

Estimación de los parámetros de calidad de las variaciones de tensión de corta duración en sistemas de distribución aplicando técnicas de minería de datos.

Andrés Felipe Manrique Barbery
Luis Fabian Monroy Pedraza

Trabajo de Grado para Optar el título de Ingeniero Eléctrico

Director

Ph.D Gabriel Ordoñez Plata
Ph.D en Ingeniería Industrial

Codirector

Ph.D Jairo Blanco Solano
Ph.D en Ingeniería (Ing. Eléctrica, Electrónica & Desarrollo)

Universidad Industrial de Santander

Facultad de Físico- mecánicas

Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones

Bucaramanga

2019

Dedicatoria

A Dios por haber sido mi guía a lo largo de mi carrera y darme la oportunidad de llegar a ser un gran profesional.

A mis padres Amanda Barbery y Carlos Manrique, quienes con su dedicación y amor me formaron como ser humano, resaltando su esfuerzo por brindarme las oportunidades para cumplir mi sueño. Gracias por su apoyo y amor incondicional en todo momento, son y serán mi motivo para salir adelante.

A mis hermanos por su amor y apoyo incondicional siempre.

A Andrea, quien con su amor hizo de mí una mejor persona, apoyándome en cada situación difícil y dándome la motivación para seguir adelante.

A mis compañeros y amigos que aportaron en mi desarrollo personal y profesional a lo largo de toda esta etapa, estoy agradecido por la oportunidad de haber podido conocer a cada uno de ellos.

Gracias infinitas.

Andrés Felipe Manrique Barbery

Dedicatoria

A ti madre querida. porque aunque partiste, desde la distancia me cuidas, me hace cada día más fuerte y me proteges en cada paso que doy. “Todo lo puedo por esa fuerza que sigue viva y que trasciende en el infinito”.

A mi padre Israel Monroy, por su apoyo incondicional, por guiarme hacer lo correcto y enseñarme el valor de las cosas.

A mi nona Elvira Malagón por ese amor incondicional que a diario me brinda.

A Lidia Buitrago por acogerme con mucho amor y cariño como parte esa hermosa familia que ha construido.

A mi hermano Diego, por sus ocurrencias y ese amor que transmite cada vez que nos vemos.

A mis tíos José, Benilda, Clemencia, Tránsito, Jesús y Segundo por su cariño.

A mi tío Felix y a mi primo Lucho por interceder por mí en momentos difíciles de mi vida.

A todos mis primos.

A la señora Ana por el apoyo brindado y por haber hecho de mi padrino un ser maravilloso en sus últimos años.

A María Alejandra por ser mi compañía en este proceso.

A mis amigos y compañeros que me acompañaron en mi proceso de formación profesional.

Luis Fabian Monroy Pedraza

Tabla de Contenido

Introducción	11
1. Objetivos	14
1.1 Objetivo General	14
1.2 Objetivos Específicos	14
2. Minería de datos	15
2.1. Clasificación de las técnicas de minería de datos	16
2.2. Minería de datos en la calidad de la potencia eléctrica	17
3. Métodos para la estimación del índice sarfi ₉₀	22
3.1. Técnicas de minería de datos	23
3.1.1 Redes Neuronales (NN)	23
3.1.2 Máquinas de soporte vectorial (MSV)	24
3.1.3 Vecinos cercanos (KNN)	24
3.1.4 Árbol de decisión (DT)	25
3.2. Adquisición de los datos	25
3.3. Selección de datos	27
3.4. Transformación de datos	28
3.5. Aplicación de las técnicas de minería de datos	29
4. Resultados	31
4.1 Resultados ante las condiciones probabilísticas de falla iniciales	31
4.2. Resultados ante variaciones de las condiciones probabilísticas de falla	33
5. Conclusiones	37
Referencias	38

Lista de Tablas

Tabla 1. <i>Datos estadísticos usados para la simulación</i>	26
Tabla 2. <i>Contribución de cada tipo de falla en los diferentes conjuntos de entrenamiento.</i> ...	28
Tabla 3. <i>Tiempo en segundos para cada conjunto de entrenamiento</i>	30
Tabla 4. <i>Errores medios y desviaciones estándar de la estimación del índice SARFI₉₀</i>	32
Tabla 5. <i>Datos estadísticos usados para la nueva validación.</i>	34
Tabla 6. <i>Errores medios y desviaciones estándar de la estimación del índice SARFI₉₀ nuevo escenario.</i>	34

Resumen

Título: Estimación de los parámetros de calidad de las variaciones de tensión de corta duración en sistemas de distribución aplicando técnicas de minería de datos*

Autores: Andrés Felipe Manrique Barberly, Luis Fabian Monroy Pedraza**

Palabras clave: minería de datos, aprendizaje supervisado, calidad de la potencia, fallas de red, hundimientos de tensión, variaciones de tensión de corta duración.

Descripción:

En este trabajo se presenta una metodología para resolver el problema de estimación del índice SARFI₉₀ en un sistema de distribución a partir de las tensiones registradas en un número limitado de nodos monitorizados, aplicando cuatro técnicas de minería de datos: redes neuronales, máquina de soporte vectorial, vecinos cercanos, árboles de decisión en Matlab 2019^a. Las técnicas propuestas son implementadas en el sistema de distribución IEEE 123 Node Test Feeder donde se revisa el desempeño de cada técnica a través del error de estimación. Las técnicas son validadas considerando dos escenarios probabilísticos de un conjunto de fallas simuladas, los resultados demuestran la relevancia de las técnicas en la estimación del índice SARFI₉₀.

* Trabajo de Grado.

** Facultad de Ingeniería Físico-Mecánicas. Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones

Abstract

Title: Estimation of the quality parameters of short-term voltage variations in distribution systems applying data mining techniques*

Authors: Andrés Felipe Manrique Barbery, Luis Fabian Monroy Pedraza **

Keywords: data mining, supervised learning, power quality, network failures, voltage subsidence, short-term voltage variations.

Description:

This research work presents a methodology to solve the problem of estimating the SARFI₉₀ index in a distribution system based on the tensions registered in a limited number of monitored nodes, applying four data mining techniques: neural networks, vector support machine, nearby neighbors, decision trees in MATLAB 2019^a. The proposed techniques are implemented in the IEEE 123 Node Test Feeder distribution system where the performance of each technique is reviewed through the estimation error. The techniques are validated considering two probabilistic scenarios of a set of simulated faults, the results demonstrate the relevance of the techniques in the estimation of the SARFI₉₀ index.

* Bachelor Thesis.

** Facultad de Ingeniería Físico-Mecánicas. Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones

Introducción

Con la evolución de los sistemas de distribución y dada la influencia que tienen las cargas en los sistemas eléctricos, la calidad de la potencia eléctrica (CPE) ha tomado una mayor importancia en la última década. Un importante número de normas y aspectos regulatorios conllevan a que las empresas distribuidoras de energía eléctrica dispongan de sistemas de medición y registro eficiente para la detección de las perturbaciones de la CPE.

