

UNIVERSIDAD DE EL SALVADOR
FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA DE INGENIERÍA ELÉCTRICA



**CÁLCULO DE INDICADORES SAIFI, SAIDI, TDF Y TMR
UTILIZANDO HERRAMIENTAS DE MANEJO DE GRANDES
VOLÚMENES DE DATOS**

PRESENTADO POR:

**CASTRO CONTRERAS, FIDEL ERNESTO
ROMERO PALACIOS, HAROLD ERNESTO**

PARA OPTAR AL TÍTULO DE:
INGENIERO ELECTRICISTA

CIUDAD UNIVERSITARIA, MAYO DE 2022

UNIVERSIDAD DE EL SALVADOR

RECTOR:

MSC. ROGER ARMANDO ARIAS ALVARADO

SECRETARIO GENERAL:

ING. FRANCISCO ANTONIO ALARCON SANDOVAL

FACULTAD DE INGENIERIA Y ARQUITECTURA

DECANO:

PhD. EDGAR ARMANDO PEÑA FIGUEROA

SECRETARIO:

ING. JULIO ALBERTO PORTILLO

ESCUELA DE INGENIERÍA ELÉCTRICA

DIRECTOR:

ING. ARMANDO MARTÍNEZCALDERÓN

UNIVERSIDAD DE EL SALVADOR
FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA DE INGENIERÍA ELÉCTRICA

Trabajo de Graduación previo a la opción al Grado de:

INGENIERO ELECTRICISTA

Título:

**CÁLCULO DE INDICADORES SAIFI, SAIDI, TDF Y TMR
UTILIZANDO HERRAMIENTAS DE MANEJO DE
GRANDES VOLÚMENES DE DATOS**

Presentado por:

CASTRO CONTRERAS, FIDEL ERNESTO

ROMERO PALACIOS, HAROLD ERNESTO

Trabajo de Graduación Aprobado por:

Docente Asesor:

DR. CARLOS EUGENIO MARTINEZ CRUZ

SAN SALVADOR, MAYO DE 2022

Trabajo de Graduación Aprobado por:

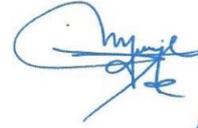
Docente Asesor:

DR. CARLOS EUGENIO MARTINEZ CRUZ

NOTA Y DEFENSA FINAL

En esta fecha, miércoles 30 de marzo de 2022, en la Sala de Lectura de la Escuela de Ingeniería Eléctrica, a las 5:00 p.m. horas, en presencia de las siguientes autoridades de la Escuela de Ingeniería Eléctrica de la Universidad de El Salvador:

1. Ing. Armando Martínez Calderón
Director



Firma



2. MSc. José Wilber Calderón Urrutia
Secretario



Firma

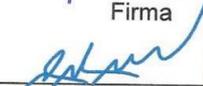
Y, con el Honorable Jurado de Evaluación integrado por las personas siguientes:

- DR. CARLOS EUGENIO MARTINEZ CRUZ
(Docente Asesor)



Firma

- ING. JOSE ROBERTO RAMOS



Firma

- ING. ANA MARIA FIGUEROA DE MUNGUJA



Firma

Se efectuó la defensa final reglamentaria del Trabajo de Graduación:

CÁLCULO DE INDICADORES SAIFI, SAIDI, TDF Y TMR UTILIZANDO HERRAMIENTAS DE MANEJO DE GRANDES VOLÚMENES DE DATOS

A cargo de los Bachilleres:

- CASTRO CONTRERAS FIDEL ERNESTO

- ROMERO PALACIOS HAROLD ERNESTO

Habiendo obtenido en el presente Trabajo una nota promedio de la defensa final: 8.3
(OCHO PUNTO TRES)

AGRADECIMIENTOS

A mi madre **Marta Contreras** que le debo absolutamente todo, por acompañarme en cada uno de los pasos que he dado a lo largo de mi vida.

A mis hermanos **Ana** y **Ricardo** por haberme apoyado en cada una de mis etapas.

A mis abuelos **Ana Castro**, **Luis Contreras** y **Marcos Sanabria** por haber creído en mi y haberme cuidado de niño.

A nuestro asesor de tesis y amigo Dr. Carlos Martínez por todos los conocimientos compartidos a lo largo de la carrera y en este trabajo de investigación.

A todos mis amigos, a los nuevos y a los de toda la vida, sobre todo a Diego Vásquez por haberme enseñado que a pesar de lo duras que sean las circunstancias que podamos vivir, siempre podemos elegir hacer lo correcto.

A la Escuela de Ingeniería Eléctrica con una mención especial a **Reina Vides** por ser parte fundamental de este proceso, por los consejos recibidos y todas esas palabras de ánimo tan necesarias.

A la **Asociación de Estudiantes de Ingeniería Eléctrica y la Rama Estudiantil IEEE de la Universidad de El Salvador** por haberme brindado un espacio para desarrollar y conocer muchas de mis habilidades, adquirir conocimiento y compartirlo con los demás.

¡Hacia la libertad por la cultura!

Fidel Castro

AGRADECIMIENTOS

Agradezco a Dios todo poderoso: por brindarme la sabiduría, el discernimiento y la fortaleza para sobrellevar todos los retos a lo largo del camino de convertirme en ingeniero.

A mi familia: grandemente por haberme brindado el sustento diario y el apoyo económico para cumplir este sueño.

A mi madre Dinora de Romero por creer en mí y siempre apoyarme en todo lo que me propongo.

A mi padre Miguel Romero por enseñarme el gran valor del trabajo, por mostrarme que no hay nada mejor que luchar y esforzarse por lo que se desea.

A mi hermano Michael Romero por no dejar de verme como alguien luchador y por inspirarme para convertirme en ejemplo.

A mis abuelos por su amor incondicional, en especial a Ángel Romero por ser mi ejemplo a seguir.

A Nohely Mejía: por ser mi apoyo incondicional, por darme palabras de amor en los momentos difíciles, por tenerme paciencia y por siempre impulsarme a sacar lo mejor de mí.

A la UES: por mostrarme que las cosas que valen la pena no son fáciles, por moldear mi carácter y convertirme no solo en un profesional sabio sino consciente y con voluntad de servicio.

A mi asesor de tesis: Dr. Carlos Martínez por creer en nosotros, por brindar su sabiduría y siempre estar dispuesto a ayudarnos, por darnos esperanza, por enseñarnos que las cosas pueden cambiar si se lucha por ello.

A todos los catedráticos de la escuela de ingeniería eléctrica, de todos me he llevado sabiduría para la vida profesional. Podría mencionar frases de cada uno que me marcaron no solo como estudiante sino como persona y lo agradezco.

Al personal administrativo como el Sr. Posada, al siempre alegre y dispuesto a ayudar el Sr. Juancito, en especial a la Sra. Reina Vides por ser alguien muy entregada a su trabajo y por su gran labor con los estudiantes. La cual nunca nos dejó perecer, siempre estuvo dispuesta a ayudar y a darnos buenos consejos para salir adelante.

A mi compañero de tesis Fidel Castro por ser un amigo, un hermano y un gran colega. Por no dejarme caer y ayudarme desde la primera vez que formamos parte de un grupo de estudio.

A mis compañeros de estudios y a todos los que forman parte de ASEIE, mil gracias por ser buenos amigos.

A mis amigos: los de la infancia, los de la colonia, los de la iglesia y a la familia Mejía por darme su apoyo incondicional y por sus buenos consejos.

A mis compañeros de trabajo por apoyarme este último año en salir a delante en especial a mi buen amigo Cristian Rodríguez, a mis jefes Arq. Ana Castro y Ing. Gustavo Rivas por creer en mí, por sus consejos, por darme un buen ejemplo de lo que es ser profesional, por su amistad, por sus buenos deseos a mi persona y por ayudarme a superarme.

Gracias por todo.

Harold Romero.

ÍNDICE GENERAL

AGRADECIMIENTOS.....	6
ÍNDICE GENERAL.....	8
ÍNDICE DE TABLAS	10
ÍNDICE DE FIGURAS	11
CAPÍTULO I- INTRODUCCIÓN.....	12
1.1 INTERÉS DE LA INVESTIGACIÓN.....	13
1.2 ANTECEDENTES	13
1.3 OBJETIVOS	15
1.3.1 OBJETIVO GENERAL.....	15
1.3.2 OBJETIVOS ESPECÍFICOS.....	15
1.4 MOTIVACIÓN PARA REALIZAR EL PROYECTO.....	16
1.5 ALCANCES.....	16
1.6 ORGANIZACIÓN	17
CAPÍTULO II – ASPECTOS GENERALES	18
2.1 INDICADORES	19
2.1.1 SAIFI Y SAIDI	20
2.1.2 TDF.....	20
2.1.3 TMR	21
2.2 CRITERIOS PARA EL CÁLCULO DE INDICADORES.....	21
2.3 INDICADORES EN EL SALVADOR.....	23
2.4 SAIFI Y SAIDI EN EL SALVADOR Y EN LA REGION LATINOAMERICANA.....	27
CAPÍTULO III - CÁLCULO DE LOS INDICADORES.....	29
3.1 CÁLCULO DE LOS INDICADORES.....	30
3.1.1 SAIFI Y SAIDI	31
3.1.2 Tasa de Falla	35
3.1.3 Tiempo Medio de Reparación.....	39
3.2 RESUMEN DE INDICADORES.....	43
3.3 ANÁLISIS DE INDICADORES.....	45
CAPÍTULO IV – MINERÍA DE DATOS.....	46
4.1 ANÁLISIS DE SERIES DE TIEMPO	46
4.2 ANÁLISIS POR REGRESIONES LINEALES	48
4.3 ANÁLISIS POR RED NEURONAL FEEDFORWARD	49
4.4 PROYECCIÓN DE INDICADORES 2021.....	51

4.5 COMPROBACIÓN DEL MODELO PARA EMPRESA XYZ	55
CAPÍTULO V – CONCLUSIONES Y LÍNEAS FUTURAS.....	60
5.1 CONCLUSIONES.	60
5.2 LÍNEAS FUTURAS.	61
BIBLIOGRAFIA.....	62
ANEXOS.	63
2 RESUMEN DE INDICADORES.....	63
3 MINERÍA DE DATOS.....	67
3.1 CRISP – DM.....	67
3.2 TÉCNICAS DE MINERÍA	68

ÍNDICE DE TABLAS

Tabla 1. Clasificación de las interrupciones según acuerdo 192-E-2014	18
Tabla 2. Tabla resumen por tipo de interrupción según [5].	22
Tabla 3. Tabla SIGET con causales de Fuerza Mayor o Emergencia	23
Tabla 4. distribuidoras eléctricas en El Salvador	24
Tabla 5. Límites SAIFI y SAIDI	24
Tabla 6. SAIFI y SAIDI por zona de densidad de carga empresa CAESS	25
Tabla 7. Desempeño del SAIFI-SAIDI 2020	26
Tabla 8. Top 10 mejor y peor desempeño en Latinoamérica en 2020	27
Tabla 9. Segmento de código para el cálculo de SAIFI.	32
Tabla 10. Segmento de código para el cálculo de TDF.....	37
Tabla 11. Segmento de código para el cálculo de TMR.	42
Tabla 12. Resumen de indicadores 2020 ABC.....	44
Tabla 13. Resumen de indicadores 2020 XYZ.....	44
Tabla 14. Comparativa de datos reales vs valores proyectados para SAIFI empresa ABC	50
Tabla 15. Comparativa de datos reales vs valores proyectados para empresa para SAIDI, SAIDI, TDF Y TMR empresa ABC	50
Tabla 16. Comparativa de datos reales vs valores proyectados de SAIFI para empresa ABC ...	51
Tabla 17. Comparativa de datos reales vs valores proyectados de SAIDI para empresa ABC..	52
Tabla 18. Comparativa de datos reales vs valores proyectados de TDF para empresa ABC.....	53
Tabla 19. Comparativa de datos reales vs valores proyectados de TMR para empresa ABC.....	54
Tabla 20. Comparativa de datos reales vs valores proyectados para empresa para SAIFI, SAIDI, TDF Y TMR empresa XYZ	55
Tabla 21. Comparativa de datos reales vs valores proyectados de SAIFI para empresa XYZ ...	55
Tabla 22. Comparativa de datos reales vs valores proyectados de SAIDI para empresa XYZ...	56
Tabla 23. Comparativa de datos reales vs valores proyectados de TDF para empresa XYZ.....	57
Tabla 24. Comparativa de datos reales vs valores proyectados de TMR para empresa XYZ.....	58
Tabla 25. INDICADORES SAIFI-SAIDI-TDF-TMR "EMPRESA ABC.....	64
Tabla 26. INDICADORES SAIFI-SAIDI-TDF-TMR "EMPRESA XYZ"	66

ÍNDICE DE FIGURAS

figura 1. Mapa de cobertura de distribuidoras en El Salvador	23
figura 2. Gráfico de SAIFI-SAIDI carga alta para CAESS 2020.....	25
figura 3. Gráfico de SAIFI-SAIDI carga alta para distribuidoras 2020.	26
figura 4. Gráfico de top 10 Latinoamérica de mejor y peor desempeño SAIFI-SAIDI.	28
figura 5. Proceso de filtrado de base de datos.	30
figura 6. Proceso de cálculo de SAIFI y SAIDI.	31
figura 7. SAIFI 2017-2021 Empresa ABC.....	32
figura 8. SAIFI 2017-2021 Empresa XYZ.....	33
figura 9. SAIFI ABC vs XYZ 2017-2021.	33
figura 10. SAIDI 2017-2021 Empresa ABC	34
figura 11. SAIDI 2017-2021 Empresa XYZ	34
figura 12. SAIFI ABC vs XYZ 2017-2021	35
figura 13. Proceso de cálculo de TDF.....	36
figura 14. TDF 2017-2021 Empresa ABC	38
figura 15. TDF 2017-2021 Empresa XYZ	38
figura 16. TDF ABC vs XYZ 2017-2021.	39
figura 17. Proceso de cálculo de TMR.....	39
figura 18. TMR 2017-2021 Empresa ABC	42
figura 19. TMR 2017-2021 Empresa ABC	43
figura 20. TMR ABC vs XYZ 2017-2021.	43
figura 21. Tendencia Serie de Tiempo SAIFI ABC	47
figura 22. Tendencia Serie de Tiempo SAIFI ABC	47
figura 23. Análisis por regresiones lineales SAIFI ABC.	48
figura 24. Valores Reales (rojo) vs Valores Predichos (verde) de SAIFI para empresa ABC....	51
figura 25. Valores Reales (rojo) vs Valores Predichos (verde) de SAIDI para empresa ABC. ..	52
figura 26. Valores Reales (rojo) vs Valores Predichos (verde) de TDF para empresa sABC.....	53
figura 27. Valores Reales (rojo) vs Valores Predichos (verde) de TMR para empresa ABC.	54
figura 28. Valores Reales (rojo) vs Valores Predichos (verde) de SAIFI para empresa XYZ....	56
figura 29. Valores Reales (rojo) vs Valores Predichos (verde) de SAIDI para empresa XYZ. ..	57
figura 30. Valores Reales (rojo) vs Valores Predichos (verde) de TDF para empresa XYZ	58
figura 31. Valores Reales (rojo) vs Valores Predichos (verde) de TMR para empresa XYZ.	59
figura 32. Fases del modelo de referencia CRISP–DM [17].....	67
figura 33. Descomposición Serie de Tiempo SAIFI XYZ	70
figura 34. Regresión polinómica desde grado 1 a grado 10 SAIFI empresa XYZ.....	71

CAPÍTULO I- INTRODUCCIÓN

Parte del día a día de una empresa distribuidora es el constante monitoreo de los indicadores claves de desempeño del servicio técnico ofrecido (KPI por sus siglas en inglés), en El Salvador el ente encargado de regularlos es la Superintendencia General de Electricidad y Telecomunicaciones (SIGET).

En la búsqueda de una mejora constante al servicio ofrecido a los clientes y evitar el pago de compensaciones, las distribuidoras llevan un seguimiento estricto de los indicadores SAIFI (Índice de frecuencia de interrupción promedio del sistema), SAIDI (índice de duración promedio de interrupción del sistema), TDF (Tasa de Falla) y TMR (Tiempo Medio de Reparación) por esto se hace necesaria una herramienta capaz de realizar el cálculo de estos indicadores de manera rápida y confiable.

Hoy por hoy el proceso de comparación de los valores de que tan bien o tan mal es desempeño de los indicadores suele ser muy engorroso, un indicador puede desviarse fácilmente del comportamiento esperado por múltiples razones, estas pueden ser aumento de lluvias y vientos, baja de personal ocasionada por una emergencia sanitaria o problemas ocasionados por la red de transmisión nacional o la interconexión regional.

En el desarrollo de este trabajo se describe una propuesta para el cálculo de los indicadores SAIFI, SAIDI, TDF y TMR considerando fallas internas de empresas distribuidoras que operen en El Salvador, utilizando herramientas de manejo de altos volúmenes de datos de Python siguiendo las normativas descritas en la norma de calidad de servicio de los sistemas de distribución [1] y el Manual de Confiabilidad (MC) [4] de SIGET y se realizara una aproximación al comportamiento futuro de los indicadores utilizando técnicas de minería de datos en el entorno de Google Colab a fin de utilizar estos resultados como valores de referencia para evaluar el comportamiento real para los años futuros.

1.1 INTERÉS DE LA INVESTIGACIÓN.

La investigación surge del interés por desarrollar una herramienta capaz de calcular rápidamente el valor de los indicadores de calidad del servicio técnico utilizado por empresas distribuidoras en El Salvador y que a su vez permita la proyección de indicadores futuros con el objetivo de establecer metas con las que las empresas puedan medir el desempeño del servicio técnico ofrecido.

Actualmente el criterio para establecer metas de indicadores suele ser muy arbitrario, las empresas en búsqueda de una mejora constante del servicio se fuerzan a cerrar al año en curso mejor que el anterior, es por esto por lo que se hace necesario el desarrollo de una herramienta capaz de estimar la tendencia y que establezca las métricas con lo que se pueda comparar el desempeño del año.

Por el momento el campo de *machine learning* sigue siendo inexplorado a nivel de cálculo de indicadores de calidad del servicio técnico en El Salvador, es por esto que es del interés de esta investigación desarrollar un modelo utilizando técnicas de *machine learning* que permita la estimación de comportamientos futuros de estos.

1.2 ANTECEDENTES.

En primer lugar, se tiene que, en 2017 fue presentado en la facultad de ingeniería de la Pontificia Universidad Javeriana de Colombia, el trabajo de post grado ***“MODELO DE MINERÍA DE DATOS PREDICTIVA PARA EL PRONÓSTICO DE INDICADORES DE CALIDAD DE SUMINISTRO SAIDI Y SAIFI”*** por Corredor Reyes, Claudia M, como requisito para optar al título de Magister en Ingeniería de sistemas y computación.

El trabajo describe el proceso empleado para realizar un modelo predictivo de los indicadores SAIFI y SAIDI para la empresa de distribución eléctrica colombiana Codensa S.A. ESP, creando un conjunto de modelos predictivos que permitan predecir los valores futuros de estos indicadores, realizando análisis de series temporales, estudiando su estacionalidad y tendencia históricas de estos.

La investigación adopta el modelo de desarrollo basado en CRISP-DM, que contempla las etapas de entendimiento de negocio, entendimiento de datos, modelamiento y evaluación.

A partir de los modelos construidos, se procedió a su validación desde una perspectiva matemática, empleando el Error porcentual absoluto Medio (MAPE), para medir la calidad y el nivel de confianza del modelo. Sin embargo, la prueba se aplica a cada uno de los modelos por separado y al final se aplica al valor total del pronóstico para cada indicador.

Para la validación del modelo se genera el pronóstico de 7 meses, obteniendo una confianza total del 81,104% para el SAIDI y del 88,714% para el SAIFI,

En la Escuela de Ingeniería Eléctrica de la Universidad de El Salvador en el año 2008, se realizó el trabajo de grado ***“Evaluación de la Normas de Calidad del Servicio en los Sistemas de Distribución de SIGET”*** donde como parte de la investigación se hace una comparativa de la reglamentación de los indicadores SAIFI y SAIDI entre El Salvador, Guatemala y Panamá , dando como resultado que a partir de la reglamentación empleada en 2008 El Salvador cuenta con indicadores menos estrictos que en el resto de países en estudio.

1.3 OBJETIVOS.

1.3.1 OBJETIVO GENERAL.

- Desarrollar una herramienta capaz de calcular y predecir mediante técnicas de análisis de datos el valor de los indicadores SAIFI, SAIDI, TDF y TMR.

1.3.2 OBJETIVOS ESPECÍFICOS.

- Utilizar el módulo Pandas del lenguaje de programación Python para cargar, procesar y analizar una base de datos de interrupciones asociadas a dispositivos de protección de la red de distribución eléctrica de empresas distribuidoras en El Salvador.
- Visualizar el comportamiento de los indicadores SAIFI, SAIDI, TDF y TMR de empresas distribuidoras de manera mensual.
- Predecir mediante técnicas de procesamiento de datos los valores futuros de los indicadores SAIFI, SAIDI, TDF y TMR.

1.4 MOTIVACIÓN PARA REALIZAR EL PROYECTO.

La presente investigación tiene por objeto desarrollar una herramienta que calcule de forma rápida y confiable el cálculo de los indicadores SAIFI, SAIDI, TDF y TMR siguiendo las normativas que los regulan.

La elección del tema responde a la necesidad de desarrollar una aplicación que facilite el proceso de cálculo de estos indicadores, que hoy por hoy se realiza con diversos programas de ofimática y la búsqueda de mejorar la distribución del recurso humano que es limitado.

1.5 ALCANCES.

Calcular, analizar y predecir los indicadores SAIFI, SAIDI, TDF y TMR para empresas de distribución de energía eléctrica que opere en El Salvador utilizando el módulo Pandas de Python y la herramienta Colab de Google.

