que sigue la serie, lo que requiere:

- Decidir qué transformaciones aplicar para convertir la serie observada en una serie estacionaria.
- Determinar un modelo ARMA para la serie estacionaria, es decir, los órdenes p y q de su estructura autorregresiva y de media móvil.



**3.13.2 Segunda fase**

Seleccionado provisionalmente un modelo para la serie estacionaria, se pasa a la segunda etapa de estimación, donde los parámetros AR y MA del modelo se estiman por máxima verosimilitud y se obtienen sus errores estándar y los residuos del modelo [16].

**3.13.3 Tercera fase**

Es el diagnóstico, donde se comprueba que los residuos no tienen estructura de dependencia y siguen un proceso de ruido blanco. Si los residuos muestran estructura se modifica el modelo para incorporarla y se repiten las etapas anteriores hasta obtener un modelo adecuada [16].

**3.13.4 Cuarta fase**

Es la predicción, una vez que se ha obtenido un modelo adecuado se realizan predicciones con el mismo análisis de series temporales es un conjunto de técnicas estadísticas que permite describir y prever el comportamiento de una serie temporal y modelizar el proceso estocástico del que estas provienen con el objetivo de hacer predicciones. La familia de modelos ARIMA es ampliamente utilizada y presenta buenos resultados para horizontes de predicción cercanos en el tiempo de series temporales que presentan comportamientos estacionales [16].

**3.14 Modelo estacional autorregresivo integrado de medias móviles (SARIMA)**

Este modelo es usado cuando la serie temporal tiene estacionalidad. La notación es similar a la del ARIMA:

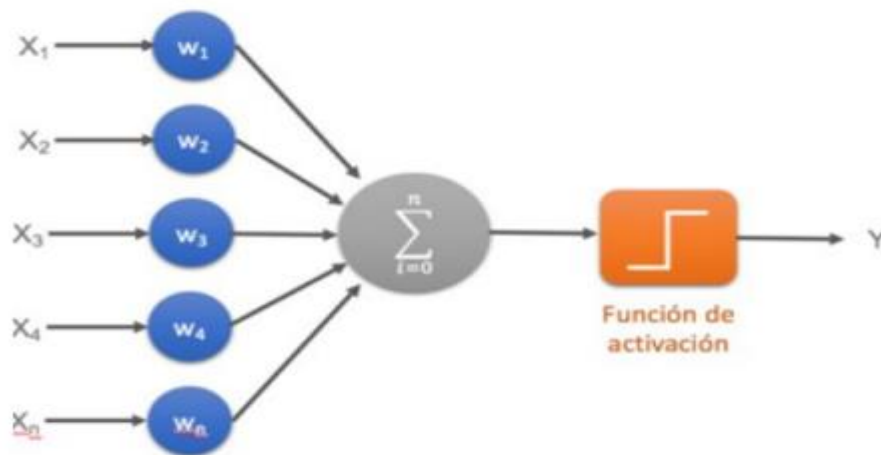
$$ARIMA(p, d, q) (P, D, Q)m$$

Los parámetros p, d, q corresponden a AR, a la parte integrada y a MA, respectivamente.

Los parámetros P, D, Q simplemente representan a (p, d, q) pero para la parte estacional de la serie temporal. El parámetro m indica el número de períodos en cada estacionalidad [14].

### 3.15 REDES Neuronales Artificiales

Las Redes Neuronales Artificiales (RNA) son dirigidos a partir de los datos, es decir, son capaces de encontrar relaciones (patrones) de forma inductiva por medio de los algoritmos de aprendizaje Figura 3.16, basado en los datos existentes más que requerir la ayuda de un modelador para especificar la forma funcional y sus interacciones. Las RNA son un método de resolver problemas, de forma individual o combinadas con otros métodos, para aquellas tareas de clasificación, identificación, diagnóstico, optimización o predicción en las que el balance datos/conocimiento se inclina hacia los datos y donde, adicionalmente, puede haber la necesidad de aprendizaje en tiempo de ejecución y de cierta tolerancia a fallos [18].



**Figura 3. 16.** Arquitectura de una neurona artificial [19].

#### 3.15.1 Rol de una neurona

El corazón de una red neuronal artificial es un nodo matemático o neurona. Es el elemento principal de todo el proceso. Las neuronas de la capa de entrada reciben la información, la cual la procesan mediante funciones matemáticas para luego distribuirla a las neuronas de la capa oculta. Esta información se transforma en las capas intermedias hasta finalmente llegar a las neuronas de salida.

La clave es que la información es procesada mediante una función de activación, una vez hecho esto se tiene el resultado ponderado y distribuido en las neuronas de la siguiente capa, por lo que las neuronas se activan a sí mismas con los pesos ponderados. Esto hace que la fuerza con la que dos neuronas están conectadas depende del peso de la información que se envía.

### 3.16 Modelo matemático de la red neuronal artificial

Para el planteamiento de los primeros modelos de red neuronal artificial se basaron en el modelo de tipo McCulloch-Pitts, propuesto por Warren McCulloch y Walter Pitts en 1943. Donde se basaron en el funcionamiento de las neuronas biológicas el modelo es considerado como un procesador elemental que está compuesta por  $X_n$  entradas y una única salida  $y_i$ . El ingreso total hacia la neurona es calculado como la suma ponderada de todas las entradas ponderadas compuesto por ciertos valores. Para lograr obtener la ecuación matemática de la neurona presentada en la Figura 3.16, presenta un grupo de entradas  $x_1 \dots x_n$ , comúnmente se conoce como las sinapsis de una neurona biológica cada una de ellas se encuentran multiplicadas por un peso al que está asociado  $w_1 \dots w_n$  [20].

En la Figura 3.16, se puede apreciar el símbolo matemático ( $\Sigma$ ) este es el resultado de la sumatoria ponderada de todas las entradas y como resultado se obtiene una salida como se muestra en la ecuación 3.4.

$$\sum_{i=1}^n w_i x_i + w_n \tag{3.8}$$

Donde:

$x_1 \dots x_n$  son las entradas.

$w_1 \dots w_n$  son los pesos simpáticos.

Todas las señales que ingresan a la neurona pueden ser vistas como una única entrada si se describe utilizando la ecuación.

$$u = \sum_{i=1}^n w_{ij} x_i + \theta_j \tag{3.9}$$

Donde:

$u$  única entrada.

$\theta_j$  umbral o sesgo.

$\sum_{i=1}^n w_{ij} x_i + \theta_j$  sumatoria de todas las entradas.

Dependiendo del signo del peso se habla de una excitación para el caso positivo y de una inhibición para el caso negativo. Si se añade la variable tiempo a la ecuación 3.6 esta se reescribe como lo muestra la ecuación 3.6.

$$u(t) = \sum_{i=1}^n w_{ij}x_i + \theta_j \quad (3.10)$$

Donde:

$u$                       única entrada.

$\theta_j$                     umbral o sesgo.

$\sum_{i=1}^n w_{ij}x_i + \theta_j$     sumatoria de todas las entradas.

$t$                         tiempo.

Los pesos son los encargados de guardar el conocimiento de la red neuronal sobre cierta tarea en cuestión y al proceso mediante el cual se ajustan estos pesos se denomina aprendizaje o entrenamiento. Por lo general el sesgo  $\theta$  se lo considera como otro peso más conectado a una entrada que siempre tiene el valor de uno es por eso que la salida de cada neurona se la puede reescribir como la ecuación 3.7.

$$y = f(u) \quad (3.11)$$

Donde:

$u$                       única entrada

$f$                       función de activación

$y$                       salida de la neurona

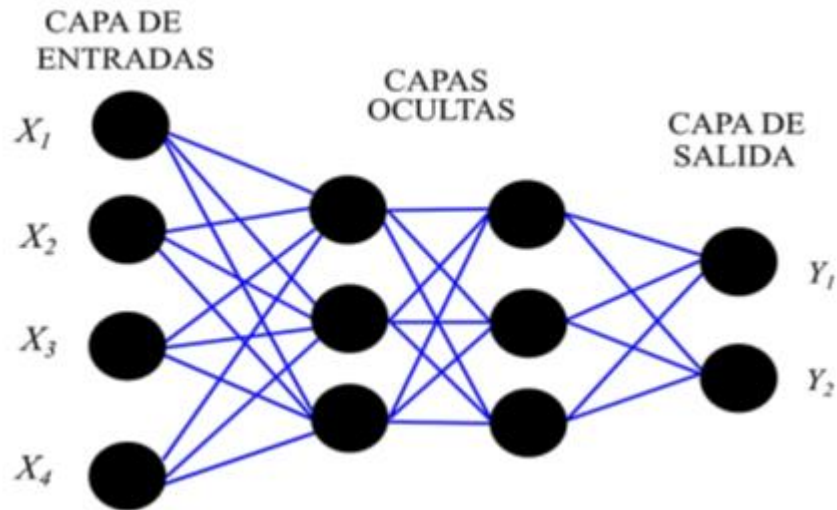


Figura 3. 17. Esquema de una red neuronal artificial [20].

La Figura 3.17, muestra como las redes neuronales artificiales poseen una estructura redundante por lo que destacan por su elevada capacidad de producir salidas correctas para entradas no vistas en el entrenamiento [20].

### 3.17 Células LSTM

La red neuronal recurrente Long Short Term Memory es semejante a las redes de Elman y de Jordan con la diferencia que sustituye las neuronas y las capas ocultas por un bloque de memoria, este bloque se denomina célula LSTM. En la siguiente Figura 3.18 se puede apreciar cómo es una célula LSTM en su interior [21].

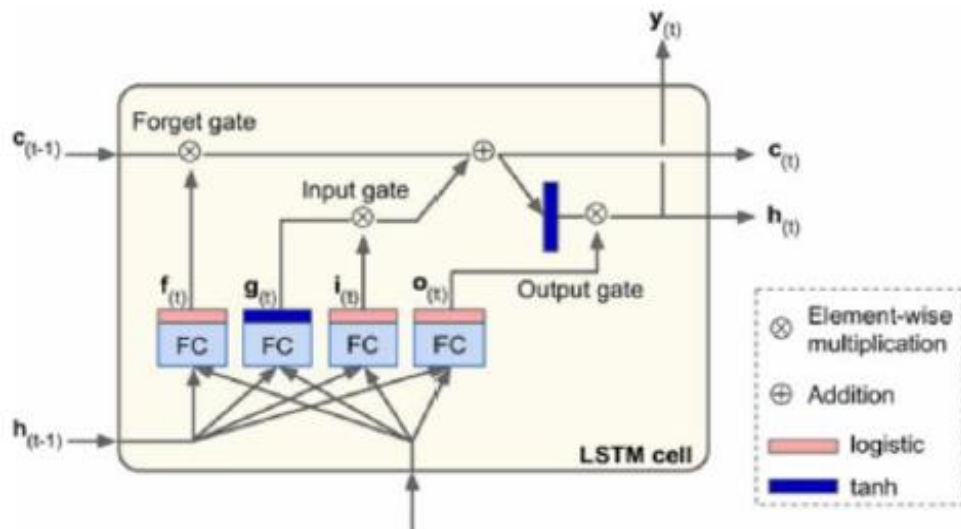


Figura 3. 18. Estructura interna de una célula LSTM [21].

Como se puede observar en la Figura 3.18, el interior del bloque que representa una célula LSTM la cual es similar a una célula normal con la diferencia que su estado ahora se divide en dos vectores,  $h(t)$  y  $c(t)$  (la letra  $c$  viene de “célula”). El término  $h(t)$  hace referencia al estado de corto plazo y  $c(t)$  al de largo plazo. La idea principal de la célula es que la red logre aprender qué información almacenar en el estado de largo plazo, qué desechar y qué leer de ellos. Si se logra fijarse bien en el término de largo plazo  $c(t-1)$  atraviesa la red de izquierda a derecha, primero atravesando por una puerta de olvido (forget gate), de esta forma se deshace de algunos datos almacenados, y luego añade nueva información que ingresa por la puerta de entrada (input gate) y el resultado  $c(t)$  se envía directamente fuera de la célula dejando la información intacta y esto servirá de entrada para la siguiente época [21].

### 3.18 Medidas de centralización

Son los valores centrales de los datos obtenidos. Las más usuales son: la media, la moda y la mediana.

#### 3.18.1 La media aritmética

Es el valor medio ponderado de la serie de datos o valores que toma la variable estadística ecuación (3.31). La media no es más que la suma de todos los valores de una variable dividida entre el número total de datos de los que se dispone [22]

Y se calcula como

$$\bar{x} = \frac{\sum_{i=1}^k x_i n_i}{N} = \frac{x_1 n_1 + \dots + x_k n_k}{N} \quad (3.12)$$

### 3.18.2 Mediana

Es uno de los cálculos más representativos de la muestra. La mediana es el valor del elemento intermedio cuando todos los elementos se ordenan. La mediana se calcula ordenando los datos de menor a mayor y tomando el valor del medio que es el que deja un 50% de observaciones a su izquierda y un 50% a su derecha [22].

## 3.19 CRITERIOS PARA LA VALIDACIÓN DE LOS RESULTADOS APLICANDO UN MÉTODO DE ANÁLISIS.

### 3.19.1 Error medio cuadrático (MSE – mean square error)

Este error representa el promedio de todas las diferencias entre la salida obtenida y la salida que se deseada obtener [23]. Esta medida se penaliza los errores mayores en la predicción, debido a que cada error es elevado al cuadrado. La predicción es mejor cuando más pequeño es el valor del MSE [24].

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2 \quad (3.13)$$

### 3.19.2 Raíz del Error Medio Cuadrático (Root Mean Squared Error – RMSE)

Este error es simplemente el resultado de extraer la raíz del Error Medio Absoluto (MSE) que se menciona a continuación.

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2} \quad (3.14)$$

**3.19.3 Error Medio Absoluto (Mean Absolute Error MAE)**

Este error está representado por el promedio del valor absoluto de la diferencia entre la salida de los datos reales y la salida de los datos pronosticados.

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}| \quad (3.15)$$

**3.19.4 Error Medio Absoluto Porcentual (Mean Absolute Percentage Error - MAPE)**

Este error está relacionado con el promedio de los errores absolutos el resultado de este error está expresado en términos de porcentaje [23].

$$MAPE = \sum_{i=1}^N \left| \frac{y_i - \hat{y}}{y_i} \right| * 100\% \quad (3.13)$$

**3.19.5 Error residual (EMM)**

El error estándar múltiple, mide la dispersión de los datos obtenidos en la predicción [45]. Su cálculo está representado por la siguiente ecuación:

$$EMM = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y})^2}{(N - k - 1)}} \quad (3.17)$$

Donde:

$y_i$  valor histórico real de la variable independiente

$\hat{y}$  Valor estimado de la variable independiente

$N$  Cantidad de datos utilizados

$k$  Número de variables independientes



### **3.19.6 Desviación media absoluta (MAD)**

La desviación absoluta media es una extensión simple de la varianza absoluta. Suma las varianzas absolutas y divide el resultado entre el número de registros. La desviación absoluta media es una estadística de error que calcula el promedio de la distancia ente cada par de puntos de datos reales y ajustados [25].

### **3.19.7 El criterio de información de Akaike (AIC)**

Compara la calidad de un conjunto de modelos estadísticos entre sí una vez que haya seleccionado el mejor modelo, considere ejecutar una prueba de hipótesis para averiguar la relación entre las variables de su modelo y el resultado de interés [26].

### **3.19.8 Desviación estándar**

La desviación estándar se interpreta como una medida de incertidumbre. La desviación estándar de un grupo repetido de medidas que da la precisión de éstas. Cuando se va a determinar si un grupo de medidas está de acuerdo con el modelo teórico, la desviación estándar de esas medidas es de vital importancia: si la media de las medidas está demasiado alejada de la predicción (con la distancia medida en desviaciones estándar), entonces se considera que las medidas contradicen la teoría.

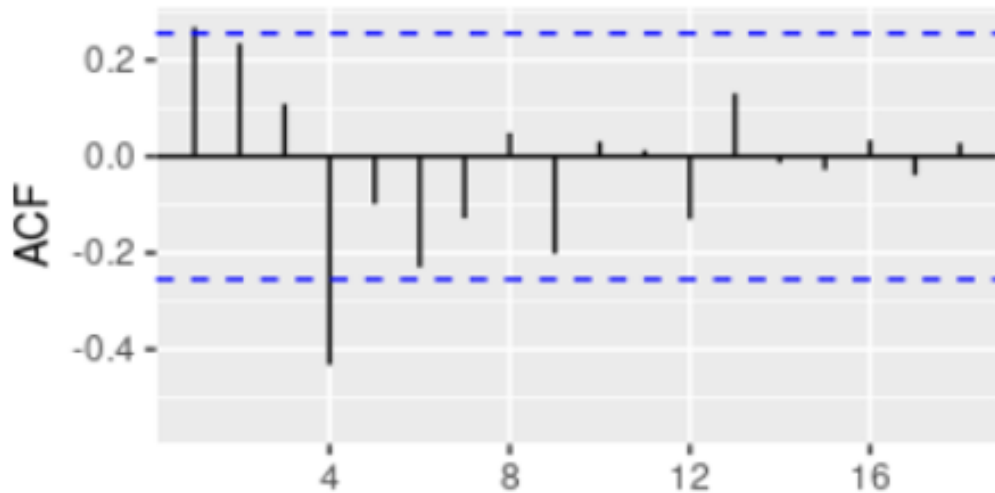
Esto es coherente, ya que las mediciones caen fuera del rango de valores en el cual sería razonable esperar que ocurrieran si el modelo teórico fuera correcto. La desviación estándar es uno de tres parámetros de ubicación central; muestra la agrupación de los datos alrededor de un valor central (la media o promedio) [27].

### **3.19.9 Varianza**

El análisis de la varianza [28] es un método estadístico para determinar si diversos conjuntos de muestras aleatorias de una determinada variable proceden de la misma población o de poblaciones distintas. En general, cada conjunto muestral se caracteriza por estar afectado por un tratamiento específico, que eventualmente puede influir en los valores que tome la variable objeto de estudio.

### 3.19.10 Función de correlación automática (ACF)

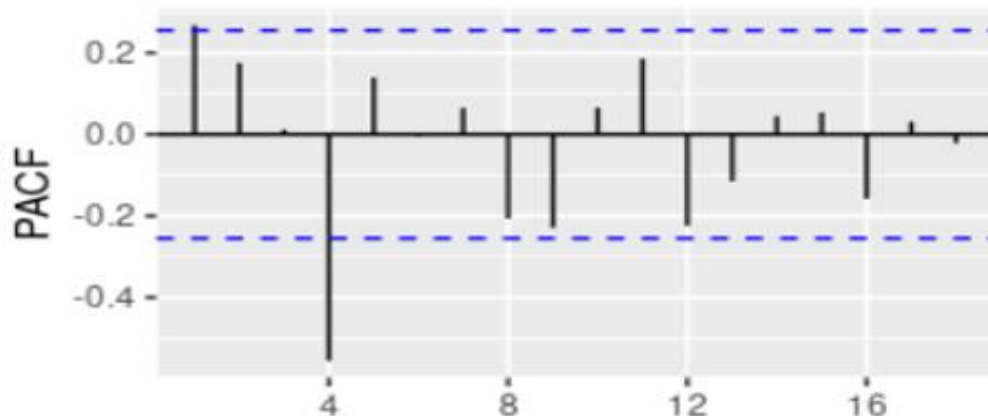
Es una función de autocorrelación completa puede proporcionarnos el valor de autocorrelación de cualquier serie con valor de retraso, describe el grado de correlación entre el valor actual de la secuencia y su valor pasado Figura 3.25. Las series de tiempo pueden incluir tendencias, estacionalidad, periódico y residuos [29].



**Figura 3.20.** Gráfica de función de correlación automática [30].

### 3.19.11 Función de autocorrelación parcial (PACF)

Es una función de autocorrelación completa puede proporcionarnos el valor de autocorrelación de cualquier serie con valor de retraso, describe el grado de correlación entre el valor actual de la secuencia y su valor pasado Figura 3.26. Las series de tiempo pueden incluir tendencias, estacionalidad, periódico y residuos [29].



**Figura 3.21.** Gráfica de función de autocorrelación parcial [30].

## **3.20 Lenguaje de programación Python**

El lenguaje de programación Python es el lenguaje elegido para muchas aplicaciones en el análisis de datos avanzados por una buena razón. Es uno de los lenguajes de programación avanzada accesible para todo tipo de programación. Python es una forma sencilla de realizar una gran variedad de investigaciones relacionadas con el análisis de datos [31]

### **3.20.1 STATSMODELS en Python**

Statsmodels es un módulo de Python que proporciona clases y funciones para la estimación de muchos modelos estadísticos diferentes, así como para realizar pruebas estadísticas y exploración de datos estadísticos. Hay disponible una extensa lista de estadísticas de resultados para cada estimador. Los resultados se comparan con los paquetes estadísticos existentes para garantizar que sean correctos [29].

## **3.21 Librerías necesarias para el uso del software de Python**

### **3.21.1 Librería Numpy**

La librería Numpy es el paquete de Python principal para la informática científica. Donde proporciona estructuras de datos (realizando matrices y matrices multidimensionales) en la que se tiene varios elementos que están almacenado es sus respectivas ubicaciones de memoria, y con los que se pueden trabajar con todas las operaciones que puedan ser involucradas [32].

### **3.21.2 Librería Pandas**

Pandas es una librería para facilitar la manipulación y el análisis de los datos. Consta de dos estructuras de datos: series y dataframes. Está implementada sobre la librería NumPy. Algunas de las muchas características que tiene es de agrupar datos, filtrarlos, convertirlos, unirlos o usar series temporales [33].

### **3.21.3 Librería Matplotlib**

Corresponde a un módulo en Python que además de crear gráficos, histogramas, visualizaciones de líneas, barras y diagramas de dispersión, entre otros, puede hacer visualizaciones en 2D o 3D dependiendo su aplicación. Es la librería más utilizada en Python, para ciencia de datos. Su éxito está basado en la facilidad que brinda a los desarrolladores, a la hora de diseñar visualizaciones con datos a partir de muy pocas líneas de código, mezclada con la posibilidad de incluir esos gráficos en cualquier proyecto web [34].

### **3.21.4 Librería Scikit-Learn**

Scikit Learn es una librería de código abierto, proporciona una gama de algoritmos de aprendizaje supervisados y no supervisados que implementa una gran variedad de algoritmos de aprendizaje automático (regresión, clasificación, agrupación). Esta librería está construida sobre SciPy e incluye soporte para varias librerías o paquetes, como Numpy, Pandas, SciPy, Matplotlib, IPy, SymPy, etc [32].

### **3.21.5 Librería Keras**

Esta librería dispone de una frontend para la modelización de sistemas neuronales y un backend computacional para el entrenamiento de éstos. La modelización es el punto fuerte de esta librería, ya que ofrece una API sencilla y modular para la definición de sistemas neuronales. Es interesante observar la evolución y supervivencia de esta librería gracias a su simplicidad a la hora de definir modelos y cómo se ha convertido en parte fundamental de otras librerías para la definición de modelos de alto nivel.

### **3.21.6 Librería Tensorflow**

Es una librería de código abierto para el aprendizaje automático desarrollada por Google. La librería Tensorflow es un paquete de Python para aprendizaje automático. El nombre de la librería hace referencia a cómo ésta procesa la información, ya que un tensor es una matriz de números y el flujo de éstas son las operaciones que se realizan para transformarlas [35]. Algunas características de esta librería se mencionan a continuación:

- Fácil construcción de modelos. - Se cuenta con múltiples niveles de abstracción para que pueda elegir el más adecuado para sus necesidades. Permite crear y entrenar modelos utilizando la API de keras de alto nivel.
- Robusta producción de aprendizaje automático. - Permite entrenar e implementar su modelo fácilmente, sin importar el idioma o la plataforma que utilice.
- Potente experimentación para la investigación. – Permite construir y entrenar modelos de última generación sin sacrificar velocidad o rendimiento. TensorFlow cuenta con un ecosistema de potentes bibliotecas [19].

### **3.21.7 Uso de variables exógenas: ARMAX, ARIMAX y SARIMAX**

Una variable exógena es aquella variable que no está afectada por otras variables en un sistema. Exo significa “fuera”. Por otro lado, una variable endógena es aquella que sí se ve influenciada por otros factores de un sistema.

Usando pues un modelo ARMA con variables exógenas (además de las endógenas), se habla de un ARMAX; si se usa un ARIMA, éste será entonces un ARIMAX, finalmente, si se emplea un SARIMA, con variables exógenas se obtendrá un SARIMAX [14].

## **4 METODOLOGÍA**

Una vez comprendido la relación existente entre el método ARIMA y los datos de la central ILLUCHI II. Esta sección tiene como objetivo especificar las fases que trae consigo la construcción del método con los datos de producción que se emplea en el presente documento. Analizando del modelo construido con sus respectivas gráficas de pronóstico con los datos reales y de esta manera solucionar el problema planteado.

### **4.1 TIPO DE INVESTIGACIÓN**

Este proyecto de investigación refleja la condición actual del sistema de generación de la central hidroeléctrica ILLUCHI II de la cual se toma los valores diarios, mensuales y anuales conjuntamente con la predicción a mediano plazo de la producción de energía eléctrica, para la cual se ocupa un orden de herramientas para analizar el comportamiento de los datos reales con los datos pronosticados.

#### **4.1.1 Investigación descriptiva**

Para la investigación se considera las características, comportamiento por ende se considera una investigación descriptiva, se debe conocer sus históricos de la producción de la central ILLUCHI II. Consiguiente se analiza los datos que servirán para la investigación.

#### **4.1.2 Investigación diagnóstica**

Esta forma de investigación ofrece acceso a las características del problema planteado mediante datos de campo, ayudando a comprender la situación actual por la cual se encuentra sometido la central hidroeléctrica ILLUCHI II, en los sistemas de generación, identificando los factores analizar para adentrar más al problema que ha planteado, ayudando a obtener un análisis mediante los datos históricos para obtener una alternativa que brinde una solución al problema de la producción de la energía eléctrica.

#### **4.1.3 Investigación aplicada**

En esta parte de la investigación se pondrá en práctica todos los conocimientos técnicos y científicos que se han adquirido, mediante uso del programa computacional Python que ayuda a

desarrollar los análisis necesarios y la programación de un algoritmo que permita realizar el pronóstico de la producción de energía eléctrica para obtener resultados suficientemente aproximados a los valores reales.