Dado a que existen una gran cantidad de fenómenos que intervienen en la CPE, el estándar IEEE 1159 (Committee, Power, and Society 2009) los agrupa según sus características en siete categorías diferentes, siendo la categoría de las variaciones de tensión de corta duración (VTCD) la de mayor relevancia debido a su presencia e impactos en el sistema eléctrico.

En un estudio realizado por el EPRI, a partir de más de 500 investigaciones, se determina a las VTCD como la causa principal de los problemas de CPE con un aporte del 48% (Framework and Power 2011).

Aunque las VTCD no son tan perjudiciales como las interrupciones de larga duración, al ocurrir con mayor frecuencia, son las principales contribuyentes a las pérdidas económicas sobre los usuarios afectados por los problemas de CPE (Framework and Power 2011). Esto se puede constatar en el estudio realizado en 2006 por *Leonardo Power Quality Initiative*, proyecto de relevancia para establecer que las VTCD contribuyen en aproximadamente el 60% de las pérdidas económicas totales de los problemas de CPE en los sectores industriales (Targosz and Chapman 2015).

En su preocupación por dar cumplimiento a los aspectos regulatorios, las empresas se han visto obligadas a la instalación de dispositivos de monitorización de la calidad que también permiten la recolección de gran información de utilidad para la identificación y diagnóstico de problemas en la CPE, la predicción de anomalías o fallas y la alarma de situaciones críticas en el sistema (Lin et al. 2016b); no obstante, una limitante es el elevado costo de la implementación masiva de estos instrumentos por parte de las empresas para, llevar a cabo un control de las VTCD. Esto ha conllevado al desarrollo de diferentes métodos para determinar la cantidad mínima de equipos de medición y su ubicación en el sistema, de manera que se puedan observar los hundimientos de tensión en todo el sistema. Una revisión de algunos métodos, se puede encontrar en (Mohamed, Shareef, and Zayandehroodi 2013).

La reducción de medidores en el sistema ha obligado a la creación de diferentes metodologías, para lograr conocer las VTCD en los nodos en donde no hay medidores instalados. La información almacenada por estos medidores tiene valor solo cuando es comprensible y esto se logra si se combina el ingenio humano con el software (Business Software Alliance 2015). Las empresas prestadoras del servicio cuentan con un sistema de monitorización, el cual permite la recolección de datos que diariamente están en aumento; muchas veces utilizan esta información almacenada únicamente para la generación de reportes ante los organismos encargados de vigilar los índices de calidad del servicio de energía eléctrica, olvidando el valor agregado que se puede generar al aplicar herramientas tecnológicas como Big data, minería de datos entre otras; las cuales permiten ser un apoyo en la toma de decisiones en tiempo real, generar patrones y predicciones con el fin de crear ventajas, mejorar procesos, ser más confiables y competitivos en el mercado.

Es así como en esta investigación se implementan unas técnicas de minería de datos, para estimar el índice SARFI₉₀ en los nodos no monitorizados del sistema de distribución IEEE 123 Node Test Feeder que no cuenta con una monitorización en todos sus nodos.

El documento está organizado de la siguiente manera: En la Sección 2 se describe la clasificación de las técnicas de minería de datos y algunas aplicaciones en la CPE. En la Sección 3 se presenta una descripción de las técnicas propuestas de minería de datos y su aplicación al sistema de distribución IEEE 123 Node Test Feeder. En la Sección 4 se presenta la validación de las técnicas propuestas y finalmente, en la Sección 5 se presentan las conclusiones del trabajo.

1.Objetivos

1.1 Objetivo General

Implementar un método basado en minería de datos para el diagnóstico de las variaciones de tensión de corta duración en media tensión mediante la cuantificación de los índices de la calidad de la potencia asociados a esta clase de perturbaciones.

1.2 Objetivos Específicos

El cumplimiento del objetivo general del trabajo de grado requiere los siguientes objetivos específicos:

- Realizar una revisión bibliográfica del estado del arte en relación con el estudio de la calidad de la potencia eléctrica en sistemas de distribución aplicando técnicas de minería de datos.
- Implementar un método para la estimación de índices de las variaciones de tensión de corta duración en un sistema de distribución (IEEE 123 nodos) con base en una técnica de minería de datos seleccionada aplicando criterios de precisión, velocidad, manejo de incertidumbres y manejo de error.
- Validar el método de estimación de las variaciones de tensión mediante el estudio por simulación del sistema de prueba IEEE 123-nodes, aplicando análisis estadísticos para la obtención de las bases de datos de entrenamiento y generando escenarios aleatorios de prueba para la comparación de resultados.

2. Minería de datos

La minería de datos (DM, por sus siglas en inglés) hace referencia al proceso de descubrir patrones y conocimientos interesantes a partir de grandes cantidades de datos.

La minería de datos se puede entender como una parte del proceso del descubrimiento de conocimiento en bases de datos. En la Figura 1 se puede apreciar la representación esquemática de este, el cual consiste en una secuencia iterativa de pasos (Han, Jiawei; Kamber, Micheline; Pei 2012).

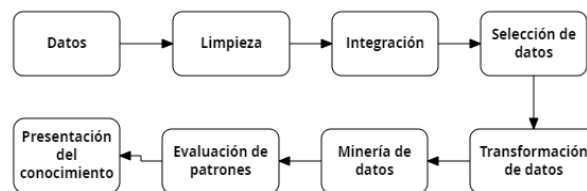


Figura 1. Pasos del descubrimiento del conocimiento.

Los pasos que seguir para el proceso de descubrimiento del conocimiento en bases de datos se enlistan a continuación:

Limpieza: Se realiza la eliminación del ruido, datos inconsistentes y faltantes.

Integración: Se combinan los datos que pueden tener procedencia de múltiples fuentes de datos.

Selección de datos: Los datos relevantes para la tarea de análisis se extraen de la base de datos.

Transformación de datos: Se transforman y consolidan en formas apropiadas para la minería realizando operaciones de resumen o agregación.

Minería de datos: Se aplican métodos inteligentes para extraer patrones de datos.

Evaluación de patrones: Se identifican los patrones que verdaderamente representan el conocimiento basado en medidas de interés.

Presentación del conocimiento: Se presenta el conocimiento minado a los usuarios.

Los datos son la parte central de la minería de datos. Cada conjunto de datos está formado por una serie de variables, las cuales se denominan atributos. Estos datos pueden ser de tipo *etiquetados*, si presentan un atributo especialmente designado, y de tipo *no etiquetados* en caso de no presentar un atributo especialmente designado (Bramer 2016).

2.1. Clasificación de las técnicas de minería de datos. Las técnicas de minería de datos se clasifican en dos grandes categorías: *supervisadas*, y *no supervisadas* (Molina, José;García 2006). La primera se aplica a los datos etiquetados y su objetivo es utilizar los datos dados para predecir el valor del atributo designado para nuevos datos. La segunda categoría se aplica a los datos no etiquetados y aquí el objetivo es extraer la mayor cantidad de información posible de los datos disponibles.