Visualizar el comportamiento mensual de los indicadores SAIFI, SAIDI, TDF y TMR para una empresa distribuidora que opere en El Salvador a través del módulo Matplotlib de Python.

Proveer una herramienta capaz de realizar estimaciones del comportamiento mensual futuro de los indicadores SAIFI, SAIDI, TDF y TMR de empresas distribuidoras que operen en El Salvador utilizando las herramientas del módulo sklearn de Python.

1.6 ORGANIZACIÓN.

A continuación, presentamos la distribución del contenido de esta investigación la cual se basa en 5 capítulos los cuales nos mostraran el camino que se ha seguido para poder realizarlo, en los capítulos I y II encontraremos la introducción y las bases teóricas esto preparándonos para la implementación de los mismos en los siguientes capítulos. En el capítulo III y IV visualizamos las herramientas que serán utilizadas para el procesamiento de los datos y su visualización en el capítulo V.

- Capítulo I: Se presenta una reseña de lo que conlleva esta investigación, los motivos y la descripción general del tema a su vez los objetivos que se pretenden alcanzar.
- Capítulo II: Describe la base teórica, los conceptos generales e importancia del cálculo de indicadores de rendimiento.
- Capítulo III: Describe el proceso de exploración, filtrado de la base datos y cálculos de los indicadores SAIFI, SAIDI, TDF y TMR.
- Capítulo IV: Se presentan los resultados obtenidos de la estimación de valores de los indicadores utilizando series de tiempo.
- Capítulo V: Resultados y líneas futuras.

CAPÍTULO II – ASPECTOS GENERALES

Parte del día del día de las empresas de distribución eléctrica consiste en la estimación de Indicadores Claves de Desempeño (KPI por sus siglas en inglés) que permiten medir la calidad servicio técnico proporcionado a sus clientes, en este capítulo se presentan las ecuaciones que describen cuatro de estos indicadores emitidas por la Superintendencia General de Electricidad y Telecomunicaciones.

La norma de calidad de servicio de los sistemas de distribución de SIGET [1] define una Interrupción, como la falta de suministro de energía eléctrica en el punto de entrega al usuario la cual puede clasificarse como:

- **Momentánea:** Interrupciones que tienen una duración limitada hasta de tres minutos, el cuál es el tiempo requerido para restablecer los dispositivos de control y protección.
- **Sostenida:** Cualquier interrupción no clasificada como momentánea. (Definición utilizada a lo largo de este estudio, para el análisis de calidad de energía: “Mayores a 3min”)

Clasificación	Duración
Momentánea	Menor a los 3 minutos.
Sostenida	Mayos a los 3 minutos.

Tabla 1. Clasificación de las interrupciones según acuerdo 192-E-2014 [1].

El motivo de las interrupciones se debe a múltiples orígenes entre ellos a operaciones de mantenimiento de la red o por fallas en los sistemas de potencia y equipos de control. Las interrupciones de tensión, incluso las instantáneas, pueden provocar el mal funcionamiento o la desconexión de equipos electrónicos o de los equipos de iluminación o cualquier elemento conectado a la red. Las interrupciones producen la mayoría de las veces la parada de los controladores electrónicos, el mal funcionamiento de las fuentes de alimentación, computadores, equipos de control de máquinas eléctricas y pueden producir también la desconexión de contactores en motores de inducción.

La forma de registrar estas interrupciones es tradicionalmente a través de hojas de cálculo obtenidas a través de sistemas de registro de eventos (de forma automática en caso se trate de fallas equipos tele contralados con acceso a un sistema SCADA o manualmente si los dispositivos accionados son tradicionales). Toda esta información debe ser procesada de una rápida y efectiva, de tal forma que pueda ser utilizada para toma de decisiones y direccionar el recurso de mantenimiento a las zonas de mayor necesidad.

A continuación, se describen los indicadores de calidad del servicio técnico objetos del estudio de esta investigación.

2.1 INDICADORES.

Como se mencionó con anterioridad, las empresas distribuidoras generan grandes volúmenes de datos día tras día, los cuales no son de utilidad si no se pueden interpretar u ocupar de alguna manera.

Como toda industria la planificación y medición son claves para el éxito, por ende, se necesita evaluar estrategias para corregir errores y detectar oportunidades que mejoren la calidad del servicio ofrecido [2].

Actualmente para poder interpretar la información generada, se hace uso de indicadores. un indicador es una unidad de medida que permite el seguimiento y evaluación periódica de las variables clave de una organización, mediante su comparación en el tiempo con los correspondientes referentes externos o internos [3]. KPI's hacen referencia a una serie de métricas que se utilizan para sintetizar la información sobre la eficacia y productividad de las acciones que se lleven a cabo en un negocio, con el fin de poder tomar decisiones y determinar aquellas que han sido más efectivas a la hora de cumplir con los objetivos marcados en un proceso o proyecto concreto [2].

En la industria de distribución de energía eléctrica se busca medir la continuidad del servicio ofrecido por la empresa a sus clientes. En El Salvador La Superintendencia General de Electricidad y Telecomunicaciones exige a las empresas distribuidoras llevar un estricto seguimiento de los indicadores de calidad de energía. Descrito en la normativa de calidad de servicio de los sistemas de distribución.

Lo que se busca con los indicadores es medir la continuidad de la red, cuando la red se vuelve discontinua ocurre una interrupción. Esta se define como la pérdida absoluta de la tensión durante un periodo de tiempo.

Por estos motivos se adoptan indicadores que se basan en las interrupciones de suministro ocurridas en el sistema de distribución. Los indicadores permiten desde un punto de vista empresarial evaluar el desempeño en cuanto a calidad y productividad, logrando definir criterios validos de comparación de empresas basados en las características del mercado, tipo de área de concesión, características de la red.

Del acuerdo 192-E-2014 extraeremos los primeros indicadores o métricas que formaran parte de nuestro trabajo de investigación.

Los parámetros sugeridos por SIGET son:

- SAIFI: Índice de Frecuencia de Interrupción Promedio del Sistema
- SAIDI: Índice de Duración de Interrupción Promedio del Sistema

SIGET exige a las empresas distribuidoras llevar un estricto seguimiento de los indicadores de rendimiento SAIFI y SAIDI utilizados para medir la calidad del servicio técnico brindado hacia los usuarios, en los cuales se menciona que si sobrepasar los límites establecidos por SIGET conlleva al pago de compensaciones económicas hacia los usuarios y hacia el ente regulador

A partir del 1 de enero de 2008 en adelante se aplicaron todas las compensaciones individuales y globales por las transgresiones a los límites establecidos en la norma de Calidad del Servicio de los Sistemas de Distribución; además de entrar en vigencia las compensaciones por perturbaciones.

Las empresas distribuidoras continuarán desarrollando las inversiones que consideren necesarias para el cumplimiento de las exigencias de Calidad del Servicio Técnico, Calidad del Producto y Calidad del Servicio Comercial. [2] Para el análisis de confiabilidad se emplearán los valores de tasa y duración de falla informados en el “Manual de Confiabilidad” (MC) [5].

A continuación, se describen los indicadores.

2.1.1 SAIFI Y SAIDI.

SAIFI: Este indicador refleja la frecuencia de las interrupciones experimentadas por el cliente promedio en un periodo determinado.

Matemáticamente el SAIFI está dado por la ecuación:

$$SAIFI = \frac{\sum \text{CLIENTES INTERRUMPIDOS}}{\text{TOTAL DE CLIENTES SERVIDOS}} = \frac{\sum N_i}{N_t} \quad \text{EC 1}$$

Donde:

N_i = cantidad de clientes afectados por la interrupción.

N_t = cantidad de clientes totales servidos por la empresa

SAIDI: Este indicador de confiabilidad refleja la duración total de las interrupciones para el promedio de clientes en un periodo determinado y es medido en horas de interrupción.

Matemáticamente el SAIDI está dado por la ecuación:

$$SAIDI = \frac{\sum \text{DURACION DE INTERRUPCIONES DE LOS CLIENTES}}{\text{TOTAL DE CLIENTES SERVIDOS}} = \frac{\sum r_i N_i}{N_t} \quad \text{EC 2}$$

Donde:

r_i = duración de la interrupción.

N_i = cantidad de clientes afectados por la interrupción.

N_t = cantidad de clientes totales servidos por la empresa.

2.1.2 TDF.

Según el manual de confiabilidad de SIGET, la tasa de falla es el número medio de fallas de un componente por unidad de tiempo de exposición [4]. Usualmente el tiempo de

exposición esta expresado en años. Las empresas objeto de estudio en esta investigación llevan el control de este indicador de manera mensual y las unidades de medida son fallas/100 km de red, de acuerdo con la siguiente fórmula:

$$TDF = \frac{\text{Numero de fallas}}{\text{Longitud de la red}/100} \quad \text{EC 3}$$

2.1.3 TMR.

El Tiempo Medio de reparación es la duración media desde la ocurrencia de la falla hasta el instante en que es restituido el servicio, ya sea por reparación del componente fallado o por reemplazo de este. Este tiempo incluye el diagnóstico y localización de componentes fallados, reparación o remplazo, ensayos y restablecimiento del servicio [4].

Para tener un valor objetivo de este indicador, se descartan el 5% por ciento de interrupciones de mayor duración y el 5% de menor duración, esta práctica ha sido tomado de acuerdo con procedimientos internos de empresas distribuidoras [6].

La fórmula para el cálculo de este indicador es la siguiente:

$$TMR = \frac{\sum \text{Tiempo de reposiciones}}{\text{Total de reposiciones}} \quad \text{EC 4}$$

En la siguiente sección se describen los criterios que terminan si una interrupción impacta o no en los indicadores.

2.2 CRITERIOS PARA EL CÁLCULO DE INDICADORES.

Como parte de la operación, las empresas distribuidoras documentan un alto volumen de eventos asociados a interrupciones; sin embargo, no todos estos eventos deberían de considerarse para el cálculo de los indicadores, ya que por su naturaleza podrían ser eventos de fuerza mayor, interrupciones ocasionadas por otra distribuidora que afectaron el desempeño de la red, programadas por la misma distribuidora como parte de las estrategias de mantenimiento y mejoras a la red.

A continuación, se describen cada uno de los criterios que deben ser tomados en cuenta para determinar si una interrupción impacta en los indicadores de acuerdo con el Anexo B Metodología para el Control de la Calidad del Servicio Técnico de la SIGET [5].

INTERRUPCIÓN	CRITERIO
Interrupciones Internas y Externas	Para el cálculo de los indicadores Individuales y Globales serán consideradas únicamente las interrupciones internas a la red de distribución.
Interrupciones Menores o Iguales a los 3 minutos	Las interrupciones menores o iguales a tres minutos las cuales son denominadas interrupciones momentáneas deberán quedar registradas en la base de datos de interrupciones por usuario e informadas a la SIGET, pero no serán consideradas para la determinación de los indicadores
Interrupciones Programadas	Las interrupciones programadas por la Distribuidora serán consideradas para la determinación de los indicadores y las compensaciones a los usuarios; sin embargo, para el objetivo de esta investigación será excluidos puesto que no son de producto de fallas.
Interrupciones por Causas de Fuerza Mayor y Caso Fortuito	Se considerarán todas las interrupciones, salvo las que sean aceptadas o estén en proceso de aceptación por la SIGET como originadas en causales de Fuerza Mayor y Caso Fortuito (ver tabla 3)
Interrupciones Originadas por Defectos en las Propias Instalaciones del Usuario	Si como consecuencia de los defectos en las instalaciones de un usuario se produjeran interrupciones que afecten a otros usuarios, las mismas deberán ser computadas a efectos de la determinación de los indicadores de Calidad de Servicio Técnico de estos últimos.
Corte de Suministro por Morosidad u Otras Causas	Las interrupciones relacionadas con domicilios de usuarios en situación de corte de suministro ordenado por el propio distribuidor basado en la Ley General de Electricidad no serán computadas para el cálculo de los indicadores y de las compensaciones de Calidad del Servicio Técnico.
Interrupciones por Maniobras	Las interrupciones que tienen origen en maniobras realizadas por la distribuidora para reponer parcial o totalmente a clientes afectados por una interrupción forzada, se denominan interrupciones por maniobra.
Interrupciones de Emergencia.	En la operación del sistema de distribución de energía eléctrica también se presentan interrupciones intencionales (apertura de elementos), que se realizan con el fin de eliminar riesgos al personal, terceros o riesgos a instalaciones eléctricas propias o particulares

Tabla 2. *Tabla resumen por tipo de interrupción según [5].*

CAUSAL	DESCRIPCION
011	Terceros-Vandalismo
012	Terceros-Cometas u otros objetos suspendidos de las líneas aéreas
013	Terceros-Alambrados/Conductores
014	Terceros-Cables Telefónicos/Televisión por Cable/Otros Servicios
015	Terceros-Trabajos en la vía Pública
016	Terceros-Embestidas/Colisiones
017	Terceros-Poda de árboles
021	Condiciones Climáticas Extremas-Vientos
022	Condiciones Climáticas Extremas-Inundaciones
024	Condiciones Climáticas Extremas-Descargas Atmosféricas
030	Incendio
040	Solicitado por el cliente
050	Autorizado por SIGET o autoridad competente
061	Afectación instalaciones subterráneas-Obstrucción de desagües
062	Afectación instalaciones subterráneas-Filtración por rotura de desagüe de agua.

Tabla 3. *Tabla SIGET con causales de Fuerza Mayor o Emergencia [5].*

Hasta el momento únicamente se han mencionado el marco legal que rige el cálculo de los indicadores, pero no se han mencionado quienes serían los beneficiarios directos del aplicativo capaz de calcular y estimar estos índices. En la siguiente sección se describen de manera general los indicadores de las principales empresas distribuidoras en El Salvador.

2.3 INDICADORES EN EL SALVADOR.

Las distribuidoras son las empresas encargadas de poseer y operar las instalaciones de distribución de la energía eléctrica, cuyo objetivo es entregar energía eléctrica en la red de baja tensión. Estas empresas operan en condiciones reguladas en tarifa y calidad entregada de suministro. Estas empresas en el salvador están reguladas por SIGET, la competencia en la distribución está permitida, tanto como comercializar y distribuir energía eléctrica, a través de sus redes hacia los centros de consumo (usuarios), el mantenimiento y la expansión de la red es parte de las funciones de las empresas distribuidoras.

- CAESS
- CLESA
- EEO
- DEUSEM
- DELSUR
- EDESAL
- B&D
- ABRUZZO



figura 1. Mapa de cobertura de distribuidoras en El Salvador [7].

La tabla resume la cantidad de clientes según [7] y [8] al cierre de 2020 para las principales empresas distribuidoras en El Salvador.

NOMBRE DE LA EMPRESA	NÚMERO DE CLIENTES (2020)	ÁREA QUE CUBRE	ZONA
CAESS	616,398	4,628 km ²	Chalatenango, Cuscatlán, Cabañas, Zona norte de San Salvador
CLESA	426,535	5,175 km ²	Santa Ana, Sonsonate, Ahuachapán y parte del Departamento de La Libertad
EEO	324,913	6,505 km ²	San Miguel, Morazán, La Unión, San Vicente y parte de Usulután
DEUSEM	85,147	1,584 km ²	Usulután
DELSUR	405,025	5,703 km ²	San Salvador, La Libertad, La Paz, Cuscatlán y San Vicente

Tabla 4. distribuidoras eléctricas en El Salvador

En el boletín de estadísticas eléctricas número 22 se presentan los resultados de los indicadores SAIFI y SAIDI para las empresas distribuidoras de El Salvador para el año 2020 [10].

En este estudio se hace la distinción de indicadores por zona de densidad de carga, en la Tabla 6 puede observarse el comportamiento de SAIFI y SAIDI acumulado reportado a SIGET de la empresa CAESS mientras que en la Tabla 7 se observa los indicadores reportados por todas las empresas de la Tabla 4, para todas las empresas SIGET se establece como límite los presentados en la Tabla 5.

INDICADOR	LIMITE
SAIDI (CARGA ALTA)	14
SAIDI (CARGA BAJA)	24
SAIFI (CARGA ALTA)	14
SAIFI (CARGA BAJA)	18

Tabla 5. Límites SAIFI y SAIDI [1]

EMPRESA: CAESS				
MES	SAIDI (CARGA ALTA)	SAIDI (CARGA BAJA)	SAIFI (CARGA ALTA)	SAIFI (CARGA BAJA)
ENERO	0.54	0.76	0.32	0.26
FEBRERO	1.05	1.63	0.55	0.75
MARZO	1.67	2.58	0.96	1.15
ABRIL	2.44	3.86	1.49	1.71
MAYO	3.00	5.7	1.8	2.31
JUNIO	3.96	8.93	2.16	3.36
JULIO	4.83	11.46	2.55	4.05
AGOSTO	5.66	14.29	2.95	4.86
SEPTIEMBRE	6.30	18.43	3.29	5.64
OCTUBRE	7.16	20.46	3.7	6.38
NOVIEMBRE	7.70	21.6	3.94	6.62
DICIEMBRE	8.01	22.29	4.08	6.78

Tabla 6. SAIFI y SAIDI por zona de densidad de carga empresa CAESS [10].

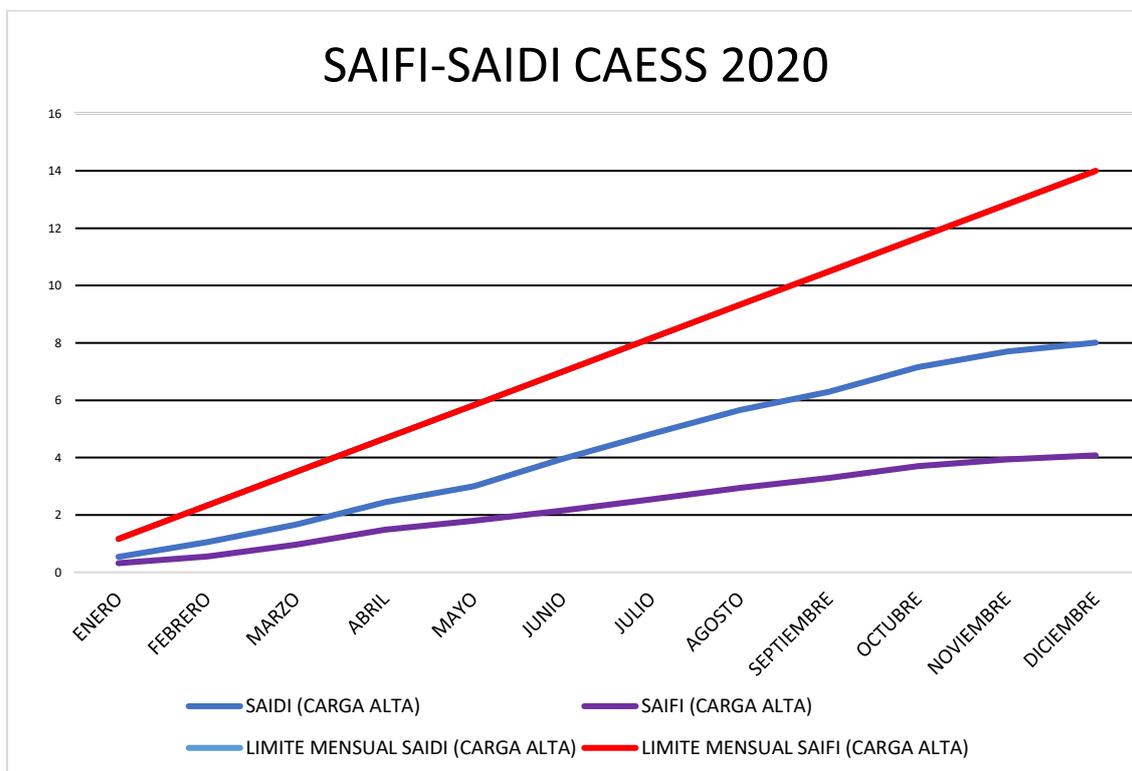


figura 2. Gráfico de SAIFI-SAIDI carga alta para CAESS 2020.

EMPRESA	SAIDI (CARGA ALTA)	SAIDI (CARGA BAJA)	SAIFI (CARGA ALTA)	SAIFI (CARGA BAJA)
CAESS	8.01	22.29	4.08	6.78
CLESA	8.32	26.11	3.25	5.55
DEUSEM	8.95	17.66	2.19	3.66
DELSUR	8.44	22.35	4.87	9.48
EEO	11.37	37.11	3.57	7.77

Tabla 7. Desempeño del SAIFI-SAIDI 2020 [10].