## **4.2 MÉTODO DE INVESTIGACIÓN**

### **4.2.1 Metodología de la investigación bibliográfica**

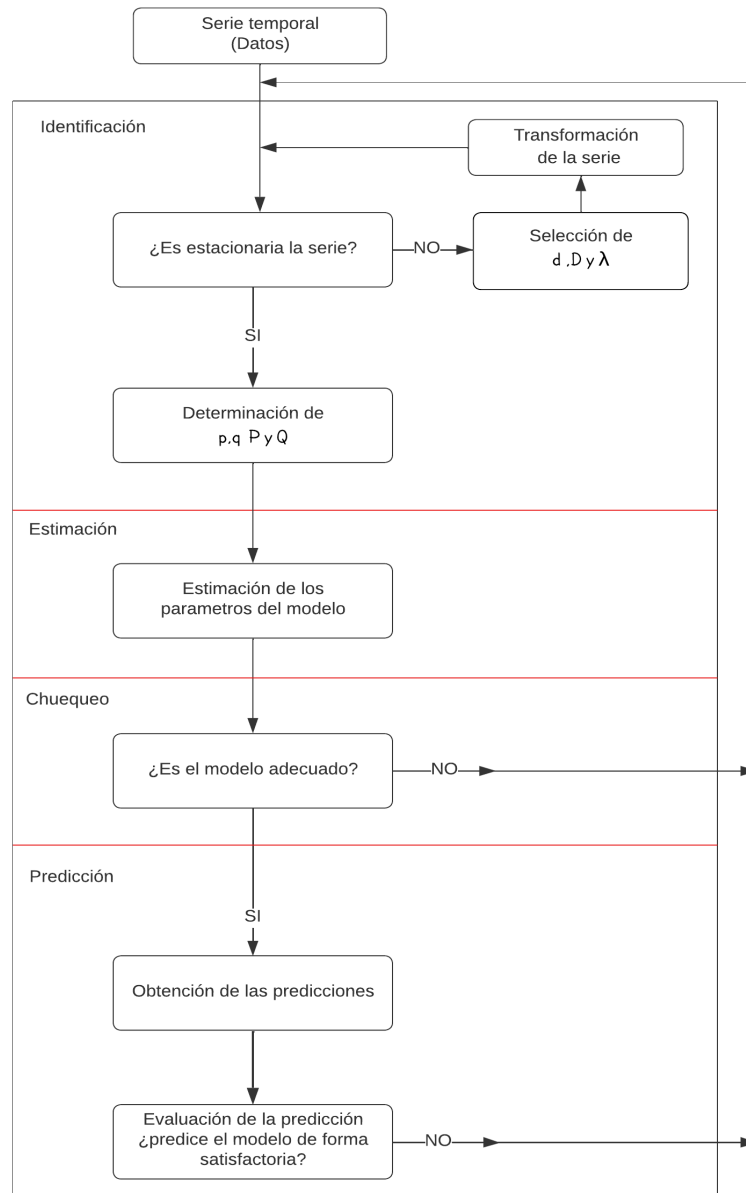
La investigación bibliográfica ayuda con el desarrollo del análisis, el cual se basa en indagar en documentos físicos o digitales como: artículos científicos, libros, trabajos de fin de grado y publicaciones en internet, que estén relacionados con la producción de energía eléctrica, la aplicación de redes neuronales artificiales e información necesaria sobre el programa computacional Python. Lo que se ha proporcionado a la formulación de teorías, hipótesis, análisis y resultados, este es un procedimiento necesario para lograr solucionar el problema planteado [36].

### **4.2.2 Inductivo – deductivo**

Este método de investigación examina de forma descriptiva todos los temas y subtemas que se emplearon en el desarrollo del presente trabajo de investigación, este procedimiento ayuda a formular las conclusiones y recomendaciones que surgieron a través de todo el proceso de investigación. En relación a los métodos aplicados se analizó cada uno de las metodologías existentes para la predicción de la producción de energía eléctrica, a partir del análisis de investigaciones desarrolladas en diferentes áreas, se desarrolló la implementación de un algoritmo que permitió obtener los datos pronosticados, de esta forma elaborar nuestros propios criterios para recomendar trabajos futuros que pueden ser elaborados y hasta mejorarlos en este tipo de investigación.

### 4.3 TÉCNICAS E INSTRUMENTOS

#### 4.3.1 Pasos a seguir mediante un diagrama de flujo para realizar la predicción de la producción eléctrica de la central eléctrica ILLUCHI II mediante el método ARIMA.



**Figura 4.1.** Flujograma metodológico empleado para la predicción de la producción de la energía eléctrica utilizando el modelo ARIMA.



El flujograma presenta los pasos a seguir para la predicción de la producción mediante el método ARIMA en el software Python:

**Paso 1.-** Como primer paso, se debe comprender el funcionamiento de la central ILLUCHI II, se procede a buscar información del funcionamiento de la central diariamente, mensualmente y anualmente. Con ello se obtiene un histórico de la base de datos que se utilizara para la realización del pronóstico de la producción consumida.



**Figura 4.2.** Ubicación del Tanque de almacenamiento central de generación ILLUCHI II [37].

La central hidroeléctrica ILLUCHI II cuenta con diferentes características con respecto a la central hidroeléctrica ILLUCHI I, debido a que es una central de paso. Esta central fue construida en el año 1984 por lo tanto sus equipos tienen mayor vida útil, y hasta la fecha se encuentra funcionando de manera adecuada. Esta central utiliza el agua turbinada que sale de la central hidroeléctrica

ILLUCHI I para producir la energía eléctrica; esto con el fin de aprovechar la energía potencial del agua tal como se presenta en las Figuras 4.3 y 4.4 [36].

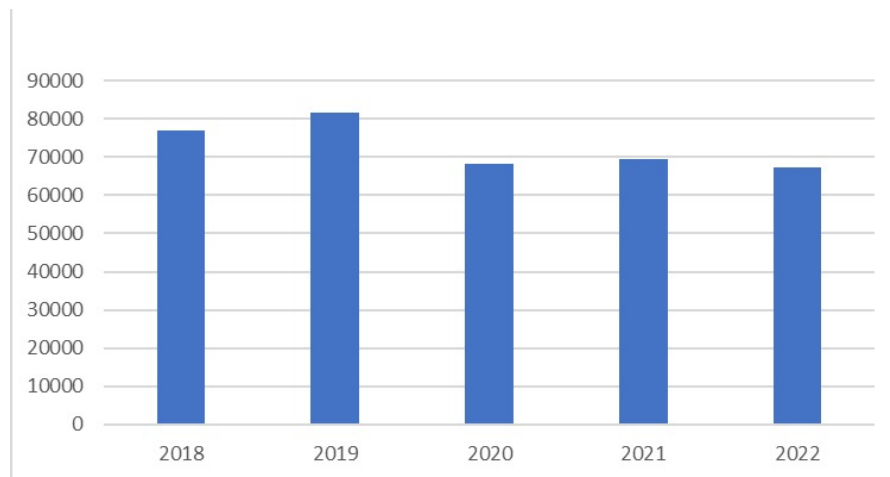


**Figura 4.3.** Tubería de presión perteneciente a la central hidroeléctrica ILLUCHI II.



**Figura 4.4.** Tubería de presión perteneciente a la central hidroeléctrica ILLUCHI II [37].

Es fundamental para el correcto desarrollo de la investigación obtener una base de datos reales, desde la cual se procede al empleo del modelo ARIMA para predicción. Como se observa en la Figura 4.5 el comportamiento de la predicción anual desde al año 2018 hasta el año 2022. Se realizo desde al año 2018 para que el estudio sea a medio plazo y de igual manera la predicción de 5 años.



**Figura 4.5.** Producción media diaria anual de la base de datos.

**Paso 2.-** En este paso se realiza el llamado de las librerías que se va a ocupar en la realización del modelo ARIMA para la predicción que brinda Python. En esta sección se emplea las librerías tanto para la manipulación y tratamiento de datos, la visualización, la modelación y las gráficas.

```

# Manipulación y tratamiento de Datos
import numpy as np
import pandas as pd

# Visualización de datos
import plotly.express as px
import matplotlib.pyplot as plt
%matplotlib inline
plt.style.use('ggplot')

# Modelación Arima
from statsmodels.tsa.statespace.sarimax import SARIMAX
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
from statsmodels.tsa.seasonal import seasonal_decompose
from statsmodels.tsa.stattools import adfuller

# Métrica de Evaluación
from sklearn.metrics import mean_squared_error
from statsmodels.tools.eval_measures import rmse
from sklearn import metrics

# No presentar advertencias
import warnings
warnings.filterwarnings("ignore")

import seaborn as sns
sns.set()

from datetime import datetime
from datetime import timedelta
from datetime import date

from plotly import __version__
from plotly.offline import download_plotlyjs, init_notebook_mode, plot, iplot
import plotly.graph_objects as go
from plotly.subplots import make_subplots
import cufflinks as cf
init_notebook_mode(connected=True)
cf.go_offline()
import plotly.express as px
import chart_studio.plotly as py

pd.set_option('display.max_columns', None)

```

**Figura 4.6.** Librerías empleadas en Python.

**Paso 3.-** En este paso se selecciona los datos para generar nuestra data para programación en el software y esto se realiza en base al paso 1, se plantea seleccionar las variables de entradas las cuales influyan en la variación de la producción de energía eléctrica en la central hidroeléctrica ILLUCHI II y permita un correcto pronóstico.

Para la selección de las variables de entrada que se presenta en la tabla 4.1, utilizados para la predicción de la producción de la central se ocupa la base de datos proporcionada por personal de ELEPCO S.A, en una base de datos en Excel desde el año 2018 hasta el año 2022.

**Tabla 4.1** Selección preliminar de las variables para la predicción de la producción de energía eléctrica.

	<b>Descripción</b>	<b>Unidad de medida</b>
<b>Año</b>	Tiempo de consumo de energía eléctrica	<b>Año</b>
<b>Mes</b>	Tiempo de consumo de energía eléctrica	<b>Mes</b>
<b>Día</b>	Tiempo de consumo de energía eléctrica	<b>Día</b>
<b>Energía consumida</b>	Energía activa consumida en Kilovatio Hora	<b>KW-H</b>

**Paso 4.-** en el tercer paso se procede a validar y procesar nuestra base desde nuestra data y consiguiente se nombra como df y realiza el llamado en Python mediante la librería pd.read\_excel tal como se muestra en la Figura 4.7.

```
df= pd.read_excel('DATA/df_kwh-total.xlsx')
df
```

	<b>FECHA</b>	<b>PRODUCCIÓN KW-H</b>
<b>0</b>	2018-01-01	101400.0
<b>1</b>	2018-01-02	102200.0
<b>2</b>	2018-01-03	85400.0
<b>3</b>	2018-01-04	67500.0
<b>4</b>	2018-01-05	67000.0
<b>...</b>	<b>...</b>	<b>...</b>
<b>1822</b>	2022-12-28	78000.0
<b>1823</b>	2022-12-29	80400.0
<b>1824</b>	2022-12-30	73200.0
<b>1825</b>	2022-12-31	64800.0
<b>1826</b>	2023-01-01	64800.0

**Figura 4.7.** Validación y preprocesamiento de la información procedente de una base de datos.

En la Figura 4.8, se observa cuantos datos tiene la base de datos, el peso, los datos de fechas, y los datos de objeto con el fin de obtener una data time en Python y luego se realiza la comprobación de no tener (nans) los cuales vienen a ser los espacios en blanco.

```
In [4]: df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1827 entries, 0 to 1826
Data columns (total 2 columns):
# Column Non-Null Count Dtype
---
0 Date 1827 non-null datetime64[ns]
1 kw 1827 non-null float64
dtypes: datetime64[ns](1), float64(1)
memory usage: 28.7 KB

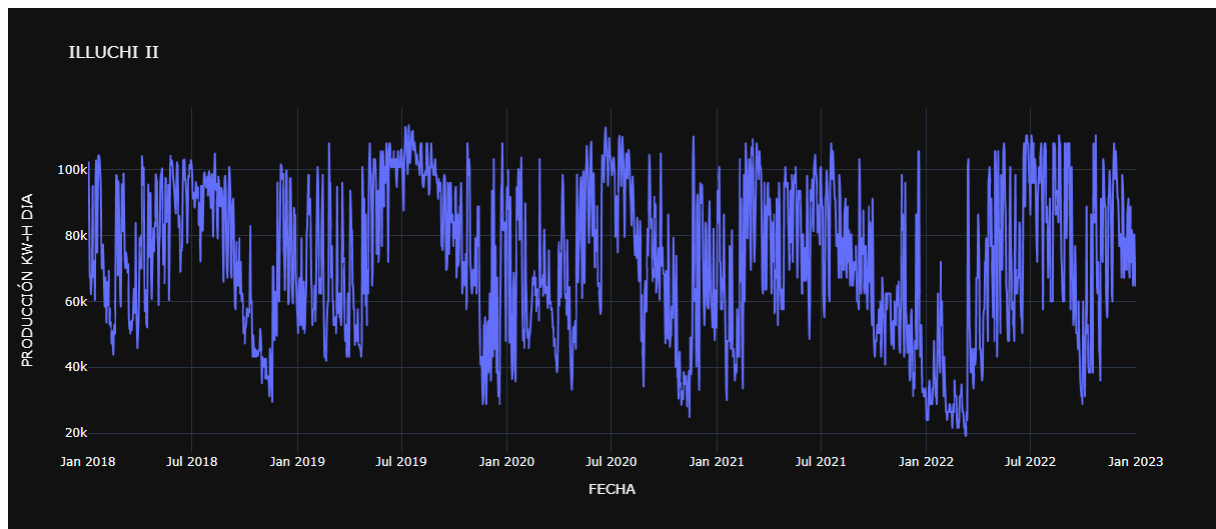
In [5]: #Verificación de nans
df.isnull().values.any()
df.isnull().sum().sum()

df.isnull().sum()

Out[5]: Date 0
kw 0
dtype: int64
```

**Figura 4.8.** Información de nuestro archivo data.

**Paso 5.-** En el cuarto paso se gráfica los datos reales de la base de dato ingresada anteriormente. En la Figura 4.9, comparando la media móvil simple de 12 meses junto con la serie se realiza para determinar la tendencia.



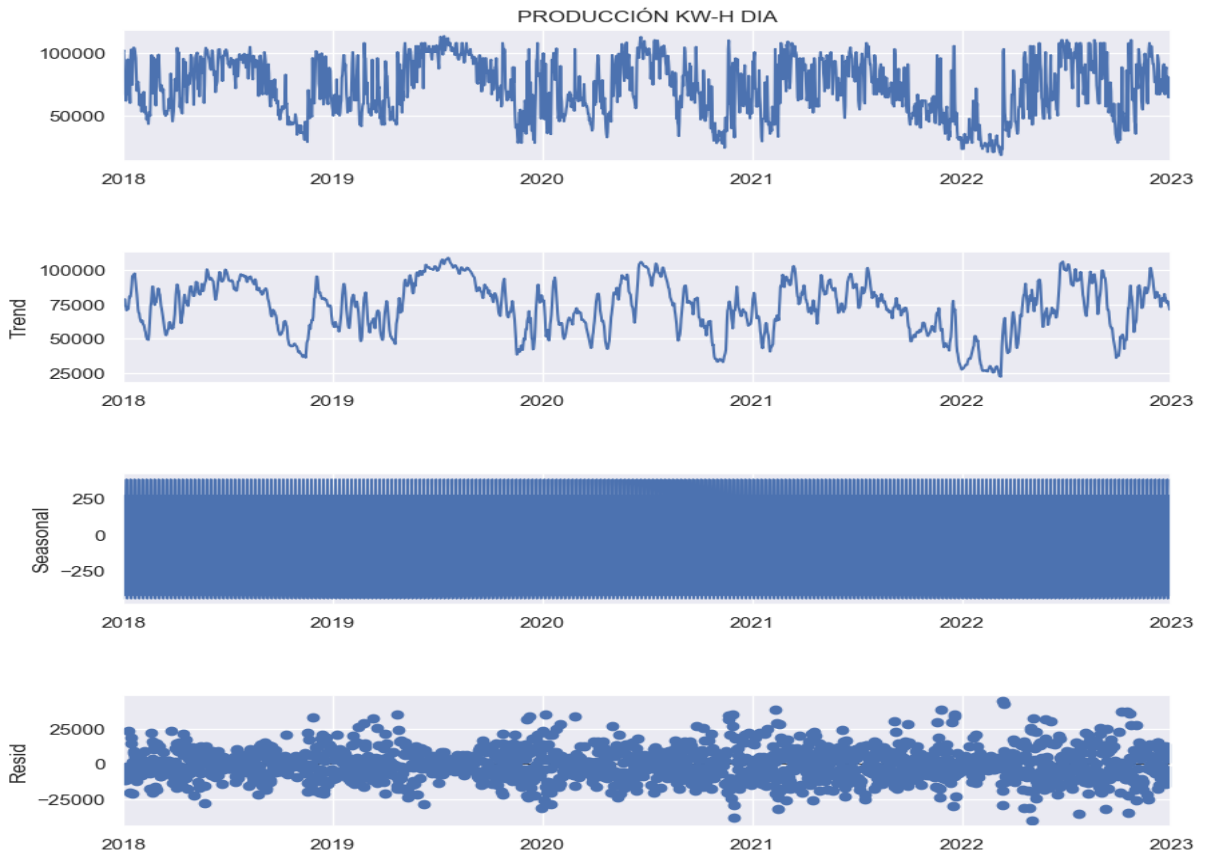
**Figura 4.9.** Gráfica de los datos reales para determinar la tendencia.

Es necesario trazar la desviación estándar para ver si hay alguna varianza o no, como se muestra en la Figura 4.10, que en base a nuestros datos se observa los 12 meses, la potencia y la desviación estándar. Como se observa en la Figura 4.10, la desviación estándar no varía mucho, por lo que no hay mucha varianza.



**Figura 4.10.** Gráfica de los datos reales para determinar la tendencia.

**Paso 6.-** En este paso se inicia con lo que es la modelación ARIMA se centra en descomponer los datos mediante la librería `seasonal_decompose`, como se observa en la Figura 4.11.



**Figura 4.11.** Gráfica de la descomposición de los datos.

La Figura 4.11 muestra los datos observados de la base de datos, (Trend), muestra el movimiento general hacia arriba o hacia abajo de los puntos de datos, en el tercer (Seasonal) es el patrón mensual, anual de los puntos de los datos, por último, los (Resid) que vienen a ser los residuos de la potencia durante los 5 años de investigación.

Una vez descompuestos los datos se realiza el llamado de la librería SARIMAX que viene a ser (El promedio móvil integrado autorregresivo estacional con factores exógenos) el cual es una versión actual del modelo ARIMA, por ende, SARIMAX es un modelo equivalente estacional como SARIMA y Auto ARIMA. Esta característica del modelo difiere de otros modelos juntamente con otras librerías que ayudan a ejecutar los modelos estadísticos con el propósito de analizar los datos y ajustar el modelo con sus coeficientes como se observa en la Figura 4.12.

```
In [11]: import statsmodels.api as sm
         from sklearn.linear_model import LinearRegression
         from sklearn.metrics import mean_squared_error, r2_score

         from statsmodels.tsa.statespace.sarimax import SARIMAX
         from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
         from statsmodels.tsa.stattools import adfuller
```

**Figura 4.12.** Librería empleada para el modelo.

**Paso 7.-** En este paso se emplea la prueba de DICKEY-FULLER en donde se Prueba si la serie temporal es estacionaria mediante la prueba ADF. Es decir, un proceso en donde valores de la estadística de prueba ADF son negativos. Los valores más bajos de ADF indican un rechazo más fuerte de la hipótesis nula en una muestra de serie de tiempo.

```
def Prueba_Dickey_Fuller(series , column_name):
    print (f'Resultados de la prueba de Dickey-Fuller para columna: {column_name}')
    dftest = adfuller(series, autolag='AIC')
    dfoutput = pd.Series(dftest[0:4], index=["Test Estadístico", 'Valor P', 'Sin retrasos', 'Número de observaciones utilizadas'])
    for key,value in dftest[4].items():
        dfoutput['Valor crítico (%s)'%key] = value
    print (dfoutput)
    if dftest[1] <= 0.05:
        print("Conclusión:====>")
        print("Rechazar la hipótesis nula")
        print("Los datos son estacionarios")
    else:
        print("Conclusión:====>")
        print("No se puede rechazar la hipótesis nula")
        print("Los datos no son estacionarios")

Prueba_Dickey_Fuller(df["PRODUCCIÓN KW-H"], "PRODUCCIÓN KW-H")
```

**Figura 4.13.** Librería empleada para la aplicación de la prueba de DICKEY-FULLER.

En la anterior gráfica se observa la librería para realizar la comprobación de nuestra base de datos en la cual si el valor de la prueba es menor que es 0.05 es estacionaria. Esto se lo logra con

el llamado de las librerías para el análisis y con las condiciones que se agrega al programa y sus respectivas impresiones de resultados.

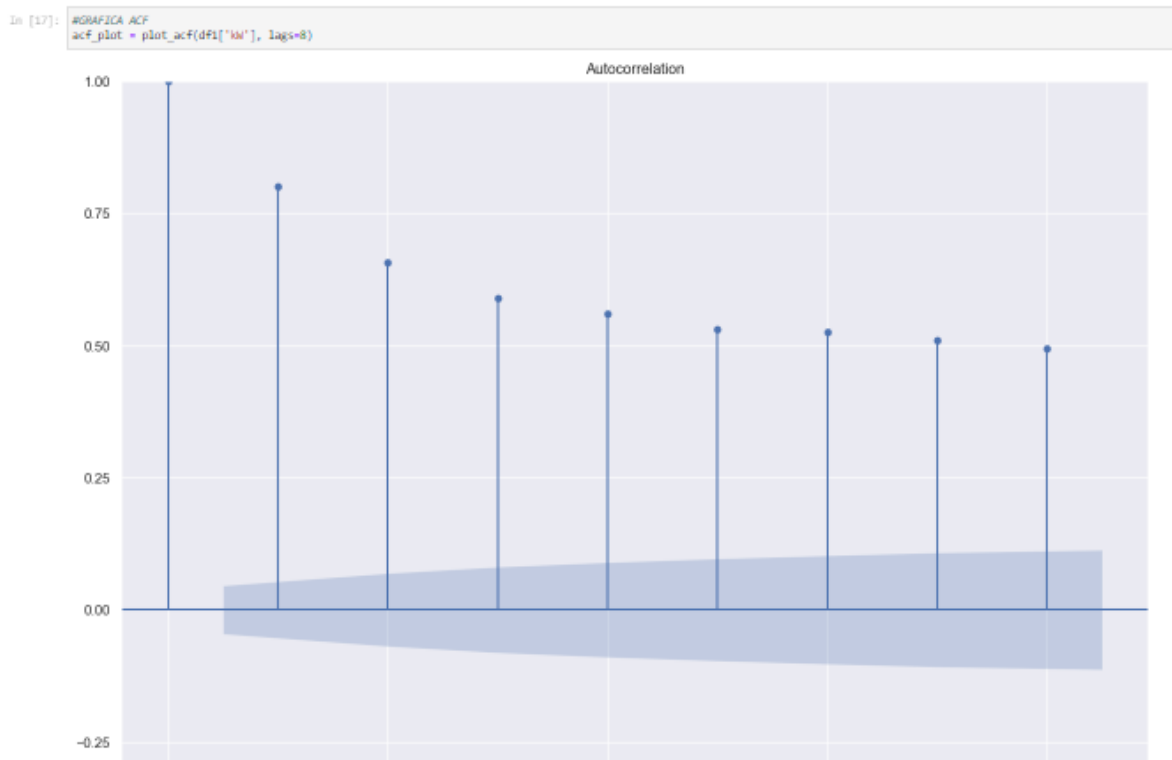
```

Resultados de la prueba de Dickey-Fuller para columna: PRODUCCIÓN KW-H
Test Estadístico          -4.988721
Valor P                   0.000023
Sin retrasos              12.000000
Número de observaciones utilizadas  1814.000000
Valor crítico (1%)       -3.433960
Valor crítico (5%)       -2.863135
Valor crítico (10%)      -2.567619
dtype: float64
Conclusion:====>
Rechazar la hipótesis nula
Los datos son estacionarios
    
```

**Figura 4.14.** Gráfica de los resultados al aplicar la prueba de DICKEY-FULLER.

En la Figura 4.14 se observa que al momento de aplicar la prueba nuestra p no sobrepasa los 0.05 por ende nuestros datos son estacionarios y no se tiene que realizar ningún tratamiento extra a la base de datos antes estudiada nombrada como df, consiguiente se toma los valores críticos para la evaluación.

**Paso 8.-** Como siguiente paso se obtiene las gráficas del ACF que se denomina la función de autocorrelación simple y el PACF es la función de autocorrelación parcial las cuales brindan información esencial para el estudio de nuestra base de datos.

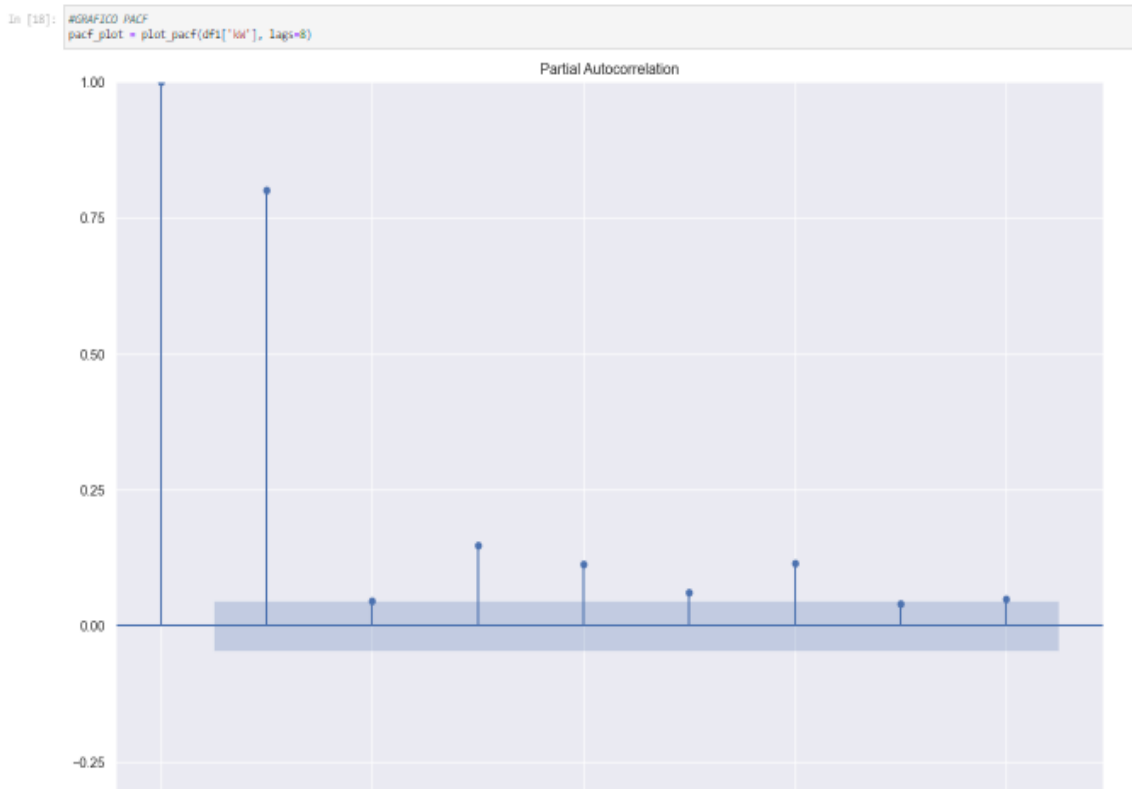


**Figura 4.15.** Gráfica ACF.



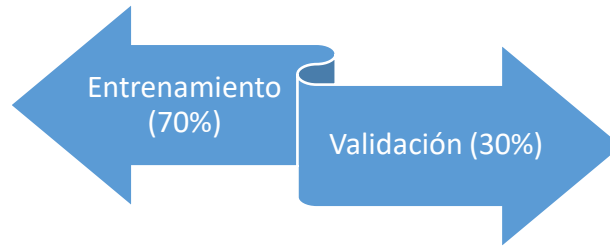
Como se observa en la Figura 4.15, el grafico del ACF con respecto a nuestra base datos y una vez ya comprobado que son datos estacionarios, con dicha información se observa que las líneas de color azul indica que la autocorrelación decae exponencialmente por ende el primer, segundo y demás rezagos se encuentran fuera de la línea de importancia por ende en esta sección se evalúa la Media móvil (MA) (valor q). La autocorrelación en el retraso 0 se incluye por defecto, lo que siempre toma el valor 1, ya que representa la correlación entre los datos y ellos mismos [38].

Con esta función en la Figura 4.16, de correlación parcial PACF, ayuda a identificar el número de coeficientes de autorregresión (AR) (valor p) en un modelo ARIMA, validan los periodos estacionales porque los coeficientes de la ACF para retardos múltiplos del período estacional de la serie son significativamente distintos de cero.



