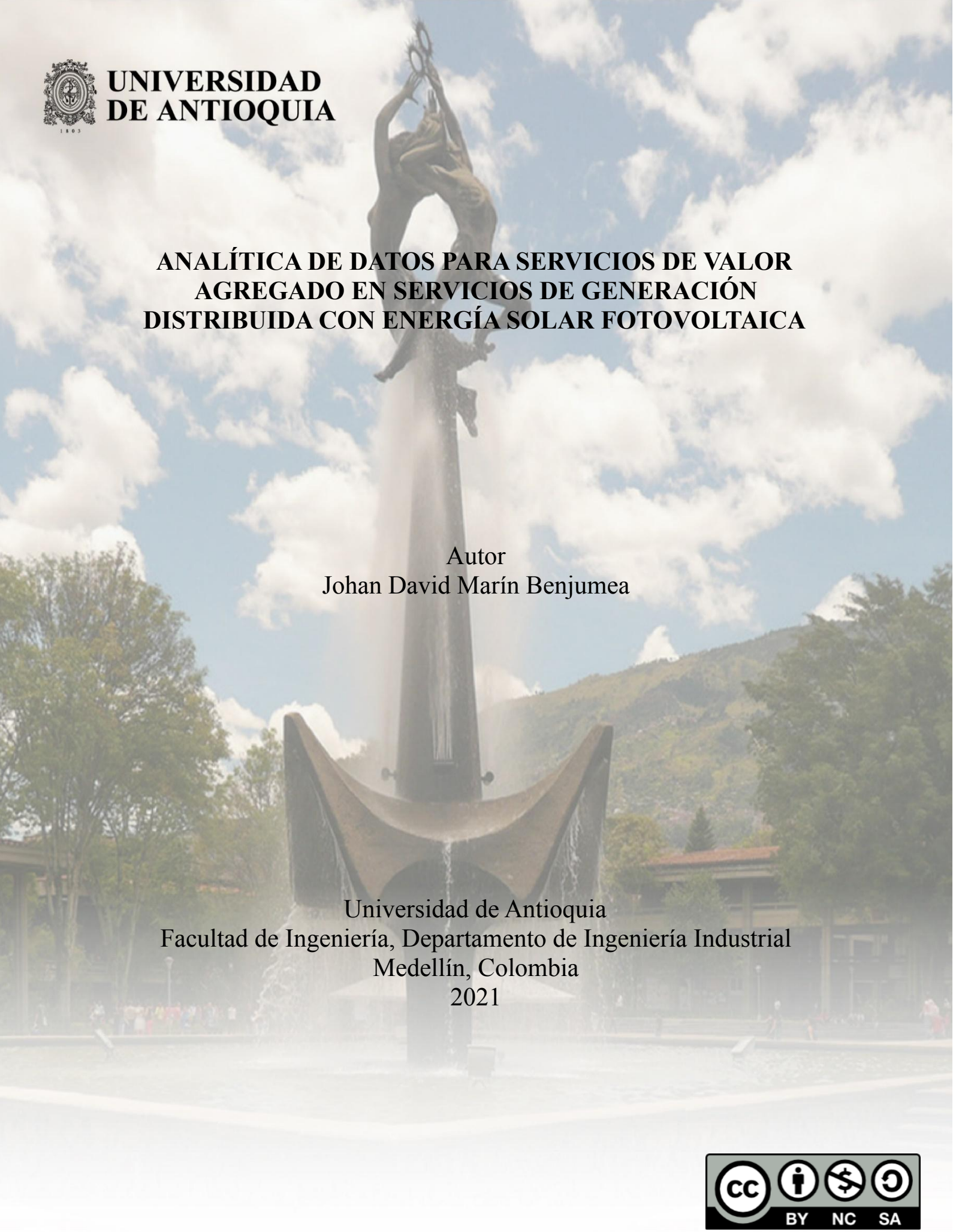

UNIVERSIDAD DE ANTIOQUIA

ANALÍTICA DE DATOS PARA SERVICIOS DE VALOR AGREGADO EN SERVICIOS DE GENERACIÓN DISTRIBUIDA CON ENERGÍA SOLAR FOTOVOLTAICA

Autor
Johan David Marín Benjumea

Universidad de Antioquia
Facultad de Ingeniería, Departamento de Ingeniería Industrial
Medellín, Colombia

2021

Analítica de Datos para Servicios de Valor Agregado en Servicios de Generación Distribuida con Energía Solar Fotovoltaica

Johan David Marín Benjumea

Proyecto de práctica presentado como requisito parcial para optar al título de:
Ingeniero Industrial

Asesores:
Juan Guillermo Villegas Ramirez
Ph.D. Ingeniería/Optimización de sistemas industriales

Ramón Alberto Leon Candela
M.Sc. Ingeniería Eléctrica

Línea de Investigación:
Análisis de datos

Universidad de Antioquia
Facultad de Ingeniería, Departamento de Ingeniería Industrial
Medellín, Colombia
2021

Tabla de Contenido

Introducción **5**

Objetivos **7**
Objetivo General 7
Objetivos Específicos 7

Marco Teórico **8**

Metodología y Resultados **9**
Gobierno de datos 10
Gestión de metadatos. 10
Calidad de datos. 15
Integración de datos. 20
Conocimiento del negocio 21
Análisis exploratorio de los datos. 21
Definición de indicadores. 25
Indicadores de operación. 25
Indicadores ambientales. 28
Indicadores económicos. 28
Creación de modelos. 29
Selección de la información. 34
Indicadores de operación. 35
Indicadores ambientales. 37
Indicadores económicos. 38
Estandarización y automatización del proceso 38

Conclusiones **41**

Referencias Bibliográficas **42**

Anexos **45**

Lista de Tablas

Tabla 1: Resumen gestión de metadatos
Tabla 2: Variables descargadas de las bases de datos
Tabla 3: Formato definido para cada tipo de variable
Tabla 4: Métricas de desempeño para los modelos ajustados insolación solar
Tabla 5: Métricas de desempeño para los modelos para predecir la potencia esperada
Tabla 6: Tabla para identificar paradas

Lista de Figuras

Figura 1: Diagrama metodología utilizada
Figura 2: Comportamiento de la insolación solar durante 5 días consecutivos
Figura 3: ACF y PACF insolación solar
Figura 4: insolación promedio y radiación por hora y por mes.
Figura 5: Modelo medias móviles (MM 2) insolación solar
Figura 6: Modelo AR(2) insolación solar
Figura 7: Modelo de regresión lineal insolación solar
Figura 8: Promedio de generación por hora
Figura 9: potencia generada vs insolación
Figura 10: Correlación energía vs insolación y temperatura de módulo
Figura 11: Potencia generada vs insolación paradas
Figura 12: potencia generada vs insolación producción de energía inferior a la esperada
Figura 13: Consumos y costos de energía promedio por hora.
Figura 14: Distribución del coeficiente, antes y después de eliminar outliers
Figura 15:Regresión lineal potencia vs insolación
Figura 16: Regresión lineal potencia vs temperatura módulo
Figura 17: Distribución de coeficiente
Figura 18: ACF y PACF coeficiente potencia generada/insolación solar
Figura 19: Potencia generada, en función de la insolación y el coeficiente AR(2)
Figura 20:Potencia generada, en función de la insolación y el coeficiente MM(2)
Figura 21: Energía generada y energía esperada por mes (informe)
Figura 22: Energía cubierta por la planta promedio y total mes
Figura 23: Pérdidas de energía por fallas
Figura 24: Emisiones de CO2 evitadas por mes gracias a la planta solar
Figura 25: Ahorros por mes en COP
Figura 26: Pérdidas en dinero por fallas
Figura 27: Home GUI de paquete power
Figura 28: Arquitectura de automatización del proceso

ANALÍTICA DE DATOS PARA SERVICIOS DE VALOR AGREGADO EN SERVICIOS DE GENERACIÓN DISTRIBUIDA CON ENERGÍA SOLAR FOTOVOLTAICA

Resumen

Este proyecto presenta el primer piloto de servicios de valor agregado basados en datos, para clientes de Interconexión eléctrica S.A (ISA), con sistemas de generación distribuida basados en plantas de energía solar fotovoltaica. Como requerimiento principal del proyecto, se tuvo que la solución propuesta, fuera replicable para diferentes clientes, fuera utilizable por cualquier miembro de la compañía y funcionará de forma automática para el usuario.

Para solucionar el problema se realizó un proceso de gestión de metadatos sobre las bases de datos, un proceso de calidad de datos para estandarizar e integrar los datos, un análisis de datos para definir y desarrollar indicadores para el servicio. Finalmente se desarrolló y se compiló un paquete en Python, que permite replicar la metodología automáticamente, acompañado de una interfaz gráfica que facilita el uso para usuarios sin conocimiento en programación.

Introducción

"Históricamente ha prevalecido la idea de que un sistema eléctrico eficiente debe basarse en grandes plantas de generación y largas líneas de transmisión "(Quintero, 2008). Sin embargo, el desarrollo de nuevas tecnologías en el sector eléctrico, avances en tecnologías de la información y la comunicación, y la relevancia cada vez mayor de la conservación del medio ambiente, están generando una transformación del sector eléctrico (Cozzi & Gould, 2016).

Uno de estos cambios es la integración de sistemas de generación distribuida (GD) a la red eléctrica. La GD hace referencia a la producción de electricidad cerca del lugar de consumo, en lugar de grandes centrales (hidroeléctricas, térmicas o nucleares) alejadas del cliente. La GD minimiza las pérdidas por transporte, incrementa la eficiencia y la confiabilidad del sistema, optimiza el uso de los recursos, disminuye la contaminación ambiental y reduce el tamaño de las plantas de generación (Dulau, Abrudean, & Bica, 2014).

En Colombia, un país que en 2013 tenía aproximadamente un 96% de generación centralizada (Consorcio Hart-Re, 2014), incluso las industrias que no pertenecen al sector eléctrico están optando por este nuevo modelo, principalmente con la instalación de sistemas de GD (González, Daza, & Urueña, 2008). Este fenómeno es impulsado por la creciente necesidad energética, la regulación de emisiones cada vez más estricta y principalmente por el abaratamiento de los sistemas de generación de energía renovable.

En 2019 los sistemas de generación de energía solar fotovoltaica alcanzaron costos competitivos para algunas empresas y se espera que estos costos continúen disminuyendo a futuro (García, Manan, & Manghan, 2019). Esta reducción de costos ha hecho que en los últimos dos años empresas en Colombia y otros países de Latinoamérica, instalen plantas de generación de energía solar fotovoltaica en sus predios, (Bnamericas, 2019), tendencia que se espera continúe.

Con la instalación de sistemas para GD, los antiguos consumidores de energía se convierten en generadores, que pueden no solo producir su energía sino también vender en el mercado de energía los excesos de energía producida. Estos cambios implican nuevas necesidades para las empresas, desde la instalación y mantenimiento del sistema de generación, hasta la venta de energía de exceso y la operación bajo la regulación (Perez, 2018), necesidades que están impulsando el surgimiento de un nuevo mercado.

Interconexión Eléctrica SA (ISA), empresa del sector de la transmisión de energía eléctrica que opera en Colombia y otros países de Latinoamérica, está explorando incursionar en este mercado, ofreciendo soluciones de GD para empresas conectadas a niveles de tensión 2 y 3, es decir entre 30 – 57.5 kV y 57.5 - 220 kV,

respectivamente. Entre sus estrategias, ISA pretende brindar servicios de valor agregado (SVA) basados en datos; como factor diferenciador, aprovechando la experiencia de la compañía y su conocimiento del mercado eléctrico colombiano. Para lograrlo, ha abierto una nueva línea de negocio enfocada en recursos energéticos distribuidos (RED), entre los que se encuentra la GD.

Sin embargo, el desarrollo de estos servicios trae consigo nuevos retos, cuatro de los cuales se abordarán en esta práctica: (i) se requiere de la unión de varias fuentes de datos; (ii) se requiere del procesamiento y análisis de estos datos para la creación posterior de indicadores, usando modelos predictivos y prescriptivos.; (iii) es necesario depurar la información relevante para cada cliente y presentarla en un informe corto. Y, por último, (iv) se busca la estandarización y automatización de los tres procesos anteriores.

Por otro lado, cabe mencionar que para el desarrollo de esta práctica los retos arriba mencionados se abordaron en un estudio piloto enfocado en la planta de generación de energía solar ubicada en las instalaciones de ISA, de modo que sea la base para su implementación con los clientes de la compañía en el futuro.

Objetivos

Objetivo General

Desarrollar una herramienta computacional que permita medir automáticamente el comportamiento de la planta solar fotovoltaica de la compañía en términos, ambientales, económicos y de operación.

Objetivos Específicos

Automatizar la recolección de información de la planta solar fotovoltaica de la compañía, junto con la información del mercado de energía eléctrica colombiano.

Definir indicadores de desempeño de la planta solar basados en la información recolectada para describir el comportamiento con respecto a las dimensiones ambiental, financiera y de operación.

Generar automáticamente reportes integrados a los sistemas de información de la compañía de manera que la información sea accesible a distintos miembros de la organización.