Dado que se desea estimar el índice SARFI en los nodos no monitorizados a partir de valores de las magnitudes de tensión residual en los nodos medidos, la investigación requiere el uso de técnicas de clasificación supervisada. Este tipo de técnicas tiene como finalidad, dado un conjunto de datos de entrenamiento previamente etiquetados, lograr determinar el valor de la

etiqueta (salida) para nuevos datos de entrada. Dependiendo del número de valores que pueda tomar este atributo, se distinguen entre clasificación *binaria*, si la salida solo puede pertenecer a una de dos clases posibles, y clasificación *multiclase*, si la salida puede pertenecer a un conjunto limitado de clases (Charte 2015).

Adicional a los dos tipos de clasificación mencionadas anteriormente, existe una tercera, conocida como clasificación *multietiqueta*, que se caracteriza porque el conjunto de datos está asociado a múltiples etiquetas (salidas) simultáneamente.

Existen diferentes métodos para la solución a los problemas de clasificación de etiquetas múltiples que pueden ser agrupados en dos categorías principales, los métodos de transformación de problemas y los métodos de adaptación de algoritmos. La primera categoría agrupa aquellos métodos que transforman el problema de múltiples etiquetas en uno o más problemas de etiqueta única, permitiendo el uso de cualquier algoritmo de clasificación de etiquetas única. Por otra parte, la segunda agrupa aquellos algoritmos para etiqueta única, adaptados para problemas de etiquetas múltiples (Tsoumakas and Katakis 2009).

2.2. Minería de datos en la calidad de la potencia eléctrica. El análisis de datos en la calidad de la energía se ha venido explorando desde finales de la década de los 90, en búsqueda de la identificación automática de las causas de diferentes perturbaciones de la CPE. Dado el incremento en los volúmenes de datos relacionados con la energía, han limitado la aplicación de métodos tradicionales de análisis de datos. Las técnicas de minería de datos han demostrado ser útiles en la industria del sector energético, dada la posibilidad de extraer patrones y tendencias ocultas generando así oportunidades en las empresas. Además permiten el modelado de los datos de energía eléctrica sin incurrir en ninguna pérdida de información como resultado de las

transformaciones de datos (Ghodsi 2016), sirviendo para el respaldo de las decisiones de evaluación de la seguridad de operación de los sistemas de energía eléctrica.

Cuando se habla de CPE imposible no pensar en las fallas de red y su impacto. En el ámbito de la localización de fallas, diferentes estudios han aplicado las técnicas de minería de datos. Un trabajo relevante (Hurtado-cortés, Villarreal-lópez, and Villarreal-lópez 2016) refleja la extensa literatura sobre propuestas metodológicas y aplicaciones de las técnicas de inteligencia artificial tales como; inferencia difusa y neurodifusa, redes neuronales y los sistemas inmunes, siendo organizados teniendo en cuenta las características de un sistema de detección y diagnóstico de fallas. Como resultado se tiene que las estrategias que no utilizan modelos son más eficaces en la detección de fallas. De igual manera describen algunas técnicas que descomponen el sistema total en distintos subsistemas más simples de analizar además identificar el origen de una falla favoreciendo la interpretación, pero que requieren de un costo computacional más alto.

Para las empresas es muy importante identificar las diferentes perturbaciones en el sistema para conocer las posibles causas, evaluarlas y tomar decisiones para mitigarlas. A continuación, se enuncian algunos trabajos que muestran la relevancia de las bases de datos y cómo a partir de diferentes estrategias de análisis se abordan los problemas.

En Bogotá se instalaron 290 medidores con el propósito de medir elevaciones, caídas, desbalances y, parpadeos; con esta información se realiza un análisis por medio de simulación en un sistema de distribución y transmisión (Romero, Gallego, and Pavas 2011), agrupando los diferentes tipos de hundimientos de tensión por medio del método *Cluster* y con el algoritmo K-medias se ubican las posibles zonas donde se presenta la mayor ocurrencia de fallas. Otro aspecto para resaltar es el uso de las redes neuronales artificiales profundas. *Seongmin He and*

Jay H. Lee (Heo et al. n.d.) abordan el problema de detección y clasificación de fallas aplicando diferentes restricciones, pesos o tasas de aprendizaje (Barrero 2019) (híper-parámetros) en el número de capas ocultas y el número de neuronas en la última capa oculta para generar diferentes patrones de datos. Esto se realiza con miras en examinar la capacidad de las redes neuronales profundas en la clasificación de fallas y posteriormente compararlas con otros métodos usados para el análisis de datos. Se debe resaltar la importancia del trabajo realizado por los analistas de los centros de control, ya que cuando se presenta una falla en el sistema eléctrico deben tomar decisiones inmediatas para restablecer el servicio y diagnosticar el funcionamiento de los equipos involucrados. A causa de esto, se propone (Calderón 2008) una herramienta para la clasificación automática de las fallas por medio de una red neuronal con aprendizaje por regularización bayesiana y finalización temprana. Empleando información de corrientes, tensiones de las tres fases medidas durante el evento encuentran un desempeño satisfactorio en cuanto al error de clasificación resultante en diferentes escenarios de entrenamiento y validación.

Los fenómenos de CPE se pueden clasificar en continuos y discretos; en un trabajo característico (Nourollah and Moallem 2011) han propuesto un método unificado para obtener dos índices globales de la calidad de la energía, mediante el algoritmo de análisis rápido de componentes (FAST-ICA).

Abordajes importantes son presentados en estudios posteriores (Blanco-Solano, Petit-Suárez, and Ordóñez-Plata 2015) y (Velasco and Arnedo 2008), en donde el análisis mediante el software ATP (Alternative Transients Program) permiten la integración de una serie de programas complementarios y aplicaciones específicas para la visualización de resultados (Antonio 1984), aplicando reglas de decisión se realizó un análisis a un sistema de

distribución IEEE 34 nodos para la ubicación relativa de caídas de tensión y la localización de las fuentes generadoras de hundimientos de tensión únicamente con medidas de tensión; también, utilizando el método de Monte Carlo y con el soporte de los programas informáticos EXCEL y ACCES se llevó a cabo un estudio de sensibilidad a un sistema de media y baja tensión, el cual permite al usuario consultar un nodo particular de la red, proceder con una estimación de los hundimientos de tensión y conocer características estadísticas y eléctricas de los dos niveles de tensión.

Con las técnicas de minería de datos se puede observar la influencia de diferentes factores en la calidad de la potencia y la identificación de patrones. A manera de ejemplo, en (Sikorski and Borkowski 2019) se hace una evaluación del impacto de la generación distribuida en los parámetros de calidad de la potencia mediante el método de *Cluster*, adicionalmente en China (Lin et al. 2016a), los autores integraron datos de monitorización de la calidad de la potencia, datos meteorológicos y datos de operación con el propósito de revisar la correlación entre los índices de calidad, la influencia de la temperatura, realizar un pronóstico e identificar patrones de calidad existentes para predecir la tendencia futura. Asimismo ocurre en (Asheibi, Stirling, and Robinson 2007), donde en un sistema de distribución de MT con diferentes tipos de carga se emplea el método de Cluster y algoritmo A Priori, con el fin de identificar patrones de las diferentes componentes armónicas en cada sitio, evaluar el efecto de cada componente armónica y determinar la fuente más significativa de distorsiones. En un nuevo escenario (Sikorski 2016), el método de Cluster asocia los datos de diferentes niveles de demanda de carga y plantas de cogeneración medidos en las bahías de una subestación; con el propósito de evaluar la influencia de los cambios de carga y las variaciones de las plantas en los parámetros CPE.