**Figura 4.16.** Gráfica PACF.

**Paso 9.-** En este paso se realiza el entrenamiento y testeo de los datos en nuestro caso de investigación se realizó con pasos de 12 con una base de 70 % de la base de datos para intentar predecir la salida partir de los datos de entrada.



**Figura 4.17.** Partición de los datos en subconjuntos para la predicción.

```
train_data = df[:len(df)-12]
test_data = df[len(df)-12:]
test=test_data.copy()

test_data.head()
```

PRODUCCIÓN KW-H	
FECHA	
2022-12-21	67200.0
2022-12-22	86400.0
2022-12-23	88800.0
2022-12-24	72000.0
2022-12-25	81600.0

**Figura 4.18.** Partición de los datos en subconjuntos para la predicción.

**Paso 10.-** En este paso se ejecuta el Modelo Auto-ARIMA la cual, es una extensión de ARIMA que admite explícitamente datos de series temporales univariadas con un componente estacional. Hay cuatro elementos estacionales que no forman parte de ARIMA que deben configurarse; ellos son:

P: orden autorregresivo estacional.

D: Orden de diferencia estacional.

Q: Orden promedio móvil estacional.

m: El número de pasos de tiempo para un solo período estacional.

Mediante este modelo se procede a realizar de manera automática la asignación a nuestras variables p, d, q y P, D, Q conjuntamente con la que se nombra de la siguiente manera nuestros datos para que se ejecuten de manera automática como se muestra a continuación:

- 7 para los días
- 12 para los meses

- 52 para las semanas
- 4 trimestral.
- 1 anual

```

: modelo_auto=auto_arima(train_data,start_p=0,d=1,start_q=0,
  max_p=4,max_d=2,max_q=4, start_P=0,
  D=1, start_Q=0, max_P=2,max_D=1,
  max_Q=2, n=12, seasonal=True,
  error_action='warn',trace=True,
  suppress_warnings=True,stepwise=True,
  random_state=20,n_fits=50)
print(modelo_auto)

Performing stepwise search to minimize aic
ARIMA(0,1,0)(0,1,0)[12] : AIC=40742.427, Time=0.15 sec
ARIMA(1,1,0)(1,1,0)[12] : AIC=40285.065, Time=1.89 sec
ARIMA(0,1,1)(0,1,1)[12] : AIC=inf, Time=5.18 sec
ARIMA(1,1,0)(0,1,0)[12] : AIC=40715.910, Time=0.44 sec
ARIMA(1,1,0)(2,1,0)[12] : AIC=40099.556, Time=3.57 sec
ARIMA(1,1,0)(2,1,1)[12] : AIC=inf, Time=11.27 sec
ARIMA(1,1,0)(1,1,1)[12] : AIC=inf, Time=7.23 sec
ARIMA(0,1,0)(2,1,0)[12] : AIC=39984.222, Time=2.55 sec
ARIMA(0,1,0)(1,1,0)[12] : AIC=40248.688, Time=0.92 sec
ARIMA(0,1,0)(2,1,1)[12] : AIC=inf, Time=9.56 sec
ARIMA(0,1,0)(1,1,1)[12] : AIC=inf, Time=5.34 sec
ARIMA(0,1,1)(2,1,0)[12] : AIC=40073.817, Time=3.66 sec
ARIMA(1,1,1)(2,1,0)[12] : AIC=inf, Time=16.75 sec
ARIMA(0,1,0)(2,1,0)[12] intercept : AIC=40127.449, Time=3.47 sec

Best model: ARIMA(0,1,0)(2,1,0)[12]
Total fit time: 71.988 seconds
ARIMA(0,1,0)(2,1,0)[12]
    
```

**Figura 4.19.** Ejecución del modelo automático de ARIMA.

En la Figura 4.19, se observa la declaración de nuestras variables en la sección de librerías, se analiza nuestro método Auto-ARIMA, luego se eligen los mismos minimizando el AIC, el cual viene a ser el criterio de información de Akaike, que se encarga de la comparación del conjunto de modelos que brinda la librería, en donde, se efectúa la comparación de p, d, q y el tiempo que toma el mismo en ejecutarse, por ende indica que, el mejor modelo: ARIMA(4, 1, 0) (2, 1, 0) [12] que viene a ser (p, d, q) (P, D, Q) [12], con un tiempo de 174.608 segundos, con esto no se tiene que estar probando uno a uno los posibles conjuntos y refleja el mejor escenario del método.

```

print(modelo_auto.summary())

=====
SARIMAX Results
=====
Dep. Variable:          y          No. Observations:      1815
Model:                SARIMAX(0, 1, 0)x(2, 1, 0, 12)  Log Likelihood        -19989.111
Date:                 Sun, 12 Feb 2023              AIC                   39984.222
Time:                 23:51:56                      BIC                   40000.712
Sample:               01-01-2018                    HQIC                  39990.309
Covariance Type:     opg

=====
              coef    std err          z      P>|z|      [0.025    0.975]
-----
ar.S.L12    -0.6704     0.018    -36.694     0.000    -0.706    -0.635
ar.S.L24    -0.3676     0.018    -20.369     0.000    -0.403    -0.332
sigma2      2.528e+08    8.97e-12    2.82e+19     0.000    2.53e+08    2.53e+08
=====
Ljung-Box (L1) (Q):          25.76   Jarque-Bera (JB):          642.98
Prob(Q):                     0.00    Prob(JB):                  0.00
Heteroskedasticity (H):      1.48   Skew:                      0.62
Prob(H) (two-sided):         0.00    Kurtosis:                  5.65
=====
    
```

**Figura 4.20.** Resultados modelo AUTO-ARIMA.

De la Figura 4.20, se destaca el AIC ya que se basa en este resultado para elegir el modelo idóneo para nuestras variables.

**Paso 11.-** En este paso se realiza la implementación del modelo manualmente con la selección de nuestras variables  $p$ ,  $d$ ,  $q$ . Para validar nuestro método Auto-ARIMA se selecciona los valores iniciales, medios y el método idóneo para realizar la comparación en base al AIC y en base a las gráficas que proporciona las librerías de Python para su respectiva comparación.

### ESCENARIO 3 $p d q = (4,1,0)$ , $P D Q = (2,1,0,12)$

```
]: arima_model = SARIMAX(train_data["PRODUCCIÓN KW-H"], order = (4,1,0), seasonal_order = (2,1,0,12))
   arima_result = arima_model.fit()
   arima_result.summary()
```

**Figura 4.21.** Escenario 1 con nuestros datos inferiores.

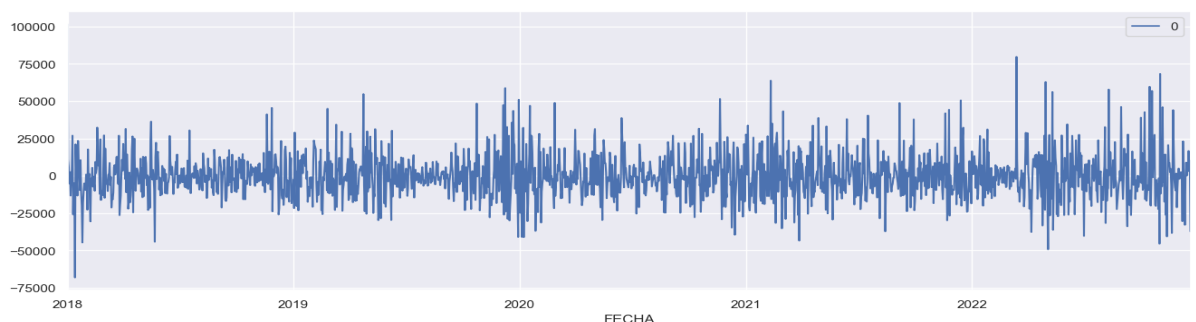
### ESCENARIO 2 $p d q = (2,1,1)$ , $P D Q = (1,1,0,12)$

```
: arima_model1 = SARIMAX(train_data["PRODUCCIÓN KW-H"], order = (3,1,1), seasonal_order = (1,1,0,12))
   arima_result1 = arima_model1.fit()
   arima_result1.summary()
```

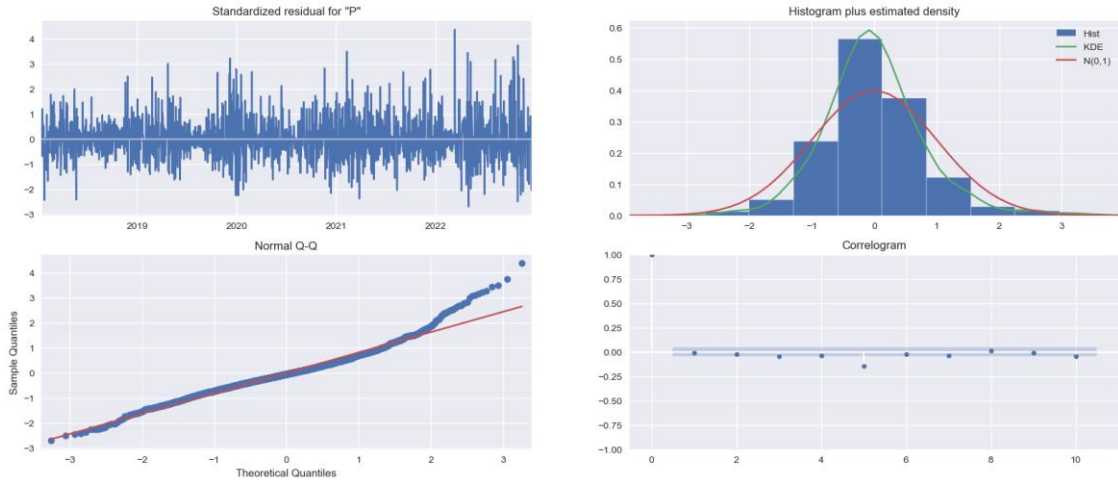
**Figura 4.22.** Escenario 2 con nuestros datos inferiores.

En las Figuras 4.21 y 4.22 se muestran los escenarios que se obtiene al cambiar las variables del método y consiguiente se analizara en la sección de análisis y resultados las gráficas de la potencia actual con la predicción y seguidamente las gráficas del AIC. Una vez especificado, se analiza las gráficas en base al mejor modelo del AIC.

El modelo con el AIC más bajo es el mejor modelo. En base a esta afirmación se procede a graficar los residuos como se observa en la Figura 4.23.



**Figura 4.23.** Gráfico de línea de errores residuales.



**Figura 4.24.** Gráfico de los residuos.

En la gráfica 4.24 en la parte superior izquierda: los errores residuales parecen fluctuar alrededor de una media de cero y tienen una varianza uniforme.

En la parte superior derecha: la gráfica de densidad sugiere una distribución normal con media cero.

En la parte inferior izquierda: todos los puntos deben estar perfectamente alineados con la línea roja. Cualquier desviación significativa implicaría que la distribución está sesgada.

En la parte inferior derecha: el correlograma, también conocido como gráfico ACF, muestra que los errores residuales no están auto correlacionados. Cualquier autocorrelación implicaría que existe algún patrón en los errores residuales que no se explican en el modelo. Por lo tanto, deberá buscar más X (predictores) en el modelo.

**Paso 12.-** En este paso se realiza la predicción de nuestra base de datos en la cual se obtiene la comparación de las gráficas con nuestro modelo que se obtuvo del modelo Auto-ARIMA y con el respectivo uso de las librerías como se observa en la Figura 4.25.

```

: arima_pred = arima_result.predict(start = len(train_data), end = len(df)-1, typ="levels").rename("ARIMA Predictions")
arima_pred

2022-12-21    76432.679139
2022-12-22    84448.860632
2022-12-23    77312.743143
2022-12-24    74466.727407
2022-12-25    71938.830034
2022-12-26    60069.910077
2022-12-27    56324.416770
2022-12-28    69042.725909
2022-12-29    79491.357390
2022-12-30    81462.559387
2022-12-31    75793.375195
2023-01-01    76820.239657
Freq: D, Name: ARIMA Predictions, dtype: float64

: arima_pred2 = arima_result.predict(start='2018-01-01',end='2023-01-01', typ="levels").rename("ARIMA Predictions")
arima_pred2

2018-01-01     0.000000
2018-01-02    88876.767462
2018-01-03    80532.215102
2018-01-04    72847.885053
2018-01-05    64990.017771
...
2022-12-28    69042.725909
2022-12-29    79491.357390
2022-12-30    81462.559387
2022-12-31    75793.375195
2023-01-01    76820.239657
Freq: D, Name: ARIMA Predictions, Length: 1827, dtype: float64

: #pd.DataFrame(arima_pred2).to_excel('PREDICCIONES/Arima_pred_2018_2023.xlsx')

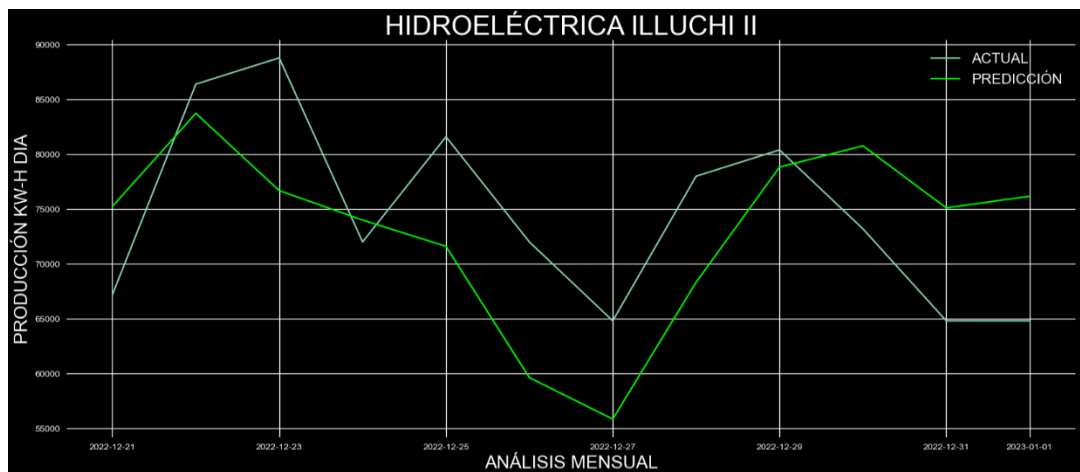
: plt.style.use('dark_background')
plt.rcParams["figure.figsize"] = (20, 8)

plt.plot(test_data["PRODUCCIÓN KW-H"], label="ACTUAL")
plt.plot(arima_pred, color="lime", label="PREDICCIÓN")
plt.title("HIDROELÉCTRICA ILLUCHI 2", fontsize=30);
plt.xlabel("ANÁLISIS MENSUAL", fontsize=20)
plt.ylabel("PRODUCCIÓN KW-H", fontsize=20)
plt.legend( fontsize=16);
plt.show();

```

**Figura 4.25.** Gráfico de la predicción.

Una vez que se usa el arima\_pred para obtener el resultado de la predicción con el método ARIMA se genera una base de datos que a continuación se van a trazar en la Figura 4.26 en donde se observa la potencia actual con la predecida.



**Figura 4.26.** Gráfica de la comparación de la potencia actual con la predicción.

En la Figura 4.27, se realiza el análisis del error en donde en base a nuestra evaluación métrica se obtiene un Mae de 326.84 y un MAPE de 12.81. Se debe tomar en cuenta en especial la evaluación del error MAPE ya que se realiza un análisis del 100% en el cual como error se tiene un 12.81 % realizando una comparación la eficiencia del programa es de un 87.19%.

```

arima_pred3 = arima_result.predict(start='2022-01-01',end='2023-01-01', typ="levels").rename("ARIMA Predictions")
arima_pred3

2022-01-01    7911.784469
2022-01-02    15642.034208
2022-01-03    22837.095898
2022-01-04    16914.561983
2022-01-05    20256.512265
...
2022-12-28    69042.725909
2022-12-29    79491.357390
2022-12-30    81462.559387
2022-12-31    75793.375195
2023-01-01    76820.239657
Freq: D, Name: ARIMA Predictions, Length: 366, dtype: float64

pd.DataFrame(arima_pred3).to_excel('PREDICCIONES/Arima_pred_2022_2023.xlsx')
    
```

**Figura 4.27.** Gráfica del Análisis del error.

**Paso 13.-** En este paso se realiza el análisis del comportamiento de los datos reales vs los predcidos. En base a las librerías y sobre todo la gráfica 4.28 en donde se evidencia la eficacia de la predicción de los datos actuales con los pronosticados.

```

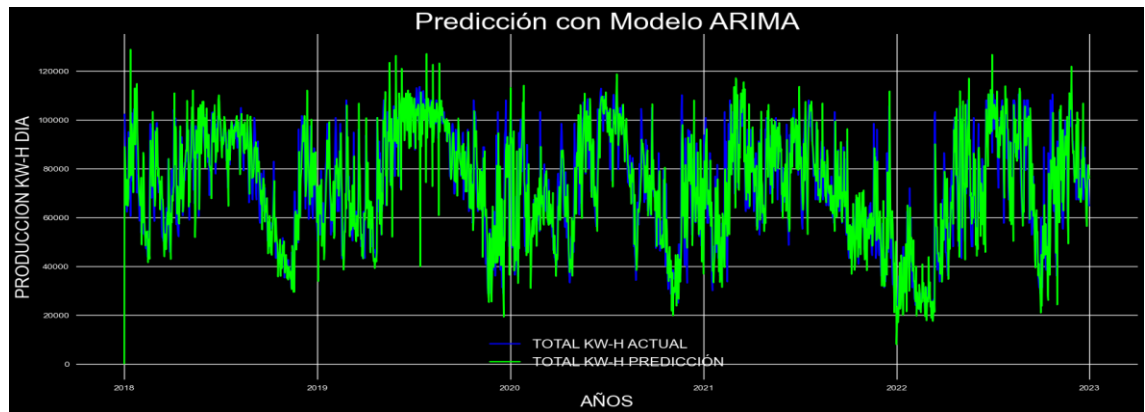
i]: dataf= pd.read_excel('DATA/Comparacion_Arima_2018_2023.xlsx', index_col='FECHA')
dataf.index = pd.to_datetime(dataf.index)
dataf

i]:

```

FECHA	TOTAL KW-H ACTUAL	TOTAL KW-H PREDICCIÓN
2018-01-01	101400.0	0.000000
2018-01-02	102200.0	88876.767462
2018-01-03	85400.0	80532.215102
2018-01-04	67500.0	72847.885053
2018-01-05	67000.0	64990.017771
...	...	...
2022-12-28	78000.0	69042.725909
2022-12-29	80400.0	79491.357390
2022-12-30	73200.0	81462.559387
2022-12-31	64800.0	75793.375195
2023-01-01	64800.0	76820.239657

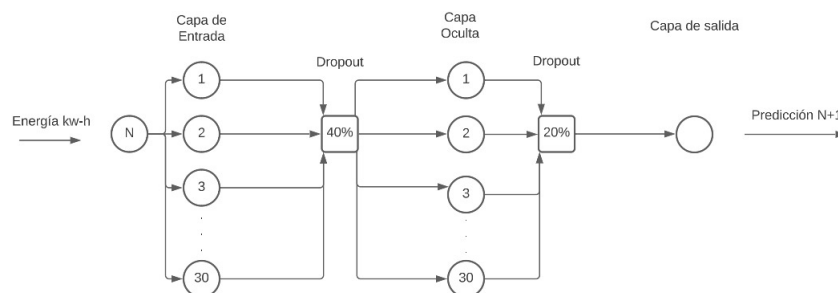
**Figura 4.28.** Gráfica de las librerías del comportamiento de los datos.



**Figura 4.29.** Gráfica del comportamiento de los datos.

En la Figura 4.29, se observa claramente el comportamiento de los datos actuales con la predicción, en este caso al ser una gráfica no se observa de manera específica como se lo puede realizar en el programa colocando el cursor en cada sección de la gráfica y refleja la fecha y la potencia de tal día, gracias a que se realiza la gráfica mediante la librería plot.ly que es mediante el uso de internet.

#### 4.3.2 Estructura de la red neuronal artificial para el desarrollo del algoritmo de predicción.



**Figura 4.30:** Estructura de la red neuronal artificial implementado en el algoritmo de predicción.

En el interior de cada red hay un gran número de parámetros con los que experimentar hasta obtener mejores resultados ya que se puede jugar con las funciones de activación de cada neurona, con las funciones de pérdida a minimizar, con el número de variables explicativas, con el número de capas y el número de neuronas en cada capa (hay algunas recomendaciones según la topología de la red, pero ninguna es 100 % efectiva), etc. En este trabajo de investigación se detallan las redes, con los mejores resultados que han proporcionado a nuestra red neuronal propuesta. Antes de centrarse en cómo construir redes neuronales recurrentes en Python se explicará la técnica del *Dropout*. El término de *Dropout* se refiere a “eliminar” o “perder” neuronas en una red neuronal. Es decir, ignorar un conjunto de neuronas (elegidas aleatoriamente) durante la fase de entrenamiento. De una manera más técnica, el *Dropout* hace que en cada iteración del



entrenamiento algunas neuronas no se tengan en cuenta con probabilidad  $1 - p$  y así se obtiene una red más reducida. El *Dropout* es necesario para prevenir el sobreentrenamiento (overfitting) ya que una capa totalmente conectada ocupa la mayor parte de los parámetros y por tanto las neuronas desarrollan codependencia entre ellas, lo que restringe la potencia individual de cada una y produce este sobreentrenamiento.

No hay un valor de probabilidad  $p$  para el *Dropout*, pero después de pruebas preliminares se puede concluir que el valor 0,4 y 0,2 son los que mejor funciona. Una vez explicada esta técnica que se usa (se probó sin utilizarla y se obtuvieron peores resultados por lo que se decidió utilizar esta técnica), los modelos de redes que se implementarán serán sencillos y con pocas capas. Como ya se ha dicho, se utiliza el modelo *Sequential* para crear las redes. Además, todas nuestras redes tendrán la capa de entrada en la que se indica el número de variables explicativas  $x$  que entran al modelo y la capa de salida con una única neurona que hace referencia a nuestra predicción de energía y [36].

**Tabla 4.2.** Selección preliminar de las variables para la predicción de la producción de energía eléctrica.

<b>Técnica de modelo</b>	<b>Red Neuronal recurrente aplicando células LSTM.</b>
<b>Número de capas</b>	3
<b>Número de neuronas</b>	30
<b>Función de activación</b>	Tangente hiperbólica, Sigmoidal.
<b>Optimizador</b>	RMSprop
<b>Épocas de entrenamiento</b>	100

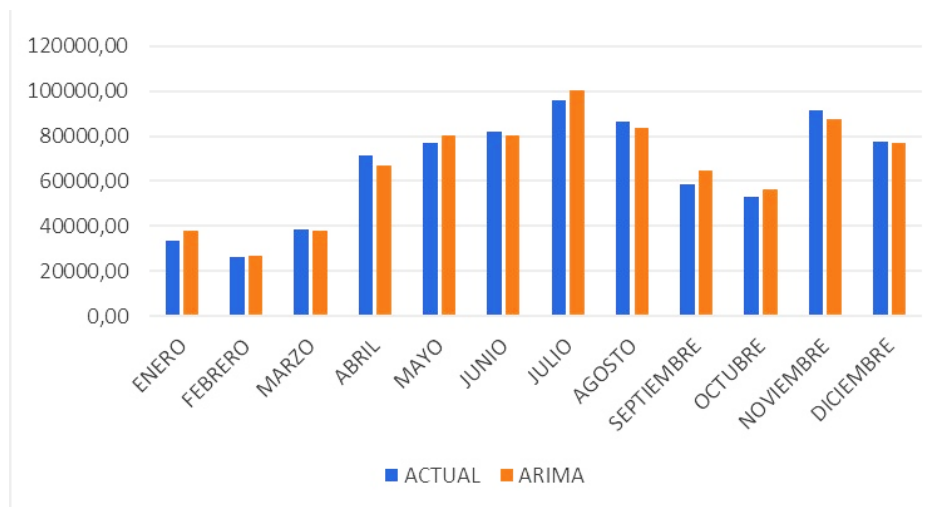
Los datos históricos ingresados en los parámetros considerados, varios factores como: el número de retrasos aplicados a la red, el número de neuronas elegidas es 30, y las épocas o número de entrenamientos también es 100, de esta manera se considera que nuestro recurrente la red neuronal permanece en estado de ajuste normal, después de pruebas las cuales iniciaron con el número de neuronas que fueron 10 y épocas 10 como se lo realizó en la programación original del manual. También se puede mencionar que nuestra red consta de una capa de congelación de neuronas llamada Dropout, que congela el 40% de las neuronas para evitar el sobreajuste, ingresar datos específicos y lograr un entrenamiento correcto de la implementación del modelo predictivo.

## 5 ANÁLISIS DE RESULTADOS

Como en el capítulo 4, se propone definiciones como: el modelo ARIMA y en este capítulo se presentan los resultados de la validación del modelo ARIMA propuesta y creada a través del lenguaje de programación Python. Para el aprendizaje y validación de se utiliza datos de la central ILLUCHI II. Además, se compara la predicción de la producción de energía eléctrica utilizando redes neuronales artificiales y finalmente se presenta los resultados de la predicción de la producción de energía eléctrica.

### 5.1 ANÁLISIS DE LA CENTRAL HIDROELÉCTRICA ILLUCHI II

Una vez que se analizó el comportamiento de la producción utilizando el total de la base de datos, se procede a realizar la comparación de la predicción 2023-2028 como se muestra en la Figura 5.1, los datos tomados son a partir de enero del 2018 que esta es nuestra base de datos para la predicción.



**Figura 5.1.** Predicción 2023-2028 con base de datos desde el 2018.

## 5.2 ANÁLISIS DE CORRELACIÓN DE LA INFORMACIÓN DE LA CENTRAL HIDROELÉCTRICA ILLUCHI II

La correlación que se presentan al momento de depurar la base de datos original es conocer si las variables pueden tener un comportamiento similar, es decir si están correlacionadas una a la otra. Esto se muestra en la Figura 5.2, las variables de año, mes, día, hora de la producción de energía eléctrica.