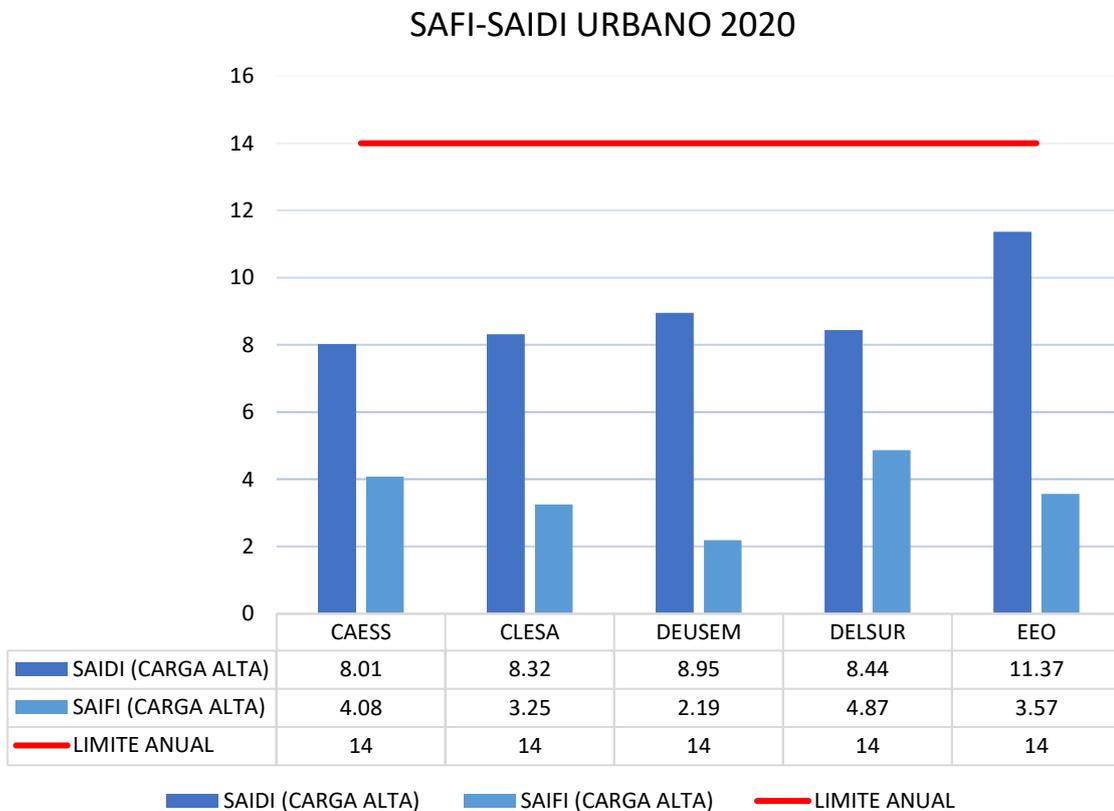


figura 3. Gráfico de SAIFI-SAIDI carga alta para distribuidoras 2020.

Tanto en la Tabla 6 y 7 hacen referencia a los indicadores reportados a SIGET, esto hace discrepancia con los calculados en el capítulo 3, ya que el objetivo de esta investigación es la estimación de indicadores KPI por tanto únicamente se toman en cuenta las interrupciones que son responsabilidad de las empresas.

2.4 SAIFI Y SAIDI EN EL SALVADOR Y EN LA REGION LATINOAMERICANA.

De la tabla 7 podemos resumir los siguientes resultados para El Salvador para el año 2020:

- SAIDI urbano: 8.48 horas
- SAIDI rural: 22.29 horas
- SAIFI urbano: 3.97 interrupciones
- SAIFI rural: 7.08 interrupciones

En la tabla 8 se resumen los valores de SAIFI y SAIDI del top 10 de países con mejor y peor desempeño en 2020

PAÍS	SAIDI	SAIFI
Costa Rica	0.5	0.2
Santa Lucia	0.2	0.3
Dominica	0.6	0.3
México (Monterrey)	0.4	0.4
México (Cd Mx)	0.6	0.9
Chile	2.9	1.5
Perú	6.4	1.7
Jamaica	27.1	10.7
Belice	14.2	13.8
Argentina	4.5	14.4
Paraguay	21.9	22.8
Honduras	32.5	23.4
Nicaragua	93.6	45.7

Tabla 8. Top 10 mejor y peor desempeño en Latinoamérica en 2020. [11].

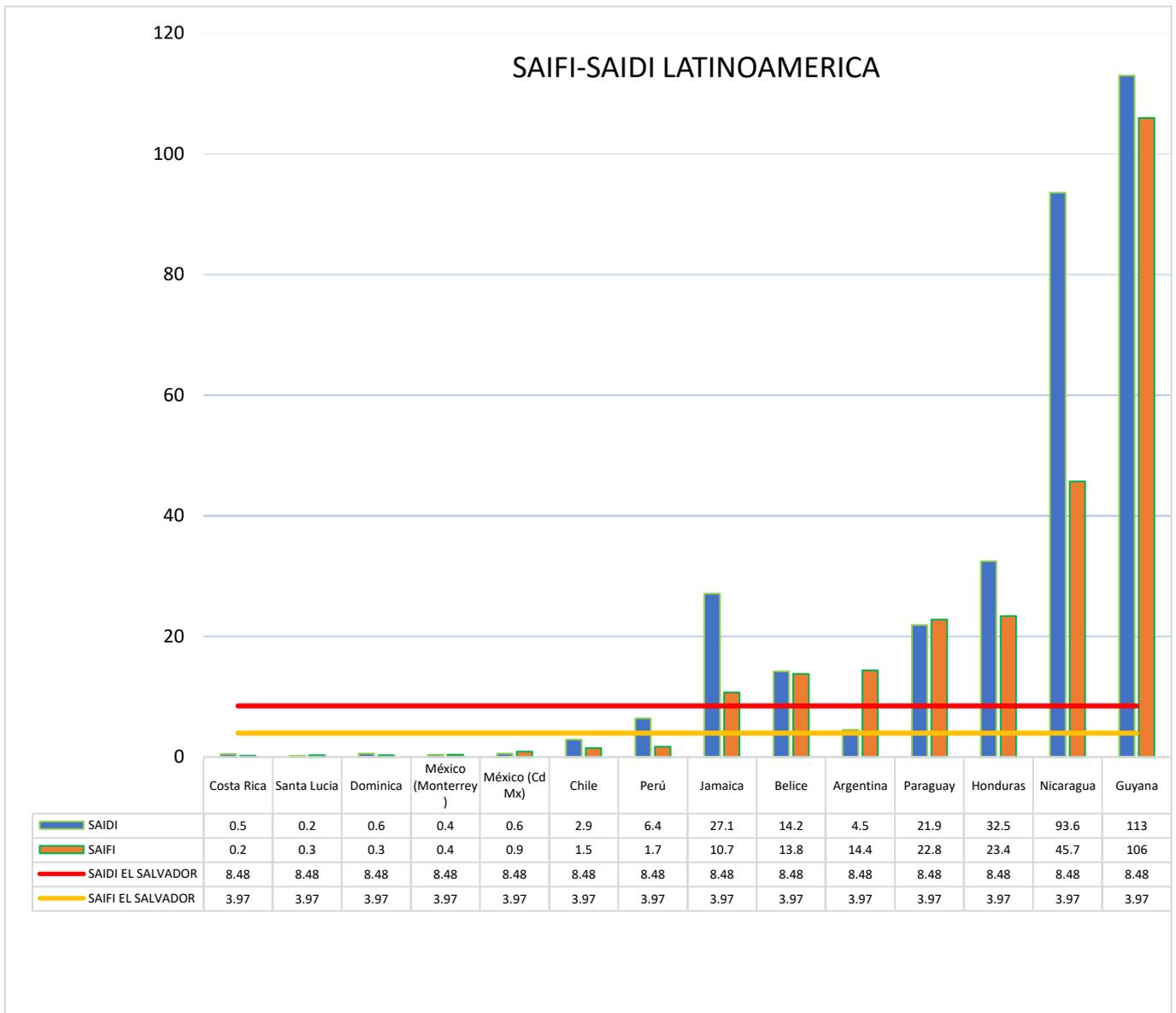


figura 4. Gráfico de top 10 Latinoamérica de mejor y peor desempeño SAIFI-SAIDI.

Tomando como medida comparativa los valores de SAIFI y SAIDI urbanos, El Salvador se encuentra muy por encima de Costa Rica que representa el país latinoamericano con mejor desempeño a nivel de los indicadores con un SAIFI de 0.2 interrupciones y un SAIDI de 0.5 horas.

Hasta el momento se ha hecho una revisión bibliográfica de las normativas que regulan los indicadores y un repaso general de los indicadores reportados al ente regulador para el año 2020 para el caso de El Salvador y los documentados por CEPAL, en el siguiente capítulo se describe el proceso del cálculo de indicadores SAIFI, SAIDI, TDF y TMR a fin de evaluar el desempeño históricos de estos.

CAPÍTULO III - CÁLCULO DE LOS INDICADORES

El proceso de análisis de datos parte con una exploración inicial a la base de interrupciones disponibles, para este trabajo se han utilizado datos históricos desde enero de 2017 hasta agosto de 2021 de dos empresas distribuidoras que operan en El Salvador (por motivos de confidencialidad llamaremos a estas empresas “ABC” y “XYZ”)

Los datos disponibles en las bases de datos son los siguientes.

- **Nombre de la empresa:** Hace referencia a la empresa en la que ha surgido la interrupción puede ser ABC, XYZ.
- **Numero de interrupción:** Detalla el número único con el que se identifica la interrupción.
- **Tipo de interrupción:** Puede ser forzada o programada en caso forme parte de las interrupciones realizadas como parte de las estrategias para mejorar la red eléctrica.
- **Origen de la interrupción:** Puede ser interna o externa dependiendo si la causa es propia o ajena a la distribuidora.
- **Fuerza mayor:** Puede ser “Si” en caso la interrupción fuese causada por razones fortuitas como condiciones climáticas extremas, terremotos, árboles caídos, choque de postes entre otros.
- **Elemento fallado:** Detalla el ID del elemento causante de la interrupción.
- **Tipo de afectación:** Detalla si la interrupción afecto la red de media o baja tensión.
- **Fecha de inicio de la interrupción.**
- **Fecha de fin de la interrupción.**
- **Duración de la interrupción:** Especifica la duración la interrupción en horas.
- **Clientes afectados:** Detalla la cantidad de clientes afectados por la interrupción.
- **Causa:** Especifica el motivo de la interrupción.

Esta base datos se guarda en un archivo .CSV y se carga en el entorno de Google Colab con la librería pandas de Python, una vez cargado los datos se procede a iniciar con la depuración de los datos según las normativas y procedimientos descritos en el capítulo anterior, para esto agregaremos las siguientes columnas al dataframe:

- **"Mayor_a_3_min":** Se coloca “Si” cuando la duración sea mayor a 3 minutos y “No” en caso contrario.
- **“SAIFI”:** Se calcula el valor de SAIFI de la interrupción, realizando la división de los clientes afectados entre el total de clientes de la empresa en cuestión.
- **“SAIDI”:** Se calcula el valor de SAIDI de la interrupción.
- **“Dia”:** Se almacena el número del día en que ocurrió la interrupción.
- **“Mes”:** Se almacena el número del mes en que ocurrió la interrupción.
- **“Año”:** Se almacena el año en el que ocurrió la interrupción.
- **“Suma”:** Se coloca “True” si la causa impacta en los indicadores y “False” si no.

A continuación, se eliminan todos los registros que no impactan en el indicador, ya sea por la duración de la interrupción, causa, tipo de interrupción o sea una fuerza mayor

usando la función drop de pandas, con los registros actuales se accede a el cálculo de los indicadores que servirán como entrenamiento y validación de los modelos de *machine learning*.

En general el proceso del filtrado de los datos se describe en la figura 5.

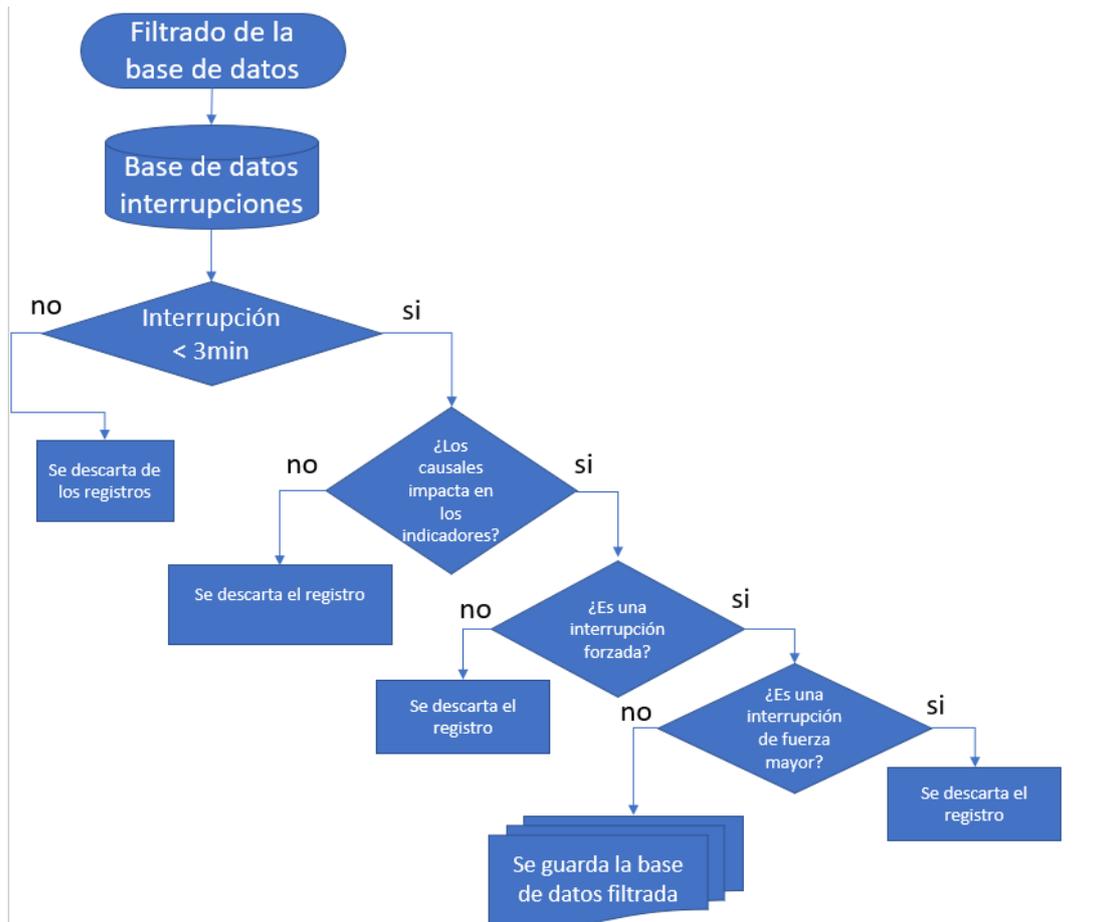


figura 5. Proceso de filtrado de base de datos.

3.1 CÁLCULO DE LOS INDICADORES.

Hasta el momento se han depurado la base de datos de todos los registros que no impactaban en los indicadores, en los apartados siguientes se procederá al cálculo de los indicadores de manera mensual durante el periodo de estudio.

3.1.1 SAIFI Y SAIDI.

Con la base de datos depurada del apartado anterior se inicia el cálculo de los indicadores SAIFI y SAIDI, para esto hacemos uso de las herramientas de Pandas set_index y resample.

Haciendo esto, obtenemos los siguientes resultados para SAIFI y SAIDI de manera mensual durante el periodo de enero de 2017 a agosto de 2021.

La figura 6 describe el proceso de obtención de los valores de SAIFI y SAIDI.

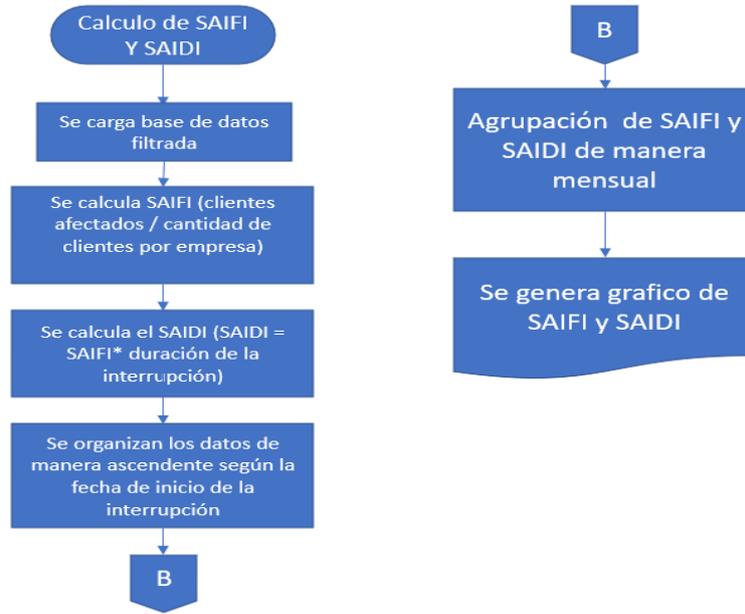


figura 6. Proceso de cálculo de SAIFI y SAIDI.

El segmento de código que hace esta agrupación se presenta en la siguiente tabla:

CÓDIGO	COMENTARIO
<code>KPI_MES=dA[["FECHA_INI_INTERR", "SAIFI", "SAIDI"]]</code>	Del DataFrame depurado únicamente se conservan los registros Fecha de inicio de la interrupción, SAIFI y SAIDI calculados.
<code>KPI_MES = KPI_MES.set_index('FECHA_INI_INTERR')</code>	Se establece la fecha de inicio como índice del DataFrame
<code>KPI_MES.sort_index(inplace=True)</code>	Se organizan los datos a partir del índice con orden ascendente.
<code>KPI_MES = KPI_MES.resample('M').sum()</code>	Se agrupan los datos con una sumatoria de forma mensual.

<pre>KPI_MES=KPI_MES.rename_axis("FECHA").reset_index()</pre>	<p>Se pasa el valor de la fecha a una columna</p>
<pre>plt.figure(figsize=(16,8)) plt.grid() plt.plot(KPI_MES.FECHA,KPI_MES.SAIFI, color="BLUE") plt.title("SAIFI ABC",fontsize=18) plt.xlabel('AÑO', fontsize=18) plt.ylabel("SAIFI CLIENTE/CLIENTE", fontsize=18) plt.legend(["SAIFI"], fontsize=18)</pre>	<p>Código para generar grafica de SAIFI calculado.</p>

Tabla 9. Segmento de código para el cálculo de SAIFI.

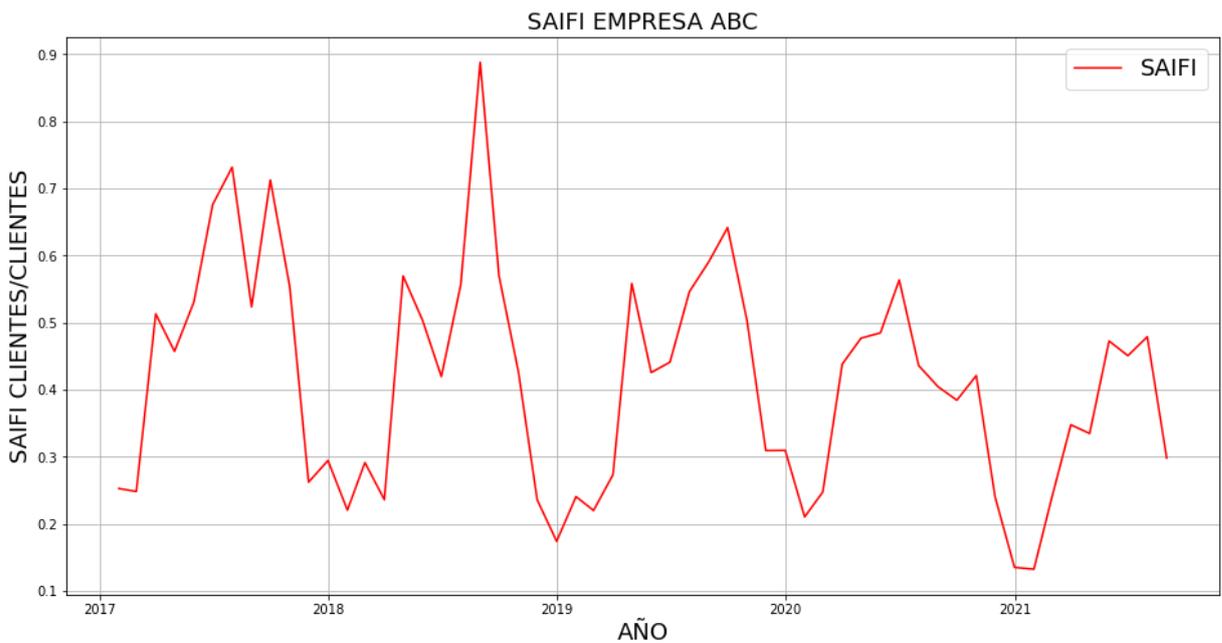


figura 7. SAIFI 2017-2021 Empresa ABC

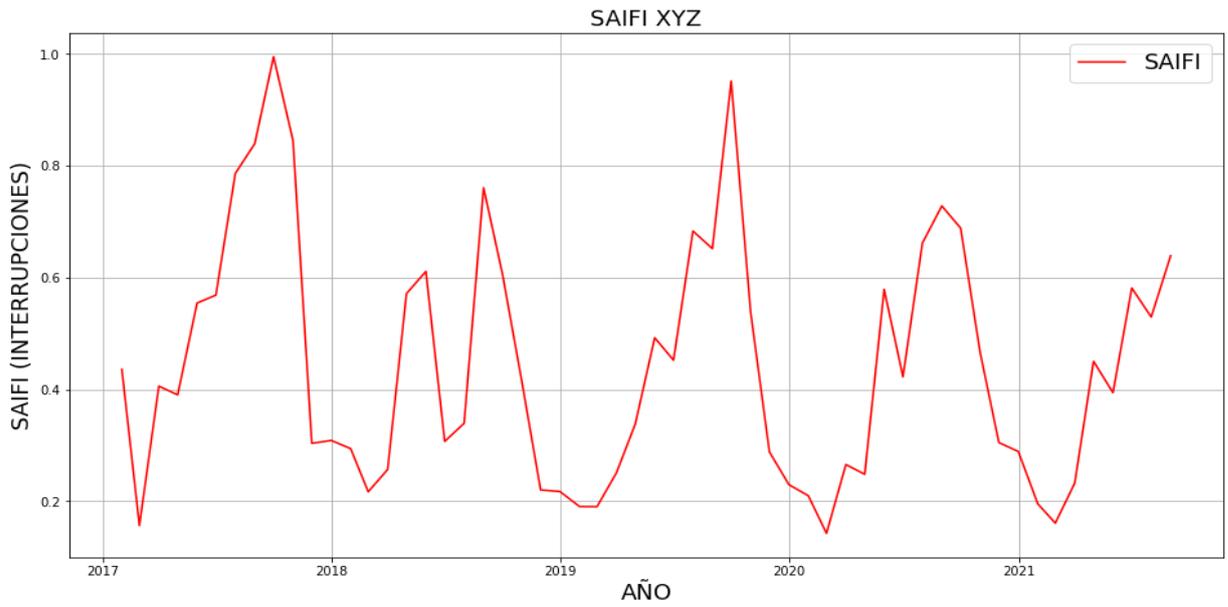


figura 8. SAIFI 2017-2021 Empresa XYZ

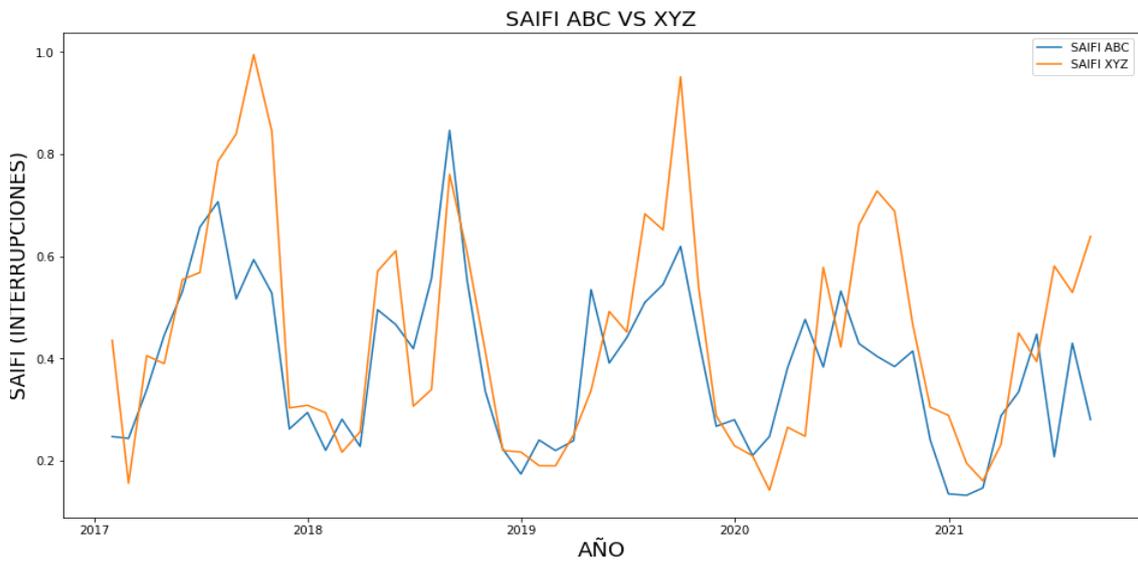


figura 9. SAIFI ABC vs XYZ 2017-2021.