El tema de identificación de perturbaciones ha sido abordado mediante el uso de la transformada de Wavelet, que permite la extracción de características de las señales de perturbación y la incorporación del algoritmo de árbol de decisión C4.5 (Yinghui Kong, Jinsha Yuan 2007). Sin embargo, la transformada de Wavelet discreta y otros algoritmos basados en la lógica difusa, modelo oculto de Márkov no son lo suficientemente robustos como clasificadores. A causa de esto se presentan las máquinas de soporte vectorial y los árboles de decisión (Upadhyaya and Mohanty 2016) para la localización y detección de diferentes perturbaciones de la CPE.

En (Chen, Luo, and Shen 2018) se propone un método para extraer y filtrar datos de caídas de tensión, observar la severidad en el evento de hundimiento de tensión mediante un algoritmo A Priori mejorado. La función principal es reducir el volumen, acelerando más la operación y ocupando menos espacio para el tratamiento de los datos, en conjunto con la teoría del objetivo gris que establece un modelo de coincidencia entre el escenario de falla y la falla real; adicionalmente para la extracción y discretización de datos se utiliza un algoritmo adaptativo de transformación de la nube gaussiana encargado de la extracción y discretización de los datos. Mientras tanto, mediante el software WEKA (Kiranmai and Laxmi 2018) se evalúa el rendimiento de diferentes algoritmos de extracción de datos (árbol de decisión, árbol aleatorio y bosque aleatorio) en la clasificación de elevaciones de tensión, disminución de tensión, interrupciones y desequilibrios. Sintetizando, los algoritmos se validan en dos escenarios diferentes. El primero con tensiones trifásicas y el segundo con tensiones máximas, mínimas, promedio con el propósito de averiguar con cual es más eficiente la clasificación.

3. Métodos para la estimación del índice SARFI₉₀

Esta investigación tiene como finalidad para un sistema de distribución que no cuente con el monitoreo de la totalidad de los nodos, aplicar el uso de 4 técnicas de minería de datos para la estimación del SARFI por fase en los nodos no monitorizados del sistema, a partir de las tensiones registradas en los nodos monitorizados.

El sistema tomado como caso de estudio corresponde con el sistema de distribución IEEE 123 Node Test Feeder (Figura 3), operado bajo un nivel de tensión de 4,16 kV. Además, contenido por 118 líneas aéreas y subterráneas con configuración monofásica trifilar, bifásica trifilar, trifásica tetrafilar, 4 reguladores de tensión, 4 bancos de capacitores y 2 transformadores. Los datos del sistema se encuentran en (Baughman 1921).

El modelo de medición óptima utilizado en el sistema corresponde al propuesto en (Jairo 2018), dando lugar a la instalación de un número mínimo de 14 medidores dispuestos en los nodos (150, 8, 28, 40, 50, 57, 56, 61, 66, 79, 82, 91, 450, 300), los cuales corresponderán a los nodos monitorizados del sistema.

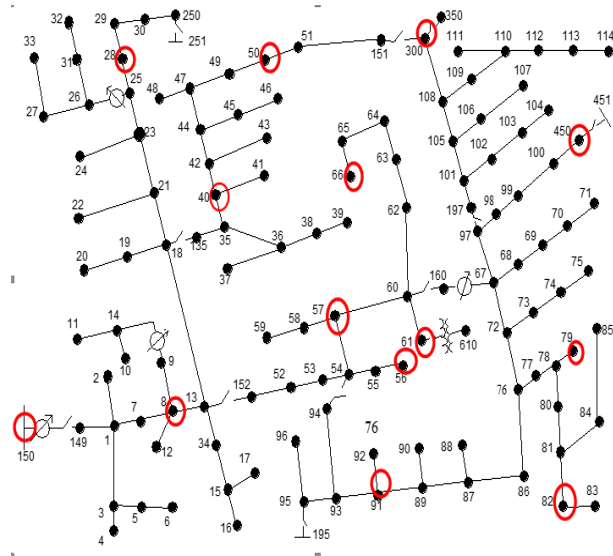


Figura 2. IEEE 123 Node Test Feeder.

3.1. Técnicas de minería de datos.

El software utilizado para la aplicación de las técnicas de minería de datos corresponde con Matlab 2019a, se incorporaron diferentes tipos de algoritmos. A continuación, se presenta una breve descripción de cada algoritmo seleccionado:

3.1.1 Redes Neuronales (NN). Las redes neuronales son sistemas de procesamiento de información inspirados en el funcionamiento del cerebro humano (González et al. 1995). Consiste en un conjunto de nodos o neuronas organizadas en varias capas. Cada neurona está interconectada con otras neuronas mediante un enlace de comunicación donde cada enlace tiene asociado un peso. El peso representa la información que utilizará la red para solucionar un problema. Las redes neuronales pertenecen al conjunto de algoritmos de aprendizaje supervisado los cuales aprenden mediante un entrenamiento con ejemplos. Dentro de las capas ocultas se

encuentran las neuronas que son funciones matemáticas lineales encargadas de procesar la información de la capa de entrada y entregarla a la capa de salida. Cuando se tiene una gran cantidad de neuronas se vuelve más complejo el conocimiento a elaborar. Por ello, se utilizan las funciones de activación que mediante funciones no lineales permiten encadenar de forma efectiva las neuronas y generar un patrón de salida que represente la información almacenada (Ruiz and Matich 2001).

3.1.2 Máquinas de soporte vectorial (SVM): Las máquinas de soporte vectorial son unas técnicas de clasificación lineal basados en la minimización del riesgo estructural (SRM) (J. 2014), funcionan discriminando entre miembros positivos y negativos de una clase de vectores por medio de un hiperplano de separación que equidista de los ejemplos más cercanos de cada clase a partir de las funciones matemáticas llamadas Kernel encargadas de convertir un problema de clasificación no-lineal, a un sencillo problema de clasificación lineal.

3.1.3 Vecinos cercanos (KNN). Es una de las técnicas de clasificación más conocidos que utiliza aprendizaje supervisado (Gonzalo 2018). A diferencia de las redes neuronales la técnica de vecinos cercanos no aprende, sino que memoriza los datos de entrenamiento mediante un vector, junto a una clase asociada a cada muestra de este conjunto (Telecomunicación, Cambronero, and Moreno n.d.). Luego, se calculan las distancias de cada dato de entrenamiento a las clases almacenadas en el vector a fin de buscar las k muestras más cercanas. Posteriormente ordena dichas distancias de menor a mayor para ir seleccionando la clase a la cual pertenece. En

la etapa de validación el algoritmo debe realizar una clasificación final con los datos de validación y buscar el valor de k más dominante.