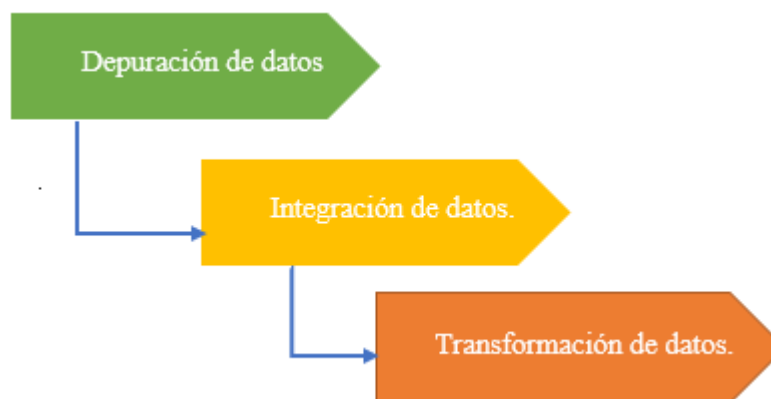
HORAS	GENERADOR No.1							GENERADOR No.2							TOTAL				
	I (A)	P (MW)	EL (MWH)	Eg (MWH)	Iex (A)	Q (MVAR)	T (oC)	I (A)	P (MW)	EL (MWH)	Eg (MWH)	Iex (A)	Q (MVAR)	T (oC)	V <sub>B</sub> (V)	Cos φ	E (KWH)	PT (KW)	QT (KVAR)
0:00	544	2,4	81.402,5	2,4	57	0,49	47	474	2,0	76.298,9	2,0	54	0,41	16	2.560	0,98	4.400	4.000	893
0:30	554	2,4			55	0,49	47	484	2,1			52	0,43	16	2.560	0,98	4.400	4.400	914
1:00	554	2,4	81.404,9	2,4	55	0,49	47	484	2,1	76.301,0	2,1	52	0,43	16	2.560	0,98	4.500	4.500	914
1:30	540	2,3			55	0,47	47	486	2,1			52	0,43	16	2.560	0,98	4.400	4.400	893
2:00	540	2,3	81.407,2	2,3	55	0,47	47	486	2,1	76.303,1	2,1	52	0,43	16	2.560	0,98	4.400	4.400	893
2:30	542	2,3			55	0,47	47	482	2,1			52	0,43	16	2.560	0,98	4.400	4.400	893
3:00	542	2,3	81.409,5	2,3	55	0,47	47	482	2,1	76.305,2	2,1	52	0,43	16	2.560	0,98	4.400	4.400	893
3:30	544	2,3			55	0,47	47	484	2,1			52	0,43	16	2.560	0,98	4.400	4.400	893
4:00	544	2,3	81.411,8	2,3	55	0,47	47	484	2,1	76.307,3	2,1	52	0,43	16	2.560	0,98	4.400	4.400	893
4:30	542	2,3			55	0,47	47	486	2,1			52	0,43	16	2.560	0,98	4.400	4.400	893
5:00	542	2,3	81.414,1	2,3	55	0,47	47	486	2,1	76.309,4	2,1	52	0,43	16	2.560	0,98	4.400	4.400	893
5:30	480	2,1			52	0,43	47	488	2,1			54	0,43	16	2.550	0,98	4.200	4.200	853
6:00	480	2,1	81.416,2	2,1	52	0,43	47	488	2,1	76.311,5	2,1	54	0,43	16	2.550	0,98	4.200	4.200	853
6:30	482	2,2			52	0,45	47	478	2,1			54	0,43	16	2.550	0,98	4.300	4.300	873

Figura 5.2. Variables de datos obtenidos por ELEPCO S.A.

Cada uno de los archivos de Excel proporcionados por ELEPCO S.A. son datos obtenidos cada media hora, el cual su depuración se lo hizo para cada día con la ayuda de la media aritmética proporcionando una base de datos detallada por cada día. Hay que mencionar que se tomó datos desde el 2018 ya que los datos de anteriores años eran distorsionados.

## 5.3 PREPROCESAMIENTO DE LA INFORMACIÓN DE LA CENTRAL HIDROELÉCTRICA ILLUCHI II

El preprocesamiento de datos es uno de los procesos que son utilizados en cualquier aplicación de aprendizaje, debido a que debe disponer de datos de calidad este paso es fundamental para poder obtener buenos resultados. Desafortunadamente los datos procedentes de la vida real suelen presentar ciertas deficiencias originadas por múltiples causas, como errores de escritura, errores de actualización de los datos, errores en la transmisión de los datos, etc. Si los datos no son preparados previamente el algoritmo de aprendizaje de la red neuronal no podrá funcionar, y si funciona lo más seguro es que no presente resultados aceptables [39].



**Figura 5.2.** Secuencia de pasos a seguir para el preprocesamiento de la información procedente de una base de datos [39]

### 5.3.1 Depuración de datos de la CENTRAL HIDROELÉCTRICA ILLUCHI II

Tener un registro almacenado de la información de un sistema siempre ha sido muy importante en un sistema eléctrico. Archivos con valores exactos de las señales y mediciones del sistema son necesarios para satisfacer requerimientos técnicos, legales y otros con regulaciones y permitirán proveer y tener un respaldo de la información para futuras operaciones del sistema además de aprovechar esta información para realizar una mejor planificación del mantenimiento de cada uno de los componentes del sistema [40].

La información de los grupos generadores de la CENTRAL HIDROELÉCTRICA ELÉCTRICA ILLUCHI II, están generados por los operarios encargados de obtener registros como: potencia, corriente, energía, etc. La frecuencia de almacenamiento de los datos es de aproximadamente cada media hora.

Los operarios encargados diariamente verifican la información y los almacenan en un archivo en formato Excel tal como se muestra en la Figura 5.8, especificando las mediciones con respecto a las horas de trabajo [36]. Una vez que finalice el mes de trabajo, operación y almacenamiento de todas las mediciones, esta base de datos es enviada a las oficinas de ELEPCO S.A. para para verificación.

# UNIVERSIDAD TÉCNICA DE COTOPAXI – CARRERA DE ELECTRICIDAD

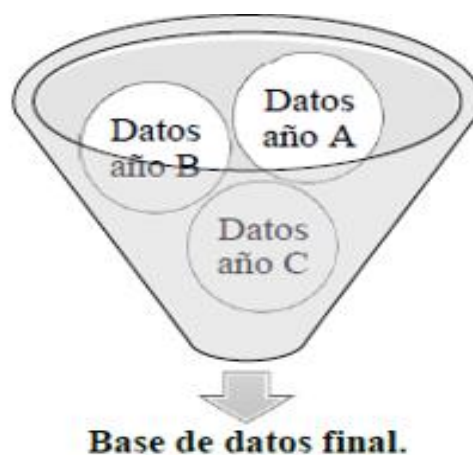
HORAS	GENERADOR No.1								GENERADOR No.2								TOTAL				
	I (A)	P (MW)	E <sub>L</sub> (MWh)	E <sub>G</sub> (MWh)	I <sub>EX</sub> (A)	Q (MVAR)	T (°C)		I (A)	P (MW)	E <sub>L</sub> (MWh)	E <sub>G</sub> (MWh)	I <sub>EX</sub> (A)	Q (MVAR)	T (°C)	V <sub>e</sub> (V)	Cos φ	E (kWh)	PT (kWh)	QT (kWh)	
0:00	544	2.4	81.402,5	2,4	57	0.49	47	474	2.0	76.298,9	2,0	54	0.41	16	2.560	0.98		4.400	4.000	893	
0:30	554	2.4			55	0.49	47	484	2.1					52	0.43	16	2.560	0.98		4.400	914
1:00	554	2.4	81.404,9	2,4	55	0.49	47	484	2.1	76.301,0	2,1	52	0.43	16	2.560	0.98		4.500	4.500	914	
1:30	540	2.3			55	0.47	47	486	2.1					52	0.43	16	2.560	0.98		4.400	893
2:00	540	2.3	81.407,2	2,3	55	0.47	47	486	2.1	76.303,1	2,1	52	0.43	16	2.560	0.98		4.400	4.400	893	
2:30	542	2.3			55	0.47	47	482	2.1					52	0.43	16	2.560	0.98		4.400	893
3:00	542	2.3	81.409,5	2,3	55	0.47	47	482	2.1	76.305,2	2,1	52	0.43	16	2.560	0.98		4.400	4.400	893	
3:30	544	2.3			55	0.47	47	484	2.1					52	0.43	16	2.560	0.98		4.400	893
4:00	544	2.3	81.411,8	2,3	55	0.47	47	484	2.1	76.307,3	2,1	52	0.43	16	2.560	0.98		4.400	4.400	893	
4:30	542	2.3			55	0.47	47	486	2.1					52	0.43	16	2.560	0.98		4.400	893
5:00	542	2.3	81.414,1	2,3	55	0.47	47	486	2.1	76.309,4	2,1	52	0.43	16	2.560	0.98		4.400	4.400	893	
5:30	480	2.1			52	0.43	47	488	2.1					54	0.43	16	2.550	0.98		4.200	853
6:00	480	2.1	81.416,2	2,1	52	0.43	47	488	2.1	76.311,5	2,1	54	0.43	16	2.550	0.98		4.200	4.200	853	
6:30	482	2.2			52	0.45	47	478	2.1					54	0.43	16	2.550	0.98		4.300	873
7:00	482	2.2	81.418,4	2,2	52	0.45	47	478	2.1	76.313,6	2,1	54	0.43	16	2.550	0.98		4.300	4.300	873	
7:30	480	2.1			52	0.43	47	480	2.1					54	0.43	16	2.550	0.98		4.200	853
8:00	480	2.1	81.420,5	2,1	52	0.43	47	480	2.1	76.315,7	2,1	54	0.43	16	2.550	0.98		4.200	4.200	853	
8:30	504	2.1			58	0.69	47	496	2.1					60	0.69	18	2.550	0.95		4.200	1.380
9:00	504	2.1	81.422,6	2,1	58	0.69	47	496	2.1	76.317,8	2,1	60	0.69	18	2.550	0.95		4.200	4.200	1.380	
9:30	508	2.1			58	0.83	47	502	2.1					62	0.83	17	2.540	0.93		4.200	1.660
10:00	508	2.1	81.424,7	2,1	58	0.83	47	502	2.1	76.319,9	2,1	62	0.83	17	2.540	0.93		4.200	4.200	1.660	
10:30	504	2.1			58	0.69	46	500	2.1					62	0.69	17	2.540	0.95		4.200	1.380
11:00	504	2.1	81.426,8	2,1	58	0.69	46	500	2.1	76.322,0	2,1	62	0.69	17	2.540	0.95		4.200	4.200	1.380	
11:30	506	2.1			58	0.69	46	504	2.1					62	0.69	17	2.540	0.95		4.200	1.380
12:00	506	2.1	81.428,9	2,1	58	0.69	46	504	2.1	76.324,1	2,1	62	0.69	17	2.540	0.95		4.200	4.200	1.380	
12:30	492	2.1			57	0.61	47	494	2.1					60	0.61	18	2.540	0.96		4.200	1.225
13:00	492	2.1	81.431,0	2,1	57	0.61	47	494	2.1	76.326,2	2,1	60	0.61	18	2.540	0.96		4.200	4.200	1.225	
13:30	504	2.1			59	0.69	47	498	2.0					62	0.66	18	2.540	0.95		4.100	1.348
14:00	504	2.1	81.433,1	2,1	59	0.69	47	498	2.0	76.328,2	2,0	62	0.66	18	2.540	0.95		4.100	4.100	1.348	
14:30	506	2.1			59	0.76	47	506	2.0					63	0.73	17	2.540	0.94		4.100	1.488
15:00	506	2.1	81.435,2	2,1	59	0.76	47	506	2.0	76.330,2	2,0	63	0.73	17	2.540	0.94		4.100	4.100	1.488	
15:30	502	2.1			58	0.61	47	496	2.1					60	0.61	17	2.540	0.96		4.200	1.225
16:00	502	2.1	81.437,3	2,1	58	0.61	47	496	2.1	76.332,3	2,1	60	0.61	17	2.540	0.96		4.200	4.200	1.225	
16:30	502	2.1			58	0.61	47	496	2.1					60	0.61	17	2.540	0.96		4.200	1.225
17:00	502	2.1	81.439,4	2,1	58	0.61	47	496	2.1	76.334,4	2,1	60	0.61	17	2.540	0.96		4.200	4.200	1.225	
17:30	502	2.1			58	0.61	47	496	2.1					60	0.61	17	2.540	0.96		4.200	1.225
18:00	502	2.1	81.441,5	2,1	58	0.61	47	496	2.1	76.336,5	2,1	60	0.61	17	2.540	0.96		4.200	4.200	1.225	
18:30	488	2.1			54	0.43	48	484	2.1					56	0.43	17	2.550	0.98		4.200	853
19:00	488	2.1	81.443,6	2,1	54	0.43	48	484	2.1	76.338,6	2,1	56	0.43	17	2.550	0.98		4.200	4.200	853	
19:30	488	2.1			54	0.43	48	484	2.1					56	0.43	17	2.550	0.98		4.200	853
20:00	488	2.1	81.445,7	2,1	54	0.43	48	484	2.1	76.340,7	2,1	56	0.43	17	2.550	0.98		4.200	4.200	853	
20:30	488	2.1			54	0.43	48	484	2.1					56	0.43	17	2.550	0.98		4.200	853
21:00	488	2.1	81.447,8	2,1	54	0.43	48	484	2.1	76.342,8	2,1	56	0.43	17	2.550	0.98		4.200	4.200	853	
21:30	488	2.1			54	0.43	48	484	2.1					56	0.43	17	2.550	0.98		4.200	853
22:00	488	2.1	81.449,9	2,1	54	0.43	48	484	2.1	76.344,9	2,1	56	0.43	17	2.550	0.98		4.200	4.200	853	
22:30	488	2.1			54	0.43	48	484	2.1					56	0.43	16	2.550	0.98		4.200	853
23:00	488	2.1	81.452,0	2,1	54	0.43	48	484	2.1	76.347,0	2,1	56	0.43	16	2.550	0.98		4.200	4.200	853	
23:30	488	2.1			54	0.43	48	484	2.1					56	0.43	16	2.550	0.98		4.200	853
24:00	488	2.1	81.454,1	2,1	54	0.43	48	484	2.1	76.349,1	2,1	56	0.43	16	2.550	0.98		4.200	4.200	853	
MAXIMOS	554	2.4				0.83	48	506	2.1					63	0.83	18	2.560	0.98		4.500	1.660
MINIMOS	480	2.1				0.43	46	474	2.0					52	0.41	16	2.540	0.93		4.000	853
E. TOTAL				54.0						52.2								106.200			

**Figura 5.3.** Muestra de la base de datos del proceso de generación de la CENTRAL HIDROELÉCTRICA ILLUCHI II cada media hora enero 2018.

La depuración de datos consiste en filtrar la información que contiene la base de datos, eliminando datos que son irrelevantes o innecesarias. Esta nueva información fue suministrado a través de un conjunto de hojas de cálculo en Excel, el objetivo de este paso es obtener un archivo final que contenga las variables seleccionadas, con las series de datos relacionados con la producción de energía eléctrica, es te paso es largo debido a que se lo realiza de forma manual.

### 5.3.2 Integración de datos de la central HIDROELÉCTRICA ILLUCHI II

En este proceso se procede a la unión datos procedentes de diferentes bases de datos como se presenta en la Figura 5. 4, este proceso debe ser cuidadoso para no mezclar cada uno de los datos.



**Figura 5.4.** Integración de la información hacia una nueva base de datos [36]

### 5.3.3 Transformación de datos de la central ILLUCHI II

Para la transformación de datos de la central se realiza una depuración de cada uno de los datos ya que por tener la lectura cada media hora se realiza una media aritmética la cual se tiene el resultado por día, y así se realizará para cada uno de los días divididas por cada mes como se indica en la Figura 5.5, el cual se realizará por cada uno de los años hasta enero 2023.

Esta misma depuración sirve para la modelación del modelo ARIMA en Python y se procede al llamado correspondiente de cada uno de los años para la proyección.

FECHA	PRODUCCIÓN KW-H DIA
1/1/2018	101400
2/1/2018	102200
3/1/2018	85400
4/1/2018	67500
5/1/2018	67000
6/1/2018	62200
7/1/2018	66000
8/1/2018	69000
9/1/2018	95000
10/1/2018	67500
11/1/2018	71200
12/1/2018	72400
13/1/2018	60400
14/1/2018	96500
15/1/2018	102800
16/1/2018	90400
17/1/2018	74800
18/1/2018	93200

**Figura 5.5.** Muestra de la base de datos depurada año 2018 del proceso de generación de la CENTRAL HIDROELÉCTRICA ILLUCHI II.

## 5.4 SELECCIÓN DE LAS VARIABLES DE ENTRADA PARA EL DESARROLLO DEL ALGORITMO DE PREDICCIÓN

La selección de las variables de entrada se obtiene mediante el tratamiento de la base de datos para filtrar la información útil.

- En la matriz inicial se identificó las variables de entrada clave, es decir los datos que proporcionarán la información hacia el modelo ARIMA.
- Para este caso la información de los datos que formarán parte de las nuevas variables de entrada se describe en la tabla 5.1.

**Tabla 5.1.** Selección final de las variables de entrada para la predicción de la producción de energía eléctrica.

Variable	Descripción	Unidad de medida
Fecha (año, día, mes, hora)	Tiempo de consumo de energía eléctrica	Tiempo
Energía consumida	Energía activa consumida en Kilovatio-Hora-Día	KWh-Día

## 5.5 VALIDACIÓN Y PREPROCESAMIENTO DE DATOS DE ENTRADA PARA EL DESARROLLO DEL ALGORITMO DE PREDICCIÓN

Los datos empleados para el modelamiento se basaron en datos históricos de la CENTRAL HIDROELÉCTRICA ILLUCHI II, que inicia desde el año 2018, para este proceso se toma como referencia los 5 años en específicos 2018-2022.

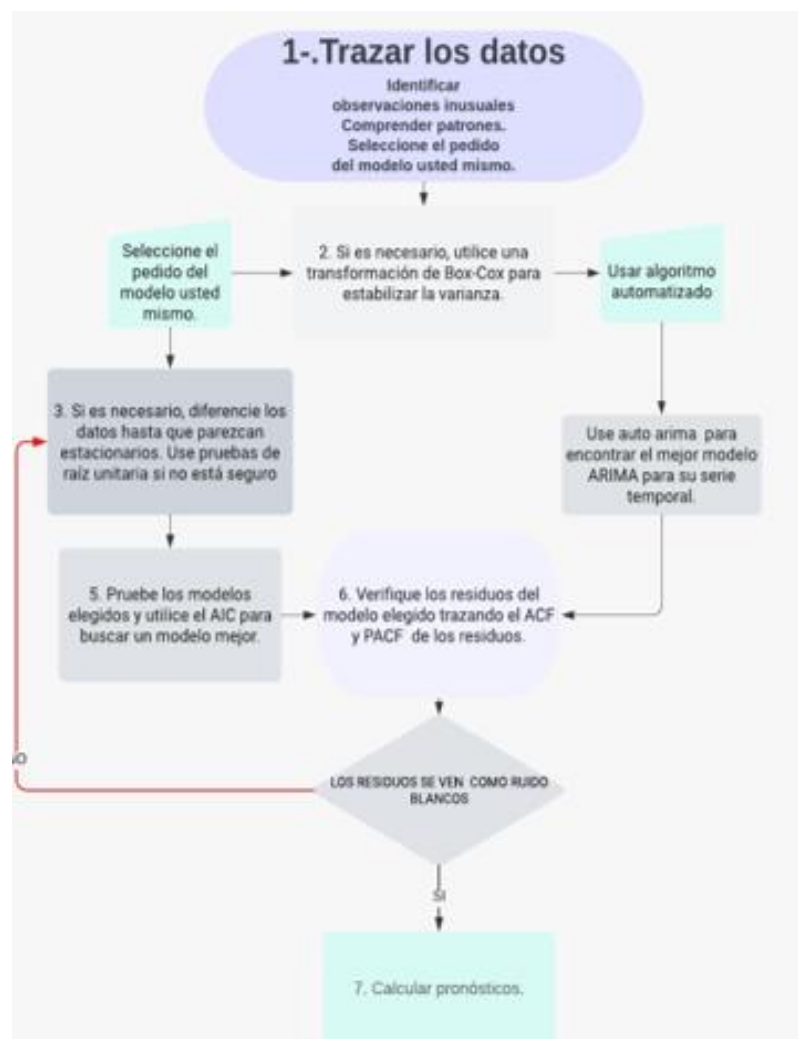
Una de las formas para definir correctamente el conjunto de datos del modelo, deben cumplir con dos características fundamentales:

- El conjunto de datos debe ser lo suficientemente grande como para obtener resultados significativos.
- Todo el conjunto de datos debe ser representativos, es decir, no se debe elegir un conjunto de datos de validación con características diferentes al conjunto de datos del modelo.

### 5.5.1 Desarrollo del algoritmo de predicción en Python

Una vez visto y analizado cómo es nuestra serie temporal se procede a crear las librerías que ayuden a predecir la producción eléctrica durante el tiempo a predecir.

Statsmodels es un módulo de Python que proporciona clases y funciones para la estimación de muchos modelos estadísticos diferentes, así como para realizar pruebas estadísticas y exploración de datos estadísticos. Hay disponible una extensa lista de estadísticas de resultados para cada estimador. Los resultados se comparan con los paquetes estadísticos existentes para garantizar que sean correctos [29].



**Figura 5.6.** Diagrama de flujo en base del análisis de Datos para el modelo ARIMA [41].



```

# Manipulación y tratamiento de Datos
import numpy as np
import pandas as pd

# Visualización de datos
import plotly.express as px
import matplotlib.pyplot as plt
%matplotlib inline
plt.style.use('ggplot')

# Modelación Arima
from statsmodels.tsa.statespace.sarimax import SARIMAX
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
from statsmodels.tsa.seasonal import seasonal_decompose
from statsmodels.tsa.stattools import adfuller

# Métrica de Evaluación
from sklearn.metrics import mean_squared_error
from statsmodels.tools.eval_measures import rmse
from sklearn import metrics
import statsmodels.api as sm
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score

# No presentar advertencias
import warnings
warnings.filterwarnings("ignore")

import seaborn as sns
sns.set()

from datetime import datetime
from datetime import timedelta
from datetime import date

from plotly import __version__
from plotly.offline import download_plotlyjs, init_notebook_mode, plot, iplot
import plotly.graph_objects as go
from plotly.subplots import make_subplots
import cufflinks as cf
init_notebook_mode(connected=True)
cf.go_offline()
import plotly.express as px
import chart_studio.plotly as py

pd.set_option('display.max_columns', None)

```

**Figura 5.7.** Aplicación de las librerías de Python para implementar el algoritmo de predicción.

### 5.5.2 Preprocesado de los datos en Python

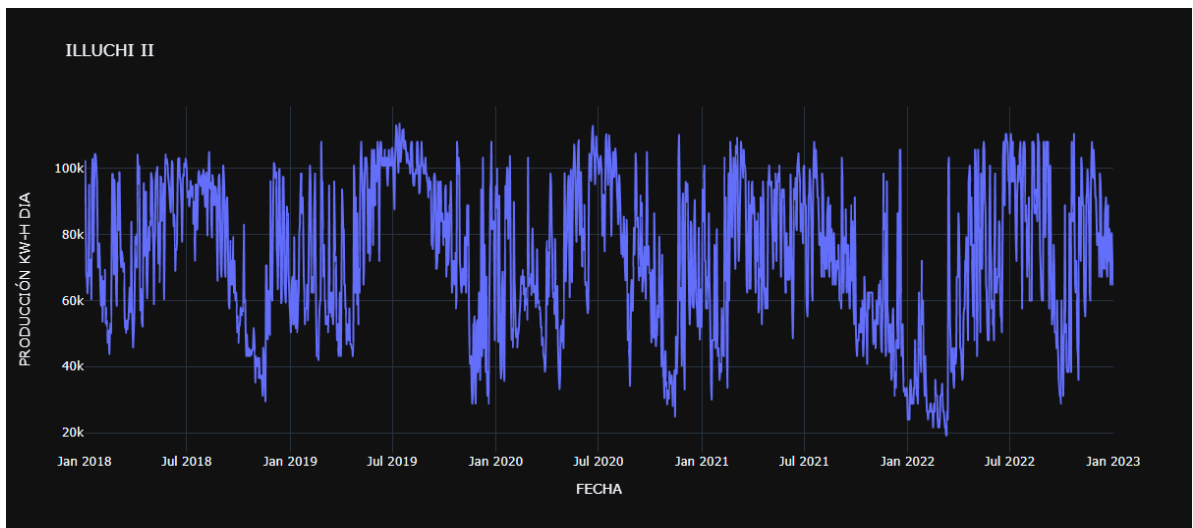
Con la depuración de datos obtenidos se tiene y se crea un dataframe con el formato de fecha correcto y el valor de KWH-Día, esta estructura se observa en la Figura 5.8.

	FECHA	PRODUCCIÓN KW-H
0	2018-01-01	101400.0
1	2018-01-02	102200.0
2	2018-01-03	85400.0
3	2018-01-04	67500.0
4	2018-01-05	67000.0
...	...	...
1822	2022-12-28	78000.0
1823	2022-12-29	80400.0
1824	2022-12-30	73200.0
1825	2022-12-31	64800.0
1826	2023-01-01	64800.0

1827 rows × 2 columns

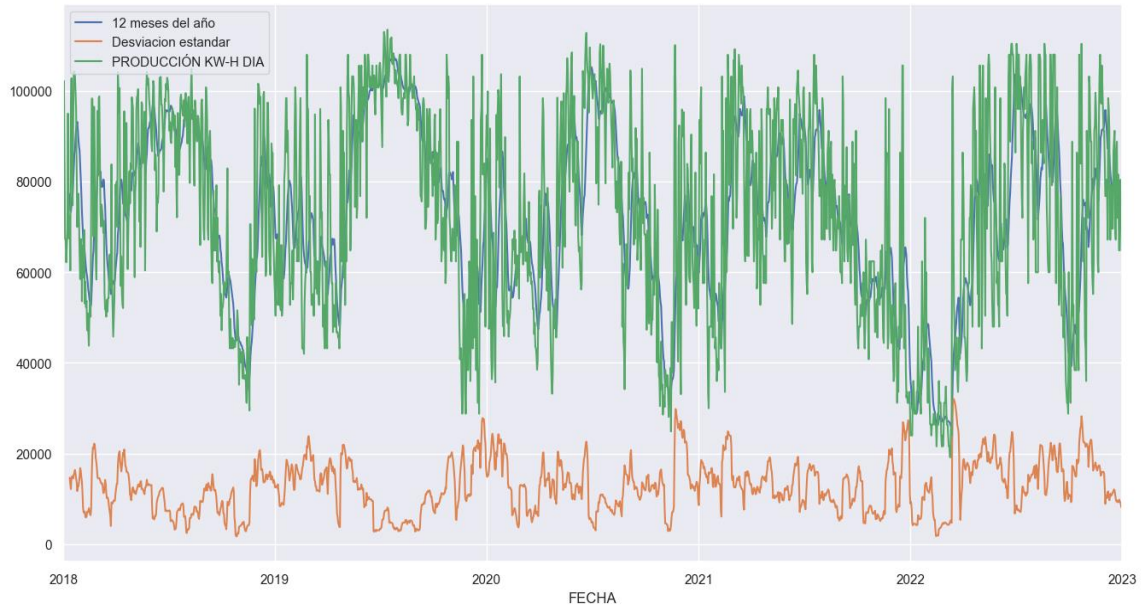
**Figura 5.8.** Datos creados por el nuevo Dataframe para la implementación del algoritmo de predicción.

Con plotly.express se puede observar la serie de tiempo a partir de los datos ejecutados en la Figura 5.9, en esta librería se puede apreciar la producción por el intervalo de tiempo.



**Figura 5.9.** Serie Temporal de la predicción junio2018-junio2023.

La desviación estándar también se traza para ver si hay alguna varianza o no. Como se ve en la Figura 5.10, la desviación estándar no varía mucho, por lo que no hay mucha varianza.