Tanto la figura 7 y 8 demuestran el comportamiento del indicador SAIFI durante el periodo de estudio para las empresas ABC y XYZ.

La figura 9 presenta una comparativa entre los valores anuales

Modificando el último segmento de código de la tabla 9, por los registros que contienen SAIDI se obtiene las siguientes graficas de las figuras 6 y 7.

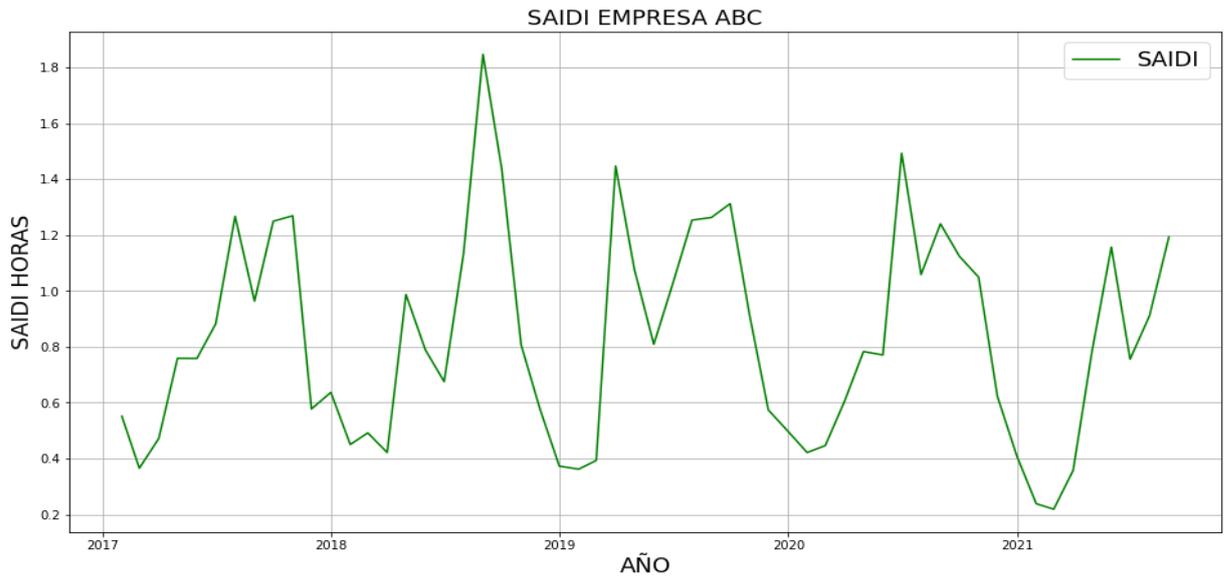


figura 10. SAIDI 2017-2021 Empresa ABC

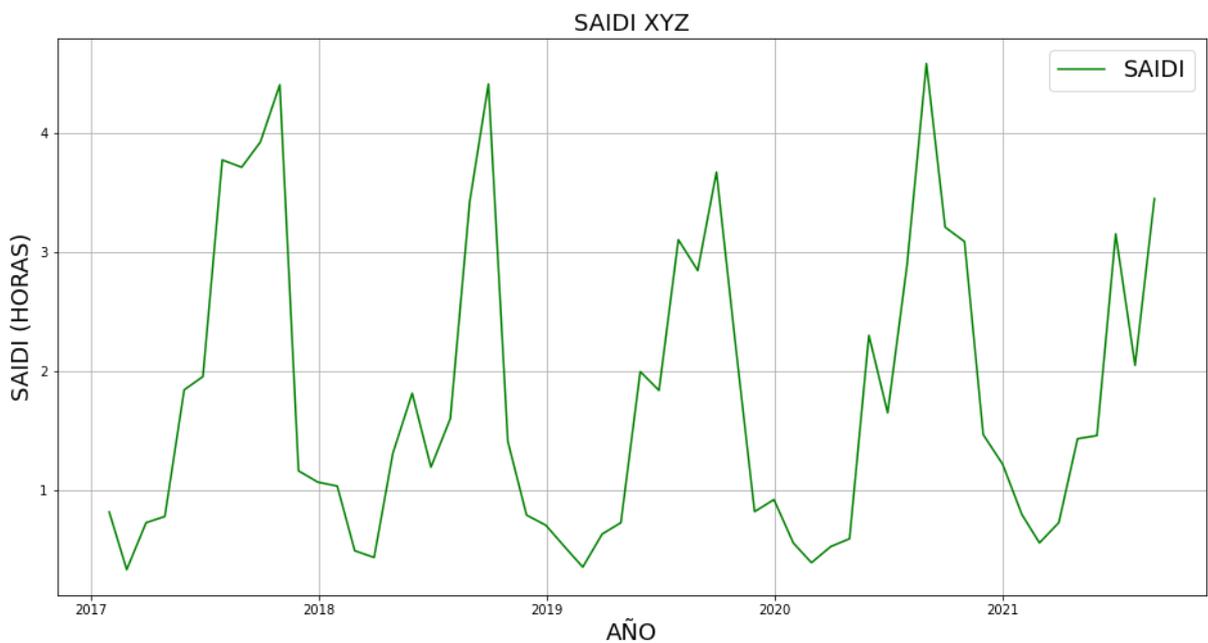


figura 11. SAIDI 2017-2021 Empresa XYZ

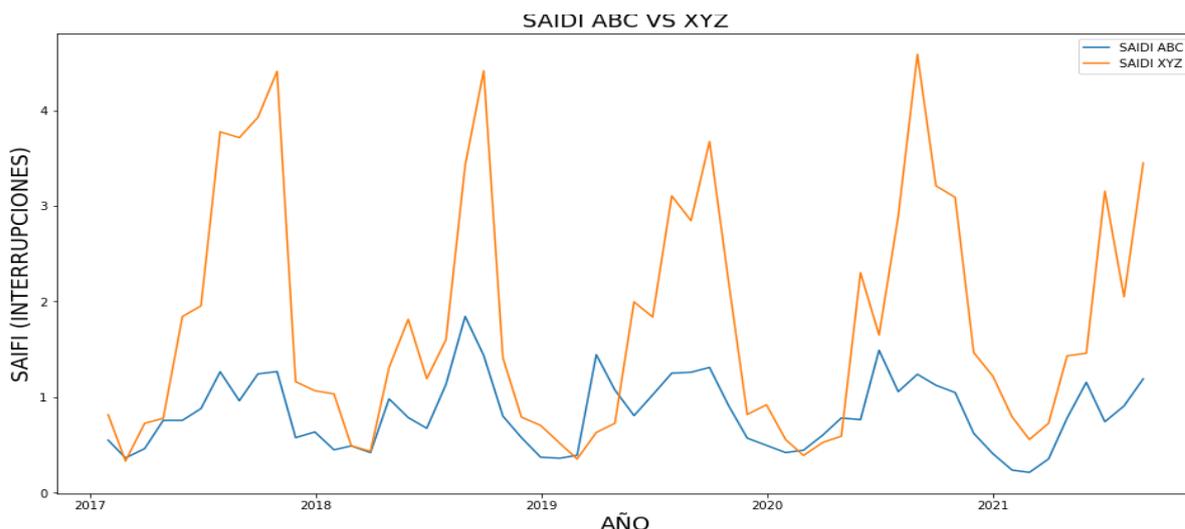


figura 12. SAIFI ABC vs XYZ 2017-2021

Finalizada esta etapa se procede al cálculo del indicador Tasa de Falla.

3.1.2 Tasa de Falla.

El procedimiento para el cálculo de tasa de falla es bastante similar al descrito para los indicadores SAIFI y SAIDI de la sección anterior, para este indicador hay que tener en cuenta los siguientes aspectos:

- Se contabilizarán únicamente las fallas que afectan la red de Media Tensión.
- No se contabilizarán las interrupciones que tengan horas de inicio repetida, puesto que el objetivo de este indicador es cuantizar los eventos que pueden provocar una o más interrupciones.
- Las interrupciones que tengan como inicio cargas puntuales no deberán de ser incluidas.

En la siguiente tabla se presenta el segmento de código que calcula la Tasa de Falla en media tensión y genera el grafico de comportamiento mensual.

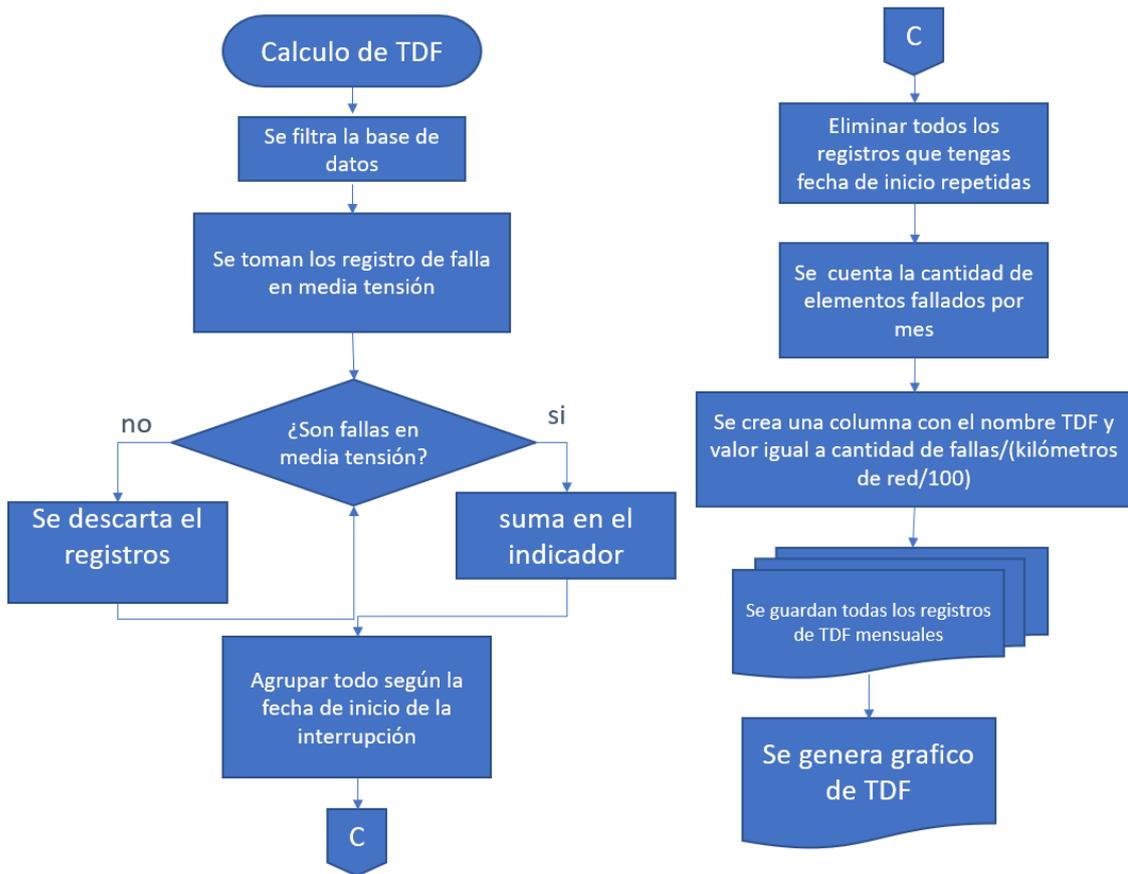


figura 13. Proceso de cálculo de TDF

CÓDIGO	COMENTARIO
<pre> filter1=TDF_MES.TIPO_ELEM_INI_INTERR!="Load" filter2= filter2*TDF_MES.TIPO_AFECTACION=="Transformador" </pre>	<p>Se preparan los filtros para los segmentos que no tengan como elemento de inicio cargas puntuales y que la afectación de la interrupción sea en transformador.</p>
<pre>TDF_MES_MT=TDF_MES[filter2]</pre>	<p>Los filtros anteriores se cargan al</p>

	DataFrame de interrupciones.
<pre>TDF_MES_MT = TDF_MES_MT.set_index('FECHA_INI_INTERR') TDF_MES_MT.sort_index(inplace=True)</pre>	Se establece la fecha de inicio como índice del DataFrame y se ordena de mayor a menor.
<pre>TDF_MES_MT=TDF_MES_MT.rename_axis("FECHA").reset_index()</pre>	Se pasa el índice a una columna del DataFrame y se le da el nombre de FECHA.
<pre>TDF_MES_MT = TDF_MES_MT.set_index('FECHA') TDF_MES_MT = TDF_MES_MT.resample('M').count()</pre>	Nuevamente se fija la fecha como índice del DataFrame y se cuentan la cantidad de elementos fallados por mes.
<pre>TDF_MES_MT.insert(loc=1, column="TDF"value=(TDF_MES_MT.PLACA_INI_INTERR/KM_ MT_ABC</pre>	Se crea una columna llamada TDF y se le da el valor de la cantidad de fallas por mes entre los kilómetros de red entre 100.
<pre>plt.figure(figsize=(16,8)) plt.grid() plt.title("TASA DE FALLA ABC",fontsize=18) plt.xlabel('AÑO', fontsize=18) plt.ylabel("TDF FALLAS/KM", fontsize=18) plt.legend(["TDF"], fontsize=18) print(TDF_MES_MT) plt.plot(TDF_MES_MT.TDF, color="RED") plt.legend(["TDF"], fontsize=18)</pre>	Código para generar grafica de TDF calculada.

Tabla 10. Segmento de código para el cálculo de TDF.

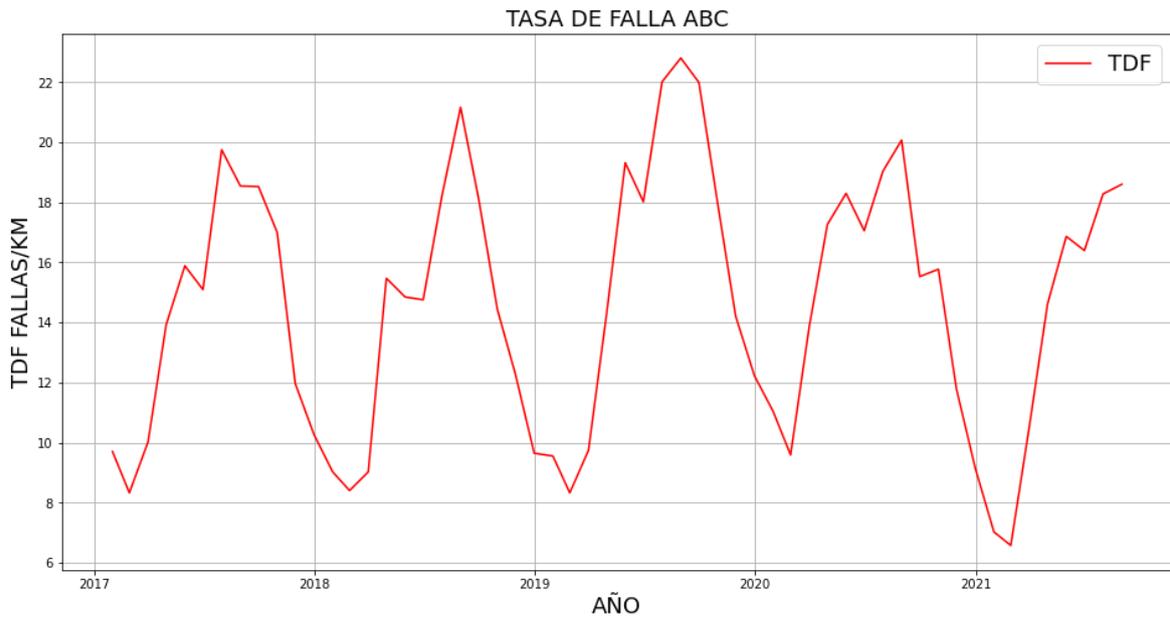


figura 14. TDF 2017-2021 Empresa ABC

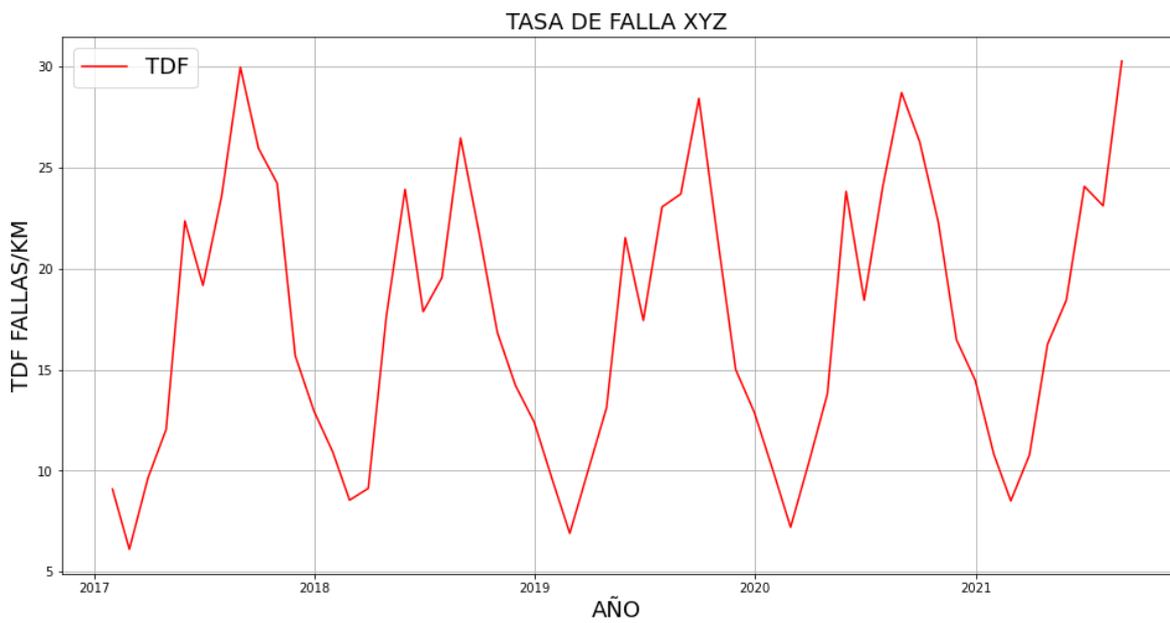


figura 15. TDF 2017-2021 Empresa XYZ

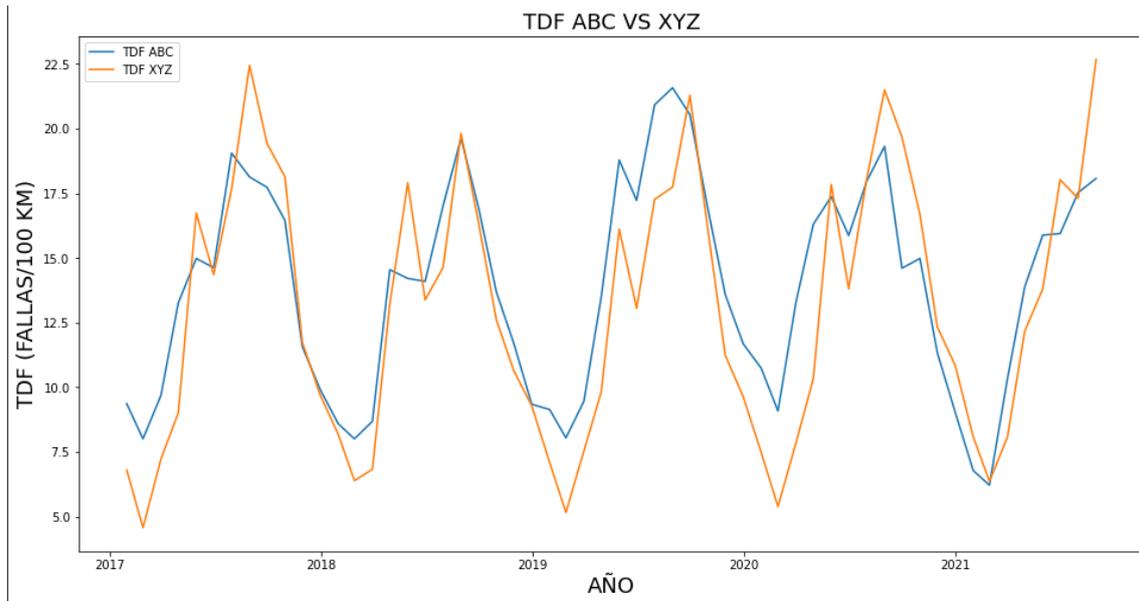


figura 16.TDF ABC vs XYZ 2017-2021.