3.1.4 Árbol de decisión (DT). El árbol de decisión se utiliza para estructurar el proceso de toma de decisiones, y su análisis se basa en la teoría de probabilidades (José 1993). Contiene nodos hoja que representan regiones clasificadas de acuerdo a una clase y nodos internos o splits que permiten decidir a qué subregión va cada elemento que llega a dicho nodo. El modelo subdivide el espacio de datos de entrada para generar regiones disjuntas, de tal manera que todos los elementos que pertenezcan a una región sean de la misma clase, la cual representa dicha región. Si una región contiene datos de diferentes clases es subdividida en regiones más pequeñas siguiendo el mismo criterio, hasta particionar todo el espacio de entrada en regiones disjuntas que solamente contengan elementos de una sola clase (Gironés Roig Jordi, Casas Roma Jordi, Minguillón Alfonso Julia 2017).

3.2. Adquisición de los datos. La adquisición de los datos para el desarrollo de este proyecto de investigación, correspondientes a las tensiones residuales en magnitud por fase en cada nodo del sistema, son el resultado del modelamiento realizado en la tesis doctoral (Jairo 2018) a través del software OpenDSS mediante la simulación de un gran conjunto de fallas generadas aleatoriamente aplicando el método de Monte Carlo, de acuerdo al escenario probabilístico descrito en la Tabla 1. Son simulados un total de 500 años equivalentes a 193973 fallas de red, esta cantidad de fallas cumple con un índice de confianza del 95% y un error inferior al 2,5% en los valores esperados del SARFI₉₀.

Tabla 1.
Datos estadísticos usados para la simulación

Prob.	Rf	Prob. falla (fallas/año)	Tipo de falla	Total fallas simuladas
65%	$\mu=1$ $\sigma=0,2$		LT 73 %	
25%	$\mu=5$ $\sigma=1$	Nodos 0,08	LL 6%	500 años 193 973 fallas
10%	$\mu=10$ $\sigma=2$	Líneas 3,7	LLT 17%	
			LLL/LLLT 4%	

Esta investigación al tener como propósito la estimación del SARFI₉₀ por fase en cada nodo no monitorizado, se decidió trabajar con tres (3) conjuntos de datos, uno por cada fase, con el fin de entrenar cada método o técnica de minería de datos de manera independiente para cada fase, permitiendo así un mejor resultado en la estimación. Por tal motivo las tensiones residuales fueron separadas en tres (3) conjuntos de datos o matrices, uno para cada fase, cada matriz estaba conformada por m (filas) y n (columnas). Las n columnas hacen referencia a los nodos del sistema, y las m filas corresponden a las fallas simuladas. Es preciso señalar que las m filas son iguales en cada fase, sin embargo, las n columnas no son iguales para cada fase, dada la aparición de diferentes configuraciones en las líneas del sistema seleccionado.

A partir de los conjuntos de datos de las tensiones residuales, se calculó el SARFI₉₀ en los nodos no monitorizados. Se crearon tres matrices (una por cada fase) donde se asignó el valor de

uno (1) para aquellos nodos donde ocurra una caída de tensión menor al 90% de su tensión nominal y cero (0) para aquellos donde no ocurra una caída de tensión. Este proceso se realizó para todas las tensiones residuales correspondientes a las fallas simuladas. De tal manera, que cada matriz de SARFI se formaba por m(filas) y n (columnas). Las n columnas indicaban los nodos no monitorizados del sistema, y las m filas corresponden a las fallas simuladas.

3.3. Selección de datos. Se seleccionaron los atributos de interés para el estudio, o dicho de otro modo las entradas y etiquetas para el aprendizaje de los modelos. Las tensiones residuales en magnitud por fase en los nodos monitorizados corresponden a las entradas de los modelos y el SARFI₉₀ en los nodos no monitorizados por cada fase, equivale a las etiquetas de los modelos.

Para la selección de los datos de entrenamiento se plantearon cuatro divisiones, con el objetivo de observar el comportamiento de cada algoritmo con diferentes cantidades de datos en el entrenamiento. De esta manera se optaron por cuatro conjuntos de entrenamiento correspondientes a 5, 10 ,15 y 20 años. En la Tabla 2 se observa con detalle cómo está conformado cada conjunto de entrenamiento.

Tabla 2.

Contribución de cada tipo de falla en los diferentes conjuntos de entrenamiento.

Tipo de falla	Años usados en entrenamiento			
	5 años	10 años	15 años	20 años
LT	81%	81%	81%	81%
LL	4%	4%	4%	4%
LLT	12%	12%	12%	12%
LLL/LLLT	3%	3%	3%	3%
Fallas totales	1891	3807	5748	7632

Como se observa en la Tabla 2, los diferentes conjuntos de entrenamiento tienen una distribución muy similar para los diferentes tipos de falla, esto es debido a que los datos fueron simulados de acuerdo con un escenario estadístico determinado.

3.4. Transformación de datos. Las herramientas de Matlab permiten que las redes neuronales puedan trabajar conjuntos de datos con múltiples etiquetas, siendo innecesario la transformación de los datos cuando se use esta técnica.

Por otra parte, el árbol de decisión, las máquinas de soporte vectorial y vecinos cercanos, no están adaptados para conjuntos de datos de múltiples etiquetas, por ello fue necesario transformar

cada conjunto de entrenamiento; se seleccionó el método de transformación *Label Powerset (LP)*, dada la simplicidad y ventaja que presenta en comparación con otros métodos de transformación al tener en cuenta las correlaciones entre etiquetas (Alazaidah 2016). Este método convierte el problema de etiqueta múltiple a uno de multiclase, tratando cada combinación de las etiquetas presentes en el conjunto de entrenamiento como una clase; de esta manera, el número de clases dependerá del número de combinaciones de etiquetas que aparezcan en los conjuntos de entrenamiento.

3.5. Aplicación de las técnicas de minería de datos. El Toolbox *nnstart* de Matlab se utiliza para el entrenamiento de la red neuronal, y se propone una red de dos capas: una capa oculta y una capa de salida. El número de neuronas en la capa de salida dependerá del número de nodos no monitorizados en cada fase. Para seleccionar el número de neuronas en la capa oculta se utilizaron diferentes cantidades (7, 10, 15 y 20 neuronas), presentándose el mejor resultado con 10 neuronas.

Después del entrenamiento, se obtuvo la red neuronal entrenada por fase, la cual es guardada como una función para su posterior validación.

Mediante la aplicación *Classification Learner* de Matlab se exploraron otros métodos de aprendizaje de clasificación supervisado DT, KNN, MSV, fueron usadas 5 capas de validación cruzada para prevenir el sobreajuste. Como se mencionó en la etapa de transformación los algoritmos arboles de decisión, máquinas de soporte vector y vecinos cercanos, fueron entrenados como un problema de etiqueta única con múltiples clases.

Cada método de transformación ha de contar también con una técnica que le permita construir los labelset resultantes a partir de las predicciones devueltas por los clasificadores tradicionales, generando así la salida final del clasificador (Charte 2015).