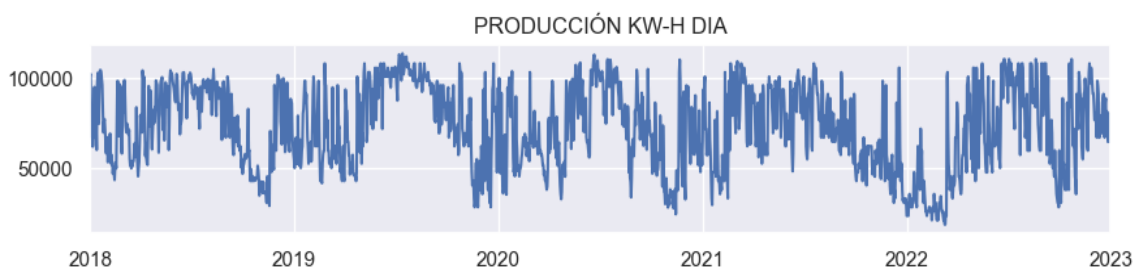


**Figura 5.10.** Desviación Estándar por el intervalo de tiempo 2018-2023.

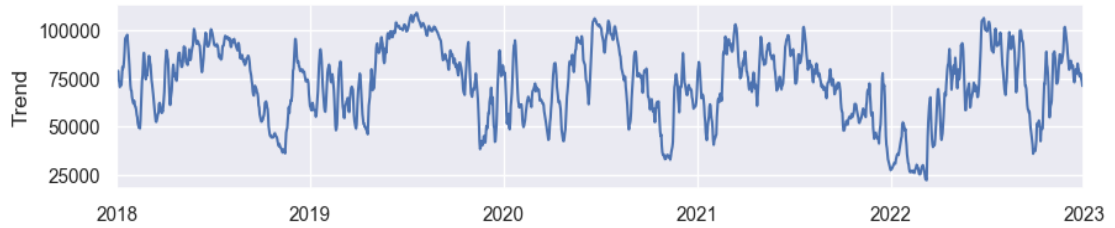
### 5.5.3 Descomposición de los datos de series temporales a su tendencia, estacionalidad y componentes residuales.

La descomposición de datos se lo realiza con la librería `from statsmodels.tsa.seasonal import seasonal_decompose`, y se obtiene gráficas las cuales son:

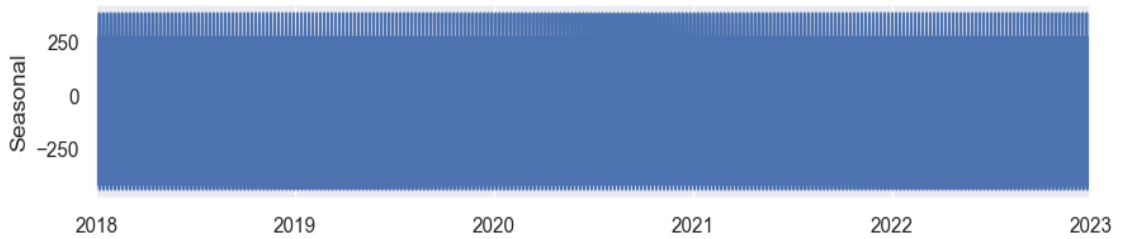
- Serie temporal deducida por los datos modelados.
- La tendencia que sigue la serie temporal.
- La estacionalidad de la serie temporal.
- Los residuos de la serie temporal.



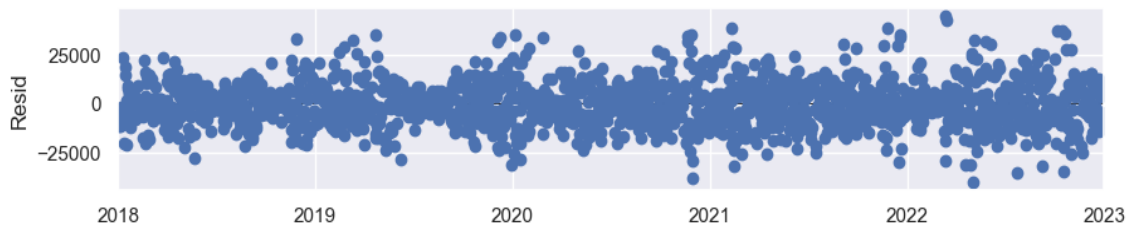
**Figura 5.11.** Serie temporal deducida por los datos modelados 2018-2023.



**Figura 5.12.** Tendencia que sigue la serie temporal desde el 2018-2023.



**Figura 5.13.** La estacionalidad de la serie temporal datos del 2018-2023.



**Figura 5.14.** Los residuos de la serie temporal datos del 2018-2023.

## 5.6 DEFINIR LA ESTRUCTURA PARA QUE LA SERIE SEA ESTACIONARIA

Para la verificar si la serie es estacionaria se re realiza la prueba de Dickey Fuller Aumentada, en otras palabras, es una forma para determinar si existen raíces unitarias en las series de tiempo y por tanto comprobar si hay o no estacionariedad. Para esta prueba se nombra la librería adfuller como se indica en la Figura 5.15.

```
def Prueba_Dickey_Fuller(series , column_name):
    print (f'Resultados de la prueba de Dickey-Fuller para columna: {column_name}')
    dftest = adfuller(series, autolag='AIC')
    dfoutput = pd.Series(dftest[0:4], index=['Test Estadístico', 'Valor P', 'Sin retrasos', 'Número de observaciones utilizadas'])
    for key,value in dftest[4].items():
        dfoutput['Valor critico (%s)'%key] = value
    print (dfoutput)
    if dftest[1] <= 0.05:
        print("Conclusion:====>")
        print("Rechazar la hipótesis nula")
        print("Los datos son estacionarios")
    else:
        print("Conclusion:====>")
        print("No se puede rechazar la hipótesis nula")
        print("Los datos no son estacionarios")
```

**Figura 5.15.** Estructura para el procesamiento de la prueba Dickey Fuller Aumentada.

La prueba ADF es que la serie de tiempo do es estacionaria Entonces si el valor de la prueba es menor que el nivel de significancia (0.05), entonces determina que la serie que la serie de tiempo es estacionaria Figura 5.16.

Entonces, en nuestro caso, si  $P > \text{Value } 0.05$  seguiría adelante, con la búsqueda del orden de diferenciación.

```
Resultados de la prueba de Dickey-Fuller para columna: PRODUCCIÓN KW-H
Test Estadístico                -4.988721
Valor P                          0.000023
Sin retrasos                     12.000000
Número de observaciones utilizadas 1814.000000
Valor critico (1%)               -3.433960
Valor critico (5%)               -2.863135
Valor critico (10%)              -2.567619
dtype: float64
Conclusion:====>
Rechazar la hipótesis nula
Los datos son estacionarios
```

**Figura 5.16.** Resultados de la prueba Dickey Fuller Aumentada en Python.

Después de determinar los resultados de Dickey Fuller y esclarecer que la serie de tiempo es estacionaria se determina el siguiente paso. La cual es la diferenciación entre los datos kW.

```
df1=df.copy()
#Tomamos la primera diferencia entre los datos kw
df1['PRODUCCIÓN KW-H_diff'] = df['PRODUCCIÓN KW-H'].diff()

#se eliminan los datos nulos, comando dropna
df1.dropna(inplace=True)

# Se crea un nuevo df
df1
```

**Figura 5.17.** En caso de no ser estacionarios se realiza la diferenciación entre los datos.

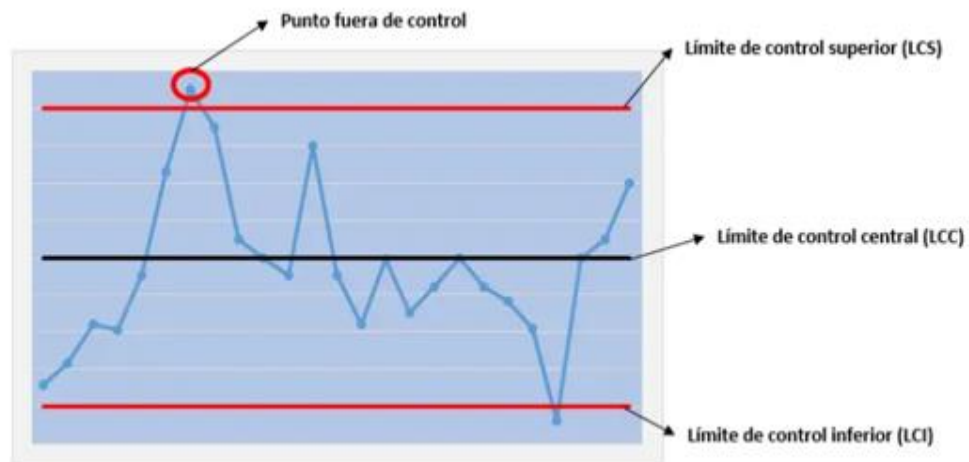
## 5.7 ESCENARIOS APLICADOS CON EL MODELO ARIMA

Por los escenarios a determinar se aplica con la gráfica de control. Creado por Walter Andrew Shewhart en 1920, el diagrama o gráfico de control, también conocido como diagrama de Shewhart, carta de control o diagrama de comportamiento de proceso, básicamente consiste en los datos de un proceso plasmados de forma cronológica en una gráfica con límites establecidos, lo que permite determinar cuándo una variación no es normal [42].

Es un diagrama que muestra los valores producto de la medición de una característica de calidad, ubicados en una serie cronológica. Se establece una línea central o valor nominal, que suele ser el objetivo del proceso o el promedio histórico, junto a uno o más límites de control, tanto superior como inferior, usados para determinar cuándo es necesario analizar una eventualidad.

**Límite superior de control:** Es el valor más grande aceptado en el proceso. En español LSC, en inglés UCL (Upper control limit).

**Límite inferior de control:** El opuesto al superior, es decir, el valor más pequeño. LIC en español, LCL en inglés (Lower control limit) [42].



**Figura 5.18.** Partes de un gráfico de control [42].

Para la ejecución en la programación de Python se toma el valor más alto y el valor bajo como se presenta en la gráfica de control Figura 5.18.

### 5.7.1 Escenario 1

En el escenario, límite superior se obtiene por la gráfica de control con los intervalos.

$$p d q = (0,1,0), P D Q = (2,1,0,12)$$

SARIMAX Results						
Dep. Variable:	PRODUCCIÓN KW-H DIA	No. Observations:	1815			
Model:	SARIMAX(0, 1, 0)x(0, 1, 0, 12)	Log Likelihood	-20325.314			
Date:	Tue, 28 Feb 2023	AIC	40652.628			
Time:	04:16:25	BIC	40658.124			
Sample:	01-01-2018	HQIC	40654.656			
	- 12-20-2022					
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
sigma2	3.643e+08	8.82e+06	41.314	0.000	3.47e+08	3.82e+08
Ljung-Box (L1) (Q):	16.40	Jarque-Bera (JB):	228.12			
Prob(Q):	0.00	Prob(JB):	0.00			
Heteroskedasticity (H):	1.76	Skew:	0.09			
Prob(H) (two-sided):	0.00	Kurtosis:	4.73			

Figura 5.19. Programación de la gráfica de control escenario 1 en Python.

### 5.7.2 Escenario 2

En el escenario 2, límite inferior se obtiene por la gráfica de control con los intervalos.

$$p d q = (2,1,0), P D Q = (1,1,0,12)$$

SARIMAX Results						
Dep. Variable:	PRODUCCIÓN KW-H DIA	No. Observations:	1815			
Model:	SARIMAX(2, 1, 0)x(1, 1, [1], 12)	Log Likelihood	-19834.426			
Date:	Tue, 28 Feb 2023	AIC	39678.853			
Time:	04:16:44	BIC	39706.336			
Sample:	01-01-2018	HQIC	39688.997			
	- 12-20-2022					
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	-0.1116	0.039	-2.831	0.005	-0.189	-0.034
ar.L2	-0.2227	0.040	-5.556	0.000	-0.301	-0.144
ar.S.L12	-0.0071	0.043	-0.165	0.869	-0.092	0.078
ma.S.L12	-0.9722	0.016	-61.208	0.000	-1.003	-0.941
sigma2	3.437e+08	3.15e-11	1.09e+19	0.000	3.44e+08	3.44e+08
Ljung-Box (L1) (Q):	2.17	Jarque-Bera (JB):	925.30			
Prob(Q):	0.14	Prob(JB):	0.00			
Heteroskedasticity (H):	1.66	Skew:	0.87			
Prob(H) (two-sided):	0.00	Kurtosis:	6.05			

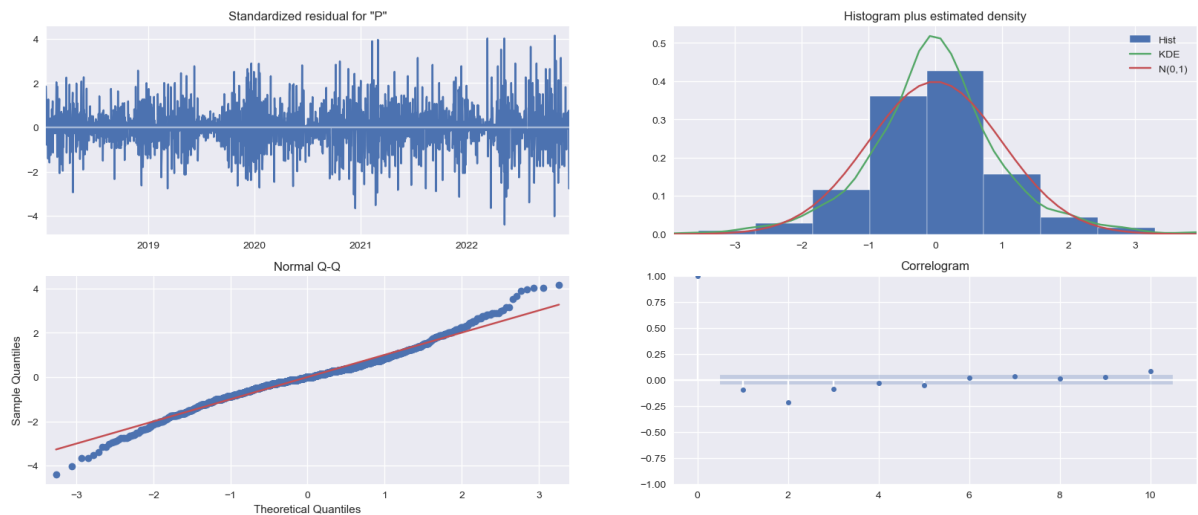
Figura 5.20. Programación de la gráfica de control escenario 2 en Python.

Con la librería `arima_result.plot_diagnostics` se puede apreciar los datos que se muestrean.

En la Figura 5.21 y 5.22 en la parte superior izquierda indica los errores residuales parecen fluctuar alrededor de una media de cero y tienen una varianza uniforme tanto como para el escenario 1 y 2.

En la parte superior de la derecha la Figura 5.21 y 5.22, se tiene que la de densidad sugiere una distribución normal con media cero tanto para el escenario 1 y 2.

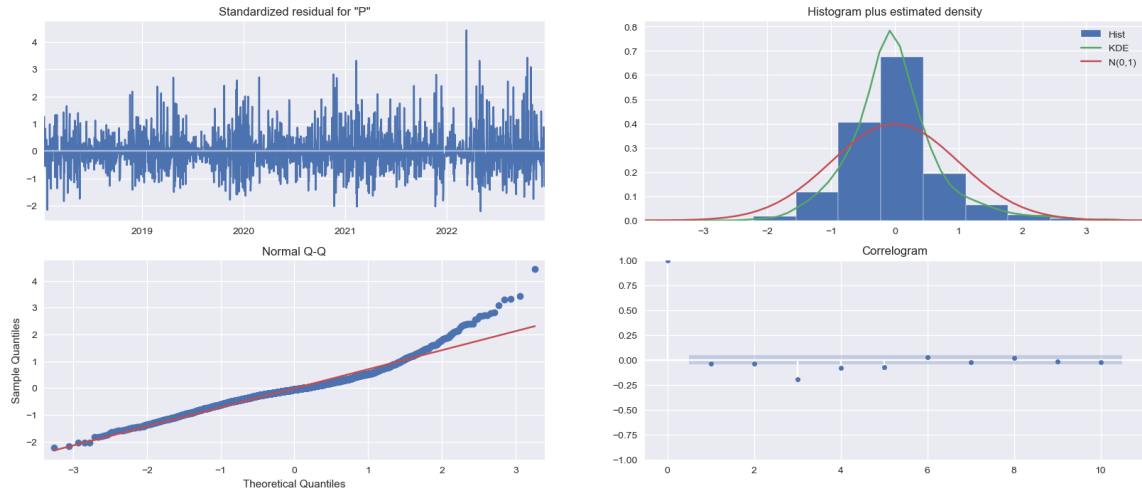
En la parte inferior derecha en la Figura 5.21 y 5.22, determina que todos los puntos deben estar perfectamente alineados con la línea roja. Cualquier desviación significativa implicaría que la distribución está sesgada.



**Figura 5.21.** Errores residuales escenario 1.

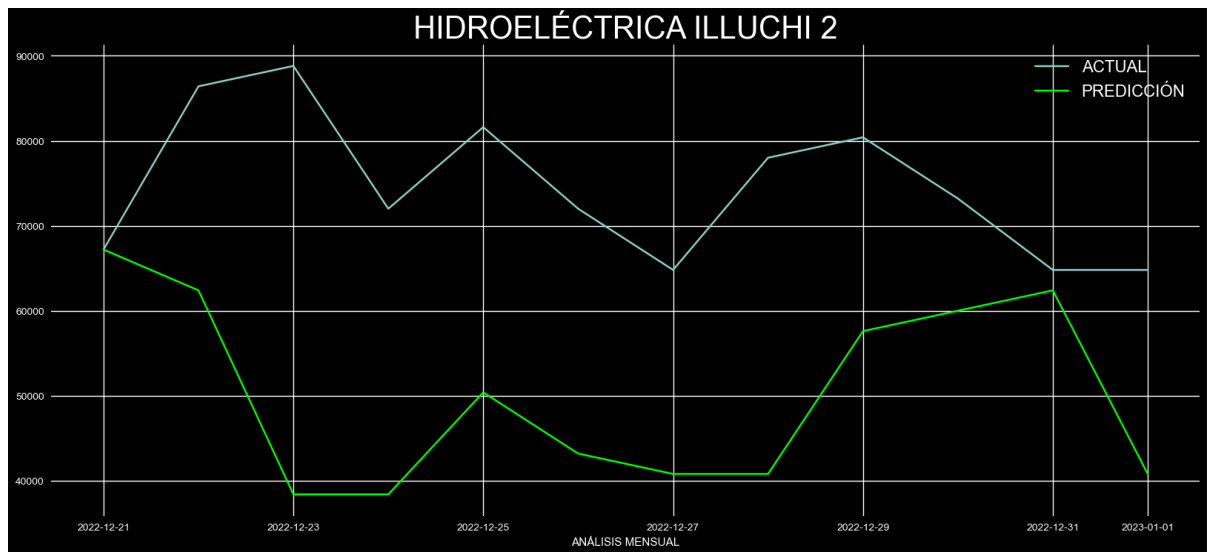
En la parte inferior izquierda de Figura 5.21 y 5.22, el correlograma, también conocido como gráfico ACF, muestra que los errores residuales no están autocorrelacionados. Cualquier autocorrelación implicaría que existe algún patrón en los errores residuales que no se explican en el modelo. Por lo tanto, deberá buscar más X (predictores) en el modelo.



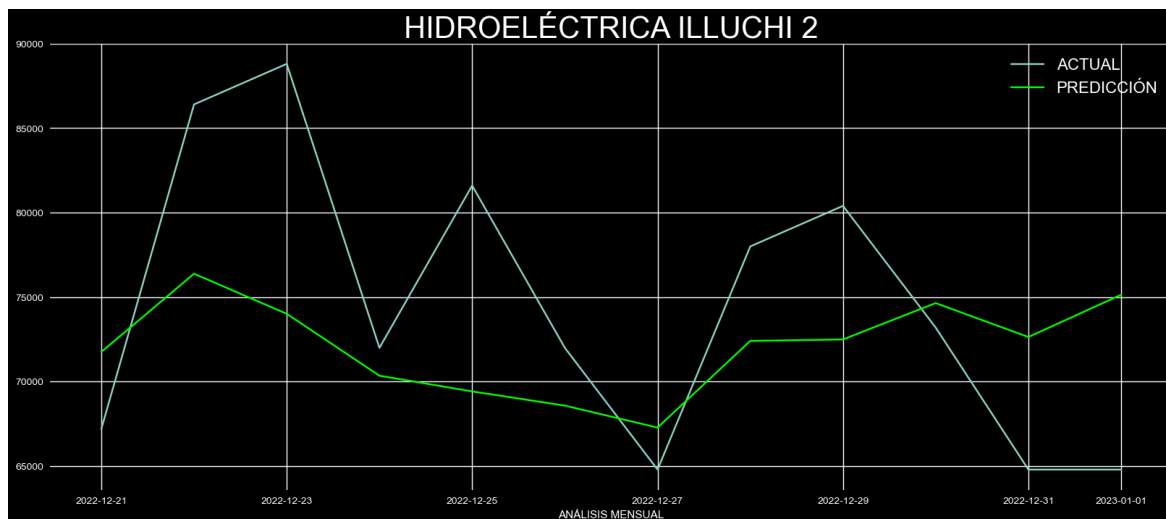


**Figura 5.22.** Los Errores residuales escenarios 2.

En la Figura 5.23 y la Figura 5.24 demuestran la predicción con los valores superior como inferior, que se verifica que no se comparan con los valores actuales el cual indica que independiente por los valores no se acercan a la predicción ya que se observa la diferenciación de la producción por la gráfica de control.



**Figura 5.23.** Relación entre la predicción del escenario 1 frente al actual análisis mensual 2022.



**Figura 5.24.** Relación entre la predicción del escenario 2 frente al actual análisis mensual 2022.

## 5.8 COMPARACIÓN DEL MODELO ARIMA FRENTE A REDES NEURONALES ARTIFICIALES (RNA)

### 5.8.1 Pronóstico de la producción de energía mediante modelo matemático ARIMA

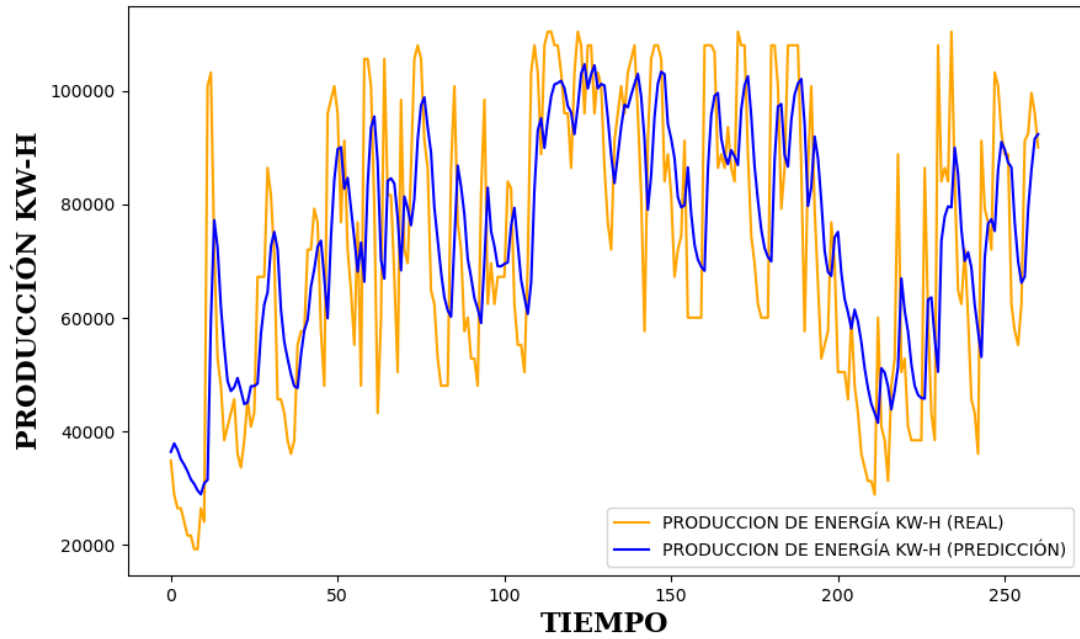
Para el pronóstico de la producción se realiza la tabla donde se realiza las medias de error para determinar la validación del modelo donde se detalla en la tabla 5.1 de cada uno de los meses de los años predecidos.

**Tabla 5.1.** Pronóstico de la producción de energía mediante modelo matemático ARIMA 2022.

VALIDACIÓN DE LA PRODUCCIÓN DE ENERGIA MEDIANTE MODELO MATEMÁTICO ARIMA										
n	Mes	Producción Real	P. Pronosticada	Error Abs.	Error % Abs.	∑ Error Abs.	MAD	Error Nor.	∑ Error Nor.	TS
1	Enero	33600,00	37797,75	4197,752125	12,49%	4197,75213	4197,8	-4197,752	4197,7521	-1
2	Febrero	26400,00	27082,08	682,0754415	2,58%	4879,82757	2439,9	-682,0754	4879,8276	-2
3	Marzo	38400,00	37996,62	403,3826123	1,05%	5283,21018	1761,1	403,38261	-4476,445	-2,5
4	Abril	71400,00	67034,95	4365,054816	6,11%	9648,26499	2412,1	4365,0548	111,39014	0,0
5	Mayo	76800,00	80311,11	3511,108968	4,57%	13159,374	2631,9	-3511,109	3622,4991	-1,4
6	Junio	82200,00	80219,60	1980,401666	2,41%	15139,7756	2523,3	1980,4017	1642,0974	-0,7
7	Julio	96000,00	100376,62	4376,616352	4,56%	19516,392	2788,1	-4376,616	6018,7138	-2,2
8	Agosto	86400,00	83838,24	2561,764566	2,97%	22078,1565	2759,8	2561,7646	3456,9492	-1,3
9	Septiembre	58800,00	64943,47	6143,474906	10,45%	28221,6315	3135,7	-6143,475	9600,4241	-3,1
10	Octubre	52800,00	56484,04	3684,038223	6,98%	31905,6697	3190,6	-3684,038	13284,462	-4,2
11	Noviembre	91200,00	87524,01	3675,98793	4,03%	35581,6576	3234,7	3675,9879	9608,4744	-3
12	Diciembre	77400,00	77066,49	333,5085996	0,43%	35915,1662	2992,9	333,5086	9274,9658	-3,1
				SUMA %	58,63%					
				n	12					
				MAPE	5%					

Como se puede observar el MAPE que se encuentra la tabla 5.1, es del 5% el cual es un porcentaje adecuado para la validación del modelo ARIMA.

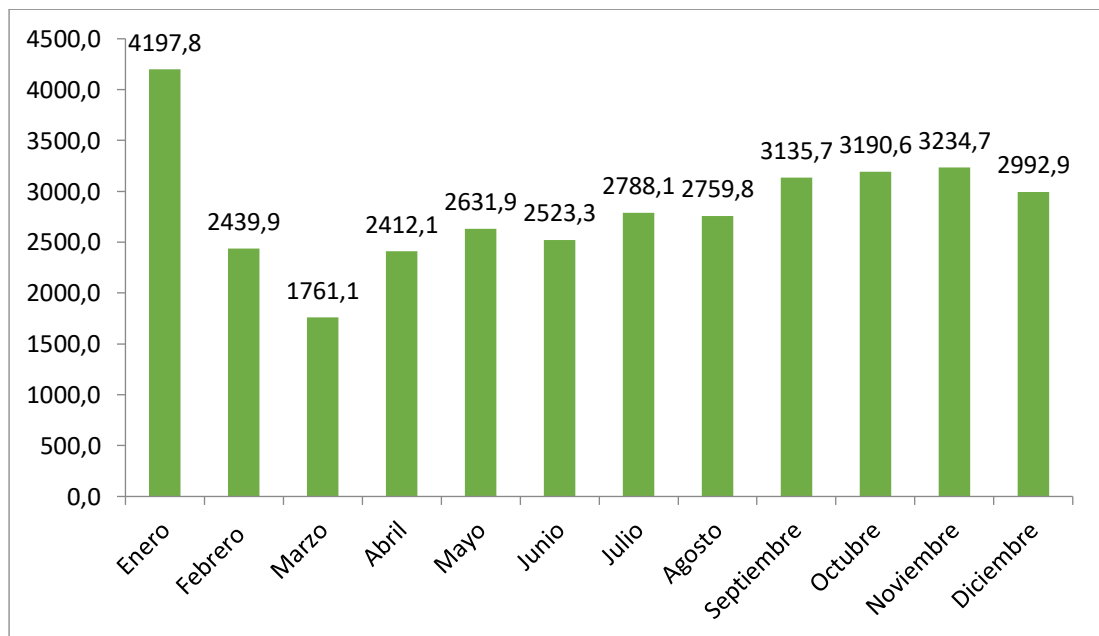
Para la comparación de modelo ARIMA y las redes Neuronales Artificiales se lo realiza conjuntamente con las gráficas una comparación. La cual determina que el método ARIMA Figura 5.25, tiene mayor coincidencia para la predicción realizada ya que al tener variables aprovecha los rezagos de la serie y otros parámetros que se observa.