En la siguiente sección se presenta el último de los indicadores en estudio

3.1.3 Tiempo Medio de Reparación.

Para el cálculo de TMR se deben establecer las condiciones características del indicador:

- En cada mes se deberá excluir el 10% de las interrupciones (5% con mayor duración y 5% con menor duración).
- Los causales de falla son los mismos para el cálculo de SAIFI y SAIDI.
- Solo se contabilizan interrupciones con duración mayor a tres minutos.

En la siguiente tabla se presenta el segmento de código que hace el cálculo del indicador TMR.

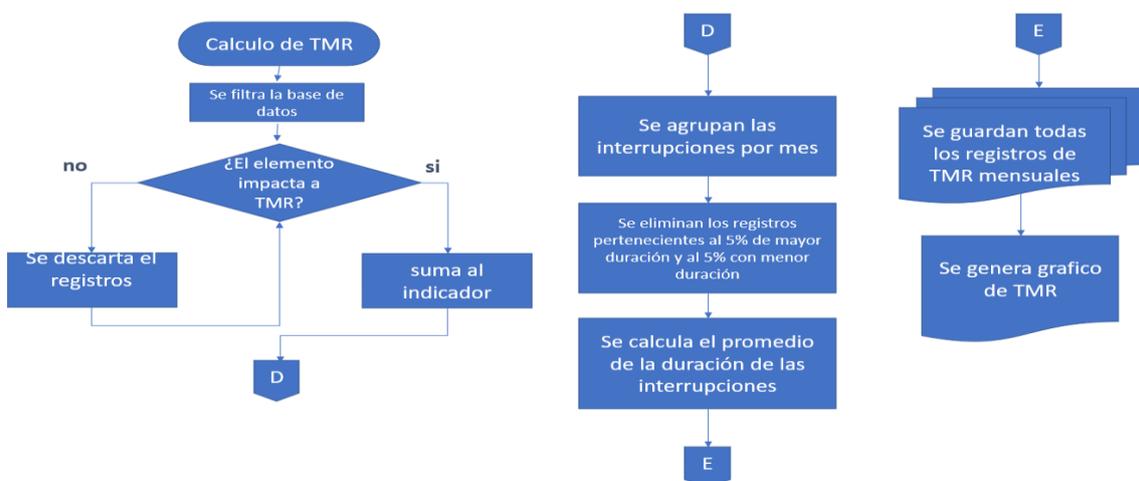


figura 17. Proceso de cálculo de TMR.

CÓDIGO	COMENTARIO
<pre>for i in range (2017,2022): for m in range (1,13):</pre>	<p>Se inician dos ciclos for el exterior para el movimiento de los años y el interior para el movimiento de los meses.</p>
<pre>dTMR=dA</pre>	<p>Se crea un DataFrame auxiliar para el cálculo de TMR</p>
<pre>dTMR["Check_Causa"]=dTMR.CAUSA.isin(d_CausaTMR.causal)</pre>	<p>Se comprueba si las interrupciones pertenecen a los causales de TMR.</p>
<pre>dTMR=dTMR.drop(dTMR[dTMR["Check_Causa"]!=True].index)</pre>	<p>Se eliminan las interrupciones que no sean causal de TMR.</p>
<pre>dTMR["Check_Elemento"]=dTMR.ELEMENTO.isin(d_ElementsTMR["ELEMENTO_FALLADO"])</pre>	<p>Se comprueba si las interrupciones son con elemento de inicio de TMR.</p>
<pre>dTMR=dTMR.drop(dTMR[dTMR["Check_Elemento"]!=True].index)</pre>	<p>Se eliminan las interrupciones que no tienen de elemento de inicio elemento de TMR.</p>
<pre>dTMR=dTMR.sort_values(by=['DURACION_HRS'])</pre>	<p>Se ordena las interrupciones según el tiempo de duración.</p>
<pre>dTMR=dTMR.drop(dTMR[dTMR["año"]!=i].index) dTMR=dTMR.drop(dTMR[dTMR["mes"]!=m].index)</pre>	<p>Se conservan solo los registros que sean del mes y año de interés.</p>
<pre>dTMR=dTMR.drop(dTMR[dTMR["TIPO_AFECTACION"]!="Transformador"].index)</pre>	<p>Se conservan solo las interrupciones</p>

	con afectación de transformador.
<code>dTMR["Posicion"]=np.arange(1, len(dTMR)+1)</code>	Se agrega una columna para guardar la posición.
<code>numero_de_no_cuenta_datos=round(len(dTMR.index)*5/100)</code>	Se calcula la cantidad de interrupciones a eliminar (5%).
<code>cantidad=round(len(dTMR.index))</code> <code>Numeros_Sup= np.array(range(cantidad-numero_de_no_cuenta_datos,cantidad+1,1))</code> <code>Numeros_Inf= np.array(range(1,numero_de_no_cuenta_datos+1,1))</code> <code>Numeros={"Registro":np.concatenate((Numeros_Inf, Numeros_Sup), axis=0)}</code> <code>dN=pd.DataFrame(Numeros</code>	Se contabilizan la cantidad total de elementos actual en DataFrame, se guardan las posiciones de los registros de 5% superior e inferior, se concatena y se guarda en un DataFrame
<code>dTMR["Check"]=dTMR.Posicion.isin(dN.Registro)</code> <code>dTMR=dTMR.drop(dTMR[dTMR["Check"]!=False].index)</code>	Se comprueba si los elementos pertenecen al 10% a eliminar y se elimina.
<code>BD_TMR = pd.concat([BD_TMR, dTMR])</code>	Se crea un DataFrame y se guardan todas las interrupciones que suman al TMR.
<code>TMR_CALC= BD_TMR[['FECHA_INI_INTERR', 'DURACION_HRS']]</code> <code>TMR_CALC = TMR_CALC.set_index('FECHA_INI_INTERR')</code> <code>TMR_CALC = TMR_CALC.resample('m').mean()</code> <code>TMR_CALC.sort_index(inplace=True)</code>	Se crea un DataFrame auxiliar para guardar el registro de fecha y duración de las interrupciones. Posteriormente se agrupan los datos por medio de un promedio mensual.

```

plt.figure(figsize=(16,8))

plt.grid()

plt.title("TIEMPO MEDIO DE REPARACIÓN
ABC",fontsize=18)

plt.xlabel('AÑO', fontsize=18)

plt.ylabel("TMR (HORAS)", fontsize=18)

plt.legend(["TMR"], fontsize=18)

#print(TDF_MES_MT)

plt.plot(TMR_CALC, color="RED")

```

Código para generar grafica de TMR calculado.

Tabla 11. Segmento de código para el cálculo de TMR.

El código completo del procesamiento se encuentra disponible en el anexo 1.

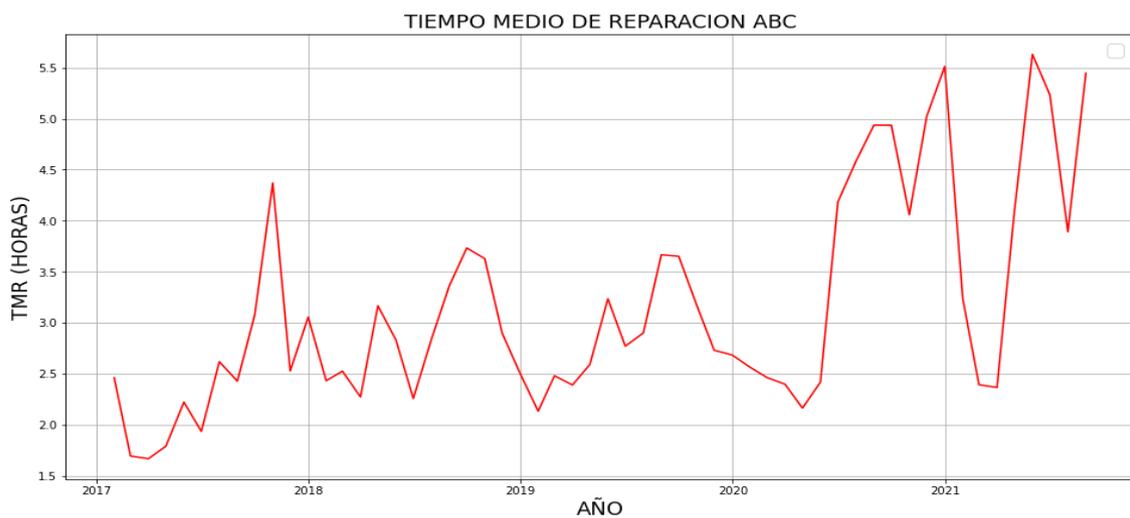


figura 18. TMR 2017-2021 Empresa ABC

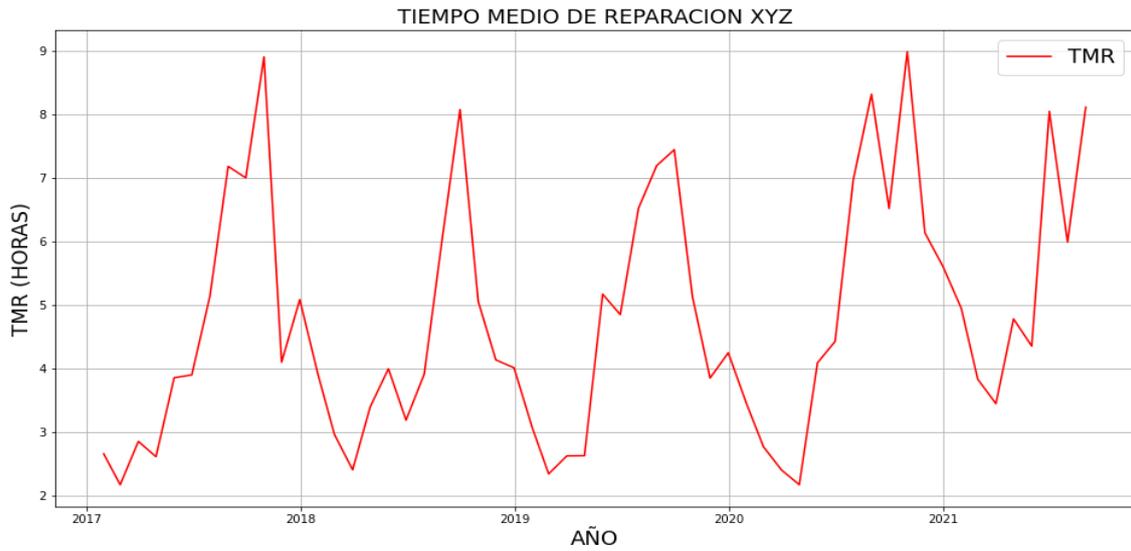


figura 19. TMR 2017-2021 Empresa ABC

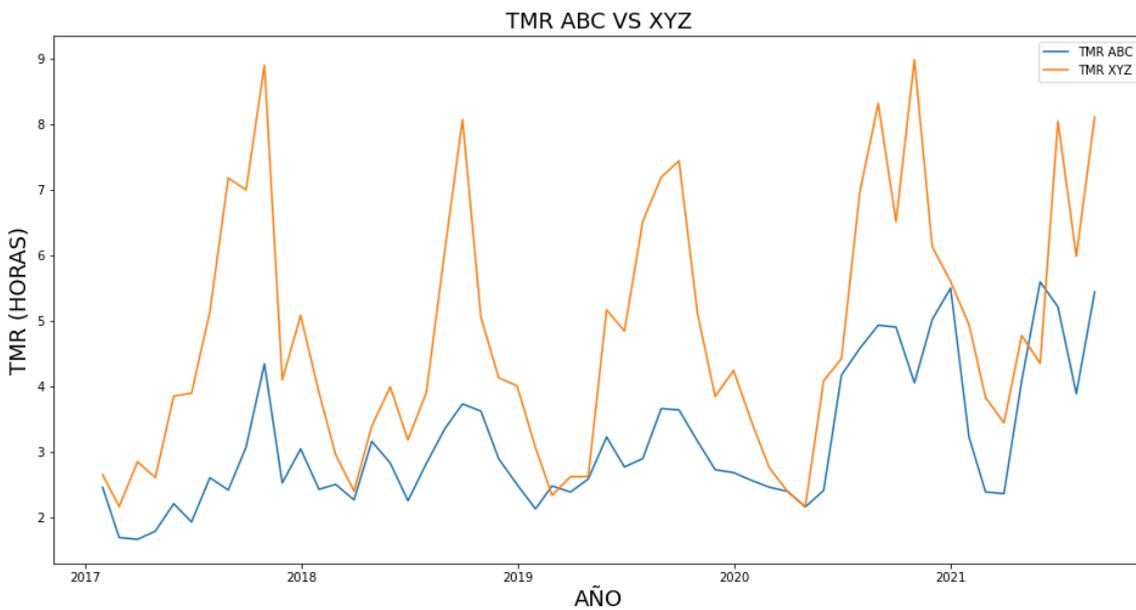


figura 20. TMR ABC vs XYZ 2017-2021.

En las siguientes dos secciones se presentan un resumen de los indicadores para las empresas en estudio y un análisis del comportamiento de estos.

3.2 RESUMEN DE INDICADORES.

En las tres secciones anteriores se han calculado los cuatro indicadores objeto de estudio en esta investigación, a continuación, se presenta el resumen de los indicadores para 2020

de la empresa ABC y XYZ, por simplicidad el resto de los años estará disponibles en el anexo 2.

INDICADORES SAIFI-SAIDI-TDF-TMR "EMPRESA ABC"				
FECHA	SAIFI (INTERRUPCIONES)	SAIDI (HORAS)	TDF (FALLAS/100 KM)	TMR (HORAS)
ENE-20	0.2034	0.4210	10.74	2.56
FEB-20	0.2187	0.4447	9.08	2.46
MAR-20	0.3625	0.6025	13.29	2.40
ABR-20	0.4387	0.7807	16.30	2.16
MAY-20	0.3285	0.7626	17.38	2.42
JUN-20	0.4968	1.4891	15.87	4.18
JUL-20	0.3630	1.0545	17.96	4.58
AGO-20	0.3853	1.2392	19.32	4.94
SEP-20	0.3056	1.1196	14.61	4.94
OCT-20	0.3826	1.0470	14.98	4.06
NOV-20	0.2081	0.6208	11.35	5.02
DIC-20	0.1252	0.4088	9.03	5.51

Tabla 12. Resumen de indicadores 2020 ABC

INDICADORES SAIFI-SAIDI-TDF-TMR "EMPRESA XYZ"				
FECHA	SAIFI (INTERRUPCIONES)	SAIDI (HORAS)	TDF (FALLAS/100 KM)	TMR (HORAS)
ENE-20	0.195	0.5562	7.48	3.45
FEB-20	0.1422	0.3912	5.39	2.76
MAR-20	0.2654	0.5268	7.85	2.4
ABR-20	0.2478	0.5924	10.33	2.17
MAY-20	0.5787	2.3013	17.84	4.09
JUN-20	0.4224	1.6499	13.81	4.42
JUL-20	0.6617	2.8902	18.04	6.97
AGO-20	0.7154	4.5835	21.50	8.32
SEP-20	0.6884	3.2099	19.68	6.52
OCT-20	0.4394	3.0876	16.69	8.99
NOV-20	0.3046	1.4674	12.34	6.13
DIC-20	0.2887	1.2193	10.84	5.6

Tabla 13. Resumen de indicadores 2020 XYZ

3.3 ANÁLISIS DE INDICADORES.

A nivel de SAIFI y SAIDI junio de 2020 fue el mes con mayor impacto en los indicadores para la empresa ABC, sin embargo, no es el mes de 2020 con mayor tasa de falla lo que da a entender que la razón principal del aumento de los indicadores es la cantidad de clientes afectados por interrupción.

El caso de la empresa XYZ muestra un considerable desvío en SAIFI, SAIDI y TDF para el mes de agosto de 2020, con esta información puede concluirse que debido a la época lluvia hubo un considerable aumento en la cantidad de eventos ocurridos y estos, a su vez fueron de altas cantidades de clientes.

En el caso de tasa de falla agosto 2020 para ABC fue el mes mayor impacto, pero se observa un ligero aumento en los indicadores SAIFI y SAIDI del mes anterior, esto quiere decir que si bien es cierto la cantidad de fallas han aumentado estas no afectaron tantos clientes como en meses anteriores.

El caso del TMR en ABC demuestra que diciembre fue el mes con peor desempeño en este indicador con 5.51 horas; sin embargo, al analizar los valores de SAIFI, SAIDI y TDF concluimos que este indicador no es resultado de un aumento de fallas ni representa grandes cantidades lo que podría indicar que las fallas han sido distribuidas en zonas de difícil acceso para el personal de atención o por la tendencia histórica de la disminución de fallas se cuenta con menos personal.

En XYZ el TMR tuvo a octubre de 2020 como el mes de peor desempeño llegando casi a 9 horas el promedio de atención de reclamos, al observar tanto SAIFI, SAIDI y TDF, puede concluirse que, si bien es cierto, no es el mes con mayor cantidad de fallas, no dejan de ser valores considerables, el aumento del tiempo de reparación puede ser asociado a la dispersión de los reclamos lo que provoca que las parejas de atención de fallas tengan que desplazarse mayores distancias.

Para un análisis completo de los indicadores es necesario establecer métricas que sirvan de comparación del comportamiento real con el esperado, es por esto que en el siguiente capítulo se iniciara una exploración al análisis de los indicadores a través de series temporales a fin de obtener métricas que sirvan para establecer parámetros de comparación, en la industria estas métricas son conocidas como metas.

CAPÍTULO IV – MINERÍA DE DATOS

Este apartado describe el proceso empleado para el cálculo y estimación de las proyecciones para los indicadores.

Las generalidades de la técnica de minería de datos utilizando el modelo CRISP – DM se encuentra detallada en el anexo 3.

4.1 ANÁLISIS DE SERIES DE TIEMPO.

En las series de tiempo cada instancia representa un momento en el tiempo y los valores asociados con ese tiempo, generalmente las instancias se encuentran separadas por un intervalo regular de tiempo [12].

Las series de tiempo son útiles para modelar diferentes procesos naturales y económicos como con comportamiento estocástico: comportamientos del mercado de valores, comportamientos poblacionales, observaciones científicas, entre otros [13].

Generalmente las series de tiempo se usan para realizar pronósticos de regresión, y para tal fin se descompone la serie algunos de los siguientes elementos:

- **Tendencia:** Esta indica la dirección en la que la serie de ha ido moviendo en un periodo de tiempo determinado.
- **Estacionalidad:** Son patrones casi idénticos que aparecen con cierta regularidad dentro de la serie de tiempo, generalmente con duración similar en los diferentes periodos.
- **Comportamiento cíclico:** Son patrones que se presentan en la serie de tiempo, pero sin un comportamiento enmarcado dentro de un periodo fijo, es similar a la estacionalidad solo que puede comprender periodos de tiempo más extensos.
- **Comportamiento aleatorio:** Estos representan los cambios esporádicos que tienen un comportamiento estocástico y que son causados por elementos externos a los atributos que componen la serie de tiempo.

El análisis de series de tiempo permite visualizar si el comportamiento de los datos tiene componentes estacionales [14], para este análisis se selecciona el algoritmo naive por medio de la descomposición STL (Seasonal and Trend descomposition using Loess), que utiliza Loess como método estadístico para estimar relaciones no lineales, usando la función `tsa.seasonal_decompose` del módulo `statsmodels.api` de Python.

Como puede verse en la figura el análisis por serie de tiempo permite descomponer el comportamiento de una señal en sus diversas componentes (tendencia y estacionalidad).