Marco Teórico

La analítica de datos juega un papel cada vez más relevante para la gestión de la red eléctrica (Ilic, Black, & Prica, 2007), impulsada por la integración de una capa de información, para la recopilación, el almacenamiento y el análisis de datos en las redes eléctricas convencionales (Yang, Tao, & Bompard, 2018). Análisis como la detección de fallas y el mantenimiento basado en servicios; para mejorar la confiabilidad del servicio, la predicción de generación con energías renovables, o detección de pérdidas no técnicas. Para una revisión más detallada de estas aplicaciones el lector puede remitirse a Yang, Tao, & Bompard (2018).

La analítica también se ha utilizado en la gestión de microrredes y GD para: selección del sistema de generación (Ruan, et al., 2009), mejorar la operación, identificación de fallos y fluctuaciones (Seyedi, Karimi, & Grijalva, 2019), gestión de la carga (Ozcanli, Yaprakdal, & Baysal, 2020) y modos flexibles, mejorar la rentabilidad en los mercados (Kakran & Chanana, 2018; Mitra & Suyanarayanan, 2010) y respuesta rápida a los cambios en la regulación (Ilic, Black, & Prica, 2007). A continuación, se presenta brevemente una revisión de herramientas utilizadas en la capa de información en redes y microrredes inteligentes.

Desde las herramientas computacionales se encontró que los sistemas distribuidos (Munshi & Mohamed, 2017), las redes computacionales dedicadas y soluciones en la nube (Simmhan, et al., 2013; Diamantoulakis, Kapinas, & Karagiannidis, 2015) son las herramientas más comunes para la recopilación, procesamiento y el almacenamiento de datos en redes inteligentes, en combinación con

diferentes lenguajes de programación y lenguajes de consulta como HIVE o IMPALA (Diamantoulakis, Kapinas, & Karagiannidis, 2015).

Por otra parte en cuanto a las herramientas matemáticas utilizadas para el análisis, se encontró una gran diversidad de estas desde redes neuronales para detección de fallas, predicción de generación y consumo (Ozcanli, Yaprakdal, & Baysal, 2020), modelos de programación entera para operar una microrred conectada a la red (Parisio, Rikos, & Glielmo, 2014), hasta el uso de teoría de grafos y datos de redes sociales para detectar eventos de emergencia como cortes de energía (Bauman, Tuzhilin, & Zaczynski, 2017). En particular Yang, Tao, & Bompard (2018), las clasifican en seis categorías: estadística, aprendizaje de máquina, aprendizaje profundo, minería de datos, reconocimiento de patrones e inteligencia artificial.

Conscientes de la necesidad de establecer un proceso de gestión de datos, para garantizar la escalabilidad de la capa de información en "redes inteligentes", Munshi & Mohamed (2017) desarrollan un marco, para la gestión de los datos en cinco pasos; adquisición de datos, almacenamiento distribuido de datos, procesamiento de datos distribuidos, consulta de datos y análisis de datos. Mientras que (Simmhan, et al., 2013) estructuran el desarrollo de una plataforma de software escalable, basada en la nube.

Metodología y Resultados

Teniendo en cuenta la importancia de desarrollar el proyecto desde un marco general de gestión de datos (Munshi & Mohamed, 2017), se decide abordar el desarrollo, utilizando los tres pasos de la metodología propuesta por Brackett, Earley, & Henderson (2009); Gestión de metadatos, Calidad de datos y conocimiento del negocio, seguida de un proceso de automatización.

Primero un proceso de gobierno de datos que incluye la exploración de las fuentes de datos, la identificación de formatos y variables para interconexión de distintas fuentes. Seguido de una fase de conocimiento del negocio donde se identifica y se extrae información de negocio y se definen indicadores; haciendo uso de diferentes técnicas de análisis de datos. Finalmente una etapa de estandarización y automatización, en la que se definen medios para hacer accesible tanto los datos como

la información y se automatiza el proceso de extracción de información como se muestra en la figura 1.

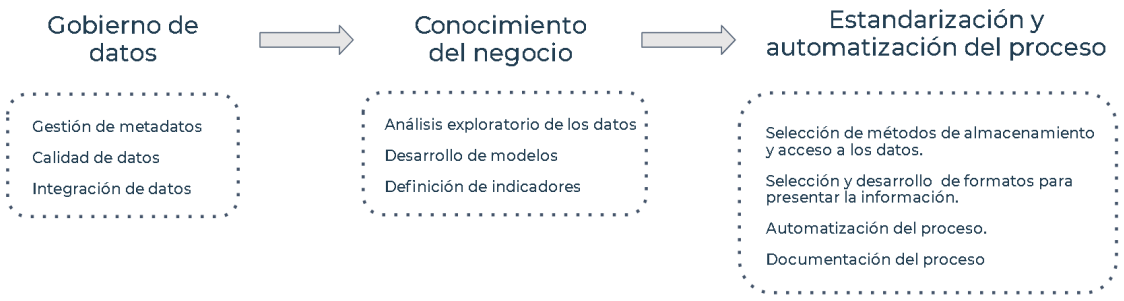

Figura 1: Diagrama metodología utilizada.

Gobierno de datos

Para el desarrollo del proyecto se cuenta con tres fuentes de datos, la información de la planta solar se obtendrá a partir de un software de monitoreo, que se usa en la operación de la planta (conocida como Meteocontrol), La información de consumos de energía desde la red eléctrica nacional, que se obtiene desde EPM y la información del mercado de energía, que se obtiene utilizando la API de XM, el operador del Sistema eléctrico colombiano.

Gestión de metadatos.

Primero se identificaron los datos disponibles en cada fuente y el formato en el que estos están almacenados. Se identificó que en todas las fuentes, la información está almacenada en tablas con formato matricial, a excepción de la información proveniente de la aplicación de Meteocontrol, que también tiene datos en formato data frame (3NF). Con miras a la posterior integración de los datos, se identificaron variables que puedan servir como clave de integración entre las fuentes (clave foránea). En este caso se identificó que la única variable común en las tres fuentes es la fecha y hora, y algunas variables para integrar tablas de información de la misma fuente (claves primarias). Ratificando la fecha y la hora como clave común entre tablas y el inversor para algunas tablas provenientes de Meteocontrol. Sin embargo, se encontraron datos registrados con

diferentes frecuencias de muestreo, incluso en la misma fuente. Surgió así la necesidad de transformar frecuencias al momento de la integración. La tabla 1 resume la gestión de metadatos, con los datos disponibles, las frecuencias de muestreo, las claves, los formatos de las tablas y la forma de extraer los datos.

Tabla 1: Resumen gestión de metadatos. Datos disponibles frecuencias y variables para integración

	Meteocontrol	**EPM**	**XM**
Datos Disponibles	+ Datos meteorológicos + Datos de producción de energía por inversor y de la planta + Pronósticos generación	+ Datos de consumo de la red a 13.1 kWh y a 44 kWh + Precios por kWh	+ Datos de generación consumo y precios de diferentes actores en el SIN
Frecuencia de muestreo	+ 5 min	+ Horaria	+ Horaria + Diaria + Anual
Claves Foráneas	+ Fecha y hora	+ Fecha y hora	+ Fecha y hora
Claves Primarias	+ Fecha y hora + Inversor		+ Fecha y hora + código de sub mercado + Tipo de generador
Formato de datos	+ Matricial	+ Dataframe + Matricial	+ DataFrame + Matricial
Acceso a datos	+ Archivos csv + HTML aplicación web	+ Archivos csv y xlsx	+ API + Archivos xlsx

Se eligen como formatos para el acceso a los datos los archivos .csv para EPM y Meteocontrol y el `DataFrame` entregado por la API para los datos de XM. Además. En el caso de XM, se puede obtener información entre dos fechas cualquiera según la frecuencia propia de la variable. Mientras que para Epm, se pueden extraer datos por un año con frecuencia horaria, una tabla por variable y finalmente en el caso de Meteocontrol de la misma forma que con Epm es una tabla por variable y además solo se pueden obtener los datos para un día con frecuencia de 5 minutos.

Teniendo en cuenta que la fecha y hora es la clave de integración seleccionada, se observa el formato en el que cada una de las fuentes la presenta, en el caso de XM las fecha y hora usan el formato datetime de python, con la fecha en cada registro y columna por cada hora. Por su parte, para Epm, la fecha se encuentra en formato dd/mm/aaaa y la hora en formato entero 1, 2; donde 1 es el tiempo transcurrido entre las 00:01 am y las 01:00 am y así sucesivamente, con la fecha frente a cada registro y una columna por cada hora. Finalmente en Meteocontrol la fecha tiene el formato aaaa/mm/dd y la hora hh:mm, la fecha solo aparece en la parte superior del archivo csv y de la página web, y en el campo fecha hora aparece hora y minuto.

Adicionalmente se observó que los datos de XM están anonimizados con la clave de submercado, y que al igual que Meteocontrol utilizan la coma como indicador de decimal, que Epm presenta datos separados para dos conexiones, debido a que la compañía tiene dos conexiones al SIN, y Meteocontrol utiliza la ‘x’ para indicar valores perdidos.

Particularmente en Meteocontrol se encontraron varias inconsistencias: como que en algunas fechas no se registran datos climáticos entre las 18:30 y las 5:00 horas, mientras en otras fechas se registran las 24 horas, la incorporación de nuevas variables posteriores a la fecha de inicio de registro, como en el caso de la corriente AC que el 22 de

agosto de 2020 pasó de ser una variable a tres, una por fase. Para esta fuente de datos en particular estos cambios de formato se dan con frecuencia así como cambios en las tablas en las que presenta la información y la falta de algunos registros de insolación en los datos de la planta que sí estaban presentes cuando se miraban los registros de insolación registrada por el satélite de Meteocontrol. Esto generó varios reprocesos en el desarrollo del proyecto. En particular, fue necesario actualizar los códigos creados para la extracción automática de la información.

Las variables extraídas de las fuentes de datos fueron definidas con ayuda de los miembros del grupo de nuevos negocios de la compañía. Teniendo en cuenta lo siguiente. todos los datos (variables) de XM se consultan y se procesan (automáticamente) en el momento de su uso, para evitar utilizar espacio de almacenamiento extra, para los datos de Epm, se descargan las variables relacionadas con la energía, consumo y costo por unidad (KWh) de energía consumida. Finalmente en el caso de Meteocontrol, se decidió descargar todas las variables que fueran independientes (la potencia se descargó debido a que hay muchos datos de tensión faltantes); debido a que en ese momento los datos estaban almacenados solamente en los servidores del proveedor y el tener los datos en la compañía aporta independencia y facilita el acceso a los datos para otros análisis. Para todos los datos se utilizó la frecuencia más alta de muestreo que tuviesen disponible, cada 5 minutos para Meteocontrol y horario para Epm. Las variables se presentan en la tabla 2 con su respectivo formato y fuente de datos. Además de los datos simulados durante la formulación del proyecto y los cobros mensuales al cliente.

Tabla 2: Variables descargadas de las bases de datos

Variable	**Fuente**	**Formato**
Consumo por hora kWh	Epm	matricial .xlsx
Costo por KWh consumido	Epm	matricial .xlsx
Potencia AC inversor	Meteocontrol	matricial .csv
Corriente AC inversor	Meteocontrol	matricial csv
Tensión AC inversor	Meteocontrol	matricial .csv
Potencia CC inversor	Meteocontrol	matricial .csv
Corriente CC inversor	Meteocontrol	matricial .csv
Tensión CC inversor	Meteocontrol	matricial .csv
Frecuencia inversor	Meteocontrol	matricial .xlsx
Temperatura inversor	Meteocontrol	matricial .xlsx
Fasor inversor	Meteocontrol	matricial .csv
Potencia reactiva inversor	Meteocontrol	matricial csv
Potencia aparente inversor	Meteocontrol	matricial .csv
Entradas inversor	Meteocontrol	matricial .csv
Insolación solar sensor	Meteocontrol	3NF .csv
Insolación solar satélite	Meteocontrol	3NF .csv
Temperatura de módulo	Meteocontrol	3NF .csv
Potencia esperada planta	Meteocontrol	3NF .csv
Datos de simulación	PVSyst	3NF .csv
Datos de cobros	Contrato	3NF .csv

Calidad de datos.

Teniendo en cuenta la información recolectada en el paso anterior, se define el formato de cada uno de los datos, el cual se presenta en la tabla 3. Además como se había definido previamente en la metodología se organiza la información en formato 3NF, utilizando solo una columna por variable .

Tabla 3: Formato definido para cada tipo de variable, todas en formatos estándar de python

Tipo Variable	Estándar
Fecha y Hora	**Formato:** datetime (yyyy-mm-dd HH:MM:SS) **Nota:** Se ponen ambas en una misma variable
Numéricas	**Formato:** float **Nota:** punto es decimal
Valores perdidos	**Formato:** float('nan')
Carácter	**Formato:** str

Para realizar este proceso se desarrolló un módulo en python para cada fuente de datos que permite realizar la lectura de los datos desde cada una de las fuentes, transformarlas al formato definido en la tabla 2 y poner los datos estandarizados en una tabla en formato 3NF.