Cada conjunto de datos por fase se sometió al proceso de minería de datos para los diferentes algoritmos. La Tabla 3 recopila los diferentes tiempos de entrenamiento para cada algoritmo, los tiempos corresponden a la suma de cada entrenamiento para cada fase. Las características de los datos son un factor relevante en el tiempo de entrenamiento de las diferentes técnicas. Se puede evidenciar que, a mayor cantidad de datos, mayor será el tiempo de entrenamiento; adicionalmente, en DT, KNN, SVM donde los datos son de etiqueta única con múltiples clases, requieren menor capacidad de cómputo, haciendo que el tiempo de entrenamiento sea más rápido. Por otro lado, las NN al utilizar múltiples etiquetas exigirán mayor capacidad de computación haciendo más lento el entrenamiento.

Tabla 3.
Tiempo en segundos para cada conjunto de entrenamiento

Método	5	10	15	20
	años	años	años	años
DT	11,17	14,60	15,62	17,81
KNN	3,73	6,40	7,33	11,37
SVM	142,18	327,17	480,89	740,36
NN	2448	3996	7363,2	10800,6

Todas las tareas de cómputo se realizaron en un computador portátil ASUS K555L con un procesador Intel Core i7 (CPU 2,40 GHz 2 núcleos, 4 procesadores lógicos) y 8GB de RAM. Es preciso aclarar que estos tiempos varían conforme a la capacidad del equipo de computación con que se cuente.

4. Resultados.

Una vez entrenado cada modelo con los diferentes algoritmos, se validó su precisión a partir del conjunto de datos total de tensiones residuales simuladas bajo el escenario probabilístico descrito en la Tabla 1 (escenario usado en entrenamiento), correspondiente a 500 años con un total de 193973 fallas. Además, se presenta una segunda validación, para un escenario con variaciones de las condiciones probabilísticas de falla inicialmente usadas. A continuación, se presentan los resultados para los dos escenarios.

4.1 Resultados ante las condiciones probabilísticas de falla iniciales. Buscando enfocar la validación de cada modelo a una aplicación más práctica, se evaluó el índice SARFI90 año a año, para cada fase. Por ello en cada modelo ya entrenado se ingresaron las tensiones residuales por fase de los nodos monitorizados año a año, obteniendo el índice de SARFI90 estimado en los nodos no monitorizados. Esta estimación fue comparada con la respuesta real esperada, encontrando así el error de estimación en cada nodo por fase, estos errores fueron promediados para obtener un error por cada año.

La Tabla 4 recopila la media y la desviación estándar de los errores promedios de los 500 años para cada método por fase, de acuerdo a los años usados en el entrenamiento. En esta tabla se puede apreciar que cada método presenta una respuesta diferente y que el aumento en los datos de entrenamiento no necesariamente asegura una disminución en el error.

Los modelos empleados presentan un error medio aceptable, sin embargo, aquellos métodos que requirieron una transformación previa de los datos (DT, KNN y SVM) presentan un error mayor comparado al presentado por las NN, cuyo procesamiento no requirió transformación.

Este incremento de error se debe en gran medida a la pérdida de información presentada en la etapa de transformación, dado que el método usado solo permite estimar aquellas combinaciones que aparecieron en el conjunto de entrenamiento, por tal motivo aquellas combinaciones no vistas en el conjunto de entrenamiento no las podrá estimar. Se evidencia que las redes neuronales fueron las que tuvieron el menor error medio de estimación. Adicionalmente su baja desviación estándar permite observar que los errores anuales no están tan dispersos respecto del error medio, haciendo que la técnica sea más precisa.

Tabla 4.
Errores medios y desviaciones estándar de la estimación del índice SARFI₉₀

Métodos		5 años	10 años	15 años	20 años
		Error (%)	Error (%)	Error (%)	Error (%)
DT	A	$\mu=1,91$	$\mu=1,52$	$\mu=1,57$	$\mu=1,61$
		$\sigma=0,89$	$\sigma=0,80$	$\sigma=0,87$	$\sigma=0,83$
	B	$\mu=3,43$	$\mu=2,61$	$\mu=2,83$	$\mu=3,09$
		$\sigma=1,64$	$\sigma=1,39$	$\sigma=1,51$	$\sigma=1,60$
	C	$\mu=3,29$	$\mu=3,30$	$\mu=3,18$	$\mu=3,05$
		$\sigma=1,34$	$\sigma=1,36$	$\sigma=1,36$	$\sigma=1,28$
KNN	A	$\mu=2,02$	$\mu=1,79$	$\mu=1,74$	$\mu=1,71$
		$\sigma=0,87$	$\sigma=0,85$	$\sigma=0,86$	$\sigma=0,85$
	B	$\mu=2,77$	$\mu=2,48$	$\mu=2,68$	$\mu=2,78$
		$\sigma=1,41$	$\sigma=1,29$	$\sigma=1,40$	$\sigma=1,48$
	C	$\mu=3,11$	$\mu=3,22$	$\mu=3,19$	$\mu=3,17$
		$\sigma=1,28$	$\sigma=1,32$	$\sigma=1,33$	$\sigma=1,33$

Continuación Tabla 4.

Errores medios y desviaciones estándar de la estimación del índice SARFI₉₀

Métodos	5 años	10 años	15 años	20 años	
	Error (%)	Error (%)	Error (%)	Error (%)	
SVM	A	$\mu=2,35$	$\mu=2,05$	$\mu=1,87$	$\mu=1,86$
		$\sigma=0,98$	$\sigma=0,93$	$\sigma=0,88$	$\sigma=0,88$
	B	$\mu=3,06$	$\mu=2,56$	$\mu=2,62$	$\mu=2,53$
		$\sigma=1,51$	$\sigma=1,33$	$\sigma=1,36$	$\sigma=1,34$
	C	$\mu=4,02$	$\mu=3,68$	$\mu=3,75$	$\mu=3,70$
		$\sigma=1,53$	$\sigma=1,47$	$\sigma=1,48$	$\sigma=1,50$
NN	A	$\mu=0,94$	$\mu=0,99$	$\mu=0,79$	$\mu=0,83$
		$\sigma=0,29$	$\sigma=0,31$	$\sigma=0,25$	$\sigma=0,26$
	B	$\mu=0,93$	$\mu=0,92$	$\mu=0,58$	$\mu=0,86$
		$\sigma=0,35$	$\sigma=0,35$	$\sigma=0,26$	$\sigma=0,31$
	C	$\mu=0,95$	$\mu=0,73$	$\mu=0,88$	$\mu=0,78$
		$\sigma=0,29$	$\sigma=0,24$	$\sigma=0,30$	$\sigma=0,25$

4.2. Resultados ante variaciones de las condiciones probabilísticas de falla. Con la finalidad de ver la efectividad del entrenamiento de los diferentes modelos, se realiza una segunda validación, a partir de la simulación de fallas bajo un escenario estadístico diferente al usado en el entrenamiento. En la Tabla 5 se pueden observar las estadísticas usadas para el nuevo escenario.

Tabla 5.
Datos estadísticos usados para la nueva validación.