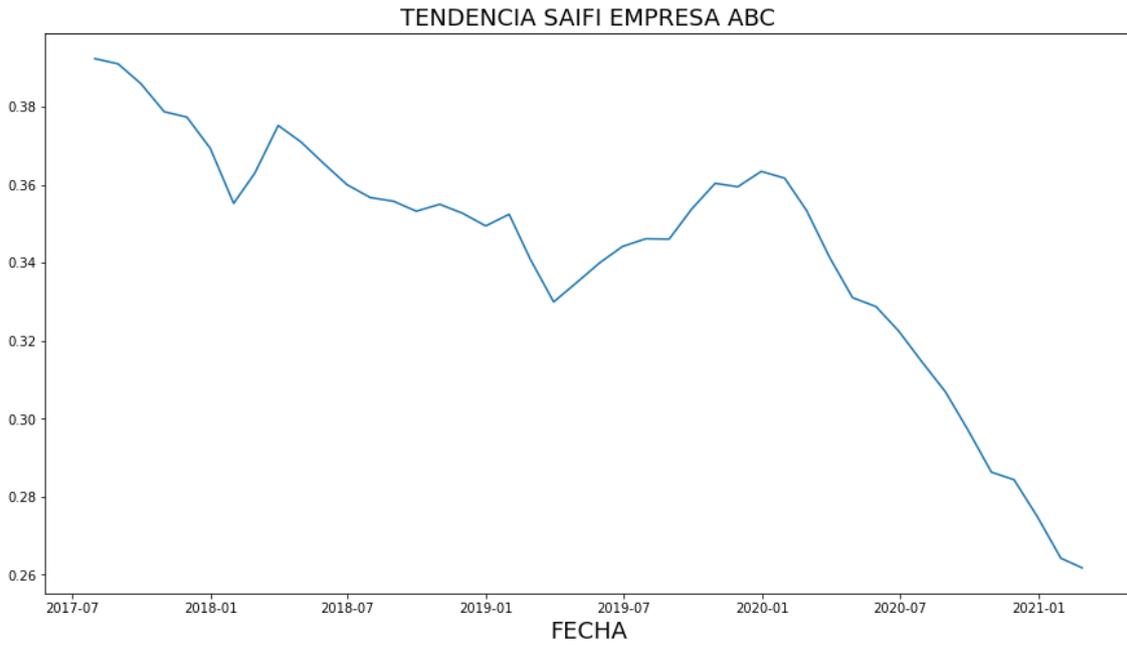


figura 21. Tendencia Serie de Tiempo SAIFI ABC

En la figura 14 se visualiza una clara tendencia a la baja en el indicador SAIFI para la empresa ABC, esto producto de las mejoras implementadas en la red eléctrica por la distribuidora manifestándose en la disminución de fallas en trocales principales. Mientras que el comportamiento estacional observado en la figura 15 tiene un marcado periodo de aumento correspondiente a los meses de junio, julio y agosto que coinciden con los meses con mayor afectación climática por lluvias.

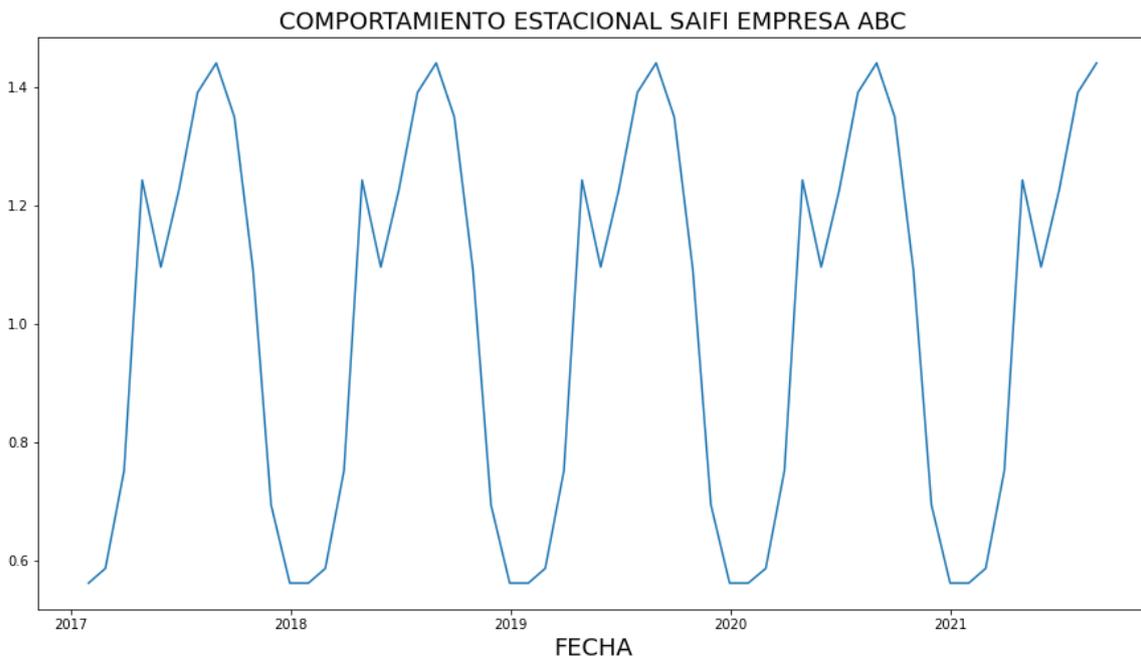


figura 22. Tendencia Serie de Tiempo SAIFI ABC

Una vez realizado este análisis se concluye que la serie de tiempo de SAIFI tiene un claro comportamiento estacional por que pueden emplazarse métodos de machine learning que describan su comportamiento, en la siguiente sección se realizaran aproximaciones a los conjuntos de datos a través de regresiones lineales.

4.2 ANÁLISIS POR REGRESIONES LINEALES.

Un análisis de datos por regresión lineal polinómica proporciona una estimación de los valores dentro del conjunto disponible, para este cálculo haremos uso de las herramientas disponibles en el módulo sklearn de Python, para la estimación se ocupará el 30% de los datos como datos de prueba y el resto será utilizado como datos de entrenamiento.

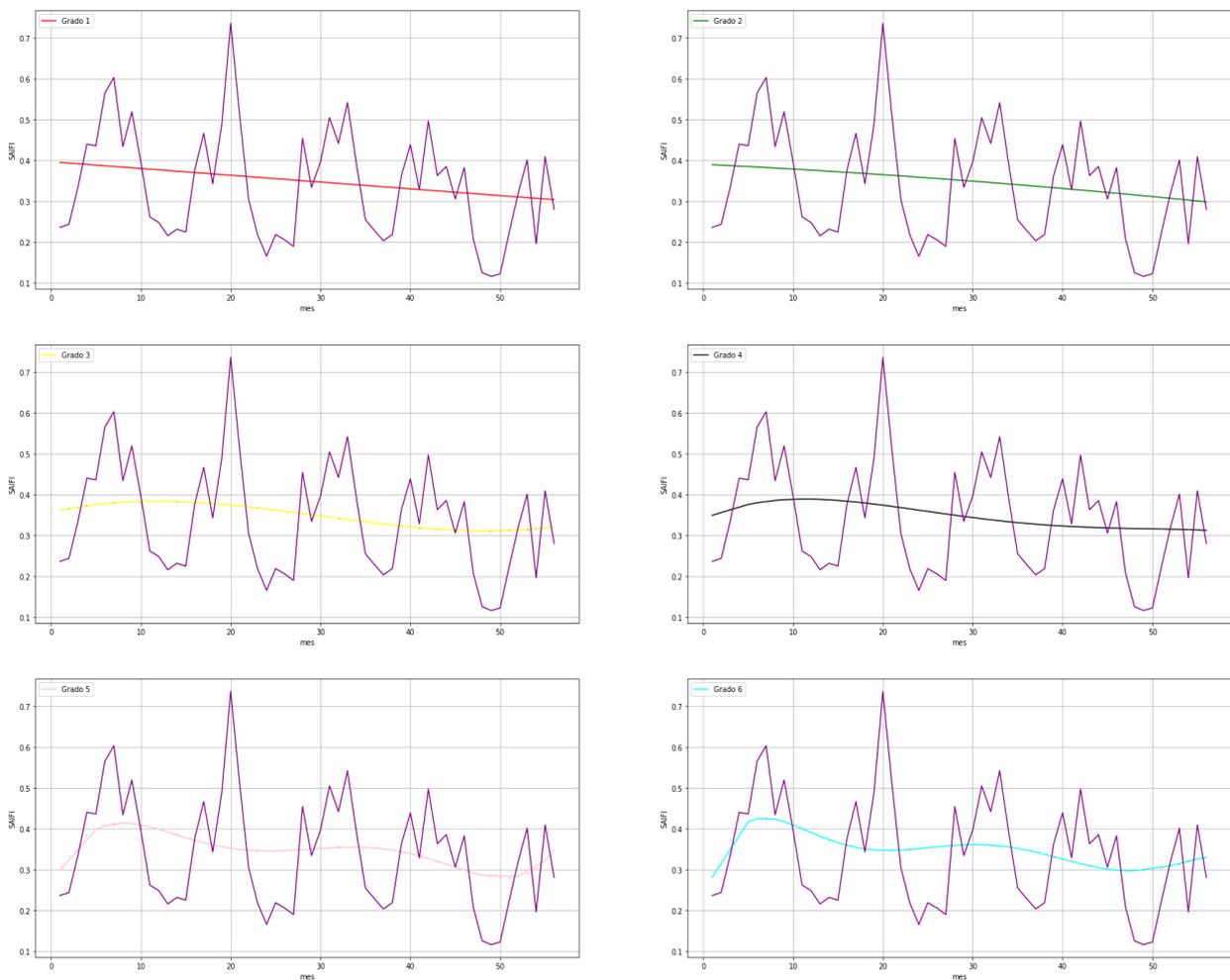


figura 23. Análisis por regresiones lineales SAIFI ABC.

Como puede verse en la figura 16 a medida el grado del polinomio aumenta la aproximación por medio de la curva es mayor; sin embargo, este tipo de aproximaciones son de utilidad en caso se desee estimar un valor dentro del conjunto de datos y no para

estimar valores futuros, para este objetivo en la siguiente sección se describirá la proyección usando una red neuronal MLP.

Tanto el análisis por serie de tiempo y regresión polinómica de empresa XYZ se encuentra disponible en los anexos 1 y 4.

4.3 ANÁLISIS POR RED NEURONAL FEEDFORWARD.

El objetivo de la implementación de una red neuronal es poder encontrar un modelo que describa el compartimento estacional de una serie de tiempo y a su vez proporcione una visualización del posible comportamiento futuro, se seleccionó la red neuronal FeedForward por su simplicidad y debido a que la naturaleza de los datos de indicadores se adapta fácilmente a las características de esta.

Para la creación de la red neuronal se utilizarán 40 meses como datos de entrenamiento y el resto será utilizado como datos de validación.

A continuación, se describe el proceso para la obtención del indicador SAIFI de la empresa ABC; sin embargo, es similar para el resto de los indicadores.

En resumen, la arquitectura de la red neuronal será la siguiente:

- Entrada de 12 datos.
- Una capa oculta con 20000 neuronas.
- La salida será una sola neurona.
- La función de activación será atreves de tangente hiperbólica.
- Como optimizador se utilizará RMRPROP y métrica de perdida Mean Absolute Error.
- El cálculo de eficiencia será atreves de Mean Squared Error.
- Como datos de salida se tendrán el valor del indicador en los últimos cuatro meses disponibles con lo que se hará la comparativa.

Para entrenar la red se usarán 3000 iteraciones, en busca del menor margen de error, el detalle del código en Python de red neuronal se encuentra disponible en el anexo 1.

Consideramos dos métricas de rendimiento para evaluar la precisión del pronóstico: estas son el error cuadrático medio (RMSE) y el error porcentual absoluto medio (MAPE) utilizando las siguientes ecuaciones.

$$RMSE = \frac{1}{\sqrt{N}} \sqrt{\sum_{i=1}^N (I_{(pred,i)} - I_{(meas,i)})^2} \quad EC 5$$

$$MAPE = 100 * \frac{1}{N} \sum_{i=1}^N \left| \frac{I_{(meas,i)} - I_{(pred,i)}}{I_{(meas,i)}} \right| \quad EC 6$$

Donde:

N = Número total de muestras.

I_{pred} = Valor calculado.

I_{meas} = Valor real.

Aplicando las ecuaciones 5 y 6 se obtienen un error cuadrático medio de 0.1379 clientes/clientes y un error porcentual absoluto medio de 44%.

INDICADOR	DATO	VALOR REAL	VALOR PROYECTADO	RMSE	MAPE
SAIFI	1	0.4477	0.4096	0.1436	46%
	2	0.2078	0.4659		
	3	0.4300	0.3474		
	4	0.2804	0.3674		

Tabla 14. Comparativa de datos reales vs valores proyectados para SAIFI empresa ABC

En la siguiente tabla se muestra el resumen de las estimaciones realizadas para el resto de los indicadores:

INDICADOR	MES DE VALIDACIÓN	VALOR REAL	VALOR PROYECTADO	RMSE	MAPE
SAIFI	1	0.4477	0.4096	0.1436	46%
	2	0.2078	0.4659		
	3	0.4300	0.3474		
	4	0.2804	0.3674		
SAIDI	1	1.1557	1.0866	0.4029	38%
	2	0.7435	1.4652		
	3	0.9087	1.2228		
	4	1.1923	1.3503		
TDF	1	15.8887	16.1111	0.7955	4%
	2	15.9452	15.4243		
	3	17.5285	18.7407		
	4	18.0750	17.2142		
TMR	1	5.5940	5.2580	0.7670	13%
	2	5.2133	5.4051		
	3	3.8896	5.3214		
	4	5.4415	5.0502		

Tabla 15. Comparativa de datos reales vs valores proyectados para empresa para SAIDI, SAIDI, TDF Y TMR empresa ABC

La siguiente etapa consiste en la estimación de valores de indicador fuera del conjunto de datos, para esto se utilizarán únicamente los registros de los años 2019 y 2020 dentro del modelo.

4.4 PROYECCIÓN DE INDICADORES 2021.

Más allá de la validación del modelo, para la operación de las empresas distribuidoras sirve de poco la utilización del modelo para valores conocidos, es por esto en esta sección se presentan los resultados para 2021, utilizando únicamente como datos de entrada los indicadores de 2019 y 2020.

Con esta implementación se obtienen los datos de la siguiente tabla para SAIFI:

MES	REAL	PROYECTADO	RMSE	MAPE
ENERO	0.1322	0.1913	0.0852	29%
FEBRERO	0.1467	0.1955		
MARZO	0.2880	0.2770		
ABRIL	0.3346	0.3453		
MAYO	0.4477	0.3856		
JUNIO	0.2078	0.4109		
JULIO	0.4300	0.3713		
AGOSTO	0.2804	0.3388		
SEPTIEMBRE		0.2825		
OCTUBRE		0.2473		
NOVIEMBRE		0.1922		
DICIEMBRE		0.1785		

Tabla 16. Comparativa de datos reales vs valores proyectados de SAIFI para empresa ABC

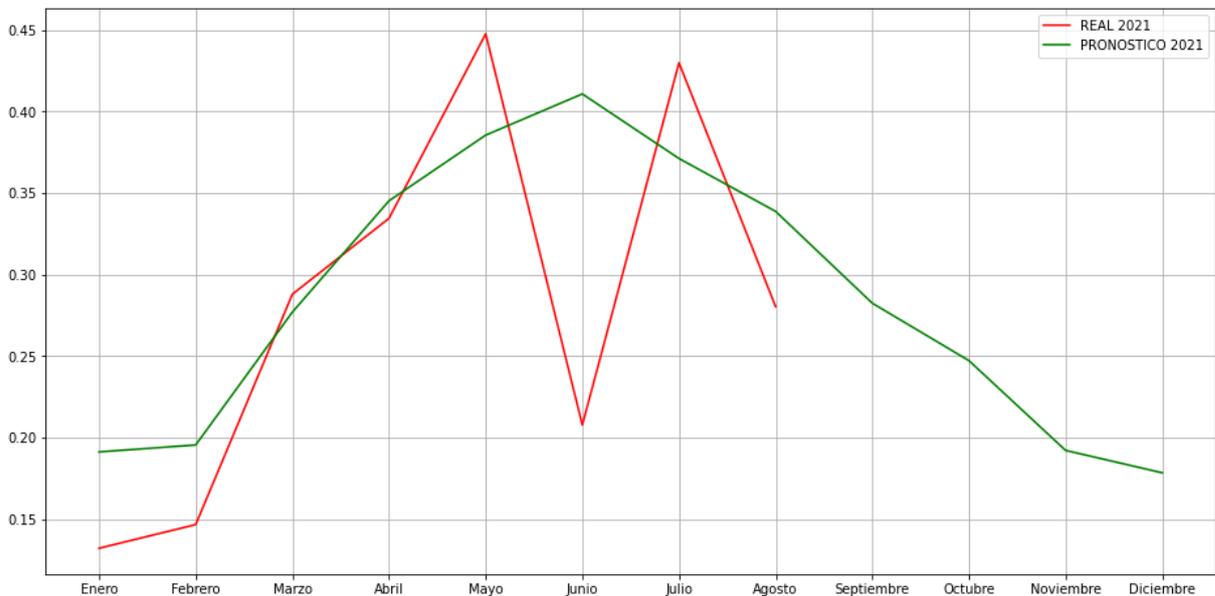


figura 24. Valores Reales (rojo) vs Valores Predichos (verde) de SAIFI para empresa ABC.

Siguiendo el mismo proceso para la proyección de SAIFI, se procede a estimar el resto de los indicadores para 2021, obteniendo los siguientes resultados.

MES	REAL	PROYECTADO	RMSE	MAPE
ENERO	0.2384	0.4269	0.2569	44%
FEBRERO	0.2142	0.4194		
MARZO	0.3545	0.4847		
ABRIL	0.7821	0.7085		
MAYO	1.1557	0.9087		
JUNIO	0.7435	1.2026		
JULIO	0.9087	1.2851		
AGOSTO	1.1923	1.3134		
SEPTIEMBRE		1.1321		
OCTUBRE		0.8275		
NOVIEMBRE		0.5197		
DICIEMBRE		0.4266		

Tabla 17. Comparativa de datos reales vs valores proyectados de SAIDI para empresa ABC

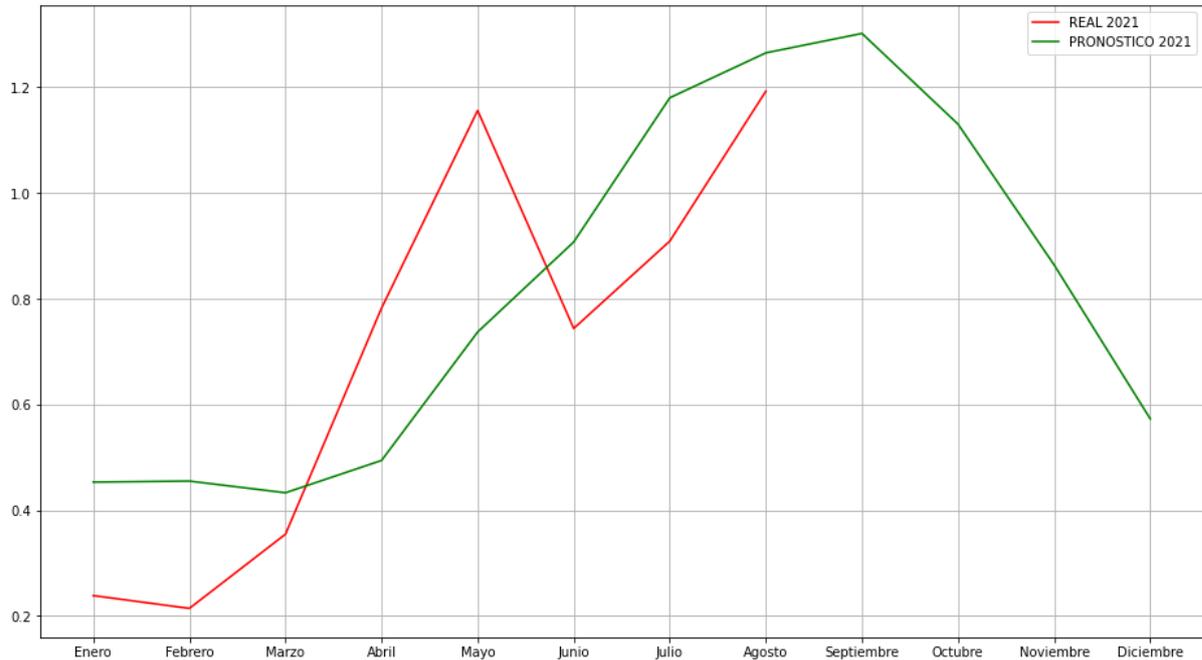


figura 25. Valores Reales (rojo) vs Valores Predichos (verde) de SAIDI para empresa ABC.

MES	REAL	PROYECTADO	RMSE	MAPE
ENERO	6.7852	8.1614	1.1086	10%
FEBRERO	6.2198	8.3138		
MARZO	10.3286	9.9546		
ABRIL	13.8908	12.6470		
MAYO	15.8887	15.3561		
JUNIO	15.9452	14.8242		
JULIO	17.5285	17.8829		
AGOSTO	18.0750	17.6279		
SEPTIEMBRE		16.4055		
OCTUBRE		15.2205		
NOVIEMBRE		12.5680		
DICIEMBRE		10.1110		

Tabla 18. Comparativa de datos reales vs valores proyectados de TDF para empresa ABC



figura 26. Valores Reales (rojo) vs Valores Predichos (verde) de TDF para empresa sABC.

MES	REAL	PROYECTADO	RMSE	MAPE
ENERO	3.2286	4.4251	0.7412	17%
FEBRERO	2.3899	3.0799		
MARZO	2.3643	2.6093		
ABRIL	4.0897	3.6702		
MAYO	5.5940	4.9956		
JUNIO	5.2133	5.3266		
JULIO	3.8896	5.2544		
AGOSTO	5.4415	5.3088		
SEPTIEMBRE		5.4596		
OCTUBRE		5.4750		
NOVIEMBRE		5.4193		
DICIEMBRE		5.1375		

Tabla 19. Comparativa de datos reales vs valores proyectados de TMR para empresa ABC.