Con los expertos del área de nuevos negocios de la compañía, se definió la necesidad de que los registros de la variable insolación estuviesen completos para utilizarlos en análisis posteriores, por lo que fue necesario realizar la imputación de los mismos. Teniendo en cuenta que la variable insolación solo presentó correlación con las variables energía y temperatura del módulo; ambas dependientes de la primera, para definir un método de imputación para los valores de insolación faltantes se consideran la fecha y hora, y la misma variable. Utilizando los datos

registrados por satélite de Meteocontrol, ya que presentaron un menor número de registros faltantes.

La figura 2 muestra el comportamiento de la variable insolación donde se puede evidenciar una ciclicidad, aproximadamente cada 24 horas, y teniendo en cuenta la frecuencia 5 minutal de los datos esto quiere decir una ciclicidad cada 288 registros, lo que se termina de confirmar en la figura 4a. Por otro lado, los gráficos de autocorrelación (ACF) y autocorrelación parcial (PACF) de la figura 3, sugieren el uso de un modelo AR(2), finalmente se observa que el mes tiene efecto sobre la radiación promedio durante una hora Figura 4b.

Con una sola variable para la imputación se plantean tres modelos autorregresivos (Makridakis & Hibon, 1997), el primero, un modelo de medias móviles, con retraso 2 (MM(2)), el segundo un modelo AR(2), finalmente para el tercero se intentó ajustar un modelo SARIMA sin éxito en la convergencia de parámetros, por lo que se reemplazó por una regresión lineal sobre la insolación dos periodos antes, la media histórica del periodo y el promedio de radiación durante el mes en este mismo periodo de tiempo.

Se dividen los datos de entrenamiento y test, y para el test, teniendo en cuenta que el objetivo del modelo no es predecir valores futuros, sino imputar los valores faltantes, los tres modelos tuvieron como entradas los datos reales para predecir cada valor. Las figuras 4, 6 y 7 muestran el ajuste de los modelos y el comportamiento de los errores, Tanto el modelo de medias móviles como el modelo de regresión presentaron un buen ajuste, con un comportamiento muy similar a los datos reales. El modelo AR(2) en cambio a pesar de seguir el comportamiento de los datos tiende a sobreestimar los valores,figura 6, debido a esto aunque el error tiene un comportamiento ruido blanco para los tres modelos, en el caso del modelo AR(2) tiene una media de -177 W/m^2 mientras que los valores para la media móvil y la regresión son 0 W/m^2 y 0.6 W/m^2 respectivamente.

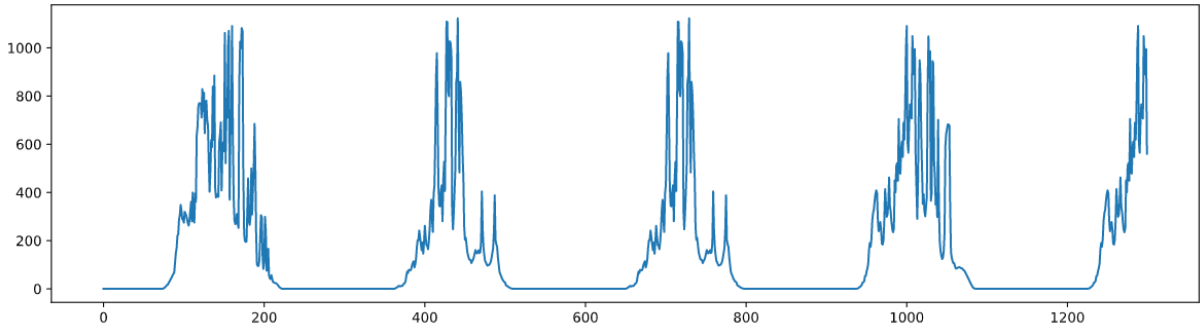

Figura 2: Comportamiento de la insolación solar durante 5 días consecutivos.

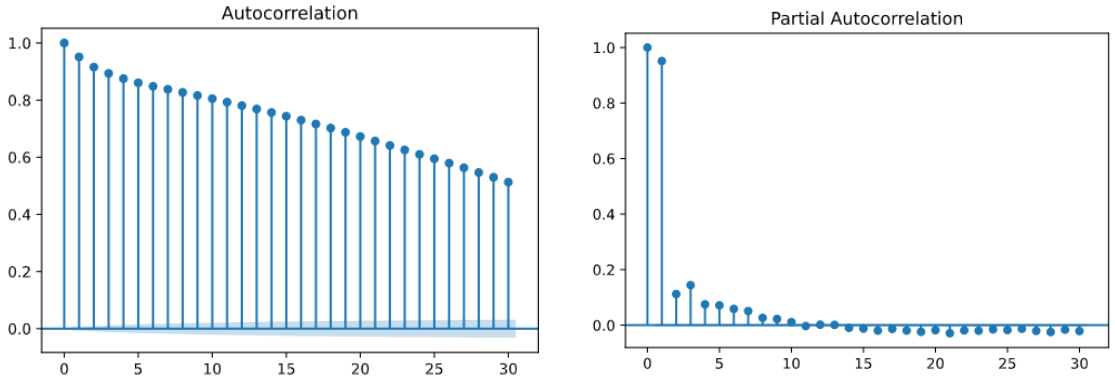

Figura 3: ACF(Auto correlation function) y PACF (Partial Auto Correlation Function) insolación solar con frecuencia 5 minutal.

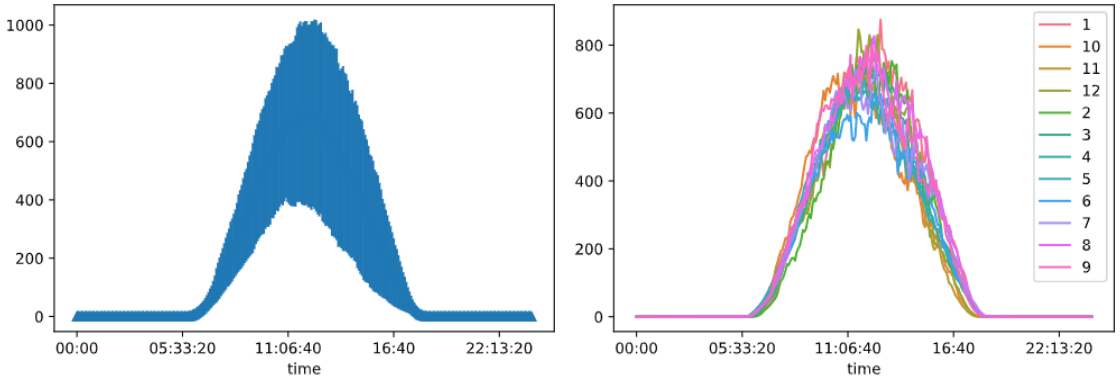

Figura 4: a la izquierda insolación promedio y radiación por hora, insolación promedio por hora y por mes.

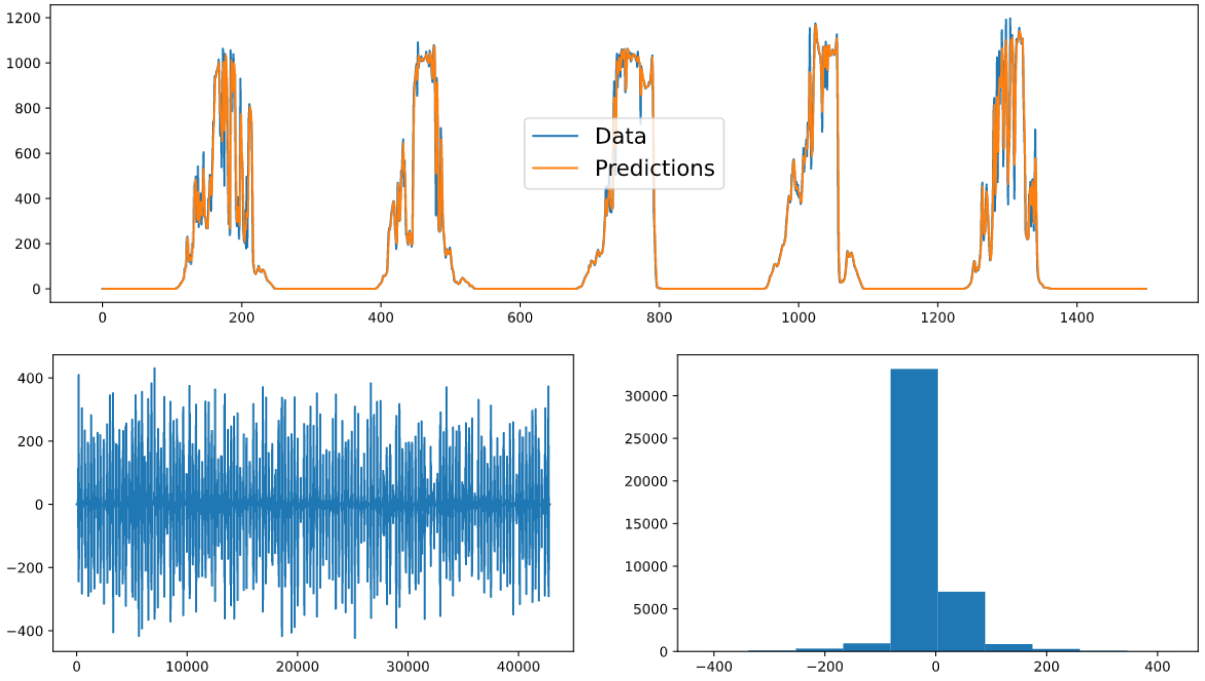

Figura 5: Modelo medias móviles (MM) con retardo 2: arriba, en azul los valores reales y en naranja los valores predichos, abajo los errores y el histograma de los mismos.

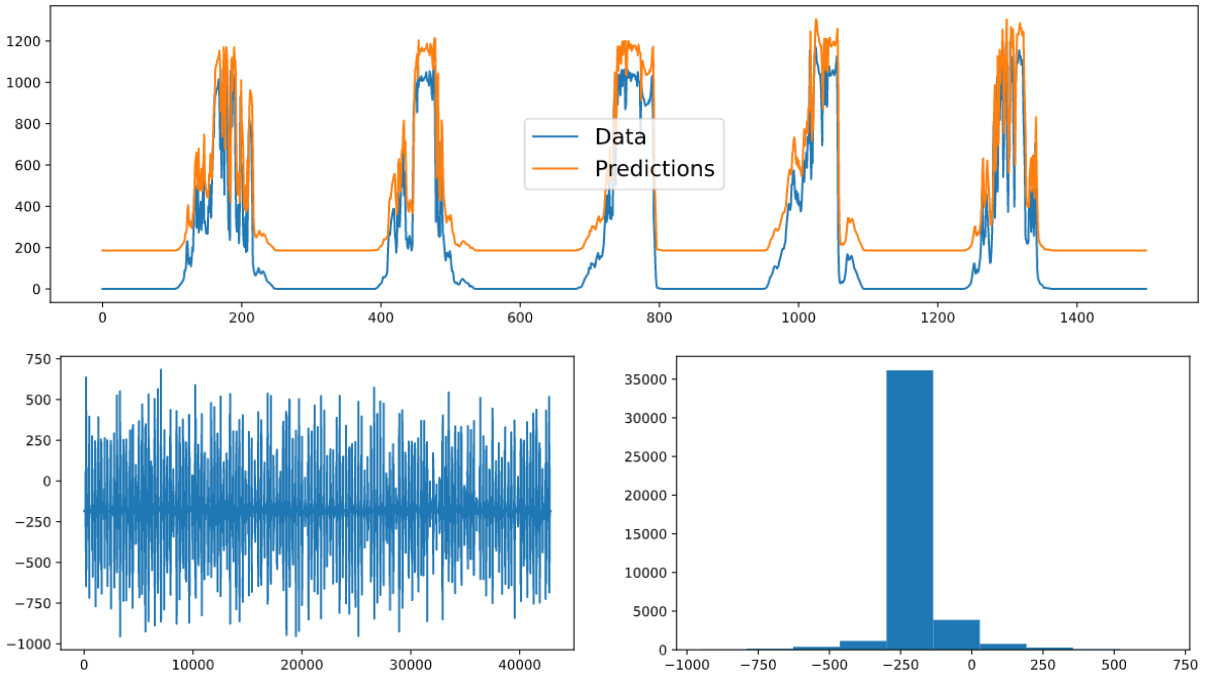

Figura 6: Modelo AR 2: arriba, en azul los valores reales y en naranja los valores predichos, abajo los errores y el histograma de los mismos.