**Figura 5.25.** Gráfica de producción actual y la predicción con el método ARIMA 2022.

### 5.8.1.1 MAD (Desviación Media Absoluta)

Representa el error promedio del pronóstico en valores absolutos. Mide la dispersión de un valor observado que se aleja del valor esperado. Se obtiene mediante el valor absoluto de las diferencias entre la producción real y la pronosticada, dividida entre la cantidad de errores [43].



**Figura 5.26.** Gráfica del MAD con modelo ARIMA 2022.

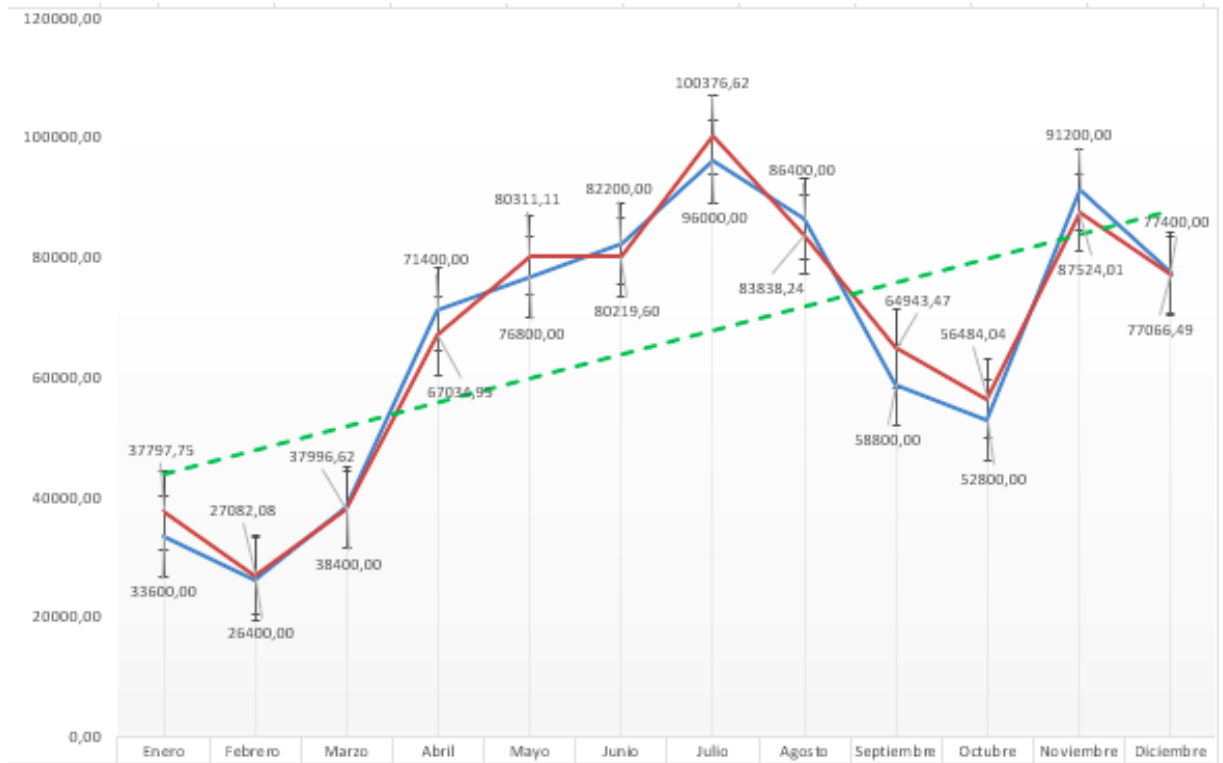


Figura 5.27. Producción Real frente al método ARIMA 2022.

### 5.8.2 PRONÓSTICO DE LA PRODUCCIÓN DE ENERGÍA MEDIANTE RNA-LSTM

Para el pronóstico de la producción de energía mediante RNA-LSTM, se debe evaluar con los errores para determinar si es favorable la predicción como se indica la tabla 5.2.

Mientras que la gráfica 5.28, de la potencia actual con la predicción con el modelo de redes neuronales. Con la visualización de los gráficos se determina que el óptimo y parecido es el método ARIMA el cual tiene la aproximación de los datos.

Como se puede observar el MAPE tiene una media de error del 11% e indica que tiene un significativo margen de error.

Tabla 5.2. Pronóstico de la demanda de energía mediante modelo RNA-LSTM 2022.

VALIDACIÓN DE LA PRODUCCIÓN DE ENERGIA MEDIANTE MODELO MATEMÁTICO ARIMA																
n	Mes	Producción Real	P. Pronosticada	Error Abs.	Error % Abs.	∑ Error Abs.	MAD	Error Nor.	∑ Error Nor.	TS						
1	Enero	33600,00	45058,48	11458,48437	<b>34,10%</b>	11458,4844	<b>11458,5</b>	-11458,48	-	-1						
2	Febrero	26400,00	27500,36	1100,359375	<b>4,17%</b>	12558,8437	<b>6279,4</b>	-1100,359	-	-2						
3	Marzo	38400,00	31023,11	7376,890625	<b>19,21%</b>	19935,7344	<b>6645,2</b>	7376,8906	-	0,8						
4	Abril	71400,00	72801,73	1401,734375	<b>1,96%</b>	21337,4687	<b>5334,4</b>	-1401,734	-	1,2						
5	Mayo	76800,00	78978,63	2178,625	<b>2,84%</b>	23516,0937	<b>4703,2</b>	-2178,625	-	1,9						
6	Junio	82200,00	86976,56	4776,558594	<b>5,81%</b>	28292,6523	<b>4715,4</b>	-4776,559	-	2,9						
7	Julio	96000,00	95121,79	878,2109375	<b>0,91%</b>	29170,8633	<b>4167,3</b>	878,21094	-	-3						
8	Agosto	86400,00	86416,45	16,45312495	<b>0,02%</b>	29187,3164	<b>3648,4</b>	-16,45312	-	3,5						
9	Septiembre	58800,00	53229,38	5570,617187	<b>9,47%</b>	34757,9336	<b>3862,0</b>	5570,6172	-	1,8						
10	Octubre	52800,00	62267,31	9467,3125	<b>17,93%</b>	44225,2461	<b>4422,5</b>	-9467,312	-	3,7						
11	Noviembre	91200,00	78598,64	12601,36328	<b>13,82%</b>	56826,6094	<b>5166,1</b>	12601,363	-	0,8						
12	Diciembre	77400,00	97604,07	20204,07031	<b>26,10%</b>	77030,6797	<b>6419,2</b>	-20204,07	-	3,8						
<table border="1"> <tr> <td>SUMA %</td> <td>136,35%</td> </tr> <tr> <td>n</td> <td>12</td> </tr> <tr> <td>MAPE</td> <td><b>11%</b></td> </tr> </table>					SUMA %	136,35%	n	12	MAPE	<b>11%</b>						
SUMA %	136,35%															
n	12															
MAPE	<b>11%</b>															

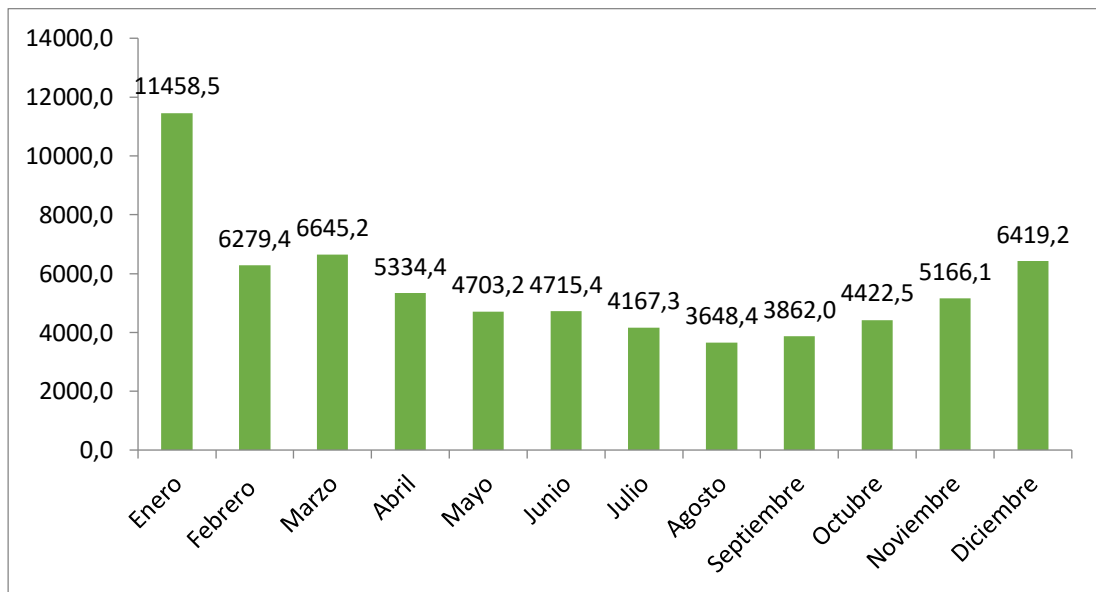
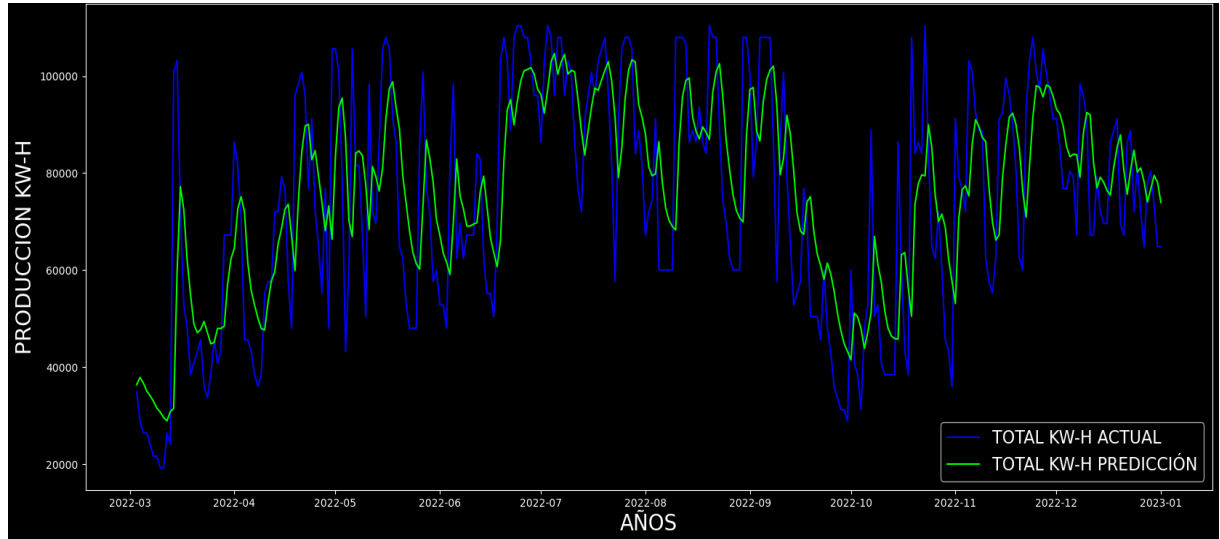


Figura 5.28. Gráfica del MAD con RNA-LSTM 2022.

Al momento de determinar los valores dados por Python se determina que la RNA, no relaciona con los datos obtenidos en el transcurso de la predicción con relación del método ARIMA el cual se asemeja a los datos que se tiene en la base de datos, determinando que en nuestro caso el modelo ARIMA.



**Figura 5.29.** Gráfica de producción actual y la predicción con el método RNA 2022.

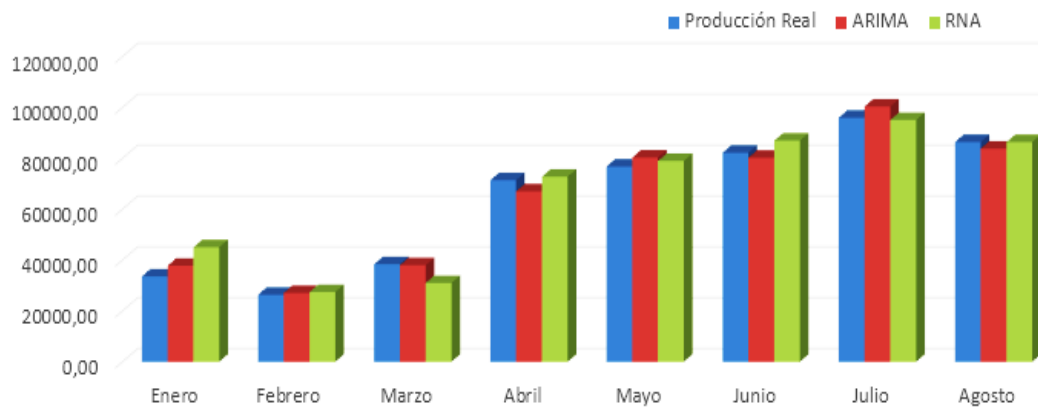
### 5.9 COMPORTAMIENTO DE PRODUCCIÓN KW PARA SU VALIDACIÓN

El comportamiento de la producción para su validación se determina mediante la relación de los datos actuales con los datos generados por Python, en la Figura 5.30, se puede observar que RNA-LSTM aumenta considerablemente frente a la demanda real y en el caso del método ARIMA no tiene mucha diferenciación en cuanto a la demanda real.

Esto indica que el método ARIMA es válido debido a que las demandas en intervalos de meses como se puede observar en la tabla 5.2, el MAPE es más alto a comparación de la tabla 5.1 del modelo ARIMA ya que es el resultado de las sumas de los errores absolutos de cada mes dividido por el número de meses.

**Tabla 5.3.** Comparación de la producción Real con el modelo ARIMA y RNA-LSTM 2022.

Mes	Producción Real	ARIMA	RNA-LSTM
Enero	33600,00	37797,75	45058,48
Febrero	26400,00	27082,08	27500,36
Marzo	38400,00	37996,62	31023,11
Abril	71400,00	67034,95	72801,73
Mayo	76800,00	80311,11	78978,63
Junio	82200,00	80219,60	86976,56
Julio	96000,00	100376,62	95121,79
Agosto	86400,00	83838,24	86416,45
Septiembre	58800,00	64943,47	53229,38
Octubre	52800,00	56484,04	62267,31
Noviembre	91200,00	87524,01	78598,64
Diciembre	77400,00	77066,49	97604,07



**Figura 5.30.** Gráfica de la producción actual con la predicción con el método RNA-LSTM y ARIMA enero-agosto 2022.

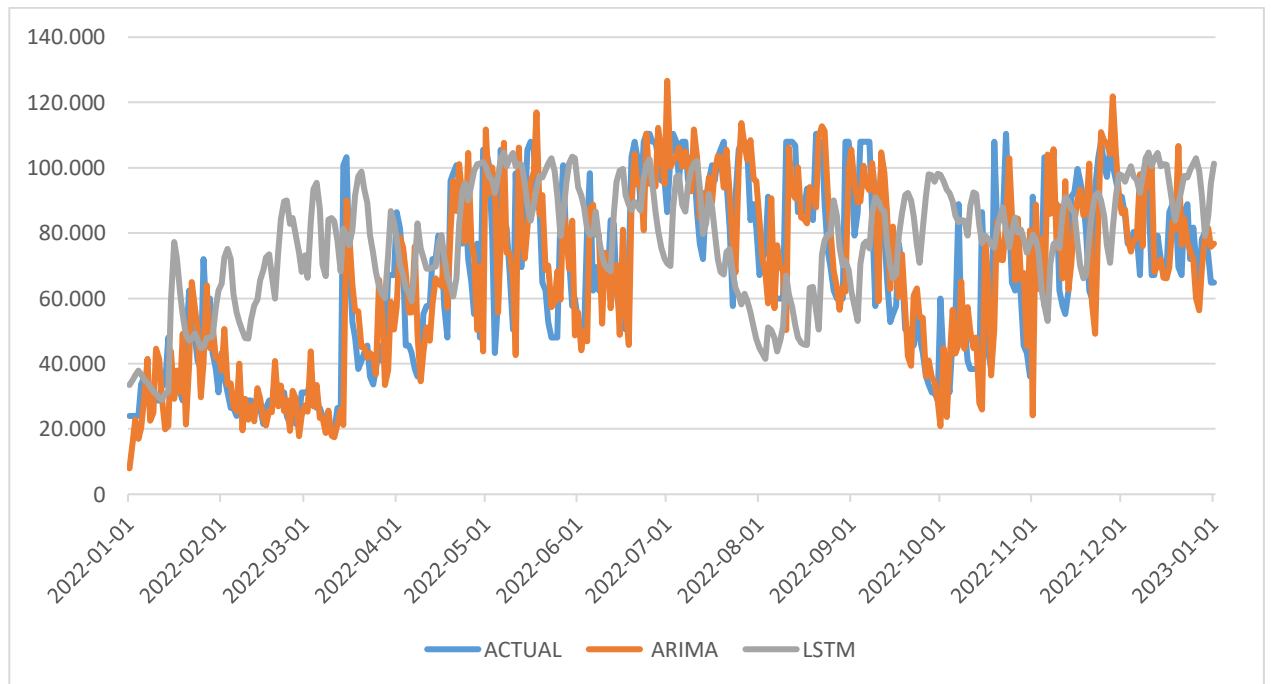
Los pronósticos de las series realizadas mediante modelos ARIMA Y RNA en niveles se muestran que puede observar que el modelo obtenido por RNA pronostica los datos en las primeras 60 observaciones y después pierde la tendencia; mientras que, después, el modelo ARIMA sigue más de cerca la tendencia original [44]



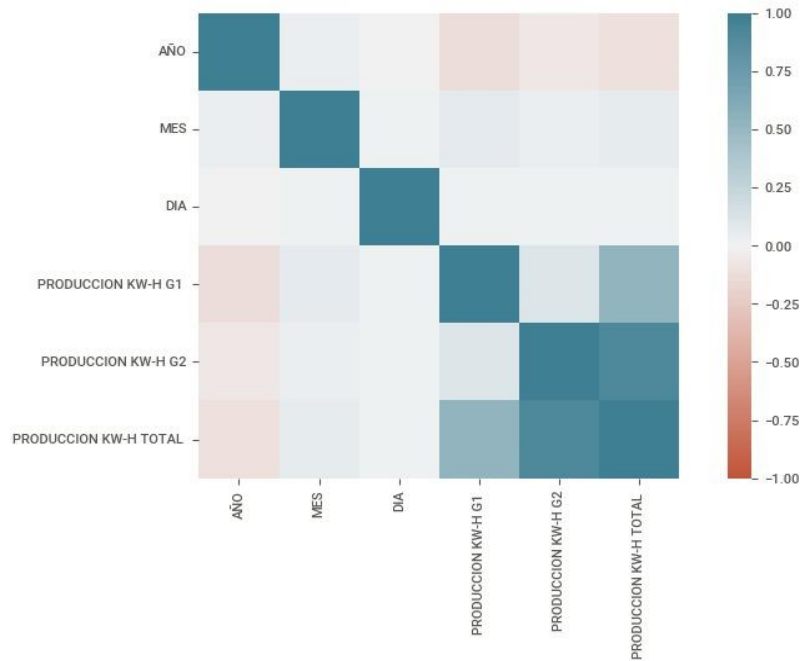
**Tabla 5.3.** Comparación de errores modelo ARIMA y el modelo RNA-LSTM 2022.

Mes	Errores con el modelo ARIMA				Errores con el modelo RNA-LSTM			
	Error Abs.	Error % Abs.	$\Sigma$ Error Abs.	MAD	Error Abs.	Error % Abs.	$\Sigma$ Error Abs.	MAD
Enero	4197,752125	12,49%	4197,752125	4197,8	11458,4844	34,10%	11458,48437	11458,5
Febrero	682,0754415	2,58%	4879,827567	2439,9	1100,35937	4,17%	12558,84375	6279,4
Marzo	403,3826123	1,05%	5283,210179	1761,1	7376,89063	19,21%	19935,73437	6645,2
Abril	4365,054816	6,11%	9648,264995	2412,1	1401,73438	1,96%	21337,46875	5334,4
Mayo	3511,108968	4,57%	13159,37396	2631,9	2178,625	2,84%	23516,09375	4703,2
Junio	1980,401666	2,41%	15139,77563	2523,3	4776,55859	5,81%	28292,65234	4715,4
Julio	4376,616352	4,56%	19516,39198	2788,1	878,210938	0,91%	29170,86328	4167,3
Agosto	2561,764566	2,97%	22078,15655	2759,8	16,4531249	0,02%	29187,31641	3648,4
Septiembre	6143,474906	10,45%	28221,63145	3135,7	5570,61719	9,47%	34757,93359	3862,0
Octubre	3684,038223	6,98%	31905,66967	3190,6	9467,3125	17,93%	44225,24609	4422,5
Noviembre	3675,98793	4,03%	35581,6576	3234,7	12601,3633	13,82%	56826,60937	5166,1
Diciembre	333,5085996	0,43%	35915,1662	2992,9	20204,0703	26,10%	77030,67969	6419,2
	SUMA %	58,63%			SUMA %	136,35%		
	n	12			n	12		
	MAPE	5%			MAPE	11%		

Para la validación del modelo ARIMA, se realizó la comparación con el método RNA-LSTM con los mismos datos depurados y se constató que el modelo ARIMA obtuvo un menor error frente al modelo RNA-LSTM, por medio del MAPE y el MAD correspondiente al año 2022 tabla 5.3.

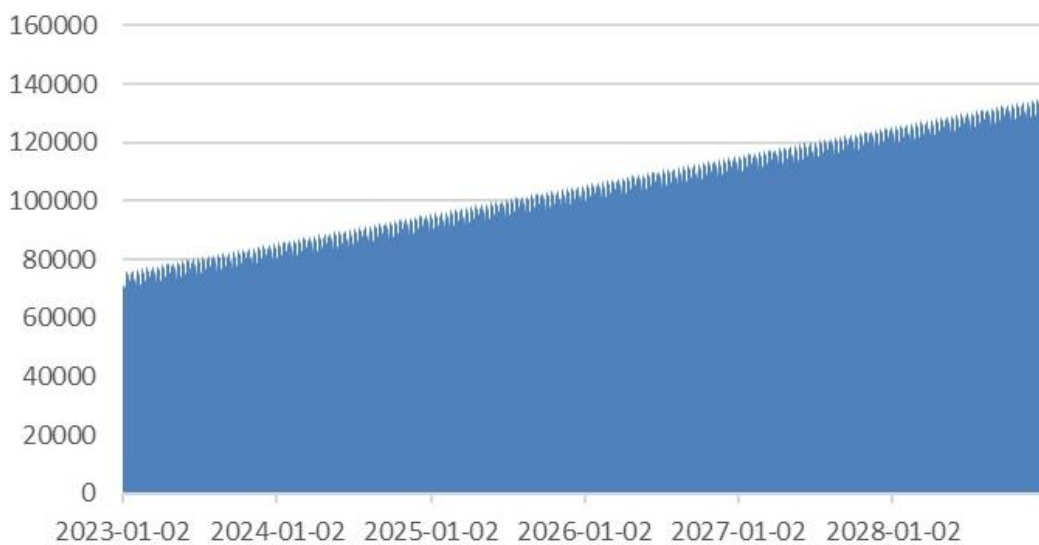


**Figura 5.31.** Gráfica de la producción actual con la predicción con el método RNA-LSTM y ARIMA 2022-2023.



**Figura 5.32.** Mapa de calor de la producción de la central ILLUCHI II.

El mapa de calor de la Figura 5.32 muestra la relación entre múltiples variables, las áreas marcadas en azul permiten ver que están perfectamente correlacionadas, lo cual es una de las principales razones para elegir las variables que ayudan a hacer predicciones, y también se puede analizar el comportamiento de la producción general del grupo en relación con la producción de electricidad mediante un mapa de calor.



**Figura 5.33.** Proyección de producción de la central ILLUCHI II desde el 2023-2028.

## 6 CONCLUSIONES Y RECOMENDACIONES.

### 6.1 CONCLUSIONES

Del trabajo de investigación realizado a los “Escenarios de producción de energía eléctrica en una pequeña central hidroeléctrica utilizando series sintéticas”, se presenta las siguientes conclusiones:

En virtud de lo investigado y expuesto en el estado del arte, es posible caracterizar el comportamiento de las series sintéticas de la producción de energía de una pequeña central hidroeléctrica en base a los datos históricos de generación, se ha tomado como escenario de la investigación el valor más alto y bajo de acuerdo con la gráfica de control del diagrama de Shewhart, lo que permite determinar cuándo una variación no es normal.

El análisis estadístico permitió tomar una decisión del modelo escogido para la generación de la serie sintética, el cual cumple con los parámetros requeridos de normalidad, varianza constante y aleatoriedad, teniendo el modelo ARIMA(0, 1, 0) (2, 1, 0), con un tiempo de 26.009 segundos, con esto no se tiene que estar probando uno a uno los posibles conjuntos y refleja el escenario adecuado del método.

Se obtuvieron medias de error con el modelo ARIMA del 5% frente a la producción real mientras que la RNA-LSTM fue del 11%, lo que indica que este modelo pronostica los datos en las primeras 60 observaciones y después pierde la tendencia; mientras que el modelo ARIMA sigue más de cerca la serie original.

Los resultados de la simulación con el modelo ARIMA, comparados con los datos observados, muestran un ajuste adecuado de los valores mínimos y máximos. Esto permite concluir que, aunque este modelo no simula el comportamiento exacto en el tiempo, es una buena herramienta para generar las series sintéticas para aproximar eventos mínimos y máximos.

En síntesis, mediante los modelos de proyección ARIMA y predicción RNA LSTM estadística permiten mediante la simulación generar series sintéticas para tener aproximaciones de los comportamientos de la producción de energía de una pequeña central hidroeléctrica, estas metodologías constituyen una herramienta, para observar a los cambios en la generación evidenciándose una gran ventaja para la planificación y operación de la central.

## 6.2 RECOMENDACIONES

Se recomienda tomar el estudio del caudal para una investigación consecuente a la de la presente investigación.

Los datos proporcionados por parte de ELEPCO S.A deben llevar un tratamiento exhaustivo en la depuración de los datos, y elaborar nuevas bases de datos que permitan ser llamadas de manera eficaz en el programa.

El uso de Python como software en la presente investigación trae múltiples ventajas, siempre y cuando se descargue de manera correcta y se realice el llamado de la librería de manera puntual.

El método ARIMA tiene muchas varianzas por ende se recomienda para futuras investigaciones indagar las posibles formas de comparar entre el método ARIMA, de igual manera en el uso de software, se puede realizar una comparación con código R entre otros.

El tratamiento de la base de datos es primordial a la hora de trabajar con un modelo predictivo, el cual se basa en los históricos para demostrar la estacionariedad de estos para lograr una mayor efectividad en la predicción de la producción de energía eléctrica.