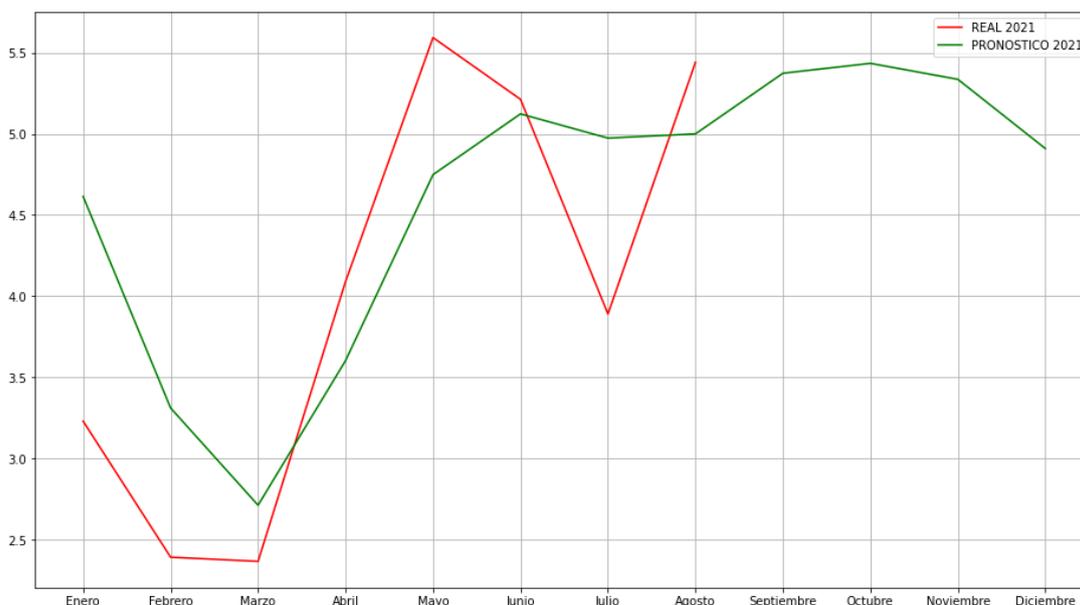


figura 27. Valores Reales (rojo) vs Valores Predichos (verde) de TMR para empresa ABC.

Como pudimos ver el método funciona y debería ser aplicable a cualquier distribuidora eléctrica, con el fin de verificar lo anterior mencionado se procederá al cálculo de indicadores a un banco de datos distinto mediante el modelado de proyección 2021 para una empresa denominada XYZ

4.5 COMPROBACIÓN DEL MODELO PARA EMPRESA XYZ.

Entrenando el modelo con data de interrupciones de 2017 a agosto de 2021 se obtienen los siguientes resultados:

INDICADOR	MES DE VALIDACIÓN	VALOR REAL	VALOR PROYECTADO	RMSE	MAPE
SAIFI	1	0.3939	0.4409	0.1093	16%
	2	0.5813	0.3906		
	3	0.5294	0.6237		
	4	0.6391	0.6223		
SAIDI	1	1.4593	1.1699	0.6454	16%
	2	3.1538	1.8974		
	3	2.0497	2.0895		
	4	3.4496	3.3998		
TDF	1	13.8054	18.1250	2.9064	16%
	2	18.0261	15.2450		
	3	17.3062	18.9689		
	4	22.6702	20.5184		
TMR	1	4.3490	4.9025	1.3464	16%
	2	8.0447	7.4848		
	3	5.9851	8.5596		
	4	8.1085	8.0490		

Tabla 20. Comparativa de datos reales vs valores proyectados para empresa para SAIFI, SAIDI, TDF Y TMR empresa XYZ

Se obtienen los datos de la siguiente tabla para el SAIFI:

MES	REAL	PROYECTADO	RMSE	MAPE
ENERO	0.1951	0.2178	0.1091	15%
FEBRERO	0.1604	0.1662		
MARZO	0.2321	0.2326		
ABRIL	0.4501	0.2366		
MAYO	0.3939	0.3526		
JUNIO	0.5813	0.3677		
JULIO	0.5294	0.5234		
AGOSTO	0.6391	0.5982		
SEPTIEMBRE		0.5203		
OCTUBRE		0.4093		
NOVIEMBRE		0.3084		
DICIEMBRE		0.2450		

Tabla 21. Comparativa de datos reales vs valores proyectados de SAIFI para empresa XYZ

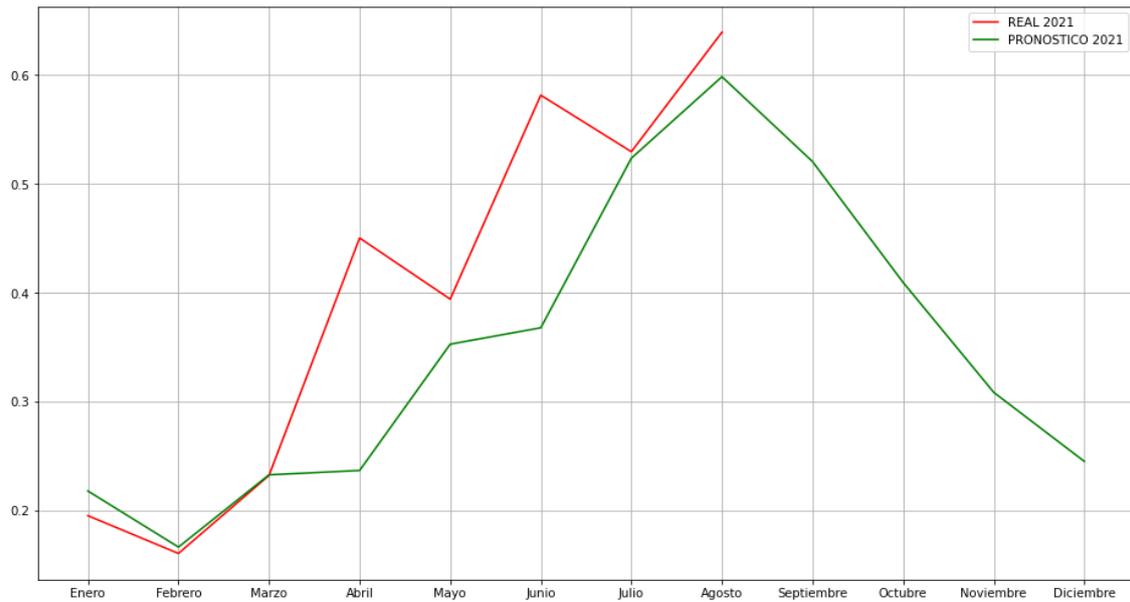


figura 28. Valores Reales (rojo) vs Valores Predichos (verde) de SAIFI para empresa XYZ

Habiendo calculado el SAIFI y observando que obtenemos valores bajos de MAPE, se procede a estimar los demás indicadores para 2021 de la empresa XYZ.

MES	REAL	PROYECTADO	RMSE	MAPE
ENERO	0.7947	0.7441	0.5841	24%
FEBRERO	0.5577	0.3899		
MARZO	0.7281	0.5930		
ABRIL	1.4327	0.5016		
MAYO	1.4593	1.0858		
JUNIO	3.1538	1.8674		
JULIO	2.0497	1.9444		
AGOSTO	3.4496	3.5402		
SEPTIEMBRE		2.7759		
OCTUBRE		3.0051		
NOVIEMBRE		2.0004		
DICIEMBRE		1.0975		

Tabla 22. Comparativa de datos reales vs valores proyectados de SAIDI para empresa XYZ

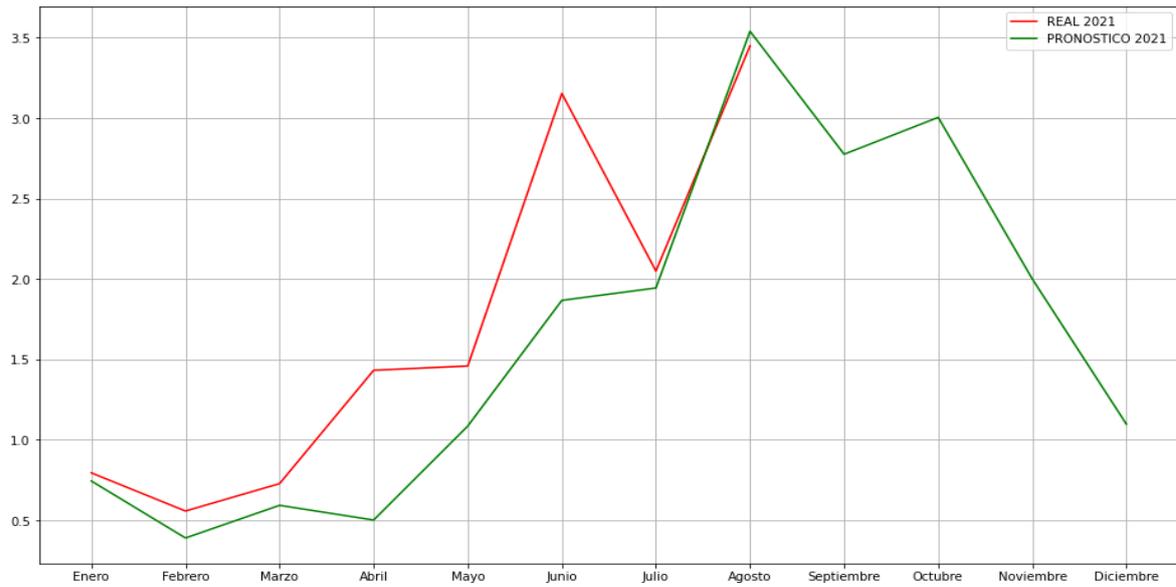


figura 29. Valores Reales (rojo) vs Valores Predichos (verde) de SAIDI para empresa XYZ.

MES	REAL	PROYECTADO	RMSE	MAPE
ENERO	8.0885	8.0187	1.9890	8%
FEBRERO	6.3663	6.3053		
MARZO	8.0743	8.1985		
ABRIL	12.1821	12.2128		
MAYO	13.8054	18.1217		
JUNIO	18.0261	16.4777		
JULIO	17.3062	19.3543		
AGOSTO	22.6702	20.1404		
SEPTIEMBRE		18.6397		
OCTUBRE		14.6614		
NOVIEMBRE		11.8872		
DICIEMBRE		11.1328		

Tabla 23. Comparativa de datos reales vs valores proyectados de TDF para empresa XYZ

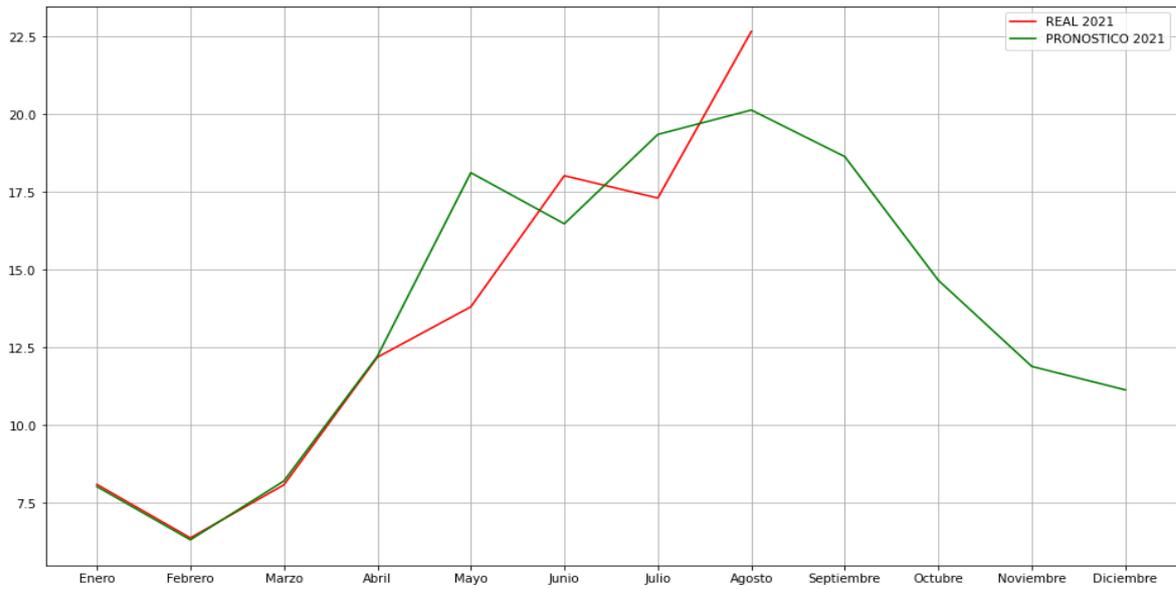


figura 30. Valores Reales (rojo) vs Valores Predichos (verde) de TDF para empresa XYZ

MES	REAL	PROYECTADO	RMSE	MAPE
ENERO	4.9435	4.8359	0.9593	10%
FEBRERO	3.8293	3.8188		
MARZO	3.4433	3.8631		
ABRIL	4.7775	4.5756		
MAYO	4.3490	4.8471		
JUNIO	8.0447	7.6707		
JULIO	5.9851	8.5505		
AGOSTO	8.1085	8.5141		
SEPTIEMBRE		8.8310		
OCTUBRE		8.7256		
NOVIEMBRE		8.6040		
DICIEMBRE		8.4730		

Tabla 24. Comparativa de datos reales vs valores proyectados de TMR para empresa XYZ

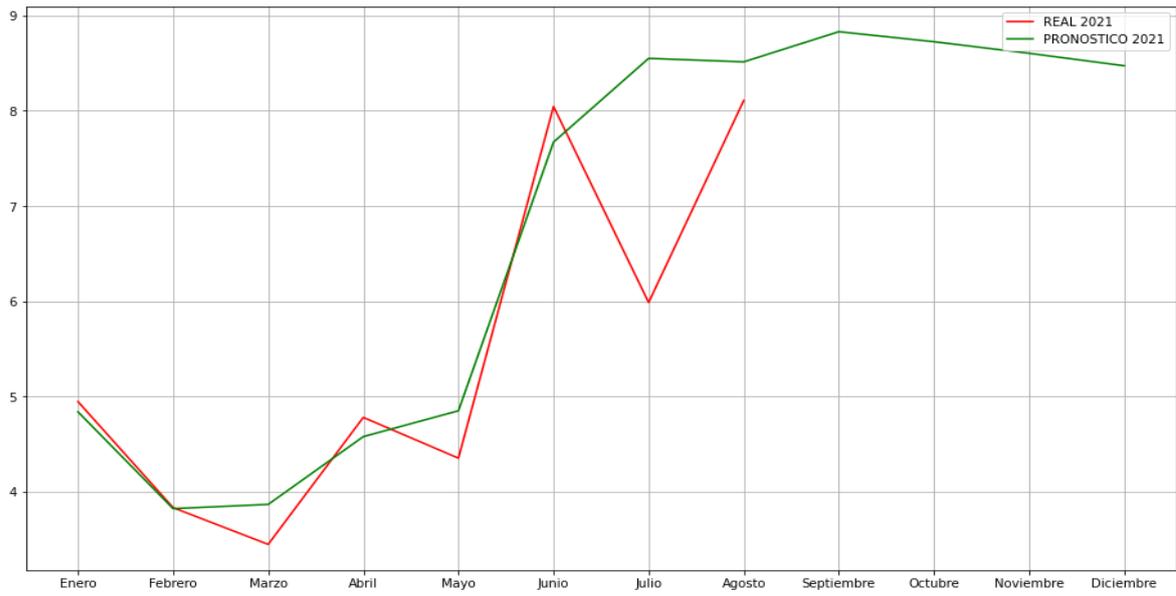


figura 31. Valores Reales (rojo) vs Valores Predichos (verde) de TMR para empresa XYZ.

CAPÍTULO V – CONCLUSIONES Y LÍNEAS FUTURAS.

5.1 CONCLUSIONES.

- Actualmente en El Salvador, las empresas distribuidoras hacen estos análisis actualizando diversas hojas de cálculo, Con el programa propuesto, el valor de los indicadores puede ser calculado en cuestión de segundos, la utilización de esta herramienta sin duda mejorara los procesos.
- A pesar de que el porcentaje de error entre los valores reales y proyectados para 2021 de los indicadores SAIFI y SAIDI fue inferior al 50% para la empresa ABC, estas estimaciones pueden ser utilizadas para el establecimiento de metas, con el objetivo de medir el desempeño de estos indicadores.
- Tanto los indicadores TMR y TDF de la empresa ABC presentaron porcentajes de error inferiores al 20% por tanto podemos dar las como aceptables de los comportamientos reales para el año 2021 y podrían ser utilizados como referencias.
- Los valores de TMR y TDF se adaptan mejor al modelo, presentan valores de error inferiores a los de SAIFI y SAIDI estos indicadores dependen de los clientes, debido a el tipo de datos es más difícil predecir cuantos clientes se afectarán, pero las cantidades de falla y los tiempos de atención son datos puntuales esto favorece a las predicciones de estos indicadores.
- Con el fin de obtener un pronóstico con un nivel de confianza superior al obtenido y con base en los trabajos relacionados, se recomienda empezar a almacenar datos históricos sobre la topología eléctrica, elementos de maniobra, ubicación, si son tele controlados, cronogramas de poda y manteamientos.
- Para el cálculo de la empresa XYZ, no fue utilizada la red neuronal entrenada por los cálculos previos de la empresa ABC, esto con el fin de evitar inferir en el cálculo ya que se trata de una base de datos total mente distinta y lo que se buscaba era validar el uso de la red neuronal.
- La red Para la empresa XYZ los errores en la proyección fueron SAIFI de 15% SAIDI de 24% vemos una disminución en el error del SAIFI en comparación al 29% de la empresa ABC, pensar que son comparables entre empresas es equivocado debido a que el error calculado solo nos dice que tan lejos estamos de cada dato y al ser bases de datos distinta esto se vuelve incomparable.
- Se logra constatar el cumplimiento del objetivo general de este trabajo que era desarrollar una herramienta capaz de calcular y proyectar el valor de los indicadores SAIFI, SAIDI, TDF y TMR. Para ambas empresas los errores de proyección de los datos fueron inferiores al 50%.

- La red neuronal fue pensada para proyectar datos de una sola empresa a la vez, debido a que cada empresa representa una base de datos.

5.2 LÍNEAS FUTURAS.

- Agregar variables como la cantidad de recurso de atención de fallas y las proyecciones climáticas permitirían un análisis y proyección de los indicadores más certeras.
- Implementar una interfaz gráfica de consultas de tal forma que permita a los usuarios consultar de manera ágil el indicador calculado para un mes en específico.

BIBLIOGRAFIA

- [1] SIGET, «Norma de calidad del servicio de los sistemas de distribución» 2014.
- [2] ESERP Business & Law school, «¿que son los KPIs ?, definición y ejemplos» 2021.
https://es.eserp.com/articulos/que-son-los-kpis/?_adin=02021864894
- [3] CEPAL, «Indicadores de desempeño en el sector público,» 2005.
- [4] SIGET, «ACUERDO N°86-E-2002» 2002.
- [5] SIGET. «Anexo B Metodología para el Control de la Calidad del Servicio Técnico de la SIGET»2014.
- [6] AES, «Calculo de indicadores TMR-aes El Salvador» 2016.
- [7] M. E. S. L. J. J. Morán Ortiz, «Análisis de las tarifas eléctricas en el salvador y su comparación con los países centroamericanos» Universidad de El Salvador facultad de ingeniería y arquitectura escuela de ingeniería eléctrica, 2021, p. 21.
- [8] AES, «Memoria de Sostenibilidad» AES-El SALVADOR, 2021.
<https://www.aeselsalvador.com/sites/default/files/202107/AES%20EI%20Salvador%20Memoria%20de%20Sostenibilidad%202020%20%20Sustainability%20Report.pdf>
- [9] DELSUR, «Reporte de Sostenibilidad» 2019.
<https://www.delsur.com.sv/wp-content/uploads/2020/03/REPORTE-DE-SOSTENIBILIDADELSUR-2019.pdf>
- [10] SIGET, «Boletín de estadísticas eléctricas número 22» 2021.
- [11] CEPAL, «Desarrollo de indicadores de pobreza energética en América» 2021.
https://repositorio.cepal.org/bitstream/handle/11362/47216/4/S2100433_es.pdf
- [12] SIGET, «Mercado Eléctrico de El Salvador 2020» 2020.
- [13] J. J. C. A. Levy, «Calidad y confiabilidad de los servicios eléctricos en América Latina» Banco Interamericano de Desarrollo, 2020.
<https://publications.iadb.org/publications/spanish/document/Calidad-y-confiabilidad-de-losservicios-electricos-en-America-Latina.pdf>
- [14] F. E. y. H. M. W. Ian, «Data mining Practical Machine Learning Tools and Techniques,» Burlington, 2011.
- [15] M. K. y. J. P. J. Han, « Data mining Concepts and Techniques» Waltham, 2012.
- [16] [2] C. M. Corredor Reyes, MODELO DE MINERÍA DE DATOS PREDICTIVA PARA EL PRONÓSTICO DE INDICADORES DE CALIDAD DE SUMINISTRO SAIDI Y SAIFI, PONTIFICIA UNIVERSIDAD JAVERIANA, 2017.
- [17] P. C. B. y. N. R. P. G. Shmueli, Data Mining for Business Analytics: Concepts, Techniques, and Applications with XLMiner,, John Wiley & Sons, Inc., 2016
- [18] D. T. L. y. C. D. Larose, Data Mining and Predictive Analytics, New Jersey: ohn Wiley & Sons, Inc, , 2015.
- [19] J. C. R. K. T. K. T. R. C. S. P. Chapman, CRISP-DM 1.0 Step-by-step data mining guide, SPSS Inc, 2000.

ANEXOS.