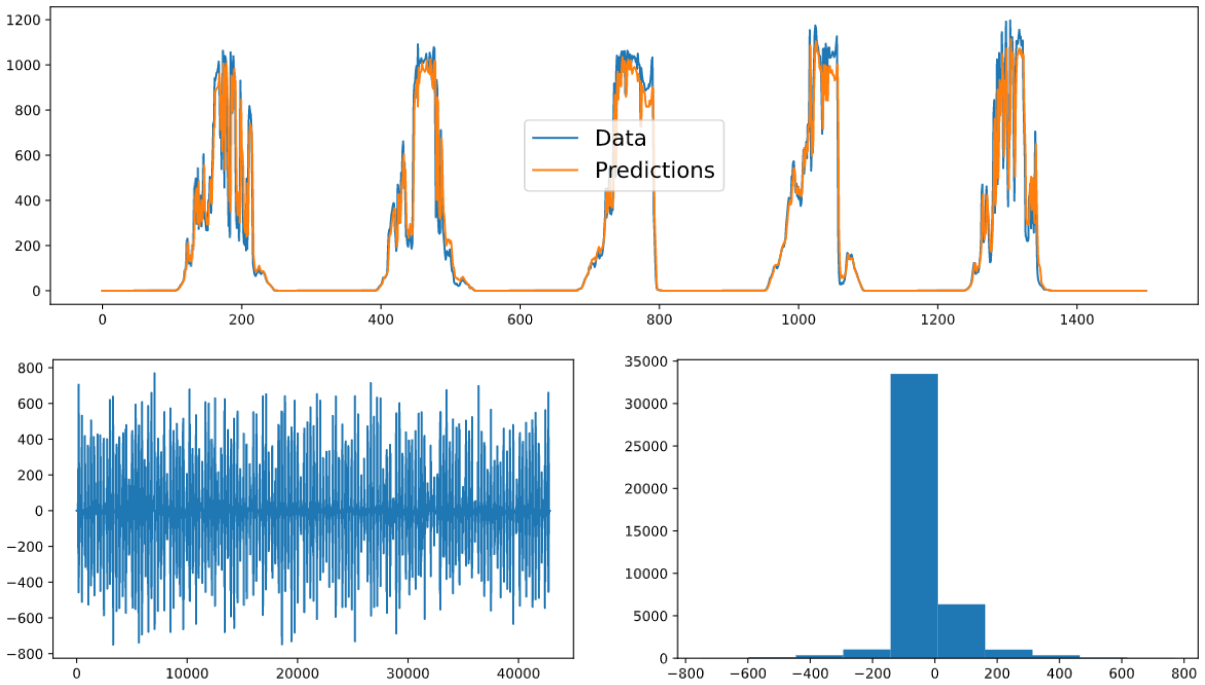

Figura 7: Modelo de regresión lineal: arriba, en azul los valores reales y en naranja los valores predichos, abajo los errores y el histograma de los mismos.

Los modelos ajustados tienen la forma expresada en las ecuaciones 1, 2 y 3 respectivamente. En la tabla los valores de las métricas de desempeño de los modelos; MAPE (Mean absolute percentage error), MSE (Mean squared error), AIC (Akaike information criteria) y BIC (Bayesian information criteria), presentadas en la tabla 4. Donde puede observarse que el modelo media móvil es el más adecuado, con un mejor desempeño en todas las métricas, sin embargo cuando hay muchos valores faltantes seguidos, tiende a generar una línea de tendencia, por lo que en estos casos específicos se utiliza el modelo de regresión lineal.

Tabla 4: Métricas de desempeño para los modelos ajustados insolación solar

Modelo	MAPE	MSE	AIC	BIC
Medias móviles	1.017963	45.902120	163931.74433	163931.744331
AR 2	1284.350963	199.413129	226860.02333	226860.023337
Regresión lineal	1.053306	86.763599	191212.93835	191212.938350

$$I_t = \frac{1}{2}(I_{t-1} + I_{t-2}) \quad (1)$$

(Modelo MM)

$$I_t = 0.8450\, I_{t-1} + 0.1111\, I_{t-2} + 186.1 \quad (2)$$

(Modelo AR 2)

$$I_t = 0.7693\, I_{t-1} + 0.41\, I_{t-2} + 0.2812\, \bar{I}_{tm} - 0.094\, \bar{I}_t \quad (3)$$

(Modelo de regresión)

Donde I_t el la irradiación en el periodo I_{t-1} es la irradiación un periodo antes e I_{t-2} dos periodos antes, $\bar{I}_t$ es el promedio general de radiación en el periodo y $\bar{I}_{tm}$ la radiación promedio para el periodo en el mes m.

Integración de datos.

Después de tener los datos de cada variable en el formato determinado, al módulo anteriormente mencionado se le agrega la función de integrar datos de diferentes variables (tablas) en una sola tabla. Como resultado se obtiene un DataFrame para Epm, un DataFrame para Xm y dos para Meteocontrol; uno para los inversores y uno para la planta. Además se le agregó la capacidad de guardar los datos en archivos csv en caso de que el usuario quiera exportarlos para su uso en otra aplicación.

Para la integración entre datos de las distintas fuentes se define la frecuencia horaria para integrar los datos y realizar análisis para el cliente y la frecuencia de cada dato al momento de realizar análisis de variables individuales. Esto debido al estándar del mercado de calcular consumos y generación por periodos de una hora. Con este fin se construyó un módulo en Python para transformar los datos con frecuencias de 5 minutos de Meteocontrol a datos con frecuencia horaria, donde el valor de una hora es el promedio de los valores registrados desde el primer minuto j pasada hora anterior $i-1$ hasta la hora que se está calculando i, como se muestra en la ecuación 4.

$$Valor_i = \frac{1}{n} \sum_{j \in (i-1)}^{n} j \quad \forall\ i \in horas \tag{4}$$

Además, se construyeron funciones para filtrar la información por fechas indicadas por el usuario e integrar la información que se requiere analizar en conjunto; por ejemplo consumos y precios desde Epm, datos de la planta e información de las simulaciones desde Meteocontrol y precios del mercado desde XM.

Conocimiento del negocio

Teniendo en cuenta que la planta estuvo en proceso de test entre el 28 de febrero de 2020 y el 22 de marzo del mismo año y su funcionamiento continuo empezó el 19 de junio de 2020, los datos utilizados para estos análisis van entre esta fecha y el 30 de abril de 2021. Debido a que los análisis se empezaron antes de esta última fecha mencionada, se utilizaron funciones que permitiera replicar fácilmente el análisis para cualquier fecha, cuyos códigos en Python se encuentran disponibles en el módulo **analice_data** del paquete **power**, desarrollado durante este proyecto y descrito en el apéndice.

Análisis exploratorio de los datos.

Por medio de análisis exploratorio se encontró que la planta solar genera energía entre las 5:15 y las 18:35 horas, es decir en ninguna época del año la planta genera energía entre las 18:40 pm y las 5:10 am. De esta manera se propone este como el horario ideal para los procesos de mantenimiento con el fin de reducir las pérdidas por detener el funcionamiento de la planta.

Explorando los datos extraídos desde Meteocontrol, se encontró que entre la 11 am y la 1 pm se presenta el mayor promedio de generación, 204.04 kWh, como puede verse en la figura 8.

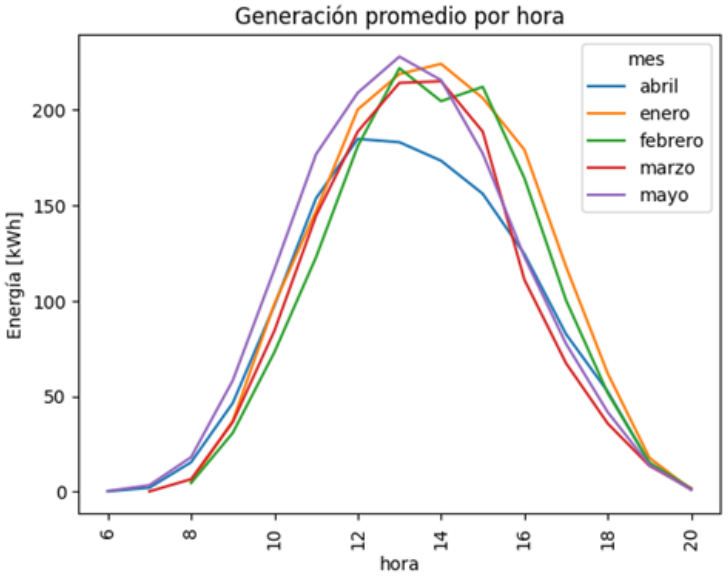

Figura 8: Promedio de generación por hora.

Debido a la relación matemática entre variables como, potencia, voltaje, corriente, etc, que las hace "linealmente dependientes" y por lo tanto no fue de interés analizarlas; para mayor profundidad se puede revisar Hoer (1972). Por lo tanto se procede a explorar la relación entre las variables climáticas de temperatura de módulos e insolación solar con la energía generada, para las horas de generación.. Como puede verse en la figura 9 la energía generada por los módulos es un reflejo de la insolación.

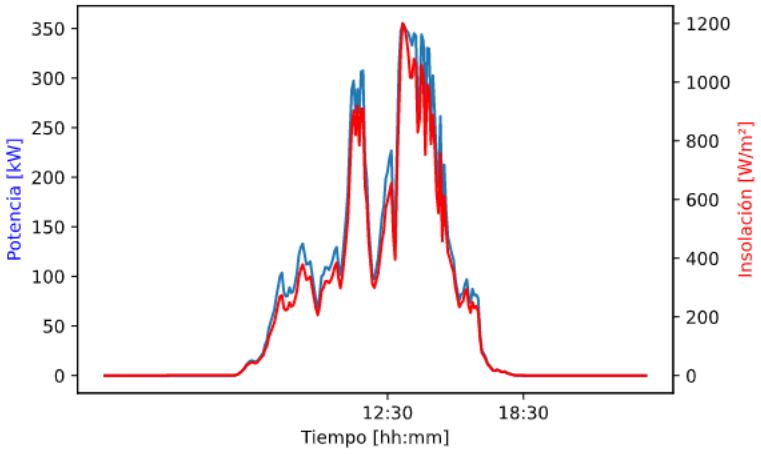

Figura 9: potencia generada en azul vs insolación en color rojo..

Al analizar la correlación, se encontró un valor de 0,979 y 0,942 para la correlación de la energía generada con la insolación solar y la temperatura de los módulos respectivamente y una correlación de 0.966 entre las dos últimas.

La figura 10 muestra los gráficos de correlación entre las variables mencionadas. En el gráfico de correlación entre la insolación y la energía generada, pueden observarse varias tendencias lineales, lo que se explica debido a la degradación de los módulos (Witteck et al., 2017). Esto hace que a medida que pasa el tiempo, menor sea la energía generada con la misma cantidad de irradiación. En los gráficos entre la energía y la temperatura de los módulos, y la temperatura de módulos con la insolación, puede observarse una relación directa pero con un comportamiento ligeramente cóncavo, lo que indica que la tasa de calentamiento de los módulos debido a la radiación, disminuye a medida que aumenta la radiación.

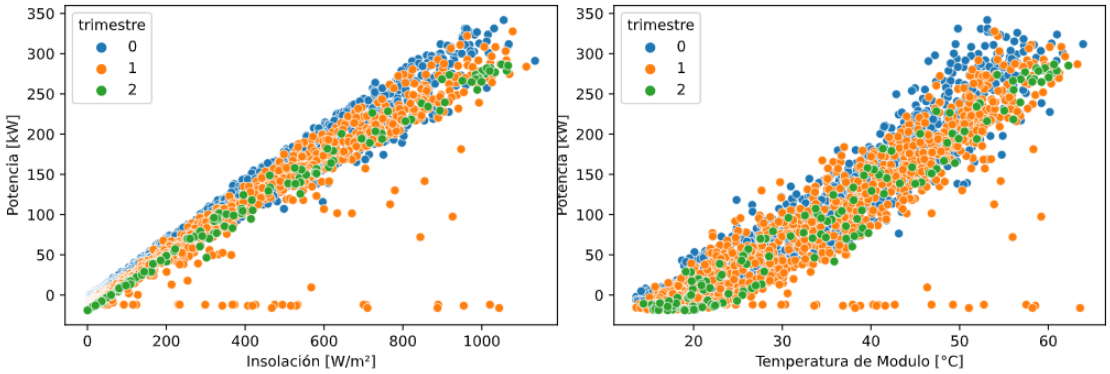

Figura 10: Gráficos de correlación, de izquierda a derecha, energía generada vs Insolación,energía generada vs temperatura módulos.

Utilizando solo registros de datos entre las 5:15 y las 18:35 horas se encontraron periodos en los que la planta no registraba generación de energía, a pesar de que había radiación suficiente para generarla, como pueden observarse en la figura 11, .La figura de la izquierda muestra un

paro programado mientras que la figura de la derecha un paro por mantenimiento.

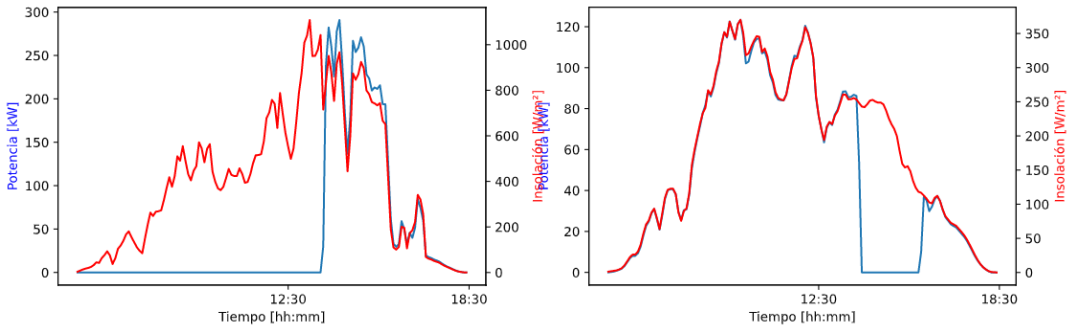

Figura 11: potencia generada en azul vs insolación en color rojo. La figura muestra el comportamiento de paradas de la planta cuando la línea azul indica un valor de cero.