## 7 REFERENCIAS

- [1] ARCONEL, “Estadística del Sector Eléctrico Ecuatoriano 2015,” p. 300, 2016.
- [2] G. Salazar and B. Panchi, “Análisis de la Evolución de la Demanda Eléctrica en el Ecuador Considerando el Ingreso de Proyectos de Eficiencia Energética,” *Revista Politécnica*, vol. 33, no. 1, 2014, [Online]. Available: [https://revistapolitecnica.epn.edu.ec/ojs2/index.php/revista\\_politecnica2/article/view/218](https://revistapolitecnica.epn.edu.ec/ojs2/index.php/revista_politecnica2/article/view/218)
- [3] IIGE, “Balance Energético Nacional 2021,” *Ministerio de Energía y Minas*, p. 181, 2022.
- [4] C. Pacheco and R. Herrera, “Universidad técnica de cotopaxi,” *Universidad técnica de cotopaxi*, vol. 1, p. 101, 2021, [Online]. Available: <http://repositorio.utc.edu.ec/bitstream/27000/4501/1/PI-000727.pdf>
- [5] L. Quisimalín, “Predicción de la Demanda Eléctrica en barra de 69kV del Sistema Eléctrico de Subtransmisión de la Empresa Eléctrica Riobamba S.A,” p. 150, 2012, [Online]. Available: <http://bibdigital.epn.edu.ec/bitstream/15000/14623/1/CD-6793.pdf>
- [6] Samuel Ramirez Castaño, “Redes de Distribución de Energía,” *Redes de Distribución de Energía*, vol. 3, p. 926, 2012, [Online]. Available: <http://blog.espol.edu.ec/econde/files/2012/08/libro-redes-de-distribucion.pdf>
- [7] Fredy O. Pèrez Ramirez, *Introducción a las series de tiempo. Métodos paramétricos*. 2007.
- [8] M. P. G. Casimiro, “Análisis de series temporales: Modelos ARIMA,” *Departamento de Economía Aplicada III (Econometría y Estadística) Facultad de Ciencias Económicas y Empresariales Universidad del País Vasco (UPV-EHU)*, 1992.
- [9] C. M. Chinlli, “Modelización de Series Temporales modelos clásicos y SARIMA,” *Universidad de Granada Departamento de Estadística e Investigación Operativa*, 2021.
- [10] C. Allaapetukhina, “Statistics and Computing Applied Time Series Analysis and Forecasting with Python.”
- [11] P. Carlos, “Jorge Guerra Rodr Fundamentos y variantes de los modelos ARIMA para el análisis de series temporales . Aplicación a la estadística universitaria . for time series analysis . Applications to university,” 2022.
- [12] C. E. Catalán, “SERIES TEMPORALES,” 2000.
- [13] John Villavicencio, “Introducción a Series de Tiempo,” 2011.

- [14] Ángela Fernández Pascual, “Predicción de energía eólica con modelos autorregresivos,” Madrid, 2019. Accessed: Feb. 01, 2023. [Online]. Available: [https://repositorio.uam.es/bitstream/handle/10486/689127/torrijos\\_moreno\\_jaime\\_tfg.pdf?sequence=1](https://repositorio.uam.es/bitstream/handle/10486/689127/torrijos_moreno_jaime_tfg.pdf?sequence=1)
- [15] P. Gonz, *Análisis de series temporales : Modelos ARIMA*.
- [16] S. de la F. Fernandez, “Metodología De Box - Jenkins,” *Facultad Ciencias Economicas y Empresariales*, vol. 1, 2010.
- [17] Mena Vélez Luis Adrian and Calo Machay Diego David, “Modelo ARIMA para el pronóstico de la producción de cacao en el Ecuador para el periodo 2020-2024,” 2020.
- [18] F. Izaurieta and C. Saavedra, “Redes Neuronales Artificiales,” *Charlas de fisica*, pp. 1–15, 1999, doi: 10.1016/S0210-5691(05)74198-X.
- [19] A. F. M. VEGA, “Predicción de la demanda de energía eléctrica en la producción de petróleo de los campos de PETROAMAZONAS EP utilizando redes neuronales artificiales,” *ESCUELA POLITÉCNICA NACIONAL FACULTAD DE INGENIERÍA ELÉCTRICA Y ELECTRÓNICA*, 2020.
- [20] Xavier Edison Yambay Chafla, “Análisis del rendimiento de algoritmos de entrenamiento de redes neuronales artificiales, aplicadas al modelamiento dinámico de represas hidroeléctricas, mediante el error de predicción del nivel de embalse de agua,” RIOBAMBA, 2019. Accessed: Feb. 01, 2023. [Online]. Available: <http://dspace.esPOCH.edu.ec/bitstream/123456789/9697/1/20T01148.pdf>
- [21] M. C. Manchado, “Implementación de redes neuronales recurrentes en Python.,” 2018.
- [22] J. M. González de Buitrago, “6. Estadística básica,” *Técnicas y Métodos de Laboratorio Clínico*, pp. 73–83, 2010, doi: 10.1016/b978-84-458-2029-2.50006-9.
- [23] D. A. Alvarez Chiriboga, “Modelo de predicción de la producción de energía de la Central Hidroeléctrica Coca Codo Sinclair, basado en técnicas de aprendizaje computacional,” *UNIVERSIDAD DE LAS FUERZAS ARMADAS*, pp. 1–100, 2020.
- [24] Adriana Marcela Ariza Ramírez, “Métodos utilizados para el pronóstico de demanda de energía eléctrica en sistemas de distribución,” *UNIVERSIDAD TECNOLÓGICA DE PEREIRA FACULTAD DE INGENIERÍAS: ELÉCTRICA, ELECTRÓNICA, FÍSICA Y CIENCIAS DE LA COMPUTACIÓN*, 2013, Accessed: Feb. 02, 2023. [Online]. Available:

<https://repositorio.utp.edu.co/server/api/core/bitstreams/ee2fb283-daa7-4987-b747-faa1e42e9711/content>

- [25] EPM Information Development Team, “Previsión y descripciones estadísticas de Planificación predictiva Feedback,” *Oracle Fusion Cloud EPM Trabajo con Planning*, 2014.
- [26] L. Benites, “riterio de información de Akaike: definición, fórmulas,” 2021. <https://statologos.com/criterio-de-informacion-de-akaikes/>
- [27] U. N. D. C. CALLAO and FACULTAD DE CIENCIAS ADMINISTRATIVAS, “Desviación Estandar,” *Universidad Nacional Del Callao Facultad de Ciencias Administrativas*, pp. 229–271, 2006, [Online]. Available: [http://www.unac.edu.pe/documentos/organizacion/vri/cdcitra/Informes\\_Finales\\_Investigacion/IF\\_JUNIO\\_2012/IF\\_CALDERON\\_OTOYA\\_FCA/capitulo\\_6\\_y\\_7.pdf](http://www.unac.edu.pe/documentos/organizacion/vri/cdcitra/Informes_Finales_Investigacion/IF_JUNIO_2012/IF_CALDERON_OTOYA_FCA/capitulo_6_y_7.pdf)
- [28] Santiago. Fuente, “Modelos De Análisis De La Varianza,” *Universidad Autónoma de Madrid Facultad Ciencias Económicas y Empresariales Departamento de Economía Aplicada*, p. 107, 2013, [Online]. Available: <http://www.estadistica.net/ECONOMETRIA/ANALISIS-VARIANZA/analisis-varianza.pdf>
- [29] S. Seabold and J. Perktold, “Statsmodels: Econometric and Statistical Modeling with Python,” no. Scipy, pp. 92–96, 2010.
- [30] Rob J Hyndman y George Athanasopoulos, *Pronóstico: principios y práctica*, Segunda ed. 2018.
- [31] B. Baka, *Python data structures and algorithms: improve the performance and speed of your applications*. 2017. Accessed: Feb. 02, 2023. [Online]. Available: <https://edu.anarcho-copy.org/Programming%20Languages/Python/Python%20Data%20Structures%20and%20Algorithms.pdf>
- [32] Aimacaña Chuquimarca Jefferson Paúl and Columba Guanoluisa Alexander Ramiro, “Análisis comparativo de algoritmos de Machine Learning para la detección de plagas en los cultivos representativos de la sierra ecuatoriana,” *Universidad Central del Ecuador*, 2021, Accessed: Feb. 02, 2023. [Online]. Available: <http://www.dspace.uce.edu.ec/handle/25000/24228>
- [33] JUAN JOSÉ MARTÍN MIRALLES, “Librería de Python para el trazamiento y la animación de trayectorias de GPS almacenadas en ficheros con formato GPX,” *Escola Politècnica*

*SuperiorUniversitat de les Illes Balears*, 2017, Accessed: Feb. 02, 2023. [Online]. Available: [https://dspace.uib.es/xmlui/bitstream/handle/11201/151783/Memoria\\_EPSU0860.pdf?sequence=1&isAllowed=y](https://dspace.uib.es/xmlui/bitstream/handle/11201/151783/Memoria_EPSU0860.pdf?sequence=1&isAllowed=y)

- [34] Isabela San José García, “Diseño e implementación de un módulo Python para representar datos geográficos en Dataframes,” Leganés, 2019. Accessed: Feb. 02, 2023. [Online]. Available: <https://e-archivo.uc3m.es/handle/10016/31974>
- [35] Omar Caja García, “Librería Python para el aprendizaje y la implementación de redes neuronales,” 2019. Accessed: Feb. 02, 2023. [Online]. Available: <https://riunet.upv.es/handle/10251/152226>
- [36] Bastidas Cashicana Wilson Roberto and Moya Cabezas Angelo Mauricio, “Desarrollo de un sistema de predicción con redes neuronales artificiales para determinar la demanda de energía eléctrica de una central hidroeléctrica,” *Universidad Técnica de Cotopaxi, Ingeniería Eléctrica en Sistemas Eléctricos de Potencia*, 2021.
- [37] Edgar Roberto Salazar Achig, “Simulación de los tanques de almacenamiento de la central Illuchi de la empresa eléctrica provincial Cotopaxi ELEPCO S.A mediante dinámica de fluidos computacional,” pp. 1–130, 2022.
- [38] Fernando Campos, “Análisis de series temporales en R. ARIMA,” 2018. <https://www.diegocalvo.es/analisis-de-series-temporales-en-r-arima/> (accessed Feb. 02, 2023).
- [39] D. M. Baer, “Escuela Politécnica Superior Memoria del Trabajo de Fin de Grado,” 2019.
- [40] F. Mauricio and C. Freire, “Estudio técnico-económico de la implementación del centro de control de las centrales hidroeléctricas ILLUCHI I e ILLUCHI II de ELEPCO S.A.,” 2007.
- [41] Subhasree Chatterjee, “Análisis de series temporales utilizando el modelo ARIMA en R,” 2018. <https://datascienceplus.com/time-series-analysis-using-arima-model-in-r/> (accessed Jan. 27, 2023).
- [42] Betancourt Diego Fernando, “Cómo hacer un gráfico de control: Ejemplo resuelto en calidad,” 2016. [www.ingenioempresa.com/grafico-de-control](http://www.ingenioempresa.com/grafico-de-control). (accessed Jan. 31, 2023).
- [43] M. Cristina, B. Gomez, and R. B. Palacio, “FOR-TSM: DESARROLLO DE UNA HERRAMIENTA DE PRONÓSTICOS CON MODELOS DE SERIES DE TIEMPO,” 2011.



- [44] Arturo Morales Castro and Iván Cruz Torres, “Pronóstico mediante redes neuronales artificiales y modelosm ARIMA,” *COMERCIO EXTERIOR* , vol. 58, 2018.


**INDICE DE ANEXOS**

**ANEXO A.** Información depurada de la base de datos desde el año 2018 hasta el año 2022....

**ANEXO B.** Estructura del método ARIMA en Python para la predicción. ...

**ANEXO C.** Datos pronosticados.....

**Anexo A: INFORME ANTI PLAGIO PROYECTO DE TITULACION**

<b>Facultad:</b>	Ciencias de la Ingeniería y Aplicadas												
<b>Carrera:</b>	Ingeniería Eléctrica												
<b>Nombre del docente evaluador que emite el informe:</b>	Ing. Salazar Achig Roberto Ms.C												
<b>Documento evaluado:</b>	Propuesta de Investigación presentado previo a la obtención del Título de Ingeniero Eléctrico en Sistemas Eléctricos de Potencia.												
<b>Autor del documento:</b>	Sr. Carrillo Carrillo Manuel Francisco Sr. Tapia Guerrero Jipson Danilo												
<b>Programa de similitud utilizado:</b>	Sistema URKUND												
<b>Porcentaje de similitud según el programa utilizado</b>	7%												
<b>Observaciones:</b> Calificación de originalidad atendido a los siguientes criterios: <ul style="list-style-type: none"> <li>• El documento cumple los criterios de originalidad, sin observaciones.</li> <li>• El documento cumple los criterios de originalidad, con observaciones.</li> <li>• El documento no cumple los criterios de originalidad.</li> </ul>	-X-  ---												
<b>Fecha de realización del informe:</b>	16/02/2023 14:38:00												
<b>Captura de pantalla del documento analizado:</b>													
<p><b>Document Information</b></p> <hr/> <table> <tr> <td><b>Analyzed document</b></td> <td>TESIS_CARRILLO_TAPIA.pdf (D158796002)</td> </tr> <tr> <td><b>Submitted</b></td> <td>2/16/2023 2:38:00 PM</td> </tr> <tr> <td><b>Submitted by</b></td> <td>Roberto</td> </tr> <tr> <td><b>Submitter email</b></td> <td>edgar.salazar7619@utc.edu.ec</td> </tr> <tr> <td><b>Similarity</b></td> <td>7%</td> </tr> <tr> <td><b>Analysis address</b></td> <td>edgar.salazar7619.utc@analysis.urkund.com</td> </tr> </table>		<b>Analyzed document</b>	TESIS_CARRILLO_TAPIA.pdf (D158796002)	<b>Submitted</b>	2/16/2023 2:38:00 PM	<b>Submitted by</b>	Roberto	<b>Submitter email</b>	edgar.salazar7619@utc.edu.ec	<b>Similarity</b>	7%	<b>Analysis address</b>	edgar.salazar7619.utc@analysis.urkund.com
<b>Analyzed document</b>	TESIS_CARRILLO_TAPIA.pdf (D158796002)												
<b>Submitted</b>	2/16/2023 2:38:00 PM												
<b>Submitted by</b>	Roberto												
<b>Submitter email</b>	edgar.salazar7619@utc.edu.ec												
<b>Similarity</b>	7%												
<b>Analysis address</b>	edgar.salazar7619.utc@analysis.urkund.com												
 <b>Ing. Salazar Achig Roberto Ms.C</b> <b>Director del Proyecto de la Propuesta de Investigación</b>													



### Document Information

Analyzed document	TESIS_CARRILLO_TAPIA.pdf (D158796002)
Submitted	2/16/2023 2:38:00 PM
Submitted by	Roberto
Submitter email	edgar.salazar7619@utc.edu.ec
Similarity	7%
Analysis address	edgar.salazar7619.utc@analysis.arkund.com

### Sources included in the report

<b>W</b>	URL: <a href="http://www.estadisticas.gobierno.pr/iepr/LinkClick.aspx?fileticket=4_BxecUaZmg%3D">http://www.estadisticas.gobierno.pr/iepr/LinkClick.aspx?fileticket=4_BxecUaZmg%3D</a> Fetched: 11/19/2020 3:24:21 PM		1
<b>SA</b>	<b>ARIMA GRUPO 1 (1).docx</b> Document ARIMA GRUPO 1 (1).docx (D77791654)		6
<b>SA</b>	<b>JannethYumiceba.pdf</b> Document JannethYumiceba.pdf (D129713252)		4
<b>SA</b>	<b>1446105915_TESIS SHIRLEY AGUIRRE final.docx</b> Document 1446105915_TESIS SHIRLEY AGUIRRE final.docx (D15905726)		1
<b>SA</b>	<b>Tesis- Monica_Erick.pdf</b> Document Tesis- Monica_Erick.pdf (D142464322)		1
<b>SA</b>	<b>UNIVERSIDAD TÉCNICA DE COTOPAXI / Tesis_Gonzalez_Hidalgo.pdf</b> Document Tesis_Gonzalez_Hidalgo.pdf (D128801482) Submitted by: edgar.salazar7619@utc.edu.ec Receiver: edgar.salazar7619.utc@analysis.arkund.com		1
<b>SA</b>	<b>Memoria TFM.pdf</b> Document Memoria TFM.pdf (D113709971)		2
<b>SA</b>	<b>1614705685_TESIS MARÍA ROSA REYES ACOSTA 03-03-2021.pdf</b> Document 1614705685_TESIS MARÍA ROSA REYES ACOSTA 03-03-2021.pdf (D97011935)		2

### Entire Document

**ANEXO B. BASE DE DATOS DESDE EL AÑO 2018 HASTA EL AÑO 2022.**

Date	kW
1/1/2018	4.200
2/1/2018	4.200
3/1/2018	3.450
4/1/2018	2.700
5/1/2018	2.700
6/1/2018	2.600
7/1/2018	2.750
8/1/2018	2.950
9/1/2018	4.000
10/1/2018	2.800
11/1/2018	2.800
12/1/2018	2.800
13/1/2018	2.500
14/1/2018	2.700
15/1/2018	4.000
16/1/2018	4.350
17/1/2018	3.800
18/1/2018	3.000
19/1/2018	4.000
20/1/2018	4.400
21/1/2018	4.350
22/1/2018	4.200
23/1/2018	4.000
24/1/2018	3.300
25/1/2018	3.350
26/1/2018	3.250
27/1/2018	3.200
28/1/2018	2.900
29/1/2018	2.650
30/1/2018	2.400
31/1/2018	2.750

**Figura A.1.** Datos depurados del año 2018 mes de enero.

1/1/2022	24000
2/1/2022	24000
3/1/2022	24000
4/1/2022	24000
5/1/2022	33600
6/1/2022	36000
7/1/2022	33600
8/1/2022	28800
9/1/2022	28800
10/1/2022	28800
11/1/2022	28800
12/1/2022	33600
13/1/2022	33600
14/1/2022	48000
15/1/2022	38400
16/1/2022	36000
17/1/2022	31200
18/1/2022	31200
19/1/2022	28800
20/1/2022	45600
21/1/2022	62400
22/1/2022	50400
23/1/2022	48000
24/1/2022	40800
25/1/2022	38400
26/1/2022	72000
27/1/2022	52800
28/1/2022	60000
29/1/2022	43200
30/1/2022	38400

**Figura A.2.** Datos depurados del año 2022 mes de enero.

**ANEXO B.** Estructura del método ARIMA en Python para la predicción.

```

# Manipulación y tratamiento de Datos
import numpy as np
import pandas as pd

# Visualización de datos
import plotly.express as px
import matplotlib.pyplot as plt
%matplotlib inline
plt.style.use('ggplot')

# Modelación Arima
from statsmodels.tsa.statespace.sarimax import SARIMAX
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
from statsmodels.tsa.seasonal import seasonal_decompose
from statsmodels.tsa.stattools import adfuller

# Métrica de Evaluación
from sklearn.metrics import mean_squared_error
from statsmodels.tools.eval_measures import rmse
from sklearn import metrics

# No presentar advertencias
import warnings
warnings.filterwarnings("ignore")

import seaborn as sns
sns.set()

from datetime import datetime
from datetime import timedelta
from datetime import date

from plotly import __version__
from plotly.offline import download_plotlyjs, init_notebook_mode, plot, iplot
import plotly.graph_objects as go
from plotly.subplots import make_subplots
import cufflinks as cf
init_notebook_mode(connected=True)
cf.go_offline()
import plotly.express as px
import chart_studio.plotly as py

pd.set_option('display.max_columns', None)

```

**Figura B.1.** Librerías ejecutadas en Python.

```

def evaluacion_metrica(y_true, y_pred):

    def mean_absolute_percentage_error(y_true, y_pred):
        y_true, y_pred = np.array(y_true), np.array(y_pred)
        return np.mean(np.abs((y_true - y_pred) / y_true)) * 100
    print('Resultados de la métrica de evaluación:-')
    print(f'MAE ES : {metrics.mean_absolute_error(y_true, y_pred)}')
    print(f'MAPE ES : {mean_absolute_percentage_error(y_true, y_pred)}')

```

**Figura B.2.** Librerías para determinar errores ejecutadas en Python.

**ANEXO C.** Tratamiento de datos depurados de la central IILLUCHI II.

```
df = pd . read_excel ( 'DATOS/df_kwh-total.xlsx' )
df
```

	FECHA	PRODUCCIÓN KW-H DIA
0	2018-01-01	101400.0
1	2018-01-02	102200.0
2	2018-01-03	85400.0
3	2018-01-04	67500.0
4	2018-01-05	67000.0
...	...	...
1822	2022-12-28	78000.0
1823	2022-12-29	80400.0
1824	2022-12-30	73200.0
1825	2022-12-31	64800.0
1826	2023-01-01	64800.0

1827 filas × 2 columnas

**Figura C.1.** Llamado de datos depurados.

```
d.f. _ información ()
<clase 'pandas.core.frame.DataFrame'>
RangeIndex: 1827 entradas, 0 a 1826
Columnas de datos (total 2 columnas):
# Columna Recuento no nulo Dtype
-----
0 FECHA 1827 no nulo datetime64[ns]
1 PRODUCCIÓN KW-H DIA 1827 non-null float64
tipos: datetime64[ns](1), float64(1)
uso de memoria: 28,7 KB

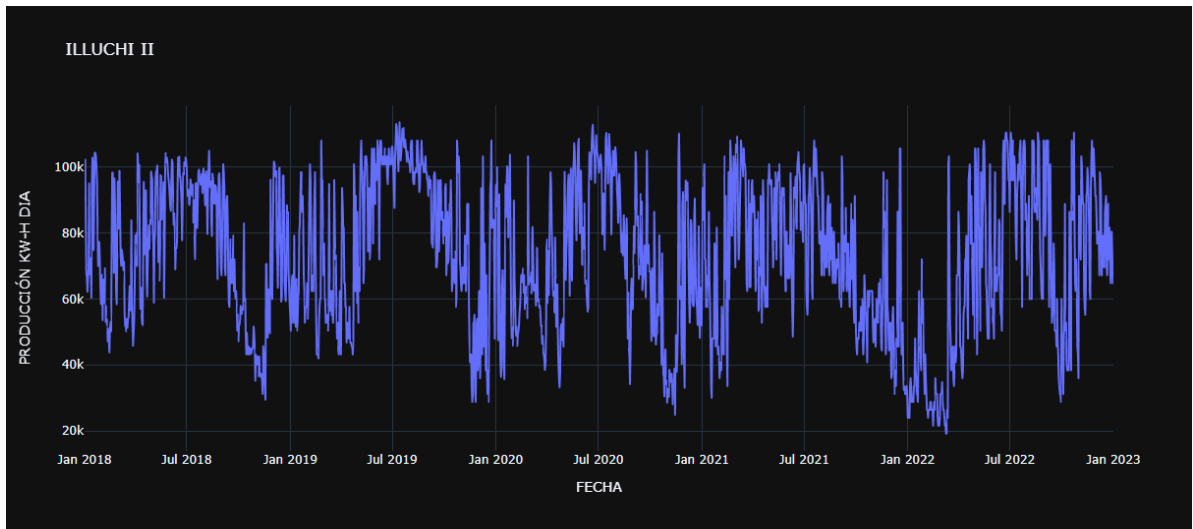
#Verificación de nans
df . es nulo () . valores _ cualquier ()
df . es nulo () . suma () . suma ()

d.f. _ es nulo () . suma ()

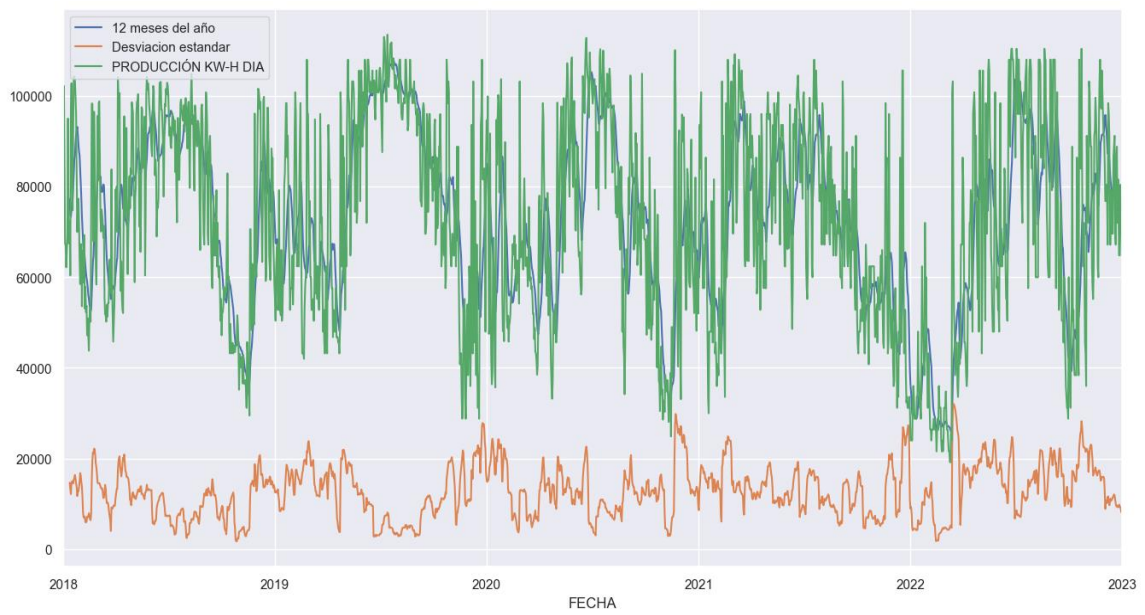
FECHA 0
PRODUCCIÓN KW-H DIA 0
tipo: int64

df = df . set_index ( "FECHA" )
df
```