1 CÓDIGOS UTILIZADOS.

Los códigos en Python utilizados en este trabajo de investigación se encuentran disponibles en el siguiente enlace <https://bit.ly/3EEOMdc>

2 RESUMEN DE INDICADORES.

INDICADORES SAIFI-SAIDI-TDF-TMR "EMPRESA ABC"				
FECHA	SAIFI (INTERRUPCIONES)	SAIDI (HORAS)	TDF (FALLAS/100 KM)	TMR (HORAS)
ene-17	0.2362	0.5506	9.37	2.46
feb-17	0.2437	0.3651	8.01	1.69
mar-17	0.334	0.4627	9.69	1.67
abr-17	0.4401	0.758	13.27	1.79
may-17	0.436	0.7527	14.98	2.22
jun-17	0.5653	0.8761	14.63	1.94
jul-17	0.603	1.2599	19.06	2.62
ago-17	0.434	0.9586	18.13	2.43
sep-17	0.5192	1.2394	17.74	3.08
oct-17	0.3966	1.2604	16.45	4.37
nov-17	0.262	0.5775	11.57	2.53
dic-17	0.2481	0.6341	9.91	3.06
ene-18	0.2156	0.4499	8.59	2.43
feb-18	0.2316	0.4884	8.01	2.53
mar-18	0.2249	0.4209	8.69	2.27
abr-18	0.3763	0.9768	14.55	3.17
may-18	0.4665	0.7872	14.21	2.83
jun-18	0.3429	0.6711	14.10	2.26
jul-18	0.4868	1.1301	17.02	2.84
ago-18	0.736	1.8387	19.64	3.36
sep-18	0.5106	1.4341	16.96	3.73
oct-18	0.3047	0.8006	13.68	3.63
nov-18	0.2178	0.5772	11.70	2.9
dic-18	0.1653	0.3723	9.35	2.5
ene-19	0.2186	0.3607	9.14	2.13
feb-19	0.2057	0.3923	8.05	2.48
mar-19	0.1896	1.4424	9.46	2.39

abr-19	0.4541	1.0709	13.50	2.59
may-19	0.3341	0.8041	18.79	3.24
jun-19	0.3966	1.0211	17.23	2.77
jul-19	0.5047	1.2513	20.92	2.9
ago-19	0.4418	1.2552	21.58	3.67
sep-19	0.5418	1.3069	20.54	3.65
oct-19	0.391	0.9077	16.91	3.17
nov-19	0.2547	0.5711	13.59	2.73
dic-19	0.2286	0.4937	11.69	2.68
ene-20	0.2034	0.421	10.74	2.56
feb-20	0.2187	0.4447	9.08	2.46
mar-20	0.3625	0.6025	13.29	2.4
abr-20	0.4387	0.7807	16.30	2.16
may-20	0.3285	0.7626	17.38	2.42
jun-20	0.4968	1.4891	15.87	4.18
jul-20	0.363	1.0545	17.96	4.58
ago-20	0.3853	1.2392	19.32	4.94
sep-20	0.3056	1.1196	14.61	4.94
oct-20	0.3826	1.047	14.98	4.06
nov-20	0.2081	0.6208	11.35	5.02
dic-20	0.1252	0.4088	9.03	5.51
ene-21	0.1161	0.2376	6.79	3.24
feb-21	0.1223	0.2127	6.22	2.39
mar-21	0.223	0.3508	10.33	2.37
abr-21	0.3193	0.7812	13.89	4.1
may-21	0.4011	1.1531	15.89	5.63
jun-21	0.1959	0.7428	15.95	5.23
jul-21	0.4092	0.9076	17.53	3.89
ago-21	0.2804	1.1923	18.08	5.44

Tabla 25. INDICADORES SAIFI-SAIDI-TDF-TMR "EMPRESA ABC

INDICADORES SAIFI-SAIDI-TDF-TMR "EMPRESA XYZ"				
FECHA	SAIFI (INTERRUPCIONES)	SAIDI (HORAS)	TDF (FALLAS/100 KM)	TMR (HORAS)
ene-17	0.3831	0.8139	6.80	2.66
feb-17	0.156	0.3328	4.57	2.17
mar-17	0.4028	0.7273	7.23	2.85
abr-17	0.374	0.7784	9.02	2.61
may-17	0.5322	1.8419	16.74	3.86
jun-17	0.5054	1.9515	14.36	3.91
jul-17	0.7282	3.7716	17.67	5.15
ago-17	0.6439	3.7021	22.44	7.2
sep-17	0.9062	3.9207	19.44	7.01
oct-17	0.7854	4.403	18.14	8.91
nov-17	0.3022	1.1627	11.74	4.1
dic-17	0.3042	1.0677	9.70	5.08
ene-18	0.2526	1.0327	8.19	3.91
feb-18	0.204	0.491	6.39	2.96
mar-18	0.2424	0.4342	6.83	2.41
abr-18	0.4484	1.3033	13.21	3.4
may-18	0.5712	1.8131	17.91	4.01
jun-18	0.3067	1.1937	13.38	3.18
jul-18	0.3389	1.6019	14.64	3.91
ago-18	0.7145	3.4202	19.82	6.06
sep-18	0.607	4.4128	16.39	8.07
oct-18	0.4142	1.4113	12.61	5.05
nov-18	0.2105	0.7907	10.64	4.14
dic-18	0.2169	0.7053	9.29	4.01
ene-19	0.1901	0.5186	7.09	3.07
feb-19	0.1899	0.3542	5.17	2.34
mar-19	0.2506	0.6317	7.54	2.62
abr-19	0.3378	0.7273	9.82	2.62
may-19	0.4722	1.9953	16.12	5.17
jun-19	0.4521	1.8386	13.06	4.84
jul-19	0.6689	3.104	17.26	6.52
ago-19	0.6518	2.846	17.74	7.19
sep-19	0.7743	3.6633	21.29	7.44
oct-19	0.5381	2.1987	16.11	5.13
nov-19	0.2879	0.82	11.24	3.85
dic-19	0.2293	0.9212	9.64	4.25
ene-20	0.195	0.5562	7.48	3.45

feb-20	0.1422	0.3912	5.39	2.76
mar-20	0.2654	0.5268	7.85	2.4
abr-20	0.2478	0.5924	10.33	2.17
may-20	0.5787	2.3013	17.84	4.09
jun-20	0.4224	1.6499	13.81	4.42
jul-20	0.6617	2.8902	18.04	6.97
ago-20	0.7154	4.5835	21.50	8.32
sep-20	0.6884	3.2099	19.68	6.52
oct-20	0.4394	3.0876	16.69	8.99
nov-20	0.3046	1.4674	12.34	6.13
dic-20	0.2887	1.2193	10.84	5.6
ene-21	0.1951	0.7947	8.09	4.94
feb-21	0.1604	0.5577	6.37	3.83
mar-21	0.2195	0.7274	8.07	3.44
abr-21	0.4501	1.4327	12.18	4.78
may-21	0.3876	1.459	13.81	4.35
jun-21	0.5634	3.1528	18.03	8.04
jul-21	0.5294	2.0497	17.31	5.99
ago-21	0.6352	3.4494	22.67	8.11

Tabla 26. INDICADORES SAIFI-SAIDI-TDF-TMR "EMPRESA XYZ"

3 MINERÍA DE DATOS.

En la actualidad no existe un consenso claro respecto al significado de la minería de datos y su relación con términos que el entorno ha introducido a lo largo de la historia como analítica o modelamiento predictivo, aprendizaje automático y ciencia de los datos [15]. De acuerdo con [15], la minería de datos hace referencia al conjunto de métodos de análisis de negocio que van más allá de las técnicas descriptivas, estadística básica, informes y reglas de negocio. Mientras que en [16], es el proceso de descubrir patrones y tendencias útiles en grandes conjuntos de datos, para hacer estimaciones hacia el futuro. El uso de la minería ha sido impulsado por la era de los grandes volúmenes de datos, debido a la necesidad de extraer información de forma rápida y automática [15].

3.1 CRISP – DM.

Es un marco de referencia desarrollado en 1996, que define el ciclo de vida para proyectos de minería de datos. La Figura 30 muestra las fases del modelo de referencia CRISP–DM.

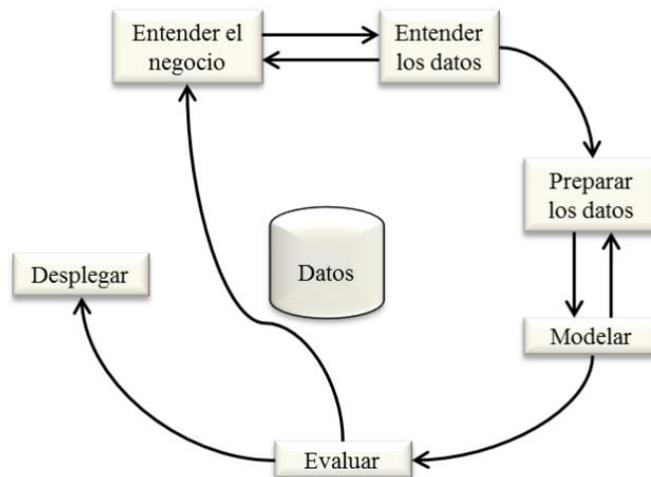


figura 32. Fases del modelo de referencia CRISP–DM [17]

La metodología se divide en 6 fases, que no necesariamente deben ejecutarse de forma consecutiva y que además se desarrollan de forma iterativa con el objetivo de mejorar los resultados [17]. CRISP-DM es empleado como guía para la definición de la metodología empleada en el presente proyecto.

3.1.1 ENTENDER EL NEGOCIO.

No solo es la primera fase del modelo, sino que además es la más importante, porque se enfoca en el entendimiento del propósito y los objetivos del proyecto [17]. En esta fase se establecen además los criterios que determinan el éxito del proyecto acorde con los lineamientos establecidos por el negocio y se genera el plan de proyecto.

3.1.2 ENTENDER LOS DATOS.

Esta fase comienza con la recolección de datos, seguido de su exploración para poder determinar la cantidad y calidad de la información, su completitud, entre otros. El acercamiento a los datos permite tener un conocimiento más claro de ellos y su relevancia

respecto al problema que se pretende resolver, no solo para detectar posibles problemas o necesidad de información adicional, sino que además permiten generar hipótesis iniciales [14].

3.1.3 PREPARAR LOS DATOS.

Acorde con los resultados obtenidos de la fase anterior, el objetivo en esta es construir el set o conjunto de datos que se utilizan posteriormente para crear los modelos de minería. La preparación de datos en sí misma incluye un conjunto de actividades que pueden realizarse más de una vez para garantizar que el conjunto de datos generado sea el que se requiere para el modelo [14].

3.1.4 MODELAR.

En esta fase se aplican las técnicas de minería necesarias para dar respuesta a la problemática. Cada técnica tiene que los datos sean ingresados en un formato específico para su correcto funcionamiento. Además, se deben configurar y calibrar los parámetros del algoritmo seleccionado para cada técnica específica [14].

3.1.5 EVALUAR.

En esta fase es importante garantizar que el modelo construido cumple con los objetivos y requerimientos del negocio, antes de proceder a realizar el despliegue del mismo. En esta fase puede llegar a detectarse que hay cuestiones de negocio que no han sido consideradas en su totalidad, por lo cual debe tomarse una decisión respecto a los resultados del modelo [14].

3.1.6 DESPLEGAR.

A pesar de ser la última fase del modelo de referencia, no significa que el ciclo de vida del proyecto de minería finalice en este punto. Puede que la empresa requiera que el modelo sea ejecutado periódicamente y se vaya ajustando a medida que los requerimientos del negocio y el mercado van cambiando, para garantizar que el modelo sea evolutivo. La fase de despliegue puede significar generar un documento o generar un proceso repetible [14].

3.2 TÉCNICAS DE MINERÍA.

3.2.1 DESCRIPTIVA.

El objetivo de las técnicas de minería no supervisada es descubrir patrones o relaciones entre el conjunto de variables suministradas en el set de datos. A través de estas técnicas es posible realizar una exploración del conjunto de datos y describir un comportamiento presente en los mismos, no orientado a la predicción de valores futuros de una variable objetivo [14].

3.2.2 PREDICTIVA.

Los modelos predictivos o minería supervisada se aplican cuando existe una variable objetivo y un conjunto de atributos descriptivos o informativos que permiten reducir el nivel de incertidumbre y por tanto permiten estimar un valor previamente desconocido para la variable objetivo.

Los valores de la variable objetivo se conocen como clases, y deben estar claramente identificadas, sin sobrellavarse entre sí, mientras que los atributos informativos deben estar correlacionadas con la variable a predecir.

Las técnicas predictivas se basan en el concepto de inducción y deducción. Para ello, se emplea un subconjunto del set de datos disponible, denominado set de entrenamiento, con el cual los modelos pueden generar reglas inductivas que generalizan las particularidades encontradas en los datos respecto la variable objetivo, por lo cual el conjunto de datos debe suministrar la clase para cada registro. Otra porción de los datos, diferentes al set de entrenamiento, se utiliza para realizar las pruebas del modelo. Este conjunto de datos no debe contener los valores de la variable a estimar, ya que el modelo toma las reglas generadas y a partir de ellas deduce o predice el valor de la variable objetivo. Finalmente, se evalúa el modelo con la porción de datos del set inicial [14].

A continuación, se mencionan dos técnicas de minería predictiva.

3.2.2.1 SERIES DE TIEMPO.

En las series de tiempo cada instancia representa un momento en el tiempo y los valores asociados con ese tiempo, generalmente las instancias se encuentran separadas por un intervalo regular de tiempo [18].

Las series de tiempo son útiles para modelar diferentes procesos naturales y económicos como con comportamiento estocástico: comportamientos del mercado de valores, comportamientos poblacionales, observaciones científicas, entre otros [19].

Generalmente las series de tiempo se usan para realizar pronósticos de regresión, y para tal fin se descompone la serie algunos de los siguientes elementos:

- **Tendencia:** Esta indica la dirección en la que la serie de ha ido moviendo en un periodo de tiempo determinado.
- **Estacionalidad:** Son patrones casi idénticos que aparecen con cierta regularidad dentro de la serie de tiempo, generalmente con duración similar en los diferentes periodos.
- **Comportamiento cíclico:** Son patrones que se presentan en la serie de tiempo, pero sin un comportamiento enmarcado dentro de un periodo fijo, es similar a la estacionalidad solo que puede comprender periodos de tiempo más extensos.
- **Comportamiento aleatorio:** Estos representan los cambios esporádicos que tienen un comportamiento estocástico y que son causados por elementos externos a los atributos que componen la serie de tiempo.

3.2.2.1 REGRESIÓN LINEAL.

Es un modelo matemático que busca obtener la relación de dependencia entre dos o más variables, donde una es la variable dependiente y las demás variables independientes.

La regresión lineal proporciona una descripción adecuada de como los insumos (variables independientes) afectan la salida (variable dependiente). Son usados para

realizar pronósticos donde se presente una relación de linealidad entre las variables, es necesario determinar el coeficiente de la correlación [14].

4 ANÁLISIS POR DE SERIE DE TIEMPO Y REGRESIÓN LINEAL XYZ.

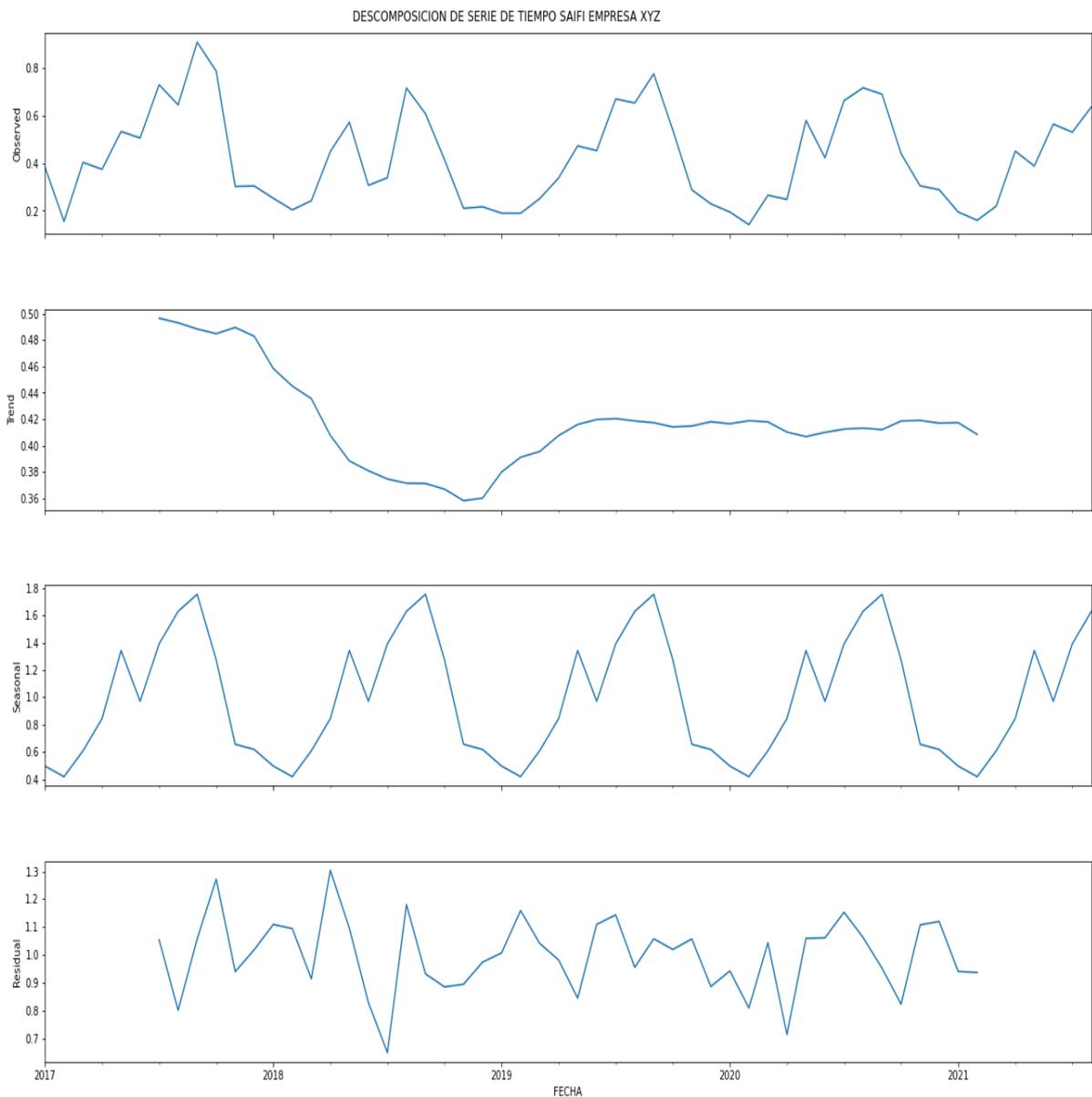


figura 33. Descomposición Serie de Tiempo SAIFI XYZ

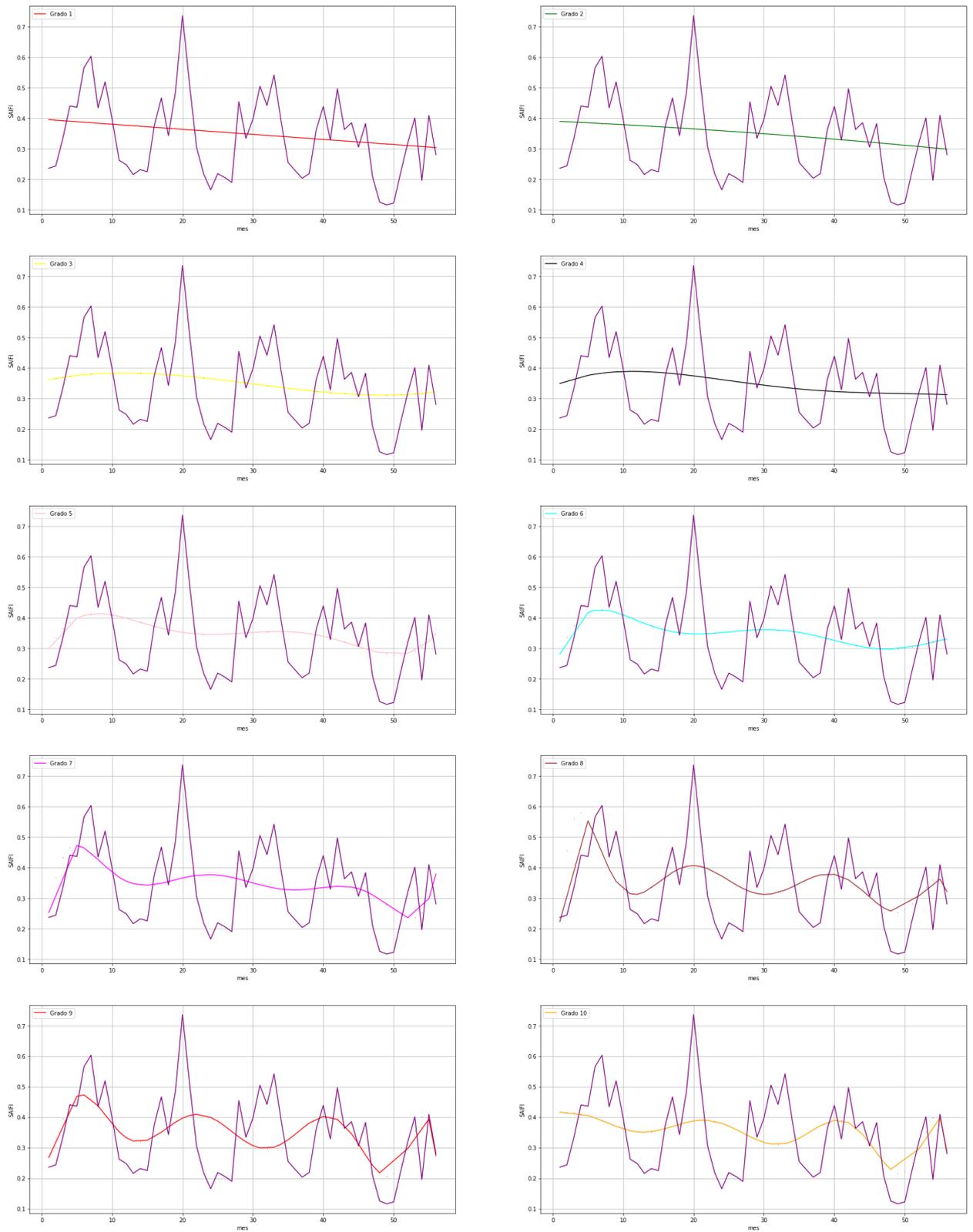


figura 34. Regresión polinómica desde grado 1 a grado 10 SAIFI empresa XYZ.