Teniendo en cuenta la relación entre la insolación solar y la potencia generada mostrada en la figura 9, se observaron momentos en que aunque había generación esta no era acorde a la insolación registrada. En la figura 12 se observa un momento en el que a las 10:00 se presentó una potencia por debajo de la esperada con base en la relación entre variables descrita previamente.

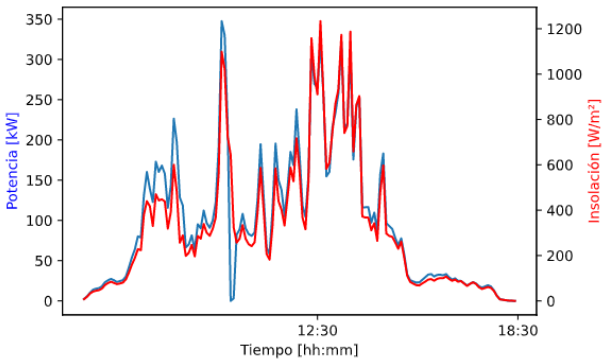

Figura 12: potencia generada en azul vs insolación en color rojo. La figura muestra una generación que no corresponde con la insolación.

Finalmente, explorando los datos de EPM y del mercado, se encontró que la compañía presenta mayor consumo de energía promedio entre las 5 y las 6 pm y un mayor precio promedio kWh consumido de la red entre las 9 y 10 am y entre las 6 y las 9 pm, ver figura 13. Además que

durante el periodo analizado, la planta cubrió el 11% de la energía consumida, en el periodo de observación, con un máximo del 73% de cobertura en una hora. A partir de esta exploración y de comunicación con expertos, se pasó al proceso de definición de indicadores.

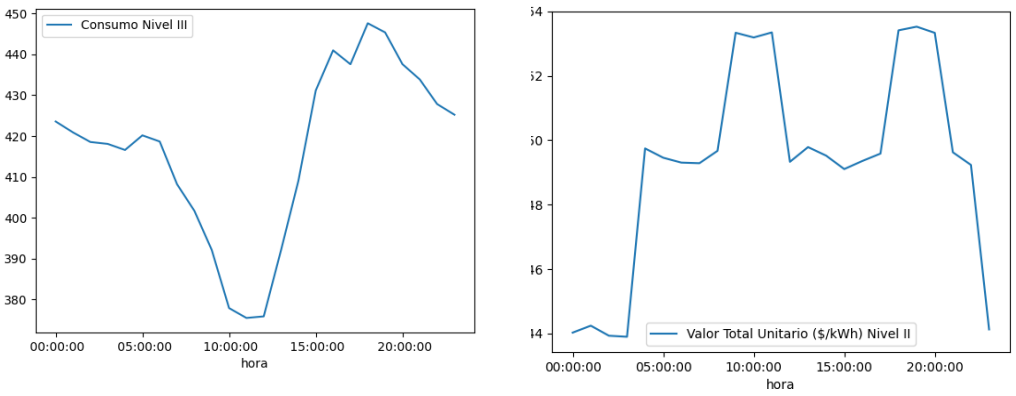

Figura 13: A la izquierda consumo promedio por hora. A la derecha costo promedio por KWh consumido de la red por hora.

Definición de indicadores.

De acuerdo a los intereses de la empresa y en compañía de expertos en el tema se definieron tres tipos de indicadores, operacionales, ambientales, y económicos. Como se había definido anteriormente en los objetivos. Resaltar que los indicadores deben permitir monitorear el comportamiento de la planta en el tiempo, y/o evaluar el desempeño de la misma con respecto al proyectado en las simulaciones realizadas durante la formulación y evaluación del proyecto de implementación de la planta solar.

Indicadores de operación.

El primer indicador es la comparación entre la energía generada por la planta en un periodo de tiempo y la energía pronosticada para este mismo periodo, la comparación se realiza de forma porcentual como se indica en la ecuación 5.

$$C_e = \frac{E_r}{E_s} \cdot 100\% \quad (5)$$

Donde C_e es el cumplimiento en la energía prometida, E_r es la energía real producida y E_s la energía simulada.

Se plantea un indicador para monitorear qué porcentaje de la energía consumida por la compañía es producida por la planta, un indicador que es también de utilidad en caso de determinar el tamaño de una ampliación del generador o la instalación de un sistema de almacenamiento. En este caso se plantean dos indicadores el primero indica cuál es el porcentaje de energía que cubre la planta del total consumido; ver ecuación 6 y el segundo permite observar el porcentaje máximo de energía que ha cubierto la planta en un periodo determinado, ver ecuación 7.

$$Cov_e = \frac{\sum_i E_{ir}}{\sum_i E_{Ci}} \cdot 100\% \quad (6)$$

$$Cov_{max} = max\left(\frac{E_{ir}}{E_{Ci}}\right) \cdot 100\% \quad (7)$$

Donde Cov_e es la cobertura total de la planta y Cov_{max} Es la cobertura máxima en un periodo de tiempo, E_{ir} energía producida por la planta en el tiempo i y E_{ci} Energía consumida por la planta en el tiempo i.

Como se expuso anteriormente en el análisis exploratorio, se lograron detectar paradas de la planta, ver figura 11; momentos en los que no se reportó generación pesa a haber una irradiación solar suficiente para que la planta funcionase. Tras una inspección a las causas de esto se encontró que en su mayoría obedecen a

mantenimientos realizados en horas con buena irradiación solar o procesos de reparación necesarios.

Teniendo en cuenta la relación entre la insolación y la potencia generada, se plantean cuatro indicadores de operación más, los primeros dos identifican los momentos en que la planta estuvo parada y en los que la producción fue inferior a la pronosticada, correspondientes a las ecuaciones 8 y 9, respectivamente. Los dos últimos estiman la energía que dejó de generarse, Tanto por paradas como por producción inferior a la esperada (normal).

$$P_i = \begin{matrix} 1 & si \quad E_i = 0 \ \wedge \ \hat{E}_i > 0 \\ 0 & en\ otro\ caso \end{matrix} \tag{8}$$

$$LP_i = \begin{matrix} 1 & si \quad E_i < \alpha \hat{E}_i \\ 0 & en\ otro\ caso \end{matrix} \tag{9}$$

Donde P_i identifica si hay una parada en el periodo i, E_i es la energía generada en el periodo i, $\hat{E}_i$ es la energía esperada en el periodo i, α es el factor de flexibilidad frente a la predicción y va entre 0 y 1 y LP_i identifica si la energía producida es inferior en el periodo i.

Los últimos dos corresponden a la energía que dejó de generarse por parada de la plata y por producción inferior a la esperada como se muestra en las ecuaciones 10 y 11.

$$E_{Plost} = \sum_i P_i \hat{E}_i \tag{10}$$

$$E_{LPlost} = \sum_i LP_i (\hat{E}_i - E_i) \tag{11}$$

Donde P_i y LP_i identifican si hay una parada en el periodo i y si la energía producida es inferior a la esperada en el periodo i respectivamente. E_i es la energía generada en el periodo i, $\hat{E}_i$ es la energía esperada en el periodo i y, E_{Ploss} y E_{LPloss} son las pérdidas de energía por paradas y por producción inferior a la esperada.

Indicadores ambientales.

Siendo el proyecto el desarrollo de un producto mínimo viable inicialmente se plantea como indicador ambiental, las toneladas de dióxido de carbono (CO_2), que se dejaron de emitir al utilizar energía de la planta en lugar de energía de la red, para realizar este cálculo se multiplica la energía generada por el factor de emisión en Colombia $0.166\ tonCO2eq/MWh$ (Resolución No. 000385 de 2020), tal y como se muestra en la ecuación 12.

$$CO2_{ev} = \sum_i E_{ir} F_{em} \tag{12}$$

Donde $CO2_{ev}$ Es el CO_2 que se dejó de emitir al utilizar energía de la planta y no de la red, E_{ir} energía producida por la planta en el tiempo$i$ y F_{em} es el factor de emisión por unidad de energía producida en Colombia.

Indicadores económicos.

Para monitorear la rentabilidad de la planta se calculan los ahorros en pesos colombianos (COP) teniendo en cuenta el costo de energía de la red y el costo de la energía de la planta ecuación 13.

$$Ah = E_t * Ce_t - Cp_t \tag{13}$$

Donde Ah son los ahorros en COP E_t es la energía generada por la planta en el periodo t, Ce_t es el costo por unidad de energía en el periodo t y Cp_t Son los costos de la planta en el periodo t.

De la misma forma que en los indicadores de operación se definió la importancia de contabilizar también en dinero las pérdidas ocasionadas por paradas de la planta y por producción inferior a la esperada, como se muestra en las ecuaciones 14 y 15.

$$M_{Plost} = \sum_i P_i \hat{E}_i C_i \tag{14}$$

$$M_{LPlost} = \sum_i LP_i * (\hat{E}_i - E_i) * C_i \tag{15}$$

Donde P_i y LP_i identifican si hay una parada en el periodo i y si la energía producida es inferior a la esperada en el periodo i respectivamente. E_i es la energía generada en el periodo i, $\hat{E}_i$ es la energía esperada en el periodo i, C_i es el costo por unidad de energía en el periodo i y, M_{Ploss} y M_{LPloss} son las pérdidas en dinero por paradas y por producción inferior a la esperada.

Creación de modelos.

Observando las ecuaciones 8, 9, 10 , 11, 14 y 15 se observa que es necesario conocer la energía esperada en el periodo ($\hat{E}_i$), para el cálculo de los indicadores. Desde Meteocontrol se obtiene una predicción de la energía esperada durante un periodo, sin embargo muchos de los datos faltantes en esta variables coinciden con las paradas de la planta, por lo que es necesaria la implementación de un modelo que ayude a predecir la energía que debía haber generado la planta en cada periodo.

Teniendo en cuenta el análisis exploratorio se tienen en cuenta dos factores para el ajuste de modelos; la relación entre la potencia y las variables climáticas y la degradación que tienen los módulos solares con el tiempo y la insolación recibida. Se plantean entonces tres modelos, dos de

estos, basados en la regresión lineal, relacionan la potencia generada, con la insolación y la temperatura de módulo, respectivamente; no se planteó un modelo utilizando las dos variables climáticas como predictoras, debido a que como se observa en la figura 9, estas son linealmente dependientes. El tercer modelo se basa en predecir el coeficiente β entre la potencia generada y la insolación, y multiplicarlo por esta última.

Para la limpieza de los datos, primero se eliminan los registros en los cuales la insolación y/o la radiación son valores faltantes. Luego se calcula el coeficiente βque se utiliza para eliminar valores atípicos, como se observa en la figura 14.

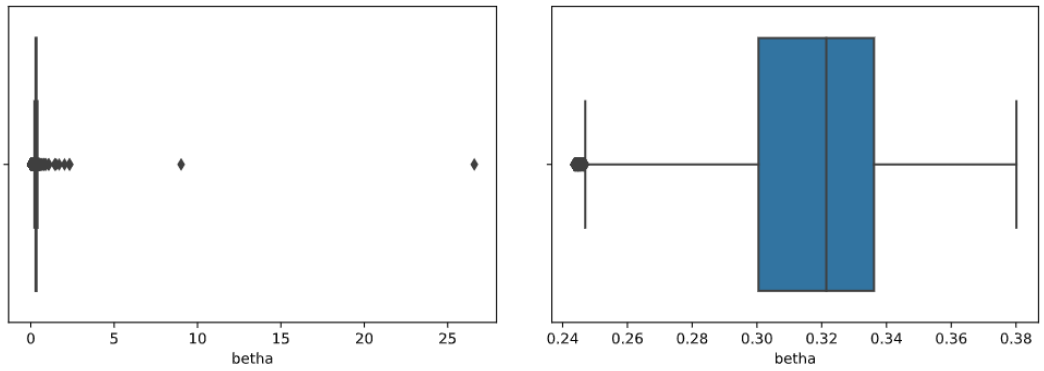

Figura 14: distribución del coeficiente, antes y después de eliminar outliers una vez.

Teniendo los datos limpios se procede a ajustar los dos modelos de regresión lineal, como se muestran en las ecuaciones 16 y 17, con los ajustes mostrados en las figuras 12 y 13.

$$P_t = 0.3013\, I_t + 2.7892 \tag{16}$$

(Modelo regresión (potencia~insolación))

$$P_t = 6.8784\, Tm_t - 127.042 \tag{17}$$

(Modelo regresión (potencia~temperatura módulo))

Donde P_t es la potencia generada en el periodo t. I_{t^r} es la insolación solar en el periodo y Tm_t la temperatura del módulo en el período t.