Prob.	Rf	Prob. falla (fallas/año)	Tipo de falla	Total fallas simuladas
55%	$\mu=0,5$ $\sigma=0,1$		LT 68 %	
35%	$\mu=3$ $\sigma=1$	Nodos 0,2	LL 6%	500 años 162922 fallas
		Líneas 3	LLT 22%	
15%	$\mu=6$ $\sigma=2$		LLL/LLLT 4%	

La Tabla 6 recopila la media y la desviación estándar de los errores promedios de los 500 años del nuevo escenario para cada método por fase, de acuerdo a los años usados en el entrenamiento. En esta tabla se puede apreciar que a pesar de que en NN el error medio aumenta la estimación sigue siendo la mejor.

Tabla 6.

Errores medios y desviaciones estándar de la estimación del índice SARFI₉₀ nuevo escenario.

Métodos		5 años	10 años	15 años	20 años
		Error	Error	Error	Error
DT	A	$\mu=2,03$	$\mu=1,85$	$\mu=2,27$	$\mu=2,29$
		$\sigma=0,84$	$\sigma=0,76$	$\sigma=1,05$	$\sigma=1,05$
	B	$\mu=1,78$	$\mu=1,32$	$\mu=1,18$	$\mu=1,52$
		$\sigma=0,66$	$\sigma=0,57$	$\sigma=0,29$	$\sigma=0,73$
	C	$\mu=3,26$	$\mu=2,66$	$\mu=2,62$	$\mu=2,55$
		$\sigma=1,05$	$\sigma=0,98$	$\sigma=0,54$	$\sigma=0,91$
KNN	A	$\mu=2,66$	$\mu=2,73$	$\mu=2,48$	$\mu=2,46$
		$\sigma=1,07$	$\sigma=1,16$	$\sigma=0,93$	$\sigma=1,06$
	B	$\mu=1,32$	$\mu=1,26$	$\mu=1,21$	$\mu=1,20$
		$\sigma=0,52$	$\sigma=0,58$	$\sigma=1,07$	$\sigma=0,59$
	C	$\mu=2,95$	$\mu=2,69$	$\mu=2,73$	$\mu=2,55$
		$\sigma=0,99$	$\sigma=0,98$	$\sigma=0,58$	$\sigma=0,96$
SVM	A	$\mu=2,78$	$\mu=2,85$	$\mu=2,53$	$\mu=2,54$
		$\sigma=1,04$	$\sigma=1,08$	$\sigma=1,01$	$\sigma=1,02$
	B	$\mu=1,69$	$\mu=1,58$	$\mu=1,55$	$\mu=1,53$
		$\sigma=0,67$	$\sigma=0,69$	$\sigma=1,01$	$\sigma=0,68$
	C	$\mu=3,27$	$\mu=2,82$	$\mu=2,79$	$\mu=2,79$
		$\sigma=1,06$	$\sigma=0,95$	$\sigma=0,68$	$\sigma=0,94$

Continuación Tabla 6.

Errores medios y desviaciones estándar de la estimación del índice SARFI₉₀ nuevo escenario.

Métodos		5 años	10 años	15 años	20 años
		Error	Error	Error	Error
NN	A	$\mu=1,36$	$\mu=1,47$	$\mu=1,12$	$\mu=1,20$
		$\sigma=0,36$	$\sigma=0,42$	$\sigma=0,31$	$\sigma=0,33$
	B	$\mu=1,09$	$\mu=0,99$	$\mu=0,80$	$\mu=0,89$
		$\sigma=0,41$	$\sigma=0,34$	$\sigma=0,32$	$\sigma=0,30$
	C	$\mu=0,96$	$\mu=0,97$	$\mu=1,18$	$\mu=1,13$
		$\sigma=0,29$	$\sigma=0,31$	$\sigma=0,38$	$\sigma=0,33$

5. Conclusiones

Los resultados obtenidos evidencian el eficiente desempeño de las técnicas de minería de datos implementadas, como metodología para la estimación del índice SARFI₉₀ en nodos no monitorizados del sistema de distribución. Los resultados están lo suficientemente cerca al valor real, a pesar que solo se haya tenido en cuenta el número de neuronas para las NN y las capas de validación cruzada para DT, KNN, SVM.

Se demuestra que las NN obtuvieron un mejor comportamiento ante los dos escenarios de validación, pero, requieren gran cantidad de tiempo en su entrenamiento. Por otro lado, el método de transformación de datos permite el uso de técnicas que requieren menor capacidad computacional (DT, KNN, SVM), reduciendo significativamente el tiempo empleado en el entrenamiento y elevando el porcentaje de error.

Aunque en este proyecto de grado solo se tuvo en cuenta el índice SARFI₉₀, esta metodología se puede aplicar para otros índices de SARFI_x.

Referencias

- A. Asheibi, D. Stirling, and D. Robinson, "Identification of load power quality characteristics using data mining," *Can. Conf. Electr. Comput. Eng.*, no. May, pp. 157–162, 2007.
- A. Mohamed, H. Shareef, and H. Zayandehroodi, "A Review of Power Quality Monitor Placement Methods in Transmission and A Review of Power Quality Monitor Placement Methods in Transmission and Distribution Systems," no. May 2014, 2013.
- "ATPDraw Home." [Online]. Available: <https://www.atpdraw.net/>. [Accessed: 20-Oct-2019].
- B. A. Gonzalo, "Implementación del algoritmo de los k vecinos más cercanos (k-NN) y estimación del mejor valor local de k para su cálculo," p. 475, 2018.
- B. S. Jairo, "Estimación de las Variaciones Tensión de Corta Duración en Sistemas de Distribución en el Marco de las Redes Eléctricas Inteligentes," 2018.
- Blanco-Solano, J. Petit-Suárez, J. F. and Ordóñez-Plata, G. (2015). Methodology for relative location of voltage sag source using voltage measurements only | Metodología para la localización relativa de la fuente de hundimientos de tensión usando únicamente mediciones de tensión. *DYNA*, vol. 82, no. 192, 94–100.

Business Software Alliance, “¿Por qué son tan importantes los datos?,” *Resumen ejecutivo*, 2015. .

C. A. Ruiz and D. J. Matich, “Redes Neuronales: Conceptos Básicos y Aplicaciones.,” 2001.

C. S. E. J., “Tutorial sobre máquinas de vectores de soporte,” p. 25, 2014.

C. Q. R. Gironés Roig Jordi, Casas Roma Jordi, Minguillón Alfonso Julia, “Capítulo 13 Árboles de decisión,” in *Minería de datos modelos y algoritmos*, no. 2017, 2017, pp. 209–228.

Calderón, J. A. (2008). Neural network for fault classification in transmission lines based on osciloperturbography records, 99–107

Alazaidah, Raed. 2016. “Trending Challenges in Multi Label Classification.” 7(10): 127–31.

Antonio, Rodriguez Gonzales Juan. 1984. “ATP-EMTP.” : 357–411.

Asheibi, Ali, David Stirling, and Duane Robinson. 2007. “Identification of Load Power Quality Characteristics Using Data Mining.” *Canadian Conference on Electrical and Computer Engineering* (May): 157–62.