**Figura C.1.** Información de los datos vacíos y espacios.

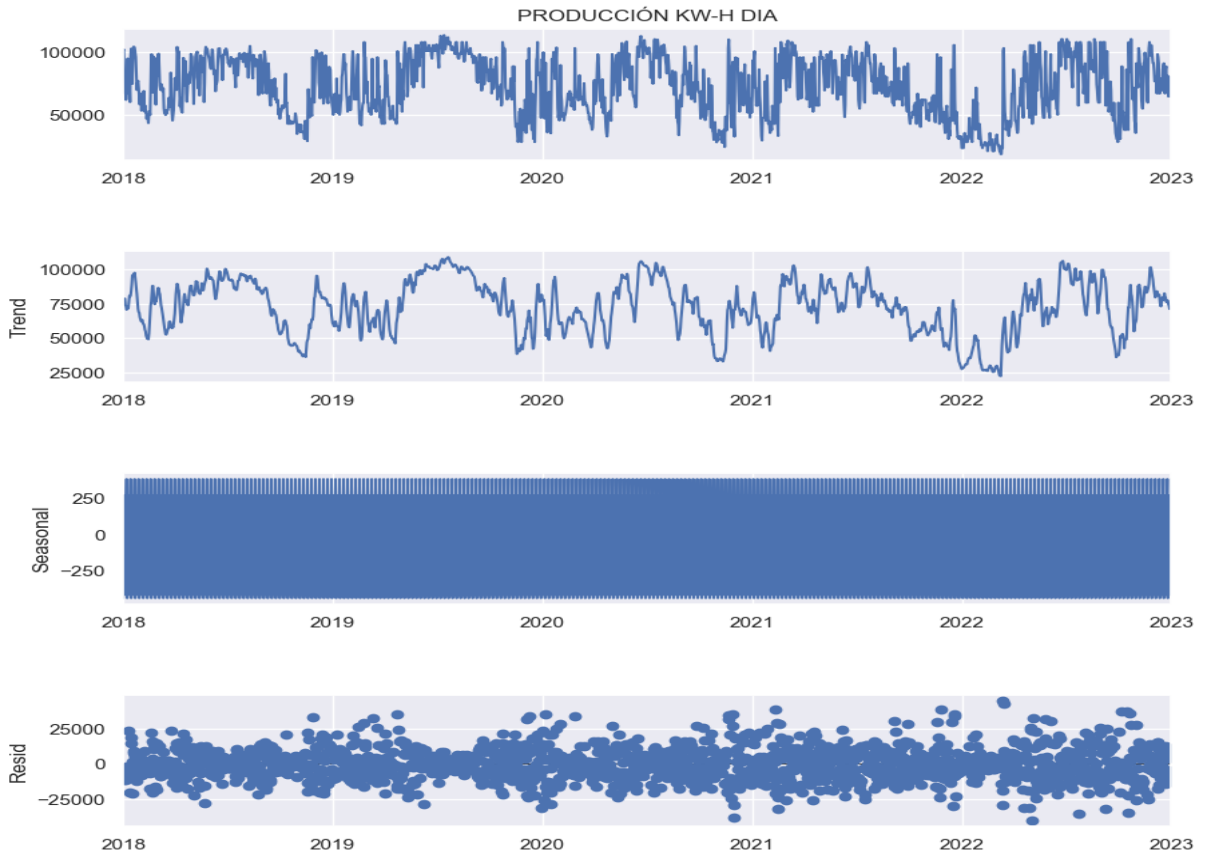


**Figura C.2.** Serie Temporal de la predicción junio2018-junio2023.



**Figura C.3.** Desviación Estándar por el intervalo de tiempo 2018-2023.





**Figura C.4.** Gráfica de la descomposición de los datos.

**ANEXO D. Prueba Dickey Fuller.**

```
def Prueba_Dickey_Fuller(series , column_name):
    print (f'Resultados de la prueba de Dickey-Fuller para columna: {column_name}')
    dftest = adfuller(series, autolag='AIC')
    dfoutput = pd.Series(dftest[0:4], index=['Test Estadístico','Valor P','Sin retrasos','Número de observaciones utilizadas'])
    for key,value in dftest[4].items():
        dfoutput['Valor crítico (%s)'%key] = value
    print (dfoutput)
    if dftest[1] <= 0.05:
        print("Conclusión:====>")
        print("Rechazar la hipótesis nula")
        print("Los datos son estacionarios")
    else:
        print("Conclusión:====>")
        print("No se puede rechazar la hipótesis nula")
        print("Los datos no son estacionarios")

Prueba_Dickey_Fuller(df["PRODUCCIÓN KW-H"],"PRODUCCIÓN KW-H")
```

**Figura D.1.** Librería empleada para la aplicación de la prueba de DICKEY-FULLER.

```
Test Estadístico -4.928157
Valor P 0.000031
Sin retrasos 12.000000
Numero de observaciones usadas 1814.00000
Valor crítico (1%) -3.433960
Valor crítico (5%) -2.863135
Valor crítico (10%) -2.567619
tipo: float64
Conclusión:====>
Rechazar la hipótesis nula
Los datos son estacionarios
```

**Figura D.2.** Resultados de la prueba de Dickey-Fuller.

```
df1 = df . copy ()
#Tomamos La primera diferencia entre Los datos kW
df1 [ 'PRODUCCIÓN KW-H DIA_diff' ] = df [ 'PRODUCCIÓN KW-H DIA' ] . diferencia ()

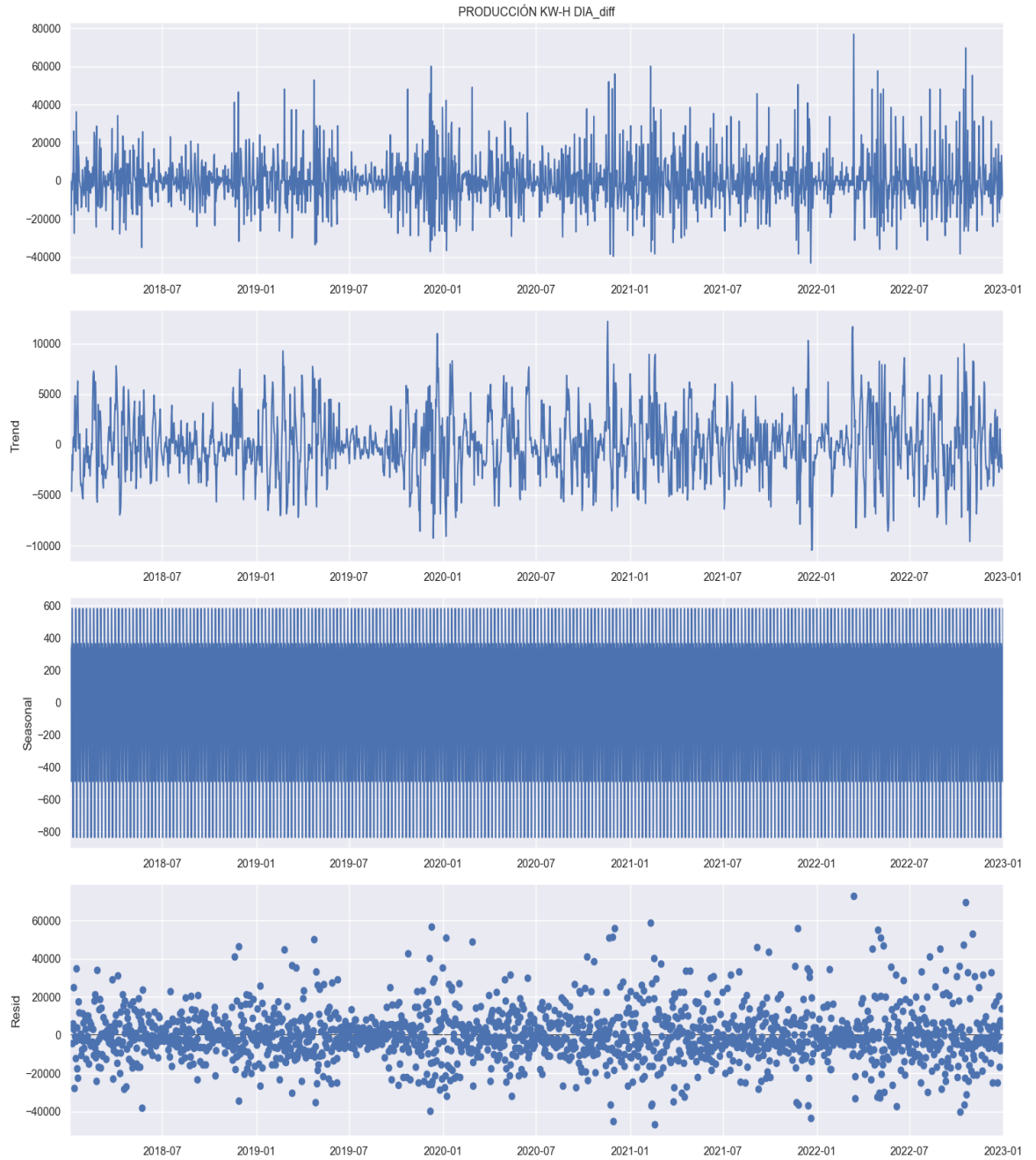
#se eliminan datos nulos, comando dropna
df1 . dropna ( en su lugar = Verdadero )

# Se crea un nuevo df
df1
```

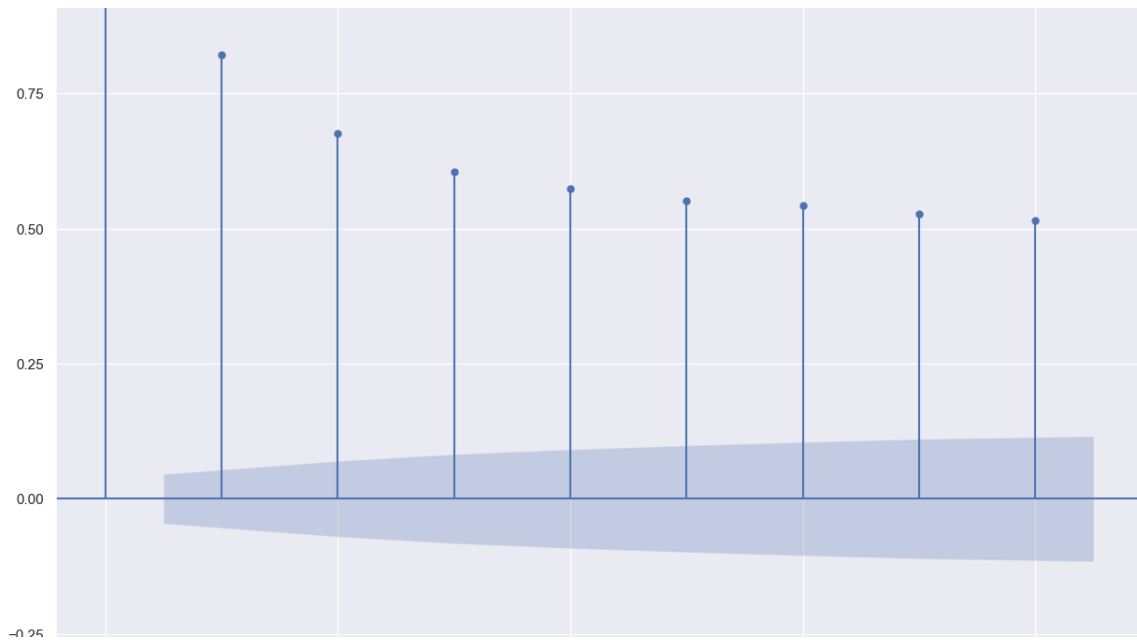
FECHA	PRODUCCIÓN KW-H DIA	PRODUCCIÓN KW-H DIA_diff
2018-01-02	102200.0	8.000000e+02
2018-01-03	85400.0	-1.680000e+04
2018-01-04	67500.0	-1.790000e+04
2018-01-05	67000.0	-5.000000e+02
2018-01-06	62200.0	-4.800000e+03
...	...	...
2022-12-28	78000.0	1.320000e+04
2022-12-29	80400.0	2.400000e+03
2022-12-30	73200.0	-7.200000e+03
2022-12-31	64800.0	-8.400000e+03
2023-01-01	64800.0	3.274181e-08

1826 filas × 2 columnas

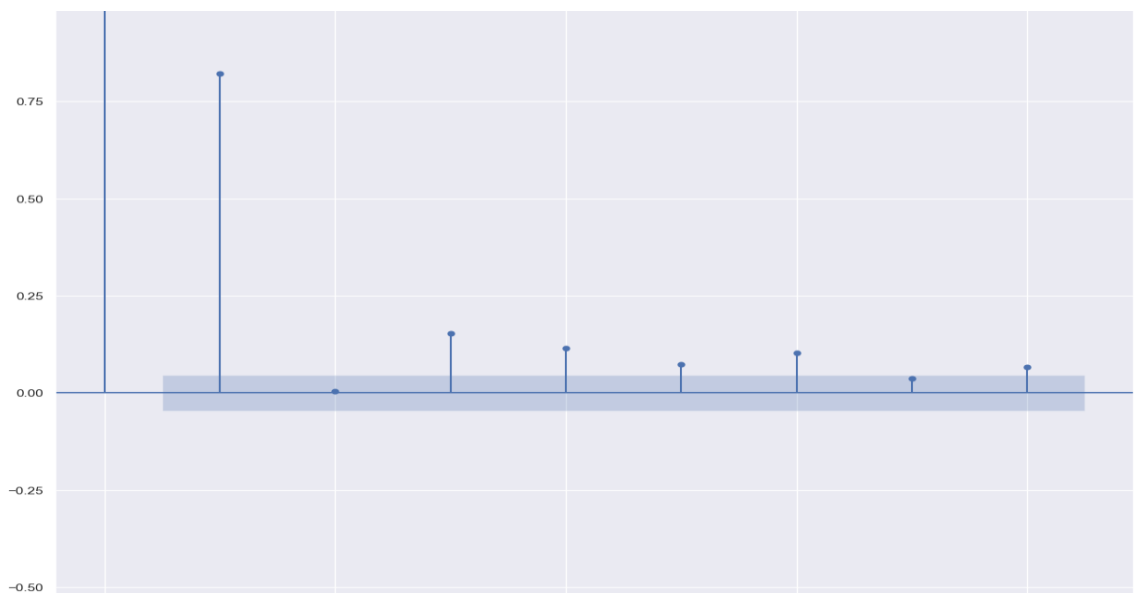
**Figura D.3.** Diferenciación de la producción KWH-DIA 2018-2023.



**Figura D.4.** Descomposición de la serie temporal 2018-2023.



**Figura D.5.** Autocorrelación simple de la serie temporal.



**Figura D.6.** Autocorrelación parcial de la serie temporal.

```
tren_datos = df [: len ( df ) - 12 ]  
prueba_datos = df [ len ( df ) - 12 : ]  
prueba = prueba_datos . copiar ()  
  
prueba_datos . cabeza ()
```

:

PRODUCCIÓN KW-H DIA	
FECHA	
2022-12-21	67200.0
2022-12-22	86400.0
2022-12-23	88800.0
2022-12-24	72000.0
2022-12-25	81600.0

**Figura D.7.** División de datos de entrenamiento y testeo.

## ANEXO E. Modelo AUTO-ARIMA.

```
# Modelo Auto-Arima
de pmdarima import auto_arima
```

Ejecutemos la función `auto_arima()` para obtener los mejores parámetros de p,d,q,P,D,Q

**Nota:** El parámetro *metrose* refiere al número de periodo para cada estacionalidad

- 7 para los días
- 12 para los meses
- 52 para las semanas
- 4 trimestrales
- 1 anual

**Figura E.1.** Modelamiento AUTO-ARIMA.

```
modelo_auto = auto_arima ( train_data , start_p = 0 , d = 1 , start_q = 0 ,
                          max_p = 4 , max_d = 2 , max_q = 4 , start_P = 0 ,
                          D = 1 , start_Q = 0 , max_P = 2 , max_D = 1 ,
                          máx_Q = 2 , m =12 , estacional = Verdadero ,
                          error_action = 'advertir' , rastro = Verdadero ,
                          suprimir_advertencias = Verdadero , paso a paso = Verdadero ,
                          estado_aleatorio = 20 , n_ajustes = 50 )
imprimir ( modelo_auto )
```

Realización de búsqueda paso a paso para minimizar aic

```
ARIMA(0,1,0)(0,1,0)[12] : AIC=40652.628, Tiempo=0.31 seg
ARIMA(1,1,0)(1,1,0)[12] : AIC=40206.080, Tiempo=0.76 seg
ARIMA(0,1,1)(0,1,1)[12] : AIC=inf, Tiempo=2.16 seg
ARIMA(1,1,0)(0,1,0)[12] : AIC=40636.255, Tiempo=0.16 seg
ARIMA(1,1,0)(2,1,0)[12] : AIC=40028.569, Tiempo=1.33 seg
ARIMA(1,1,0)(2,1,1)[12] : AIC=inf, Tiempo=4.09 seg
ARIMA(1,1,0)(1,1,1)[12] : AIC=inf, Tiempo=1.37 seg
ARIMA(0,1,0)(2,1,0)[12] : AIC=39908.300, Tiempo=1.05 seg
ARIMA(0,1,0)(1,1,0)[12] : AIC=40159.707, Tiempo=0.40 seg
ARIMA(0,1,0)(2,1,1)[12] : AIC=inf, Tiempo=3.79 seg
ARIMA(0,1,0)(1,1,1)[12] : AIC=inf, Tiempo=2.23 seg
ARIMA(0,1,1)(2,1,0)[12] : AIC=40010.052, Tiempo=1.45 seg
ARIMA(1,1,1)(2,1,0)[12] : AIC=inf, Tiempo=5.56 seg
Intercepción ARIMA(0,1,0)(2,1,0)[12] : AIC=40046.743, Tiempo=1.34 seg
```

Mejor modelo: ARIMA(0,1,0)(2,1,0)[12]

Tiempo total de ajuste: 26.009 segundos

```
ARIMA(0,1,0)(2,1,0)[12]
```

**Figura E.2.** Búsqueda paso a paso para minimizar AIC.

```
imprimir ( modelo_auto . resumen ())
```

```

                                Resultados SARIMAX
=====
dep. Variable: y No. Observaciones: 1815
Modelo: SARIMAX(0, 1, 0)x(2, 1, 0, 12) Logaritmo de probabilidad -19951,150
Fecha: jue, 23 feb 2023 AIC 39908.300
Hora: 10:53:46 BIC 39924.790
Muestra: 01-01-2018 HQIC 39914.387
                                - 20-12-2022

Tipo de covarianza: opg
=====
                                coef std err z P>|z| [0,025 0,975]
-----
ar.S.L12 -0.6656 0.019 -35.581 0.000 -0.702 -0.629
ar.S.L24 -0.3590 0.019 -18.899 0.000 -0.396 -0.322
sigma2 2.422e+08 4.6e-12 5.26e+19 0.000 2.42e+08 2.42e+08
=====
Ljung-Box (L1) (Q): 16,38 Jarque-Bera (JB): 491,92
Prob(Q): 0,00 Prob(JB): 0,00
Heterocedasticidad (H): 1,75 Sesgo: 0,65
Prob(H) (bilateral): 0,00 Curtosis: 5,21
=====

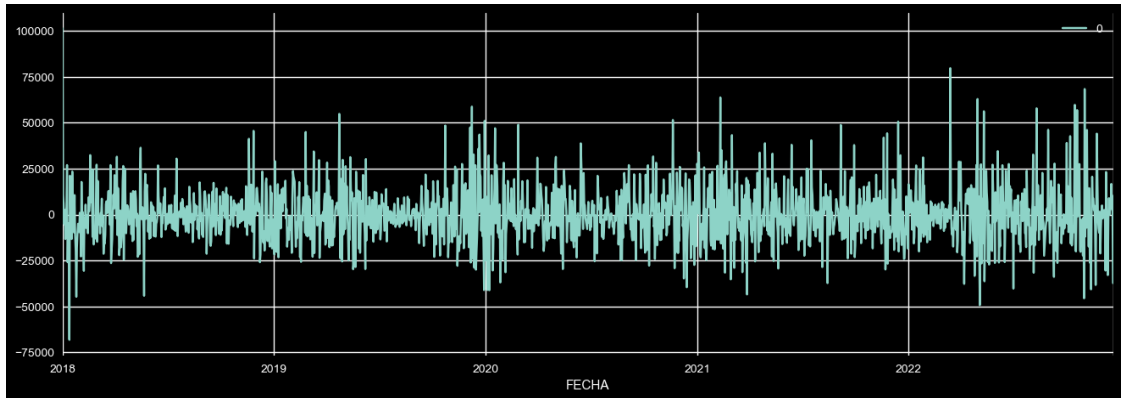
```

**Figura E.3.** Resultados de Testeo modelo AUTO-ARIMA.

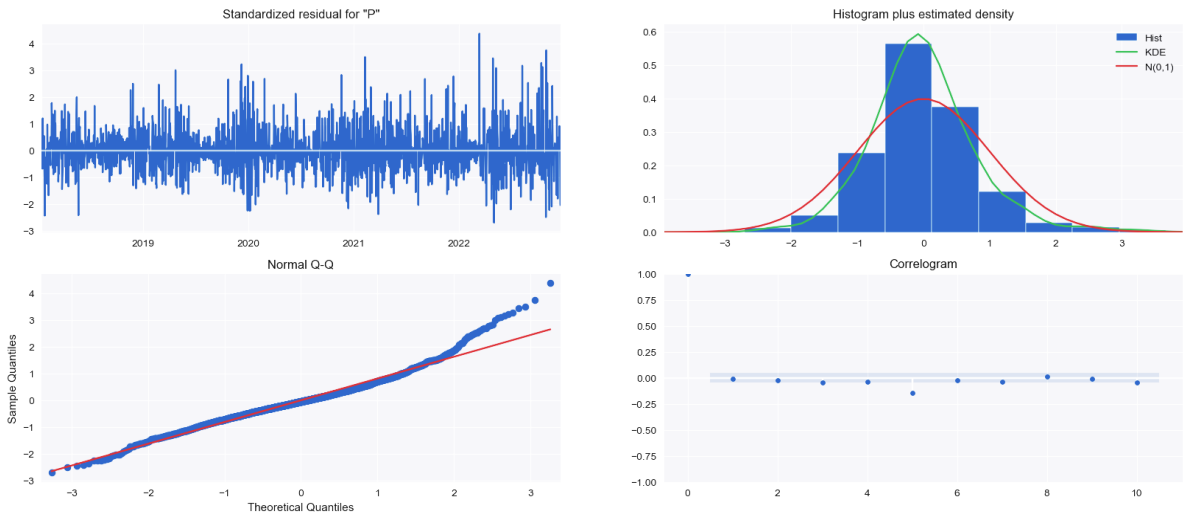
Resultados SARIMAX						
dep. Variable:	PRODUCCIÓN KW-H DIA		Nº Observaciones:	1815		
Modelo:	SARIMAX(4, 1, 0)x(2, 1, 0, 12)		Probabilidad de registro	-19930.868		
Fecha:	jue., 23 feb. 2023		AIC	39875.737		
Tiempo:	10:53:51		BIC	39914.214		
Muestra:	01-01-2018		HQIC	39889.939		
				- 20-12-2022		
Tipo de covarianza:	opg					
	coef	error estándar	z	P> z	[0.025	0.975]
ar.L1	-0.1706	0.032	-5.395	0.000	-0.233	-0.109
ar.L2	-0.2662	0.032	-8.264	0.000	-0.329	-0.203
ar.L3	-0.1694	0.033	-5.169	0.000	-0.234	-0.105
ar.L4	-0.1108	0.030	-3.667	0.000	-0.170	-0.052
ar.S.L12	-0.6614	0.028	-23.483	0.000	-0.717	-0.606
ar.S.L24	-0.3508	0.028	-12.336	0.000	-0.407	-0.295
sigma2	3.316e+08	2.41e-11	1.38e+19	0.000	3.32e+08	3.32e+08
Caja Ljung (L1) (Q):	0.27	Jarque Bera (JB):	454.88			
Prob(Q):	0.60	Prob(JB):	0.00			
Heterocedasticidad (H):	1.69	Sesgar:	0.62			
Prob(H) (bilateral):	0.00	Curtosis:	5.12			



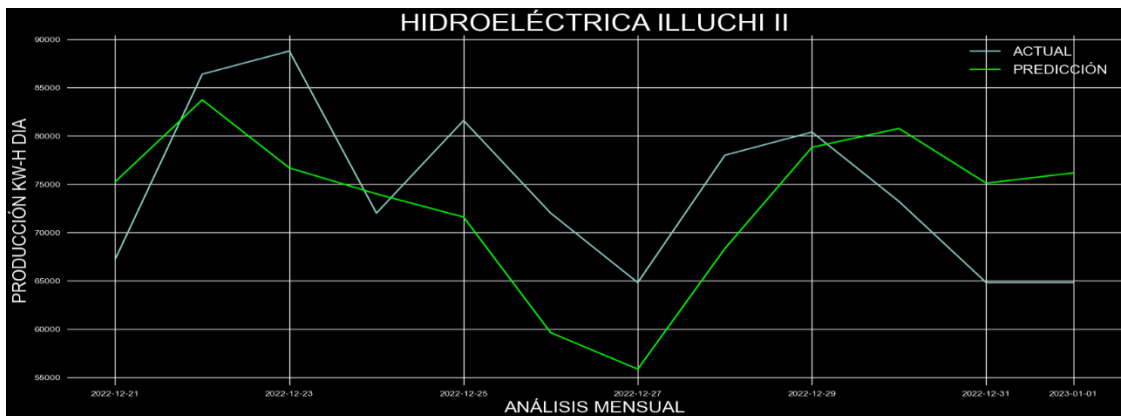
**Figura E.4.** Implementación del modelo.



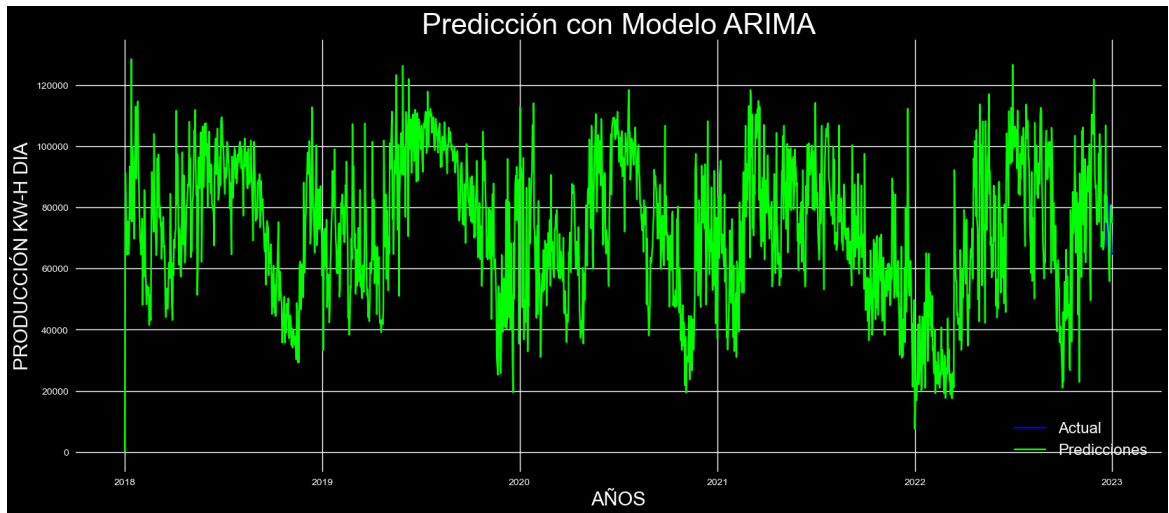
**Figura E.5.** Gráfico de línea de residuos AIC.



**Figura E.6.** Modelo de diagnóstico.



**Figura E.7.** Comparación del modelo ARIMA frente a la predicción análisis mensual.



**Figura E.7.** Predicción con el Modelo ARIMA análisis anual.

ANEXO F. Estructura del método ARIMA en los Escenarios.

```
# Manipulación y tratamiento de Datos
import numpy as np
import pandas as pd

# Visualización de datos
import plotly.express as px
import matplotlib.pyplot as plt
%matplotlib inline
plt . estilo _ uso ( 'ggplot' )

# Modelación Arima
de statsmodels.tsa.statespace.sarimax import SARIMAX
de statsmodels.graphics.tsaplots import plot_acf , plot_pacf
de statsmodels.tsa.seasonal import season_decompose
de statsmodels.tsa.stattools import adfuller

# Métrica de Evaluación
from sklearn.metrics import mean_squared_error
from statsmodels.tools.eval_measures import rmse
from sklearn import metrics

# No presentar advertencias
import warnings
warnings . advertencias de filtro ( "ignorar" )

importar seaborn como sns
sns . conjunto ()

from datetime import datetime
from datetime import timedelta
from datetime import fecha de importación
```

Figura F.1. Librerías ejecutadas en Python.

```
df = pd . read_excel ( 'DATOS/df_kwh-total.xlsx' )
df
```

	FECHA	PRODUCCIÓN KW-H DIA
0	2018-01-01	101400.0
1	2018-01-02	102200.0
2	2018-01-03	85400.0
3	2018-01-04	67500.0
4	2018-01-05	67000.0
...	...	...
1822	2022-12-28	78000.0
1823	2022-12-29	80400.0
1824	2022-12-30	73200.0
1825	2022-12-31	64800.0
1826	2023-01-01	64800.0

1827 filas × 2 columnas

Figura F.2. Llamado de datos depurados.