De los resultados, se evidencia un mejor desempeño en la predicción de la potencia generada, en función de la insolación que cuando se utiliza la temperatura de módulo, esto se debe a que la temperatura tiene un efecto acomulativo alrededor del día y un cambio más lento, lo que genera una predicción retrasada, como se puede ver en la figura 14, en cambio cuando se utiliza la insolación, se observa una subestimación de la potencia generada los primeros meses y una sobre estimación de los generados en los últimos meses, esto debido a que no tiene en cuenta la degradación de los módulos.

Para hallar el coeficiente entre la potencia generada y la insolación, se procede a calcular el cociente entre ambas y se realiza un análisis exploratorio de esta nueva variable β. Teniendo en cuenta que la insolación depende de la hora, por medio de gráficos boxplot; ver figura 17, se comprueba la independencia del coeficiente con respecto a la hora y a la insolación acumulada (en percentiles). Por lo que se procede a analizar si es relevante en la degradación de los módulos y por lo tanto en el valor de este coeficiente. Tras realizar los gráficos ACF y PACF de la figura 18, se concluye que β está relacionado con su valor dos periodos anteriores, por lo que se procede a calcular el valor de β por medio de un modelo AR(2) y un modelo MM(2).

Se grafican los valores reales contra los esperados, de cada modelo; ver figuras 15, 16, 19 y 20, y se calculan el MSE, MAPE, AIC y BIC, de cada uno; tabla 5, Y con base en esto se elige el modelo $\beta(MM(2))$, ecuación 18, ya que, los modelos de regresión no incluyen la degradación por lo que presentan el peor desempeño y no son confiables a largo plazo, por otro lados es más simple y presenta un mejor desempeño que el modelo $\beta(AR(2))$ ecuación 19.

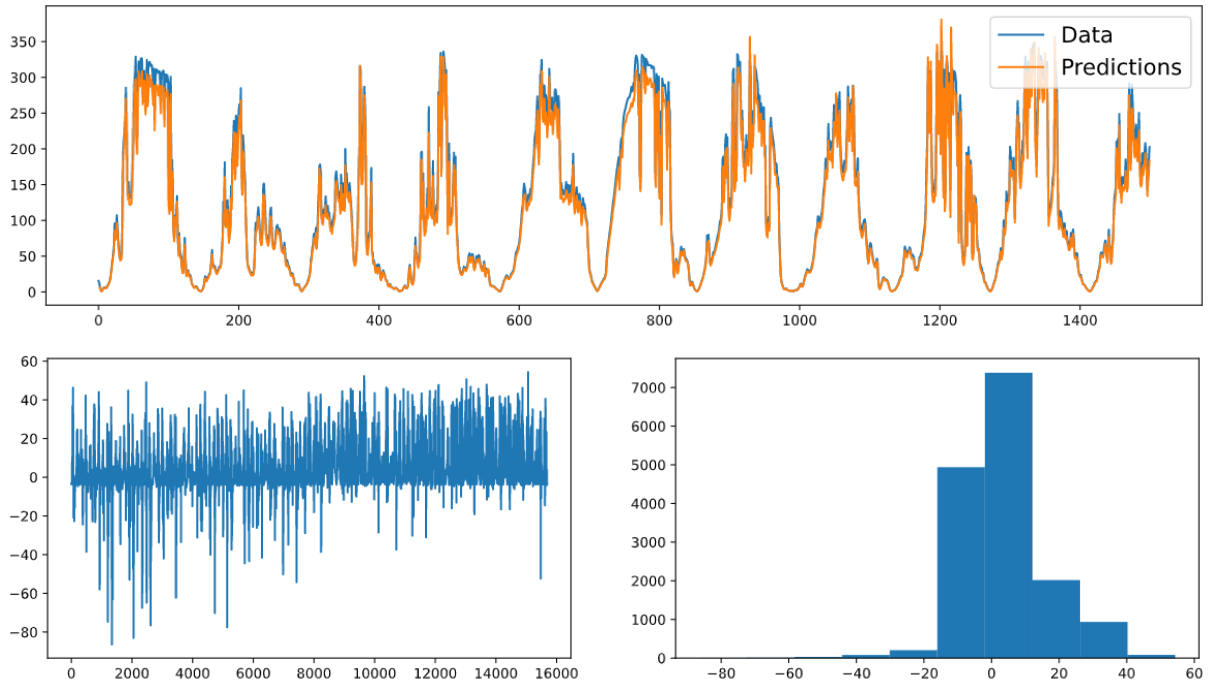

Figura 15: Modelo de regresión lineal: potencia en función de la insolación. Arriba, en azul los valores reales y en naranja los valores predichos, abajo los errores y el histograma de los mismos.

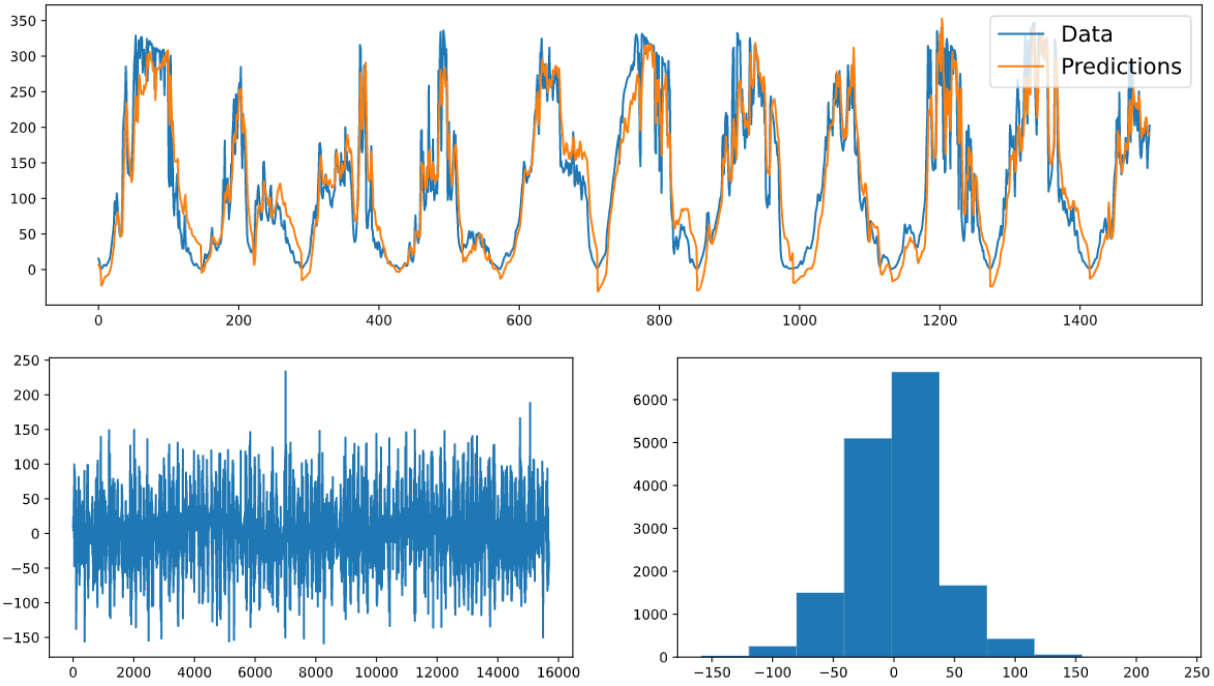

Figura 16: Modelo de regresión lineal: potencia en función de la temperatura del módulo. Arriba, en azul los valores reales y en naranja los valores predichos, abajo los errores y el histograma de los mismos.

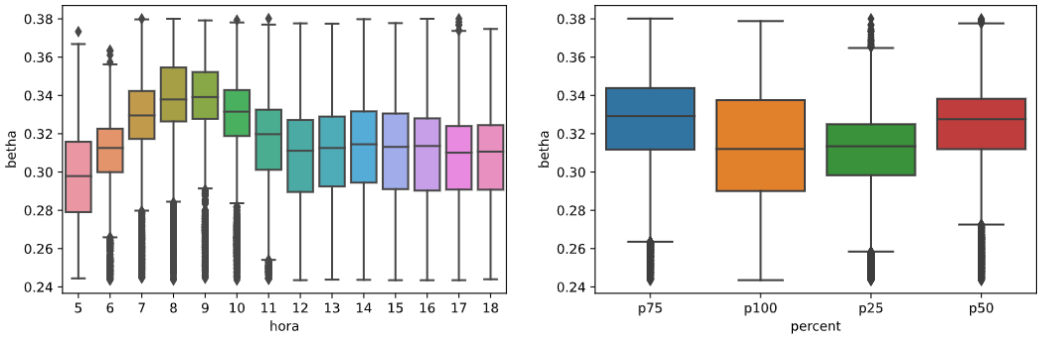

Figura 17: distribución de coeficiente, diferenciada por hora; a la izquierda y por radiación (Cuartil), a la derecha.

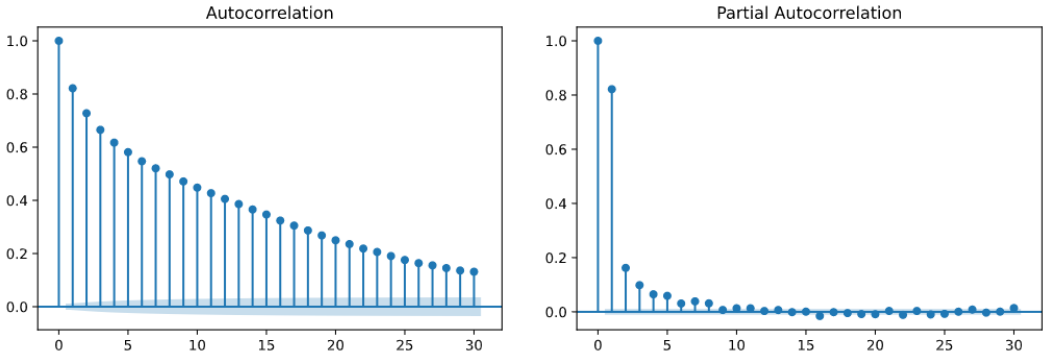

Figura 18: ACF y PACF coeficiente potencia generada/insolación solar.

$$\beta_t = 0.7035\,\beta_{t-1} + 0.2187\,\beta_{t-2} + 0.03136 \quad (18)$$

(Modelo AR(2))

$$\beta_t = (\beta_{t-1} + \beta_{t-2})/2 \quad (19)$$

(Modelo MM(2))

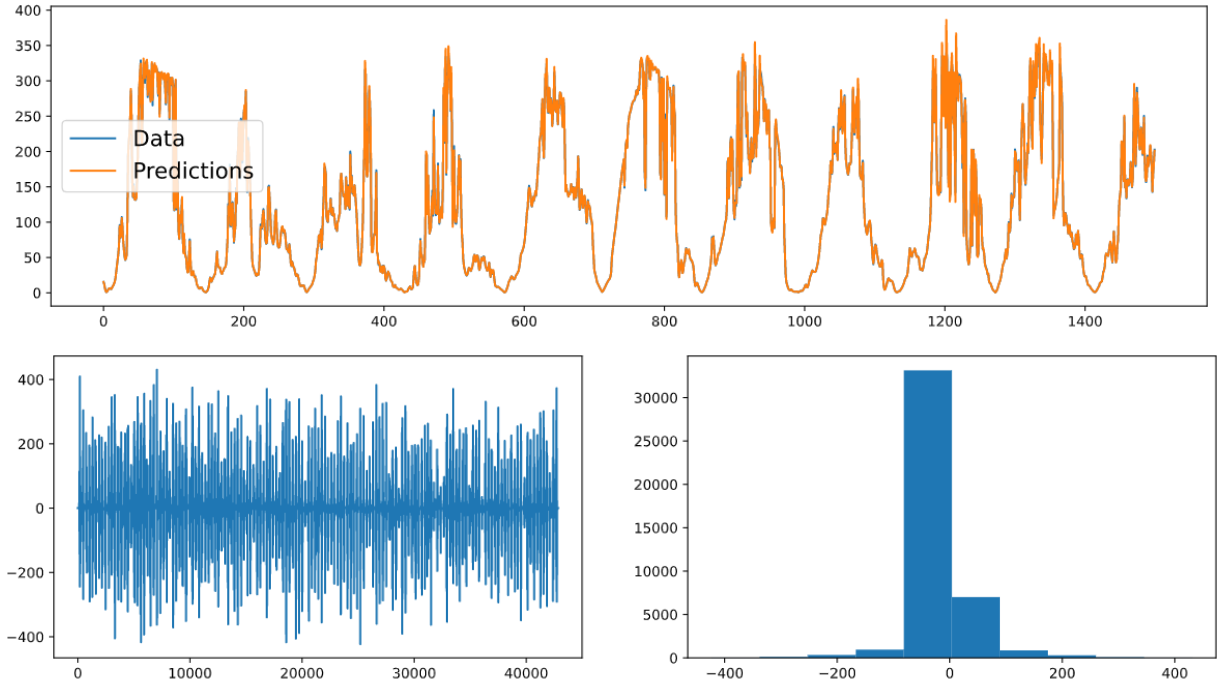

Figura 19: Potencia generada, en función de la insolación y el coeficiente calculado. Arriba, en azul los valores reales y en naranja los valores predichos, abajo los errores y el histograma de los mismos.