Barrero, Giovanni. 2019. “Optimización de Hiperparámetros de Algoritmos de Aprendizaje Automático Usados Para El Análisis de La Calidad Del Software.” (January).

Baughman, Martin L. 1921. “*IEEE PES Power & Energy Society , 'IEEE 123 Node Test Feeder,' Distribution Test Feeders.*” .

Blanco-Solano, J., J.F. Petit-Suárez, and G. Ordóñez-Plata. 2015. “Methodology for Relative

- Location of Voltage Sag Source Using Voltage Measurements Only | Metodología Para La Localización Relativa de La Fuente de Hundimientos de Tensión Usando Únicamente Mediciones de Tensión.” *DYNA (Colombia)* 82(192): 94–100.
- Bramer, Max. 2016. *Principles of Data Mining*. Springer-Verlag London.
- Business Software Alliance. 2015. “¿Por Qué Son Tan Importantes Los Datos?” *Resumen ejecutivo*: 40.
- Calderón, Jhon Albeiro. 2008. “Neural Network for Fault Classification in Transmission Lines Based on Osciloperturbography Records.” : 99–107.
- Charte, Francisco. 2015. “Nuevos Métodos Híbridos de Computación Flexible Para Clasificación Multietiqueta.”
- Chen, Tiantian, Ling Luo, and Xiang Shen. 2018. “An Analytical Method of Data Mining on Voltage Sag Based on Gray Target Theory and Cloud Model.” *2018 China International Conference on Electricity Distribution (CICED)* (201804250000038): 636–43.
- Committee, Distribution, Ieee Power, and Energy Society. 2009. 2009 *IEEE Std 1159TM-2009, IEEE Recommended Practice for Monitoring Electric Power Quality*.
- Framework, Economic, and For Power. 2011. “Economic Framework For Power Quality.” (June).
- Ghodsi, Mansi. 2016. “A Brief Review of Recent Data Mining Applications in the Energy Industry.” (June).
- Gironés Roig Jordi, Casas Roma Jordi, Minguillón Alfonso Julia, Cihuelas Quiles Ramon. 2017. “Capítulo 13 Árboles de Decisión.” In *Minería de Datos Modelos y Algoritmos*, , 209–28.
- González, Ramón Hilera, Victor José, Martínez Hernando, and Victor Martinez Hernando. 1995. “Redes Neuronales Artificiales : Fundamentos , Modelos y Aplicaciones / José.” (January

2017).

Gonzalo, Berástegui Arbeloa. 2018. “Implementación Del Algoritmo de Los k Vecinos Más Cercanos (k-NN) y Estimación Del Mejor Valor Local de k Para Su Cálculo.” : 475.

Han, Jiawei; Kamber, Micheline; Pei, Jian. 2012. *Data Mining: Concepts and Techniques*. Third Edit. Morgan Kaufmann Publishers.

Heo, Seongmin et al. “ScienceDirect.” *IFAC-PapersOnLine* 51(18): 470–75.
<https://doi.org/10.1016/j.ifacol.2018.09.380>.

Hurtado-cortés, Luini Leonardo, Edwin Villarreal-lópez, and Luís Villarreal-lópez. 2016. “Fault Detection and Diagnosis through Artificial Intelligence Techniques , a State of Art □ Artificial ,.” 83(December): 19–28.

J., Carmona Suárez Enrique. 2014. “Tutorial Sobre Máquinas de Vectores de Soporte.” : 25.

Jairo, Blanco Solano. 2018. “Estimación de Las Variaciones Tensión de Corta Duración En Sistemas de Distribución En El Marco de Las Redes Eléctricas Inteligentes.”

José, Salinas Ortiz. 1993. *Análisis Estadístico Para La Toma de Decisiones En Administración y Economía*.

Kiranmai, S Asha, and A Jaya Laxmi. 2018. “Data Mining for Classification of Power Quality Problems Using WEKA and the Effect of Attributes on Classification Accuracy.”

Lin, Shunfu et al. 2016a. “Monitoring Data Analysis.” *2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA)* (51207088): 338–42.

———. 2016b. “The Data Mining Application in the Power Quality Monitoring Data Analysis.” *2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA)* (51207088): 338–42.

Mohamed, Azah, Hussain Shareef, and Hadi Zayandehroodi. 2013. “A Review of Power Quality

- Monitor Placement Methods in Transmission and A Review of Power Quality Monitor Placement Methods in Transmission and Distribution Systems.” (May 2014).
- Molina, José ;García, Jesús. 2006. “Técnicas de Análisis de Datos,Aplicaciones Prácticas Utilizando Microsoft Excel y Weka.”
- Nourollah, Sara, and Mehdi Moallem. 2011. “A Data Mining Method for Obtaining Global Power Quality Index.” *2011 2nd International Conference on Electric Power and Energy Conversion Systems, EPECS 2011*: 1–7.
- Romero, M, L Gallego, and A Pavas. 2011. “Estimation of Voltage Sags Patterns with K-Means Algorithm and Clustering of Fault Zones in High and Medium Voltage Grids.” *Ingenieria e Investigacion* 31(2 SUPPL.): 131–38. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-82155166493&partnerID=40&md5=21335744b6492a5f024ccda00ed5e115>.
- Ruiz, Carlos Alberto, and Damián Jorge Matich. 2001. “Redes Neuronales: Conceptos Básicos y Aplicaciones.”
- Sikorski, Tomasz. 2016. “Cluster Analisis of Long-Term Power Quality Data.” : 0–5.
- Sikorski, Tomasz, and Klaudiusz Borkowski. 2019. “Clustering as a Tool to Support the Assessment of Power Quality in Electrical Power Networks with Distributed Generation in the Mining Industry.” 166(September 2018): 52–60.
- Targosz, Roman, and David Chapman. 2015. “Application Note the Cost of Poor Power Quality.” (October).
- Telecomunicación, Redes De, Cristina García Cambroner, and Irene Gómez Moreno. “Algoritmos de Aprendizaje: Knn & Kmeans.”
- Tsoumakas, Grigorios, and Ioannis Katakis. 2009. “Multi-Label Classification : An Overview Multi-Label Classification : An Overview.” (September).

- Upadhyaya, Swarnabala, and Sanjeeb Mohanty. 2016. “Localization and Classification of Power Quality Disturbances Using Maximal Overlap Discrete Wavelet Transform and Data Mining Based Classifiers.” *IFAC-PapersOnLine* 49(1): 437–42. <http://dx.doi.org/10.1016/j.ifacol.2016.03.093>.
- Velasco, Juan A Martínez, and Jacinto Martín Arnedo. 2008. “Aplicaciones Informáticas Para Análisis de Huecos de Tensión.”
- Yinghui Kong, Jinsha Yuan, Linlin Che and Jing An. 2007. “‘Online Power Quality Disturbances Identification Based on Data Stream Technologies,’ 2007 International Power Engineering Conference (IPEC 2007), Singapore, 2007, Pp. 585-590.” : 585–90.
- Velasco, J. A. M. and Arnedo, J. M. (2008). Aplicaciones informáticas para análisis de huecos de tensión.