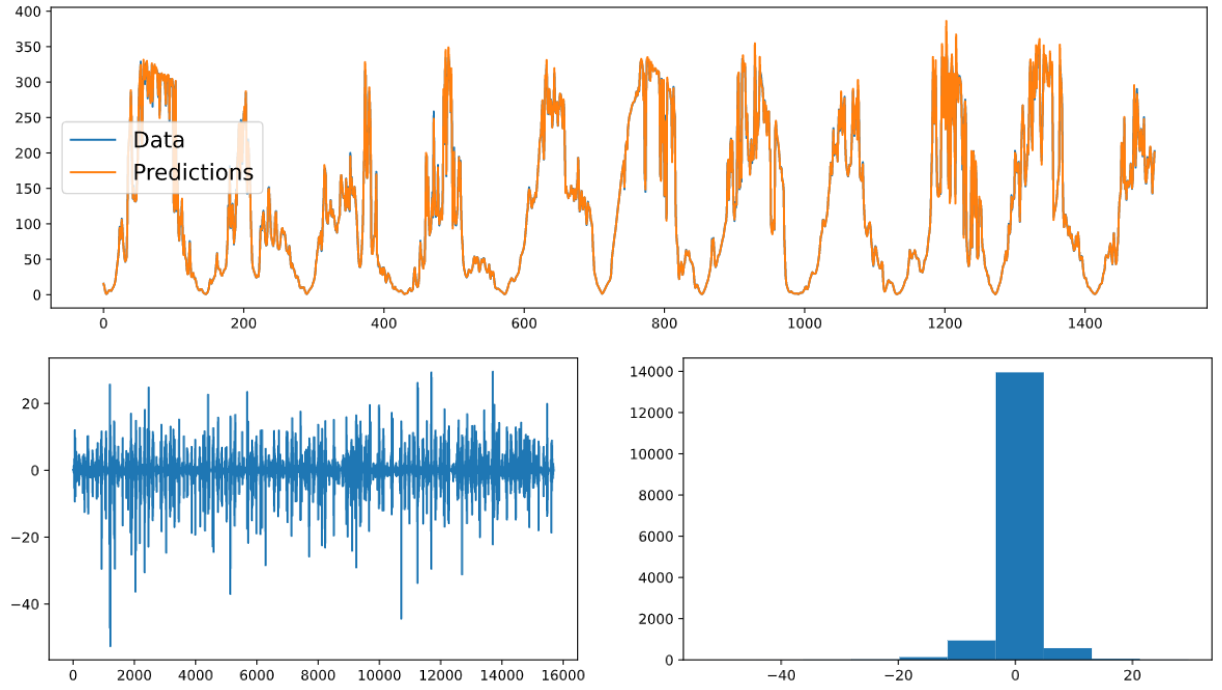

Figura 20: Potencia generada, en función de la insolación y el coeficiente calculado. Arriba, en azul los valores reales y en naranja los valores predichos, abajo los errores y el histograma de los mismos.

Tabla 5: Métricas de desempeño para los modelos para predecir la potencia esperada.

Modelo	MAPE	MSE	AIC	BIC
$P_t = f(I_t) + b$	1.137390	12.667716	39845.261137	39845.261137
$P_t = f(Tm_t) + b$	1.602509	37.611663	56918.932169	56918.932169
$P_t = \beta(AR(2)) \cdot I_t$	1.000411	12.490048	39619.660264	39619.660264
$P_t = \beta(MM(2)) \cdot I_t$	0.923048	3.274886	18621.594075	18621.594075

Selección de la información.

Teniendo en cuenta la variabilidad climática y el cálculo mensual de los costos, el informe se realiza para un periodo mínimo de un mes. utilizando el siguiente formato para presentar los indicadores en este.

Indicadores de operación.

Los indicadores de energía generada vs pronosticada se presentan en dos formatos, el primero como una suma del total de energía generada vs el total de la energía que se esperaba de acuerdo a lo planteado al inicio del proyecto; para facilitar la comparación global, y la segunda un gráfico de barras agrupadas por mes (ver la figura 21), comparando la energía generada por la planta y la esperada de acuerdo al pronóstico utilizado en el planteamiento del proyecto.

El porcentaje de energía cubierta por la planta, se muestra el máximo y el promedio como valores (números), mientras para el periodo se muestra el promedio agrupado por hora en un gráfico de barras como se muestra en la figura 22a, además se agregó un

gráfico de barras que permite ver la energía consumida por mes desde la red y de la planta, ver figura 22b.

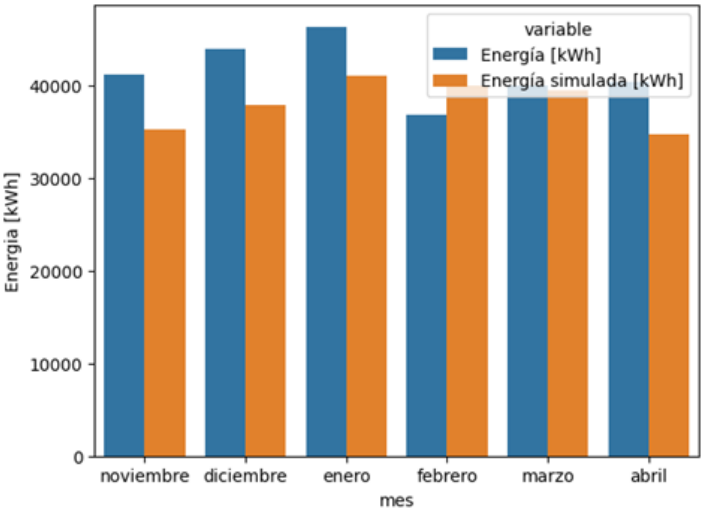

Figura 21: Energía generada y energía esperada por mes.

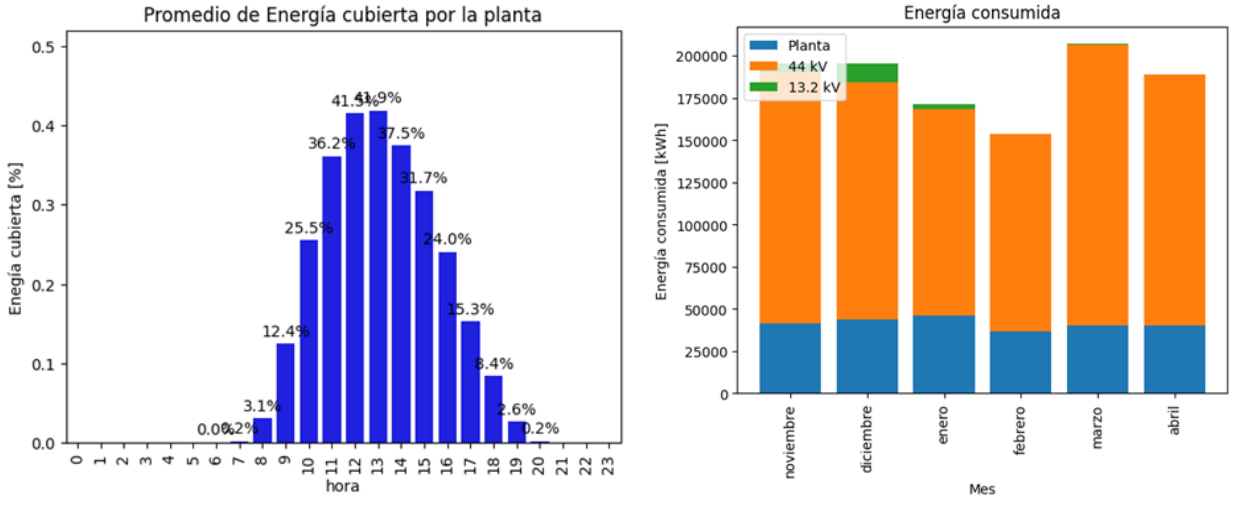

Figura 22: A la izquierda, porcentaje promedio de energía cubierta por la planta, por hora y por mes. A la derecha energía consumida por fuente por mes.

Para identificar las causas de las paradas se utiliza la fecha y hora en la que ocurrieron, permitiendo investigar las causas directamente en la planta, para facilitar la evaluación del impacto para cada una de las paradas se, se identificó cuanto tiempo de esta parada era productivo (cuánto tiempo de la parada la planta hubiese generado energía), las pérdidas de energía y dinero estimadas, esta información se presenta en formato de la tabla 6. Finalmente se decidió realizar un gráfico de barras para agrupar las pérdidas de energía por paradas y por producción inferior a la esperada, ver figura 23.

Tabla 6: Tabla para identificar paradas

Inicio	Perdida [h]	Pérdida [kWh]	Pérdida [$]
2021-04-09 16:00	2	139.374	40058

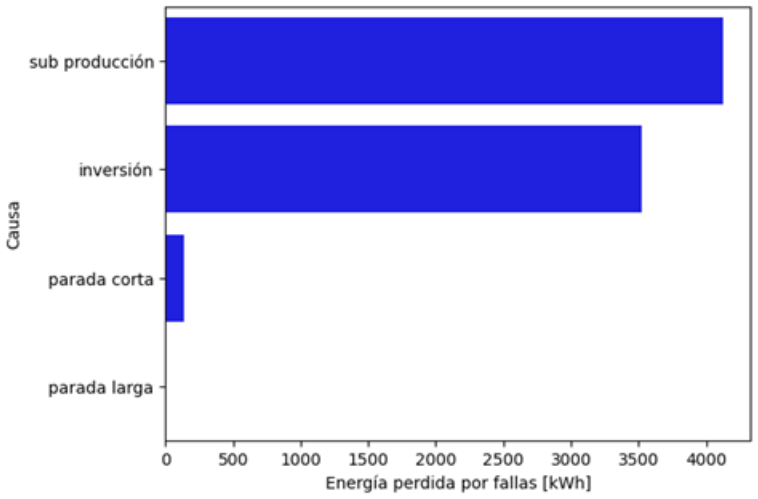

Figura 23: Pérdidas de energía por paradas y por producción inferior a la esperada.

Indicadores ambientales.

Las toneladas de CO2 evitadas, se presentan primero como el total durante el periodo indicado, además de un gráfico de barras de las toneladas de CO2 evitadas por mes ver figura 24.

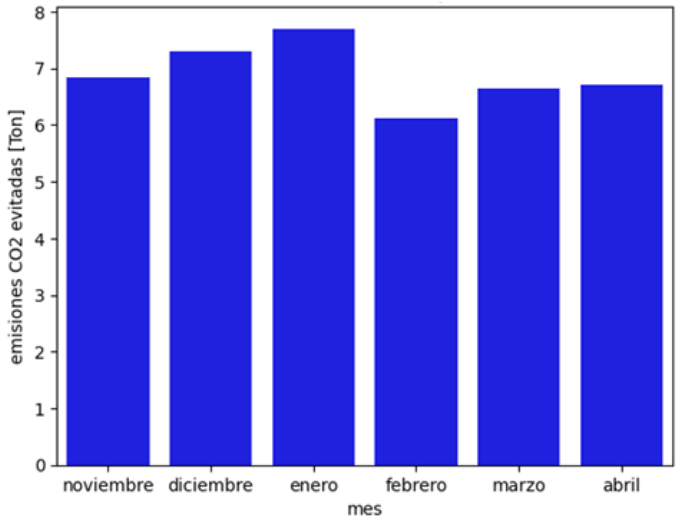

Figura 24: Emisiones de CO2 evitadas por mes gracias a la planta solar.

Indicadores económicos.

Siguiendo el formato de la energía generada y las emisiones de CO2 evitadas, los ahorros en dinero(COP) se presentan en dos como el total durante el periodo indicado, y como un gráfico de barras por mes ver figura 25. De la misma forma que para las pérdidas de energía las pérdidas estimadas en dinero por paradas y por producción inferior a la esperada, se agrupan en un gráfico ver figura 26

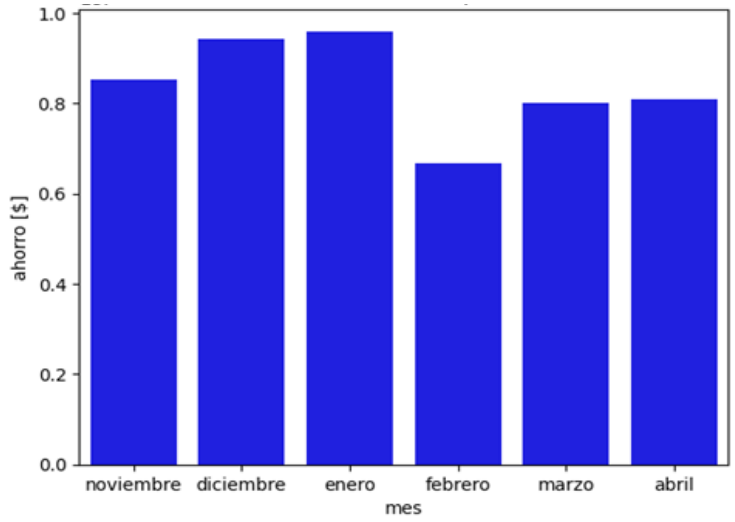

Figura 25: Ahorros por mes en COP.

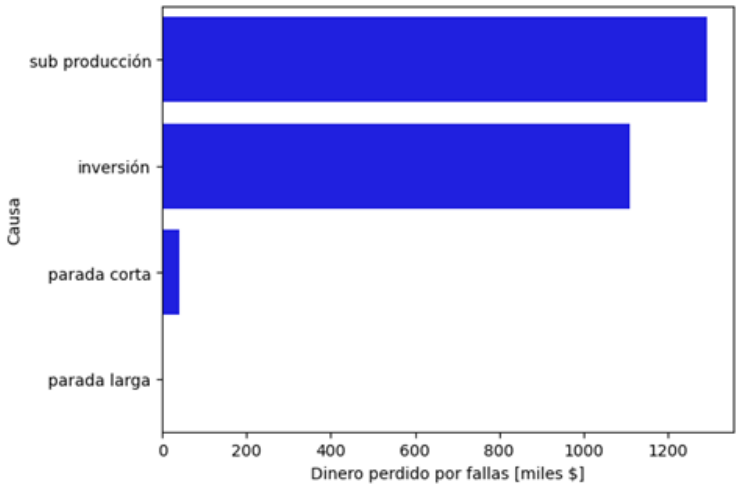

Figura 26: Pérdidas de dinero por paradas y por producción de energía inferior a la esperada.

Estandarización y automatización del proceso

Para almacenar los datos se definen dos tipos de archivo, archivos csv para almacenar los datos extraídos de las diferentes fuentes, ya que es fácil de transformar a otros formatos y permite utilizar el formato 3NF, y archivos json para almacenar datos como información de la planta y contraseñas, que al ser más flexibles permite el almacenamiento de nueva información sin afectar la información previa. Por seguridad, estas últimas (contraseñas) se encriptan automáticamente. Para que los datos estén accesibles para diferentes análisis y facilitar compartirlo con cualquier miembro de la organización que lo solicite, los datos procesados se almacenaron en sharepoint.

Habiendo desarrollado todo el proyecto utilizando módulos de Python, se decide empaquetar todo en un paquete de python llamado **power**. Para facilitar el uso del paquete con diferentes clientes, el paquete tiene un módulo que permite la creación de una base de datos para un nuevo cliente o para una nueva planta.

Para automatizar la descarga automática de los datos se programaron dos bots en Python utilizando **selenium** (Selenium, n.d.), para extraer información desde Meteocontrol y Epm; para los datos disponibles en formato csv, el bot realiza automáticamente la descarga de los archivos, para aquellos que no se realiza web scraping. Para extraer la información de XM, se utiliza la API disponible (XM, 2021) también en Python. Todos incluidos en módulos del paquete.

Posterior a la descarga otras funciones en el paquete procesan los datos automáticamente, estandarizando los valores como se indica en la tabla 3 y el almacenamiento en archivos csv. Otro módulo del paquete permite al usuario generar un informe entre dos fechas elegidas por este. Para generar el informe se utiliza un archivo en word, ya que permite editar fácilmente el informe posteriormente y es un formato estándar.

Para mejorar la experiencia por parte de usuarios sin conocimientos en programación, se desarrolló una interfaz gráfica de usuario (GUI) utilizando el paquete Dear PyGui, compuesta de 4 pestañas, como se puede ver en a figura 27, la pestaña de inicio que muestra el logo del paquete, la pestaña Nuevo cliente, que permite la configuración del paquete para un nuevo cliente, la pestaña datos que permite la descarga y procesamiento automático de los datos, adicionalmente se le agregó la opción de cargar datos de forma manual. Y la pestaña de informe que permite generar el informe entre dos fechas indicadas por el usuario.

Figura 27: Home GUI de paquete power

Finalmente para facilitar la continua gestión y actualización del paquete se utiliza git como sistema de control de versiones y se almacena en un repositorio de github, con su respectiva documentación, guía de instalación y uso. Adicionalmente tanto el paquete como la GUI, se desarrollaron de forma modular para facilitar la integración de nuevas funcionalidades. La figura 28 resume el proceso realizado por el paquete. Debido a la sensibilidad de algunos datos y métricas de negocio incluidas en el paquete, en el momento el repositorio del mismo se encuentra en modo privado y por el momento, en caso de querer acceder al mismo, debe solicitarse el acceso.

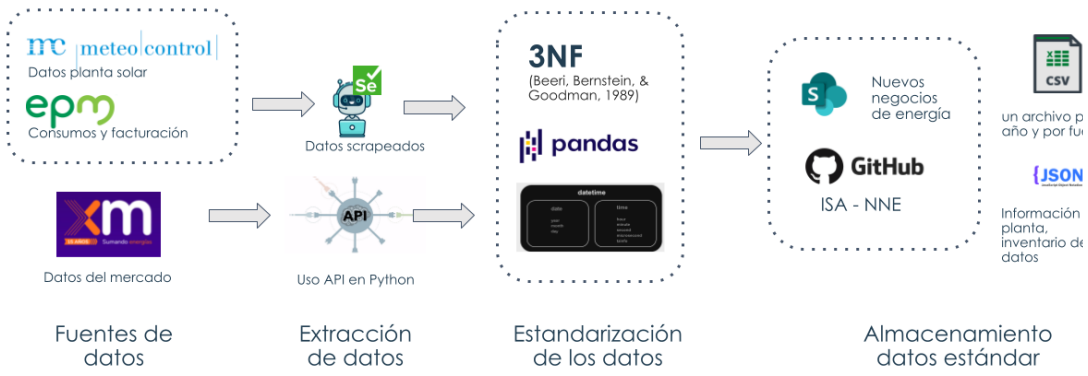

Figura 28: Arquitectura de automatización del proceso

Conclusiones

El proyecto permitió desarrollar con éxito indicadores económicos, ambientales y de operación, y con el software desarrollado se logró automatizar el proceso de generación del informe.

Cabe resaltar la importancia de realizar un proceso de gestión de metadatos, para facilitar la correcta integración de diferentes fuentes de datos y de calidad de datos para garantizar la confiabilidad de indicadores generados a partir de los mismos. De la misma manera lo es el trabajo en compañía de expertos en el negocio para facilitar la definición de indicadores.

Por otra parte, cuando se desarrollan soluciones basadas en software, es de vital importancia tener en cuenta los usuarios de las mismas, e incluir una buena documentación y guia de usuario. Optar por la inclusión de una interfaz gráfica es una buena solución para que todo tipo de usuario pueda acceder a este con mayor facilidad.

Finalmente, el desarrollo modular del software, acompañado de un sistema de control de versiones, le brinda a ISA la posibilidad de integrar nuevas funcionalidades al software e implementar procesos de despliegue continuo e integración continua en próximas versiones del paquete. Permitiendo aumentar su portafolio de servicios de valor agregado.

Como trabajo futuro se vislumbra la implementación del software para otros usuarios de la compañía, la incorporación de otras variables climáticas como la velocidad del viento y la humedad para la predicción de energía esperada con más precisión, el uso de una base de datos dedicada para mejorar la disponibilidad de los datos y la confiabilidad del servicio. Finalmente también existe la posibilidad de incluir nuevas funcionalidades, por ejemplo la optimización de costos con la incorporación de baterías.

Referencias Bibliográficas

Bauman, K., Tuzhilin, A., & Zaczynski, R. (2017). Using social sensors for detecting emergency events. ACM Transactions on Management Information Systems, 8(2–3), 1–20. https://doi.org/10.1145/3052931

Beeri, C., Bernstein, P. A., & Goodman, N. (1989). A Sophisticate's Introduction To Database Normalization Theory, This work was supported in part by the National Science Foundation under Grant MCS-77-05314. In Readings in Artificial Intelligence and Databases (pp. 468–479). Elsevier. http://dx.doi.org/10.1016/b978-0-934613-53-8.50035-2

BNamericas. (2019, December 10). El estado de la generación distribuida en Latinoamérica de cara a 2020. BNamericas. https://www.bnamericas.com/es/reportajes/el-estado-de-la-generacion-distribuida-en-latinoamerica-de-cara-a-2020

Brackett, M., Earley, S., & Henderson, D. (2009). he DAMA Guide to The Data Management Body of Knowledge: DAMA-DMBOK Guide (1st ed.). Technics Publications.

Cai, Y., & Chow, M.-Y. (2009, July). Exploratory analysis of massive data for distribution fault diagnosis in smart grids. 2009 IEEE Power & Energy Society General Meeting. http://dx.doi.org/10.1109/pes.2009.5275689

Consorcio Hart-re. (2014). Capacidad Instalada de Autogeneración y Cogeneración en Sector de Industria, Petróleo, Comercio y Público del País. Unidad De Planeación Minero Energética. https://www1.upme.gov.co/DemandaEnergetica/1_Informe_final_auto_cogeneracion.pdf

Cozzi, A., & Gould, C. (2016). Energy and Air Pollution: World Energy Outlook Special Report 2016. OECD/IEA. http://pure.iiasa.ac.at/id/eprint/13467/1/WorldEnergyOutlookSpecialReport2016EnergyandAirPollution.pdf

Diamantoulakis, P. D., Kapinas, V. M., & Karagiannidis, G. K. (2015). Big data analytics for dynamic energy management in smart grids. Big Data Research, 2(3), 94–101. https://doi.org/10.1016/j.bdr.2015.03.003

Dulău, L. I., Abrudean, M., & Bică, D. (2014). Effects of distributed generation on electric power systems. Procedia Technology, 12, 681–686. https://doi.org/10.1016/j.protcy.2013.12.549

García de Fonseca, L., Parikh, M., & Manghani, R. (2019). Evolución futura de costos de las energías renovables y almacenamiento en América Latina. Inter-American Development Bank. http://dx.doi.org/10.18235/0002101

González Matilla, J. M., Daza Duque, C. A., & Urueña Galeano, C. H. (2008). Análisis del esquema de generación distribuida como una opción para el sistema eléctrico colombiano. Revista Facultad de Ingeniería Universidad de Antioquia, 44(ISSN 2422-2844), 97–110. Scielo.

Hoer, C. A. (1972). The six-port coupler: A new approach to measuring voltage, current, power, impedance, and phase. IEEE Transactions on Instrumentation and Measurement, 21(4), 466–470. https://doi.org/10.1109/tim.1972.4314068

Makridakis, S., & Hibon, M. (1997). ARMA models and the box-jenkins methodology. Journal of Forecasting, 16(3), 147–163.

https://doi.org/10.1002/(sici)1099-131x(199705)16:3<147::aid-for652>3.0.co;2-x

Mitra, J., & Suryanarayanan, S. (2010, July). System analytics for smart microgrids. IEEE PES General Meeting. http://dx.doi.org/10.1109/pes.2010.5589700

Ozcanli, A. K., Yaprakdal, F., Baysal, M. (2020). Deep learning methods and applications for electrical power systems: A comprehensive review. International Journal of Energy Research, 44(9), 7136–7157. https://doi.org/10.1002/er.5331

Parisio, A., Rikos, E., & Glielmo, L. (2014). A model predictive control approach to microgrid operation optimization. IEEE Transactions on Control Systems Technology, 22(5), 1813–1827. https://doi.org/10.1109/tcst.2013.2295737

Perez, E. (2018). Retos de la Integración de Recursos Energéticos Distribuidos. Energética 2030.

Quintero, J. P. V. (2008). Generación distribuida: Democratización de la energía eléctrica. Criterio Libre, 6(8), 105–112. https://doi.org/https://dialnet.unirioja.es/descarga/articulo/4547088.pdf

Ruan, Y., Liu, Q., Zhou, W., Firestone, R., Gao, W., & Watanabe, T. (2009). Optimal option of distributed generation technologies for various commercial buildings. Applied Energy, 86(9), 1641–1653. https://doi.org/10.1016/j.apenergy.2009.01.016

Selenium. (n.d.). Documentation. Selenium HQ. Retrieved February 11, 2021, from https://www.selenium.dev/

Seyedi, Y., Karimi, H., & Grijalva, S. (2019). Irregularity detection in output power of distributed energy resources using PMU data analytics in smart grids. IEEE Transactions on Industrial Informatics, 15(4), 2222–2232. https://doi.org/10.1109/tii.2018.2865765

Simmhan, Y., Aman, S., Kumbhare, A., Liu, R., Stevens, S., Zhou, Q., Prasanna, V. (2013). Cloud-Based software platform for big data analytics in smart grids. Computing in Science & Engineering, 15(4), 38–47. https://doi.org/10.1109/mcse.2013.39

Resolución No. 000385 de 2020, 2 (2020). https://www1.upme.gov.co/Normatividad/385_2020.pdf

Witteck, R., Veith-Wolf, B., Schulte-Huxel, H., Morlier, A., Vogt, M. R., Köntges, M., & Brendel, R. (2017). UV-induced degradation of PERC solar modules with UV-transparent encapsulation materials. Progress in Photovoltaics: Research and Applications, 25(6), 409–416. https://doi.org/10.1002/pip.2861

XM. (2021, January 20). GET api. ASP.NET. https://serviciofederacion.XM.com.co/

Zhang, Y., Huang, T., & Bompard, E. F. (2018). Big data analytics in smart grids: A review. Energy Informatics, 1(1). https://doi.org/10.1186/s42162-018-0007-5

Anexos

Guía para uso del paquete